THREE ESSAYS IN EMPIRICAL ECONOMICS: DATA-INTENSIVE METHODS APPLIED TO MEDIA ECONOMICS AND EXECUTIVE COMPENSATION

by

Justin T Briggs
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Economics

Committee:

________________________ Director

____

____

________________________ Department Chairperson

________________________ Program Director

________________________ Dean, College of Humanities and Social Sciences

Date: ____________________ Summer Semester 2019
George Mason University
Fairfax, VA

Three Essays in Empirical Economics: Data-Intensive Methods Applied to Media Economics and Executive Compensation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Justin T Briggs
Master of Arts
George Mason University, 2010
Bachelor of Science
University of Maryland, 2001

Director: Dr. Alex T Tabarrok, Professor
Department of Economics

Summer Semester 2019
George Mason University
Fairfax, VA

Dedication

I dedicate this work to serious journalists everywhere who are striving to report the news neutrally and in depth, and whose efforts may enhance the quality of news coverage to come.

Acknowledgments

I would like to express my gratitude to all those thanks to whom this solitary exercise has been transformed into a hopefully stimulating body of work. To start with I would like to thank my advisor for the encouragement and advice he gave over the course of my study, as well as for his flexibility when I was studying part-time or remotely. In the course of an enjoyable collaboration on a joint paper, I learned a lot about what it takes to produce quality work. I was also struck by the great example he set in communicating the results and explaining the work to a broader public, something I hope to be able to emulate. I also appreciate the feedback of Tim Groseclose especially on the media work. His deep experience in the subject and caring manner have been much appreciated. Finally, the first paper has benefited from the feedback of multiple anonymous reviewers. Naturally, all remaining flaws and shortcomings are mine alone.

Thanks also go to Muriel for her proofreading and incisive critiques, and also for her forbearance in indulging my studies over these years; to my parents and sisters for their support and for those not-always-subtle nudges to get a move on; to Mary Jackson at the Department of Economics for her invaluable support in navigating the program and its requirements; and to my management and colleagues at Leidos for their flexibility in accomodating my studies. I would also like to thank Leidos for their financial support during my Masters, and the Center for Study of Public Choice for a fellowship supporting some of my dissertation research.

Finally, I would like to thank the Internet Archive for generously providing access to their repository of television recordings and closed-caption data, without which it can safely be said that this dissertation would not have been possible.

Table of Contents

List of Tables

List of Figures

Abstract

THREE ESSAYS IN EMPIRICAL ECONOMICS: DATA-INTENSIVE METHODS APPLIED TO MEDIA ECONOMICS AND EXECUTIVE COMPENSATION

Justin T Briggs, PhD

George Mason University, 2019

Dissertation Director: Dr. Alex T Tabarrok

This dissertation consists of three studies in empirical economics. The first two examine the content of television and radio news coverage using computer methods. The third uses data on public companies to examine executive compensation as a function of the companies' ownership structure and concentration.

The first paper studies the political slant of news content from all nationwide American television and radio news networks in a systematic way, including how the slant changes over time. Analyzing 270,000 news programs and nearly 1 billion phrases over the six years from 2010 through 2015, the study performs an objective computational analysis which uses the Congressional Record to learn how language usage varies with ideology. This language usage pattern is applied to each network's complete coverage (as represented by transcripts) in order to obtain a "slant index" of the news content. In terms of ordering, the slant index largely accords with public perceptions. For example, the Fox networks stand apart from all other networks as the least centrist, having the most conservative slants. However despite perceptions, all national television and radio networks are much more centrist (or balanced) than congressmen. Although there is evidence they may be becoming more ideologically diverse over time, it finds no evidence that any major networks report news with a strictly partisan slant. Moreover, all

stations respond similarly to events, indicating that it is more important for politicians to get their story in the news than it is to get it on any particular station.

The second paper examines other qualities of U.S. television and radio news, over the period 2010-2016. These get to the heart of whether the media adequately fulfill the role of informing voters in a democracy, including the topics being discussed, whether coverage is fact- or opinion- based, foreign or domestic, negative or positive, and/or emotionally charged. In an unprecendented effort, it uses cutting edge deep learning techniques to evaluate these aspects of quality at a minute-by-minute level, providing in aggregate a fine-grained summary of what the networks cover and how they cover it, plus how this changes over time. It finds that there is over twice as much negative coverage as positive coverage, and scary, shocking or outrageous coverage as pleasant coverage. The majority of cable networks' coverage is opinion, while the broadcast networks present more fact-based content. Importantly for democratic elections, only 20 percent of electoral coverage, accounting for 2.2 percent of total coverage, is on candidates' backgrounds, platforms, and speeches.

The third paper explores the relationship between company ownership concentration and executive compensation packages. It investigates the idea that, other things being equal, we should see companies with more concentrated ownership structures choose more closely optimal compensation contracts for their executives. It finds that compensation depends upon the identity of the owners as well as their concentration. Institutional ownership, by professional investment managers, is associated with much higher executive incentive pay, whereas outside ownership by others is associated with much lower pay. It points to a new hypothesis in executive compensation, whereby the principal-agent problem between investors and investment managers may be allowing higher pay packages among CEOs as an external effect.

Chapter 1: Television and Radio News Slants in America

1.1 Introduction

Trust in the media is at an all-time low. After declining steadily since the 1970s, 68% of Americans now have little or no trust in the Fourth Estate (Gallup, 2016). The Pew Research Center (2014c) finds that partisans "at both ends of the ideological spectrum stand out for their high levels of distrust in individual sources," although they disagree about which sources not to trust. Much of this mistrust seems to be driven by a belief that media are slanted against them: in a recent MRC/YouGov poll, nearly 80% of voters believed that news coverage was biased during the 2016 election. Television and radio are among the most common sources of news for Americans: 57% report they often get news from television, and 25% from radio (Pew Research Center, 2016). This compares with 38% for digital sources and 20% for print newspapers. Yet cable and broadcast mediums are among the least studied, largely because of the difficulty in obtaining and analyzing their content. This paper addresses this gap by systematically analyzing the slant of all national television and radio news networks in the United States.

Let us first recognize how influential these networks are, and how important such news sources are in a democracy like the United States. In 2013, the average network evening news broadcast had nearly 23 million viewers. No other single source comes close to reaching that many citizens.[1] Democracies depend in great part upon their news media, broadly construed, to provide low-cost information to voters about choice-relevant factors beyond their direct knowledge (Wittman, 1997). However the quality of news providers – what "news" is covered, how it is covered – can vary greatly (Djankov *et al.*, 2001; Gentzkow *et al.*, 2006; McMillan & Zoido,

[1]The top 25 newspapers *combined* had an average weekday circulation totaling 14.3 million for both print and digital (Alliance for Audited Media, 2013). Cable news average prime-time viewership was just under 3 million while NPR's top-rated 'All Things Considered' evening show reached 13 million on average (Pew Research Center, 2016).

2004). Moreover, exposure to television and radio is known to affect personal behavior (Jensen & Oster, 2009; Kearney & Levine, 2014; La Ferrara *et al.*, 2012), and news exposure more specifically can impact both voter turnout and how people actually vote (DellaVigna & Kaplan, 2007; Gentzkow, 2006; Gerber *et al.*, 2006), and consequently also public policy more generally (Stromberg, 2004). Clearly, what people see in the news affects how they view the world, and recognition of such probably underlies much current angst over the state of the media.

Broadly, this paper studies the aspect of media quality generally known as bias, or more specifically, *political slant*. It examines the content of all national television news providers, as well as the only national radio news provider in the United States, over the six year period from 2010 to 2015. The political slant of a news outlet is identified by how their pattern of language use compares with that of congressmen having known ideologies along a spectrum from liberal to conservative. This 'slant index' measurement can capture various forms of slant to at least some extent, including selective coverage and non-neutral framing and presentation.

Even assuming that news providers are slanted, if not all are slanted in the same direction, then a Bayesian consumer who is seeking the truth should be able to combine information from multiple sources in order to "triangulate" on reality – bias in the media would have little impact. But these caveats are important because there is evidence that viewers[2] do not generally seek the truth (they may be more interested in being entertained, e.g.), and that they often may choose to follow only one source (Pew Research Center, 2014a; Prior, 2013). Taken to an extreme, selective exposure to news sources can contribute to an 'echo-chamber' effect as noted by Sunstein (2001) and others, as people are exposed chiefly to information that comports with and/or does not challenge their preexisting world-view. But even if exposed to a wide variety of news, viewers may have difficulty in accurately assessing the slant of any particular story they encounter (Vallone *et al.*, 1985). Moreover, viewers may not act strictly rationally in assessing the news they consume (Bartels, 2002). Since holding false opinions about political issues may have little to no cost (and even be utility-enhancing (Caplan, 2007)), there is much evidence that individuals process political information using what Kahneman (2011) calls 'system one'

[2]"Viewers" is used without loss of generality to radio listeners.

mental processes (i.e., using heuristics and subject to various biases), or even using motivated reasoning where one of the goals in interpreting new evidence is to reach a conclusion that supports one's existing beliefs.

While the social costs of slanted media could be significant, the private decisions of news companies need not consider them. Since most are for-profit entities, the profit motive ensures that news companies attempt to maximize the engagement of viewers however they can in order to sustainably earn revenue. Given that they face competition for viewers' time and attention, it may seem logical for news providers to differentiate themselves on the basis of political slant (as Mullainathan & Shleifer (2005) predicts). But there are models that predict instead that under competition, concern for reputation could discipline bias so that they would all cluster together, assuming consumers want and expect truth from their news providers (Gentzkow & Shapiro, 2006). Indeed, every news outlet we study insists that they report the unbiased truth, or at least a balanced representation of it. Moreover, although many viewers perceive vast differences in the slants of different outlets (and some believe only one or no outlets to be "unbiased"), psychological research indicates this perception may itself be a product of their own biases (Comstock & Scharrer, 2005; Schmitt *et al.*, 2004; Vallone *et al.*, 1985). My results will suggest that there may be an element of truth in both theories.

Some go further than claim that news providers are slanted. They imply that willfully or not, they may be actively engaged in helping or hurting the prospects of one party or set of candidates, and may even be acting on behalf of one of the parties (Levendusky, 2013). Such a partisan press certainly did exist in the American past – early 19th century newspapers were often overtly affiliated with parties or politicians and did their best to back them (Gentzkow *et al.*, 2006). As noted previously, all of the national television and radio stations would ardently deny any such affiliation today. But it's possible they may have an underlying partisan preference and thus be *de facto* partisan, by under-reporting or dismissing news that is unfavorable to their preferred party or by excessively focusing on criticism of the other party while depicting their preferred party as superior. Americans seem widely to believe that at least some news providers are "closet partisans." Under such a theory, all news would be framed from the perspective of how best to help their preferred party or hurt the other. We will explore this possibility as well.

Contrary to many dire warnings, I will show that extant national television and radio news are much less extreme in their slant than most members of Congress. Nonetheless, there is clear differentiation among providers, and the entry of the Fox Networks has added a significantly more conservative angle than others.[3] Moreover, I find that polarization, or the gap between the slants of more liberal and more conservative outlets, increases modestly over the study period: stations are becoming more differentiated. However, I also explore two possible implications of the theory that networks are closet partisans, and find no evidence for them.

Although there are some other studies of television news slant using indirect or subjective methods[4], this paper presents the first systematic assessment of the slant of all national television and radio news providers based upon comprehensive news coverage, and assess changes over time.[5] It follows in the tradition of objective comprehensive analysis pioneered by Groseclose & Milyo (2005), which use citations of think tanks to proxy for slant. My method owes directly to Gentzkow & Shapiro (2010), which determines slant by analyzing newspaper text.

The paper will proceed in the next section to lay out the data and methodology used to assess slant. Section 1.3 will cover results, first on a summary for the entire study period, and then, after validating results, on how slant varies over time. A closing discussion follows.

[3]This is on the basis of their slants during the study period; I have no data with which to assess the slant of incumbent networks prior to Fox's entry.

[4]For example, there have been various efforts in the field of Journalism to assess the slant of media, but they tend to suffer from the need for human interpretation, and the necessarily limited scope this entails. Consequently results are fairly contradictory. The weaknesses of such methods are epitomized by various media watchdog groups that (while not purporting to be academic) each "claim to objectively analyze media content yet who routinely disagree on the incidence, severity and direction of bias in the media." (Baum & Groeling, 2008)

[5]Martin & Yurukoglu (2017) perform an impressive assessment of the impact of news slant, using similar methods as here, but limit their reach to three cable news networks. Castillo *et al.* (2013) also perform a computational linguistic analysis of television news during a subset of the study period. Although it is focused on other aspects of quality besides slant, they do assess the sentiment of statements relating to Barack Obama and Mitt Romney during the 2012 election campaign and find that while MSNBC and CNN are equally positive in discussing both, Fox News, and particularly Fox Business Channel are more positive when discussing Romney (other outlets are not covered). This is consistent with my findings of the partisan ordering of networks, although their methodology is quite different.

1.2 Data and Methodology

This paper analyzes a unique data set from the Internet Archive[6] which contains closed captions of television news covering the six full years from 2010 to 2015. It replicates from scratch, and then adapts, a methodology pioneered by Gentzkow & Shapiro (2010) for finding the ideological slants of newspapers; this is the first such attempt at replication. Briefly, the methodology learns what partisan political speech looks like by analyzing the Congressional Record. This includes the full text of all congressmens' speeches on the floor of Congress. In addition, I leverage work of political scientists to quantify the position of each congressman on the liberal - conservative ideological spectrum, and associate each congressman's speech with their known ideology. The idea is that partisans may deliberately use different language to talk about the same topic, such as "estate tax" versus "death tax" or "global war on terror" versus "war in Iraq," and indeed there is much evidence of partisan media strategists encouraging their party affiliates to use or not use phrases in order to gain partisan advantage. Frank Luntz is the prototypical modern example of such a word-smith, having invented the phrase "death tax" and using focus groups to hone messaging, but the practice is now commonplace (Friedenberg, 1997; PBS Frontline, 2004). In addition to capturing different ways of speaking about the same topic, the methodology will also identify the extent to which partisans may prefer to talk about different topics, with liberals relatively more likely to speak about "income inequality" and conservatives about "tort reform," to give arbitrary examples.

Once distinctively liberal, neutral, or conservative phraseologies have been identified, I analyze the language used by different news outlets. The composition of phrase usage at an outlet indicates the ideological *slant* of the outlet in question. This slant index, the primary result of this paper, can be directly compared with the ideologies of congressmen.

[6] `https://archive.org/details/tv`

1.2.1 Defining politically diagnostic language from congressional speech

The first step is to learn what comprises politically diagnostic language. To this end, the entire Congressional Record over the study period was assembled by collating the daily speech records for both the Senate and the House of Representatives, available from the Government Printing Office website.[7] This was then parsed to ignore non-speech text (including editorial content like titles, notes, and names of bills and other context like the results of votes, the page number, time and date), separate the Record into speech segments spoken by individual congressmen, and eliminate all other speech (such as by the clerk, chaplain, delegates, former congressmen, or congressmen performing formal roles like the speaker protempore). Parenthetical quotes such as the text of a bill also were removed from the speeches, as these were normally not spoken and instead were inserted into the record. The remaining text contained nearly 70 million words.

All speeches by each congressman were then compiled. A congressman's complete speech record was then separated into individual sentences. In each sentence, non-words (such as numbers and punctuation) and stopwords were removed. These stopwords are a concrete list of low-information-content words, and consisted of those identified by Fox (1989), including e.g. 'of', 'the', 'as', 'did', as well as those identified by Gentzkow & Shapiro (days of the week and names of congressional office buildings) and myself (party names and congressional organizational roles and units - 'colleague', 'speaker', 'caucus', 'committee', etc.) as not useful in this context. In order to reduce unnecessary duplication and consolidate the variations of a word, remaining words were stemmed into their roots using the Porter (1980) stemmer as implemented in NLTK, the Natural Language Toolkit (Bird *et al.*, 2009). For example: abate, abated, abatement, abatements, abates all stem to 'abat'. Finally, usage of all pairs and triples of word stems (2- and 3-grams, together referred to as *n-grams*) are compiled and counted, to produce a statistical snapshot of how each congressman speaks on the floor of Congress. In total, 4.6 million 2-grams and 12.8 million 3-grams were used at least once in the 3.6 million sentences spoken by congressmen during the six-year study period.

Of these, there are many n-grams (which I sometimes refer to as phrases for convenience)

[7] `http://www.gpo.gov`

which are either procedural language specific to Congress, or which are in very common language usage, and which do not indicate differing ideological positions of parties in Congress. Even though they are not politically meaningful, it is still possible that one party uses them more often than the other (as a consequence of being in the majority/minority, e.g., or, in the tail, even just by chance), and they are computationally costly to process. To exclude phrases of either type from further consideration, we additionally require the number of corresponding phrase usages in the news to be above some minimum but not greater than some maximum (to handle the respective cases). The exact cutoffs correspond to the quantity of data available for each type of analysis (summary or monthly), and the fact that there will be more examples of unique 2-grams than unique 3-grams. They were chosen with the goal of roughly exclude ideologically insignificant phrases while keeping ideologically significant ones; but the results are not overly sensitive to their precise values. They are listed in Table 1.2.

To uncover what n-grams are diagnostic of ideological speech, we are effectively looking for phrases that are often used by one party but rarely by the other (speech by Independents is ignored). First any n-grams that are used only once are dropped. Then, for each phrase i, a score is given of the following form, proportional to a χ^2 statistic to test the independence of phrase frequency by party:

$$\chi_i^2 \propto \frac{(f_{ir}f_{\sim id} - f_{id}f_{\sim ir})^2}{(f_{ir}+f_{id})(f_{\sim ir}+f_{\sim id})(f_{ir}+f_{\sim ir})(f_{id}+f_{\sim id})}$$

where for the parties $p \in \{r, d\}$, f_{ip} is the number of times phrase i was used by the party p, and $f_{\sim ip}$ is the number of times any other n-gram of the same size was spoken by party p. The numerator measures the extent to which phrase i is used disproportionately by one party, while the denominator normalizes this by a product of the total counts for each case: candidate phrase usage, non-candidate phrase usage, phrase usage by Republicans and by Democrats. In addition, a Bayesian smoothing factor is added to prevent incorrectly inferring partisan usage for infrequently used phrases. This ensures that both parties have used a given phrase at least twice in the equation. For the entire period, 53.4% of 2-grams and 53.5% of 3-grams were used more often by Democrats, reflecting the amount that Democrats spoke relative to Republicans. Over the period, Democrats controlled the House for one year (2010) and Republicans controlled it for the other five. By contrast, Republicans controlled the Senate for one year (2015) and Democrats for the other five.

Intuitively, a phrase with a higher χ^2 score is more likely to be ideologically partisan. Those phrases with the n highest scores are saved as politically *diagnostic phrases* and are what will be used to identify politically slanted language, as described below[8]. Top diagnostic 2- and 3-grams used by both parties are shown in Table 1.1, split by which party used them more frequently. To give an example, `'job, creator'` has one of the highest scores among republican-leaning phrases. It was used 2,334 times in Congress, with 1,934 of those usages by Republicans.[9] By contrast, one of the lowest scoring Republican-leaning 2-grams that still qualifies as diagnostic is `'price, ga'` (which comprises the actual usages "price of gas," "price at the gas," "price of that gas," "price of their gas," "prices, higher gas," "prices at a time when gas," and "prices for gas"), used 266 times, 169 of those by Republicans.[10] Correspondingly for Democrats, one of the highest scores is for `'clean,energi'`, while one of the lowest scores of a candidate phrase is `'opportun,peopl'`. While being used disproportionately by one party is indicative of ideology, I identify a phrase's ideological content more precisely as described below, after first introducing the news data.

1.2.2 News Content

News transcripts were obtained for all available television networks with nationally syndicated news, as well as the sole national radio network. The thirteen networks covered, which include all major national news programs, are: ABC, CBS, NBC, PBS, BBC, CNN, Fox News Channel (FOXNEWS), MSNBC, Al Jazeera America (ALJAZAM), CNBC, Fox Business (FBC), Bloomberg

[8]For the summary data, the actual p-value corresponding to the count cutoff for diagnostic phrases is $p <= 0.001$ for both 2-grams and 3-grams. The values of n vary by n-gram type and time period and are given in Table 1.2.

[9]An example usage comes from Republican Senator Mitch McConnell: "That is why Republicans have continued to press for policies, policies that empower job creators, not Washington."

[10]Example usages come from Republican Senator Kirk: "The cost of oil and the price of gas is high enough." and Democratic Senator Klobuchar: "We need to work toward a pragmatic solution that reforms the ethanol industry without harming jobs or driving up gas prices at a time when gas is over $3.70 a gallon."

Table 1.1: Top Diagnostic Phrases by χ^2 Score

Top Republican 2-gram stems

'presid,health', 'job,creator', 'care,bill', 'american,energi', 'trillion,debt', 'budget,amend', 'nation,debt', 'radic,islam', 'illeg,alien', 'employ,mandat', 'presid,promis', 'live,mean', 'war,terror', 'religi,freedom', 'presid,own', 'death,tax', 'feder,spend', 'enforc,law', 'presid,budget', 'energi,product', 'increas,spend', 'care,plan', 'stimulu,bill', 'keyston,xl', 'spend,trillion', 'global,war', 'free,enterpris', 'immigr,law', 'individu,mandat', 'red,tape', 'spend,debt', 'spend,control', 'repeal,obamacar', 'reduc,spend', 'medic,devic', 'xl,pipelin', 'increas,tax', 'washington,spend', 'medicar,advantag', 'debt,crisi', 'regulatori,burden', 'radic,islamist', 'rule,law', 'missil,defens', 'religi,liberti', 'kill,american', 'rule,regul', 'broken,promis', 'spend,reduct', 'children,grandchildren', 'spend,borrow', 'debt,trillion', 'train,wreck', 'energi,price', 'taxpay,dollar', 'govern,takeov', 'entitl,program', 'takeov,health', 'lower,cost', 'washington,democrat'

Top Democratic 2-gram stems

'unemploy,insur', 'clean,energi', 'public,health', 'unemploy,benefit', 'head,start', 'recoveri,act', 'comprehens,immigr', 'street,reform', 'african,american', 'pell,grant', 'reduc,deficit', 'prescript,drug', 'tax,loophol', 'tar,sand', 'domest,violenc', 'equal,pay', 'congression,black', 'women,health', 'preexist,condit', 'loan,debt', 'invest,infrastructur', 'sea,level', 'repeal,afford', 'wealthiest,american', 'violenc,women', 'invest,educ', 'extend,unemploy', 'district,court', 'fault,own', 'rais,minimum', 'dream,act', 'class,famili', 'citizen,unit', 'peopl,disabl', 'republican,senat', 'republican,refus', 'right,act', 'lost,job', 'prevent,care', 'food,tabl', 'program,help', 'student,debt', 'millionair,billionair', 'job,oversea', 'energi,effici', 'middle-class,famili', 'hope,republican', 'ryan,budget', 'koch,brother', 'women,act', 'level,rise', 'fair,shot', 'women,famili', 'clean,air', 'shut,govern', 'card,compani', 'deficit,reduct', 'protect,public', 'ship,job', 'extrem,weather'

Top Republican 3-gram stems

'presid,health,care', 'balanc,budget,amend', 'global,war,terror', 'keyston,xl,pipelin', 'health,save,account', 'promis,american,peopl', 'takeov,health,care', 'medic,devic,tax', 'feder,govern,spend', 'govern,takeov,health', 'health,human,servic', 'intern,revenu,servic', 'obama,health,care', 'borrow,cent,dollar', 'pass,balanc,budget', 'fanni,mae,freddi', 'mae,freddi,mac', 'listen,american,peopl', 'budget,amend,constitut', 'cent,dollar,spend', 'rais,debt,limit', 'largest,tax,increas', 'free,enterpris,system', 'cut,cap,balanc', 'presid,obama,promis', 'trillion,tax,increas', 'enforc,immigr,law', 'trillion,nation,debt', 'nation,debt,trillion', 'feder,independ,busi', 'nation,feder,independ', 'corpor,tax,rate', 'rais,tax,trillion', 'massiv,tax,increas', 'oil,natur,ga', 'secretari,health,human', 'tax,increas,histori', 'billion,tax,increas', 'american,peopl,understand', 'khalid,sheikh,moham', 'stop,spend,money', 'pick,winner,loser', 'tax,medic,devic', 'delay,employ,mandat', 'debt,limit,increas', 'tax,american,peopl', 'taxpay,fund,abort', 'secretari,homeland,secur', 'govern,spend,money', 'congress,american,peopl'

Top Democratic 3-gram stems

'comprehens,immigr,reform', 'wall,street,reform', 'student,loan,debt', 'repeal,afford,care', 'violenc,women,act', 'sea,level,rise', 'tax,cut,wealthiest', 'american,job,act', 'broken,immigr,system', 'ship,job,oversea', 'credit,card,compani', 'close,tax,loophol', 'equal,pay,equal', 'tax,cut,wealthi', 'extend,unemploy,insur', 'pay,fair,share', 'nation,institut,health', 'tea,parti,republican', 'class,tax,cut', 'extend,unemploy,benefit', 'women,health,care', 'middl,class,tax', 'tar,sand,oil', 'immigr,reform,bill', 'tax,cut,millionair', 'pass,comprehens,immigr', 'climat,chang,real', 'afford,health,care', 'student,loan,rate', 'cut,social,secur', 'strengthen,middl,class', 'prevent,health,care', 'impact,climat,chang', 'faith,credit,unit', 'tax,break,wealthiest', 'citizen,unit,decis', 'tax,break,millionair', 'effect,climat,chang', 'cut,middl,class', 'fix,broken,immigr', 'credit,default,swap', 'increas,minimum,wage', 'extrem,weather,event', 'head,start,program', 'district,court,judg', 'bring,job,home', 'address,climat,chang', 'center,diseas,control', 'wall,street,bank', 'tax,cut,middl'

Business, and National Public Radio (NPR). CNBC, Fox Business, Bloomberg, and Al Jazeera were not available for the whole period and estimates are provided for the period for which they are available.[11] Television transcripts came from the Internet Archive, which has recorded closed captioning (CC) data of news programs from affiliate stations in several east and west coast cities.[12] Local news programs, news programs not produced or syndicated by the network, and non-news programs (including comedy / satire) are dropped[13], as are programs which are primarily the content of politicians themselves: The Party Conventions, Inauguration, State of the Union address, and Debates. The remaining 270,442 news programs which were analyzed broadly include traditional and morning news programs, news/opinion programs including interviews, and long-format news and investigative journalism. In addition to television CC data, story transcripts were obtained from National Public Radio's website for the two nationwide programs with the largest audiences.[14] In total 69,410 NPR stories were analyzed, with between 12 and 22 stories generally being in a program.

Closed captions for television are produced for the use of hearing impaired viewers, and generally are transcribed in real time for news programs. As such they have several issues not found in print media, NPR transcripts, or the Congressional Record: notably, a number of grammatical and spelling errors. These may increase measurement noise, however there is no reason that these should have a differential effect on statistical slant estimates. Attempts to automatically correct spelling errors led to no significant difference in estimates from the original within a subsample group, and are computationally costly (also fully automated approaches are themselves highly error-prone), so they were abandoned. In addition, although most television ads do not contain CCs, some do, and they have not been removed from the data. An experiment

[11]To supplement sparse data for BBC and Al Jazeera, some programs which were broadcast on local stations (such as PBS affiliates or LINKTV) rather than their own dedicated network were attributed to the network.

[12]To avoid data issues which arose during the data capture, and to extend the time series, the local affiliates I used in the analysis changed over time. Prior to 2011, Washington, DC stations WETA, WJLA, WUSA, and WRC were used for PBS, ABC, CBS, and NBC respectively. For 2011 and after, San Francisco, CA stations KQED, KGO, KPIX, and KNTV were used.

[13]It is still possible that a small amount of local programming may sneak into the data set, particularly at the beginnings or ends of transcripts, and as local news blurbs, but also in case of unforeseen scheduling changes.

[14]The morning and afternoon programs "Morning Edition" and "All Things Considered" were scraped from `http://www.npr.org/programs`, but do not contain the hourly 5-minute news brief (which is unavailable), or any local programming which might be inserted by local affiliates.

was made attempting to automatically detect and remove ads, which succeeded in removing a significant portion of ads, but with some unintentional removal of non-ads and an unknown quantity of ads not detected or removed. Since overall results were little changed in this experiment, automatic ad detection and wiping was forgone in order to avoid the risk of imparting bias. But it's not surprising that ads proved unimportant to slant estimates for the stations, since ads rarely use political speech, and in any case represent a small quantity of text compared to that of the news programs.

Processing of both types of transcript text is similar to that of the Congressional Record. Transcripts were processed to remove extraneous characters, speaker headers, and interjections such as "(LAUGHTER)," and parsed into separate speakers and sentences. They were further split into words, stopwords were removed, words were stemmed, and 2- and 3-grams were constructed. This yielded nearly 1 billion phrases for analysis, with an average of 7.4 million 2-grams and 5.8 million 3-grams per month, and with each unique phrase usage mapped to the respective news outlets.

1.2.3 Estimating Slant

Once we have a set of disproportionately partisan *diagnostic n-grams* derived from Congress, we wish to quantify their *ideological hues*, which will tell us how indicative of ideology each phrase is. To do so, we first obtain the estimated ideology for each member of Congress in each given year from one of two outside sources discussed below.[15] Then, for each phrase i, the relative frequency that a congressman j used it is regressed on the congressman's ideology, to obtain a beta (*Phrase Hue*) and an intercept:

$$\mathbf{PhraseUsageFreq}_{ij} = Intercept_i + PhraseHue_i \mathbf{Ideology}_j + \epsilon_{ij}, \quad \forall i$$

[15]In some cases, a congressman's ideology is not available for a given year – typically because they served a partial term – in which case the congressman's speech is ignored for that year. Occasionally, this leads a phrase to have less than the minimum number of usages, in which case the phrase is also dropped for the period.

Each such regression of congressmen is estimated using a weighted least squares (WLS) estimation, with weights which are the square root of the number of sentences spoken by each congressman in Congress. WLS is the best linear unbiased estimator, more efficient than OLS, under the assumption that uncertainties of the n-gram usage frequencies vary by congressman, because the variance of the error is correlated with these uncertainties. To think of it another way, WLS is the best estimator if we consider "speech in Congress" as the unit of analysis, with speech being grouped unevenly by speaker. This regression yields a value $PhraseHue_i$ for the ideological hue of each phrase, as well as a conditional average phrase usage frequency $Intercept_i$.

Finally, to find the *speech-implied slants* of television stations as measured by this method, the stations' relative phrase frequencies are regressed on the ideological hues of the phrases which we just obtained. That is, for each station k, we regress the station's phrase usage frequency, as adjusted by that phrase's overall usage frequency as found in the first stage, on the phrase hue:

$$(\mathbf{StationUsageFreq}_{ik} - \mathbf{Intercept}_i) = SlantIndex_k \mathbf{PhraseHue}_i + \epsilon_{ik}, \quad \forall k$$

For similar reasons to before, this is also estimated using WLS, using the reciprocal of the (heteroskedasticity-robust) standard error from the prior regression as weights. This results in our estimate of a station's slant: $SlantIndex_i$, on a scale comparable to the input congressional ideology.

Prior to conducting these regressions in their final weighted (WLS) form, I run an analogous OLS regression (without the weights) for each station, in order to identify and remove outlier phrases having a studentized residual greater than 15.0 for any single station. In the summary time-frame, this results in the omission of twelve phrases having a differential impact on one or several stations: `'public, radio'` and `'public, broadcast'` which are parts of the call signs for two stations, `'pacif, northwest'`, `'east, bay'`, `'bay, bridg'` and `'medicar, patient'` which appear disproportionately in west coast affiliates of ABC, NBC, and CBS and whose audiences skew elderly, `'monetari, polici'`, `'growth, rate'` and `'market, share'` which are over-represented on business news stations, `'wide, rang'` which occurs more frequently in British English, and `'govern, forc'`, `'bashar, al-assad'` which are used disproportionately by Al-Jazeera America (other stations prefer to use "Bashar Assad"). This method is used primarily as an automated way of ensuring results for any station are not being driven by outliers in the monthly analysis described below. Outlier phrases with differential impact upon any station are removed from all, and their removal reduces the standard errors in estimation across the board. I also confirmed that there are no data points with Cook's distance greater than 0.5 in summary WLS regressions, meaning no outlying data points have excessive leverage on results; this is in large part because the weights are already discounting weakly estimated phrases from the first stage.

I use two measures of Congressional ideology common in the literature. The first are Common Space DW-NOMINATE scores (Poole, 1998; Poole & Rosenthal, 2007), which range from -1 (very liberal) to 1 (very conservative) and are based upon roll call votes by legislators. NOMINATE scores are available from the first Congress until 2014[16], are probably the most used measure of ideology in political science. The 'common space' version of these scores allows comparison both between chambers and over time. The second measure I use are Americans for Democratic Action (ADA) scores, which have been used in previous studies of media slant. ADA is a liberal advocacy group which has scored congressmen since 1947 based upon their positions on a small set of issues. Scores are nominally between 0 and 100. I have extended the data to encompass 2008 to 2015, performed an adjustment (Groseclose *et al.*, 1999) with a 1999 base year to allow comparability through time and across chambers, and for consistency with NOMINATE scores, I reverse them so that higher scores are more conservative.[17] Although the two measures were independently tabulated, they agree to a great extent, having a correlation coefficient of 0.95 over the study period.

Finally, analysis is conducted on two time scales. First is a summary of the whole period, 2010-2015. For this, all congressional speech is considered in determining diagnostic phrases

[16]For 2015, I use legislator estimates from 2014 and prior, and disregard the speech of newly-elected first-time congressmen.

[17]"Reversing" means subtracting them from 100; resulting scores may be less than 0 or greater than 100 after adjustment.

