

[Benjamin E. Lauderdale](#) and Alexander Herzog Measuring political positions from legislative speech

**Article (Accepted version)
(Refereed)**

Original citation:

Lauderdale, Benjamin E. and Herzog, Alexander (2016) Measuring political positions from legislative speech. *Political Analysis*. ISSN 1047-1987

© 2016 Society for Political Methodology

This version available at: <http://eprints.lse.ac.uk/66498/>
Available in LSE Research Online: May 2016

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Measuring Political Positions from Legislative Speech

.

Supplemental Appendix

Benjamin E. Lauderdale
London School of Economics
b.e.lauderdale@lse.ac.uk

Alexander Herzog
Clemson University
aherzog@clemson.edu

May 3, 2016

1. DATA

1.1. *Overview and Summary Statistics*

Table A1 provides an overview of the speeches and debates included in our analysis. In an average electoral period, there were 8,707 speeches across 779 debates in the Irish Dáil and 7,464 speeches across 584 debates in the US Senate. The 112th and 113th Senates stick out at as the two least productive Senates in terms of number of speeches and debates, which is most likely the result of legislative gridlock after the Republican party took over the House in the 2010 election.

The average debate in our data set consists of 12 speakers, ranging from a minimum of 5 speakers (our lower threshold) to as many as 73 speakers. At the level of individual speeches, we find an average of 683 and 552 words per speech and legislature. The longest individual speech in our data set is 37,610 words long, which is senator Ted Cruz’s (TX-R) 21-hour filibuster speech in September 2013.

1.2. *Sources*

For Ireland, we retrieved speeches from “DPSI: Database of Parliamentary Speeches in Ireland” (Herzog and Mikhaylov 2013), which includes all speeches from the Irish Dáil from 1919 to 2013. Information in this database was collected from the Houses of the Oireachtas (the Irish national parliament) and is distributed under the *Public Sector Information (PSI) Licence for Re-Use of Information*, No. 2005/08/01. Speeches from the US Senate were collected from the digital version of the Congressional Record using a web scraper and parser written in Python.

Table A1: Summary statistics for speeches and debates

<i>Number of observations by legislature</i>				
	No. of speakers*	No. of speeches	No. of debates	No. of unique words
<i>Irish Dáil</i>				
29th (2002–2007)	165	10,073	938	40,193
30th (2007–2011)	165	7,341	619	32,471
<i>US Senate</i>				
104th (1995–1997)	102	9,480	667	34,798
105th (1997–1999)	100	8,072	606	33,492
106th (1999–2001)	101	7,898	625	33,893
107th (2001–2003)	101	8,557	686	34,030
108th (2003–2005)	99	8,020	635	34,804
109th (2005–2007)	101	8,067	588	34,294
110th (2007–2009)	102	7,746	622	32,439
111th (2009–2011)	108	7,286	589	31,030
112th (2011–2013)	101	5,188	442	26,618
113th (2013–2015) [†]	104	4,327	375	26,049

<i>Number of speakers by debate and length of speeches</i>						
	Number of speakers by debate			Length of speeches by debate (<i>N</i> words)		
	mean	min	max	mean	min	max
<i>Irish Dáil</i>						
29th (2002–2007)	11	5	46	686	18	7,189
30th (2007–2011)	12	5	47	681	19	8,107
<i>US Senate</i>						
104th (1995–1997)	14	5	70	559	16	9,161
105th (1997–1999)	13	5	65	549	18	8,688
106th (1999–2001)	13	5	72	539	17	15,372
107th (2001–2003)	12	5	73	518	19	13,080
108th (2003–2005)	13	5	57	555	17	20,202
109th (2005–2007)	14	5	58	549	16	10,099
110th (2007–2009)	12	5	59	561	20	7,076
111th (2009–2011)	12	5	57	573	19	6,985
112th (2011–2013)	12	5	43	519	19	8,587
113th (2013–2015) [†]	12	5	67	599	18	37,610

Notes:

* Speakers only include members with a seat in the legislature.

[†] Data includes all speeches and debates until November 20, 2014.

2. IRISH DÁIL

2.1. *Additional Analysis of Classification and Uncertainty versus Wordfish*

In the main text, we show plots of the relationships between individual TDs Wordshoal score and coalition membership, and uncertainty estimates for the 30th Dáil. In Figure A1 we show the corresponding plots for the 29th Dáil.

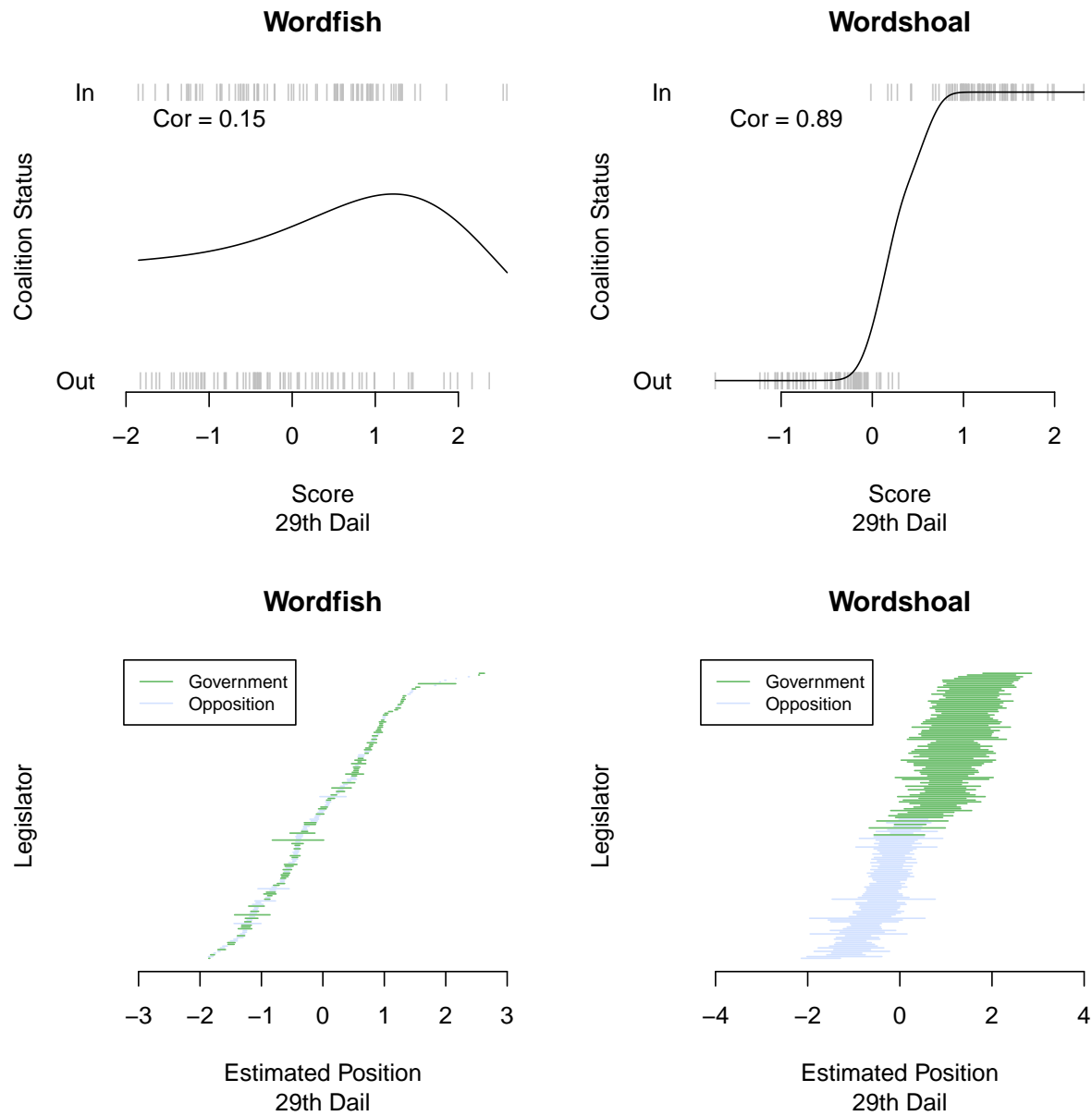


Figure A1: The association between the estimated positions of each legislator and their status as members of the coalition versus opposition, with correlation and local linear smooth, under Wordfish (left) and our approach (right), for the 29th Dáil. In the bottom row, we show the 95% intervals associated with the estimates for each legislator under Wordfish (left) and Wordshoal (right).

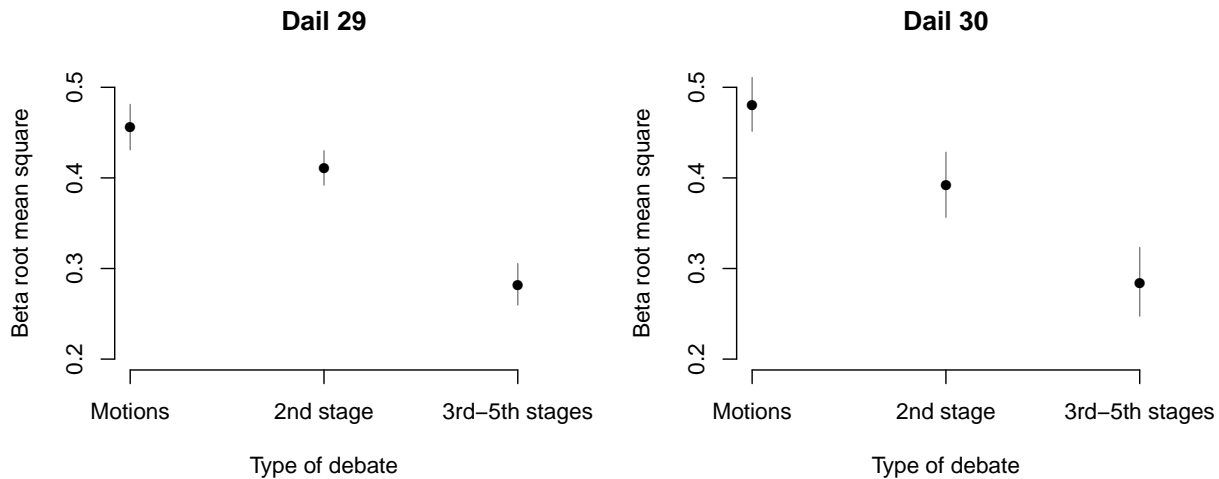


Figure A2: Weighted root mean square of β_j by type of debate for the 29th and 30th Dáil.

2.2. Analysis of Variation in Disagreement by Legislative Stage

The Wordshoal model facilitates assessments of the relative extent to which different kinds of debates align with the estimated common dimension. The greater the magnitude of the β_j , the greater the extent to which the common dimension predicts speech variation in a given debate. As our summary statistic for β_j , we use the root mean square of the β_j , weighted by the number of speeches in each debate speeches_j .

$$\sqrt{\frac{\sum_j \text{speeches}_j \cdot \beta_j^2}{\sum_j \text{speeches}_j}} \quad (1)$$

In the paper, we utilize this statistic only for the US Senate, here we assess it across different types of parliamentary debates in Ireland. The majority of debates in our data take place during the second reading of a bill, which is the most important legislative stage after which a bill is formally accepted or rejected. The second most common type are motions, which are an instrument of parliament to scrutinize the work of the government. This includes ad hoc motions on topical issues, seasonal adjournment debates, and (less frequently)

motions of confidence in the government or in individual cabinet members (Gallagher 2010).

Figure A2 shows the weighted root mean square of β_j for motions, second-stage bills, and for debates during the remaining legislative stages, which we group together because of the small number of observations at each stage. We find high government-opposition division during debates on motions, which are mostly used by the opposition parties to express grievance over government decisions, and hence elicit a strong divide between the opposition and government members. For similar reasons, we find nearly as high polarization in second-stage debates. In a parliamentary system with a majority government, legislation is almost exclusively initiated by the cabinet. The main debate of these bills then provide the opposition with another opportunity to criticize the government for its work. Once a bill has passed the second stage and is all but guaranteed to enter into law, debates become less strongly associated with the government-opposition dimension.

Dail 29 and 30

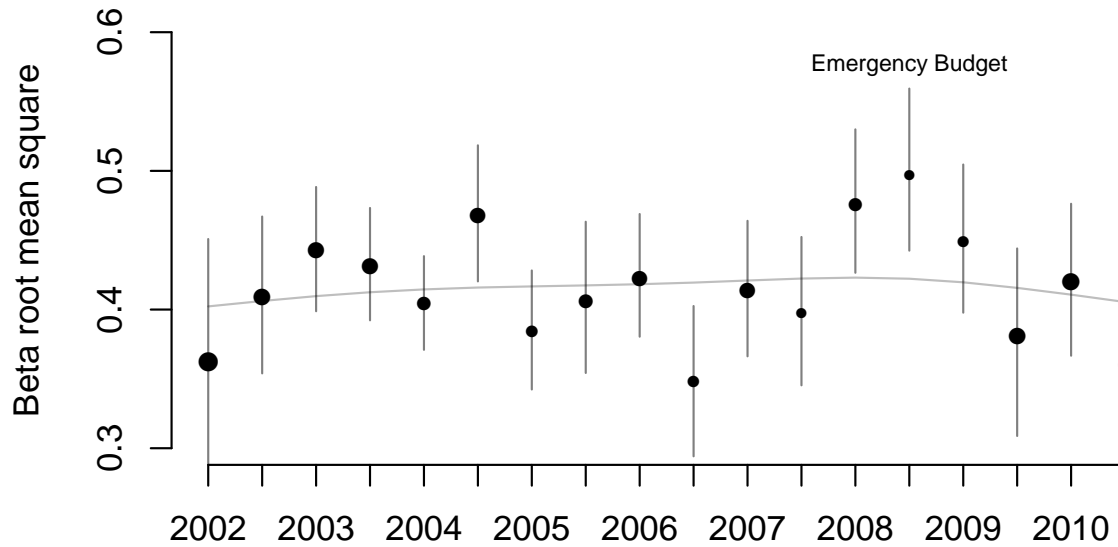


Figure A3: Weighted root mean square of β_j over six-months periods. The size of each point is proportional to the relative number of second-stage debates in each period.

2.3. Analysis of Temporal Variation in Disagreement

During the period covered by our data (2002–2011), Ireland went from boom to bust, with the end of rapid economic growth (“Celtic Tiger”) in early 2000, followed by a period of average growth until the collapse of the financial market and banking system in 2008/09. In Figure A3, we plot the weighted root mean square of β_j for six-months period for both the 29th and 30th Dáil. Because we have found above that government-opposition polarization is higher during the second reading of a bill, we plot each point in Figure A3 proportional to the relative number of second-stage debates in each period.

The plot shows an increase in the root mean square of β_j during 2008 despite a relative

small number of second-stage debates. This increase coincides with the onset of the crisis and the first emergency budget of the government in October 2008. Our analysis therefore provides evidence for an increase in government-opposition polarization when legislators started to debate solutions to the financial crisis. After 2008, polarization returns to normal levels and drops below the trend line in the second half of 2009 and even further so in 2010, shortly before the Fianna Fáil-Green coalition collapsed over internal divisions. It seems that the increasing disagreement within the governing coalition decreased the observable government-opposition divide, which is in line with findings in Herzog and Benoit (2015) based on Irish budget debates.

We also find significant deviations from the trend line in 2004/2005 and in the second half of 2006. Both time periods coincide with the electoral cycle: local and European elections (held on the same day) in June 2004 and a national election in early 2007. While more data would be needed to assess whether these are consistent patterns, there is some suggestion here that government-opposition polarization increase during local/European elections, but decrease before a national election, which also coincide with a decrease in second-stage debates.

3. US SENATE

3.1. *Wordfish applied to a Single US Senate Debate*

Here, we demonstrate how Wordfish works when applied to a single US Senate debate. Wordfish does recover useful information about the relative positions of senators in the debate; however there is clear evidence of contamination from speakers who go off topic or otherwise are not speaking like everyone else. We consider the final Senate debate before the 60-39, party-line vote on cloture for the Affordable Care Act on 23 December 2009. When we apply Wordfish to the speeches given by all the senators who spoke, we get estimates that largely reflect the party of the speakers (Figure A4). Largely, Democrats and Republicans are at opposite ends of the recovered dimension that best predicts variation in word use in this debate. However, there are two clear and substantial errors in location: the Democratic senators Baucus and Conrad are located at the far right of the dimension, despite the fact that both supported the ACA, along with all other Democrats in the Senate.

To understand why these Senators are placed on the right extreme of the recovered dimension, it is necessary to examine the word parameters for this debate. Figure A4 includes two plots showing the relationship between individual words and the recovered dimension. The word stems “health”, “coverag” and “preexist” are associated with the left of the recovered dimension, while the the words “obama”, “debt” and “deficit” are associated with the right of the dimension.