Table 1.2: Diagnostic phrase cutoffs

	Congress ct, minimum	News ct, minimum	News ct, maximum	News:Congress, minimum ratio	Diagnostic phrases selected
Summary analysis					
2-grams	20	900	19,200	1.5	3000
3-grams	10	240	9000	1.5	2400
Monthly analysis					
2-grams	10	265	4200	1.5	500
3-grams	5	50	4200	1.5	400

and finding their ideological hues, and all station content is used jointly in assessing stations' slants. The second time scale is monthly. For the purpose of determining diagnostic phrases and their hues, the congressional speech from a given month is combined with speech from two prior and one following month[18]. Because the quantity of speech per month in Congress varies widely and can be low, this provides much more reliable estimates than using a single month would. Generating diagnostic phrases on a period-by-period basis allows the methodology to capture trending polarized phrases that emphasize the topics of debate that are occurring at any given time, and can permit discovery of phrases which are just at the beginning or end of a period of debate and polarization. The monthly diagnostic phrases thus discovered are applied to the station's speech in only the single month being analyzed. The exact cutoffs used in determining diagnostic phrases for both time-frames are specified in Table 1.2.[19] All told, software developed for this project totals over 3200 lines of custom Python code, and is available from `https://github.com/jtbr/tv-news-slant`.

.

[18]The first two months utilize congressional speech from November and December 2009. The last month is omitted in the time series.

[19]In limited experiments with using a smaller or larger set of diagnostic phrases, I find broadly similar results. When using fewer phrases, results show stations to be somewhat more differentiated, and more liberal. When using more phrases, they are somewhat less differentiated and more conservative. This is likely because the marginal phrase is less partisan as the number of diagnostic phrases increases, which leads to less differentiation on average between stations. As noted previously, my preferred cutoffs were chosen to balance the inclusion of partisan phrases with the exclusion of clearly non-partisan phrases. In neither case do any substantive conclusions in this paper change.

1.3 Results

Our aim is to determine the direction and extent of slant in the mainstream media on television and radio. The next section covers overall results for the whole period of study, 2010-2015 (inclusive). After that I cover how slants vary over the period of study, and explore extensions and implications.

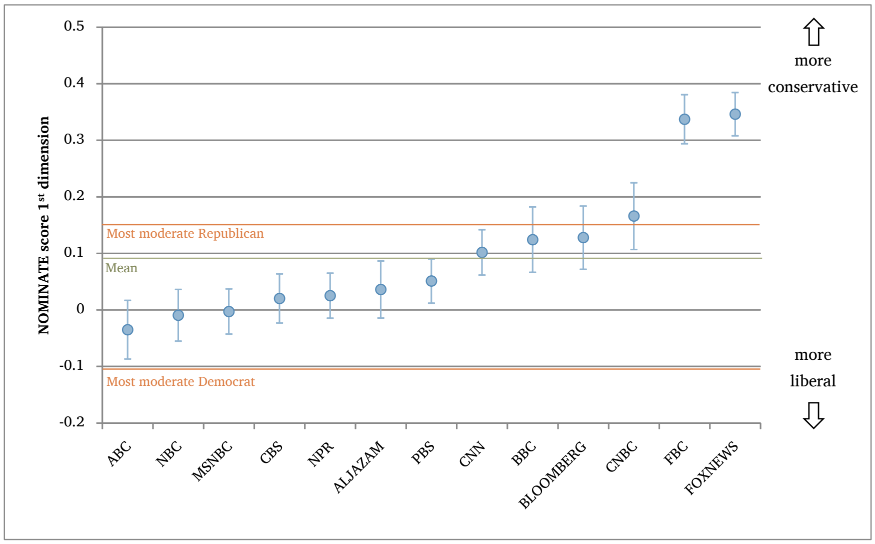
1.3.1 Summary Results

The summary results, showing my best estimate for a summary slant over the period, are shown in Figure 1.1 and listed in Table 1.3. These results are shown using both measures of ideology for comparison: common space DW-NOMINATE (Figure 1.1a) and reversed, adjusted ADA scores (Figure 1.1b), all including 95% confidence intervals. These are based upon diagnostic phrases chosen as described in Section 1.2.1, by considering the full period of congressional speech (as a whole) and applying it to the full content of each network's coverage over the study period.

Broadly, when we compare the order of news outlets' slants we discover they resemble what many media experts claim (e.g. Prior, 2013). The broadcast networks (ABC, NBC, CBS) cluster near each other and towards the liberal end of the spectrum of news providers, with cable networks being more varied in their slants. Fox News channel is substantially more conservative, while MSNBC is among the most liberal stations and CNN is almost exactly the mean station among those considered. The dedicated business channels Bloomberg, CNBC and Fox Business are more conservative than all but Fox News (with Fox Business being nearly equal); this is in part because they cover topics that Republicans speak more about than Democrats. PBS is the median station, with NPR being slightly more liberal, on par with the most conservative broadcast network (CBS) and with Al Jazeera America. BBC comes out as slightly more conservative than CNN. Given these results, it's not surprising many people alternatively believe that either MSNBC and broadcast networks or Fox News and Fox Business are slanted. These groups are fairly far apart ideologically and informed viewers can perceive a difference.

All of these relative results hold regardless of the measure of ideology we are using for slants. But the two ideologies can be analyzed in an absolute sense as well. A zero ideology on the

(a) Summary slants (DW-NOMINATE)



(b) Summary slants (reversed ADA)

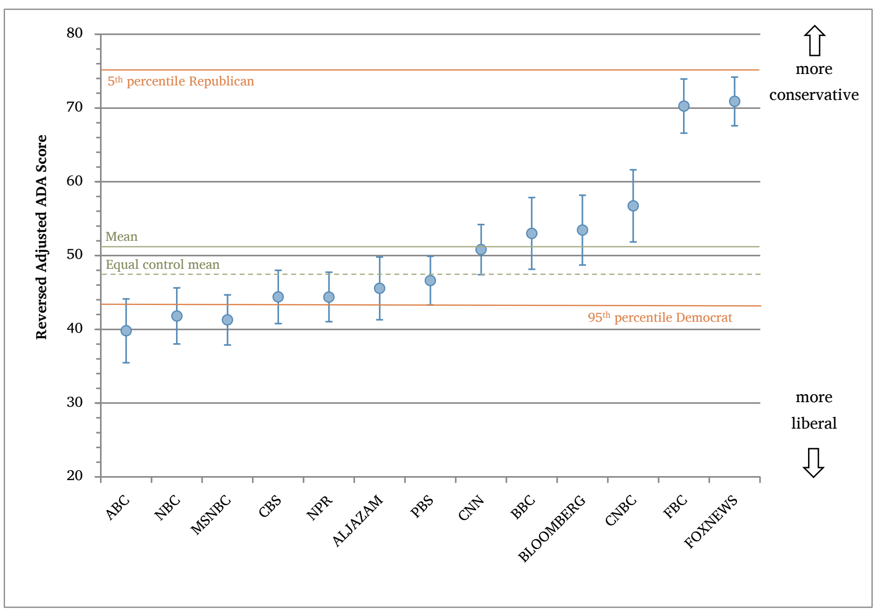


Figure 1.1: Summary slants by network for the period 2010-2015

NOTE: Estimates are shown with 95% confidence intervals. Data cover Jan. 2010 to Dec. 2015, except: FBC – available from Aug. 2012, ALJAZAM – available from Dec. 2012, BLOOMBERG – available from Dec. 2013, and CNBC – available from Jan. to May 2010 and then again from Jan. 2012.

Table 1.3: Summary Data

	ABC	NBC	CBS	PBS	CNN	FOXNEWS	MSNBC	...
2-gram count	17.4M	20.0M	12.4M	17.2M	109.4M	108.6M	100.3M	
3-gram count	13.1M	14.8M	9.6M	13.9M	84.8M	84.6M	78.9M	
diagnostic 2-gram ct	235K	265K	212K	420K	2.23M	2.99M	2.78M	
diagnostic 3-gram ct	40K	46K	42K	77K	474K	626K	644K	
% diagnostic 2-gram	1.35	1.33	1.70	2.44	2.04	2.76	2.77	
% diagnostic 3-gram	0.31	0.31	0.43	0.55	0.56	0.74	0.82	
% diagnostic phrase	0.90	0.89	1.15	1.60	1.40	1.87	1.91	
NOMINATE – summary slant	-0.035	-0.009	0.020	0.051	0.102	0.346	-0.003	
– summary slant SE	0.026	0.023	0.022	0.020	0.020	0.020	0.020	
– slant (merged monthly n-grams)	-0.018	0.027	0.012	0.090	0.098	0.278	0.023	
– slant SE (merged monthly n-grams)	0.042	0.038	0.037	0.033	0.036	0.033	0.032	
– mean monthly slant	-0.002	0.043	0.047	0.071	0.091	0.284	-0.009	
Reversed ADA – summary slant	39.80	41.82	44.39	46.60	50.81	70.89	41.28	
– summary slant SE	2.20	1.94	1.84	1.69	1.74	1.68	1.73	
– slant (merged monthly n-grams)	40.26	43.80	42.97	49.02	49.48	63.74	41.71	
– slant SE (merged monthly n-grams)	3.57	3.29	3.12	2.81	3.05	2.80	2.73	
– mean monthly slant	42.29	45.92	46.40	48.20	50.00	66.05	41.02	

	...	CNBC	FBC	BLOOMBERG	NPR	ALJAZAM	BBC
2-gram count		32.9M	28.4M	19.2M	12.2M	38.3M	3.6M
3-gram count		24.9M	22.0M	14.2M	9.9M	30.7M	2.8M
diagnostic 2-gram ct		521K	705K	281K	270K	672K	52K
diagnostic 3-gram ct		86K	135K	36K	53K	103K	6K
% diagnostic 2-gram		1.58	2.48	1.47	2.21	1.75	1.44
% diagnostic 3-gram		0.34	0.61	0.25	0.53	0.34	0.22
% diagnostic phrase		1.05	1.66	0.95	1.46	1.12	0.91
NOMINATE – summary slant		0.166	0.337	0.128	0.025	0.036	0.124
– summary slant SE		0.030	0.022	0.029	0.020	0.026	0.029
– slant (merged monthly n-grams)		0.179	0.284	0.164	0.080	0.036	0.101
– slant SE (merged monthly n-grams)		0.043	0.035	0.044	0.031	0.040	0.048
– mean monthly slant		0.172	0.289	0.146	0.046	0.101	0.124
Reversed ADA – summary slant		56.73	70.26	53.45	44.39	45.56	53.01
– summary slant SE		2.49	1.87	2.41	1.71	2.17	2.48
– slant (merged monthly n-grams)		55.61	63.78	55.13	48.09	45.74	51.68
– slant SE (merged monthly n-grams)		3.67	2.92	3.74	2.67	3.42	4.17
– mean monthly slant		56.88	65.91	53.88	46.11	50.11	52.99

NOTE: Data cover January 2010 to December 2015, except: FBC – available from August 2012, ALJAZAM – available from December 2012, BLOOMBERG – available from December 2013, and CNBC – available from January to May 2010 and then again from January 2012.

common space DW-NOMINATE scale represents the mean left-right ideology across American history.[20] I have also marked the mean congressman in the 113$^{\text{th}}$ Congress (2013-2014) with an olive green line labeled "mean." On the ADA chart, I marked a similar solid olive line for the mean representative in 2013, as well as a dashed olive line labeled "equal-control mean," which is meant to represent long-run average congressional ideology similar to NOMINATE.[21] We can see that CNN is nearly at the congressional dead center. With the exception of Fox, the business and foreign networks, all other networks are more liberal than the current congressional mean. This finding is consistent with Groseclose & Milyo's (2005) conclusion that most media besides Fox are left of current Congressional center[22]. But looking at the longer term Congressional center is more ambiguous. The conclusion holds for longer term means as well with the ADA data, but not with the NOMINATE data: no network is significantly more liberal than the long term NOMINATE mean. The difference is likely due to the lower resolution of ADA scores at more conservative points along the political spectrum.[23]

Nonetheless, much more striking than the difference between the networks is how moderate they all are compared with congressmen. On the figures, I have drawn what amounts to the ideological gap between the Democratic and Republican parties using two labeled orange lines. For DW-NOMINATE these lines represent the most liberal Republican and the most conservative Democrat in the 113th House of Representatives (2013-2014). For ADA scores there are a few outlying members, so the two analogous lines represent the 5th percentile Republican and 95th percentile Democrat representative in 2013. I can't reject the hypothesis that all stations lie within the ideological gap between the two parties for ADA scores[24]. For DW-NOMINATE

[20] Technically, it is the mean among congressmen who have served for at least 5 sessions of congress between the 1789 and 2014.

[21] Concretely it is the mean representative's ideology since the 1982 congress, the period having been selected since it covers 17 years each of Republican and Democratic control. For reference, the 113$^{\text{th}}$ House was under Republican control.

[22] Groseclose & Milyo (2005) use web articles rather than television content directly for their estimates of selected television networks.

[23] Specifically it is likely reflecting the fact that ADA data focus on a small number of topics that are of most concern to liberal activists. Since the recent shift in Congressional means is largely a product of a Republican drift towards the right (Mann & Ornstein, 2012; McCarty *et al.*, 2016), NOMINATE captures this more completely than ADA. Put another way, on the ADA scale, the ideologies of conservatives are compressed relative to liberals. Therefore the spread between the long and short term means is smaller in ADA than in DW-NOMINATE and closer to Democrats.

[24] Point estimates for the three most liberal stations lie slightly outside the gap.

scores, only the two Fox networks lie outside the gap between parties.

These results suggest that, given the number of independently owned news channels in America, competition may be disciplining bias as theorized in (Gentzkow & Shapiro, 2006). Since networks have concern for their reputation, the number of outside sources of information increase the probability that they would be exposed, at high reputational cost, if their coverage is too biased. Moreover, if in a spatial model, American viewers have a unimodal distribution of ideologies, akin to a bell curve, and a preference for confirmatory news, then we should not be surprised if networks cluster towards the central mode, and that the networks with the highest viewership should be those closest to the ideological center. This is broadly what we see.

(a) Congressional ideologies (NOMINATE) (b) Congressional ideologies (ADA)

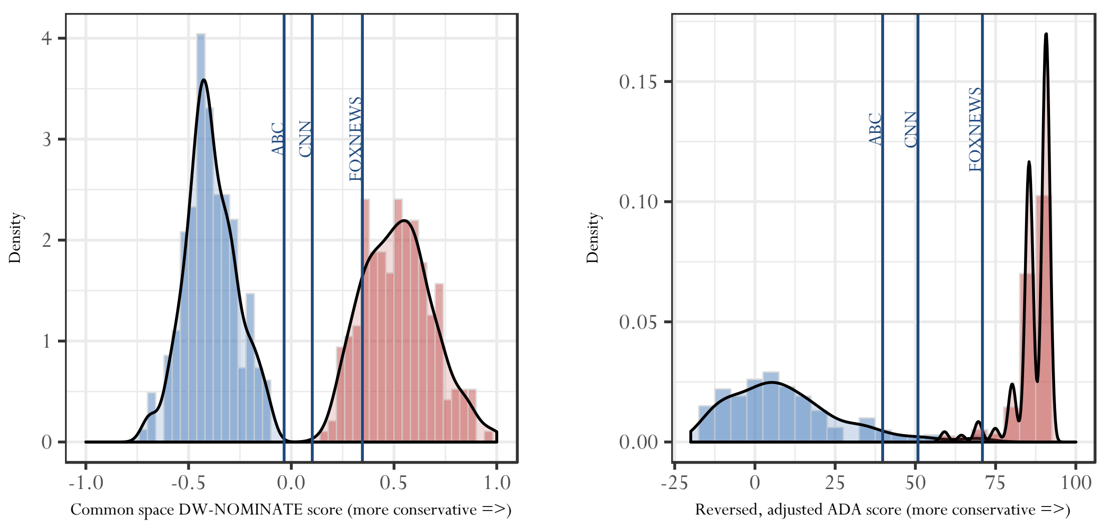


Figure 1.2: Distribution of Congressional ideologies, along with slant index for selected stations (Democrats in blue, Republicans in red; more conservative ideologies are higher)

How centrist are these networks? The *average* Republican and Democratic congressmen would lie outside the bounds of Figure 1.1. To underscore the relative moderation of television news, it is helpful to view the slants of the news networks in comparison with the distribution of congressional ideologies, as shown in Figure 1.2. The three news networks shown, ABC, CNN, and Fox News, cover the full range of television news slant, having the most Democratic, mean, and most Republican slants of any network, and all lie between the modal ideologies of the two parties. This is the case regardless of the ideological scale used. Only Fox News (and Fox Business, not shown) are even substantially within the distribution of a party. All the providers are considerably more centrist than Congress. My results strongly indicate that the existing mainstream television and radio networks, with the possible exception of Fox, occupy the center ground between the parties, and even Fox is to the left of most congressional Republicans.

Dissecting the coverage of moderate stations: balancing vs. avoiding extremes

All slanted stations are alike – they over-report one side at the expense of the other; each moderate station is moderate in its own way.[25] Scatter plots underlying the summary slant estimates are shown in Figure 1.3 for NOMINATE scores. Since most phrases, and especially the most precisely estimated (and thus highly weighted) phrases, are clustered towards the ideological center it's difficult to see what is going on in the middle without zooming in. But we can see that in the case of the Fox networks, there is less usage of the most liberal phrases and somewhat more of the most conservative. CNN, PBS and NPR have a more symmetrical distribution of phrase usages and nearly average phrase use frequency across the ideological spectrum. The foreign stations are particularly unlikely to use the most polarized phrases. ABC, NBC, BBC, CNBC, and BLOOMBERG tend use the most moderate terms more than usual.

The difference among moderate stations can be seen more clearly in Figure 1.4. Recall that diagnostic phrases are by definition the most polarized phrases spoken in Congress. This figure shows the percentage diagnostic phrase use in total speech for a station. There is wide variation even among stations with similar slants. The broadcast networks, as well as foreign networks, have particularly low diagnostic phrase usage, indicating that they spend less time discussing ideologically-polarized topics, possibly because they focus less on political news generally. They are moderate by *avoiding* coverage of extremes. For example on the day in December 2016 that CIA allegations about Russian interference in U.S. elections first surfaced, ABC and NBC

[25] With apologies to Leo Tolstoy.

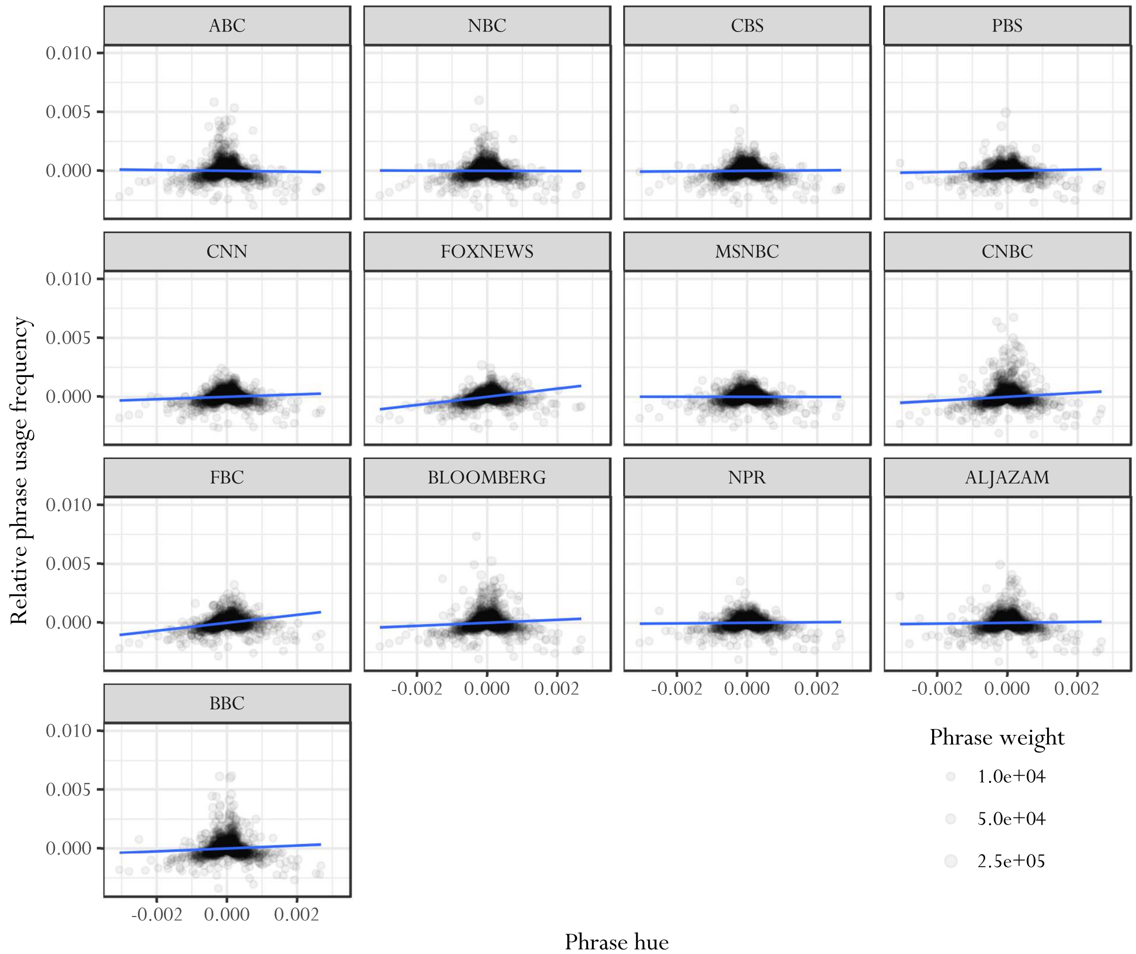


Figure 1.3: Scatter plots for final summary slant estimates (DW-NOMINATE)

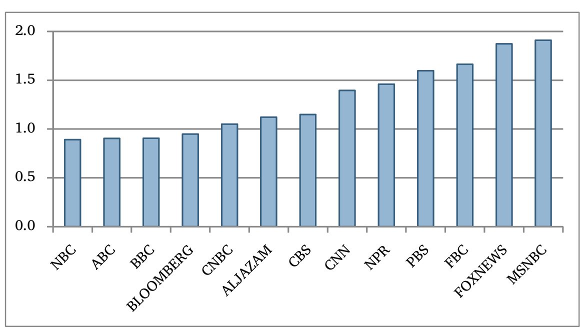


Figure 1.4: Percentage diagnostic phrase usage

led their newscast with coverage of a cold front arriving in the Midwest.[26] By contrast, Fox News and MSNBC use the highest proportion of diagnostic phrases in their coverage. If they are moderate, it is by *balancing* coverage of the extremes.[27] PBS, NPR, and CNN are somewhere in between, both balancing and avoiding extremes. Among the business networks, Fox Business uses diagnostic phrases quite a bit more than the others, reflecting a significantly greater slant.

Given the similarity between ADA and NOMINATE scores, for the remainder of the paper we will focus on NOMINATE scores. All substantial results hold regardless of the measure of ideology used.

1.3.2 Validating the Slant Index

So as to confirm that the slant index is in fact doing what we expect it to do, which is to measure political slant, I attempt to validate it against two sets of known data to verify whether it produces expected results.

As a test of internal validity, I test how well we can predict the ideologies of Congressmen using their speech alone using a procedure analogous to the estimation of news slant, regressing phrase usage on ideological phrase hues to find a slant for each congressman. Figure 1.5 shows that we can do this well – congressmen's implied slant has a correlation of 0.88 with their actual known ideologies, and a regression of ideology on implied slants has an R-squared of 0.77. Although subject to overfitting, this suggests that the diagnostic phrases we have selected, and the hues attributed to them, are indeed meaningful and predictive of ideology. Another important takeaway from Figure 1.5 is that although similar, the estimated slants have a slightly higher variance ($\sigma = 0.515$) than the actual ideologies ($\sigma = 0.455$): if anything the bias would be to exaggerate differences between stations rather than attenuate them.

We have already noted that the station ordering broadly agrees with the judgment of media experts. To further verify that these methods have external validity, I compare my overall slant estimates with audience-perceived slants. Following Gentzkow & Shapiro (2010), I obtained

[26] As noted by CBS White House correspondent Mark Knoller.

[27] MSNBC seems to be more balanced than Fox News; as noted previously, Fox is the most partisan station of those that we have analyzed, and in particular, it seems to under-use the most liberal phrases.

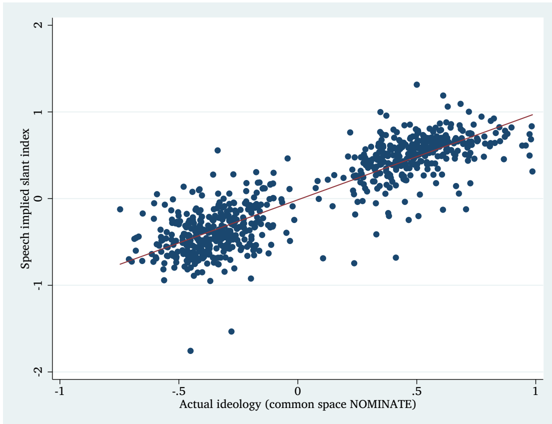


Figure 1.5: Congressional Ideologies and Implied Congressional Slants

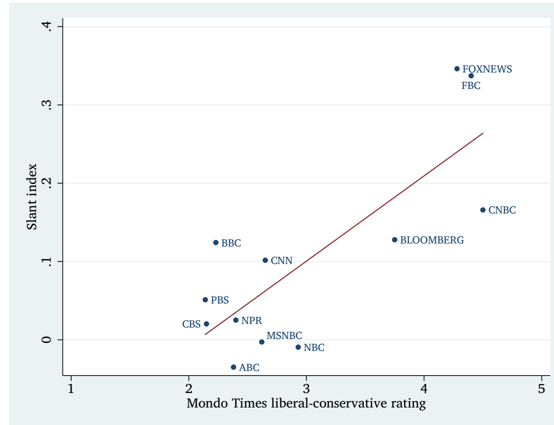


Figure 1.6: Slant Index and Mondo Times Political Bias Rating

NOTE: Mondo Times ratings are averages as of 14 July 2014 on the following scale: 1 Liberal, 2 Leans Left, 3 No Bias, 4 Leans Right, 5 Conservative

data from the website `mondotimes.com`, which allows users to assess the perceived slant of a news provider, with higher numbers signifying a more conservative tilt. Figure 1.6 shows the reasonably strong positive relationship ($\rho = .79$) between the slant indexes and the wisdom of the crowds as represented by viewer assessments on Mondo Times.

1.3.3 Time-series Variation in Slants

While my summary measure gives the average slant of a station over a long period, it might well be the case that slants vary over time, as events occur and different issues are discussed in the news. That is what we find. Although it is also possible that the stations might change relative position with respect to one another as time passes, we generally do not see this. With a few exceptions noted below, relative station slants are fairly stable. Figure 1.7 shows the three-month rolling average of slant estimates for each station based upon a set of phrases selected for each month (as described in Section 1.2.3).[28]

[28]BBC is excluded because the quantity of data are too low to support precise estimation on so short a time scale. Rolling averages are shown in order to aid interpretation; with the raw monthly data, the chart is more difficult to read.

The figures show how strikingly slants can shift over time. Viewing three-month averages actually conceals some of the shifts, but even so they are very wide. Fox News ranges from a three-month average of 0.09 to 0.55, a gap of 0.46; ABC ranges from -0.14 to 0.20, a gap of 0.34. Not only are these ranges large, amounting to substantially more than the gap between the parties in Congress, but they are on par with the total difference in slants between the stations overall. The difference between the summary slant of Fox News and that of ABC is only 0.38. Moreover, the most liberal station overall is actually more conservative at some times than the most conservative overall is at others (or conversely, the most conservative overall is at times more liberal than the most liberal overall is at other times). The key point is that shifts over time are substantial. Looking at raw data, a station shifts on average 0.12 from one month to another – enough to encompass the range of summary slants covered by NBC, MSNBC, CBS, NPR, Al Jazeera, PBS, and CNN.

Other trends can be seen as well. MSNBC is initially more centrist, and later mostly among the most liberal. Bloomberg becomes more conservative over the relatively short period for which I have data, while CNBC starts off quite conservative then becomes more centrist. Fox is consistently among most conservative at any given point in time, and more conservative than all other non-business networks by a significant and fairly steady margin. For the most part orderings between stations are rather stable. There are some shifts in the ordering of stations month to month, but mostly these reflect small absolute changes.

Covariation of stations across the ideological spectrum

In short, reviewing the chart, it appears that news cycles can shift widely from liberal to conservative. It is also apparent that all networks tend broadly to shift similarly as circumstances change, which is why the ordering of network slants remains fairly stable over time.[29]All of which suggests that news cycles can swing from liberal to conservative, irrespective of the network doing the presenting – whether because of changing events or just the whims of rhetorical debate being won and lost by both liberals and conservatives. Under the hypothesis that some

[29]In the appendix, I explore the interplay of events and diagnostic n-grams over time to better understand the dynamics at a descriptive level.

networks have a partisan affiliation upon which their reaction to news systematically depends (i.e., if they are closet partisans), we might not expect such strong covariation. Instead, we might expect networks to respond differentially to the same stimulus. Partisan networks might react more or less strongly, or in an opposite direction to neutral networks or networks with an opposing affiliation. That they don't is evidence against the hypothesis.

To make this notion of covariation more precise, let us focus on three stations spanning the ideological spectrum, Fox News, CNN, and ABC. All three are consistent with being I(1) processes at all standard levels of confidence – an augmented Dickey Fuller test cannot be rejected.[30] However they are pairwise cointegrated with a cointegration parameter of one; in each case the difference between the series is stationary – a unit root can be rejected at the 99% confidence level[31]. For example, the time series of Fox and ABC slants are both non-stationary, but they are cointegrated and the simple difference between them is stationary – that is, slants of both stations appear to respond equally to events. Thus, we find no evidence of the first potential prediction we would expect to hold if we had any "closet partisan" news providers.

Time series validation and components of variation

Since the time series and summary results are produced independently, it's helpful to compare the two sets of results. In particular we can examine the differences between the selection procedures for the summary period and for the time series, in two different ways. First, we can estimate the summary slant for the whole period using the phrases obtained by using the monthly methodology. Slants estimated using the union of monthly n-grams are shown in gray on Figure 1.8, along with the original summary slant estimates, reproduced here in blue. They are less precisely estimated, but results are broadly similar, which confirms that my selection procedure is robust and results are consistent. Second, we can look at how the average of the monthly slants for each station compares to the summary estimate for the whole period:

[30] Tests have a constant and no trend. For all tests I use 5 lags of the dependent variable, a number chosen based upon maximal information criteria.

[31] I perform a Dickey Fuller test on the differenced time series, e.g. $FOXNEWS - ABC$. Engle-Granger tests yield similar results without the assumption of a cointegration parameter. Estimates of the parameter are 0.89 for Fox and ABC, 0.80 for Fox and CNN, and 0.95 for CNN and ABC.

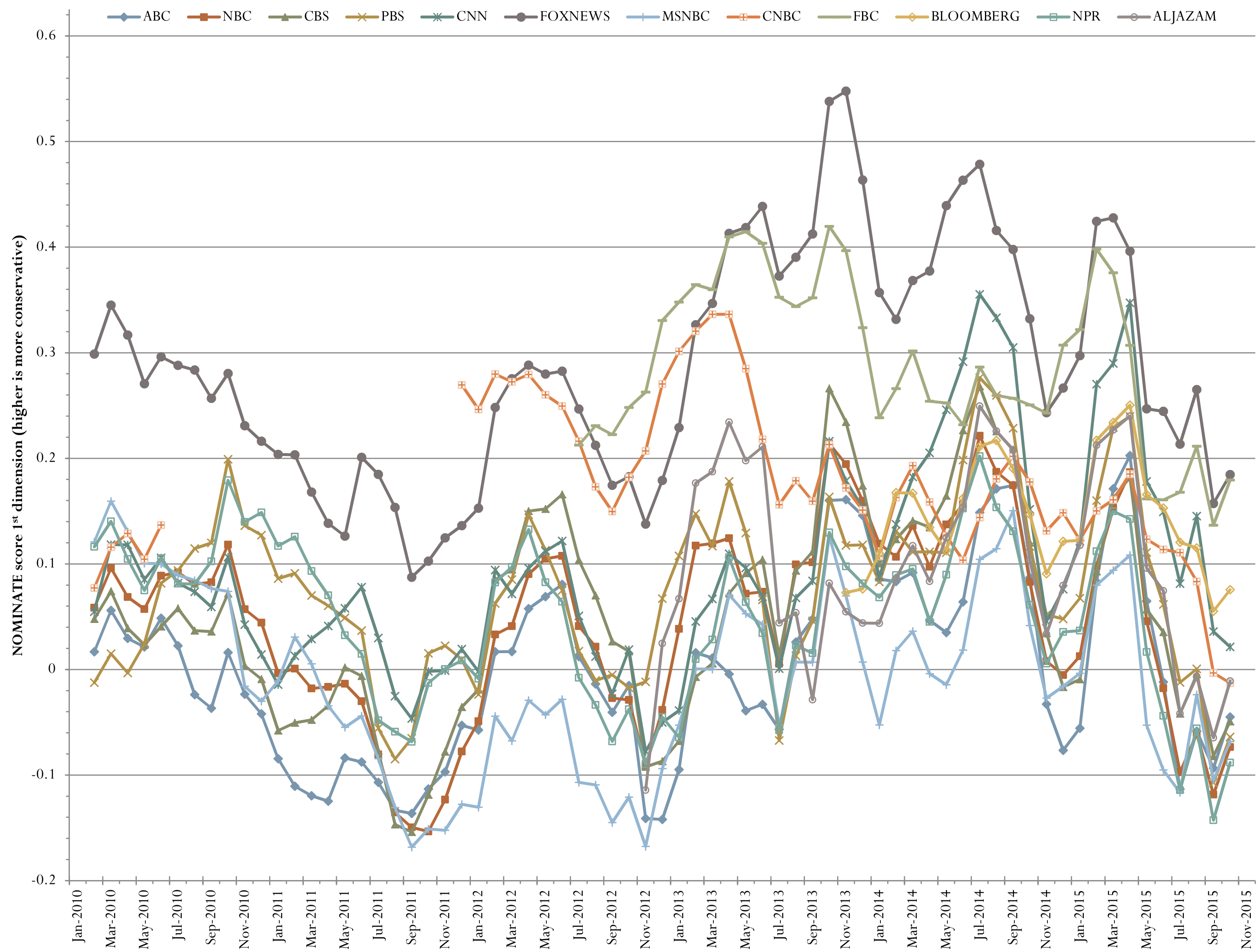


Figure 1.7: Station Slants from monthly n-grams, over 2010-2015, 3-mo. moving averages

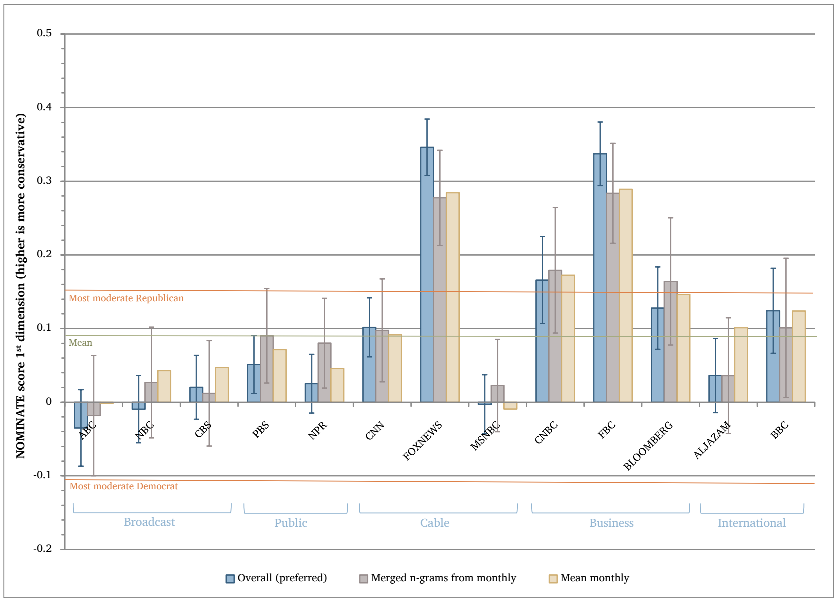
average monthly slants are shown in beige, also on Figure 1.8. Although these do not measure exactly the same thing, they should be close, and they are. This increases our confidence in the consistency of both the summary and time-series results.

To better understand what is driving the variation that we observe in slants over time, let us delve into the component parts of slant. The methodology we've used until now for time series is intended to capture changes in 1) what topics are discussed in a given month, 2) how polarizing those topics are at that time, and 3) how stations present those topics. Broadly, the first two changes are captured by the changing selection of diagnostic phrases, and by the changing hues of these phrases, respectively. How would results change if were to instead hold these two components constant, leaving only the variance in how stations choose and present topics? To see this, we can use the same fixed set of diagnostic phrases as in the summary results for the whole time series. Results are shown in Figure 1.9. Naturally, narrowing the sources of variation by using a single, (if larger) set of phrases and overall average hues, tempers the effect of varying topic salience and rhetorical resonance in a given month. This means the time-variance in station slants is lower, and each station becomes more stable around its mean. The average month-to-month shift in a station falls by half to 0.06. Some of this fall is due to a mechanical reduction in noise owing to larger sample sizes for each monthly estimation, however experiments with a much smaller single fixed sample size yielded similar reductions (and nearly identical dynamics), allowing us to estimate that only around one-fifth of the reduction in variance is due to this effect.

Rather, the reduced variation in Figure 1.9 relative to Figure 1.7 is mostly due to a combination of no longer being sensitive either to changes in the rhetorical turf of public discourse (changes in what is being discussed), or to changes in how partisan any given phrase appears to be in a given month. And between these two, much of the reduction appears to be driven by eliminating the change in phrase hues over time.[32] For example, a call to "pass a budget" has a very different hue if the congress is controlled by Democrats (2011) as opposed to Republicans

[32]This view is reinforced by an experiment using a union of all the monthly diagnostic phrases as the single set of phrases (with hues still assessed once for the whole period) to assess the slants of stations on a monthly basis. Results are very similar to those in Figure 1.9, which again suggests most of the difference is not due to variation in the salience and usage of phrases over time, but rather to variations in their hues.

(a) Summary slants (DW-NOMINATE)



(b) Summary slants (reversed ADA)

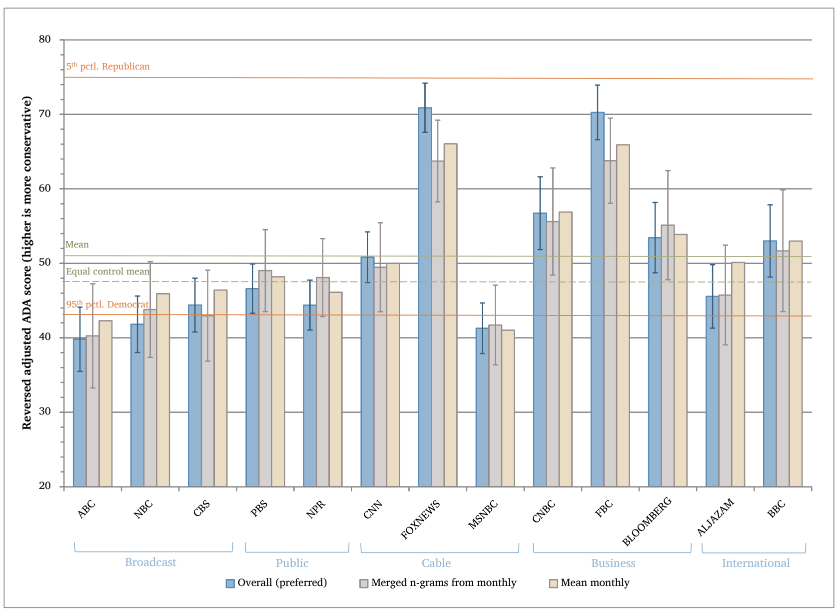


Figure 1.8: Summary slants by network for the period 2010-2015

NOTE: Estimates are shown with 95% confidence intervals when applicable. Data cover Jan. 2010 to Dec. 2015, except: FBC – available from Aug. 2012, ALJAZAM – available from Dec. 2012, BLOOMBERG – available from Dec. 2013, and CNBC – available from Jan. to May 2010 and then again from Jan. 2012. 'Mean monthly' is the arithmetic mean of the monthly time-series slants. 'Merged n-grams from monthly' is a summary estimate using a union of all n-grams used in estimating the monthly time-series.

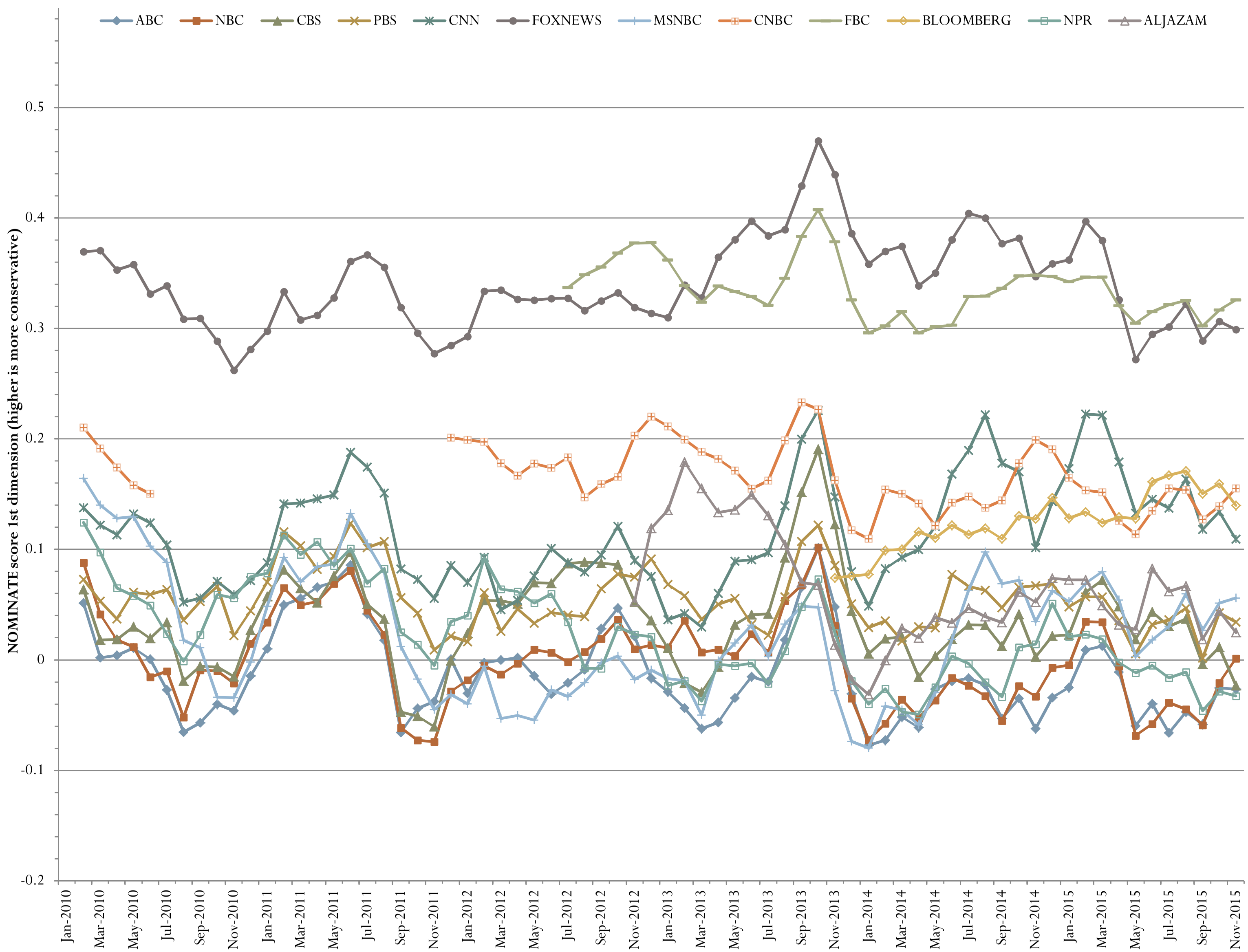


Figure 1.9: Station Slants from fixed n-grams and hues, over 2010-2015, 3-mo. moving averages

(2013). Similarly, the phrase "minimum wage" was only slightly Democratic until there was a grassroots campaign to raise it and the Minimum Wage Fairness Act was debated in 2014, when it suddenly became much more liberal. It is somewhat surprising that changes in phrase hues are a significant source of shift in news slants. But even once both these significant sources of seemingly legitimate variation[33] are eliminated – leaving the differential presentation of the same topics by stations as the main source of variation (as shown in Figure 1.9) – all substantial results in this paper hold.

Station differentiation trends and electoral season polarization

Let us now momentarily limit our attention to the seven "core" television stations for which I have complete data (ABC, NBC, CBS, PBS, CNN, FOXNEWS, MSNBC). In Figure 1.10a, I plot the path of the average station's slant over time. This illustrates the marked month-to-month shift that we noted earlier in the coverage of stations. Below we will examine the news that underlies some of these shifts. The *average* station slant ranges from a low of -0.15 (similar to Sen. Claire McCaskill) to a high of 0.35 (near Sen. Chuck Grassley), which encompasses the summary slants of all the stations. Even with the fixed n-gram and hue variation, the mean would encompass all stations except Fox and Fox Business. While the chart shows a slight trend towards slants becoming more conservative, this is not robust to variations in the model and is anyway uncertain given the high month-to-month variation. Figure 1.10b shows the path of polarization (defined as the overall distance from the mean) among stations over time. We can see there is a mild secular increase in polarization over the study period (which is robust to how n-grams are decided, and to changes in the definition of the reference point for the center). Over time, stations are becoming more diverse and less centrist, but this is happening slowly. This would make sense if viewers are becoming more polarized and demanding news to match, and there is evidence that this is beginning to happen (Pew Research Center, 2014b), since it may decrease the reputational cost of biased reporting. It is also consistent with the fact that there are increasing numbers of outlets for news, which increases the gains from specialization (or

[33]After all, the news is largely event-driven. Which topics are being discussed and capturing the zeitgeist is not exogenous, and can be hard for news providers to ignore.

conversely reduces the gains from clustering). Finally, the proliferation of new media sources and methods of obtaining news may be weakening the forces which discipline media bias, either by paradoxically making it less likely for consumers to see contradicting information, or by reducing the costs of being exposed as biased (if expectations of unbiased reporting decline, e.g.).

(a) Mean slant (b) Polarization

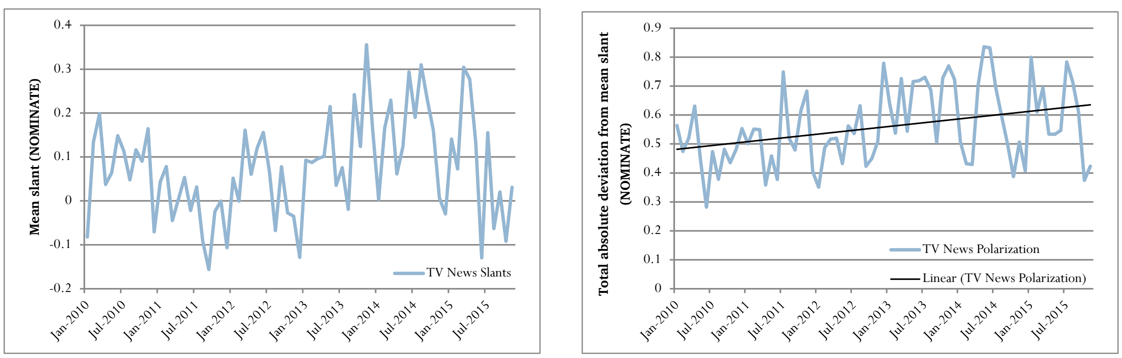


Figure 1.10: Evolution of core television station slants

Examining the trends in polarization allows us to test another potential implication of "closet partisan" networks hypothesis. If parties use more extreme (that is, polarizing) language as an election approaches, and at least some media outlets reflect the rhetoric of individual parties, then we should expect to see a wider range of slants (that is, more polarized stations) as elections approach. One reason parties may want to use more extreme language is to turn on/motivate their voters or conversely turn off/de-motivate centrists and supporters of opponents, and several scholars have noted this desire (see citations in Houser *et al.*, 2011, e.g.). A stronger version of the same theory would hold if some stations not only reflect the rhetoric of the parties but are actually advocating on their behalf, or considered themselves to be a part of the party in-group, as many partisans seem to believe. In fact, we find no evidence for this. Looking at the 10-12 months leading up to the elections in 2010, 2012, and 2014, there is no trend toward more extreme slants as elections approach.[34] There is some suggestive evidence that polarization might increase during the primary elections before dying down again for the general election, which could be the case if news channels are more focused on covering the preferred parties of their viewers during this time more than at other times. But even this tendency is not universal or strong. If anything, the trend is towards more centrist and *less* polarized coverage during election seasons. Indeed most of the peaks of polarization lie outside election years. Our data are thus inconsistent with this implication of "closet partisanship" as well.

1.4 Conclusions and Discussion

In short, we find that the ordering in slant of mainstream radio and television news networks largely comports with expectations, but that networks as a whole are substantially more neutral than Americans believe. All are highly responsive to events and their slants vary substantially over time, becoming more liberal or conservative depending upon prevailing topics or circumstances. They can all be called centrist by a reasonable definition of the term, with only the Fox Networks (Fox News and Fox Business) having a substantial distance from the ideological center point (while still remaining to the left of all but the most liberal Republicans in Congress). We also find that certain networks (including ABC, NBC, and CBS) appear to be centrist by avoiding the most partisan phrases, while others such as MSNBC do so by balancing the presentation of such phrases. We explored two implications of the theory that networks are actually "closet partisans," and while the results on this cannot be considered conclusive, we do not find any evidence for them. Networks do not respond differentially to events depending upon their slant, and they do not become more polarized as elections approach. Finally, we do find evidence that networks are differentiating more on slant and becoming slightly more polarized as the study period progresses.

[34]This is the case for the least centrist station, Fox News, as well – we see no clear trend towards more conservative-leaning rhetoric in the lead-up to these elections. In the 2012 and 2014 elections, there is some sign of increasingly conservative coverage during the primaries, but which then subsides during the general election.

Discussion

It's important to recognize this study's limitations. I assess the left-right political slant of networks. This will not capture other forms of "media bias" except to the extent that they may be correlated with political slant. Media may be biased towards covering stories which are sensational/scandalous, surprising/shocking, graphic, fear-inducing, or conflictual, all tendencies which may be more prevalent in television and radio news than political slant, but people are likely not referring to these when they say that the news is "biased." My measure may also not capture political disagreements which do not align along a liberal-conservative or Democratic-Republican continuum, although the evidence suggests that these have declined in importance in recent decades as the parties have aligned ideologically (Poole & Rosenthal, 2007). Moreover, even for capturing left-right slant, it is not perfect. Notably, it relies upon turns of phrase as a proxy for topic coverage (though itself only a part of the slant detected), and leaves ambiguous the sentiment associated with the phrases usages we count (i.e., a positive, negative, or other connotation). Although most phrases selected as diagnostic phrases are used with fairly unambiguous meaning, this is not assured, and does impart noise in the measurement, and may thus bias slants towards zero. Mocking or sarcastic usages cannot be detected, for example, and might confound the estimates, but are likely fairly rare in the data set. Finally, ideological slant may manifest itself in the media in a form, or around topics, that differ from in Congress. To the extent that this occurs, it may not be captured by the measurement. I doubt this is very great of an issue, particularly for the summary measure, where topics need only to have surfaced irregularly in Congress to be captured, but also because the presence of C-SPAN has meant that congressmen addressing Congress now have the same audience as the press: the general public. Furthermore, while we do find that networks are fairly centrist, there is nothing magical about the political center; it's just a point along the ideological spectrum. Political slant says nothing about who is correct or what is necessarily even fair. The best policy, or the truth on any particular matter, can align with rhetoric that may lie anywhere on the ideological spectrum (or may be off the public radar altogether). But the center does lie between the two political choices (in the form of the two parties) that voters are offered, so it is useful for informing electoral decisions. Also, being centrist doesn't mean that media are necessarily informative or accurate, or how much of news is substantive versus entertaining. Such other aspects of quality may well be as important as slant, but they are left for Chapter 2.

Viewers' strong perceptions of bias may differ from my conclusions for a variety of reasons. Most simply, conservatives often believe more liberal networks are biased, while liberals believe more conservative networks are (Pew Research Center, 2014c). If they are asked about "the media", it is usually those cases they are referring to; they are not normally indicting all networks equally. If the public, or at least partisan thought-leaders within the public, are as polarized as congressmen (which is not at all certain), it would not be surprising if they found networks on the opposite side of the spectrum to be "biased" simply because of the ideological distance from themselves. It is possible that viewers increasingly perceive a neutral media as biased because they themselves have drifted farther from the center. Their perceptions may be amplified since we have found that stations' slants vary greatly over time. If viewers' recollections of occasions when they believed networks were particularly slanted away from themselves are more salient, this could also explain their perceptions. Moreover, it's important to recognize that my measurements are all average estimates of a heterogeneous effect. Even for a single program on a network, a single slant estimate would necessarily average different slants for different stories or topics, and my measurements only capture the averages of multiple programs on a network over the course of at least a month. It is possible that some stations may be particularly slanted for certain particular issues or in certain programs, but that these issues and programs are balanced by others with a different slant. Again, viewers may overemphasize memories of slant for those issues to which they are most averse. Regardless of the reason, viewers perceptions of strong slant do not match my findings.

There has been much debate about the extent of media bias, whether it matters, and whether it can increase popular polarization. Theoretically, since people can rationally compensate for expected bias and seek out information with alternative slants, a slanted media need not necessarily lead to a biased information set among viewers. That, combined with lower exposure to the most slanted news (cable TV versus broadcast, e.g.) is grounds to doubt that slanted media in fact lead polarization to filter down from political elites to the general public (Prior, 2013) (or even that there is significant polarization of the public at large (Fiorina *et al.*, 2005)). But many argue that a biased media may in fact be feeding popular polarization (Baum & Groeling, 2008; Levendusky, 2013), or that it may cause even rational voters make electoral mistakes, in the sense that voters may vote differently than if they were fully informed (Bernhardt *et al.*, 2008). Other forms of media may. But this study's findings suggest that in the race towards greater partisan polarization, nationwide TV and radio networks are followers, not leaders.

Chapter 2: The Content of Television and Radio News

2.1 Introduction

The previous chapter analyzed one aspect of quality in nationwide American television and radio news: the political slant. This chapter analyzes other qualitative aspects of news coverage. These aspects of quality get to the heart of the public interest purpose of news, in particular how much information they convey to viewers or listeners, what categories of news they fall into, and how the information is conveyed.

Unprecedentedly, using a deep learning method, I assess the entirety of seven-years-worth of news content from 13 national television and radio news outlets. I find that some stations spend a majority of their time with speakers expressing opinions rather than reporting facts. Some other stations are more likely to focus on fear-inducing or outrageous news than others. No domestic station spends more than 36 percent of their airtime discussing international news. And stations spend on average only 2.2 percent of their time covering political candidates' backgrounds, platforms, or speeches.

The conclusion is that viewers of most stations are likely to end up consuming a large helping of opinion and soft news to go with the more substantive. While this paper is an examination of the supply of news, since they are competing for consumers' attention, the market seems to be concluding that people are in large part unwilling or unable to consume hard, factual, news about the whole world. It's hard to escape the belief that most news networks consider that they are principally in the entertainment business.

2.2 Methodology

This paper uses a novel method to analyze the composition of news presentations. We train the computer to recognize certain aspects about the news that we are interested in (topic, tone, etc.) in minute-by-minute chunks. We then unleash this computer method to assess the entirety of news coverage across networks.

Concretely, we will be analyzing the same closed caption and transcript data (*transcripts*) for news stations that was used in the first paper, except that the data series is extended to include programs from the year 2016. Our goal is to assess various qualitative aspects of the news which is being presented. To do so, I begin by splitting the news transcripts into short text segments (*snippets*) of roughly equal length. Each snippet is composed of three consecutive *parts*, each having one or more sentences and averaging around 100 characters in total. Each part may be from one or more speakers, and a single speaker may span multiple parts[1]. The intention is that the middle part of the snippet is the core text to be analyzed, with the before and after parts being there mostly to provide context for the interpretation of the middle one. However in practice no distinction is made between the parts, and the entirety of the snippet is considered as a whole in what follows. This method of generating snippets was developed through experimentation with the aim of producing sections of text which were generally large enough to be understood and assessed, but small enough for humans to quickly assess and to allow a fine-grained analysis of fast-flowing news presentations. I will assess the qualitative aspects of news for these snippets, and the aggregation of these assessments produces an overall picture of the coverage for each news outlet.

The dataset contains over 32 million snippets, so it is not feasible to manually assess the aspects we're interested in. Instead my approach is the following:

1. Use human intelligence to assess the various aspects of quality for a small subset of snippets

[1] I mark out a change in speaker by using the '>' symbol to preface the beginning of a new speaker's speech, so no information about changes in speakers is lost.

2. On the basis of the human input, train the computer to assess these aspects of quality automatically

3. Apply the learned computer classifiers to assess the full set of snippets

Of course, this is not trivial. Understanding human language is complicated, especially for computers. Moreover the language used in these news emissions is highly heterogeneous, the number of topics discussed is enormous, and the amount of context provided by snippets is minimal.

To increase the chance of success, I use a variety of techniques. First, I attempt to maximize the quality of the training set by facilitating correct labeling, and by using best-of-three labels for each sample. Second, I attempt to choose an optimal set of initial samples to enable the classifiers to generalize broadly, despite a small training size. Third, I use state-of-the art natural language processing (NLP) machine learning methods which leverage transfer learning from much larger and more diverse training sets to allow the machine to better understand the short snippets. Fourth, I use information about the classifiers' prediction probabilities, and not just the "best guess" for which class a snippet falls in, allowing more accurate aggregations. Finally, I had envisioned the use of iterative training which would select the most uncertain unlabeled snippets to be labeled, which (as a form of reinforcement learning) would maximize the learning rate, but results were already sufficiently accurate with the aforementioned techniques and so this was unnecessary. We discuss each of these techniques in later sections.

While developing and improving the classifier models, I assessed prediction accuracy using five-fold cross-validation, to avoid overfitting to the labeled data. This entails randomly assigning the labeled snippets to one of five "buckets." For each fold one bucket (so one fifth of the data) is treated as validation (test) data against which classifiers will be tested, and classifiers are trained on the remaining (four-fifths of the) data. Averaging the performance statistics across folds gives a good idea of how well classifiers can be expected to generalize to unseen data. The final accuracy statistics I report are based upon a hold-out set which remains out-of-sample for the entire procedure. This ensures that the reported performance statistics are unbiased, and indicative of performance on unseen data, which is essential since we wish to “generalize” to the full population.

Once prediction results are satisfactory, I apply the classifiers to the full population of snippets, which allows an overall assessment of news content presented by stations.

2.2.1 Dimensions of quality assessed

We wish to study a variety of dimensions of news quality. Firstly and most importantly, we evaluate the topics being discussed in the news. News is split into fifteen possible topic categories shown in Table 2.1 along with their descriptions. Three of these categories represent data to be excluded from further analysis (ads, transitions between stories, and content which is not classifiable or fits no other category). The categories can themselves be grouped into broader “super-groups”, or at the highest level, into “hard news” or “soft news”. For the purposes of this paper, I count the following as *‘hard news’* categories: substantive election/campaign coverage; business, finance and economics; science, technology and the environment; government affairs and politics; current events; and, social issues and cultural coverage. The remainder count as *‘soft.’* The intention in making this distinction is to count as ‘hard’ any news coverage which could potentially be valuable to viewers in their capacity as democratic citizens and voters. As such, this definition of hard news is a weak or minimal one, since it is only related to the subject of news being discussed. The topic leaves aside whether the discussion contains useful or actionable information, and even large portions of these categories may be only slightly valuable as information for citizen-voters (e.g accident reports, scientific breakthroughs, etc). Moreover, it could include lots of opinion and conjecture which may or may not be useful or actionable.

In addition to hard and soft news generally, the topics capture a distinction between substantive and non-substantive elections coverage, since much of election coverage is "meta" in the sense of covering the horse race in the polls, campaign tactics and gossip, as opposed to things that might help voters choose a preferred candidate.[2] The categories are ordered by priority,

[2]It could be argued that polls are useful to voters in case they are voting strategically, however in the U.S. most elections covered nationally are in effect two-party races, so strategic considerations do not come into play.

so that in the case the main theme of a given snippet could fit into more than one, the earlier category takes precedence.

Depending upon the topic category which best fits a snippet, other dimensions of quality may be relevant. In particular:

- *Foreign or domestic* - the geography of news coverage
- *Fact / Opinion* - whether the reporting is presented as factual or based upon evidence, or alternatively is opinion, conjecture or speculation[3]
- *Investigative* - for fact-based reporting, whether the reporting appears to be investigative or in-depth, as opposed to stories which are more superficial and inexpensive to produce.
- *Tone / sentiment* - whether the story is presented positively, negatively, or neither.
- *Appeals to emotion* - whether the reporting is either presented in such a way as to appeal to emotion, or whether the subject is itself emotional (particularly, fear-inducing / shocking / outrageous or alternatively pleasant)

These are evaluated depending upon the category. For example, for the science and technology category, we don't consider geography or fact/opinion. For ads and transitions categories, we don't consider any supplemental dimensions.

2.2.2 Supervised natural language learning approach

There are dozens of possible approaches to classifying text segments. The biggest choices are about the preprocessing (e.g. discard stopwords? punctuation? capitalization?), tokenization and numericalization (n-grams with bag of words? skip-grams? word embeddings (word2vec/doc2vec, GloVe)? language models?), and learning approach (SVM? random forests? boosted decision trees? neural networks? with RNN or CNN?). Many of these would not fare well with short text, given the challenging lack of context along with the high dimensionality of language. Among

[3]Note that this leaves aside the question of whether things that are presented as facts or fact-based actually are, or where the truth lies or even at a philosophical level whether it exists.

Table 2.1: Topic categories

Category	Description
Transitions, greetings, farewells	Beginnings and ends of segments, transitions between reports, 'teasers' for upcoming stories, chit-chat, etc. (nothing of substance)
Elections / Campaigns - substantive	Candidates' and parties' plans/platforms/policies or expert opinions about policies, candidates' records/background, interviews with or speeches by candidates, election results (U.S.)
Elections / Campaigns - fluff	The horse race (who's leading), polls/standings, speculation about outcomes, campaign tactics/strategy, talking heads/commentators (about an election), campaign updates (U.S.)
Business, Finance and Economics	News about the economy, markets, and business strategy, performance or business climate
Science, Technology & the Environment	Scientific & technological reporting; nature & environmental coverage (incl. climate, but not weather; incl. medical sciences but not health care)
Government affairs and Politics	US or International Government actions, functioning, policies, and impacts (including military, except use of force), and coverage of (non-U.S.-election-related) Politics
Entertainment / Arts & Celebrity news	News relating to celebrities and the music / arts / literature / entertainment industries
Sports	News about Sports and Games
Weather / Traffic	Current reports or forecasts of traffic or weather (but not the government response to weather, climate science, transport policy, or impacts of extreme weather events)
Consumer affairs, Products & Services	News reports about new or notable products or services, or reviews
Anecdotal and Human Interest	Stories which are perhaps interesting but mostly inconsequential. Usually quick/easy to produce. Often with only local significance and unknown or negligible wider impacts.
Current Events	Notable or consequential recent (at the time of presentation) news events. Terrorism, crime, protests, significant achievements, war/military use of force, etc.
Social Issues and Cultural Coverage	Societal or cultural themes; concerning communities, cities, health, society, religion, gender, race, human interactions, or trends (which don't fit into earlier categories)
Ads/sponsorships, self-promotion	Contents of apparently paid ads, sponsors that fund the network, or network self-promotion (such as for other programs)
None of the above / Impossible to classify	Content that doesn't fit in any of these categories; or incomprehensible, non-English, jibberish, or missing text; impossible to determine

Table 2.2: Supplemental categories

Category	Description
	Foreign / Domestic
Domestic	US domestic coverage
Foreign	Non-US, international/global coverage, or US gov't foreign relations
Unclear	Unknown/unclear/neither/both
	Fact / Opinion
Fact-based	Fact or evidence-based statements (regardless of your view of them, or whether the statements are correct)
Opinion	Editorial, opinion, conjecture, speculation, possibilities
Other	Other/neither/unknown (e.g. interviews)
	Investigative (only applicable when 'Fact-based')
Investigative or in-depth reporting	Hard-hitting or in-depth factual reports and investigations that take time and effort to produce (e.g. providing history/background/context, exposing corruption or scandals); serious policy analysis
Other	Anything else (so most stories); non-investigative reporting, including reports about others
	Tone (sentiment)
Positive tone	Positive in tone or sentiment (regardless of your view of the issue)
Neutral	Neutral in tone or sentiment
Negative tone	Negative in tone or sentiment (regardless of your view of the issue)
	Appeals to Emotion
Scary or outrageous	Appeals to emotion: Stories which are scary, shocking, or outrageous in their topic or presentation (whether or not you feel that way)
Feel-good	Appeals to emotion: Pleasant, sunny, inspirational or feel-good stories
Neither	Neither scary/outrageous nor feel-good, or no appeals to emotion (Most stories are neither)

the many options, I picked two broad approaches I expected to perform well on this problem. Both of these produce, through their learning process, a *classifier* that takes as input the snippet text and produces as output a set of likelihoods, one for each possible class. One classifier is trained for each dimension of quality (to predict the class likelihoods within that dimension).

First, to form a baseline against which I could evaluate a more cutting-edge approach, I classified snippets using a slightly older, specialized library focused on short text classification - LibShortText (Yu *et al.*, 2013). Checking against a baseline is important in light of the 'No Free Lunch' theorem in machine learning, which indicates that we can have no certainty that such a model will be better than any other for this particular problem.

LibShortText uses linear support vector machines (SVMs) to classify text using a bag-of-words model. It generates the vocabulary from the samples it trains on, so it has no outside knowledge to draw upon. As configured, it uses bigram features without stemming or stop word removal. For the training phase it uses instance-wise normalization and binary feature representation. The library trains quickly and performs quite well, given its relative simplicity (as we will see later).

I hope to improve upon this by using Howard & Ruder's (2018) state-of-the-art deep learning approach, dubbed ULMFiT. The main idea is to leverage a language model which is pre-trained on large amounts of data to dramatically accelerate learning (i.e., transfer learning). The language model itself is based upon the AWD-LSTM (Merity *et al.*, 2017) recurrent neural network (RNN) model architecture. The training process has two phases. First, I fine-tune the wikipedia-trained *language model (encoder)* to the new context of news snippets. Second, I use the encoder as the basis of classifier models which are then trained to predict the various dimensions of quality. While this is not designed specifically for short text classification, its ability to leverage a language model should help deal with the lack of context.

A language model in NLP takes as input a sequence of words[4] and attempts to predict the word that should follow. To do this successfully requires a fairly deep understanding of the underlying language, hence the name. The baseline language model was pre-trained on

[4]technically tokens, as they include punctuation and the like

the Wikitext 103 dataset (Merity *et al.*, 2016), which contains a large pre-processed subset of English-language Wikipedia (containing 28,595 articles and 103 million words). As such it has embedded a not insubstantial knowledge of both English and the world. I then use an approximately 1 percent random sample of snippets (322,364 in total) to train the language model to predict snippet text. This results in a language model which is capable of predicting the next word in a snippet with a remarkable 36.28 percent accuracy in a randomly selected validation set. From the language model's perspective, the next word might be any of hundreds of thousands of choices, yet it is able to choose the correct one over one third of the time. This indicates a fairly profound understanding of the underlying grammar and content.

This encoder is used as the basis for all of the classifiers, with the output layer of the neural network replaced with a softmax layer with the applicable number of nodes for each class. In training each classifier, I am careful to train only the last layers at first, while keeping the remainder of the network fixed, before successively allowing back-propagation to earlier layers in later rounds of training. This helps avoid "catastrophic forgetting" of previous network knowledge when transferring to the new domain of classification (snippets) (Howard & Ruder, 2018). I also make use of the latest methods used to quickly train robust deep learning models, such as one-shot training with slanted triangular learning rates (STLR), stochastic gradient descent with momentum, batch norm, and dropout (customized for RNNs (Merity *et al.*, 2017)).[5]

2.3 Implementation and validation

The fundamental unit of analysis is the "snippet", a short segment of text extracted from transcripts of television and radio news from 2010 to 2016 inclusive. News transcripts were obtained for all available television networks with nationally syndicated news, as well as the sole national radio network. The thirteen networks covered are: ABC, CBS, NBC, PBS, BBC, CNN, Fox News Channel (FOXNEWS), MSNBC, Al Jazeera America (ALJAZAM), CNBC, Fox Business (FBC),

[5]The ULMFiT classifier as used in this paper is implemented in the fast.ai library v1.0.42 using PyTorch.