When Senator Conrad spoke in this debate, this is how he began:

Mr. CONRAD. Mr. President, I rise this morning not to talk about health care but to talk about the other critical matter that faces this body before we leave this session for the holidays and that is the matter of extending the debt limit of the United States. Let me start by saying it is imperative that we extend the debt limit. If we do not, the United States would default on its debt. The

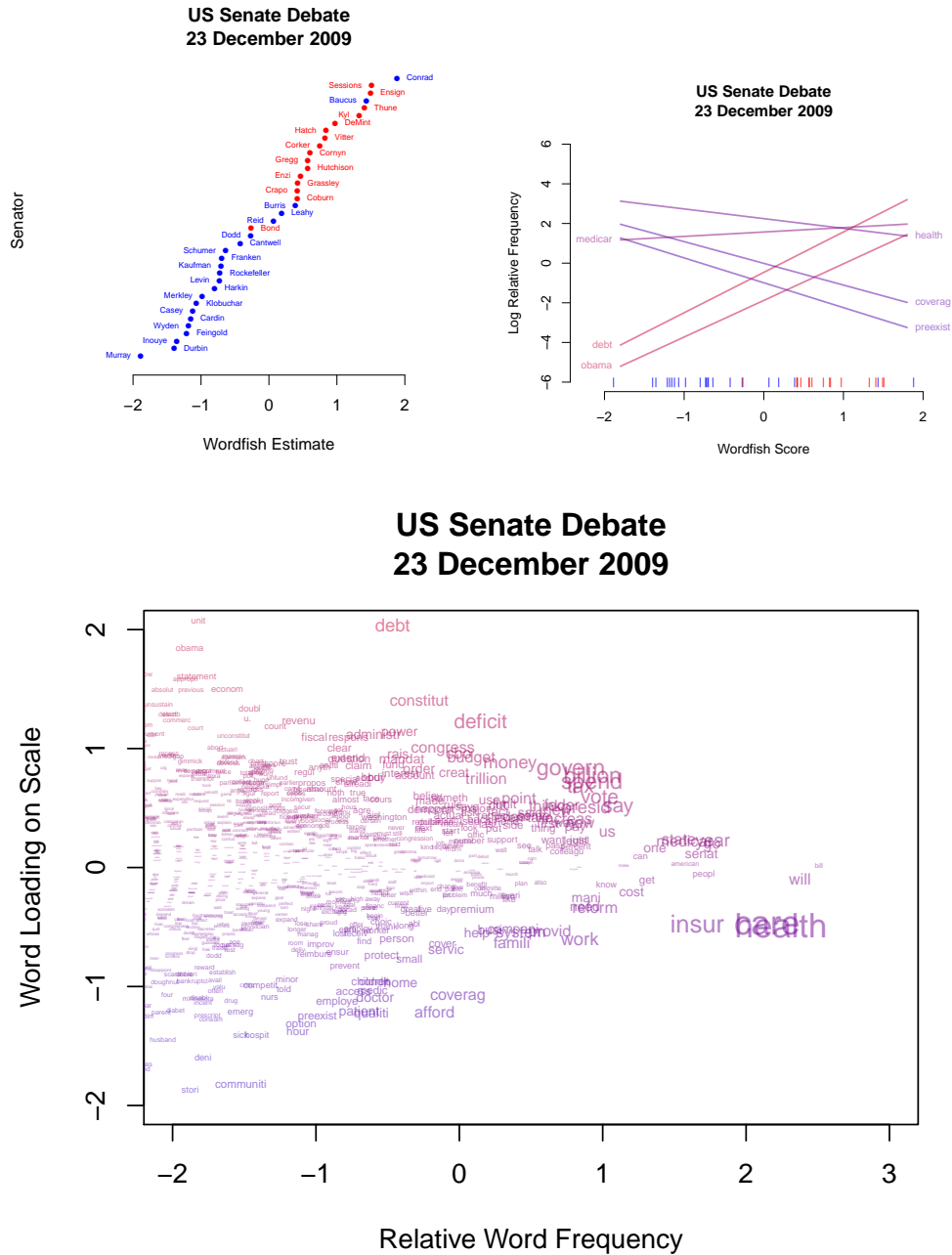


Figure A4: Wordfish estimates from the US Senate debate on 23 December 2009, preceding the vote on cloture on the Affordable Care Act. At top left, the Wordfish scores estimated for senators who spoke in the debate. At top right, the fitted log relative frequencies of several salient words in the debate, as a function of Wordfish score. At bottom, the relative word frequency and loading parameters for all the words used the debate, scaled in size to approximate aggregate influence on the estimates.

consequences for this country and the global economy would be nothing short of catastrophic.

If you think about the problems created in world markets by the fact that Dubai defaulted on \$40 billion of debt, think of what it would mean to global markets if the United States were to default on \$12 trillion of debt.

For those who say this is Obama's fault—no. This is not Obama's fault. He has been in office 11 months. I remind everyone that he walked into the biggest mess in 70 years—deficits and debt exploding, joblessness skyrocketing, economic growth plummeting. All that was happening before Barack Obama became President of the United States. He did not create the economic mess, he inherited it. He did not create the fiscal mess, he inherited it. Those are things he had to take on as the new President.

There were record deficits and a doubling of the national debt, there was the worst recession since the Great Depression, financial market and housing crises, ongoing wars in Iraq and Afghanistan, and an unsustainable long-term budget outlook with everything going in the wrong direction....

By going off-topic, and doing so in a way that involved using the word “debt”, “deficit” and other negative language repeatedly, Senator Conrad gave a speech that heavily used words that were associated with being on the right in the health care debate. This is an example of how text scaling can fail due to contamination from variation regarding topic.

Senator Baucus provides as example of how variation in style can cause similar contamination. Baucus served as the floor manager for the debate, and used his role to directly engage with the arguments that Republicans were making, especially regarding the questions of constitutionality and implications for the deficit and debt. In response to a statement by Senator Kyl (R - AZ),

Mr. BAUCUS. I ask my good friend from Arizona, is it not true that the last statement from CBO, on the degree to which the underlying legislation does or does not reduce the deficit, stated that the legislation reduces the deficit by \$132 billion—that is the last statement after addressing the deficit—and also stating that at the end of the decade, the deficit will be reduced between \$630 billion and \$1.3 trillion? Isn't that the last statement from CBO addressing the question on whether this legislation reduces or increases the deficit. Isn't that true?

And later...

Mr. BAUCUS. I wonder if the Senator is aware that CBO this morning at 9:57 sent an e-mail to all relevant staff that its estimates with regard to budget deficit reduction still stand, still hold. CBO still estimates this legislation results in a \$132 billion deficit reduction. That was an e-mail sent today. Is the Senator aware of that e-mail?

Mr. KYL. I did not see that e-mail. I assume that is the same communique about which the Senator from Alabama is talking. It shows you exactly why this is so confusing and why I am a little bit concerned about the politicization of the CBO.

And in response to Senator Hutchison (R - TX),

Mrs. HUTCHISON. Mr. President, the 10th amendment says:

The powers not delegated to the United States by the Constitution are reserved to the States.

In this bill, a State such as Texas and many other States that have taken full responsibility for insurance plans for their employees and teachers will have to justify any change in those terms to the Federal Government.

The majority claims the commerce clause gives them the power to do what is in this bill. But what they fail to mention is the power to regulate interstate commerce has not been the basis for a robust role in insurance regulation.

This is an encroachment of the Federal Government into a role left to the States in the Constitution. The 10th amendment is being eroded by an activist Congress, and it is time to stop it now.

I urge a vote to uphold this point of order.

The PRESIDING OFFICER. The Senator from Montana.

Mr. BAUCUS. Mr. President, the bill before us is clearly an appropriate exercise of the commerce clause. We further believe Congress has power to enact this legislation pursuant to the taxing and spending powers. This bill does not violate the 10th amendment because it is an appropriate exercise of powers delegated to the United States, and because our bill fundamentally gives States the choice to participate in the exchanges themselves or, if they do not choose to do so, to allow the Federal Government to set up the exchanges fully within the provisions as interpreted by the Supreme Court of the 10th amendment.

I urge my colleagues to vote against the point of order.

Because scaling models like Wordfish act on word counts, but cannot understand the context of word usage, Baucus's engagement with Republican arguments and language puts

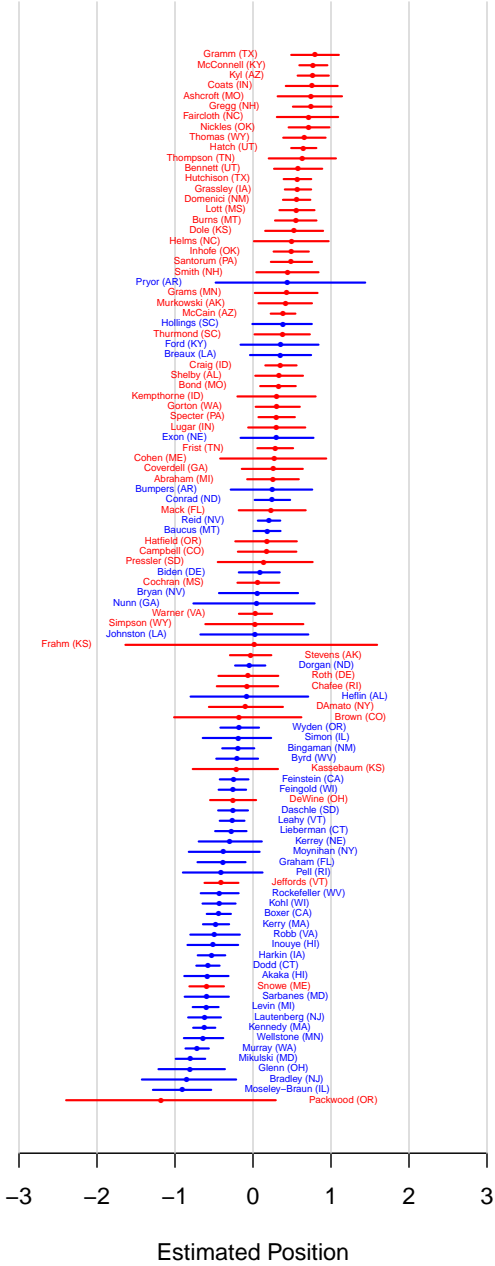
him at the right extreme of the dimension. Most Senators gave set speeches in this debate, rather than directly responding to one another. The fact that Baucus did not do this leads Wordfish to place him in a way that does not reflect his political position on the issue.

The contamination we observe for Baucus and Conrad occurs on a much larger scale if one combines texts on many topics. The logic of the approach we advocate is to extract the useful information that Wordfish recovers about positioning in each debate—recall that aside from Baucus and Conrad, the estimates are largely reasonable for this debate—and combine that information across debates. Unless particular individuals are consistently going off topic in ways that are consistently related to the words that are politically loaded in each debate, the errors we see for Baucus and Conrad will tend to average out in the aggregate. However, if certain individuals are in fact more inclined to engage with opposing arguments, this may lead to some bias towards positioning them among the opposing party.

3.2. *104th to 113th US Senate Estimates*

The following pages show the estimates for each Senate, subject to the assumption that individual senators have fixed scores. The general trend towards decreasing overlap between parties in estimated positions is readily visible across the series of plots. The apparent increase in overlap for the 113th Senate is largely due to new Senators whose locations are not estimated precisely because of the relatively small number of debates occurring during 2013-2014, and should not be taken as indicative that speech polarization has begun to decline.

Senate 104



Senate 105

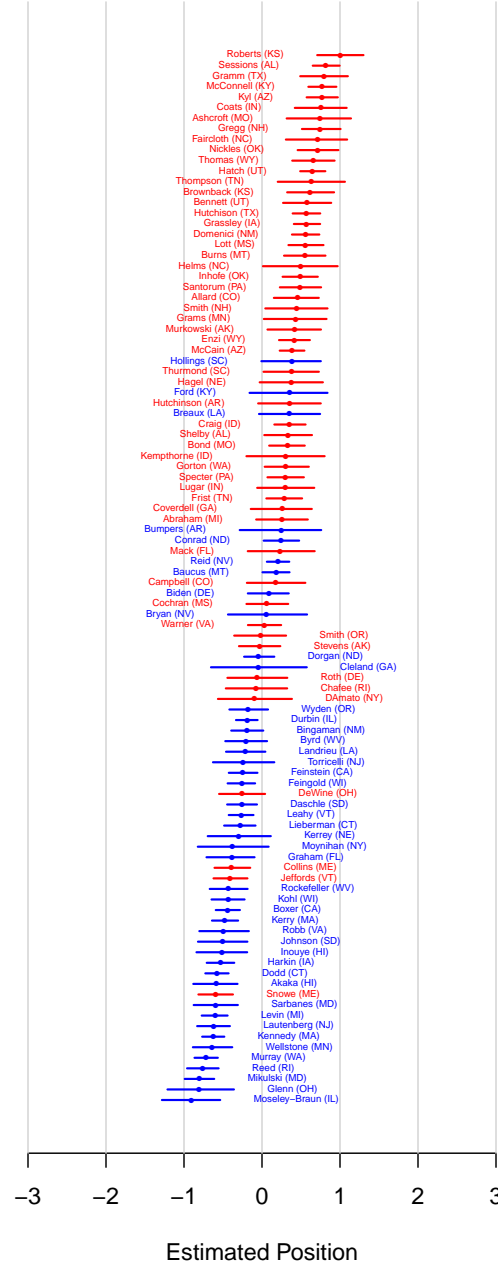
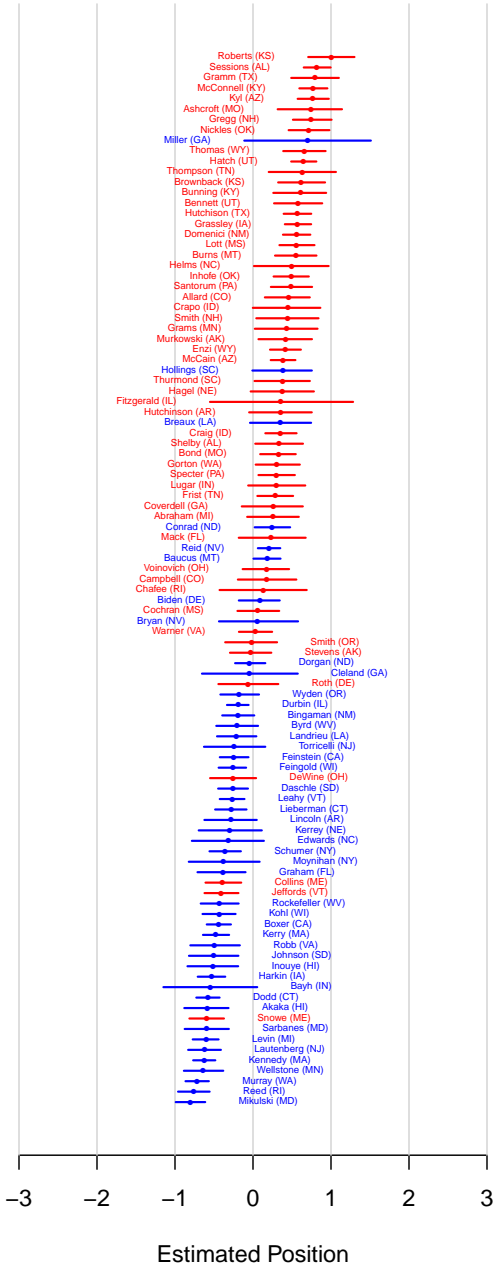


Figure A5: Wordshoal estimates for the 104th and 105th US Senates.

Senate 106



Senate 107

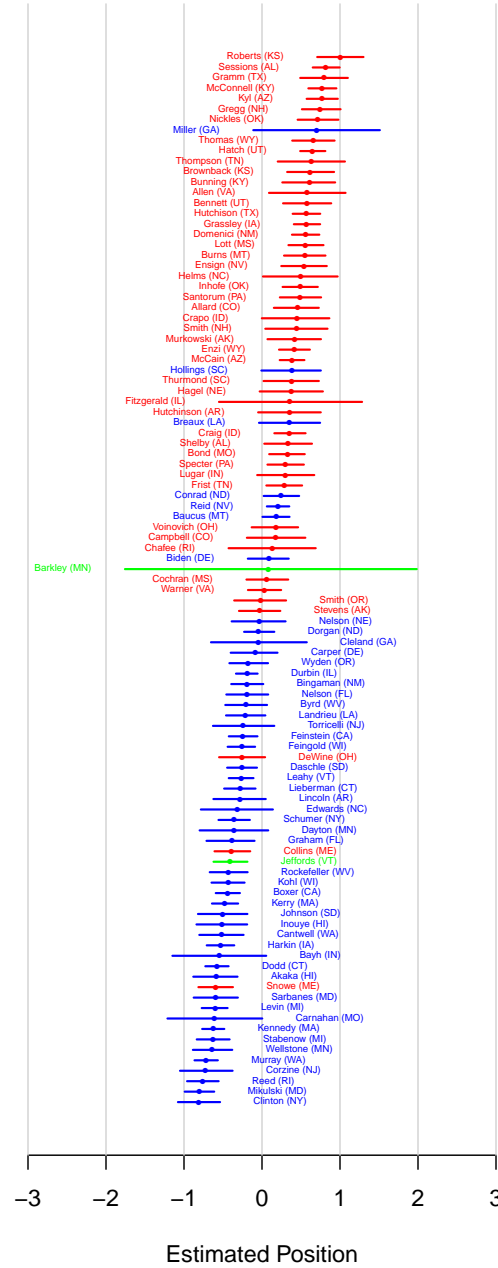
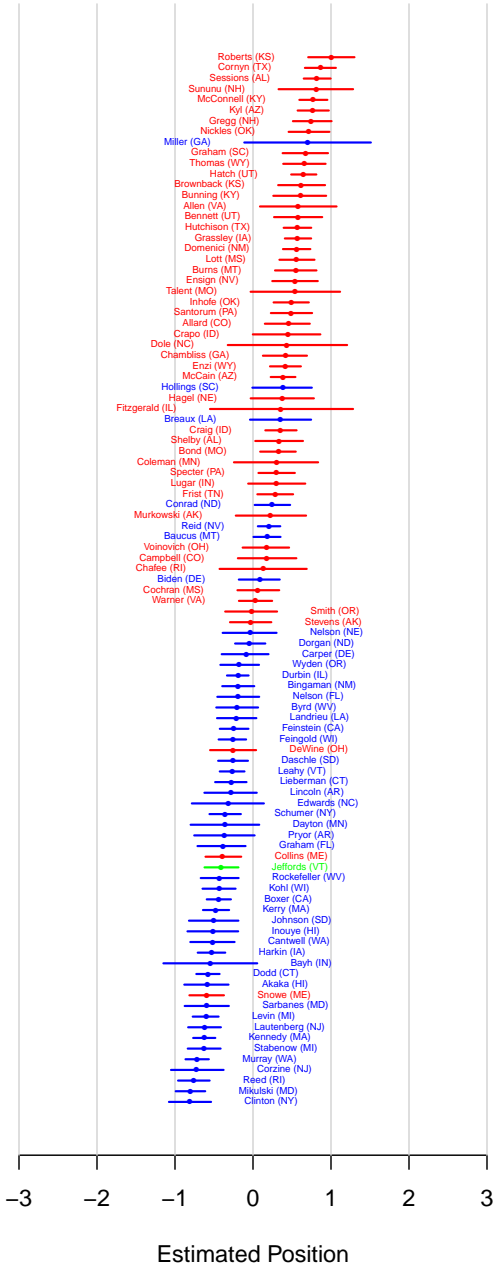


Figure A6: Wordsnoal estimates for the 106th and 107th US Senates.