Bloomberg Business, and National Public Radio (NPR).[6] For the television stations, transcripts are based upon closed captions and were obtained from the Internet Archive[7]; NPR transcripts were scraped from their website. The data were filtered in a manner akin to Chapter 1 to exclude non-news or non-national news programs. The data encompass substantially all national news-coverage for these stations for this period.

This project was mostly implemented in Python, and comprises around 5000 lines of custom code not counting portions used in chapter one. The Mechanical Turk interface discussed below is in HTML based upon Bootstrap and with custom JavaScript using JQuery. MongoDB is used for the data stores. Full source code is available at `https://github.com/jtbr/tv-news-quality`.

2.3.1 Snippet generation

Transcripts are saved for individual half-hour or hour-long television programs. For each program, the transcript text was split into speeches (segments of speech by an individual speaker). Then speakers' speeches were split into sentences. As noted previously, snippets are sections of text with three consecutive parts, meant to be long enough to be meaningful but not so long as to frequently span multiple topics or to be too time-consuming to manually label.

Snippets were constructed via a fuzzy algorithm that attempted to keep snippet parts to around 100 characters, while only splitting on complete sentences. Further, it attempted where possible to keep speeches together in the same snippet part, but long speeches are sometimes split into separate parts, and very short speeches (which can be quite common in a conversation) could be concatenated. Different speakers remain distinguishable by placing a '>' before each new person begins speaking. This algorithm led to fairly similarly sized snippets, which usually remained understandable but were not too long. Here is an example snippet:

> \> GOING THROUGH HIS MEDICAL STATEMENTS ONE DAY RICHARD WEST REALIZED HE WAS BEING BILLED FOR NURSING CARE HE WASN'T GETTING.

[6] CNBC, Fox Business, Bloomberg, and Al Jazeera were not available for the whole period and estimates provided are based upon the periods for which they are available. NPR is missing for 2016.

[7] `https://archive.org/details/tv`

> YOU WEREN'T EVEN HERE ON SOME OF THE DAYS THAT THE COMPANY ALLEGED THEY PROVIDED SERVICE FOR YOU.

> I WASN'T HERE. I HAD NO SERVICE. > AND YET HERE IT IS, IT'S BILLED.

Most closed captions are in all uppercase letters. But the deep learning model is capable of improving understanding (and classification performance) by making use of proper mixed casing. Therefore, I used a statistical true-casing algorithm (as it is known in the NLP community) together with a model trained on a wide range of contemporary English text[8] to produce mixed-case capitalization based upon sentence structure, proper nouns, etc. Results were surprisingly good, although not perfect (on other text, the model reports 98.4 percent capitalization accuracy). Although NPR transcripts were generally properly cased to begin with, I ran it through the same algorithm to avoid imparting bias. Thus the classifiers worked on machine true-cased snippets.

2.3.2 Sample selection approach

To improve the chances of choosing the most valuable initial sample of snippets to be classified by humans, I used a somewhat elaborate approach. Firstly, I chose a stratified random sample of snippets, ensuring that an equal number of snippets were randomly chosen for each station. Second, I wanted to sample across subject-matter topics that were chosen in an unsupervised fashion, since it has been shown (Taddy, 2013) that topic modeling (Blei *et al.*, 2003) can be used to sample evenly across language topics, and ultimately I need to predict both topic and other aspects of the speech. However the classical Latent Dirichlet Allocation (LDA) version of topic modeling is known not to work well with short text documents due to the high sparsity of words in documents, together with the high dimensionality of vocabulary (Jiang *et al.*, 2016). Since the LDA approach is not appropriate for snippets, I instead settled on an approach for short text clustering known as the Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model (GSDMM) (Yin & Wang, 2014). Using this approach, I clustered a stratified sample of 10,000 snippets for each of the 13 stations into 155 clusters (with the number of clusters having been

[8] `https://github.com/nreimers/truecaser/releases`

inferred in the process of clustering over 60 iterations, and tiny clusters having been combined). After verifying by hand that clusters seemed to have been meaningfully selected, I then chose a sample of 3 snippets from each cluster for manual classification along the various dimensions. I then performed a second phase of clustering from among large initial clusters (having >=100 members). This yielded 57 additional clusters, from which 8 snippets were selected each. In total 921 cluster-selected snippets were labeled and used, in addition to 2080 station-stratified randomly selected snippets (519 of which were held out for testing).

2.3.3 Obtaining labels: human assessment of quality

To speed up the process of labeling snippets[9], I used workers from Amazon Mechanical Turk (MTurk) who were paid a piece rate per snippet labeled. A customized user interface was designed to help elicit consistent responses, as described in Appendix B. This interface makes it impossible to submit incomplete answers, and it makes submitting correct answers as easy as possible.

Despite this, workers on a piece rate have an incentive to respond as quickly as possible, at the risk of lower quality answers. And high-quality labels are important to training successful classifiers which can generalize. To combat this risk, I implemented several measures. First, I only accepted MTurk workers who had completed at least 1000 MTurk tasks previously, with an acceptance rate of at least 98 percent. In addition, workers had to be U.S. residents, so they could be expected to have had some exposure to the events reported in the news. Second, I performed manual spot checks on submitted responses, in particular checking the responses which were submitted most quickly. In almost all cases, the responses were reasonable. Some individual workers were consistently better or worse than others, and worker identity was a much better indicator of quality than the speed with which they answered (anecdotally at least). But on average answers were surprisingly good despite weak monetary incentives. With the exception of one apparent bot, less than five responses were rejected. Third, I always solicited multiple responses for each snippet. Typically, I got two responses from MTurk workers for a

[9]Labeling a snippet takes on the order of 70-80 seconds on average. Thus to label 500 snippets takes on the order of 10.5 man-hours with no breaks.

snippet, and where there was not agreement, I broke the tie myself. This manual adjudication helped to improve consistency in labeling and limit outlays.

Although the labeling scheme is meant to encompass nearly all news programming, and be mutually exclusive, snippets are not always easy to classify, and there is frequently room for interpretation as to what is the best label. Sometimes, there is only enough context to understand what is being discussed with some background knowledge of news events over the past years. For example the snippet may refer to a scandal involving finances and certain people, but it may not be clear whether this is about a business or the government without outside knowledge. This background knowledge varies between labelers. Occasionally, it cannot be determined what is being discussed at all. In this case, labelers should label it "unclassifiable," but again, some labelers are better able to interpret or infer what is being discussed than others. In practice less than 2 percent of snippets were unclassifiable for one reason or another.

Another cause of difference between labelers is when a snippet encompasses multiple stories, with one ending and another beginning. Labelers are asked to pick the "primary" theme and to categorize for that one. Even then, labelers did not always agree on the primary topic. Thus in later rounds of labeling, the instructions were modified to explain that in case of uncertainty about what the primary theme is, labelers should use the middle text segment to help decide which is the "primary theme". There was also a tension in the instructions to pick the first category that fits "well". In some cases where more than one category could fit, labelers opted for what they considered a better fit later in the list, in others they picked the first category that might be considered a fit. For example a story about drug prices could arguably fit into business, science & technology (which includes medicine), government, current events, or social issues and cultural (which includes health) topic categories, depending upon other hints in the story. If they all fit equally well, the best answer should be the first (business in this case), but how "well" can be a fine distinction. It is worth noting however that in this case all of fine topic categories fit in the broader category of hard news. So when viewed at a coarser level of distinction, many of the ambiguities disappear. Finally, when it comes to some of the supplemental categories (tone, or emotional appeal for example), it can sometimes be difficult to disentangle the presentation of the news by news outlets from that of the news makers they are quoting or interviewing.

Ultimately labeling is a matter of interpretation, and one set of labels is not always more correct than another. This means that there is necessarily some uncertainty surrounding labels, and therefore in label predictions as well. Humans cannot always agree as to the best label for a snippet. But their label choices should not be expected to impart a bias that affects one station differentially from one or another. They are given no knowledge of the network whose snippet they are labeling (although in some cases the text gives that away), and they only know that snippets are from the period from 2010 onwards. They have no knowledge that their labels will even be used to distinguish different outlets. Finally, the label scheme is deliberately intended to be neutral, and not politically sensitive.

Experimental consistency of Mechanical Turk workers

To understand how much labels could vary between labelers, and how much MTurkers could be relied upon to label well, I evaluated pairs of labels by different labelers for the same snippet. I performed three analyses, as shown in Table 2.3. In the first, the researcher manually labeled a set of 520 station-stratified randomly sampled snippets. These labels can be considered to be of high quality, if not definitive. MTurk workers were then requested to label the same set. Statistics for agreement between the two sets of labels are shown in the third column of the table. For the most difficult task, topic category labels, there was 52.88 percent agreement. While this may seem fairly low, one must consider that there are 15 possible categories, so agreement by chance would only occur in 6.67 percent of cases.

Below the line break, we consider how well labels perform on broader groupings. The super-groups statistics refers to how frequently the labels concur if using a set of 5 broader categories composed of the 15 narrower ones. For example, the two elections categories are grouped with government to form one super-group (all the super-groups can be seen in Figure 2.1, to be discussed later). At this level, agreement reaches 67.31 percent. A yet broader grouping (hard/soft/neither) has 71.15 percent agreement. Looking at the individual categories of discarded data, we see 95.38, 89.62, and 95.00 percents agreement for ads, transitions, and unclassifiable respectively. MTurk labelers are clearly putting an effort into labeling the snippets. Much of the disagreement stems from genuine differences in interpretation. It is worth also noting how much difficulty labelers have in agreeing on even the relatively simple supplemental labels. The fact-opinion agreement of around 60 percent as probably indicative of how much difficulty humans in general have in distinguishing (or at least agreeing upon) fact and opinion (Rabinowitz *et al.*, 2013). But there is also often ambiguity between the underlying story which is being presented, and how it is presented. For example a story most people perceive negatively (about a kidnapping, say) may be presented neutrally. But the tone classification does not specify what the label should be based upon (in those cases where the two can be distinguished, because they often cannot be). Similarly, a news program may present the emotionally incendiary words of a news maker without being themselves incendiary. It is left to labelers to decide how to label the emotional appeal of such a snippet, and often it is difficult to tell who is doing the speaking.

Table 2.3: Snippet label agreement statistics

	MTurk-MTurk agreement, overall	MTurk-MTurk agreement, last batch	Experimenter-MTurk agreement (test batch)
topic category	47.06%	48.08%	52.88%
us-foreign	80.91%	80.15%	83.52%
fact-opinion	59.50%	61.98%	59.68%
investigative	63.79%	76.00%	64.86%
tone	55.70%	52.53%	59.63%
emotion	66.55%	63.24%	72.70%
super-groups	62.16%	63.27%	67.31%
hard-soft	69.18%	72.12%	71.15%
ads	92.83%	95.38%	95.38%
transitions	89.14%	94.42%	89.62%
unclassifiable	93.39%	91.54%	95.00%

Next, I perform two analyses to shed light on how consistently MTurk workers labeled snippets more generally. For this I analyzed all the snippets that were labeled by at least two MTurk workers, as summarized in column one. I also show, in column two, agreement statistics for the last batch of 520 snippets. Worker agreement improved over time, as I increased the pay, made a slight clarification to help choose the "primary" topic, and better chose the time of day and week that batches were submitted to workers. So the last batch has the highest level of agreement. In general, MTurk-MTurk label consistency is not much worse than between MTurk workers and the experimenter. In some cases it's actually better, most likely reflecting a difference in interpretation between the researcher and workers (the researcher likely has a deeper understanding of when a story is costly to produce in the case of investigative / non-investigative, and a slightly keener eye for discerning the presentation of fact from opinion).

The most commonly confused[10] topic categories were with government and other categories. This is not surprising considering that government has a role in so much of the news and in life, and it's often ambiguous whether that's the best fitting category for a given snippet. It is most often confused with the current events category and is also commonly confused with the two elections categories. Ads and 'Consumer affairs, products and services' are also commonly confused, as indeed it can sometimes be difficult to tell whether glowing coverage of products is paid or not. For the fact-opinion classification, it is rare to find agreement that a snippet is 'other', which is most often the case for interviews. For tone or sentiment, workers often disagreed on where the lines are between neutral and positive or neutral and negative. Similarly, but to a lesser extent, there was disagreement on where the line should be for emotional appeals to fear, shock, or outrage.

2.3.4 Label distribution

If we take the subset of labeled snippets which were derived from random sampling (with station stratification[11]), the label distributions can be viewed as a crude estimate of snippet-population

[10]I say confused as to mean there was disagreement over, not to say one or the other is right.

[11]As such, each station's distribution is given even weight in the overall distribution, even if that station is not available for the whole period, and despite wide variation in how much news coverage each station produces. In

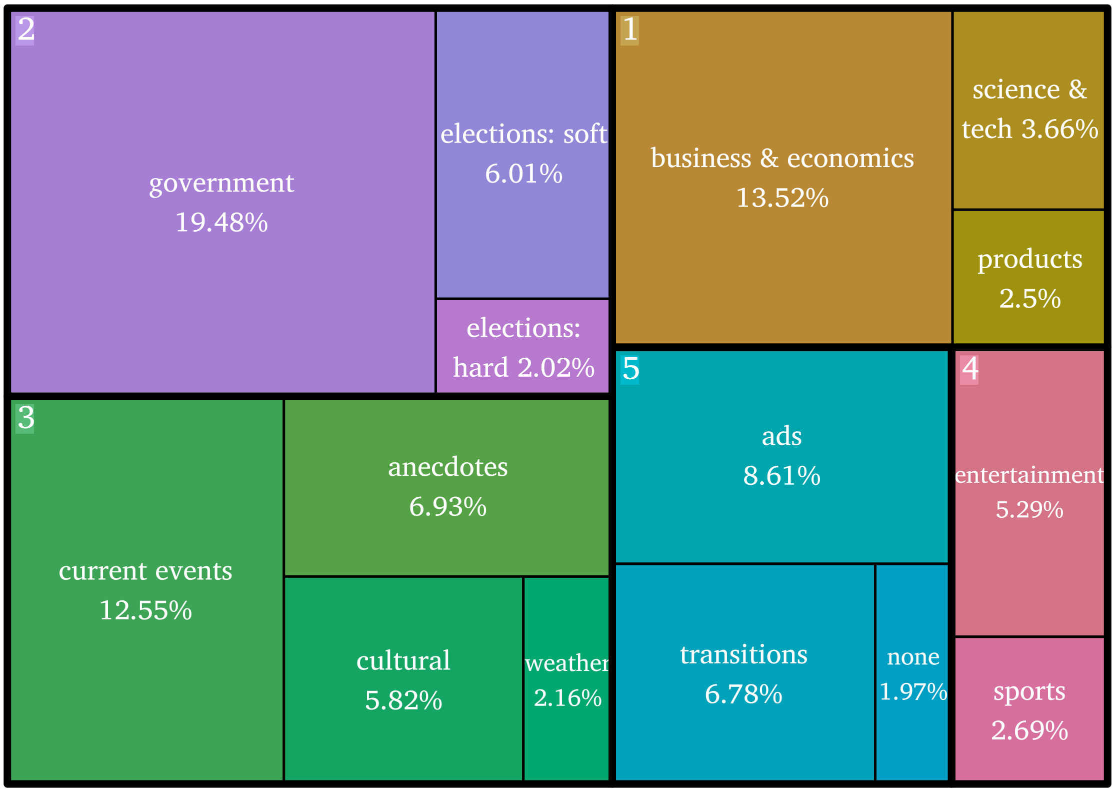


Figure 2.1: Topic category distribution of 2079 labeled snippets, by super-group

statistics. This sample includes the holdout set as well as those used for training. Figure 2.1 shows the distribution of topic categories among labeled snippets. These are grouped into the five super-groups of topics. Four categories predominate, accounting for over half of snippets: government, business/economics, current events, and anecdotes. Uninformative content (ads, transitions, and unclassifiable) together account for 17.36 percent of snippets. 8.03 percent of snippets are covering US elections, and only one quarter of those are substantive coverage. Altogether, 61.04 percent of snippets are labeled with hard news categories, before excluding uninformative categories.

Distributions of labels for the supplemental dimensions are listed in Table 2.4. A 62.35 percent majority of labeled snippets were domestic. A plurality of 48.61 percent were statements

particular this means the business networks with their more focused topic distributions, may be to some extent obscuring the patterns for more general networks of opinion, rather than fact. Of the 43.3 percent labeled as fact, only 20.8 percent were investigative stories. The clear majority of snippets were neutral and lacking in emotional appeal. But nearly twice as many were negative as were positive, and about twice as many were either scary, shocking, or outrageous as were pleasant.

Table 2.4: Label distributions of supplemental dimensions

US / foreign		Fact / opinion, investigative		Tone		Emotional appeal	
domestic	62.35 %	opinion	48.61 %	neutral	54.02 %	neither	77.82 %
foreign	26.54 %	fact, non	34.30 %	negative	29.33 %	scary/outrage	14.57 %
unk / both	11.11 %	fact, invest	9.01 %	positive	16.66 %	pleasant	7.61 %
		other	7.87 %				

NOTE: For US/foreign, n=1296, for Fact/opinion, n=1220, for Tone, n=1681, for Emotion, n=1551.

2.3.5 Classifier training and accuracy

I trained the classifiers over several iterations as I gathered more labeled data, to monitor how labeling was affecting classifier performance. Indicative accuracy for each classifier approach is shown in Table 2.5, based upon average cross-fold performance, along with accuracies we would expect from two types of random classification. The first section shows accuracy over time for the topic category, as more labels become available. The lower section shows performance for the supplemental label classifiers, based upon the full snippet data set (n=2479 snippets in total). Note that the supplemental categories are not applicable to all topics. Both classifiers show only moderate and diminishing improvement as the number of samples increase. The final batch (bringing the total number of snippets from 1959 to 2479) was a randomly-selected batch. This suggests that additional randomly-selected labels would improve the existing classifiers only modestly if at all. Instead further improvement would likely come most quickly from selecting snippets for labeling based upon classifier uncertainty, with a particular focus on under-represented classes.

When training with 2479 labeled snippets on the most challenging task (the 15-category snippet-topic classification), LibShortText achieves 52.84 percent accuracy on its “most-likely” category prediction. For comparison, a random guess would be correct only 6.67 percent of the time, while a naive classifier which was based only upon the category frequencies would be 12.05 percent accurate. In fact, this improves upon how often MTurk workers can agree upon a label, and parallels the agreement between the researcher and MTurk workers. The pattern is the same for the fact-opinion dimension, and for the emotion. These are likely areas where LibShortText can discern rules to separate these classes that are eluding (at least some) humans. By contrast, it performs more poorly on the us-foreign, most likely because it has no outside knowledge of people, places and things to draw upon.

Table 2.5: Classification accuracy (avg. across folds), by training set size

Label type	#/snippets	Random	Label-weighted random	LibShortText	ULMFiT
topic category	1503	6.67%	12.93%	51.63%	68.00%
topic category	1959	6.67%	13.45%	52.68%	69.52%
topic category	2479	6.67%	12.05%	52.84%	69.06%
US / foreign	1259	33.33%	46.73%	70.58%	76.56%
fact / opinion	1179	33.33%	43.25%	70.65%	72.95%
investigative	535	50.00%	63.21%	74.79%	75.87%
tone	1635	33.33%	40.22%	51.42%	56.06%
emotion	1491	33.33%	62.54%	77.63%	79.22%

The ULMFiT classifier improves upon the LibShortText classifier for each label set, and usually by a significant margin. This is particularly impressive in that it predicts the correct class more often than the experimenter and MTurk workers can agree in most cases, although it still falls short in the tone and us-foreign dimensions. For the tone dimension, this may be due to inconsistent interpretation of label, as noted above (tone of presentation vs affect of underlying news). It seems that ULMFiT method is able to learn and apply a more consistent set of rules for the topic than even individual humans. The relative outperformance of these classifiers on our data confirms the choice to use more recent ULMFiT approach. All further discussion of the classifiers will be using this approach.

Classifier outputs

For each label set, the classifier is optimized to provide the "best guess / most likely" class label. Ultimately the choice comes down to choosing the class having the highest likelihood among all the possibilities. This is exactly what we want when we're trying to predict a single snippet. But this is actually losing some information generated by the classifiers (if for example several classes are similarly likely, as we know to be the case with snippets) and in our case we are less interested in the correct label for any individual snippet than in the frequencies of the labels across the population of snippets. Hence, I will also explore the possibility of using the top k class likelihoods for each snippet. These likelihoods are normalized to sum to one across the k most likely classes. Then they are summed across snippet predictions to give a likelihood-weighted estimate of classes.

Performance on holdout set

Unlike in previous iterations, where each fold's classifiers were trained using a fixed procedure, the final set of classifiers was trained by hand (still on an 80 percent random sample of training snippets). This allowed me to better tune the models' hyper-parameters and to ensure the randomness of stochastic gradient descent worked in my favor, thus ensuring the deployment of high-quality classifiers. Once complete, I needed to evaluate performance before applying it to the full set of snippets. The true test of a classifier is how it performs on unseen data that has not been used at all during the training process. That is the purpose of the holdout set, a station-stratified random sample of 519 snippets, for which results are presented in Table 2.6. Looking at the (comparable) accuracy column, are in all cases higher than the average of the auto-trained cross-fold classifiers as shown in Table 2.5. Being hand tuned, it's not surprising to be able to beat the average classifier across the folds. But this is on the holdout set so such performance is very encouraging. It clearly indicates that the classifiers are not overfit to the training data (they may actually be somewhat underfit), and generalize well. For the topic category, "best guess" accuracy is 69.17 percent, as against expected accuracy of 6.67 percent for random guessing, and 10.44 percent for guessing based upon class frequencies. I am now combining the fact-opinion labels with the investigative/non-investigative labels (which are by definition fact-based), in the "fact, investigative" classifier (so this classifier has four labels: fact-investigative, fact-noninvestigative, opinion, and other).

Table 2.6: Classification accuracy on holdout set

Label type	#/snippets	Random	Label-weighted random	Accuracy	Top-k accuracy
topic category	519	6.67%	10.44%	69.17%	91.14%
us-foreign	328	33.33%	43.93%	84.45%	93.60%
fact, investigative	309	25.00%	41.37%	70.23%	89.00%
tone	417	33.33%	38.72%	58.99%	91.85%
emotion	381	33.33%	60.41%	77.17%	96.85%
super-groups	519	20.00%	22.34%	77.07%	96.53%
hard-soft	519	33.33%	40.37%	82.85%	97.50%

NOTE: Classifiers are hand tuned, except for super-groups and hard-soft, which were taken from the second-best performing fold. Top-k accuracy indicates frequency of choosing the correct class among the classifier's k most likely. $k = 3$ for topic category and $k = 2$ for all others.

The last column of the table shows the percent of snippets for which the classifier's top k most likely classes includes the correct class. All have top-k accuracies in excess of 90 percent, with the exception of the fact, investigative classifier (at 89.00 percent).

The out-of-sample holdout set can also be used to assess the accuracy of our method for generating population estimates for all snippets. We can compile snippet class frequencies generated by classifiers using both methods (top and top k). These can then be compared against the actual label frequencies in the holdout set. These are shown in Table 2.7, where the true label frequencies for each label are shown in column two (% Holdout set). Then in columns three and four are the frequency at which each class is predicted, along with the error (in frequency) relative to the true labels. Columns five and six are similar, but where the class frequency is the sum of the top k normalized likelihoods. These errors can be taken as estimate of per-class bias.

Table 2.7: Predicted frequencies for holdout set

Class	% Holdout set	Top 1 prediction %	Top 1 Error	Top k, weighted prediction %	Top k, weighted Error
Topic category					
transitions	7.13%	5.78%	-1.35%	6.29%	-0.84%
elections - hard	1.35%	0.39%	-0.96%	1.66%	0.31%
elections - soft	7.13%	6.74%	-0.39%	5.94%	-1.19%
business & economics	14.07%	15.22%	1.15%	13.81%	-0.26%
science & tech	2.70%	2.50%	-0.20%	2.84%	0.14%
government	20.42%	20.23%	-0.19%	19.61%	-0.81%
entertainment	5.78%	5.20%	-0.58%	5.18%	-0.60%
sports	3.08%	3.28%	0.20%	3.33%	0.25%
weather	3.28%	3.28%	0.00%	3.09%	-0.19%
products	2.70%	0.96%	-1.74%	1.54%	-1.16%
anecdotes	6.17%	6.74%	0.57%	6.60%	0.43%
current events	10.02%	16.19%	6.17%	15.60%	5.58%
cultural	5.59%	3.47%	-2.12%	4.54%	-1.05%
ads	9.25%	9.63%	0.38%	9.25%	0.00%
none	1.35%	0.39%	-0.96%	0.72%	-0.63%
Fact, investigative					
investigative	6.80%	5.18%	-1.62%	9.37%	2.57%
noninvestigative	32.04%	38.84%	6.80%	37.82%	5.78%
opinion	55.02%	54.37%	-0.65%	49.00%	-6.02%
other	6.15%	1.62%	-4.53%	3.81%	-2.34%
US - Foreign					
domestic	58.54%	66.77%	8.23%	64.53%	5.99%
unknown	13.41%	4.27%	-9.14%	7.14%	-6.27%
foreign	28.05%	28.96%	0.91%	28.33%	0.28%
Tone					
positive	19.18%	11.99%	-7.19%	11.58%	-7.60%
neutral	51.32%	68.59%	17.27%	59.68%	8.36%
negative	29.50%	19.42%	-10.08%	28.74%	-0.76%
Appeal to emotion					
scary	16.01%	3.94%	-12.07%	10.57%	-5.44%
neither	75.59%	95.28%	19.69%	85.56%	9.97%
pleasant	8.40%	0.79%	-7.61%	3.87%	-4.53%

NOTE: For topic categories, all 519 snippets are applicable. For fact, investigative, n=311, for US-foreign, n=328, for Tone, n=417, for Emotion, n=381. Top-k weighted prediction % indicates the summed normalized likelihoods for the k most likely classes per snippet, where $k = 3$ for topic category and $k = 2$ for all others.

For all except the “fact, investigative” classifier, the predicted top-k frequencies are more accurate on net than the top-1 predicted frequencies. As such, I will base my population frequency estimates upon the top-k accuracy except in the case of the “fact, investigative” classifier, and focus my analysis here on the same. The results indicate some not insignificant bias in the frequencies. This is most pronounced in the supplemental categories, and in all cases except “fact, investigative”, represent an inflated tendency to choose the “default” (most common) category. In essence, it appears that in case of doubt, the classifiers are biased towards choosing the most common. For “fact, investigative” it’s biased towards the noninvestigative class to the detriment of the infrequently-used “other” class.

The topic category classifier prediction distribution matches surprisingly well to the actual distribution. The main discrepancy is in somewhat-too-many predictions of the “current events” class. Particularly impressive is the ability to predict some of the less frequent classes such as “elections - hard” and “science & tech” with very nearly the correct frequency. Correctly predicting minority classes in imbalanced data sets such as this can be challenging.

Although we do see some bias in the classifiers’ frequency predictions, I see no reason to expect that these biases would vary by news outlet. The classifiers have no outside knowledge of the news outlet producing the snippets. And in fact, attempting to predict which outlet produced a given snippet is quite a difficult task. Training such a classifier yields a cross-fold average accuracy of only 26.95 percent. If a classifier cannot very well predict a station outright, it seems unlikely that it could or would have significantly different biases for any of the quality dimensions between news outlets.

2.4 Results: assessment of news content

I applied the refined classifiers to all the nearly 32 million snippets from the thirteen networks, gathering statistics for each 3 month period from 2010-2016. I was thus able to obtain a detailed picture of coverage overall, how this changed over time, and how it differed by station. I then examined interactions between the classifiers hard-topic-classified snippets, allowing us to also assess cross-dimension patterns for this subset. The results presented below exclude non-news content from three categories: transitions (accounting for 6.3 percent of total coverage), ads, sponsors, and self-promotion (10.9 percent), and content that is unclassifiable for one reason or another (0.9 percent). These fractions vary little over time. My fundamental cross-sectional results by network are presented in Table 2.8 for the topic categories, and Table 2.9 for the supplemental categories.

2.4.1 Topic coverage

The networks vary significantly in what topics they cover. This can be most easily seen in Figure 2.2, which is ordered (in ascending order from top to bottom) by the amount of business and economics coverage they have. Unsurprisingly, the business networks (CNBC, Bloomberg, and Fox Business (FBC)) have far more business and economics coverage than other networks. But there is significant variation between them; FBC actually covers much less business (29.2 percent of coverage) than the other two (52.5 and 67.3 percents). It compensates for this by including much much more government and elections coverage (amounting to nearly three times as much government and five times as much election coverage as CNBC). This is particularly true towards the end of the study period, as FBC is trending towards less business coverage, having started out in 2013-2014 with around 40 percent coverage and falling below 20 percent throughout 2016. Among the general news networks, PBS stands out for including over twice as much business and economics coverage as any other outlet (at 18.0 percent).

There is similar diversity in their coverage of current events. BBC has by far the most current events coverage at 39.7 percent, followed by Al Jazeera America (ALJAZAM) at 32.3 percent. ABC, CNN and CBS also have high proportions of current events coverage (26.2, 26.2, and 23.4 percents respectively), with CNN being the most erratic. At some times it has more coverage than any other network, and at others it is close to the average. The business networks, somewhat surprisingly, have the least coverage with 3.2 percent for CNBC, 5.0 percent for Bloomberg and 7.6 percent for FBC.

Table 2.8: Estimated topic frequency in news (percents)

class network	anec-dotes	business, economics	cultural	current events	elections - hard	elections - soft	entertain -ment	gov't	products	science & tech	sports	weather	hard	soft
ABC	16.65	4.43	4.50	26.21	0.92	4.81	14.26	9.77	3.52	3.04	4.37	7.52	48.88	51.12
ALJAZAM	5.99	6.00	9.72	32.26	0.70	2.31	2.80	28.90	0.88	5.53	3.16	1.74	83.13	16.88
BBC	5.47	5.52	6.16	39.71	0.80	3.20	3.75	28.33	0.62	4.07	1.95	0.43	84.59	15.41
BLOOMBG	4.08	52.51	3.18	5.02	1.56	5.49	4.77	16.44	2.27	2.85	1.70	0.13	81.57	18.43
CBS	12.93	6.51	5.70	23.39	1.28	5.82	10.28	16.09	2.74	4.28	4.27	6.71	57.25	42.75
CNBC	4.45	67.28	1.51	3.21	1.08	2.21	2.33	11.78	2.92	1.71	1.14	0.38	86.57	13.43
CNN	9.53	4.68	6.94	26.19	2.66	12.10	5.52	24.80	0.87	2.43	1.99	2.28	67.70	32.30
FBC	4.40	29.16	5.91	7.59	3.18	12.10	2.62	29.10	2.03	1.89	1.49	0.53	76.83	23.17
FOXNEWS	7.87	5.35	7.36	17.95	3.36	14.26	4.21	33.94	1.12	1.57	1.47	1.55	69.52	30.48
MSNBC	4.75	4.26	6.30	13.07	5.27	21.78	3.37	36.16	0.58	1.19	1.82	1.48	66.24	33.76
NBC	20.79	5.23	4.74	18.98	0.93	4.40	15.20	9.81	5.62	2.89	3.16	8.25	42.58	57.42
NPR	12.42	8.27	9.64	16.95	1.27	3.78	10.67	26.28	1.77	5.59	3.11	0.26	68.00	32.00
PBS	9.25	18.02	8.24	11.57	1.69	5.71	8.34	29.69	1.08	4.34	1.84	0.23	73.55	26.45
Overall	9.75	13.81	6.21	19.53	1.91	7.67	7.26	23.54	2.02	3.17	2.41	2.74	68.17	31.83

NOTE: Based upon top-3 probabilistic predictions over all snippets. Hard and soft are super-sets of the topics. Overall is the mean frequency across stations.

Table 2.9: Estimated supplemental category frequency in news (percents)

	Geography			Fact, investigative				Tone			Emotional appeal		
class network	domestic	foreign	unknown	fact, investigative	fact, non-investigative	opinion	other	negative	neutral	positive	neither	pleasant	scary
ABC	73.79	18.51	7.71	3.72	64.55	31.07	0.66	24.24	54.38	21.38	75.76	9.51	14.74
ALJAZAM	42.21	52.58	5.21	16.09	41.70	41.46	0.75	31.85	60.70	7.44	79.83	2.90	17.27
BBC	26.48	68.77	4.75	16.32	46.35	36.94	0.39	29.03	63.57	7.41	79.44	2.92	17.63
BLOOMBG	59.48	32.71	7.80	0.64	38.32	60.30	0.74	21.77	64.52	13.71	93.89	2.90	3.21
CBS	71.95	22.05	6.00	4.76	54.66	39.83	0.75	25.16	57.90	16.93	79.28	7.60	13.13
CNBC	75.11	15.78	9.12	0.28	35.21	63.81	0.70	22.72	63.64	13.64	95.44	1.97	2.60
CNN	69.38	23.82	6.80	4.92	34.02	60.14	0.92	31.56	59.56	8.89	84.39	3.40	12.21
FBC	79.53	13.93	6.53	0.82	22.73	75.83	0.63	35.33	56.49	8.18	91.98	1.70	6.31
FOXNEWS	77.04	17.16	5.79	2.34	25.73	71.17	0.76	37.26	55.73	7.01	86.89	2.38	10.73
MSNBC	83.10	12.76	4.13	1.84	20.97	76.32	0.87	34.22	59.44	6.33	90.55	1.89	7.56
NBC	74.98	16.75	8.27	3.32	52.43	42.96	1.30	22.65	51.57	25.78	77.68	10.71	11.62
NPR	64.95	30.91	4.13	7.56	43.97	47.52	0.94	25.10	60.68	14.22	83.08	8.24	8.68
PBS	64.76	28.93	6.30	4.31	30.50	64.01	1.17	26.85	59.60	13.55	87.71	4.99	7.30
Overall	66.87	26.86	6.27	5.15	40.06	53.96	0.83	28.45	58.64	12.91	84.43	4.99	10.59

NOTE: Based upon predictions for over all snippets using top class for fact, investigative, top-2 probabilistic for all others classifiers. Overall is the mean frequency across stations.