Senate 108



Senate 109

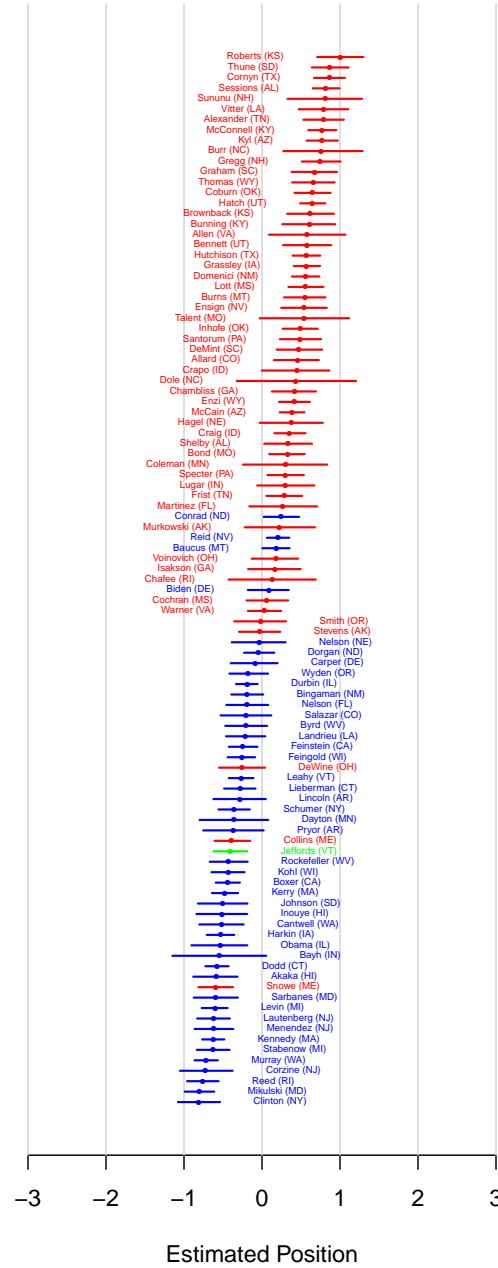
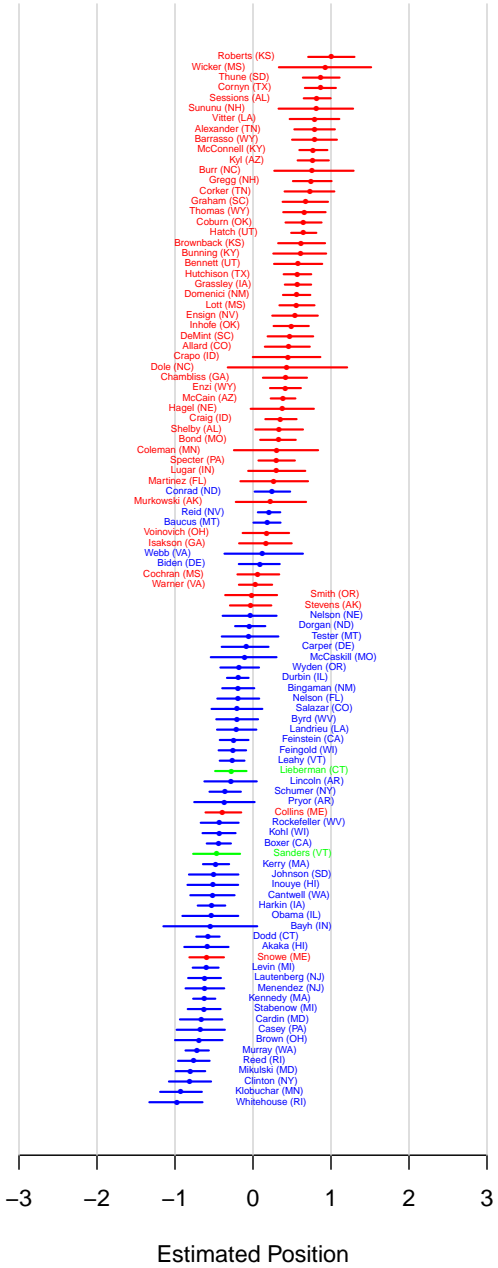


Figure A7: Wordshoal estimates for the 108th and 109th US Senates.

Senate 110



Senate 111

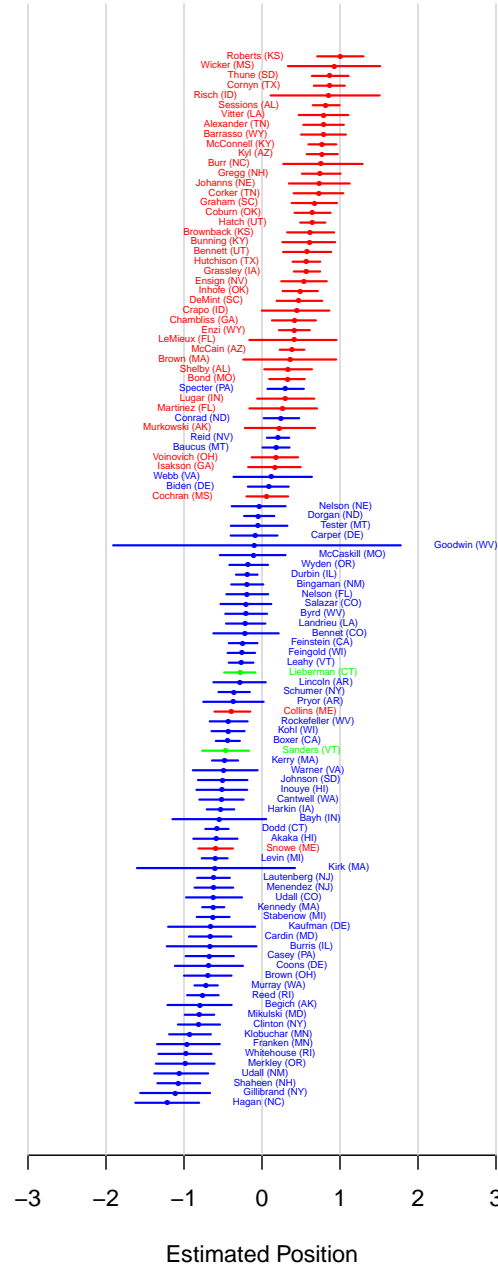
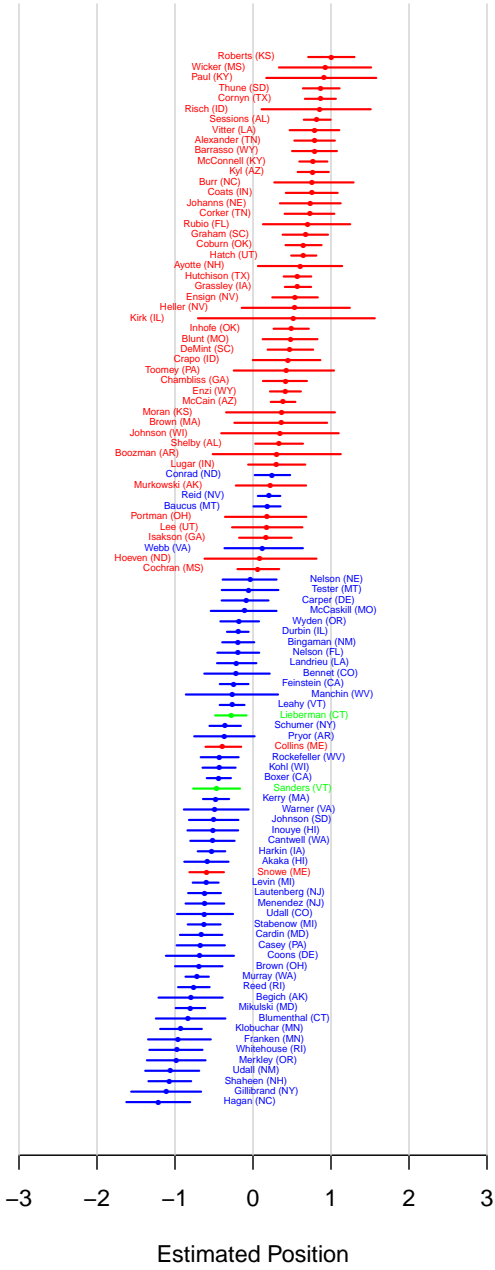


Figure A8: Wordshoal estimates for the 110th and 111th US Senates.

Senate 112



Senate 113

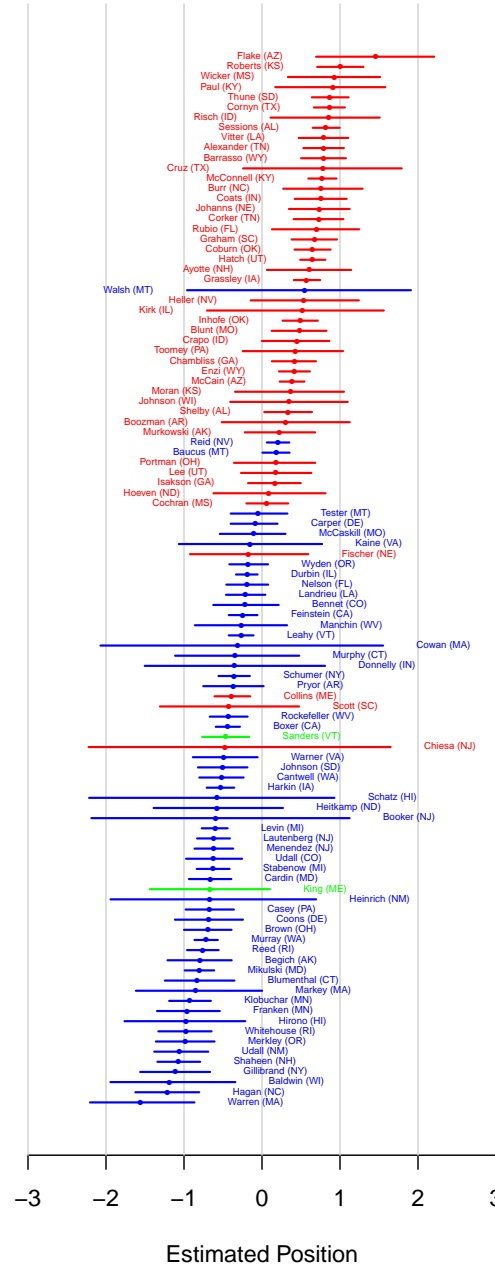


Figure A9: Wordshoal estimates for the 112th and 113th US Senates.

3.3. *Analysis of Electoral Cycles in US Senate Speeches and Votes*

In the main text we report variation in average debate loadings over time. Does this variation within Congresses follow a pattern, either for the roll-call data or the speech data? The top panel of Figure A10 shows that variation within the congressional calendar is not large, however there is some suggestion that speech polarization is generally highest in the middle of a Congress. The most distinctive feature of the speech polarization series is that it has its lowest level in the month (October) immediately preceding a congressional elections. This suggests some strategic tendency towards moderation, either in the scheduling of debates or in individual speech behavior (or both). The pattern for roll-calls (bottom panel) shows some indication of an upward trend over the congressional cycle and some evidence of lower polarization in the two months before an election relative to that trend.

Across these analyses, we find evidence that the trend in speech polarization has some of the same temporal features as the trend in polarization in voting behavior, but what is unambiguously clear is that speech polarization is far more variable over the period of time we examine. While voting behavior has been consistently highly polarized by party since 1995, the polarization of speech behavior by senators has increased substantially due to both replacement and increasingly polarized debates, as well as increasing or decreasing in response to external events, the legislative agenda, and the electoral calendar.

These patterns highlight one of the reasons that speeches are worth studying in their own right. Rhetoric can be dialed up or down. Speeches are at once more visible and less costly opportunities for senators to emphasize or deemphasize partisan differences, as the political climate dictates. Our results suggest that senators use these opportunities in response to political conditions, even as aggregate voting behavior changes relatively little.

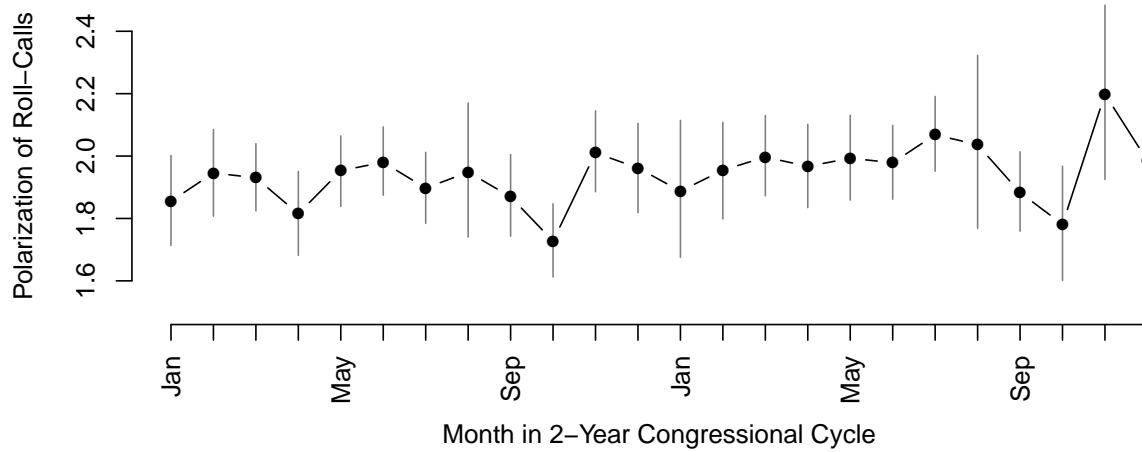
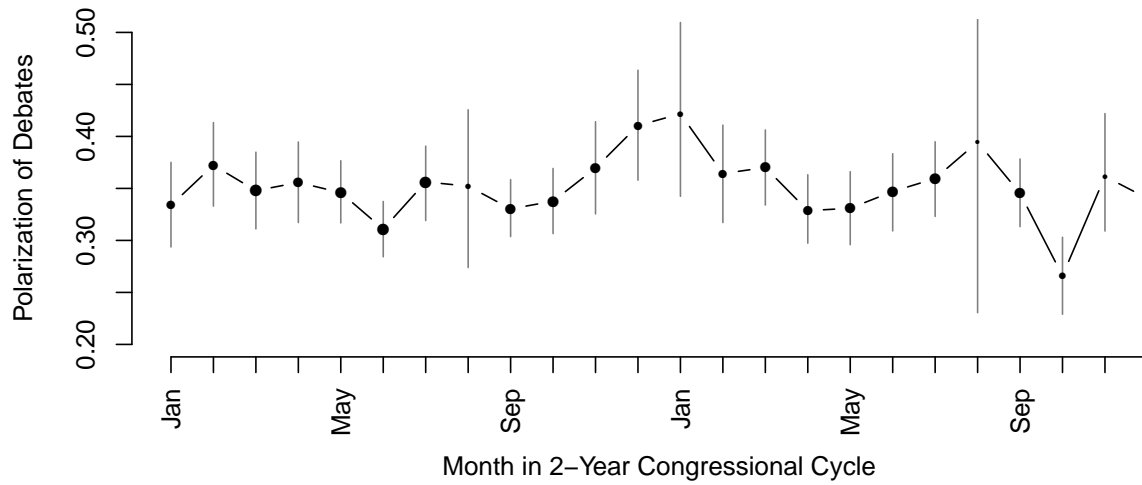


Figure A10: Average magnitude of debate loadings over the 24 months of a Congress, across the 104th to 113th Congress. The size of each point is proportional to the number of debates.

Table A2: Models predicting standardized within-party variation in average donor positions using standardized within-party variation in roll-call scores and speech scores.

Roll Call Score	0.57 (0.06)		0.40 (0.06)
Speech Score		0.56 (0.06)	0.37 (0.06)
Intercept	-0.01 (0.06)	-0.00 (0.06)	-0.01 (0.05)
R ²	0.32	0.30	0.43
Adj. R ²	0.32	0.30	0.42
Num. obs.	213	213	213

3.4. *Additional Analysis of Roll Call, Speech and Donation Scores*

In Table A2 we present regression results on standardized roll-call, speech, and donation scores. In Table A3 we present regression results on non-standardized roll-call, speech, and donation scores, with party dummy variables.

Table A3: Models predicting average donor positions using unstandardized party, roll-call scores, and speech scores. Dummy variables are included for all parties to provide party-specific intercepts, omitting the general constant.

Roll Call Score		0.37		0.25
		(0.04)		(0.04)
Speech Score			0.48	0.34
			(0.05)	(0.05)
Democrat	-0.82	-0.39	-0.62	-0.39
	(0.03)	(0.05)	(0.03)	(0.05)
Independent	-1.58	-0.80	-1.35	-0.90
	(0.33)	(0.29)	(0.28)	(0.27)
Republican	0.82	0.43	0.63	0.42
	(0.03)	(0.05)	(0.03)	(0.04)
R ²	0.86	0.90	0.90	0.92
Adj. R ²	0.86	0.90	0.90	0.92
Num. obs.	214	214	214	214

3.5. Gender Differences in Speech

There is a long-running debate about whether and how female legislators represent their constituents differently than male legislators. Some research on roll-call scores has shown that the average female representative in the US votes to the left of the average male representative from the same party; however, a recent study by Simon and Palmer (2010) suggests that this may be an artifact of the seats to which women are most likely to be elected. They construct comparisons of female members of the US House, not to all other same-party representatives, but rather to male same-party representatives who immediately preceded or succeeded them. In this analysis, there is no difference between same-party, same-seat men and women in roll-call behavior.

But roll-calls are not the only way that legislators represent their constituents, and speeches offer opportunities for representatives to differentiate themselves even when they ultimately vote similarly. We apply a variant on the Palmer and Simon identification strategy

here, comparing the 9 female Republican and 20 female Democratic senators in our data set to same-party male senators who preceded them, succeeded them, or served alongside them in the other seat from the same state. At least one such male senator exists for 22 of the 29 females, and so we restrict our analysis to those senators.¹ We compare the female senators' speech scores and roll-call scores to the average scores of the 1-3 same-party males identified through this matching procedure. As in the preceding analysis, we use scores standardized within-party.

On average, the female senators' speech scores are 1.29 to the left of their same-party, same-state male colleagues. This is a strongly significant difference: $t = -7.5$, $p \ll 0.01$, and the 95% interval runs from -1.64 to -0.93. Since these scores are standardized within-party, these indicate that the gender differences are very large relative to the within-party variation in speech scores. Nearly the entire 95% interval is greater than one standard deviation difference. Consistent with the findings of Simon and Palmer (2010), we find no such difference in roll-call voting behavior. The difference in means for roll-call scores is just 0.02, with $t = -0.1$, $p = 0.90$, and the 95% interval runs from -0.34 to 0.30. The difference in means for donation scores is -0.37 , with a 95% interval from -0.86 to 0.12. This is consistent with the results in the paper that suggested that donor behavior can be understood as falling "between" roll-call and speech behavior on average. To check that these results are not the result of a small number of outliers, we also perform a sign test of whether the fraction of women who are to the left/right of their same-party, same-state colleagues is different from 0.5. Out of the 22 female senators, only 1 has speech scores to the right of her same-state male colleagues, providing very strong evidence $p = 0.00001$ against the null hypothesis that these differences arose by chance from individual-level variation in speech behavior. That one case corresponds to a tiny difference: Murkowski

¹Within the time period we consider, there are no same-state male Democratic senators to match to Boxer (CA), Feinstein (CA), Cantwell (WA), Murray (WA), Carnahan (MO), McCaskill (MO), or Shaheen (NH).

(AK) is 0.07 to the right of Stevens (AK). No clear relationship is observed in the sign tests for roll-calls or donations, where 17 and 11 of the 22 women have roll-call and donation scores to the right of the same matched same-party, same-state men, respectively.

At its most basic level, these results indicate that across all Senate debates, female senators more frequently use the words that Democrats use more frequently, whatever those words happen to be in a given debate. A skeptical interpretation of this result is that these debate-level dimensions could just be a mixture of linguistic features of female speech and left speech, and the fact that those attributes are correlated in the population (there are more female Democrats than Republicans) is the only reason we recover a dimension that is a mix of the two. Thus it is reasonable to ask: have we have simply measured a dimension that is a mixture of left-right and male-female?

One way to address this concern is to fit a 2D model at the second level instead of the 1D model we have considered thus far. If the differences between male and female senators were due to a 1D model recovering a mixture of partisan and gender differences in speech, allowing additional dimensions could enable the model to distinguish political variation in speech from gender variation in speech. Fitting a 2D model tells us whether the male-female differences show up in the same debates as the left-right differences (indicating women speak more to the left) or whether they arise in distinct sets of debates (indicating mixing of distinct dimensions). Below, we present results showing that adding a second dimension does nothing to diminish the result that female senators speak to the left of male senators from the same states. Under the 2D model, gender differences remain very large along the axis that separates the parties, and are much smaller in the orthogonal dimension.

Figure A11 shows that in a 2D version of the Wordshoal model, the differences between male and female senators remain along the primary dimension of disagreement. The orthogonal dimension shows a difference only for Republicans. The difference on the first dimension remains greater than one standard deviation of the within-party variation in first dimension

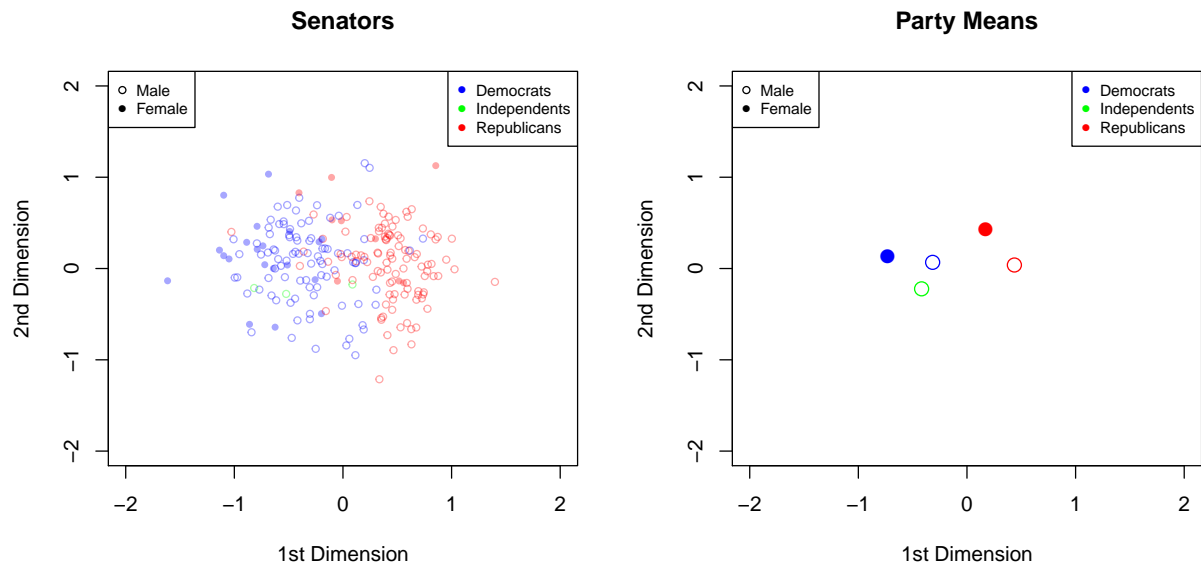


Figure A11: 2D Wordshoal speech scores for male versus female senators, by party.

scores (1.33), and very highly significant ($p = 0.000001$).

REFERENCES

- Gallagher, Michael. 2010. The Oireachtas: President and parliament. In *Politics in the Republic of Ireland*, ed. John Coakley and Michael Gallagher. Routledge chapter 7, pp. 198–229.
- Herzog, Alexander and Kenneth Benoit. 2015. “The most unkindest cuts: speaker selection and expressed government dissent during economic crisis.” *Journal of Politics* 77(4):1157–1175.
- Herzog, Alexander and Slava Mikhaylov. 2013. “DPSI: Database of Parliamentary Speeches in Ireland.” Data Repository available at <http://www.alexherzog.net>.
- Simon, Dennis M. and Barbara Palmer. 2010. “The Roll Call Behavior of Men and Women in the U.S. House of Representatives, 1937–2008.” *Politics and Gender* 6:225–246.

Measuring Political Positions from Legislative Speech*

Benjamin E. Lauderdale
London School of Economics
b.e.lauderdale@lse.ac.uk

Alexander Herzog
Clemson University
aherzog@clemson.edu

May 4, 2016

ABSTRACT

Existing approaches to measuring political disagreement from text data perform poorly except when applied to narrowly selected texts discussing the same issues and written in the same style. We demonstrate the first viable approach for estimating legislator-specific scores from the entire speech corpus of a legislature, while also producing extensive information about the evolution of speech polarization and politically loaded language. In the Irish Dáil, we show that the dominant dimension of speech variation is government-opposition, with ministers more extreme on this dimension than backbenchers, and a second dimension distinguishing between the establishment and anti-establishment opposition parties. In the US Senate, we estimate a dimension that has moderate within-party correlations with scales based on roll-call votes and campaign donation patterns, however we observe greater overlap across parties in speech positions than roll-call positions and partisan polarization in speeches varies more clearly in response to major political events.

**Authors' note:* Replication materials are available online as Lauderdale and Herzog (2016).

1. INTRODUCTION

Measuring the policy positions that parties and politicians take is a key requirement for building and testing theories of intra-party politics, polarization, representation, and policy-making. Traditionally, political scientists have used roll-call votes to estimate the positions of individual legislators (Poole and Rosenthal 1997; Clinton, Jackman and Rivers 2004; Hix, Noury and Roland 2005). Yet in most political systems, legislative votes are either not recorded or individual members seldom deviate from party-line voting because of strong party discipline (Hug 2010). Thus, if one seeks to estimate the diversity of positions taken by legislators both within and across parties, roll-call analysis is of limited use (VanDoren 1990; Carrubba, Gabel, Murrah, Clough, Montgomery and Schambach 2006; Carrubba, Gabel and Hug 2008; Proksch and Slapin 2010; Proksch and Slapin 2015).

In this paper, we propose a new strategy for estimating spatial measures of *expressed disagreement* from legislative speech. We argue that just as the natural unit for legislative voting data is the roll-call, the natural unit for legislative speech is the debate on a given bill. We capture this intuition with a hierarchical factor model for word usage in legislative debates, which we refer to as *Wordshoal* and estimate in two stages.¹ The first stage uses the existing text scaling model “Wordfish” (Slapin and Proksch 2008) to scale word use variation in each debate separately. In the second stage, we use Bayesian factor analysis to construct a common scale from the debate-specific positions estimated in the first stage.

Our method presents the first viable approach to scaling the entire speech corpus of a legislature, producing valid legislator-specific scores on one (or more) underlying general dimension(s) that can be used to study legislative behavior, intra-party politics, and polarization. One of its key innovations is that it allows the meaning and discriminatory power of a given word to vary from debate to debate. For example, the word “debt” may be important

¹A “shoal” is a group of fish, not traveling in the same direction.

to discriminate speakers in a debate on extending health care, while the same word may have little discriminatory power in a debate on the budget deficit, where it will be used heavily by most speakers. The strategy of within-debate scaling addresses a fundamental problem in the analysis of legislative speech, namely that variation in word usage between speeches is both a function of the topic of a debate and the position a legislator takes. Further, our method provides meaningful uncertainty estimates of legislators' aggregated positions, taking into account how often legislators spoke and how consistent they were in expressing their positions across debates.

Like any unsupervised scaling method, the substantive meaning of the legislator-specific scores needs to be determined *ex post* and will depend on the institutional context. We present two applications to demonstrate our approach and how it contributes to our understanding of legislative politics. In the first application, we use speeches from the Irish Dáil as an example of a multi-party parliamentary system. We show that estimated speech scores in this context strongly reflect government-opposition dynamics, but also reveal significant intra-party variation in support versus opposition towards the government between cabinet members and government backbenchers. As such, our method provides a novel way for testing theories of intra-party conflict (Giannetti and Benoit 2009), coalition governance (Strøm, Müller and Bergman 2008; Martin and Vanberg 2011), and the way government parties communicate their actions to their supporters and constituents (Martin and Vanberg 2008). When we move to a two-dimensional aggregation model, we find a second dimension dividing the opposition between establishment and anti-establishment parties.

In our second application, we compare the estimates from our model to existing scaling methods for US Senators based on roll-call votes (Poole and Rosenthal 1985; Clinton, Jackman and Rivers 2004) and campaign donations (Bonica 2014). While estimates from all three methods are positively and similarly correlated within as well as across parties, we find a much larger increase in speech polarization compared to (already high) roll-call

polarization. This increase in the extent to which Senators speak in increasingly different ways by party sheds some light on perceptions that polarization has become particularly pronounced in recent years, even though roll-call polarization has been high for much longer.

2. MEASURING PREFERENCE VARIATION FROM TEXT DATA

The fundamental difficulty in trying to estimate political positions from variation in the words used in political texts is that there are several more predictive sources of variation in word use. In rough descending order of importance, these are: (1) language, (2) style, (3) topic, and only then (4) position, preference or sentiment. Sources of variation higher on the list tend to overwhelm those lower on the list. If you have a text in German and a text in English, the variation in the frequency of different words is driven almost entirely by language. Once one has held language constant, style (or dialect) is very important: the words used in legal documents, in political speeches, and in tweets vary enormously. Similarly, variation in word use due to topic is substantial (this is why topic models work) and is comparable to differences due to dialect and style. The relative ordering of these is not important for present purposes, as the variation of interest here is that due to differences in the arguments being offered or the sentiments expressed towards a proposal, which we will refer to as *expressed preferences* or *stated positions*. This variation tends to be subtle in terms of relative word use, and therefore difficult to detect unless the more powerful sources of variation are held constant.²

Political scientists have followed one of two approaches when attempting to recover preferences from legislative speeches. One approach has been to confine the analysis to speeches

²Analogously, scaling models applied to roll-call voting data only recover plausible measure of legislator preferences when those preferences are the dominant influence on voting behavior. This is not always the case. In the UK House of Commons, almost all voting behavior is explained by whether an MP's party is in government (Spirling and McLean 2007). In the Brazilian Chamber of Deputies, voting behavior reflects a mixture of legislator ideology and membership in the governing coalition (Zucco and Lauderdale 2011).

on a single legislative act, such as a motion of confidence (Laver and Benoit 2002), contributions to the government’s annual budget debate (Herzog and Benoit 2015), or speeches on a particular bill (Schwarz, Traber and Benoit Forthcoming). While this approach (by assumption) holds topical variation constant, the resulting estimates are confined to the set of legislators who spoke and the topic on which they spoke. The opposite approach has been to combine many speeches over many legislative acts into a single document for each legislator (Giannetti and Laver 2005) or party (Proksch and Slapin 2010). Proksch and Slapin (2010), for example, scale speeches from the European Parliament by aggregating contributions across many topics by national parties. By pooling speeches across many topics, these authors have implicitly hoped that different parties would each discuss a similar mixture of topics, and therefore topical variation would cancel out. While this can work at the party-level, topical mixes vary enormously at the level of individual speakers, and in Section 4 we demonstrate the failure of this strategy for the Irish Dáil.

Our method combines these two approaches into a single estimation strategy. Similar to Laver and Benoit (2002), Herzog and Benoit (2015), and Schwarz, Traber and Benoit (Forthcoming), we use the structure of legislative debates to hold constant topic-driven word-use variation.³ If 15 speakers make statements about a single legislative proposal, the relative word counts across these texts are much more likely to vary as a function of preference variation than would be the case if one sampled 15 speeches from across all debates. Speakers may still not all talk about exactly the same aspects of that bill, some may wander off topic, or use metaphors that introduce nuisance word use variation. But using the debate structure is nonetheless a powerful form of conditioning: probably the most powerful form available in the legislative context.

³Laver and Benoit (2002) and Herzog and Benoit (2015) use the supervised scaling method “Wordscores” (Laver, Benoit and Garry 2003) to estimate positions, while we use an unsupervised scaling method, but our identification strategy shares the idea of comparing speeches only within the context of a given debate to hold topical variation constant.

Having estimated expressed positions for all speakers in a given debate, we must aggregate debate-specific dimensions that involve variable subsets of legislators into a smaller number of dimensions that include all legislators. This needs to be done in a way that is robust to the possibility that some of the debate-specific dimensions of word use variation will have no relationship with one another, either due to contamination from other sources of word use variation or due to idiosyncratic political features of the debates. In many legislatures, only a subset of 'debates' are really debates in the sense that they reveal political disagreement. For example, as Quinn, Monroe, Colaresi, Crespin and Radev (2010) document, a non-trivial fraction of speech in the US Senate consists of procedural statements or symbolic statements about notable constituents, the military, and sports. To extract the politically relevant variation, we scale the debate-specific scales, treating these debate-specific dimensions as noisy manifestations of one (or more) underlying general dimension(s).

Because this approach does not rely on word use variation in any single debate to estimate positions on a latent dimension of disagreement, it gains additional robustness against other sources of variation in word usage. All we need to discover this latent dimension is for that dimension to have general predictive power for word use variation across the set of observed debates. Crucially, the exact nature of that word use variation can be different in different debates. A word that implies a left position in one debate may imply a right position in another debate, or may imply no particular position at all. And if certain debates have speech variation that seems unrelated to other debates, the model will simply estimate that those debates fail to load strongly on the general dimension.

Like all measurement strategies, ours has no guarantees that the assumptions will hold, and so sanity checks and other forms of validation are still needed. But this is just as true in roll-call analysis, where estimated ideal points may variously reflect legislator preferences, constituency preferences, party inducements, government-opposition incentives, and other factors. Our methodological argument is fundamentally based on an empirical assumption:

that political disagreement is more clearly and consistently reflected in within-debate variation in word use than it is in across-debate variation in word use. We think this is a better assumption than those explicitly or implicitly used in previous studies, and so it is on this basis that we proceed to specify an estimation procedure.

3. SCALING TEXTS FROM SETS OF POLITICAL DEBATES

3.1. *Scaling Individual Debates*

Preference scaling of political texts projects highly multidimensional variation in word usage rates onto one (or more) continuous latent dimension(s). We begin by considering the unidimensional Poisson scaling model “Wordfish” (Slapin and Proksch 2008), as applied to a set of texts within a single political debate.