Coverage of government affairs and politics is lowest among the broadcast stations, particularly consistently ABC and NBC at 9.8 percent each, as well as CNBC and Bloomberg (11.8 and 16.4 percents respectively; CBS has 16.1 percent coverage). Perhaps the broadcast networks are trying to avoid this potentially controversial subject area, given their larger and more diverse audiences? FBC has no such qualms; as noted previously it includes a similar amount of government and politics as the broader interest cable news networks at 29.1 percent. The cable networks MSNBC and Fox News stand out for their high percentage of coverage, with 36.2 and 33.9 percents respectively. They, by contrast to broadcast networks, seem to be catering to a smaller, more tailored, audience less averse to controversial topics.

Social issues and cultural coverage (not including current events), including themes such as communities, cities, health, society, religion, gender, race, human interactions or other cultural trends, is a category that tends to be more reflective and detached from the day's events. These topics are covered most by the public networks, NPR and PBS (at 9.6 and 8.2 percents, and with both including more as time goes on), as well as ALJAZAM (9.7 percent). The cable news networks also cover more than average, with 7.4 percent for Fox News, 6.9 percent for CNN, and 6.3 percent for MSNBC.

No network spends much of their time covering science, technology and the environment. The largest shares are for NPR (5.6 percent), ALJAZAM (5.5 percent), PBS (4.3 percent), CBS (4.3 percent), and BBC (4.1 percent). Over the period of study, CNN includes less and less such coverage, while ALJAZAM shows a strong increasing trend.

While not heavy on government, business and economics, the broadcast networks instead compensate by including more lighter fare. Anecdotal and human interest stories account for 20.8, 16.6, and 12.9 percents of coverage for NBC, ABC, and CBS. The next highest are 12.4 and 9.2 percents for the public stations NPR and PBS. In addition, NBC and ABC include 15.2 and 14.3 percents coverage of entertainment, music, arts and literature. NPR, CBS, and PBS are not far behind with 10.7, 10.3, and 8.3 percents respectively. NBC and ABC also lead the way with their coverage of consumer affairs, products and services, with 5.6 and 3.5 percents, probably mostly accounted for by their morning news shows. The next highest are CNBC and

CBS with 2.9 and 2.7 percents. Broadcast networks are also the only ones to include regular breaks for local traffic and weather (at least in the mornings, and accounting for 6.7 to 8.2 percent of coverage).

2.4.2 Election coverage and hard news

In Figure 2.3, I show time trends in topic coverage. These trends represent topic averages with equal weight given to all general-interest networks. Business networks are excluded because they only become available partway through the study period, but they would also change the distribution because of their greater coverage of the business and economics topic. Their removal allows a more balanced and consistent picture of overall coverage over time.

The most noticeable trend is the overall increase in election-related coverage in the run up to the 2012 and 2016 Presidential elections. This election coverage comes at the expense of current events (especially) and other government and politics coverage. Cultural and business & economics coverage are also sacrificed, to a lesser extent, to make time for electoral coverage. Interestingly, the Congressional elections in 2010 and 2014 are but a blip compared to the Presidential election years. This can be more clearly seen in Figure 2.4, where we break down electoral coverage over time and by station. Substantive election coverage is shown in Figure 2.4a, while soft coverage (mainly meta analysis: coverage of the horse race and campaign tactics) is shown in Figure 2.4b. The total percentage of electoral coverage increases from an average of around 4 percent in early 2011 and 2015, to a height of around 19 percent during the primaries of early 2012, and over 25 percent during the primaries of early 2016. Coverage remains high for the remainder of 2012 and 2016. 2016 was even more heavily covered than 2012. By contrast the Congressional election years 2010 and 2014 max out at 8 percent and 5 percent total coverage in the quarter of the elections themselves. Since there is no nationwide election contest in the midterms, this may be for lack of a compelling national 'horse race' narrative to latch onto as there is in Presidential election years (for that is what most election coverage comes down to).

Overall, coverage of elections is overwhelmingly non-substantive (soft), i.e., unrelated to

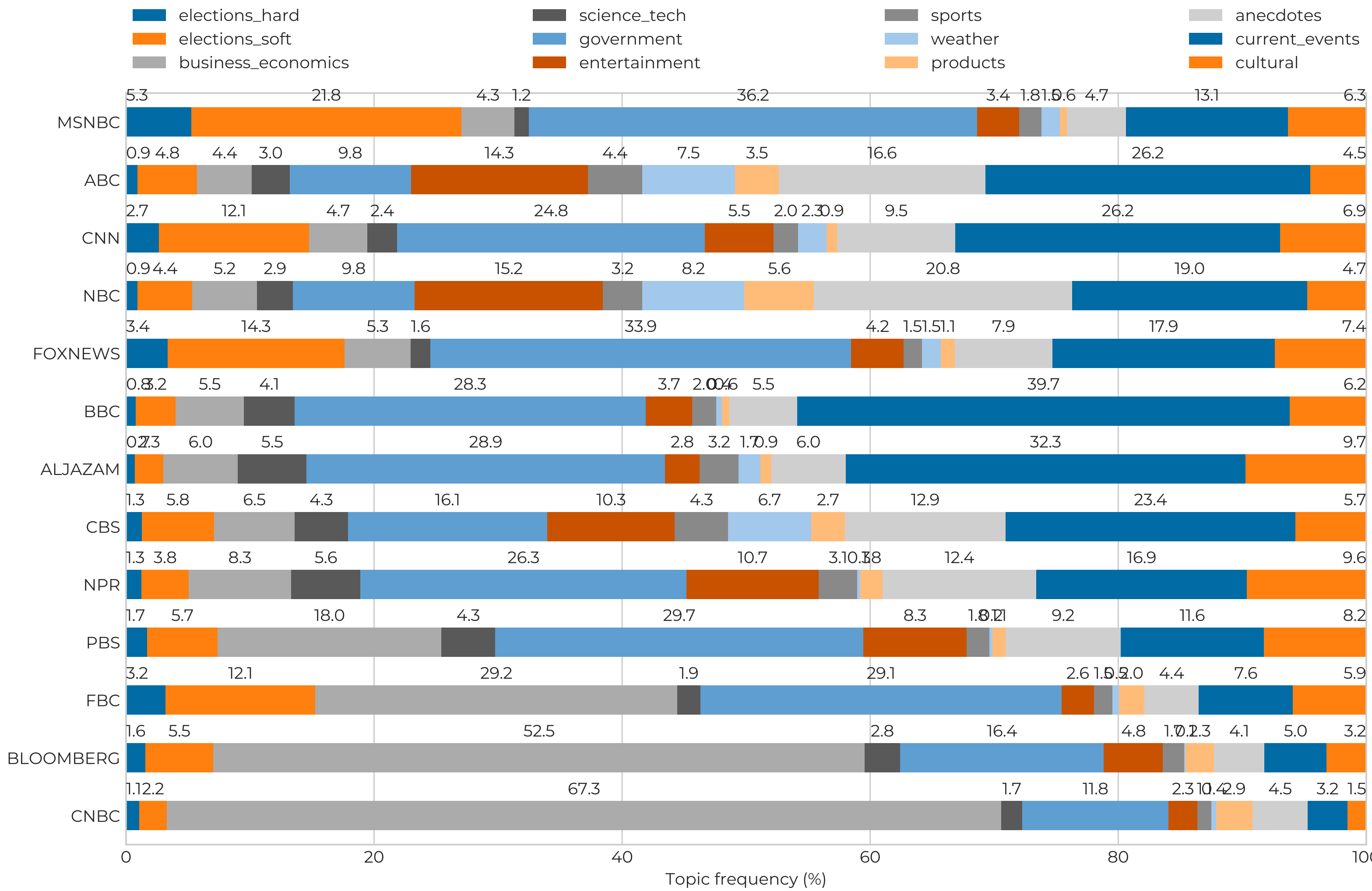


Figure 2.2: Topic coverage by network, ordered by business & economics coverage

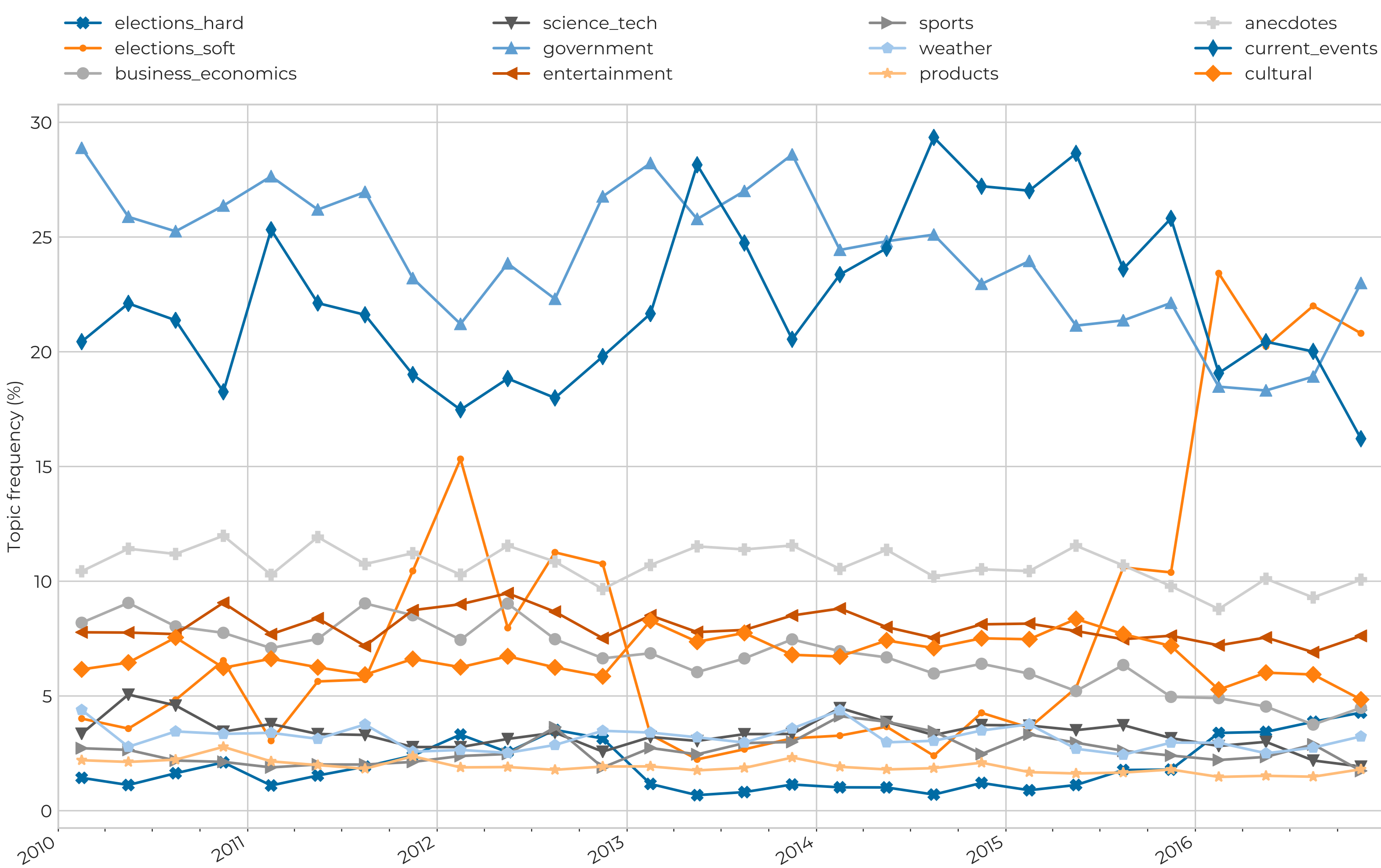


Figure 2.3: Topic coverage over time, excluding business networks (weighted by station)

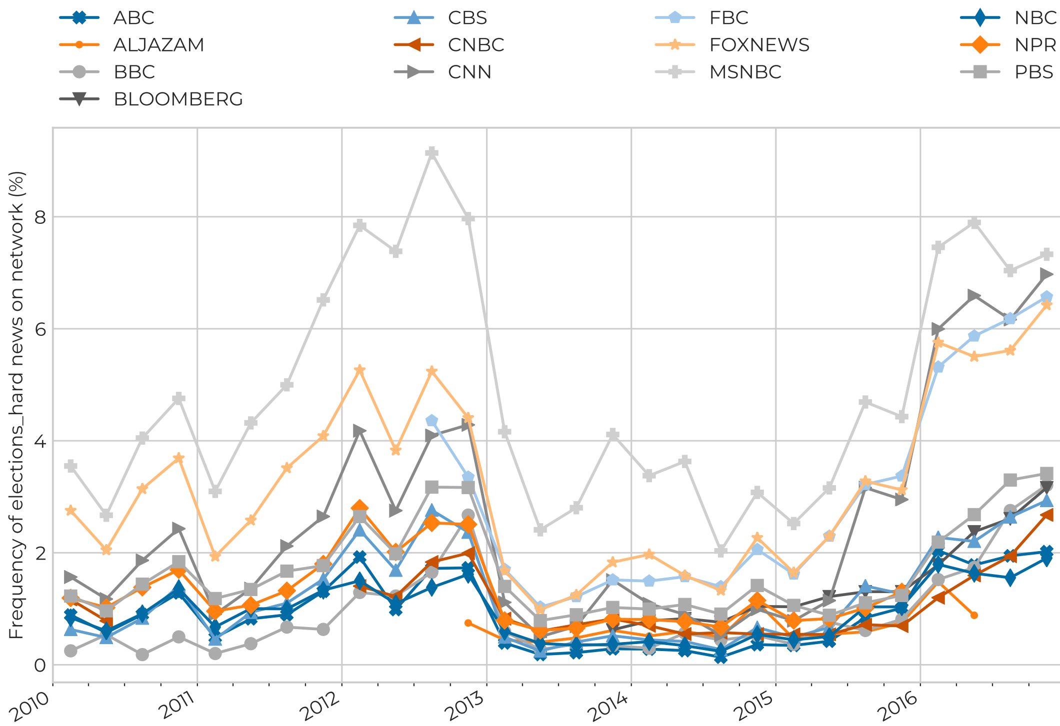
candidates' platforms, backgrounds and speeches. Overall, more than four times as much electoral coverage is soft as it is substantive (a ratio of 8.23 to 1.93 percents). No station's substantive election coverage exceeds 10 percent for any quarter, while for one quarter soft election news exceeds 50 percent for MSNBC. This is in line with what the public editor for The New York Times (an outlet known for its serious news coverage) found in reviewing two weeks of their election coverage in February 2016: 77 percent of articles could be characterized as being about the "horse race" (Sullivan, 2016). During the heart of the general election, the relative fraction of substantive news is even lower. Over the last three quarters of 2016, there is 5.4 times as much non-substantive election news as substantive (21.01 vs 3.86 percent). This means much more attention is being paid to superficial aspects of campaigns, electability, and campaigning than to the candidates' records, plans and proposals, or even the candidates' words themselves. This is true for all stations and at all times.

MSNBC provides the most election coverage, and has the most variation in its quantity of electoral coverage over time. Not far behind are the other cable news networks CNN and Fox News, as well as FBC. CNBC provides the least electoral coverage, and the most proportionately substantive.

Since election coverage is predominately soft news, election periods (primarily Presidential) crowd out hard news. As already noted, the frequency of current affairs and 'government and politics' topics decrease as election coverage increases. CNBC and Bloomberg are the only networks mostly unaffected by this crowding out, while the cable networks are particularly prone to it.

By our minimal definition of 'hard' news (based as it is only upon topic), there was a wide variation between networks in how much hard news was covered (see Table 2.8). It's not surprising that the business networks CNBC and Bloomberg have among the top proportions (86.6 and 81.6 percent) of hard news content, since 'business and economics' news counts as hard. More interestingly, BBC and ALJAZAM have a similarly high quantity (84.6 and 83.1 percents). By contrast, the broadcast networks have notably more soft news. The hard news proportions for NBC, ABC, and CBS were 42.6, 48.9, and 57.2 percents. The remaining networks range

(a) ‘Substantive’ election coverage



(b) Non-‘substantive’ election coverage

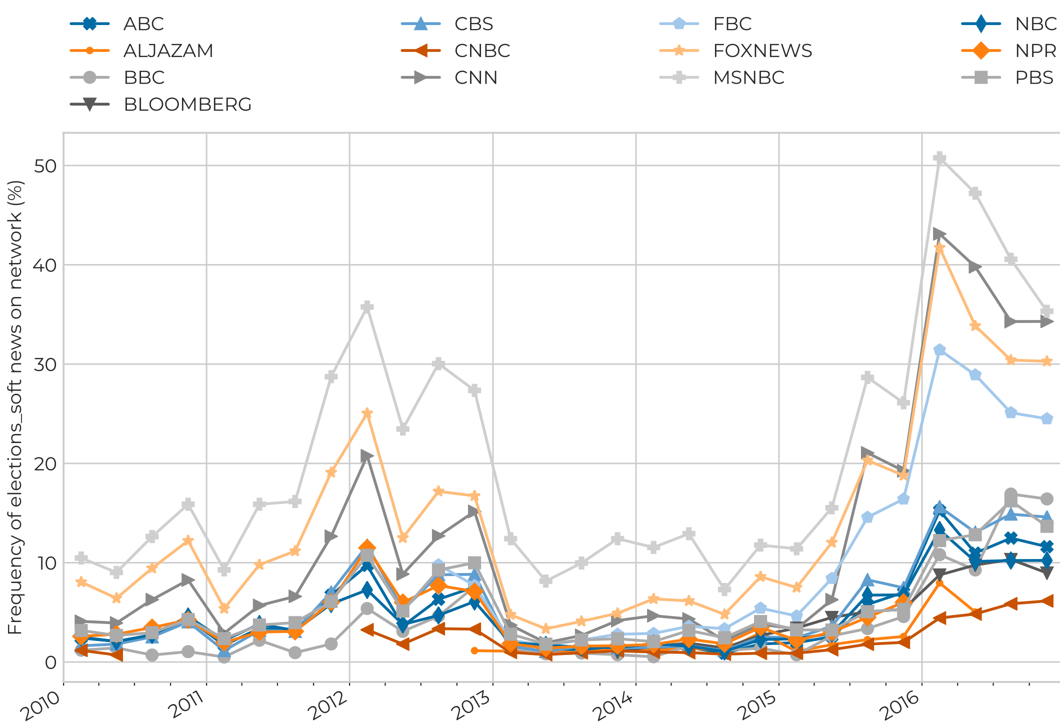


Figure 2.4: Election coverage over time by Network

from the mid 60 to the mid 70 percents.

2.4.3 Supplemental Categories

Next we will examine the supplemental dimensions of quality beyond topic. These dimensions were only assessed when the topic indicated that it was warranted. For example, if the topic was 'Weather and Traffic', no other dimensions were assessed. If it was 'Sports and games', we only assessed whether it was foreign or domestic, and the tone (positive/neutral/negative).[12] So the amount of relevant coverage applicable to these dimensions of quality differ by dimension depending upon the networks' topic distributions. But the results are for a fixed set of topics, so they are fully comparable between stations. Results for the supplemental dimensions are shown in Figure 2.5.

Figure 2.5a shows the geography of news coverage by network. BBC and ALJAZAM stand out for the most foreign coverage at 72.1 and 54.5 percents respectively. All the other networks cover predominately domestic matters – the next most foreign coverage is from Bloomberg, with 35.6, then PBS, CNN, and NPR with 35.6, 33.3 and 32.0 percents respectively. Other networks cover domestic affairs in excess of 68 percent of the time.

I cover the presentation of news in Figure 2.5b. Here there is a wide variation across stations. MSNBC, FBC, and Fox News' coverage is dominated by editorial, opinion, conjecture, and speculation, amounting to 76.3, 75.8, and 71.2 percents of relevant coverage, respectively. On the other hand, ABC's relevant coverage is 68.2 percent fact-based, followed by BBC, CBS, and ALJAZAM with 62.7, 59.5, and 57.8 percents. The cable news networks also experience an increase in opinion-based coverage during the Presidential elections, a strong indication that proportionally more of their election coverage is opinion than in their news overall (as we will later confirm). In all cases, non-fact-based, non-opinion news is negligible.

BBC and ALJAZAM are responsible for by far the most fact-based in-depth and investigative reporting of any station, proportionally speaking. For BBC 16.3 percent, and for ALJAZAM 16.1

[12] While we do not directly assess the geography of US Election coverage, it is domestic by definition, and as such is included in the domestic counts.

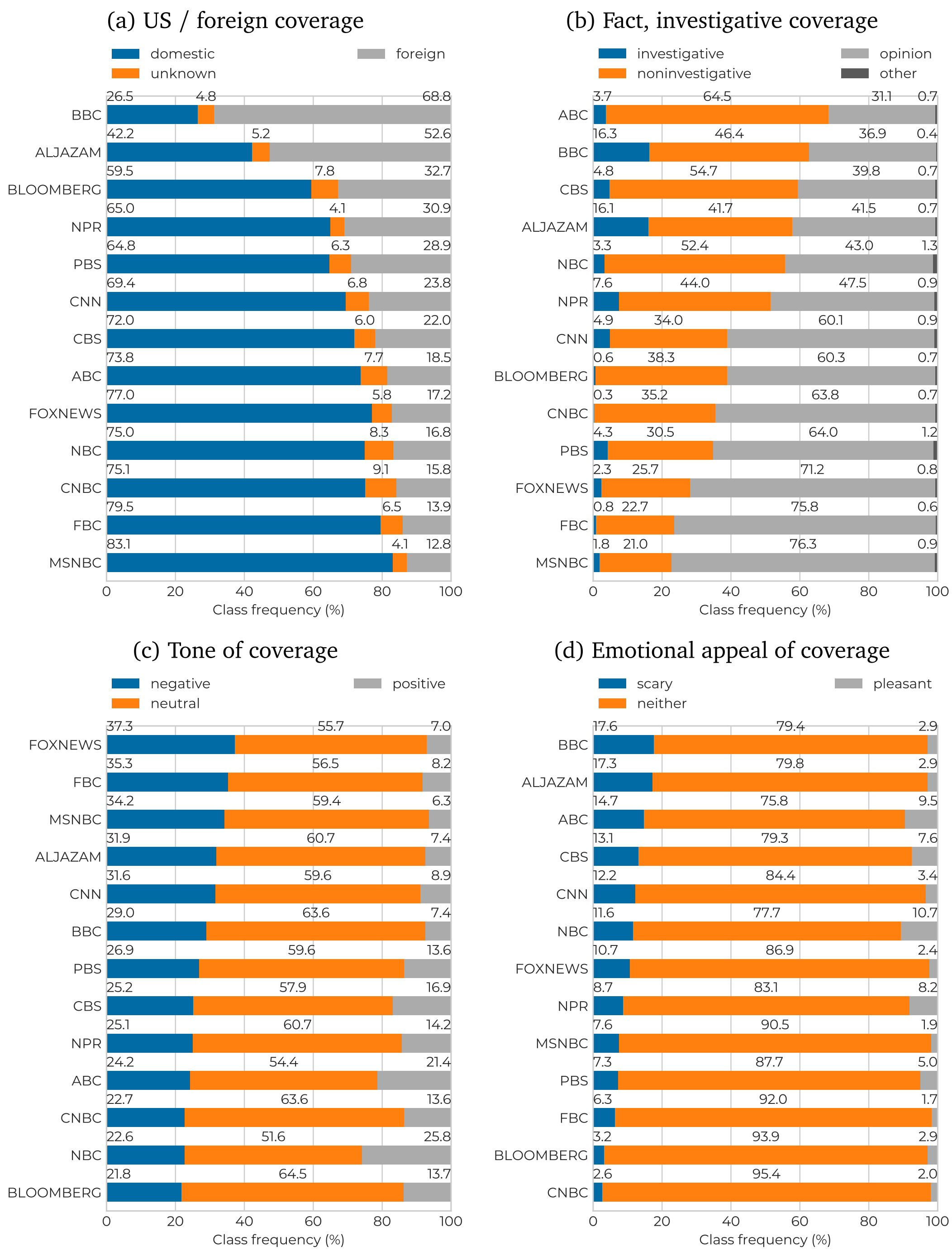


Figure 2.5: Supplemental category coverage by network

percent of their relevant reporting is investigative. The next highest are 7.6 percent for NPR, 4.9 percent for CNN, 4.8 percent for CBS, and 4.3 percent for PBS. The business networks produce almost no investigative reporting, and MSNBC and Fox News produce less than half that of CNN, which is also less than all the broadcast networks. Along with FBC, MSNBC and Fox News are really dominated by opinion, with seemingly little budgets for investigative reporting.

Fox News, FBC, and MSNBC are also responsible for the highest proportion of negatively-toned coverage among stations (Figure 2.5c), with 37.3, 35.3, and 34.2 percents of relevant coverage being negative, respectively. The broadcast networks lead the way with the most positively-toned coverage. NBC's relevant coverage is 25.8 percent positive – the only network with more positive than negative coverage. ABC is 21.4 percent positive and CBS's is 16.9 percent. ABC and NBC are producing modestly more positive-toned coverage as time passes, while CBS is producing modestly less. At Bloomberg the quantity of positive coverage has been on a significant downward trajectory since the data begins. Overall, networks present 2.2 times as much news negatively as positively, although all the stations present a majority of their coverage in a neutral tone. While there is not too much variation in the extent of networks' negative coverage over time, there is a notable spike in late 2013 around the Affordable Care Act roll out, especially among the cable news networks and FBC.

Finally, I summarize the emotional appeals in news coverage in Figure 2.5d. BBC and ALJAZAM lead the way with the most content which is deemed scary, shocking, or outrageous (17.6 and 17.3 percents). Somewhat surprisingly, they are followed by ABC and CBS with 14.7 and 13.1 percents of relevant coverage overall. While the foreign networks and the cable networks have relatively high degree of negative emotional appeals compared with positive emotional appeals, the broadcast networks and to a lesser extent the public networks balance negative emotional coverage with additional pleasant coverage. NBC has the most pleasant coverage with 10.7 percent, nearly as much as it has scary coverage (none has more pleasant than scary). NBC is followed by ABC and CBS with 9.5 and 7.6 percents, both with ratios of scary to pleasant news of less than 2:1, whereas BBC has a ratio of over 6:1. On average, the ratio of scary to pleasant news is 2.2:1, and there is little variation over time. As expected, most coverage is not overtly emotional, and CNBC and Bloomberg have particularly few stories with an emotional valence. The cable networks also have less scary / shocking / outrageous coverage than one might have expected *ex ante*.

2.4.4 The world is a scary place

It seems that the there is a relationship between the amount of foreign coverage, the amount of current events coverage, and the amount of factual coverage (see Figure 2.6). There seems to be a clear covariation between the time series of these aspects of news coverage.

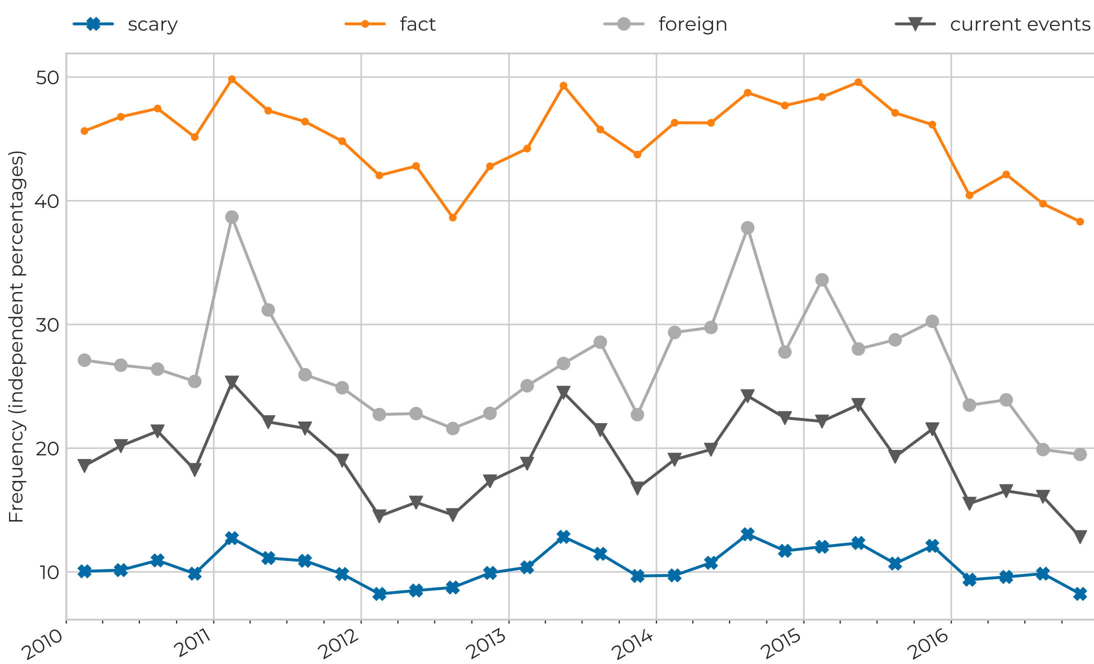


Figure 2.6: Fact-based, foreign, current events, and 'negative' emotional coverage over time

There were notable spikes in the amount of foreign coverage in early 2011, which coincided with the Arab Spring, and in late 2014-15, when the US returned to Iraq to fight Islamic State, and the Charlie Hebdo attacks occurred in Paris. The spike in factual and scary news in the second quarter of 2013 coincides with the Boston Marathon bombing, and the first use of chemical weapons use in an intensifying Syrian civil war.

Although speculative, the most likely causal mechanism for these co-movements seems to be that foreign coverage is most likely to increase when there are current events abroad that the networks deem worth covering. This is more likely when the events are scary, shocking, or outrageous. And coverage of current events tends disproportionately to be factual compared with coverage overall.

Thus the coverage that the public sees of the rest of the world largely tend to reinforce the view that the outside world is a scary place.

2.4.5 Hard topics interacted with supplemental categories

A further analysis was performed to help determine how the supplemental categories interact with the topic categories. This required tracking the topics and the various supplemental categories examined on a snippet-by-snippet level, which is memory intensive. Thus, only certain cross-products were examined. In particular, I looked only at snippets which were classified as most likely to fit in a hard topic. I then examined three subsets on a per-topic basis – 1) the tone by topic, 2) fact (investigative or non-investigative) / opinion by topic, 3) on the subset of snippets that were also classified as having scary / shocking / outrageous emotional appeal, the geography (US / domestic) by topic. These findings are summarized in Table 2.10, averaged over stations and time.

This analysis allowed us to confirm that current events are by far the most likely topic category to have content labeled as scary / shocking / outrageous – 94.4 percent of snippets coming from hard topics with this emotional valence were current events. Of those, 47.8 percent are foreign, against 42.3 percent domestic. This compares with 26.9 percent foreign and 66.9 percent overall. Current events are also much more likely to be fact-based than average (75.3 percent compared with 45.2 percent of all relevant news). Altogether, this gives some explanation for the finding of co-movements of scary, factual, foreign, and current events in the news as shown in Section 2.4.4; news presented as scary is 1.78 times more likely to be foreign than overall, and current events are more likely to be scary and factual.

Next, looking at findings about tone by topic category, we find that certain categories are

Table 2.10: Estimated supplemental category frequency by hard topic category (percents)

hard topic interacted class	*(overall prevalence)*	*(all hard)*	business & economics	cultural	current events	elections-hard	government	science & tech
(overall prevalence)		68.17	13.81	6.21	19.53	1.91	23.54	3.17
				Tone				
negative	28.45	31.45	24.29	36.32	32.87	31.95	35.78	25.48
neutral	58.64	62.75	63.24	51.44	65.13	64.58	61.98	61.15
positive	12.91	5.80	12.47	12.24	2.00	3.47	2.24	13.37
				Fact, investigative				
investigative	5.15	5.58	0.10	2.24	13.88	0.00	2.14	5.90
noninvestigative	40.06	40.80	43.80	16.41	61.44	23.85	25.22	33.31
opinion	53.96	52.58	55.17	79.90	23.55	75.98	71.71	57.71
other	0.83	1.04	0.93	1.45	1.13	0.17	0.93	3.08
				US/Foreign among scary / shocking / outrageous				
domestic	66.87	42.31	67.08	64.65	42.45	58.88	18.74	39.06
foreign	26.86	47.81	15.85	18.15	47.97	34.69	78.83	17.45
unknown	6.27	9.88	17.07	17.20	9.57	6.43	2.43	43.50

NOTE: Based upon all snippets which were predicted to be in a hard topic category. Percents are means across stations of the percents in the whole time period for each station. "Overall prevalence" is the percent of news coverage falling in that category overall, from Tables 2.8 and 2.9, presented here for comparison purposes.

much more likely to have a positive tone: science & technology (13.4 percent), business & economics (12.5 percent), and cultural (12.2 percent) are all much more likely than other hard categories to be presented with a positive tone. Meanwhile, cultural (36.3 percent), and government & politics (35.8) are the most likely to be presented negatively, followed by current events (32.9 percent). Cultural is thus by far the most likely topic to be presented non-neutrally. Current events, government and politics, substantive elections coverage, are particularly unlikely to be presented in a positive manner. Overall, 31.5 percent of hard news was presented in a negative manner, compared with 28.5 percent of all news for which tone is assessed, while 5.8 percent is presented positively, compared with 12.9 percent overall. The Fox networks are outliers in their tone of presentation for hard news topics. They have the most negative coverage about government & politics, at 45.4 and 42.7 percents for FBC and Fox News. They also have the most negative (over 45 percent) and the least positive (less than 7.5 percent) cultural news. Finally, Fox News stands out with by far the most negative coverage of business & economics during the study period (at 35.1 percent).

Finally, and perhaps most importantly, we find that the topics that are most important to citizen-voters in their civic capacity – substantive election and government & politics, are predominately covered with presentations of opinion rather than fact. They are 76.0 and 71.7 percent opinion on average, respectively. Cultural and societal issues coverage, which is arguably the next most important, is even more dominated by opinion (79.9 percent on average). Substantive, investigative reporting relating to candidates and elections is so rare as to be nearly undetectable. Some stations are even more opinion-dominated - CNBC, FBC, and MSNBC's 'substantive' election coverage is over 84 percent opinion, followed by Fox News and CNN (82.0 and 81.2 percents respectively). ABC, BBC, NPR, and CBS have relatively less opinion (and more fact-based) coverage of substantive election topics (from 66 to 68 percent opinion, so still a majority). There are similar variations among stations for the government & politics topic. As noted previously, current events coverage by contrast is the most likely to be fact-based generally, and the most likely to be in-depth and investigative (13.9 percent). Overall, hard news topics are 52.9 percent opinion, calling into question whether it is even right to call them that.