For all the following discussion, we index individuals $i \in 1, 2, \dots, N$, index debates $j \in 1, 2, \dots, M$, and index words $k \in 1, 2, \dots, K$.

$$w_{ijk} \sim \mathcal{P}(\mu_{ijk}) \tag{1}$$

$$\mu_{ijk} = \exp(\nu_{ij} + \lambda_{jk} + \kappa_{jk}\psi_{ij}) \tag{2}$$

That is, the frequency that legislator i will use word k in debate j depends on a general rate parameter ν_{ij} for individual i ’s word usage in debate j , word-debate usage parameters $\lambda_{jk}, \kappa_{jk}$ and the individual’s debate-specific position ψ_{ij} . The ν_{ij} parameters capture the baseline rate of word usage in a given speech, which is simply a function of the length of the speech. The λ_{jk} capture variation in the rate at which certain words are used. The κ_{jk} capture how word usage is correlated with the individual’s debate-specific position ψ_{ij} . This describes a standard text-scaling model, which could be applied to (1) all speeches given

in a legislative session, (2) the aggregated speeches of each legislator, or (3) applied to the speeches in a specific debate. Lowe (2008) shows that correspondence analysis provides an approximation to a Poisson ideal point model for text data. Lowe (2013) argues that in most applications it does not make much difference which model is used; however we have found that the Poisson scaling model is more robust when a single legislator gives a speech that is very different than his/her colleagues, which happens not infrequently in the legislatures we examine. Therefore, in the analysis that follows, we use the Poisson scaling model as our debate-level scaling model.⁴

3.2. *Aggregating Debate-Level Scales*

The Poisson scaling model (Wordfish) applied to each debate results in a debate-specific estimate, ψ_{ij} , of each speakers' relative position. In the second stage, we treat these estimates as data and use factor analysis to aggregate them into one (or more) general latent position θ_i for each legislator. Because not all legislators speak in each debate, the legislator-debate matrix containing all ψ_{ij} will have a large number of “missing observations”, which means the simplest factor analysis methods do not apply. We therefore adopt a fully Bayesian treatment of the linear factor model to recover θ_i , treating the ψ_{ij} as data and the missing ψ_{ij} as missing at random.

This assumption about the missing ψ_{ij} implies that the positions that legislators express are unrelated to their decisions to participate in a debate.⁵ Because of this assumption, the measures we recover should be interpreted as summaries of the positions actually taken

⁴Our identification and estimation strategies are slightly different than those used by Slapin and Proksch (2008) or by Lowe in the R package “austin”. We place normal priors with mean zero on all of the sets of the parameters in the model, with standard deviation 1 for the debate-specific positions ψ_{ij} and 5 for the other model parameters.

⁵Like the assumptions that make up the Wordfish model itself, this is an obviously wrong, but nonetheless useful assumption.

by legislators, relative to their peers, in the debates they participated in. These may be unrepresentative of their broader positions, if we could observe them in all debates. We discuss what is known about selection into legislative speech in several institutional contexts, and what that implies about extending our approach to model selection in Section 6.

The above assumptions imply a model for the debate-specific estimates ψ_{ij} that is linear as a function of a single latent dimension θ_i , with a normally distributed error.

$$\psi_{ij} \sim \mathcal{N}(\alpha_j + \beta_j \theta_i, \tau_i) \tag{3}$$

$$\theta_i \sim \mathcal{N}(0, 1) \tag{4}$$

$$\alpha_j, \beta_j \sim \mathcal{N}\left(0, \left(\frac{1}{2}\right)^2\right) \tag{5}$$

$$\tau_i \sim \mathcal{G}(1, 1) \tag{6}$$

This specification means that the primary dimension of word-usage variation in individual debates ψ can be more or less strongly associated with the aggregate latent dimension θ being estimated across all debates, with either positive or negative polarity for any particular debate. Essentially, this allows the model to select out those debate-specific dimensions that reflect a common dimension (large estimated values of β_j), while down-weighting the contribution of debates where the word-usage variation across individuals seems to be idiosyncratic ($\beta_j \approx 0$). The priors on θ_i and β_j allow the model to remain agnostic about the relative polarity of individual debate dimensions, while constraining the common latent dimension of interest to a standard normal scale. This 1D aggregation model can be extended to 2D by replacing $\alpha_j + \beta_j \theta_i$ with $\alpha_j + \beta_{1j} \theta_{1i} + \beta_{2j} \theta_{2i}$ in the above equations, adding corresponding priors for the additional parameters, and fixing the orientation of the latent space through appropriate constraints on parties or individual legislators (Rivers 2003).

A large number of quantities of interest can be calculated from the parameters of this model, some of which are summarized in Table 1. Most of these are functions of parameters of

Table 1: Quantities of interest that can be calculated from the debate-level and aggregate-level model parameters.

Quantity	Unit	Statistic	Description
Position on general scale	Speaker	θ_i	Speech position of legislator i on general scale (can be averaged over parties or other legislator characteristics).
Debate-specific position	Speaker	$\psi_{ij} \cdot \beta_j$	Speech position of legislator i on debate j (calibrated to the general scale).
Debate loading	Set of Debates	$\sqrt{\frac{\sum_j n_j \cdot \beta_j^2}{\sum_j n_j}}$	Strength of association of debate-scales with general scale across debates (root mean square, weighted by number of speeches n_j in each debate j)
Word loading	Set of Debates	$\frac{\sum_j n_{kj} \cdot \kappa_k \beta_j}{\sum_j n_{kj}}$	Association of word with general scale across debates (mean, weighted by frequency of word appearance n_{jk} in each debate).

the second-level model; however the debate-level parameter estimates can also be revealing, particularly when used in combination with the second-level parameters. For example, we can leverage the fact that a given word can have different political alignments in different debates to track how word use varies over time or as a function of some other feature of debates (see Section 5.3).

3.3. Implementation

In this paper, we present results based on estimating the Wordfish model for each debate, and then using those estimates as data for the second-stage aggregation model. The central benefit of breaking the estimation problem into two stages is computation speed, enabling us to quickly estimate the model on the word frequency matrices for each of the hundreds or thousands of debates that occur in a legislative term. For example, we are able to estimate

the model on recent sittings of the Irish Dáil, with about 1,000 debates, 10,000 speeches and 40,000 unique words, in a few minutes. The Wordfish first-stage model is estimated using an EM estimation procedure with Newtonian optimization steps based on the derived gradient and hessian of the log-posterior. This is implemented in C++ to speed estimation (Eddelbuettel and François 2011), and is 20-40 times faster than the fastest previous implementation in the R package “austin” (Lowe 2015). The second-stage factor analysis can be similarly estimated using an EM algorithm taking the first stage estimates as data, however in this paper we present results based on full Bayesian posteriors for the second-stage model estimated using JAGS (Plummer 2014). Our implementation of the Wordfish model is now available in the text analysis package “quanteda” (Benoit, Nulty, Barber, Watanabe and Lauderdale 2016), and the two-stage implementation of the Wordshoal model will be upon publication.

An alternative estimation strategy would be to estimate a fully hierarchical model in which the Wordfish parameters across debates are modeled as random coefficients. This approach would impose significant computational difficulties with very little estimation efficiency gain. While we further discuss the costs and benefits of combining the two estimation stages into a single model in Section 6, here we simply note that given the model we are fitting, the two-stage estimation approach will yield approximately the same point and uncertainty estimates as the hierarchical approach. The reason for this is that the Wordfish likelihood leads to very precise point estimates for the debate-specific position (document) parameters (see Section 4.3, as well as Lowe and Benoit 2011). This means that, even under the hierarchical approach, the debate-level positions are effectively data because their uncertainty is very small compared to the variation in the relative debate-specific positions for a given legislator across debates.

For the same reason, the two-stage approach also does not meaningfully understate estimation uncertainty versus the hierarchical model because nearly all of the uncertainty is in

the second-stage, not the debate-level scalings. Confidence in the estimated positions θ_i on the general scale(s) increases with the number of debates, with the extent to which sets of speakers overlap in different debates, and with the extent to which legislators are consistent in the positions they express in their speeches across debates. This is as it should be: these features of the data are the most meaningful ones if one is trying to assess whether a speaker is *generally* to the right or left of another speaker across a set of debates on heterogeneous topics.

3.4. Data

The substantive meaning of the dimension that our method recovers will depend on political context because the structure of legislators' preferences and their motivations to speak vary by political context. To illustrate this, we examine speech data from two very different institutional contexts: the Irish Dáil as an example of a multi-party parliamentary system with strong voting unity (Hansen 2009), and the US Senate as an example of a two-party system with weaker voting unity.⁶ The US Senate example also enables comparisons to spatial measures based on roll-call votes (Poole and Rosenthal 1985; Clinton, Jackman and Rivers 2004) and campaign donations (Bonica 2014).

The Irish data includes two complete legislative sessions, the 29th Dáil (2002–2007) and the 30th Dáil (2007–2011). Data for the US Senate includes all speeches from the 104th to the 113th Senate, which covers almost 20 years of legislative debates (January 1995–November 2014). We collected all speeches from existing databases of legislative debates or from official parliamentary records (see the Supplemental Appendix for further details). Before we scaled speeches and debates, we removed contributions from the person officially presiding over the chamber. In Ireland, this is either the Ceann Comhairle (speaker) or Leas-Cheann Comhairle

⁶Replication materials are available online as Lauderdale and Herzog (2016).

(deputy speaker). In the US Senate, we removed speeches from the Presiding Officer. We further removed procedural debates, such as the discussion of the meeting agenda, prayers, tributes, elections of the speaker, points of order, and any other discussions concerning the rules of parliamentary procedure. Finally, we removed punctuation, numbers, stop words and reduced words to their stem.

A key step in organizing the data was to identify speeches that belong to the same debate. We defined a debate as a set of speeches with the same title (as reported in the official parliamentary records) and that were held on the same day and included at least five speakers. Of course legislative debate on a single question can sometimes span multiple days or even weeks. However, even setting aside the relative difficulty of operationalizing this kind of broader definition, we nevertheless think it is preferable to limit the definition of a debate to a single day because the content and context of a debate can change from one day to the next. Within each debate, we combined all contributions of a legislator into a single composite speech, excluding contributions with less than 50 words because they are usually interruptions. For further details on the numbers of debates, speakers, speeches, and unique words, see the Supplemental Appendix.

4. IRISH DÁIL

In this section, we use legislative debates from the 29th and 30th Irish Dáil (Ireland's lower house) as an example of a multiparty parliamentary system with strong party discipline to demonstrate the usefulness of our approach in estimating individual TD's (*Teachta Dála*, an Irish member of parliament) expressed preferences. We first demonstrate that our approach outperforms an alternative strategy for scaling speeches from a legislative session: applying Wordfish to speeches aggregated across all debates in the entire legislative session into a single text for each of the 165 members. We further demonstrate that the primary dimension we

recover with our method represents TDs' relative levels of support and opposition to the government rather than left-right ideological positions, with a second dimension distinguishing between the establishment and anti-establishment opposition parties. This result is hardly surprising, given the weakness of ideology in Irish politics and the fact that in a coalition system like Ireland the fate of the government depends on acting unified. Nevertheless, there is substantial and meaningful *intra-party* variation along the government-opposition dimension. We illustrate this finding with an analysis of preference divergence between cabinet ministers and government backbenchers, and discuss opportunities for future research to use our estimates to study the tensions and conflicts that parties and coalition members face in policy-making.

During both legislative sessions included in our analysis, a coalition government led by Fianna Fáil (FF)—the largest party at that time—was in office. During the 29th Dáil, it was joined by the Progressive Democrats (PD), a small center-right/liberal party that formed in 1985 and dissolved in 2009, with its remaining members joining FF. The 30th Dáil added the Green Party to the coalition. The largest opposition party in both parliaments was Fine Gael (FG), the second largest party after FF at that time. Both FF and FG are centrist parties with similar policy positions that have historically been divided over Ireland's relationship with the United Kingdom (Benoit and Laver 2006; Weeks 2010). The other main opposition party was the Labour Party (LAB), a social-democratic party that has frequently formed coalitions with FG. The remaining opposition parties included Sinn Féin (SF), an anti-establishment party with the primary goal to unify Ireland, and the Socialist Party that was represented by a single TD in the 29th Dáil.

4.1. *Party Locations on the Primary and Secondary Dimension*

What are the primary factors that explain what positions legislators take in their speeches? In the absence of alternative preference estimates for Irish TDs, we first aggregate the legislator-specific estimates by parties and compare mean party positions to two benchmarks: whether the parties are in the governing coalition, and the left-right location of the parties as estimated from expert surveys (Benoit and Laver 2006).

The top row in Figure 1 shows mean party positions estimated from our approach against these two benchmarks. Based on these results, it appears that in the Irish data our approach is primarily recovering government versus opposition conflict, rather than left-right ideology. There are two ways to see this. First, while the largest parties FF and FG are generally viewed to be ideologically moderate in left-right terms, we estimate them at or near the extremes of our dimensions. Note in particular the fact that the Labour Party is estimated to be more centrist than FG, which only makes sense if we think of this as government-opposition. Second, when the Green Party joins the coalition in the 30th Dáil, it moves from having a similar average position to FG to having nearly the same position as FF.

In contrast, Wordfish estimates do not seem to consistently reflect the coalition structure of the Dáil, as is evident from the two scatterplots in the bottom row in Figure 1. The Green Party has a similar estimated position to FF, both when they are in coalition and when they are not. The Progressive Democrats are at one extreme of the dimension in the 29th Dáil and the other in the 30th, despite no change in coalition status. Neither do these estimates seem to reflect the ideological cleavages of the Dáil as assessed by expert surveys. In particular, experts do not place the Labour Party between FG and FF, but Wordfish does in both the 29th and 30th Dáil. In general, the association between the party locations from Wordfish and from the expert surveys are very weak.

When we extend the Wordshoal debate score aggregation model to 2D, we are able to

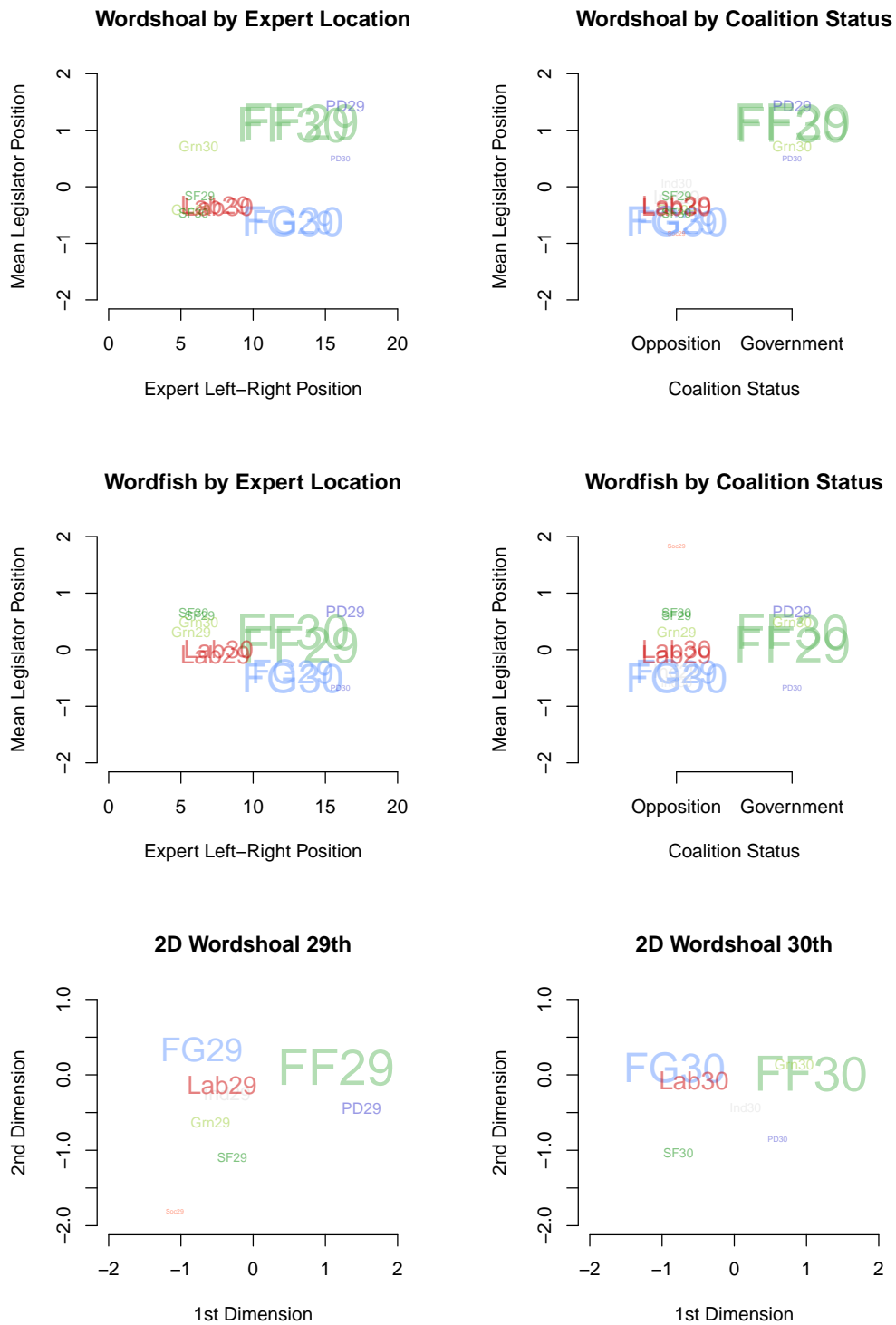


Figure 1: The top row shows the association between party average 1D Wordshoal scores and expert assessed left-right position (left) and coalition status (right). The middle two rows show the corresponding relationships for Wordfish scores. The final two rows show party average 2D Wordshoal scores for the 29th and 30th Dáil.

recover a more nuanced map of the positions of the Irish parties in these two Dáils. In order to orient the 2D space, we adopt a party-level normal prior that the average TDs from FF and FG are at 1 and -1 in the first dimension respectively, and both at 0 in the second dimension. In the final two panels of Figure 1 we show estimates of the average 2D party positions in the 29th and 30th Dáil. We see that the second dimension distinguishes between the establishment and anti-establishment opposition parties, with FG at the former end of the second dimension and SF at the latter. The single Socialist TD in the 29th, Joe Higgins, is even further out on this dimension while the Green Party is the next most anti-establishment after SF. In the 30th, when the Green Party joins a government for the first time in its history, it not only moves towards FF on the government-opposition first dimension, but also on this establishment dimension: it is difficult to maintain anti-establishment rhetoric from within a governing coalition.