If we were to exclude opinion coverage from hard news topics, then only 32.1 percent of news coverage overall would fit the description, and only 0.53 percent of coverage is substantive, non-opinion electoral coverage.

2.5 Conclusion

This is but a start in related research, so cannot be considered definitive. While there is some uncertainty as to the precise levels of any of the classes for the different dimensions reported here due to potential per-class biases[13], it is quite likely that the cross-station and over-time comparisons I have reported will hold up, as I have no reason to believe that any biases would be dependent upon station or time. It is also important to note that all assessments, though intended to be systematic and non-normative, are nonetheless based upon human judgment in their initial labels of categories. In particular, I can only assess whether coverage is presented as fact-based, not whether it actually is factual, due to the near-impossibility of determining the latter on such a large scale.

While the prior chapter showed that networks are not particularly slanted in a partisan sense, this one showed that they may also not be particularly informative as a whole.My analysis reveals substantial differences between the stations, but also significant consistencies. Broadly, the broadcast networks appear to be relatively averse to controversy, with low levels of coverage of government and politics, with the most soft news, some of the least opinion, and with balanced emotional appeals. This is significant because the broadcast networks retain by far the highest viewership of the television networks. Of the three, CBS has the most factual and investigative reporting, and most hard news reporting more generally. The cable networks, particularly MSNBC and Fox News, present predominately opinion, largely on domestic matters. They have some of the most coverage of government and politics, as well as elections, but as for other networks, the vast majority of the election coverage is not substantive, and these topics

[13]In particular, it is likely that I am overestimating the quantity of current events coverage, the amount of factual coverage relative to opinion, the amount of domestic coverage rather than unknown/global/neither. It is likely we're underestimating the amount of positively-toned news, and the amount of both scary/shocking/outrageous and pleasant content.

are even more likely than others to be presented as opinion. The public networks NPR and PBS are the most balanced and lie squarely in between the cable and the broadcast networks along most dimensions. They have wide-ranging coverage of topics, above-average foreign and investigative coverage, fairly balanced emotional appeal and middle of the road tone.

The business networks have a notably different distribution of topics covered (with business and economics coverage dominating), although on many dimensions, Fox Business more closely resembles the cable news networks in its coverage than CNBC or Bloomberg. It substantially sacrifices its business and economics coverage to cover elections, for example, and is the second most negative station overall, whereas the other business networks are the two least negative. It also has the second-most opinion content of any network, (a solid 15 percent more than the other business networks) and is trending upwards.

Only the international networks, BBC and the now-defunct Al Jazeera America, appear to systematically present significant quantities of factual, hard news coverage, or broad coverage of news beyond US borders. They dominate in the proportion of in-depth and investigative reporting, with no others close. This is particularly impressive for Al Jazeera, because they were a 24-hour news network; all the other such networks resort to large quantities of opinion to help fill the time.

In short, I have demonstrated numerous tendencies (one might call them biases) of television and radio news. Setting aside the business networks, 35 percent of coverage was deemed soft news by our weak standards, including 80 percent of all election coverage. Fully two thirds of relevant coverage was domestic, and that doesn't include most of the soft news. It would be 72 percent if we left out BBC and Al Jazeera. Over half of relevant coverage is opinion, and the same is true for what we count as hard news topics, indicating that much of hard news is perhaps not so hard after all. And only four percent of news can be considered in-depth, investigative, or providing significant history, background, or context. Put more succinctly, the news is largely parochial, superficial, non-authoritative, and myopic.

There are possible reasons for all of this. Domestic news is arguably more relevant for most viewers, as generally what is more local tends to be, and it's quite likely more interesting for most viewers. But the US is only a small part of a large world, and viewers of most stations are clearly not getting a representative sample of world news. The "news" is unsurprisingly focused on what is new and happening now. But for the vast majority of people who don't make it their habit to follow the details of what is happening every day, this leaves them with a weak understanding of where the the day's events are coming from. So there being so little reporting that provides fact-based context and background can leave viewers uncertain what to make of what they see. In a fast-changing environment, and with ever-shorter deadlines, it can be hard for journalists to be definitive and get to the bottom of the most recent developments. And it can be hard to produce deep dives on issues and try to disentangle opposing viewpoints; these are expensive and take much longer to produce than to show. Not to mention that the required complexity may necessitate more focus than most viewers are willing to give. It's much easier to get experts (or just pundits) to hash it out on the air, even if they're often just themselves given time to state opposing opinions without much depth. And media everywhere are facing a harsh competitive environment and are in many cases barely profitable. But such unreflective presentations seem to do little to bridge divides and can even amplify preexisting prejudices by simply providing viewers with further internal justifications for them.

Given the extent that the broadcast news networks focus on soft news, and the extent that cable news networks focus on opinion content, it seems clear that at least the cable news networks and the broadcast networks, operate largely as a form of entertainment (although for different market segments). Perhaps this is better than the alternative, which for many might be no news at all (Baum & Jamison, 2006). But that is no great consolation either. Many tune into these programs as their primary source of news, hoping to be well informed, but are likely being left wanting.

Considering the state of the news providers that reach the most Americans, it should be little surprise that most Americans have little to no trust in their government (Gallup, 2019). How can citizens be expected to pick a good democratic government when they are given such poor information with which to choose? How can governments and candidates be compelled to live up to a higher standard when they know how superficial television and radio news is, and how easy it can be to spin things in an environment of general ignorance, or worse bias-reinforcing outlets and algorithms (Sunstein, 2001)? And how can citizens be expected to be confident in a globalized and fast-changing world when so little coverage is given of the outside world, and of science and technology? Obviously, there remain other sources of information out there. But it is the median voter who counts most in a democracy, and he gets his news primarily from these networks.

Chapter 3: Executive Compensation under Concentrated Ownership

3.1 Introduction

In the aftermath of the Great Recession of 2007-9, followed by a prolonged period of high unemployment, and in the midst of ever increasing inequality between top and middle earners (Gordon & Murphy, 2007), there has been much interest in the compensation packages of top executives. While the literature on this issue is voluminous, much of the work was done prior to the recession, an event that might be expected to discipline the pay of even the highest earners. This chapter provides a snapshot in the aftermath of the recession, examining a large, novel data set of executive compensation data for 2009. In particular it provides support for one tentative new hypothesis in the controversial and unsettled question of what drives executives' level of total compensation, and in particular their incentive pay.

Broadly, there are two conflicting theories about the motives for executive compensation contracts. One theory explains high-powered incentives as a solution to a principal-agent problem between owners and managers, which seeks to align managers' interests with those of the owners. The second explains these incentives as a form of coordination problem among owners which is exploited by powerful managers, in order to increase their potential pay package. Other things being equal, one would expect a more concentrated ownership group to suffer from less of a coordination problem and be more motivated to improve company profitability, and thus to choose a more closely optimal employment contract for its executives. This paper tests the effect of ownership concentration on incentive compensation as a share of total compensation for company CEOs and other executive officers.

While one might expect concentrated ownership would affect pay similarly regardless of the identity of the owners, this is not what I find. Instead, my results indicate that companies largely controlled by institutional investors choose to give much more incentive pay, and higher pay in general, to executives than those that aren't. The effect is reversed for high non-institutional ownership stakes. This may be because institutional investors have superior knowledge about how best to motivate executives to improve shareholder-value. Alternatively, institutional investors may be more likely to be passive or short term investors, not seeking seats on the board or overt influence on company operations. Or, more perniciously, it may be that investment managers fail to restrain the pay packages of company executives because they themselves find little incentive to do so. They may view the executives as peers for which such discussions are not collegial; they may find little incentive to do so since the money they're investing is not their own; or they may even view it as against their self-interest, when they themselves often have similarly constructed pay packages.

This paper begins with a quick review of the literature and a review of the data we examine. Next it examines the tests of the core hypothesis, and finishes with some concluding remarks.

3.2 Literature and Motivation

In his good early survey which this review draws upon, Murphy *et al.* (1999) noted that the explosion in executive compensation levels has been surpassed only by the proliferation of research papers on the subject. The trend has surely not abated. The central question I am addressing is what is the optimal makeup of a compensation package for executives. This is not a simple question.

As noted before, the traditional explanation for incentive pay, as opposed to base pay (salary), is that it can serve to align the interests of executive managers with that of owner/shareholders, alleviating a principal-agent problem. Without incentive pay, executives are not explicitly rewarded for increasing shareholder value, and will likely put a minimum of effort to this objective, while instead pursuing ends which improve their individual utility, perhaps replacing the carpeting in the executive office suite, empire-building, and retaining earnings when there are no profitable investment opportunities. As such, incentive pay can be an alternative to direct monitoring.

At the other extreme, placing too strong of an incentive on imperfect proxies of the ultimate goal (e.g., shareholder value) can encourage manipulation to meet objectives, and may ultimately damage that goal in the long term. Even just rewarding based upon outcomes rather than behaviors can be counterproductive. Additionally, extrinsic incentive rewards may overpower intrinsic motivations to do what's right for shareholders (Frey & Osterloh, 2005). Clearly, we should expect there to be some difference across industries as well, since for example firms relying heavily on R&D will want to reward different things than, say, utilities.

There are various sorts of incentive pay, each with its own advantages and disadvantages. Bonus schemes which are based upon accounting measures are inherently backward looking and short run. This may cause managers to avoid actions that reduce current profitability at the expense of future profits, or manipulate the figures outright, by shifting discretionary expenses or accruals through time. If next year's bonus is based on this year's output, it may cause CEOs to avoid actions that might be detrimental to next year's budget. Bonus schemes tend also to have nonlinearities, such as a minimum threshold to qualify, and maximum payouts, which can have perverse effects when executives are near these thresholds.

The awarding of options are also potentially troublesome. Granting options with fixed exercise prices (as is usually done), rewards or penalizes executives for things that are outside their control (e.g., movements in the general stock market). They also penalize the payment of dividends, which means owners of options do not have the same incentives as shareholders, encouraging short-term thinking and risk taking. And when the stock price falls sufficiently below the exercise price they lose their incentive effects altogether, as during a market correction. Risk-averse executives should always prefer cash pay rather than option pay, unless their expectations are inflated. Indeed, a reasonable model by Dittmann & Maug (2007) finds that "most CEOs should not hold any stock options. Instead CEOs should have lower base salaries and receive additional shares in their companies."

Additionally, options seem suboptimal since they cost companies more than they are valued by (typically) undiversified executives, who are constrained in their ability to hedge by legal restrictions on short selling their company's stock (Hall & Murphy, 2002). Part of the explanation for their prevalence lies in a favorable tax treatment for companies and executives, and in accounting rules which before 2006 allowed companies to avoid any accounting charge for granting such options. But Jensen *et al.* (2004) notes that "managers should be held accountable for factors that are beyond their control if they can control or affect the impact of those uncontrollable factors on performance" as options do. Furthermore, Bolton, Scheinkman & Xiong (2006) create a model in which investors have heterogeneous beliefs and stock prices have a speculative component arising from a desire to potentially sell to overoptimistic investors (as bubbles form). In such a world "optimal compensation contracts may emphasize short-term stock performance, at the expense of long-run fundamental value, as an incentive to induce managers to pursue actions which increase the speculative component in the stock price," even if this may not be in the interests society writ large, or even the company itself in the long term. So the overall picture of options isn't as clear as one might hope and it is unclear what the optimal options component of executive compensation is after considering all these effects.

As we've seen above, self-interested managers who are large stockholders in their company will by definition do what they think is good for stockholders. Thus awarding managers shares of the company would appear to be an optimal way to align the interests of owners. Even this however has some downfalls; as company ownership increases as a portion of the executive's personal wealth, additional ownership will be valued at less than face value by the executive, meaning companies could likely find cheaper ways to provide an equivalently valued level of remuneration.

The question is what is the right balance of fixed salary versus incentive pay. Since the theoretical balance is unclear, it would seem to call for empirical study of what is done. In efficient labor markets, we would expect observed pay to be optimal. Unfortunately, it is not at all clear that pay is actually set in an optimal way, since the market for chief executives is far from a perfect market.

Bebchuk & Fried (2005, 2003) argue that 'managerial power' sways the pay packages of CEOs and other top executives in their favor, causing the high pay and high incentive pay that is observed. They argue that company directors have had economic incentives to support or at least go along with arrangements favorable to a company's top executives. This may be complemented by social and psychological factors such as collegiality, team spirit, desire to avoid conflict within the board, and sometimes even friendship and loyalty. According to this theory, executives may succeed in rent-seeking for better pay packages because directors' financial incentives are too weak to induce personally costly/unpleasant route of haggling too hard with CEOs. As a result, the primary constraint on directors' actions (and hence executive pay packages) is in "outrage costs" – how the arrangement is perceived by outsiders whose views matter to directors and executives. The risk is that these outsiders may cause embarrassment or harm reputations, and cause shareholder pressure to emerge. Thus the theory predicts we should see a high degree of camouflage in pay, and that the greater the CEO's power over the board, the larger the CEO's rents will be.

There has been a great deal of evidence for this hypothesis. Hwang & Kim (2009) find that social ties between executives and board directors matter. Those boards that are both conventionally and socially independent award a lower level of total compensation and have pay that is more sensitive to performance. Bertrand & Mullainathan (2001) find that consistent with the managerial power hypothesis, executive compensation is tied as much to luck (factors outside their control) as to performance. Garvey & Milbourn (2006) extend this result and find that not only do executives enjoy increased pay for good luck, but they also don't face decreased pay for bad luck – which would almost be the optimal scenario if you're deciding on your own camouflaged rent-seeking pay package.

But others dispute that managerial power and skimming are really what's driving executive pay. Murphy & Zábojník (2004) argue that market competition for the most-talented CEOs is driving the trends of higher pay, because transferable general management skills have increased in importance for a successful CEO, while the importance of firm-specific knowledge has decreased. Gabaix & Landier (2010) also argue that the rise in pay over the past three decades has been driven by market factors. Despite a small dispersion in CEO talent, the increase in size of firms over that period has caused competition for the very best executives to increase in step. Thus CEO pay is rising for similar reasons as for star athletes: as their reach has increased dramatically, even minuscule differences in talent can lead to huge differences in pay, and exponential growth for the best. But Malmendier & Tate (2010) suggest that although the skewed distribution of CEO salaries reflect the rewards to superstar-CEO status, once CEOs attain their stardom and huge pay, they tend to underperform, particularly in firms with weak governance. So the outsized pay packages may be a winner's curse for companies that award them.

Suffice it to say that there is no consensus on the causes of the rise in executive compensation, or in what optimal pay should look like in different circumstances. So this paper will examine how, in practice, compensation varies as a function of shareholder power. In particular, it will see how it varies depending upon if ownership is controlled by institutional investors or otherwise.

Prior studies have examined the empirical relationship between institutional ownership concentration and executive compensation in isolation, during the 1990s. Hartzell & Starks (2003) primarily examine the sensitivity of options compensation to institutional ownership concentration. They also find a negative relationship to the level of compensation, but only after controlling for the total share of institutional ownership. I would argue that this formulation, by holding the share of ownership fixed, obscures the true effect of concentration (especially given the often very high total share of institutional ownership). We will however find a similar non-linear effect of concentration: at the highest levels, the impact on compensation is attenuated. Khan *et al.* (2005) also examine the relationship among 224 of the largest companies and find results which contrast with mine. Having a larger cross-section of public companies than previously studied allows me to reach different conclusions and I show that examining only the S&P 500 leads to different results. Some differences may also have resulted from a change in the incentives for options compensation in 2006.

3.3 Data

The universe of companies I consider is from the Wilshire 5000 index (Wilshire, 2010) of domestic, commonly traded public companies. This is one of the broadest indices of public companies in America, and was chosen to allow a comparison among a fuller cross-section than the more commonly used S&P 500 index or other indices that are dominated by the largest companies. As of September 30, 2010 the Wilshire 5000 contained 4063 companies (identified by their stock tickers), covering the full gamut of industries and sizes. I examine these companies and their executives' pay in 2009, in the immediate aftermath of the financial crisis.

Fundamental company information for these companies is from the S&P Compustat database (Standard & Poor's, 2010b). I was able to gather such data for 3843 of the companies. This data includes company industry categorization (in three forms: North American Industry Classification System (NAICS), Standard Industrial Classification (SIC), and Global Industry Classification Standard (GICS)), headquarters state, and fundamentals from 2009: Market value (mkvalt), net sales/turnover (sales), net income/loss (ni), S&P core earnings, and number of employees (emp). From this I calculate profit margin as the ratio of net income to sales. A selection of company summary statistics by GICS business sector are presented in Table 3.1. In this study, company data are used primarily as controls.

3.3.1 Company ownership data

Next, I need to obtain information about the ownership of these public companies. Public information as to company ownership comes from two sources: reports of the holdings of large institutional investors, and the reports of the holdings of company officers and/or beneficial owners of large stakes (amounting to over 10 percent of the company). I collect from both of these sources using two data sets from Thomson Reuters. The institutional ownership dataset (Thomson Reuters, 2010b) contains quarterly information about the common equity holdings of institutional managers with $100m or more Assets under management, as reported in Form

Table 3.1: Company summary statistics by GICS sector

	mkval ($M)	sales ($M)	act ($M)	spce ($M)	emp (1000)	Π-margin (ratio)	%Δ spce (YoY)	%Δ Π-margin (YoY)	ceo total ($1k)	ofc. total ($1k)
Energy	275	278	275	278	268	270	278	269	181	181
	5204.3	4404.6	1160.3	149.0	3.1	-0.280	-0.762	-0.926	4080.2	2061.8
	23099.4	21473.2	4495.0	1531.8	9.4	0.919	2.274	3.098	5339.2	2445.9
N=278 (7.2%)	10.8	9.6	8.8	10.2	5.9	-3.568	-2.125	-2.641	2.6	2.4
Materials	165	169	168	168	167	165	167	163	130	130
	3192.5	2745.1	1177.2	88.5	6.7	-0.172	0.024	-0.328	3934.6	1872.5
	6700.4	5096.3	2398.2	398.3	10.8	0.884	2.159	3.138	3768.0	1815.3
N=169 (4.4%)	3.8	4.8	5.1	2.0	3.0	-4.630	-1.283	-2.469	1.8	3.3
Industrials	515	524	506	522	518	524	522	524	421	421
	2604.6	2839.4	960.6	115.6	12.5	-0.095	-0.293	-0.630	2902.7	1427.5
	9608.2	9346.8	2746.1	612.5	38.3	0.626	1.890	2.925	3442.9	1582.4
N=524 (13.6%)	10.6	10.2	7.5	11.4	7.7	-6.298	-2.050	-2.896	2.2	3.2
Consumer	539	543	517	539	540	542	539	542	434	434
Discret.	2515.3	3084.2	953.9	95.9	17.5	-0.033	0.080	-0.065	3825.3	1984.5
	6886.6	8465.5	2115.8	420.6	42.6	0.311	2.131	2.874	5170.9	2486.1
N=543 (14.1%)	5.6	7.5	4.4	5.1	5.4	-9.633	-1.981	-2.788	3.2	3.6
Consumer	143	149	149	148	148	148	148	148	124	124
Staples	8968.9	10773.6	2291.2	559.2	38.0	-0.036	0.188	0.086	4441.2	2181.6
	26425.0	36904.6	5498.8	1769.2	178.5	0.579	1.850	2.075	5395.4	2309.0
N=149 (3.9%)	5.0	8.7	5.0	5.3	10.6	-7.701	-1.608	-1.852	2.3	2.5
Health Care	552	556	552	551	546	520	551	508	452	452
	2772.5	2200.8	931.7	152.5	4.5	-0.820	0.017	-0.292	2872.8	1488.9
	12707.3	9822.7	4173.4	852.7	12.8	1.623	1.856	2.742	4687.4	2122.0
N=556 (14.5%)	9.5	7.5	9.1	8.9	5.1	-1.761	-2.421	-3.143	4.0	4.4
Financials	787	790	60	778	775	784	775	781	621	621
	2555.1	2205.0	1238.2	-22.3	4.1	-0.073	-0.817	-1.040	2500.5	1384.3
	11688.4	10764.0	3182.5	2954.0	21.6	0.492	2.898	4.026	3905.3	1765.5
N=790 (20.6%)	10.6	9.4	4.4	-22.7	10.5	-4.807	-1.580	-2.113	4.2	3.1
Info.	668	677	668	670	662	676	668	673	562	562
Technology	3512.1	1621.3	1105.2	139.5	5.1	-0.132	-0.300	-0.644	2561.6	1432.8
	17640.2	7407.6	4636.7	1004.2	22.2	0.615	2.416	3.405	5000.8	2125.7
N=677 (17.6%)	8.7	10.3	8.1	10.1	12.8	-5.867	-1.686	-2.642	10.0	6.7
Telecomm.	48	51	51	51	51	50	51	49	40	40
Services	7682.6	6437.5	1655.0	338.6	13.7	-0.240	0.317	0.239	4513.5	2400.9
	26997.2	22799.6	4698.8	1961.1	49.7	0.932	1.558	1.872	5498.6	3106.4
N=51 (1.33%)	5.0	4.5	4.1	5.1	4.7	-3.675	1.023	-0.154	2.7	3.0
Utilities	105	106	105	106	104	106	106	105	79	79
	4596.0	4069.2	1543.1	306.4	5.1	-0.223	0.408	0.224	4980.4	2288.0
	6248.6	4677.4	1951.2	498.8	6.1	1.160	1.172	1.678	3893.8	1550.3
N=106 (2.8%)	2.3	1.3	1.6	2.3	1.7	-3.601	0.473	-1.814	0.8	0.6
Total *count*	3797	3843	3051	3811	3779	3785	3805	3762	3044	3044
mean	3338.5	2935.0	1117.9	121.0	8.9	-0.208	-0.277	-0.535	3136.6	1630.8
s.d.	14295.9	12916.3	3772.7	1583.4	44.8	0.875	2.298	3.220	4598.7	2084.7
skewness	10.7	14.7	8.6	-28.4	29.2	-4.282	-1.921	-2.661	4.9	4.1

NOTES: All figures are for 2009. Each section has four rows: the count of applicable companies, mean, standard deviation, and skewness. mktval=Total Market Value; sales=Net Sales/Turnover; act=Current Total Assets; spce=S&P Core Earnings; emp=Number of Employees; Π-margin=Profit Margin (net income/net revenue); ceo total=CEO Total Compensation; officer total=Average Total Compensation of Officers (including CEO).

13F filings with the U.S. Security and Exchange Commission (SEC). These data give a good picture what large institutional investors are holding. Our data are from the fourth quarter 2008 holdings period, and provide data for 3818 of the Wilshire universe of companies. I chose this period on the assumption that pay packages for 2009 would usually be negotiated in advance. Companies may be missing because they are not held by any large institutions, or because they were not commonly traded under their current symbol in late 2008. From this data, I computed several summary statistics for each company across all institutions. First I compute the total percent of common equity held by all institutions (instl_share). Next, I compute an index of institutional company ownership inspired by the Herfindahl-Hirschman (H-H) Index (used to measure market concentration). It is calculated as $\sum_{i=1}^{n} r^2$ where r is the fraction of the company held by institution i, across all n institutions that hold the company (hh_instl). Finally I keep a count of the number of institutions holding a greater than 5% stake in the company, and greater than a 10% stake in the company.

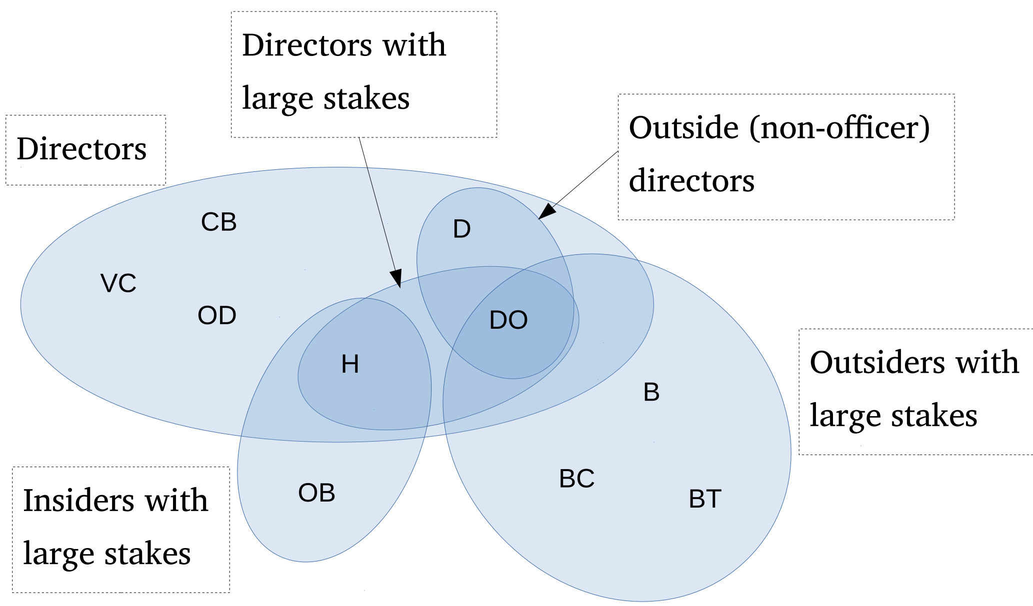


Figure 3.1: Insider-filing role codes

The second source of ownership information is from Thomson Reuters' Insider Filing dataset (Thomson Reuters, 2010a). Company officers and directors, as well as any beneficial owners of over 10% of the company's equity are required to file SEC Forms 3, 4, and 5 (which are captured by Thomson Reuters) whenever performing any transaction relating to these shares, or a minimum of once a year. I utilize the data from Table 1 of each form, which contains the shares transacted as well as (usually) the remaining shares held by the individual (both directly and indirectly) after each transaction. The goal is to extract ownership concentration information by company from this data. I consider the year centered on 31 Dec 2008, and mark those records that contain the most current report of direct holdings, as well as each of the indirect holdings (all of which are listed separately) for that individual at the time. It then attempts to remove duplicate reporting of indirect holdings (which is because multiple individuals might all be indirectly responsible for a set of shares, as for a set of trustees). Next, I flag several overlapping groups of individuals as shown in Figure 3.1: outsiders (i.e., non-company-officers) with large ($> 10\%$) stakes (who have rolecodes DO, B, BC, and BT), insiders (i.e., company officers) with large ($> 10\%$) stakes (rolecodes H, OB), all directors on the board (rolecodes CB, D, DO, H, OD, and VC), outside (non-officer) directors (rolecodes DO and D), and directors with $> 10\%$ stakes (rolecodes DO and H)[1]. Since the raw data are subject to significant noise, I use TR's cleansed fields where possible. Summary statistics are constructed using the marked entries. For each owner, I find the fraction of the company held as the sum of all ownership all the marked direct and indirect holdings divided by the number of common shares outstanding for the company (from S&P Compustat). Grouping by company, I next compute the Herfindahl-Hirschman-like index analogously to before, for all owners (hh), for outside owners only (hh_outsiders), for directors (hh_directors), and for outside directors (hh_outsidedirectors). I also keep track of both the number of large outside stakeholders (large_outsiders), large inside stakeholders (large_insiders), and directors with large holdings (large_directors). Finally, I calculate the total fraction of the company held by these groups (large_outsider_share, large_insider share,

[1]Note that in this paper, when referring to 'outside' board members I am not referring to independent directors, who are meant to have no pecuniary or familial interest in the company, but rather to board members which do not serve as officers or employees of the company.

director_share, and outside_director_share). At least some of this data is available for 3728 of the companies under consideration; companies may be missing because individuals misreported or omitted reporting of total holdings, or because there was a change in the company's trading status between 2008 and 2009.

Table 3.2: Ownership summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
large_insider_share_w	0.022	0.079	0	0.555	3713
large_outsider_share_w	0.056	0.134	0	0.807	3661
director_share_w	0.073	0.119	0	0.687	3432
outside_director_share	0.036	0.084	0	0.896	3427
instl_share_w	0.566	0.308	0.02	1.079	3735
large_insiders	0.156	0.438	0	5	3748
large_outsiders_w	0.449	0.949	0	6	3748
large_directors	0.248	0.586	0	8	3748
large5_institutions	2.297	1.774	0	10	3735
large10_institutions	0.49	0.727	0	6	3735
hh_w	0.026	0.076	0	0.587	3530
hh_largeoutsiders_w	0.015	0.056	0	0.496	3748
hh_directors_w	0.014	0.048	0	0.41	3432
hh_outsidedirectors_w	0.006	0.026	0	0.251	3427
hh_instl_w	0.027	0.023	0	0.106	3735

NOTE: Variables ending in _w have been Winsorized.

The various measures of company ownership are summarized in Table 3.2. The simplest measure of ownership is the count of stakeholders. It may also be the most robust, since the reporting requirements are simpler and less subject to error. However, more owners does not mean more concentration, so these variables are reduced to dummies – is there at least one large outsider, for example. The second measure is the fraction of the company held by different types of owner, and this is reported as 'share.' The primary disadvantage of this measure is that it makes no distinction between many small holders and several large holders, although these would likely mean different things for coordination. My preferred measure is the H-H measure of concentration. This is quite robust to the absence of small holders, and captures the importance of a few large holders. The primary downside is that it relies on correct reporting of shareholdings in a noisy dataset. To help compare these measures, I include a correlation matrix for H-H measures on Table 3.3 and ownership shares in Table 3.4.

Table 3.3: H-H correlation matrix

	HH concentration measures for:				
	all large owners	large outsiders	directors	outside directors	institutional investors
all large owners	1				
large outsiders	0.815***	1			
directors	0.747***	0.357***	1		
outside directors	0.486***	0.500***	0.622***	1	
inst'l investors	0.0994***	0.122***	0.0409*	0.121***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

NOTE: Correlations between different types of ownership concentration, all calculated as Herfindahl-Hirschman indexes.

Table 3.4: Total ownership share correlation matrix

	Share of company held by:				
	large insiders	large outsiders	directors	outside directors	institutional investors
large insiders	1				
large outsiders	0.0252	1			
directors	0.660***	0.345***	1		
outside directors	0.00647	0.508***	0.689***	1	
inst'l investors	-0.187***	-0.178***	-0.344***	-0.216***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

While there may in theory be overlap between the two sources of ownership data, i.e. that some stake might be reported in both data sets, I don't believe that is likely in practice. The question reduces to whether an institutional ownership stake might also be reported on an insider reporting form SEC 3, 4, or 5. These forms mandate reporting of beneficial ownership stakes. But a stake managed by an institutional investment manager is not beneficially owned by them or their firm, but rather by their investors. Thus an investment company employee sitting on a company's board would not report their company's investment stake on these forms. And an investment company controlling a 10 percent stake would not report that on these forms, except in the unusual circumstance that their investors are sufficiently concentrated that at least one owns 10 percent of the company through their investment at that institution. This understanding is reinforced by the fact that the computed insider ownership shares are negatively correlated with those of institutions (as shown in Table 3.4. Further, in the regressions I will play these sources against one another and find that there is little to no impact on them individually. Finally, a few spot checks of the raw data fail to show the same stakes for institutions with 10 percent stakes in the insider data.

3.3.2 Executive compensation data

On the assumption that smaller companies are more likely to have concentrated ownership (which turns out to have been correct, *ex poste*), a major effort has been undertaken to gather compensation data from primary sources for as wide a spectrum of companies as possible. The SEC requires public companies to report current and two prior years' compensation for the top five executives in a summary table format within annual definitive proxy statements (Form DEF 14A), which are archived on SEC's EDGAR service (Securities and Exchange Commission, 2010) in a hybrid XML/HTML format. Although companies are required to follow a particular form for the table, they often do not, and even when they do, there is a wide variety of compliant forms. A Matlab program was developed to systematically find and download DEF 14A forms and attempt to extract the summary compensation table from them following a multi-step process. The program is robust to common spelling errors, footnotes, comments, special characters, and multiple ways of formatting HTML tables. When successful, the program extracts the name and title of each named officer, and for each of the listed years their base salary (salary), bonus (bonus), stock awards (stock) and option awards (option) both priced at grant date fair value, non-equity incentive plan compensation (nonequityincentives), "annual change in the actuarial present value of accumulated pension benefits and above-market or preferential earnings on nonqualified deferred compensation" (Securities and Exchange Commission, 2006) (deferred), all other compensation including perquisites (other) and a total annual compensation (total).

With this assembled dataset, I compute the total amount of each individual's pay which is incentive based (=bonus+stocks+options+nonequityincentives), as well as the fraction of their total pay which is incentive pay (inc_pct),. Finally, I generate summary data for each company: the CEO's compensation figures and percentages, as well as the mean compensation figures across all named officers and percentages of those means. I was able to successfully extract this data for 3145 firms in 2009. The remaining firms have one or more flaws in their data – either there are multiple tables that meet the characteristics of a summary compensation table, the columns are ambiguously labeled with respect to the template, the table has been constructed in a highly irregular way, or no matching table was found. An analysis of the characteristics of firms unable to be retrieved shows them to be smaller on average than firms which were retrievable (by sales, total assets, and employees), but despite this underweight on small firms, the collection nonetheless contains a greater cross section of small firms than previous datasets of CEO compensation have managed. Additionally, since selection is on an arbitrary matter (ability to extract data mechanically), it should be independent of the variables of interest – ownership characteristics and compensation practices – and thus introduce no bias. A summary of overall executive compensation statistics is included in Table 3.5.

For robustness checks and historical comparison, compensation data has also been gathered for 1999-2009 from the S&P Execucomp database (Standard & Poor's, 2010a) which covers executive compensation for S&P 1500 firms also based on proxy filings. The same data as collected above is available for these firms from 2006 - 2009. Reporting prior to 2006 does not include data in the value of deferred compensation or split out non-equity incentives from bonus. The earlier data is based on S&P's estimates of stock and option value at time of grant, using a modified Black-Scholes formula. Note that I do not merge this dataset with my own, since this may risk introducing a bias into the sample[2]. I construct parallel company-level summary data to the above for CEOs and officers, including amounts and percentages of incentive pay for the year 2009, allowing comparison with the larger data set.