4.2. *Legislator-Specific Positions*

When we look at the 1D estimates for individual TDs, rather than the party means, we can see the association between our estimates and coalition status even more clearly. Figure 2 shows the relationship between the estimated legislator positions and the coalitions under both Wordfish and our estimates in the 30th Dáil (the very similar plots for the 29th are included in the Supplemental Appendix). In the 30th Dáil, the (Pearson) correlation between being in the coalition government and Wordshoal score is 0.94, versus a correlation of 0.31 with Wordfish.

Figure 2 also shows that Wordfish gives implausibly narrow uncertainty intervals. The uncertainty estimates for TDs from Wordfish reflect the relative fit of different positions in predicting *words* across all texts, given the Poisson functional form and word-level indepen-

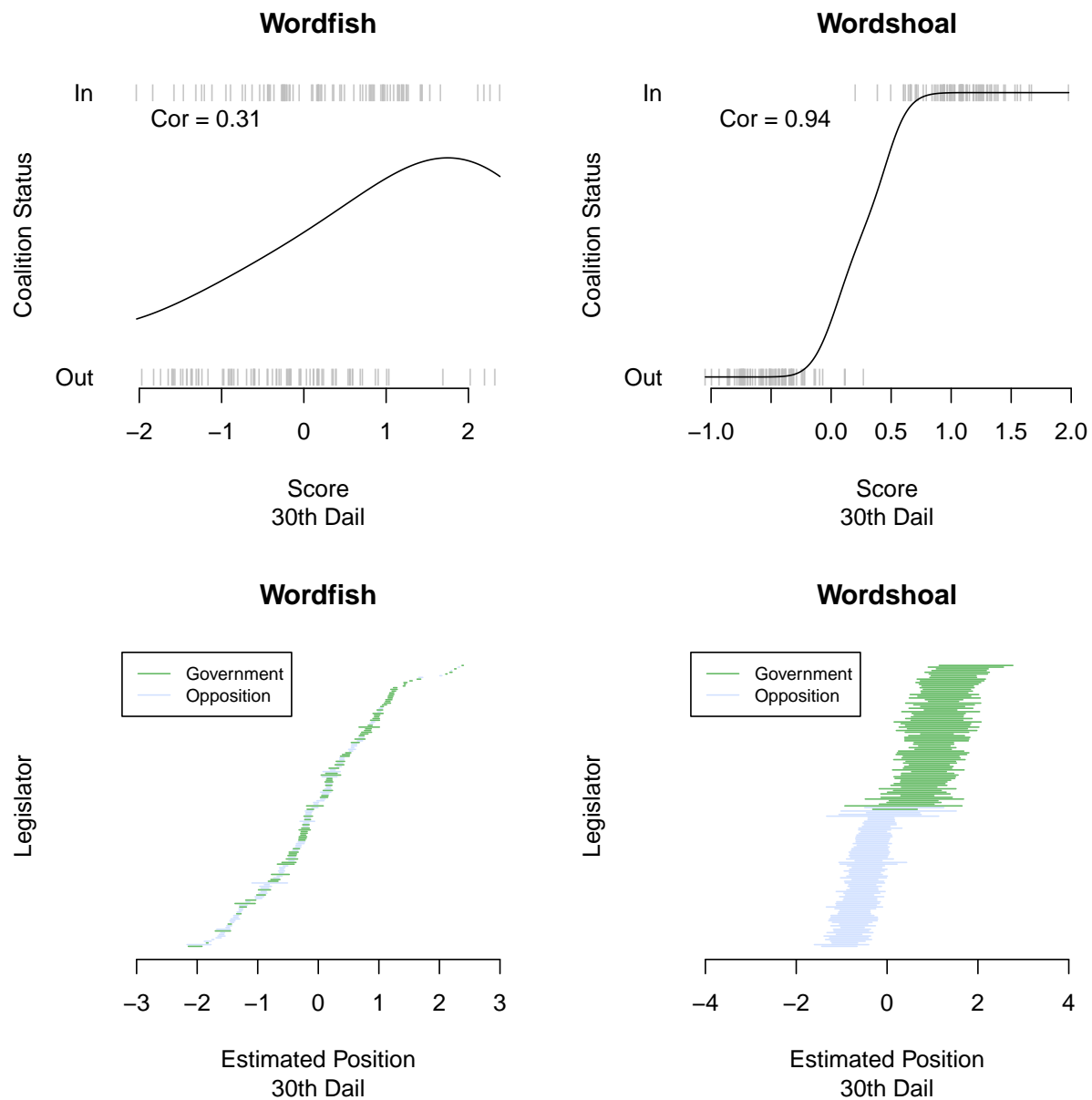


Figure 2: The association between the estimated positions of each legislator and their status as members of the coalition versus opposition, with correlation and local linear smooth, under Wordfish (left) and our approach (right), for the 30th Dáil. In the bottom row, we show the 95% intervals associated with the estimates for each legislator under Wordfish (left) and Wordshoal (right).

dence assumptions of that model.⁷ This uncertainty measure is substantively uninteresting, because resampling individual words does not capture a meaningful counterfactual sample of legislative speech. Any such counterfactual sample would involve resampling at the levels of speeches and debates, not words. The uncertainty intervals for the Wordshoal model reflect the number of debates each legislator speaks in, the extent of overlap between speakers in different debates, and the extent to which legislators are consistently ordered (by debate-level Wordfish) across the debates they speak in. This is the relevant kind of uncertainty for assessing if we have enough data to say that a particular legislator takes different positions from another legislator across a legislative session.

In sum, Wordshoal recovers point estimates that measure a meaningful quantity and provide uncertainty intervals that reflect realistic uncertainty about that quantity. Applied in the manner of previous studies, Wordfish neither recovers plausible measures of policy preferences nor plausible measures of government-opposition disagreement. Wordshoal very clearly recovers the government-opposition dimension of disagreement in Ireland. Recalling the identification strategy underlying Wordshoal, and thinking about the Irish context, this is not surprising. Our approach aims to recover the dimension that best explains within-debate variation in word use, across all debates. In a parliamentary system with strong party discipline like Ireland's, it is hardly surprising that the single factor that most consistently shapes speech behavior across every debate, is whether a legislator's party is in government or opposition.

⁷Wordfish, like LDA and other multinomial and poisson text models, is overconfident in its estimates for similar reasons to why Poisson regression coefficient estimates are overconfident when data are overdispersed.

4.3. *Intra-Party Variation in Government Support and Opposition*

Having validated the estimates as reflecting a government-opposition dimension in speech, we can begin to explore how TDs vary in position along this dimension. There is a voluminous body of research on multiparty governments, with recent work looking at the challenges that coalition partners and legislators face in day-to-day policy-making (Giannetti and Benoit 2009; Strøm, Müller and Bergman 2008; Martin and Vanberg 2008; Thies 2001; Martin and Vanberg 2011; Carroll and Cox 2012). One challenge for individual TDs is to balance the policy interests of their constituents against party demands (Kam 2009). This is particularly true in the Irish case, where the Single Transferable Vote (STV) electoral system gives TDs an incentive to cultivate a personal vote (Gallagher and Komito 2009; Marsh 2007). The intensity of this incentive will vary with electoral safety, constituency composition, and a members position within his or her party, among other things (Heitshusen, Young and Wood 2005).

Legislative speeches provide one opportunity for legislators to justify and explain their positions to supporters and party colleagues. Our legislator-specific estimates therefore provide a novel way to study what factors explain how legislators position themselves in support or opposition to the government. We here look at one potential factor that explains within-party variation in expressed positions: whether or not a legislator is a member of the cabinet. Bound by the doctrine of collective cabinet responsibility (Laver and Shepsle 1996; O'Malley and Martin 2010), cabinet members are required to publicly support decisions made by the cabinet even if they privately disagree. We hence expect ministers to more reliably defend the government position than government backbenchers. We can assess whether this is the case in our data by comparing the average locations of TDs inside and outside the cabinet.

Figure 3 shows the average Wordshoal positions for cabinet ministers, junior ministers,

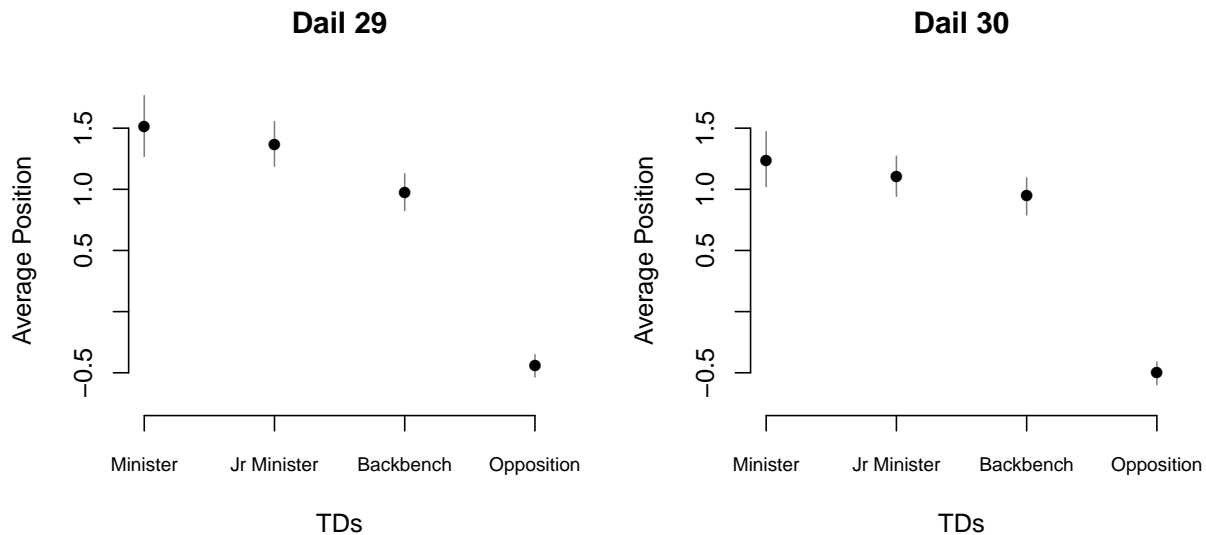


Figure 3: Mean positions of opposition speakers, government backbench TDs, junior ministers, and ministers for the 29th and 30th Dáil, with corresponding posterior intervals.

government backbench TDs, and opposition members.⁸ Consistent with the expectation of collective responsibility, we find that cabinet members are the most pro-government speakers. In the 29th Dáil, the average cabinet minister position is 1.52 versus the average position of backbench TDs at 0.98. In the 30th Dáil, the difference is slightly smaller with positions at 1.23 and 0.95, respectively. The posterior probability of these differences having these signs are both greater than 0.99. The average position of junior ministers is slightly, but less significantly, more moderate than the average minister position, indicating that junior ministers speak similarly to cabinet ministers, whether because of collective responsibility, career concerns, or some other factor.

The measured difference between ministers and backbench speakers is just one example for how our estimates can be used in secondary analysis to study within-party variation in expressed positions. The next step in analyzing these data would be to explore other

⁸If a member had multiple positions or transferred from one position to another during the legislative term, we counted the position with the longest duration.

factors that potentially explain when legislators strategically deviate from the government line or the position of their party, such as long term promotion prospects, promoting the particularistic interests of constituencies, or other factors that might motivate dissent.

This kind of legislator-specific estimate on a government-opposition dimension can be used to inform research on coalition governance and political communication. A key challenge for coalition parties is the need to compromise on policies while maintaining support from rank-and-file members, activists, and interest groups. As Martin and Vanberg (2008, 507) argue, "participation in coalition has the potential to undermine a party's carefully established profile and to erode support among constituents with a particular concern for the party's traditional goals." Legislative debates allow government members to justify and explain their positions on controversial policy decisions that potentially damage their reputation among core supporters. Martin and Vanberg (2008) offer the first empirical test of this type of political communication by looking at the *length* of legislative debates as a proxy for position-taking of government members. Our estimates, which are based on the content of legislative debates, would enable further assessment of the degree to which coalition members spoke consistently in favor of a bill in parliament. Such analysis is further enabled by another quantity of interest that can be calculated from our approach and that we illustrate in the next section: the strength of association of each debate with the general scale, which is a measure of the debate-specific degree of polarization on the primary speech dimension.

4.4. *Identifying High and Low Polarizing Debates*

The second-stage in the Wordshoal algorithm uses a Bayesian factor analysis to recover the primary dimension of word-usage variation from the debate-specific positions estimated in the first stage. This factor analysis estimates β_j , which is the strength of association of each debate with the general scale. We can use these estimates to answer the question: During

Table 2: The five debates with the highest and lowest loadings on the government versus opposition dimension, as measured by the absolute value of β_j ranging from 0 to 1.

<i>High government-opposition polarization</i>	Abs. β_j
Social Welfare and Pensions (No. 2) Bill 2009 (Second Stage)	0.942
Early Childhood Care and Education (Motion)	0.887
Private Members' Business – Vaccination Programme (Motion)	0.824
Capitation Grants (Motion)	0.819
Confidence in Government (Motion)	0.814
<i>Low government-opposition polarization</i>	
Cancer Services Reports (Motion)	0.003
Finance (No. 2) Bill 2007 (Committee and Remaining Stages)	0.002
Finance Bill 2011 (Report and Final Stages)	0.002
Private Members' Business – Mortgage Arrears (Motion)	0.002
Wildlife (Amendment) Bill 2010 (Committee and Remaining Stages)	0.001

which kinds of debates are TDs more polarized along government-opposition lines in what they say?

Table 2 shows the titles of the five most and least polarizing debates from the 30th Dáil as indicated by the absolute value of β_j . The most polarizing debate is from the second reading of a bill, which is the most important legislative stage after which the principle of a bill is formally accepted or rejected. We also find high polarization between government and opposition members during the 2009 confidence in the government motion, which prime minister Brian Cowen put forward to affirm his position as cabinet leader following poor results in local and European elections. Debates with low degrees of polarization are from committee stages and final readings at which point the outcome of a bill has usually been decided.

In the Supplemental Appendix, we explore variation along this government-opposition dimension more systematically. There we find an increase in government-opposition polarization with the onset of the economic and financial crisis in 2008, followed by a sharp

decrease in the observable government-opposition divide in 2010 before the collapse of the FF-Green coalition in the following year. This type of analysis illustrates how our method can be used to examine under what conditions and types of bills coalition partners are internally divided. Previous work in this area has relied on measures such as the length of debates (Martin and Vanberg 2008) or the duration of parliamentary scrutiny (Martin and Vanberg 2004) to assess party behavior on internally divisive issues. Our approach enables a much more direct assessment of the extent to which legislators are divided over an issue.

5. US SENATE

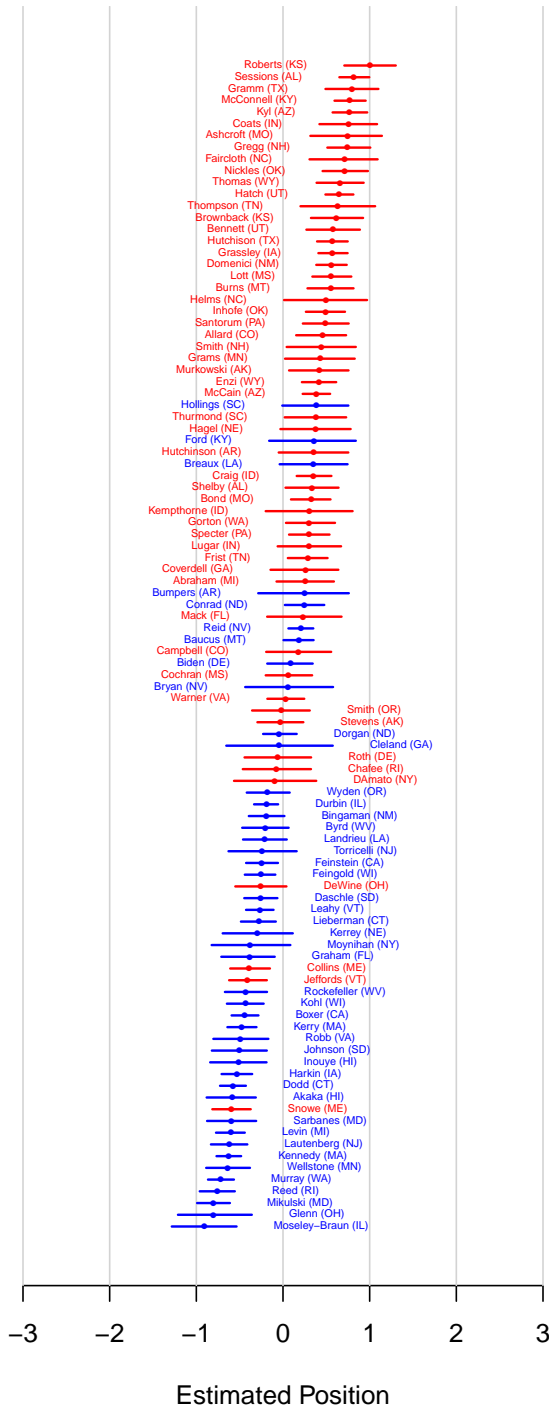
In our analysis of the US Senate, we use all debates from January 1995 to the end of October 2014, covering the 104th to the 113th Senate. We fit a model where senators are assumed to have constant positions. An analysis using the constant position assumption enables a comparison of the degree to which polarization over this period has occurred due to senator replacement versus the same senators having more partisan debates.⁹

Figure 4 shows the Wordshoal scores and 95% intervals of the senators serving in the 105th Senate (1997-1998) and the 112th Senate (2011-2012).¹⁰ The partisan polarization of senators due to replacement is visually apparent from the increased degree to which the scores correlate with party. In the 105th, Democratic senators Ford (KY), Hollings (SC), Breaux (LA), Conrad (ND), Bumpers (AR), Reid (NV), Baucus (MT), Biden (DE), Bryan (NV), and Dorgan (ND) spoke like Republicans. This list includes nearly all of the Democrats from the South as well as several from states like Montana, Nevada and North Dakota that typically voted Republican in Presidential elections and Democratic in Congressional elections in the

⁹This model cannot, however, identify whether these more partisan debates are occurring because these individuals' views have become more extreme or because they are more consistently debating on the issues that divide them.

¹⁰Similar plots for all the Senates from the 104th to the 113th are shown in the Supplemental Appendix.

Senate 105



Senate 112

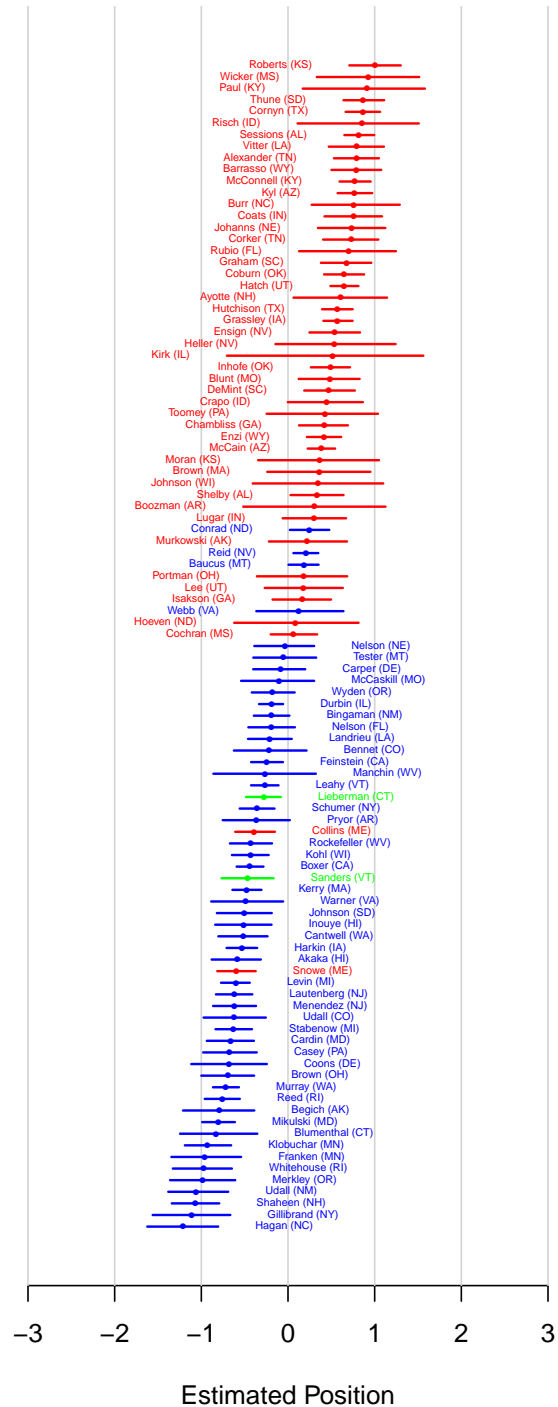


Figure 4: Wordshoal estimates for the 105th and 112th US Senates.

preceding decades. The Republicans interspersed among the Democrats on the left side of the estimated dimension—Snowe (ME), Jeffords (VT), Collins (ME), DeWine (OH), Roth (DE), D’Amato (NY), and Chafee (RI)—mostly come from the Northeast. In contrast, in the 112th, there is much cleaner separation between the parties: all five of the overlapping Senators are long-serving members of the chamber, three of whom have retired since the end of the 112th Senate.

5.1. *Speeches, Roll-Calls, and Donations*

How do our measures of US Senators’ relative positions compare to other scales of the same legislators based on different kinds of data? For roll-call votes, we fit a standard Bayesian IRT model to all votes over the same period as our speeches.¹¹ For campaign donations, we use career CFscores (Bonica 2014) for each senator.¹²

In essence, a comparison of these three measures is a comparison of roll-call behavior, speech behavior, and donor behavior, each modeled in terms of a single spatial dimension. If we calculate correlations between within-party variation in these three scores, we find that roll-call scores are correlated with speech scores at $\rho = 0.46$, while donor scores are correlated with roll-call scores at $\rho = 0.57$ and with speech scores at $\rho = 0.55$.¹³

One could make the argument that there is only a single meaningful latent political dimension, and that the three measures differ only because of measurement uncertainty.

¹¹Like the debate-score aggregation model, this is a heteroskedastic-by-legislator scaling model (Lauderdale 2010) and treats Senators’ ideal points as constant over time in order to isolate the effects of legislator replacement and avoid difficulties with inter-temporal identification.

¹²These career averages are for 1979-2012. We drop appointed Senators who never ran a campaign, for whom CFscores are not available and for whom roll call and speech scores are very imprecisely estimated.

¹³Under bootstrap resampling of senators, the differences in these correlations are plausibly attributable to unsystematic variation in legislators’ scores. The largest difference (between the 0.46 correlation of roll-calls and speech scores and the 0.57 correlation between roll-call and donor scores) has $p = 0.05$, ignoring the multiple comparisons. In the Supplemental Appendix we report linear regressions predicting standardized within-party variation in average donor positions using standardized within-party variation in roll-call scores and speech scores.

The relatively high correlations with donor behavior might simply indicate less measurement error in those scores than the other two.¹⁴ If this were the case, it would strongly indicate the value of having all three of these measures: forming a joint scale based on all three would then improve measurement of this “one true” latent dimension. In fact though, we think this is a relatively implausible account for the differences between these measures. Speech, roll-call and donor behavior are distinct political behaviors subject to distinct political forces, which becomes clearer when we examine the ways in which they differ.

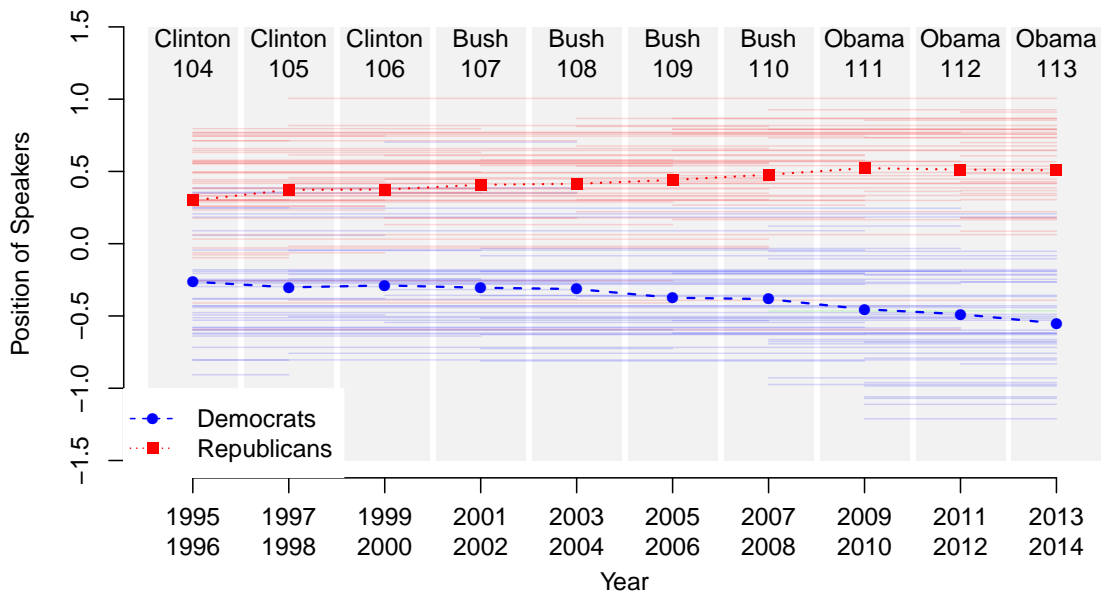
5.2. *Positions and Polarization in Speeches versus Roll Calls*

Figure 5 shows the increasing polarization due to replacement by showing senators’ Word-shoal speech positions (top panel) as well as the equivalent analysis for roll-call scores (bottom panel) from 1995 to 2014. A striking difference is that the gap between the average party positions in speeches has doubled, a far larger increase than for roll-call voting. New Republicans and Democrats have voted similarly on average to those same-party senators whom they have replaced; however, newly elected senators speak in more partisan ways than those they replace. Whereas Republican and Democratic senators used to substantially overlap in how they spoke on the floor, that overlap has mostly disappeared over the last two decades due to turnover.

Turnover is not the sole cause of increasing speech polarization. The parameter β_j describes how each debate relates to the general dimension. We can use this measure to capture how *strongly* the average debate tracks the primary dimension we recover. We do this by calculating the speech-weighted root mean square of β_j and comparing its value over time (see quantity “Polarization” in Table 1). The top panel of Figure 6 shows this quantity within 8 month periods running from January to August in the year after an election, from

¹⁴CFScores lack measures of uncertainty, so we cannot check this.

Speech Polarization of Senators



Roll-Call Polarization of Senators

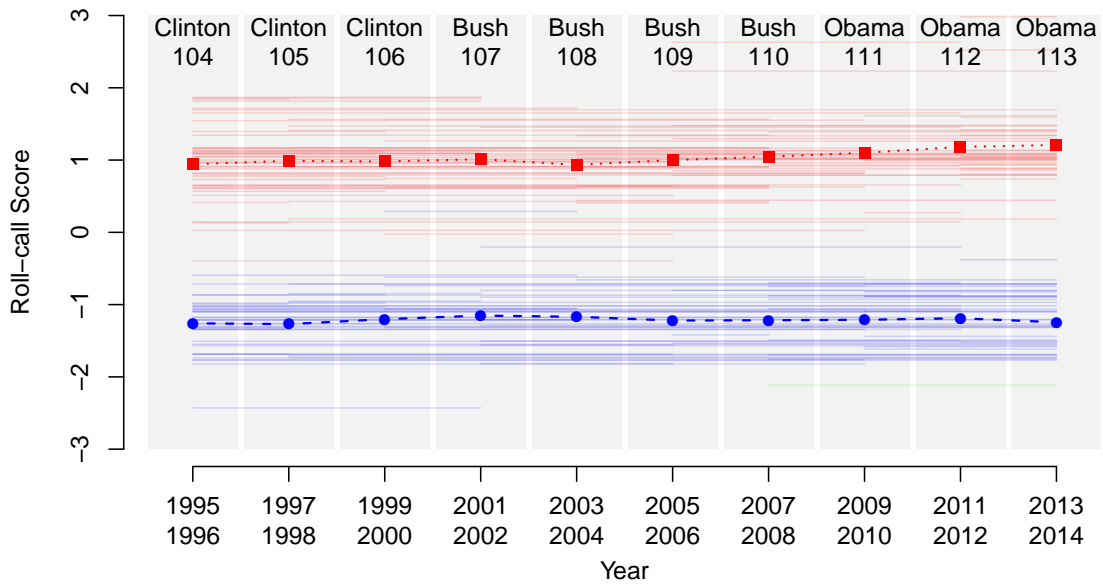
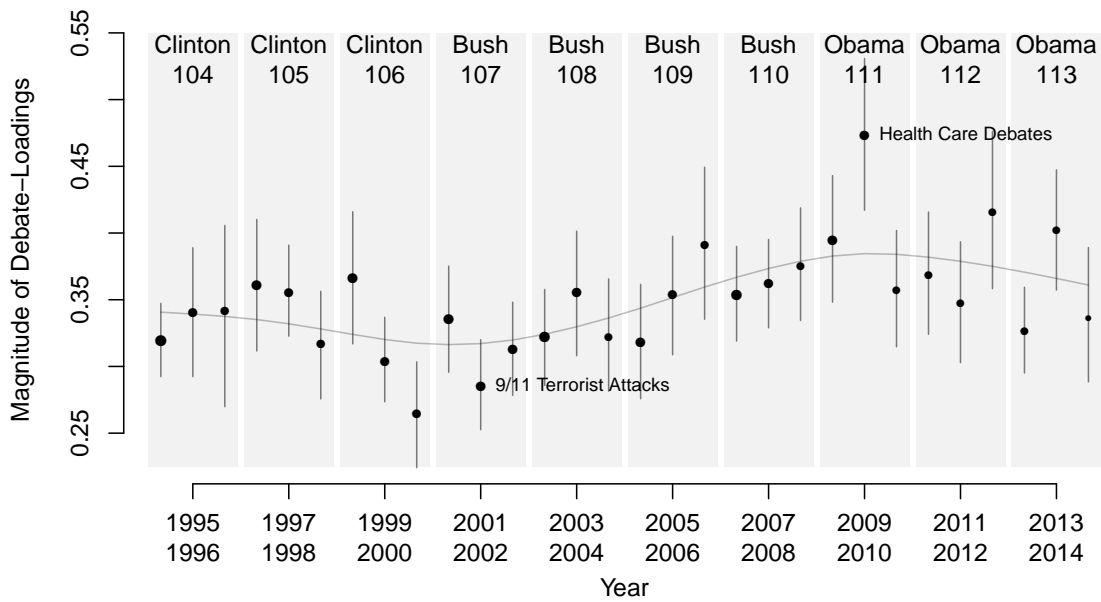


Figure 5: Average party positions in speeches (top panel) and in roll-call votes (bottom) from the 104th Senate (1995-1996) to the 113th Senate (2013-2014)

Polarization of Senate Speeches



Polarization of Senate Roll-Call Votes

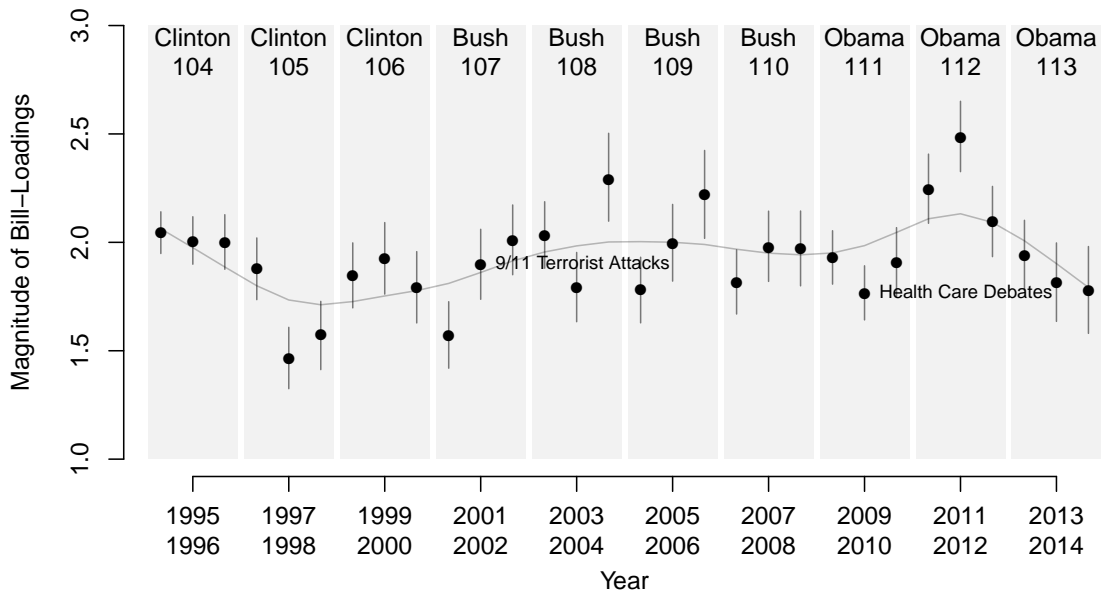


Figure 6: Average debate loadings (top panel) and average roll-call vote loadings (bottom) from the 104th Senate (1995-1996) to the 113th Senate (2013-2014), with spline smooth.

that September to April of the following year, and from May through December in the year of the next election. In general, polarization of debates, holding fixed the composition of the Senate, held steady during the Clinton administration, rose during the Bush administration, and has held steady at a higher level under Obama. As we saw earlier, increasing polarization due to senator turnover was occurring during all three periods.

Some 8-month periods deviate from this overall trend in the sense of being outliers from the spline regression fit depicted in the figure. The period from September 2001 to April 2002, which began with the terrorist attacks on September 11 and included the US invasion of Afghanistan, had speech polarization below the trend, reflecting a (brief) period of political unity after the attacks. The period with the highest polarization of debates, by far, is the period from September 2009 to April 2010 that included the Senate debates on health care legislation introduced by President Obama.

The bottom panel of the same figure shows the equivalent trajectory for roll-calls, which shows different patterns.¹⁵ The health care debates, which occupied so much Senate time in 2009-2010, fail to register because they involved relatively few roll-call votes, even as they occupied a great deal of political attention. The highest point in roll-call polarization instead comes from September 2011–April 2012, a period generating little major legislation, but occurring immediately after the debt-ceiling crisis of July-August 2011. There is some general upward trend over the period, but there is more substantial variation within Congresses.

5.3. *Evolving Patterns of Partisan Discourse*

Because our scores are based on word usage, we can examine how political language evolves in its usage over time. The first-stage scaling model estimates word-specific parameters, κ_{jk} ,

¹⁵The magnitudes of the debate loadings and bill loadings are not directly comparable, as the former are for a linear model and the latter a binary choice model.

that capture the association of a word with a legislator’s debate-specific position. Because we estimate these word-specific parameters for each debate separately, we can track whether particular words tend to signify being on the left or the right, and how that evolves across debates. Previous approaches that combine speeches into a single text assume that the association between words and positions are constant across topics and over time. Figure 7 shows how five word stems—health, constitut, tax, deficit, and preexist—load on the general dimension over time, averaging across debates weighted by the frequency that the word appears (see quantity “Word Loading” in Table 1).

These four words were all especially relevant to the discussion of the ACA in 2009-10; however their evolution also reflects their broader usage and the longer-term development of the health care issue in US politics. We see that throughout the period from 1995, the word “health” tended to be mentioned more by individuals on the left, across all debates. Interestingly, this fades after the passage of the ACA, perhaps reflecting the accomplishment of this long-standing Democratic policy goal. The stem “preexist” appears very suddenly as a highly left-leaning word in 2009-10 when Obama took office. The need to provide a mechanism for individuals with pre-existing conditions to secure health insurance was a central component of the Democratic argument for the ACA, and it was emphasized by Democrats both before and during the debates over the ACA itself. Once the legislation passed, the Democratic loading of the word stem faded, as Republicans began to acknowledge that they also supported enabling individuals with pre-existing conditions to have insurance, even as they continued to oppose the ACA.