[2]Since the S&P data is comprehensive for these firms, but non-S&P firms would only be available via mechanical extraction, merging them would impart a bias in selection of firms towards including S&P firms over non-S&P firms

Table 3.5: Executive compensation summary statistics

Variable	Mean	(Std. Dev.)	Min.	Max.
ceo_salary	588,226	(389,494)	0	8,100,000
ceo_bonus	149,598	(711,010)	0	19,891,275
ceo_stocks	830,367	(1,818,431)	-1,659,667	24,758,827
ceo_options	660,685	(2,216,199)	-15,728	78,421,000
ceo_nonequityincentives	492,426	(1,089,417)	0	14,629,074
ceo_deferred	220,426	(857,872)	-373,820	14,197,821
ceo_othercomp	129,983	(849,706)	0	43,576,932
ceo_total	3,072,505	(4,546,389)	-36,929	84,501,759
ofc_salary	382,895	(220,531)	0	5,440,197
ofc_bonus	86,173	(336,283)	0	10,129,157
ofc_stocks	400,689	(839,513)	-606,556	12,265,880
ofc_options	307,474	(856,355)	-97,444	28,107,649
ofc_nonequityincentives	239,157	(504,702)	0	7,270,608
ofc_deferred	92,440	(312,889)	-110,545	4,624,293
ofc_othercomp	95,375	(321,122)	0	10,540,630
ofc_total	1,604,786	(2,079,116)	378	30,979,400
N		3145		

All variables that appear to have been subject to error have been Winsorized at the 1% level, except for profit-margin variables, spce growth, and institutional ownership measures, which were Winsorized at the 5% level.

3.4 Empirical Results

For context, let us first examine historical patterns in executive pay in the decade leading up to the study year, shown in Figure 3.2, and based upon the S&P Execucomp data. The height of each bar is the median total annual pay for the year, while the breakdown is by mean percentage of each type. Growth in pay has continued, although seemingly at a lower rate than in the prior two decades, despite two market corrections. The reduced rate of increase owe itself to these two falls in stock market capitalization. Over the period we also see a marked secular trend towards lower options pay and higher stock pay as a percent of total. There is a change in reporting requirements between 2005 and 2006, adding the additional categories of 'non-equity incentive plan pay' and 'change in the value of deferred' pay. Much of what was previously called bonus is now called non-equity incentive pay. Base salary increased over the period, but at a lower rate than total pay, and salary made up only a third or less of total compensation throughout the period. For comparison, a line graph of The Wilshire 5000 index (approximating total stock market capitalization) is shown (right scale).[3]

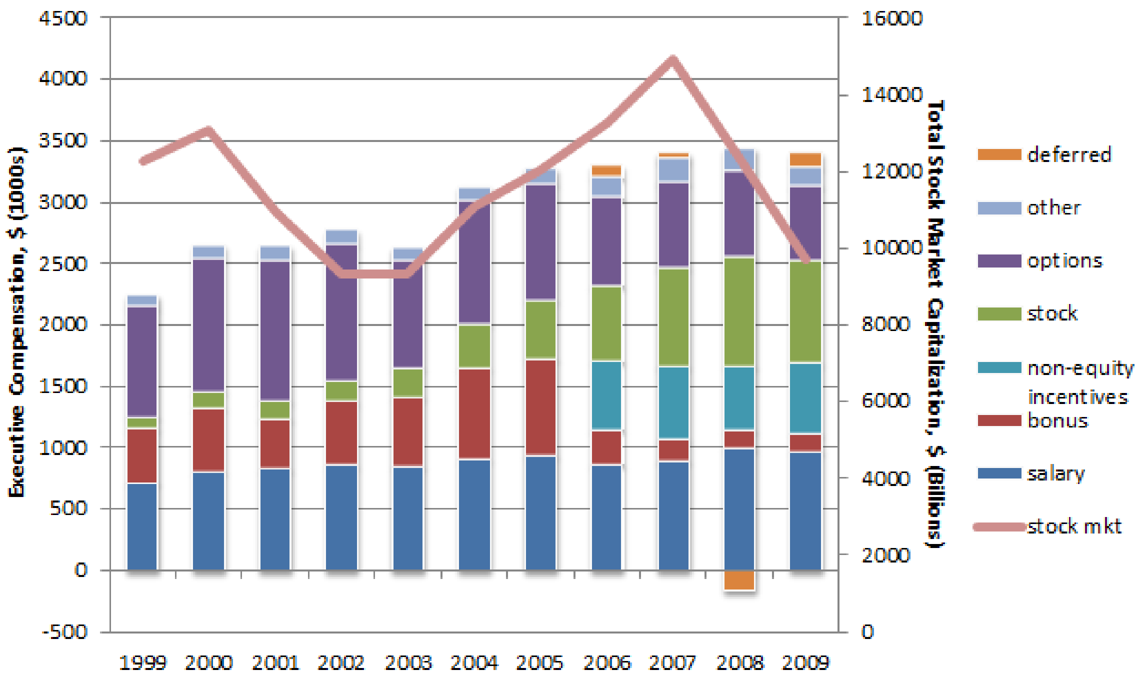


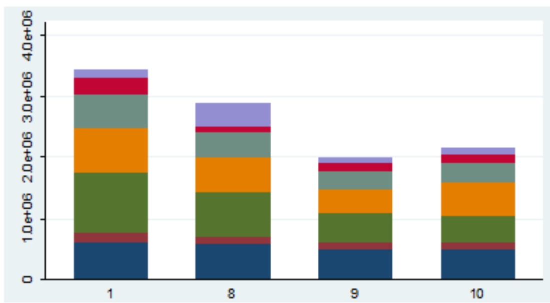
Figure 3.2: Historical compensation trends (S&P Execucomp)

Next let us visually examine the effect of different variables on the composition of CEO pay in 2009, shown in Figure 3.3. In chart a, we see a general downward trend in overall pay, and incentive pay in particular, as large outside ownership concentration increases. Note that the bins are equally sized, ranging from zero to the maximum. Thus, all of the 75 percent of companies which have no large outsiders are in bin 1. In chart b, we see that increases in institutional ownership concentration are generally associated with increases in both overall and incentive pay, although there is some evidence for nonlinearities in the effect (pay declines slightly at the

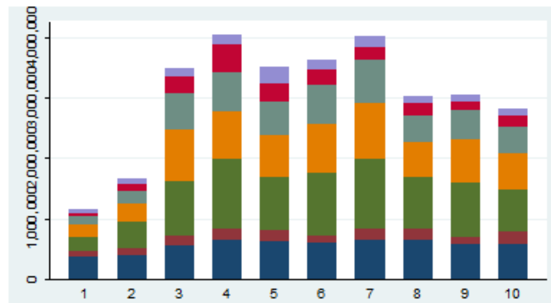
[3]Levels of the Wilshire 5000 for 1999 and 2000 were approximated using figures for other indices.

highest concentrations). In chart c, we see a quite clear fall in particularly incentive compensation as the concentration of ownership among board members increases. Finally, in chart d, we see the strong impact of company size (as measured by sales) on executive compensation which was examined to in Gabaix & Landier (2010).

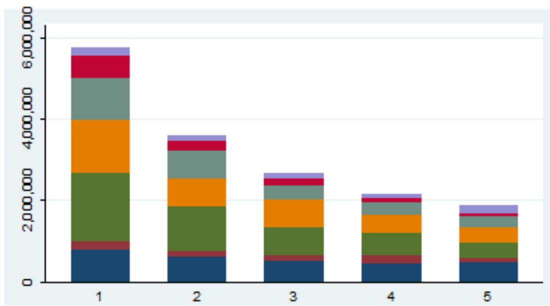
(a) By deciles of large outside ownership concentration



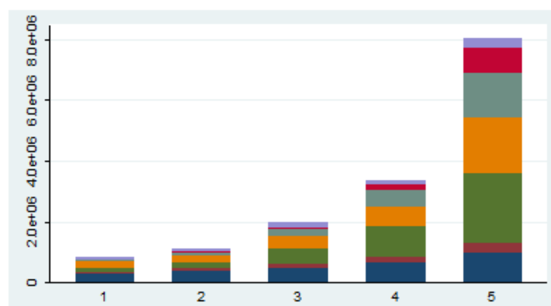
(b) By deciles of institutional ownership concentration



(c) By quintiles of company-director ownership concentration



(d) By company size quintiles: sales revenue



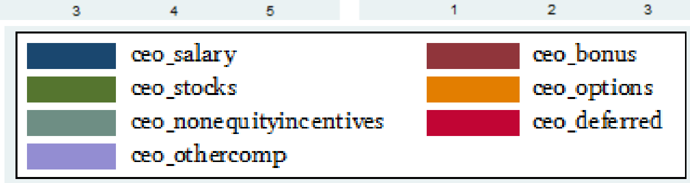


Figure 3.3: Mean components of CEO pay for 2009, by quintiles/deciles (ordered lowest to highest) of different measures of ownership concentration, and by company sales. Note 75 percent of companies have no large outsiders.

I hypothesize that we will see a differentiation in executive compensation depending upon the ownership concentration. Specifically, the more concentrated the ownership structure, the more likely compensation packages will reflect optimal practices for that firm, *ceteris paribus*.

The proposed causal mechanism is that a more concentrated set of owners faces less of a coordination problem in setting pay, and, given more of a stake in the company for each individual owner, more incentive to set executive pay optimally. A less concentrated ownership has lower incentives to, and more complications in setting optimal executive compensations. This allows executives themselves to play a greater role in setting it to their own benefit. By implication, the compensation levels and structures observed at the highest levels of concentration are likely to be lower, and closer to optimal, than that of counterparts with less ownership concentration.

The strategy I take to analyzing this problem is primarily to regress log CEO incentive pay on various measures of ownership concentration across a cross-section of companies (i) in 2009.

The basic model is of the following form:

$$\begin{aligned} ln(\textbf{ExecutivePay}_i) &= \beta_0 + \beta_1 \textbf{OwnershipConcentration}_i + \beta_2 ln(\textbf{Sales}_i) \\ &\quad + \beta_3 \textbf{CoreEarningsGrowth}_i + \beta_4 \textbf{ProfitMarginGrowth}_i \\ &\quad + \beta_5 \textbf{ProfitMarginGap}_i + X\delta + \epsilon_i \end{aligned}$$

I control for company size (as measured by the natural log of sales), and performance, which is measured by three variables: 1) the growth in the company's 'S&P core earnings' in 2009 relative to 2008, 2) the growth in profit margin in 2009 relative to 2008, and 3) the gap between the company's profit margin in 2009 and average profit margins for the company's sub-industry that year. All regressions have sub-industry fixed-effects, and cluster by state to account for some degree of regional variation in pay.

Tables 3.6 and 3.7 present the main results on CEO incentive pay (in log form). Table 3.6 presents results comparing the effect of large outside ownership concentration and institutional ownership concentration. Table 3.7 presents results of regressions of the same form, now comparing the impact of ownership specifically among company directors with that of institutional investors. In all regressions on total incentive pay, all the control variables are all significant.[4]

[4]For CEO options alone, the variables measuring growth in earnings and profit margins are insignificant. This is probably because options grants are meant to incentivize future improvements in performance, rather than reward past or current ones.

Table 3.6: Outsider and Institutional Ownership's impact on Incentive Pay

Dependent variable:	*ln*(CEO incentive pay)					*ln*(CEO options)
	(1)	(2)	(3)	(4)	(5)	(6)
hh-largeoutsiders	-1.710*** (0.529)		-1.977*** (0.579)			-1.312 (0.900)
hh-instl		8.391*** (1.226)	8.599*** (1.262)			4.204** (2.074)
large-outsider-share				-.006 (0.213)		
instl-share				1.757*** (0.133)		
large-outsiders-d					-.209*** (0.061)	
large5-institutions-d					0.441*** (0.105)	
ln(sales)	0.635*** (0.015)	0.623*** (0.015)	0.617*** (0.016)	0.481*** (0.019)	0.610*** (0.017)	0.571*** (0.045)
core-earnings-growth	0.059*** (0.017)	0.064*** (0.016)	0.066*** (0.017)	0.073*** (0.016)	0.057*** (0.016)	0.016 (0.023)
profit-margin-growth	0.028** (0.011)	0.027** (0.012)	0.026** (0.012)	0.019* (0.012)	0.026** (0.011)	0.031 (0.019)
profit-margin-gap	-.379*** (0.053)	-.379*** (0.050)	-.382*** (0.054)	-.342*** (0.049)	-.361*** (0.044)	-.431*** (0.064)
Const.	9.724*** (0.405)	10.017*** (0.107)	10.050*** (0.106)	9.710*** (0.105)	9.528*** (0.363)	9.700*** (0.271)
Sub-industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2826	2812	2767	2760	2933	2770
R^2	0.461	0.467	0.473	0.505	0.457	0.234

NOTES: Standard errors are adjusted for clusters by state to account for regional variation in executive pay. Variables ending in -d are dummies. Levels of statistical significance: * : 10% ** : 5% *** : 1%

Table 3.7: Director and Institutional Ownership's impact on Incentive Pay

Dependent variable:	*ln*(CEO incentive pay)					*ln*(CEO options)
	(1)	(2)	(3)	(4)	(5)	(6)
hh-directors	-2.965*** (0.784)					-3.898*** (0.931)
hh-outsidedirectors		-1.588* (0.943)				
hh-instl	8.081*** (1.404)	9.420*** (1.494)				2.992 (2.297)
director-share			-.776*** (0.287)			
instl-share			1.628*** (0.138)			
large-directors				-.351*** (0.056)		
large-noninstl-outsiders					-.178*** (0.067)	
large5-institutions				0.410*** (0.102)	0.426*** (0.104)	
ln(sales)	0.605*** (0.017)	0.614*** (0.015)	0.474*** (0.020)	0.606*** (0.016)	0.614*** (0.017)	0.558*** (0.044)
core-earnings-growth	0.066*** (0.015)	0.069*** (0.015)	0.072*** (0.015)	0.058*** (0.015)	0.059*** (0.015)	0.008 (0.024)
profit-margin-growth	0.027** (0.012)	0.026** (0.012)	0.020* (0.011)	0.025** (0.011)	0.025** (0.011)	0.032 (0.020)
profit-margin-gap	-.363*** (0.054)	-.370*** (0.063)	-.330*** (0.050)	-.357*** (0.044)	-.363*** (0.045)	-.398*** (0.063)
Const.	10.151*** (0.113)	9.584*** (0.335)	9.896*** (0.114)	9.598*** (0.358)	9.506*** (0.363)	9.291*** (0.493)
Sub-industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2696	2584	2696	2933	2933	2698
R^2	0.474	0.473	0.503	0.460	0.456	0.239

NOTES: Standard errors are adjusted for clusters by state to account for regional variation in executive pay. Variables ending in -d are dummies. Levels of statistical significance: * : 10% ** : 5% *** : 1%

Interestingly profit margin growth is only weakly indicative of the CEO's incentive pay, suggesting that the pay either is not corresponding well to the CEO's performance, or is a poor reward for performance. Also, surprisingly, incentive pay is lower the greater the (positive) gap between company's profit margin and peers in the same sub-industry. It may be that once the growth in profit margin is controlled for, investors do not penalize for having below average profits, and dis-incentivize increased profit margins when they are already performing well with respect to the industry. This point probably deserves further research.

Moving to the measures of ownership concentration in Table 3.6, regression one shows that as the concentration of large outside ownership increases, CEO incentive pay decreases, an effect which is significant at the 1 percent level. The second regression does an equivalent regression but instead assesses the impact of institutional ownership concentration. This has a stronger, opposite effect, also significant at the 1 percent level. Next, in regression three, we take these together, and find the results hold, and the magnitudes are even stronger. This is strong confirmation that any potential overlap between large outside investors and institutional investors is not confounding the results. Even when accounting for the presence of one another, large outside holders have the effect of restraining CEO incentive pay, while institutional investors have the *opposite* effect. Regressions four and five are robustness checks. Regression four examines the share, rather than the concentration of ownership (although this may confound broad dispersed ownership with narrow concentrated ownership and is furthermore subject to significant noise). For large outsiders, this measure is particularly subject to noise, and is no longer significant, but the direction remains the same, while the positive effect of institutions on incentive pay remains strong and significant. In regression four, we very simply assess the effect of the presence of one or more either outside investor or institutional investor on incentive pay, and the effects remain as before. All else held constant, the presence of a large outside investor has the effect of reducing CEO incentive pay by 18.9 percent, whereas the presence of an institutional investor with at least a five percent stake has the effect of increasing CEO incentive pay by 55.4 percent. Finally, we find the same impact to a slightly lower degree in regression six, but on CEO option pay only. This effect is not significant for large outsiders however. As noted previously, option pay is a much smaller portion of incentive pay since the FAS 123 accounting changes that went into effect in 2006.

It seems that large outsiders have a different set of views from institutional investors as to the appropriate level of incentive compensation. In Table 3.7, we compare the impact of ownership by members of a company's board of directors against ownership by institutional investors. In regression one, we see that the concentration of ownership among directors has the same effect of reducing incentive pay as for outsiders, but to an even stronger degree. This makes sense given their more direct ability to impact pay. The same is true in regression two for only non-officer directors (that is, members of the board who do not work at the company). Interestingly, this is slightly less strong of an effect than when inside directors are included. It may be that when the CEO or other officers are already a large shareholder, this has the effect of disciplining their incentive pay – in this case their incentives are already aligned with other shareholders, and thus there is little purpose or benefit to strong incentive pay.[5] But even the effect of only outside director ownership is significant, if only at the 10 percent level. By contrast, in both regressions, institutional investor ownership concentration has a positive impact on incentive pay of a similar magnitude as before, and is significant at the 1 percent level. Regression three uses ownership shares rather than concentrations, and finds a similar effect strongly significant. Regression four mirrors regressions one, but reduces to a dummy for the presence of a director with a large (10 percent) stake. The results are similar. All else being equal, the presence of a director with a large stake reduces CEO incentive pay by 29.6 percent, while the presence of an institutional investor with a large (five percent) stake implies a 50.7 percent increase. There is a wide variation in incentive pay among CEOs, so such large impacts are plausible. To be certain that the outside investors captured in the 13(f) data are not in fact institutional investors (or at least that they are not driving results for them), I construct a dummy for companies that have large outsiders even after subtracting all large institutional investors (both at the 10 percent ownership-stake level). In regression 5, we see that the same results hold and are significant at

[5]Although I don't show the regression, I also find that insider-directors are related to lower overall pay for CEOs. Given that such CEOs already have objectively very large stakes in their companies, their reward apparently comes directly from running the company well.

the 1 percent level. The magnitude is slightly smaller than for large outsiders overall, and the effect of large institutions is similar but slightly larger. Finally, in regression six, we see that test the impacts of ownership concentration on CEO option pay more narrowly. There is a stronger negative association of directors with option incentives, but the positive effect of institutional investors is weaker and no longer statistically significant.

These results indicate that the presence and concentration of institutional ownership are associated with higher levels of CEO incentive compensation, while concentrated ownership by non-institutions is generally associated with lower levels. We naturally wonder how robust this finding is. In Table 3.8 we examine some extensions. Firstly in regression one, we find that there is a nonlinear relationship in institutional ownership concentrations. Seeing that the squared term on institutional ownership concentration is negative and significant at the 1 percent level, we find that although higher ownership concentrations are generally associated with more incentive pay for CEOs, this effect diminishes as the ownership concentration increases, and at the highest levels of concentration the effect could potentially even reverse. This is consistent with what we saw in Figure 3.3b.

In regression two, we examine if, independent of the total level of compensation, institutions and outsiders are impacting the proportion of pay that is incentive-based (rather than a salary, say). We do this by using the *percent* of total compensation that is incentive pay as the dependent variable. Since this is a limited dependent variable, we use a logit model to estimate it. We find that ownership concentration among institutional investors is associated with higher percentage of incentive pay (significant at the 1 percent level), while more concentrated ownership by outsiders is associated with a lower percentage (significant at the 10 percent level). Next, we consider how *total* compensation is impacted by concentrated ownership. Again, institutional ownership is strongly associated with higher levels of total compensation, significant at the 1 percent level in both regressions 3 and 4. Thus institutional ownership concentration is associated both with higher levels of CEO pay, as well as with higher percents of CEO pay that are incentive-based. Thus, this cannot be a story simply about how different kinds of investors choose an optimal mix of incentive- and non-incentive-based pay, depending upon their ability

Table 3.8: Additional ownership impacts and validity checks

Pay data source	Collected from annual statements								S&P Execucomp		
Dependent var.	ln(inc. pay)	%/inc. pay	ln(total compensation)		ln(incentive pay)		ln(incentive pay)				
Pay of	CEO	CEO	CEO		Officers (mean)		CEO				
Subset of companies	Complete	Complete	Complete		Complete		S&P 1500	S&P 500	Complete	S&P 500	Complete
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
hh-largeoutsiders	-1.436**	-2.217*	-.987***		-1.209***		-2.727**	-7.045*	-5.005***	-7.148**	
	(0.569)	(1.300)	(0.284)		(0.413)		(1.177)	(3.627)	(1.387)	(3.042)	
hh-outsidedirectors				-1.044**		-1.988***					-4.390*
				(0.512)		(0.700)					(2.342)
hh-instl	24.926***	20.550***	3.751***	4.165***	3.945***	4.273***	6.737***	-1.314	4.485**	-5.516	3.269*
	(5.333)	(4.586)	(0.643)	(0.684)	(0.925)	(1.152)	(2.314)	(4.730)	(1.771)	(3.833)	(1.976)
hh-instl^2	-176.750***										
	(53.116)										
ln(sales)	0.601***	0.292***	0.460***	0.459***	0.568***	0.565***	0.552***	0.183**	0.531***	0.061	0.53***
	(0.016)	(0.047)	(0.014)	(0.015)	(0.009)	(0.011)	(0.029)	(0.090)	(0.029)	(0.090)	(0.029)
core-earnings-growth	0.067***	0.077**	0.031***	0.035***	0.054***	0.055***	0.030**	0.030	0.041**	0.049	0.030
	(0.017)	(0.034)	(0.009)	(0.008)	(0.014)	(0.014)	(0.015)	(0.050)	(0.020)	(0.043)	(0.022)
profit-margin-growth	0.025**	0.035	0.013**	0.011*	0.016	0.016	0.018	0.031	0.005	0.008	0.006
	(0.012)	(0.030)	(0.006)	(0.006)	(0.011)	(0.011)	(0.015)	(0.037)	(0.014)	(0.034)	(0.016)
profit-margin-gap	-.380***	-.067	-.306***	-.297***	-.372***	-.355***	0.037	0.160	0.107	-.057	0.150
	(0.053)	(0.070)	(0.030)	(0.035)	(0.049)	(0.051)	(0.190)	(0.286)	(0.200)	(0.343)	(0.221)
Const.	9.825***	1.622***	11.576***	11.430***	9.662***	9.315***	9.578***	14.186***	11.180***	15.214***	10.794***
	(0.132)	(0.48)	(0.092)	(0.211)	(0.086)	(0.198)	(0.227)	(0.979)	(0.230)	(0.879)	(0.313)
Sub-industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2767	2501	2769	2586	2767	2584	1332	406	1577	472	1510
R^2	0.478		0.503	0.498	0.602	0.606	0.37	0.208	0.383	0.252	0.367

NOTES: The S&P Execucomp database used in regressions 8-11 generally covers the S&P 1500 companies. Regression two uses a logit model; all others use an OLS regression. Standard errors are adjusted for clusters by state to account for regional variation in executive pay. Levels of statistical significance: * : 10% ** : 5% *** : 1%

and cost of monitoring. They also increase the level on net. Outsiders and outside directors have a negative association with total CEO compensation, also similar to the association with incentive pay. This makes sense considering that there is much more variation in incentive pay than in salaries for CEOs across similar companies. Variation in total compensation is largely being driven by variation in incentive compensation. This is also suggested in the time series shown in Figure 3.2.

Next, we examine whether these associations with ownership are limited to CEOs or extend to other company officers (CFO, CIO, etc.). Regression 5 finds the now familiar result that higher outside ownership concentration is associated with lower incentive pay, while higher institutional ownership concentration is associated with higher incentive pay, now for the mean company officer. Regression 6 also finds the same for outside directors more specifically. All of these results are significant at the 1 percent level. Thus the same effects that are behind CEO pay also seem to be impacting top executives more generally.

Finally, in order to validate my newly collected data set against an established data set that has been previously used in academic research, I perform analogous regressions using CEO salary data from the S&P Execucomp dataset. This dataset is limited to the approximately 1500 companies on the S&P 1500 index. Thus it is biased towards larger, more established companies than the Wilshire 5000 index of widely-traded public companies we have studied until now. First I run regressions on CEO incentive pay where I limit the dataset to companies which are in the S&P 1500 and 500 indexes in regressions 7 and 8 respectively. These can be compared with regression 3 in Table 3.6. In regression 7, we see that results for the S&P 1500 subset are similar to those on the complete set, with the positive effect of institutional investors being only slightly weaker. But in regression 8, the estimate for institutional investor concentration loses significance and even changes sign. In the restricted set of large, well-known companies, the effect we've robustly seen across wider population of companies disappears. Since much prior research as focused upon this set, it may explain why the institutional investor-incentive pay relationship it is not better known. The reason for this disappearance is not clear, but may be due to the lower concentrations of institutional ownership observed for these large companies, coupled with their higher levels of compensation (even after accounting for their size and levels of profits). We also find that among these restricted sets the negative association with large outside ownership concentration and incentive pay strengthens (even if it is less statistically significant in a smaller dataset).

To validate the dataset I have gathered, I compare against results using data from the commercial Execucomp dataset. Regression 9 and 10 parallel those of 7 and 8, but use Execucomp data. Regression 9 uses the complete set of Execucomp companies, while Regression 10 restricts to the S&P 500 companies. Results in both cases are similar and have equivalent interpretations. For completeness, I also include a regression using outside directors rather than all outsiders, again with similar results (compare with regression 2 in Table 3.7 on my complete dataset). Results may differ because my collection method was not able to capture data for all companies, the Execucomp dataset is not exactly the same as the S&P 1500, and there may be other discrepancies in the data as well. But the differences are not enough to affect the timbre of results, and it doesn't appear that there is any real bias being generated by the mechanical company selection engendered by the collection methods.

In short, we find that the identity of the owners matters. Where there is a high concentration in company ownership generally, executive incentive pay, and executive compensation overall, tends to be smaller. The effect is large and statistically significant. But, as the concentration of ownership by large institutional investors increases, executive incentive pay increases along with overall executive compensation, again sharply, and even more robustly. It seems likely that most large holders, including directors and non-directors, robustly call for less CEO incentive (and also option) pay, while companies with large ownership by institutional investors robustly end up with CEOs having higher incentive pay.

3.5 Conclusion

The evidence we have considered shows that the more concentrated the share ownership of board members or large stakeholders is, the lower incentive and option pay is. This effect is robust and economically significant. What is surprising is that the opposite is true of institutional ownership. The more concentrated institutional ownership is, the higher the total and incentive pay of executives are. This is important given that institutions now own over half of public companies on average (56.6 percent in our data) compared with less than 20 percent in 1965 (Coffee Jr, 1991). Despite widely publicized examples of institutional activism for disciplined CEO pay, it's only at the highest levels of concentration that institutions begin to behave like other investors in disciplining pay. While what I have demonstrated is only an association, and causation may run in the reverse direction (i.e., so that institutional investors are drawn to companies having higher-paid CEOs), it seems less likely. Indeed, other research has shown that in time series, changes in institutional ownership lead to changes in pay for performance, but not the reverse (Hartzell & Starks, 2003). Nonetheless, further research is needed to confirm this.

If these higher incentive and overall pay are in reality above optimal levels, then this suggests there may be another principal-agent problem at play, besides the commonly cited one between company owners and management. This one is between the institutional investment managers and their investors. It may be investment managers are not optimally promoting the interests of their investors by disciplining executive pay in the companies they own. One possible explanation for this could be that many institutional investment managers are themselves beneficiaries of extremely high compensation, which makes promoting lower pay for company executives self-defeating.

Appendix A: Events and shifting rhetoric

In Chapter 1, we assessed the slants of news using a large set of slant-diagnostic n-grams each having a certain partisan hue. Both what phrases were slanted, and how slanted they were, vary greatly over time, and produce a larger effect of changing slant on the set of networks under analysis than the differences between the networks. In order to understand how the slants are affected by events and shifting rhetoric, we can explore big month-month shifts in mean and peaks of slant and of polarization, and see what n-grams were highly slanted at the time. The underlying trend seems to be that the way the conversation swings politically depends largely upon the issues that get into the news, rather than how the media treats them. The conversation swings Democrat when the issues swing Democrat or events better support their case, and vice versa. The following are several examples:

- *A one month spike in how liberal stations were in December 2010.* In December, the "Don't ask, don't tell policy" in the military was repealed, a federal district judge ruled that the individual mandate in the Affordable Care Act was unconsitutional, and an agreement was reached to extend the Bush-era tax cuts for two years and emergency unemployment insurance for 18 months. Most partisan n-grams include: `‘tax, break’`, `‘tax, credit’`, `‘unemploy, insur’`, `‘serv, countri’`, `‘repeal, tell’`, `‘bush, tax, cut’`, `‘tax, increas’`, `‘rais, tax’`, `‘tax, cut, wealthi’`, `‘insur, compani’`, `‘health, care, law’`, `‘repeal, replac’`

- *Slants lurch leftwards from July to September 2011, before reverting.* This period begins with the debt ceiling crisis of 2011, which culminated in advance of an August 2 deadline. The agreement on a Budget Control Act was temporary and required further cuts to be advanced by a bipartisan Congressional committee and approved in order to avoid substantial across the board "sequestration" cuts evenly split between defense and non-defense spending. The showdown led the S&P rating agency to downgrade the credit rating of the U.S. government from AAA to AA+. Towards the end of August, Hurricane Irene hit the east coast, doing particular damage to New York and New Jersey. FEMA funding for disaster relief was caught in the bickering over a short-term funding measure, which was needed to avoid a government shutdown. The shutdown was averted on September 27. In addition, over the course of September, President Obama had introduced and promoted a large "Jobs Act" intended to reduce unemployment. Most partisan n-grams over the period include: `‘cut, cap, balanc’`, `‘faith, credit, unit’`, `‘american, job, act’`, `‘job, creator’`, `‘reduc, deficit’`, `‘reduc, spend’`, `‘natur, disast’`, `‘live, mean’`, `’pay, fair, share’`, `‘budget, amend, constitut’`, `‘women, children’`

- *Slants jump to conservative peaks in September and November 2013 before falling again.* In early September, chemical weapons were suspected of having been used in Syria, crossing a ‘red line,’ and President Obama was seeking Congressional approval for military action there. This proved not to be forthcoming, and he opted for a diplomatic course to remove said weapons. The newly elected Iranian president signaled a desire for a thaw in relations and Iranian and U.S. presidents speak for the first time in decades after a speech at the U.N. At the same time a debt ceiling crisis was looming, as the "extraordinary measures" needed to avoid surpassing it were reaching an end. Earlier in the year, the fiscal cliff had been overcome, and sequestration cuts had been enacted, but without raising the debt limit. Now, Republicans had offered to raise the debt limit enough to operate for one year, but only with a one-year delay in implementation of Obamacare, a vote on tax reform, and fast-tracking the construction of the Keystone XL pipeline. In response, Obama reiterated his unwillingness to negotiate on the issue of debt repayment. Thus deadlocked, most of the federal government shut down on October 1st, and would remain so until the projected breach of the debt limit October 17th, when an agreement was reached to suspend the debt ceiling until early 2014, along with a continuing resolution to fund government. Later, Iran nuclear talks resumed, and by the end of November, an interim agreement had been reached between it and the P5+1. Also in November, China declared an air defense zone over the East China

Sea, which was flouted by America and allies, gunmen opened fire in L.A. International Airport and a mall in New Jersey, two more states voted to approve same-sex marriage, and the Senate voted to end the filibuster on most executive and judicial branch nominees. Most partisan n-grams over the period include: `‘reopen, govern’`, `‘republican, govern, shutdown’`, `‘faith, credit, unit’`, `‘clean, continu, resolut’`, `‘repeal, afford, care’`, `‘employ, mandat’`, `‘delay, individu, mandat’`, `‘readi, prime, time’`, `‘credit, rate’`, `‘obamacar, train, wreck’`, `‘defund, obamacar’`, `‘medic, devic, tax’`, `‘presid, health, care’`, `‘secur, border’`, `‘foreign, polici’`, `‘protect, american’`, `‘allow, vote’`, `‘sexual, orient’`, `‘refus, negoti’`, `’south, korea’`

- *A peak in polarization in January 2015.* In this month, terrorists belonging to Al-Qaeda Yemen attacked the satirical newspaper Charlie Hebdo in Paris. Solidarity demonstrations are attended by over one million people and 40 world leaders. Houthi rebels take the Yemeni capital of Sanaa, forcing the President to resign. Fighting resumes in Eastern Ukraine, stoked by clandestine Russian operatives. A series of massacres occur around Baga, Nigeria as Boko Haram takes the city. New York city policemen turn their backs on the Mayor de Blasio, in protest against his stance supporting police protesters, as he delivers a eulogy for a fallen policemen. An Argentinian prosecutor dies in mysterious circumstances just before testifying in congress about an investigation into President Kirchner. The Supreme Court agrees to hear a consolidated case on same-sex marriage. Most partisan n-grams this month include: `‘afford, care, act’`, `‘presid, health, care’`, `‘african, american’`, `‘right, act’`, `‘enforc, law’`, `‘rule, law’`, `‘fund, depart, homeland’`, `‘broken, immigr, system’`, `‘execut, amnesti’`, `‘illeg, immigr’`, `‘job, creation’`, `‘human, life’`, `‘foreign, affair’`, `‘radic, islam’`, `‘world, war’`. Polarization may have resulted from the absence of any single dominant domestic theme in the news this month.

- *A short-lived surge in how conservative stations were in March 2015.* In this month, a scathing Justice Department report on the circumstances of the Ferguson police shooting was released. Following its release two policemen were shot during otherwise peaceful protests outside the police station. Tensions over police shootings, treatment of minorities, and respect for police are coming to a head. Republicans in Congress attempt to tie funding for the Department of Homeland Security to a rollback of executive actions President Obama had taken to protect certain groups of illegal immigrants from deportation, following the impasse in 2013 on comprehensive immigration reform. The news broke for the first time that the Benghazi panel had discovered that Clinton exclusively used her own private email server while Secretary of State, and these emails may not have been preserved as required by law. Clinton stated that she had done this for convenience, and announced that she had asked the State Department to release her emails. There is a crescendo in rhetoric against a nuclear deal being negotiated with Iran – Israeli Prime Minister Netanyahu address a joint session of congress and claims it would "all but guarantee that Iran gets weapons," and a large group of Senators writes to Iran warning that any deal reached could be revised by the next president "with the stroke of a pen." A controversial "Religious Freedom Restoration Act" was passed in Indiana that would allow individuals or companies to assert as a defence in legal proceedings that their exercise of religion may be burdened; many argued that it is targeted at LGBT groups. Most partisan n-grams this month include: `'african, american'`, `'tax, credit'`, `'execut, branch'`, `'religi, freedom'`, `'rule, law'`, `'depart, homeland, secur'`, `'secur, fund'`, `'comprehens, immigr, reform'`, `'play, polit'`, `'prime, minist, netanyahu'`, `'common, sens'`, `'illeg, immigr'`, `'family, busi'`, `'ballist, missil'`, `'bad, deal'`, `'red, line'`. "Department of Homeland Security" suddenly takes on a very partisan, liberal hue.

These examples can help us understand how events and diagnostic n-grams are interacting to affect slants over time.

Appendix B: Mechanical Turk user interface

I developed a graphical user interface to facilitate quick and correct labeling of snippets for users of Mechanical Turk (MTurk). As MTurk workers are paid a piece rate, they have strong incentive to answer as quickly as possible. This is however tempered by a desire by most to maintain high approval rates for their completed tasks, as this leads to more and better work in the future.

Workers can view a task before accepting to work on it, and can choose whether or not to continue working on your tasks altogether. One thing I do is to note that the work is for an academic-research purpose, in the hopes that this will generate good will and intrinsic motivation. Then, in order to attract the best workers, and elicit the best answers[1], a researcher needs to make it as easy as possible for workers to answer quickly, and at the same time, facilitate the selection of correct responses. We do this in several ways in our user interface (Figure B.1).

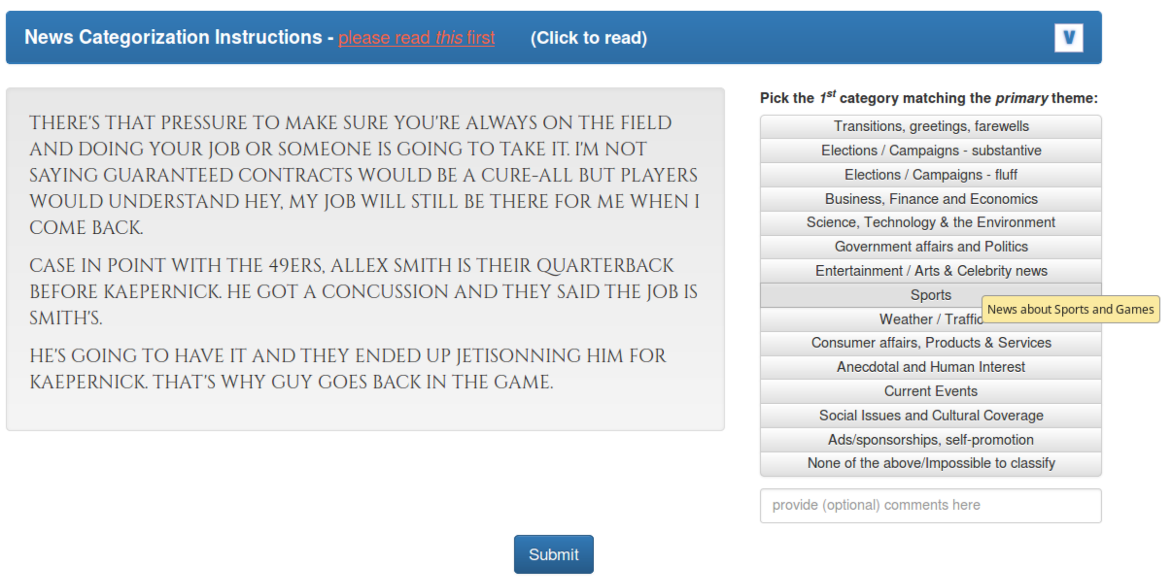


Figure B.1: Mechanical Turk user interface

First, I make it clear that it is necessary to, and make it easy to, read the instructions when completing the first task. The text "please read this first" blinks in bright red when the task is

[1] besides offering a higher piece rate

first opened. Clicking there opens the instructions as an overlay containing the instructions. Second, I take care to fit on a wide variety of screen sizes without the need for the user to scroll down to complete and submit a response, which makes the their work more efficient and thus more lucrative. Third, I take care to make the snippets as readable as possible, with clear fonts, and line heights, spacing and line wrap widths optimized for reading. As noted previously, snippets themselves are also constructed to be an easily digestible size. Fourth, I only ask questions that are applicable given the user's previous responses, so the questions are asked dynamically. Figure B.2 shows how it looks when a user is asked for supplemental labels (in this case because "current events" was the selected topic category). Fifth, we make it easy to tell what is the correct answer without having to return to the instructions. By hovering a mouse over any answer (label), the user get a tooltip giving the description for that answer. In Figure B.1, the user's mouse is over the "sports" topic category. By hovering over it, a tooltip has appeared with the description. Finally, a submission is verified to be complete before being accepted.

The instructions to MTurk workers (reproduced in Figure B.3), while somewhat long, are meant to accomplish several goals. First, it tries to give a quick understanding of the purpose of the task. Second it can serve as a reference when users are uncertain which is the best response (which is not uncommon); the rules (determine the primary topic, then choose the *first* category that fits well) are meant to clarify correct responses, thus reducing variation in response and prediction. Third, it serves to comfort those workers who would be scared off by ambiguity because they're worried their answers might get rejected, hurting their chances of getting future work. Note that importantly, the instructions should normally only need to be consulted once per user, and that the bulk of the instructions (the label descriptions) is very easily accessible by hovering over the actual answer buttons (see Figure B.1).

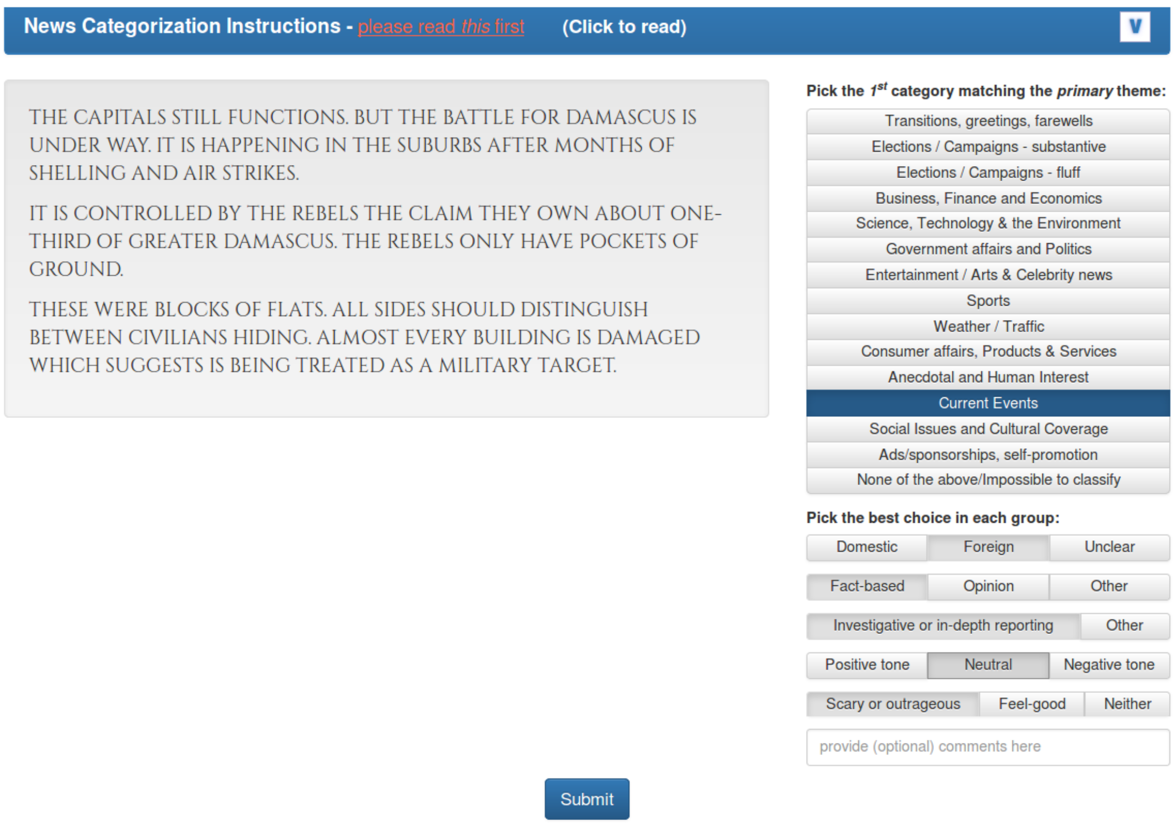


Figure B.2: User interface showing applicable supplemental questions

Pick the category that best fits the *primary* theme in the provided text (or the theme of the middle segment if it's unclear which theme predominates). **If the primary theme can fit into more than one category, choose *the first that fits, from top to bottom*** (many times the best answer is towards the bottom). For example, the awarding of a Nobel prize in Physics might fit into *Science* or *Current Events*, but since the Science category comes first, that is the correct category.

The > symbol means a new person is speaking (and it may be much later or on an unrelated topic); without a >, assume the same person is continuing. All text comes from closed captioning (CC) for U.S. television and radio news since 2010. It may have frequent errors in spelling (you can often sound it out), omissions, gaps, etc. Just do your best to categorize what you get; sometimes the right answer is debatable. You can use the comment field if you feel there is something we should know about.

The categories and their descriptions are below; you can also see the descriptions by hovering your mouse over the buttons. Depending on your answer, you may be asked to choose one or more supplemental categories. These are not meant to reflect your views on or knowledge of the news being discussed, but rather simply what is presented.

News Category	Description
Transitions, greetings, farewells	Beginnings and ends of segments, transitions between reports, 'teasers' for upcoming stories, chit-chat, etc. (nothing of substance)
Elections / Campaigns - substantive	Candidates' and parties' plans/platforms/policies or expert opinions about policies, candidates' records/background, interviews with or speeches by candidates, election results (U.S.)
Elections / Campaigns - fluff	The horse race (who's leading), polls/standings, speculation about outcomes, campaign tactics/strategy, talking heads/commentators (about an election), campaign updates (U.S.)
Business, Finance and Economics	News about the economy, markets, and business strategy, performance or business climate
Science, Technology & the Environment	Scientific & technological reporting; nature & environmental coverage (incl. climate, but not weather; incl. medical sciences but not health care)
Government affairs and Politics	US or International Government actions, functioning, policies, and impacts (including military, except use of force), and coverage of (non-U.S.-election-related) Politics
Entertainment / Arts & Celebrity news	News relating to celebrities and the music / arts / literature / entertainment industries
Sports	News about Sports and Games
Weather / Traffic	Current reports or forecasts of traffic or weather (but not the government response to weather, climate science, transport policy, or impacts of extreme weather events)
Consumer affairs, Products & Services	News reports about new or notable products or services, or reviews
Anecdotal and Human Interest	Stories which are perhaps interesting but mostly inconsequential. Usually quick/easy to produce. Often with only local significance and unknown or negligible wider impacts.
Current Events	Notable or consequential recent (at the time of presentation) news events.

Figure B.3: Labeling Instructions

	Terrorism, crime, protests, significant achievements, war/military use of force, etc.
Social Issues and Cultural Coverage	Societal or cultural themes; concerning communities, cities, health, society, religion, gender, race, human interactions, or trends (which don't fit into earlier categories)
Ads/sponsorships, self-promotion	Contents of apparently paid ads, sponsors that fund the network, or network self-promotion (such as for other programs)
None of the above/Impossible to classify	Content that doesn't fit in any of these categories (please explain in comment); or incomprehensible, non-English, jibberish, or missing text; impossible to determine

The supplemental categories and their descriptions:

Domestic	**Foreign**	**Unclear**
US domestic coverage	Non-US, international/global coverage, or US gov't foreign relations	Unknown/unclear/neither/both

Fact-based	**Opinion**	**Other**
Fact or evidence-based statements (regardless of your view of them, or whether the statements are correct)	Editorial, opinion, conjecture, speculation, possibilities	Other/neither/unknown (e.g. interviews)

Investigative or in-depth reporting	**Other**
Hard-hitting or in-depth factual reports and investigations that take time and effort to produce (e.g. providing history/background/context, exposing corruption or scandals); serious policy analysis	Anything else (so most stories); non-investigative reporting, including reports about others' investigations

Positive tone	**Neutral**	**Negative tone**
Positive in tone or sentiment (regardless of your view of the issue)	Neutral in tone or sentiment	Negative in tone or sentiment (regardless of your view of the issue)

Scary or outrageous	**Feel-good**	**Neither**
Appeals to emotion: Stories which are scary, shocking, or outrageous in their topic or presentation (whether or not you feel that way)	Appeals to emotion: Pleasant, sunny, inspirational or feel-good stories	Neither scary/outrageous nor feel-good, or no appeals to emotion (Most stories are neither)

Figure B.3: Labeling Instructions (continued)

Bibliography

ALLIANCE FOR AUDITED MEDIA, "Top 25 U.S. Newspapers for March 2013," September, *2013*, `http://auditedmedia.com/blog`.

BARTELS, LARRY M., "Beyond the Running Tally: Partisan Bias in Political Perceptions," *Political Behavior*, Vol. 24 (2), pp. 117–150, June, *2002*.

BAUM, MATTHEW A. & TIM GROELING, "New Media and the Polarization of American Political Discourse," *Political Communication*, Vol. 25 (4), pp. 345–365, *2008*.

BAUM, MATTHEW A. & ANGELA S. JAMISON, "The *Oprah* Effect: How Soft News Helps Inattentive Citizens Vote Consistently," *The Journal of Politics*, Vol. 68 (4), pp. 946–959, November, *2006*.

BEBCHUK, LUCIAN A. & JESSE M. FRIED, *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*, Cambridge, MA: Harvard University Press, *2004*.

_ "Pay Without Performance: Overview of the Issues," *Journal of Applied Corporate Finance*, Vol. 17 (4), pp. 8–23, *2005*.

BEBCHUK, LUCIAN ARYE & JESSE M. FRIED, "Executive Compensation as an Agency Problem," *The Journal of Economic Perspectives*, Vol. 17 (3), pp. 71–92, *2003*.

BERNHARDT, DAN, STEFAN KRASA, & MATTIAS POLBORN, "Political polarization and the electoral effects of media bias," *Journal of Public Economics*, Vol. 92 (5–6), pp. 1092–1104, June, *2008*.

BERTRAND, MARIANNE & SENDHIL MULLAINATHAN, "Are CEOS Rewarded for Luck? The Ones without Principals Are," *The Quarterly Journal of Economics*, Vol. 116 (3), pp. 901–932, August, *2001*.

BIRD, STEVEN, EDWARD LOPER, & EWAN KLEIN, *Natural Language Processing with Python*: O'Reilly Media Inc. *2009*.

BLEI, DAVID M., ANDREW Y. NG, & MICHAEL I. JORDAN, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3 (Jan), pp. 993–1022, *2003*.

BOLTON, PATRICK, JOSÉ SCHEINKMAN, & WEI XIONG, "Executive Compensation and Short-Termist Behaviour in Speculative Markets," *Review of Economic Studies*, Vol. 73 (3), pp. 577–610, *2006*.

BROWN, LAWRENCE D. & YEN-JUNG LEE, "The Impact of SFAS 123R on Changes in Option-Based Compensation," *SSRN eLibrary*, *2007*.

CAPLAN, BRYAN, *Myth of the Rational Voter*: Princeton University Press, *2007*.

CASTILLO, CARLOS, GIANMARCO DE FRANCISCI MORALES, MARCELO MENDOZA, & NASIR KHAN, "Says who?: automatic text-based content analysis of television news," In *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data using Natural Language Processing*, pp. 53–60: ACM Press, *2013*.

COFFEE JR, JOHN C, "Liquidity versus control: The institutional investor as corporate monitor," *Colum. L. Rev.*, Vol. 91, p. 1277, *1991*.

COMSTOCK, GEORGE A. & ERICA SCHARRER, *The Psychology of Media and Politics*: Academic Press, *2005*.

CORE, J. E, R. W HOLTHAUSEN, & D. F LARCKER, "Corporate governance, chief executive officer compensation, and firm performance," *Journal of Financial Economics*, Vol. 51 (3), p. 371–406, *1999*.

DELLAVIGNA, STEFANO & ETHAN KAPLAN, "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics*, Vol. 122 (3), pp. 1187–1234, August, *2007*.

DITTMANN, INGOLF & ERNST MAUG, "Lower Salaries and No Options? On the Optimal Structure of Executive Pay," *The Journal of Finance*, Vol. 62 (1), pp. 303–343, *2007*.

DJANKOV, SIMEON, CARALEE MCLIESH, TATIANA NENOVA, & ANDREI SHLEIFER, "Who owns the media?" Working Paper 8288, National Bureau of Economic Research, *2001*.

FIORINA, MORRIS P., SAMUEL J. ABRAMS, & JEREMY C. POPE, *Culture War? The Myth of a Polarized America*, New York: Longman, 2nd edition, *2005*.

FOX, CHRISTOPHER, "A stop list for general text," *ACM SIGIR Forum*, Vol. 24 (1–2), *1989*.

FREY, BRUNO S. & MARGIT OSTERLOH, "Yes, Managers Should Be Paid Like Bureaucrats," *Journal of Management Inquiry*, Vol. 14 (1), pp. 96–111, March, *2005*.

FRIEDENBERG, ROBERT V, *Communication consultants in political campaigns: Ballot box warriors*: Greenwood Publishing Group, *1997*.

FRYDMAN, CAROLA & RAVEN E. SAKS, "Executive Compensation: A New View from a Long-Term Perspective, 1936-2005," *Review of Financial Studies*, Vol. 23 (5), pp. 2099 –2138, May, *2010*.

GABAIX, XAVIER & AUGUSTIN LANDIER, "Why Has CEO Pay Increased So Much?*," *Quarterly Journal of Economics*, Vol. 123 (1), pp. 49–100, December, *2010*.

GALLUP, "Americans' Trust in Mass Media Sinks to New Low," September, *2016*, `http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx`.

_ "Trust in Government," January, *2019*, `https://news.gallup.com/poll/5392/trust-government.aspx`.

GARVEY, GERALD T. & TODD T. MILBOURN, "Asymmetric benchmarking in compensation: Executives are rewarded for good luck but not penalized for bad," *Journal of Financial Economics*, Vol. 82 (1), pp. 197–225, October, *2006*.

GENTZKOW, MATTHEW, "Television and Voter Turnout," *The Quarterly Journal of Economics*, Vol. 121 (3), pp. 931–972, August, *2006*.

GENTZKOW, MATTHEW, EDWARD GLAESER, & CLAUDIA GOLDIN, "The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered," In *Corruption and Reform: Lessons from America's Economic History*, Chicago: University of Chicago Press, pp. 187–230, *2006*.

GENTZKOW, MATTHEW & JESSE SHAPIRO, "Media bias and reputation," *Journal of Political Economy*, Vol. 114 (2), pp. 280–316, *2006*.

GENTZKOW, MATTHEW & JESSE M. SHAPIRO, "Competition and Truth in the Market for News," *The Journal of Economic Perspectives*, Vol. 22 (2), pp. 133–154, April, *2008*.

— "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, Vol. 78 (1), pp. 35–71, January, *2010*.

GENTZKOW, MATTHEW, JESSE M. SHAPIRO, & MICHAEL SINKINSON, "Competition and Ideological Diversity: Historical Evidence from US Newspapers," Working Paper 18234, National Bureau of Economic Research, *2012*.

GERBER, ALAN, DEAN S. KARLAN, & DANIEL BERGAN, "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," SSRN Scholarly Paper ID 903812, Social Science Research Network, Rochester, NY, *2006*.

GIBBONS, ROBERT, "Incentives in Organizations," *The Journal of Economic Perspectives*, Vol. 12 (4), pp. 115–132, *1998*.

GORDON, ROBERT J. & KEVIN J. MURPHY, "Unresolved Issues in the Rise of American Inequality," *2007*, Available from `http://www.people.fas.harvard.edu/~idew/papers/BPEA_final_ineq.pdf`.

GROSECLOSE, TIM, *Left turn: How liberal media bias distorts the American mind*: St. Martin's Press, *2011*.

GROSECLOSE, TIM, STEVEN D LEVITT, & JAMES M SNYDER, "Comparing interest group scores across time and chambers: Adjusted ADA scores for the US Congress," *American political science review*, Vol. 93 (01), pp. 33–50, *1999*.

GROSECLOSE, TIM & JEFFREY MILYO, "A Measure of Media Bias," *The Quarterly Journal of Economics*, Vol. 120 (4), pp. 1191–1237, November, *2005*.

HALL, BRIAN J. & KEVIN J. MURPHY, "Stock options for undiversified executives," *Journal of Accounting and Economics*, Vol. 33 (1), pp. 3–42, February, *2002*.

— "The Trouble with Stock Options," *The Journal of Economic Perspectives*, Vol. 17 (3), pp. 49–70, *2003*.

HARTZELL, JAY C & LAURA T STARKS, "Institutional Investors and Executive Compensation," *The Journal of Finance*, Vol. 58 (6), pp. 2351–2374, December, *2003*.

HODGE, FRANK D., SHIVARAM RAJGOPAL, & TERRY J. SHEVLIN, "How do Managers Value Stock Options and Restricted Stock?" *SSRN eLibrary*, *2006*.

Houser, Daniel, Rebecca Morton, & Thomas Stratmann, "Turned on or turned out? Campaign advertising, information and voting," *European Journal of Political Economy*, Vol. 27 (4), pp. 708–727, December, *2011*.

Howard, Jeremy & Sebastian Ruder, "Universal Language Model Fine-tuning for Text Classification," In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1, pp. 328–339, *2018*.

Hwang, Byoung-Hyoun & Seoyoung Kim, "It pays to have friends," *Journal of Financial Economics*, Vol. 93 (1), pp. 138–158, July, *2009*.

Jensen, Jacob, Ethan Kaplan, Suresh Naidu, & Laurence Wilse-Samson, "Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech," *Brookings Papers on Economic Activity*, Vol. 45 (2 (Fall)), pp. 1–81, *2012*.

Jensen, Michael C., Kevin J. Murphy, & Eric G. Wruck, "Remuneration: Where We've Been, How We Got to Here, What are the Problems, and How to Fix Them," *SSRN eLibrary*, *2004*.

Jensen, Robert & Emily Oster, "The Power of TV: Cable Television and Women's Status in India," *The Quarterly Journal of Economics*, Vol. 124 (3), pp. 1057–1094, August, *2009*.

Jiang, L., H. Lu, M. Xu, & C. Wang, "Biterm Pseudo Document Topic Model for Short Text," In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 865–872, November, *2016*.

Kadan, Ohad & Jeroen M. Swinkels, "Stocks or Options? Moral Hazard, Firm Viability, and the Design of Compensation Contracts," *Review of Financial Studies*, Vol. 21 (1), pp. 451 –482, January, *2008*.

Kahneman, Daniel, *Thinking, Fast and Slow*: Farrar, Straus and Giroux, *2011*.

Kearney, Melissa S. & Phillip B. Levine, "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing," Working Paper 19795, National Bureau of Economic Research, *2014*.

Kern, H. L. & J. Hainmueller, "Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes," *Political Analysis*, Vol. 17 (4), pp. 377–399, July, *2009*.

Khan, Raihan, Ravi Dharwadkar, & Pamela Brandes, "Institutional ownership and CEO compensation: a longitudinal examination," *Journal of Business Research*, Vol. 58 (8), pp. 1078–1088, August, *2005*.

La Ferrara, Eliana, Alberto Chong, & Suzanne Duryea, "Soap Operas and Fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, Vol. 4 (4), pp. 1–31, October, *2012*.

Lefebvre, Mathieu, Ferdinand Vieider, Mathieu Lefebvre, & Ferdinand Vieider, "Reining in Excessive Risk Taking by Executives: Experimental Evidence," *2010*, `http://econpapers.repec.org/`.

Levendusky, Matthew, *How Partisan Media Polarize America*: University of Chicago Press, pp.208, *2013*.

LEVENDUSKY, MATTHEW S., "Why Do Partisan Media Polarize Viewers?" *American Journal of Political Science*, Vol. 57 (3), pp. 611–623, July, *2013*.

MALMENDIER, ULRIKE & GEOFFREY TATE, "Superstar CEOs*," *Quarterly Journal of Economics*, Vol. 124 (4), pp. 1593–1638, November, *2010*.

MANN, THOMAS E & NORMAN J ORNSTEIN, *It's Even Worse Than It Looks: How the American Constitutional System Collided With the New Politics of Extremism*: Basic Books, *2012*.

MARTIN, GREGORY J. & ALI YURUKOGLU, "Bias in Cable News: Persuasion and Polarization," *American Economic Review*, Vol. 107 (9), pp. 2565–2599, September, *2017*.

MCCARTY, NOLAN, KEITH T POOLE, & HOWARD ROSENTHAL, *Polarized America: The dance of ideology and unequal riches*: MIT Press, *2016*.

MCMILLAN, JOHN & PABLO ZOIDO, "How to subvert democracy: Montesinos in Peru," *Journal of Economic Perspectives*, Vol. 18 (4), pp. 69–92, *2004*.

MERITY, STEPHEN, NITISH SHIRISH KESKAR, & RICHARD SOCHER, "Regularizing and Optimizing LSTM Language Models," *arXiv preprint arXiv:1708.02182*, *2017*.

MERITY, STEPHEN, CAIMING XIONG, JAMES BRADBURY, & RICHARD SOCHER, "Pointer Sentinel Mixture Models," *arXiv:1609.07843 [cs]*, September, *2016*, arXiv: 1609.07843.

MULLAINATHAN, SENDHIL & ANDREI SHLEIFER, "Media bias," Working Paper 9295, National Bureau of Economic Research, *2002*.

_ "The Market for News," *The American Economic Review*, Vol. 95 (4), pp. 1031–1053, September, *2005*.

MURPHY, KEVIN J., "Explaining Executive Compensation: Managerial Power versus the Perceived Cost of Stock Options," *The University of Chicago Law Review*, Vol. 69 (3), pp. 847–869, *2002*.

MURPHY, KEVIN J., ORLEY C. ASHENFELTER, & DAVID CARD, "Chapter 38 Executive compensation," In *Handbook of Labor Economics*, Vol. Volume 3, Part 2: Elsevier, pp. 2485–2563, *1999*.

MURPHY, KEVIN J. & JÁN ZÁBOJNÍK, "CEO Pay and Appointments: A Market-Based Explanation for Recent Trends," *The American Economic Review*, Vol. 94 (2), pp. 192–196, May, *2004*.

OLMSTEAD, KENNETH, MARK JURKOWITZ, AMY MITCHELL, & JODI ENDA, "How Americans Get TV News at Home," Technical Report, Pew Research Center for the People and the Press, *2013*, `http://www.journalism.org/2013/10/11/how-americans-get-tv-news-at-home/`.

OYER, P. & S. SCHAEFER, "Personnel economics: Hiring and incentives," *NBER Working Paper*, *2010*.

PBS FRONTLINE, "Interview with Frank Luntz," November, *2004*, `http://www.pbs.org/wgbh/pages/frontline/shows/persuaders/interviews/luntz.html`.

PEW RESEARCH CENTER, "In Changing News Landscape, Even Television is Vulnerable," Technical Report, Pew Research Center for the People and the Press, *2012*, `http://www.people-press.org/2012/09/27/in-changing-news-landscape-even-television-is-vulnerable/`.

__ "State of the News Media 2014," Technical Report, Pew Research Center for the People and the Press, *2014a*, http://www.journalism.org/2014/03/26/state-of-the-news-media-2014-key-indicators-in-media-and-news/.

__ "Political Polarization in the American Public," Technical Report, Pew Research Center for the People and the Press, *2014b*, http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/.

__ "Political Polarization and Media Habits," Technical Report, Pew Research Center for the People and the Press, *2014c*, http://www.journalism.org/2014/10/21/political-polarization-media-habits/.

__ "State of the News Media 2016," Technical Report, Pew Research Center for the People and the Press, *2016*, http://www.journalism.org/2016/06/15/state-of-the-news-media-2016/.

POOLE, KEITH T, "Recovering a basic space from a set of issue scales," *American Journal of Political Science*, pp. 954–993, *1998*.

POOLE, KEITH T & HOWARD L ROSENTHAL, *Ideology and Congress*: Transaction Publishers, 2nd edition, *2007*.

PORTER, M.F., "An Algorithm for Suffix Stripping," *Program*, Vol. 14 (3), pp. 130–137, *1980*.

PRIOR, MARKUS, "Media and Political Polarization," *Annual Review of Political Science*, Vol. 16 (1), pp. 101–127, *2013*.

RABINOWITZ, MITCHELL, MARIA ACEVEDO, SARA CASEN, MYRIAH ROSENGARTEN, MARTHA KOWALCZYK, & LINDSAY BLAU PORTNOY, "Distinguishing facts from beliefs: Fuzzy categories," *Psychology of Language and Communication*, Vol. 17 (3), pp. 241–268, *2013*.

SCHEINKMAN, JOSÉ A. & WEI XIONG, "Overconfidence and Speculative Bubbles," *The Journal of Political Economy*, Vol. 111 (6), pp. 1183–1219, December, *2003*.

SCHMITT, KATHLEEN M., ALBERT C. GUNTHER, & JANICE L. LIEBHART, "Why Partisans See Mass Media as Biased," *Communication Research*, Vol. 31 (6), pp. 623–641, December, *2004*.

SECURITIES AND EXCHANGE COMMISSION, U.S. SEC, "SEC Votes to Adopt Changes to Disclosure Requirements Concerning Executive Compensation and Related Matters," Jul, *2006*, http://www.sec.gov/news/press/2006/2006-123.htm Accessed Dec 2010.

__ "EDGAR database [Definitive Proxy Statements Form DEF 14A, 2008-9]," Accessed Oct-Nov, *2010*, Author's compilation from http://www.sec.gov/edgar.shtml.

SHLEIFER, ANDREI & ROBERT W. VISHNY, "Large Shareholders and Corporate Control," *The Journal of Political Economy*, Vol. 94 (3), pp. 461–488, June, *1986*.

STANDARD & POOR'S, "Compustat Execucomp [1999-2009]," Accessed November, *2010a*, http://www.compustat.com. Retrieved from Wharton Research Data Service.

__ "Compustat Xpressfeed North America Fundamental Annual 2009," Accessed November, *2010b*, http://www.compustat.com. Retrieved from Wharton Research Data Service.

STROMBERG, DAVID, “Radio’s Impact on Public Spending,” *The Quarterly Journal of Economics*, Vol. 119 (1), pp. 189–221, February, *2004*.

SULLIVAN, MARGARET, “Waiter, Where’s Our (Political) Spinach?” *The New York Times*, p. SR10, Mar, *2016*.

SUNSTEIN, CASS R., *Republic.com*, Princeton, NJ: Princeton University Press, *2001*.

TADDY, MATT, “Measuring Political Sentiment on Twitter: Factor Optimal Design for Multinomial Inverse Regression,” *Technometrics*, Vol. 55 (4), pp. 415–425, November, *2013*.

THOMSON REUTERS, “Insider Filing Feed [2008-9],” Accessed November, *2010a*, Retrieved from Wharton Research Data Service.

_ “Institutional (13F) Holdings [2008],” Accessed November, *2010b*, Retrieved from Wharton Research Data Service.

TOSI, HENRY L. & LUIS R. GOMEZ-MEJIA, “The Decoupling of CEO Pay and Performance: An Agency Theory Perspective,” *Administrative Science Quarterly*, Vol. 34 (2), pp. 169–189, June, *1989*.

VALLONE, ROBERT P., LEE ROSS, & MARK R. LEPPER, “The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre.,” *Journal of personality and social psychology*, Vol. 49 (3), p. 577, *1985*.

WEISBACH, MICHAEL S., “Review: Optimal Executive Compensation versus Managerial Power: A Review of Lucian Bebchuk and Jesse Fried’s "Pay without Performance: The Unfulfilled Promise of Executive Compensation",” *Journal of Economic Literature*, Vol. 45 (2), pp. 419–428, June, *2007*.

WILSHIRE, “Wilshire 5000 Index,” Accessed October, *2010*, `http://www.wilshire.com/Indexes/Broad/Wilshire5000/`.

WITTMAN, DONALD A., *The Myth of Democratic Failure: Why Political Institutions Are Efficient*, Chicago: University Of Chicago Press, *1997*.

YIN, JIANHUA & JIANYONG WANG, “A Dirichlet Multinomial Mixture Model-based Approach for Short Text Clustering,” In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’14, pp. 233–242, New York, NY, USA: ACM, *2014*.

YU, HSIANG-FU, CHIA-HUA HO, YU-CHIN JUAN, & CHIH-JEN LIN, “LibShortText: A Library for Short-text Classification and Analysis,” *2013*.

Curriculum Vitæ

Justin Thomas Briggs has a Bachelor of Science in Computer Science and Mathematics from the University of Maryland College Park (2001), a Diploma in Economics from the University of London, School of Economics and Political Science, as an external student (2008), and a Master of the Arts in Economics from George Mason University (2010), during which he attended one term at Hertford College, Oxford University.

Since 2003, he has worked at Leidos (previously SAIC), building research software and doing data science. Currently Lead Software Engineer and Research Scientist, he has worked for various customers, including DARPA, IARPA, Army Research Laboratory (ARL), Domestic Nuclear Detection Office (DNDO) and others. These projects spanned many domains, including automation, machine learning, pattern recognition, computer vision, data fusion and remote detection using a variety of phenomenologies, and with applications in areas such as transportation safety, nuclear non-proliferation, situational awareness, force protection, and airborn control systems.

Prior to that he worked at AOL to create automated tests for the AOL client, and before that at a startup company in Spain, creating cutting-edge anti-piracy and cryptography software.

He speaks English, French, and Spanish, and codes in C++, C, Python, Java, Matlab, Stata, and R, among others. He is a member of the Association of Computing Machinery (ACM) and the American Economic Association (AEA). Hobbies include travel photograpy, woodworking, soccer and cycling.

Papers in peer-reviewed journals

With Alex Tabarrok, 2014. "Firearms and suicides in US states." *International Review of Law and Economics*, 37: 180-188.

With John Kaufhold and Jaroslaw Tuszynski, 2013. "A method for automatic manifest verification of container cargo using radiography images." *Journal of Transportation Security*, 6(4): 339-356.