The Republican objections to the ACA centrally involved arguments that the bill was unconstitutional, that it was actually increasing taxes, and that it would increase the deficit and debt. All of these words reflect broader themes of Republican argumentation across a range of issues, and so they have tended to lean to the right in debates from 1995 onwards, however all became particularly loaded language during the first two years of the Obama

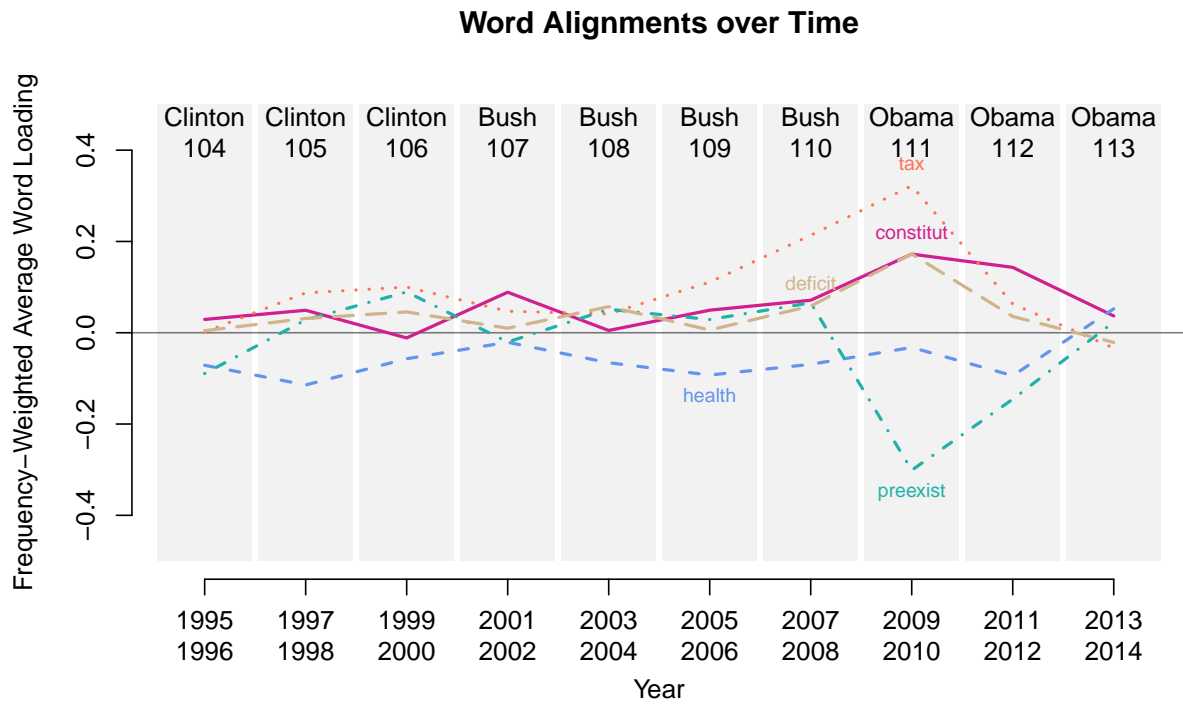


Figure 7: Average loading of the word stems health, constitut(ion), tax, deficit, and preexist(ing) across debates, weighted by frequency of appearance.

administration as the ACA was being debated and passed. Interestingly, in the most recent Congress, the 113th, occurring after Obama’s re-election, the political loading of all of these words has decayed. This reflects the fact that the Senate had essentially no debates about health care at all in that Congress, and indeed very few debates at all. With the House in Republican control, but passing very little legislation, Senate business was consumed primarily by other issues such as executive and judicial branch confirmations.

6. LIMITATIONS AND EXTENSIONS

Sparsity and Speaker Selection: Only a few legislators speak in a given debate: on average about 12 in the two legislatures we examine in this paper (see Table A1 in the Supplemental Appendix). As a result, the matrix of debate-specific Wordfish scores is sparse. The degree of sparsity will depend on the legislature, with larger legislatures (e.g., the US House, the UK House of Commons, the EU Parliament) having more severe sparsity than the smaller legislatures that we examined in this paper. Sparsity can make it difficult to measure the preferences of legislators who speak rarely, and it increases the importance of assumptions about the process by which legislators choose to speak in a given debate.

Political scientists have only recently started to examine what factors explain speaker selection in legislative debates (Proksch and Slapin 2012; Proksch and Slapin 2015; Herzog and Benoit 2015). Proksch and Slapin (2015) have shown that the degree to which party leaders exercise control over who speaks and what legislators say depends on the electoral system. In systems with strong personal vote incentives, such as Ireland’s STV system, legislators speak more freely because parties recognize the need for personal name recognition. In systems such as closed list PR, in contrast, party leaders tend to exercise greater control over the party message on the floor in order to protect the party label. Proksch and Slapin’s (2015) findings imply that the legislator-specific measures produced by our method will be

more accurate predictors of intra-party cohesion and dissident behavior in countries where the electoral systems provides strong personal vote incentives. In systems with weak personal vote incentives, our legislator scores potentially underestimate true levels of intra-party variation depending on the extent to which party leaders control speaking time across all debates included in the analysis. Proksch and Slapin (2015) show that party leaders are still more likely to allow dissenting speeches than dissenting votes, meaning that legislative speech holds valuable information that cannot be recovered from voting data alone.

In general, researchers need to be aware of the strategic incentives behind speech making when applying our method and using the quantities it produces in secondary analysis. Like any content analysis method, our approach can at best recover the “intended message” (Benoit, Laver and Mikhaylov 2009) of a speech and not a legislator’s “true” position, which is fundamentally unobservable. Of course strategic missingness can also be a problem in roll-call data (VanDoren 1990; Carrubba et al. 2006; Carrubba, Gabel and Hug 2008), but missingness is far rarer in those data than in speeches. Following recent work on jointly modeling missingness and voting in roll-calls (Rosas, Shomer and Haptonstahl 2014), one could use measurable variables that predict the decision to speak to jointly model the presence and position of speeches (Herzog and Benoit 2015). Extending our method in this direction would be a promising avenue for future research, but requires the collection of context-specific, theoretically motivated variables that explain the strategic selection of speakers.

Even without such a selection model, the approach followed in this paper still yields valuable summaries of behavior. What we recover is a summary of the speeches that were actually given. For example, the fact that Senator Snowe (R-Maine) is estimated far from her co-partisans, in the middle of the Democrats, does not mean she is “really” a Democrat, nor does it mean that our estimates are wrong. What it does mean is that when Snowe chose to speak in a Senate debate, she used similar language to Democrats who spoke in the same debates. Even if she was choosing those debates highly strategically, this is still an

important fact about the speeches she actually delivered.

Hierarchical Estimation: This paper argues that the political association of particular words depends on the debate in which those words were used. Conditioning on debate allows us to control for topic to a far greater extent than is otherwise possible. However, the two-stage procedure followed in this paper might take this logic too far. As we show in Section 5.3 some words are used similarly to denote position across many debates, and the presented approach does not take advantage of the estimation efficiencies that using this information could provide. One solution would be to estimate a full hierarchical model in which the Wordfish parameters across debates are modeled as random coefficients. Such an approach could in principle form a compromise between the assumption that the political associations of words in each debate are uninformative about other debates and the assumption that the political associations of words are identical across every debate.

As noted earlier, our main reason for not estimating the full hierarchical model is because of the additional computational burden and negligible consequences. Any efficiency gains from a full hierarchical model would stem from shrinking the Wordfish coefficients towards their population means, but almost no shrinkage could occur given the very precise estimates that result from the Wordfish first stage. To gain significant efficiency from a full hierarchical model, we would have to replace the Wordfish model with an alternative model which yielded substantially more uncertain estimates of document positions. The problem is not merely an issue of overdispersion yielding overly confident estimates from Wordfish’s Poisson likelihood function, rather the problem is that Wordfish, like nearly all text models, uses a bag-of-words assumption that each word is independently generated from some underlying distribution. This independence assumption, combined with the fact that speeches and other texts typically have a large number of words that are not actually independently generated, yields problematic likelihood-based uncertainty estimates in text analysis models more generally. There has been recent research exploring alternative simulation-based strate-

gies for estimating uncertainty in text scaling models (Lowe and Benoit 2011), however for the hierarchical model to work in our application, a likelihood function that yielded weaker inferences would be needed. There is important work to be done in this area, but it is well beyond the scope of this paper.

Multidimensionality and Dynamics In this paper, we have presented models that reduce variation in word use within debates to debate-specific dimensions, and that summarize variation in those debate-specific dimensions using one or two general dimensions. However just as we extended the second-level model to have a second dimension in the Irish example, it is also straightforward to fit multidimensional scaling models for words at the debate-level (Lowe 2013). Classic multidimensional models such as the one we presented for Ireland require identifying assumptions that have the effect of fixing the rotation and labeling of the dimensions. Because we only use the within-debate variation in word usage to estimate positions, one could use the across-debate variation in word usage to estimate which debates give us information about positions on which latent dimensions. Thus, given sufficiently rich data, we could measure how legislators’ positions vary by topic, and recover multidimensional preference estimates with topic labels, following recent work on roll-call analysis (Lauderdale and Clark 2014). As with disaggregating by topic, estimation of dynamic positions can also be achieved from a closely related model that does not change the lower-level model for the texts. To model dynamics, one could apply one of the existing techniques for modeling dynamics (Poole and Rosenthal 1997; Martin and Quinn 2002) to θ_i , allowing those parameters to vary over time.

7. CONCLUSION

It is appropriate to be skeptical about unsupervised estimators like ours that purport to turn word counts into estimates of “expressed preferences” or “stated positions”. This is

partly because of the black box process by which such models sometimes assume—rather than demonstrate—that preferences are a major source of variation in word usage. But it is also because the longer experience of scaling roll-call voting data in political science has given us a sense of the various ways that measurement models can fail to measure what we would like them to measure. Demonstrating that text scaling is a useful measurement strategy requires validation of the types provided in this paper.

The validation we have done suggests that—at least in the legislatures we examined—the primary dimension of *political disagreement*, rather than the primary dimension of *policy preferences*, is what we can measure using our speech scaling strategy. This makes sense given the way that our estimation procedure is constructed. The many debate-specific scales will reflect features of particular debates and the idiosyncratic contributions of particular legislators to those debates. However the scaling of these scales will select out the common dimension of variation across speakers that most consistently shapes word usage across all spoken debates. It is perhaps not surprising that this tends to be the government-opposition cleavage in the Westminster-style system of the Irish Dáil. In the US, where a different constitutional structure and a two party system create different incentives for legislative speech, we see patterns of speech behavior that, while different from roll-call and donor-based measures of senator positions, have a strong association with those measures both across and within parties. But there are also important differences which reflect the different processes that generate speeches, roll-calls, and donations. We find that party polarization of speeches is more responsive to political events than is roll-call behavior and that recent increases in rhetorical polarization are heavily the result of senatorial turnover.

Whereas both computer scientists and political scientists have made enormous progress in recent years at developing and refining tools for recovering topics from texts, progress on the problem of recovering continuous measures of disagreement has advanced more slowly and is more peculiar to political science. Recovering measures of relative disagreement is a more

difficult problem because of the nature of word usage, and there are fewer researchers working on this problem actively. As the preceding section indicates, there are several potential avenues for further methodological development. Legislative speech is both interesting as a proxy for more general conceptions of political position-taking as well as in its own right: it is a core component of the strategies adopted by legislators in response to the political and electoral environments that they face.

REFERENCES

- Benoit, Kenneth and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.
- Benoit, Kenneth, Michael Laver and Slava Mikhaylov. 2009. “Treating words as data with error: Uncertainty in text statements of policy positions.” *American Journal of Political Science* 53(2):495–513.
- Benoit, Kenneth, Paul Nulty, Pablo Barber, Kohei Watanabe and Benjamin Lauderdale. 2016. “quanteda: Quantitative Analysis of Textual Data.” Version 0.9.4.
URL: <https://cran.r-project.org/web/packages/quanteda/>
- Bonica, Adam. 2014. “Mapping the Ideological Marketplace.” *American Journal of Political Science* 58(2):367–386.
- Carroll, Royce and Gary W Cox. 2012. “Shadowing ministers: monitoring partners in coalition governments.” *Comparative Political Studies* 45(2):220–236.
- Carrubba, Clifford J., Matthew Gabel, Lacey Murrain, Ryan Clough, Elizabeth Montgomery and Rebecca Schambach. 2006. “Off the Record: Unrecorded Legislative Votes, Selection Bias and Roll-Call Vote Analysis.” *British Journal of Political Science* 36(4):691–704.
- Carrubba, Clifford J., Matthew Gabel and Simon Hug. 2008. “Legislative Voting Behavior, Seen and Unseen: A Theory of Roll-Call Vote Selection.” *Legislative Studies Quarterly* 33(4):543–572.
- Clinton, Joshua D., Simon Jackman and Douglas Rivers. 2004. “The Statistical Analysis of Roll Call Data.” *American Political Science Review* 98(2):355–370.
- Eddelbuettel, Dirk and Romain François. 2011. “Rcpp: Seamless R and C++ Integration.” *Journal of Statistical Software* 40(8):1–18.

- Gallagher, Michael and Lee Komito. 2009. The constituency role of Dáil deputies. In *Politics in the Republic of Ireland*, ed. John Coakley and Michael Gallagher. 5th ed. London: Routledge.
- Giannetti, Daniela and Kenneth Benoit, eds. 2009. *Intra-Party Politics and Coalition Governments*. Routledge.
- Giannetti, Daniela and Michael Laver. 2005. "Policy positions and jobs in the government." *European Journal of Political Research* 44(1):91–120.
- Hansen, Martin Ejnar. 2009. "The Positions of Irish Parliamentary Actors 1937–2006." *Irish Political Studies* 24(1):29–44.
- Heitshusen, Valerie, Garry Young and David M Wood. 2005. "Electoral Context and MP Constituency Focus in Australia, Canada, Ireland, New Zealand, and the United Kingdom." *American Journal of Political Science* 49(1):32–45.
- Herzog, Alexander and Kenneth Benoit. 2015. "The most unkindest cuts: speaker selection and expressed government dissent during economic crisis." *Journal of Politics* 77(4):1157–1175.
- Hix, Simon, Abdul Noury and Gérard Roland. 2005. "Power to the parties: cohesion and competition in the European Parliament, 1979–2001." *British Journal of Political Science* 35(02):209–234.
- Hug, Simon. 2010. "Selection Effects in Roll Call Votes." *British Journal of Political Science* 40(1):225–235.
- Kam, Christopher J. 2009. *Party Discipline and Parliamentary Politics*. Cambridge University Press.
- Lauderdale, Benjamin E. 2010. "Unpredictable Voters in Ideal Point Estimation." *Political Analysis* 18(2):151–171.
- Lauderdale, Benjamin E. and Alexander Herzog. 2016. "Replication Data for: Measuring Political Positions from Legislative Speech." Harvard Dataverse.
URL: <http://dx.doi.org/10.7910/DVN/RQMIV3>
- Lauderdale, Benjamin E. and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58(3):754–771.
- Laver, Michael and Kenneth A. Shepsle. 1996. *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies*. Cambridge: Cambridge University Press.
- Laver, Michael and Kenneth Benoit. 2002. "Locating TDs in policy spaces: the computational text analysis of Dáil speeches." *Irish Political Studies* 17(1):59–73.

- Laver, Michael, Kenneth Benoit and John Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97(2):311–331.
- Lowe, Will. 2013. “There’s (basically) only one way to do: Some unifying theory for text scaling models.” Paper presented at the American Political Science Association meeting September 2013, Chicago.
- Lowe, Will. 2015. “Austin: Do things with words.” Version 0.2.2.
URL: <http://github.org/conjugateprior/austin>
- Lowe, Will and Kenneth Benoit. 2011. “Estimating uncertainty in quantitative text analysis.” Paper presented at the Annual Conference of the Midwest Political Science Association, Chicago.
- Lowe, William. 2008. “Understanding Wordscores.” *Political Analysis* 16(4):356–371.
- Marsh, Michael. 2007. “Candidates or parties? Objects of electoral choice in Ireland.” *Party Politics* 13(4):500–527.
- Martin, Andrew D. and Kevin M. Quinn. 2002. “Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999.” *Political Analysis* 10:134–53.
- Martin, Lanny W and Georg Vanberg. 2004. “Policing the bargain: Coalition government and parliamentary scrutiny.” *American Journal of Political Science* 48(1):13–27.
- Martin, Lanny W and Georg Vanberg. 2008. “Coalition government and political communication.” *Political Research Quarterly* 61(3):502–516.
- Martin, Lanny W. and Georg Vanberg. 2011. *Parliaments and Coalitions: The Role of Legislative Institutions in Multiparty Governance*. Oxford: Oxford University Press.
- O’Malley, Eoin and Shane Martin. 2010. The government and the Taoiseach. In *Politics in the Republic of Ireland*, ed. John Coakley and Michael Gallagher. Routledge chapter 10, pp. 295–326.
- Plummer, Martyn. 2014. *rjags: Bayesian graphical models using MCMC*. R package version 3-14 ed. <http://CRAN.R-project.org/package=rjags>: .
- Poole, Keith T. and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29(2):357–384.
- Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2010. “Position Taking in European Parliament Speeches.” *British Journal of Political Science* 40(3):587–611.

- Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56(3):520–37.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate*. Cambridge: Cambridge University Press.
- Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1):209–228.
- Rivers, Douglas. 2003. "Identification of multidimensional spatial voting models." Manuscript, Stanford University.
- Rosas, Guillermo, Yael Shomer and Stephen R. Haptonstahl. 2014. "No News Is News: Nonignorable Nonresponse in Roll-Call Data Analysis." *American Journal of Political Science* .
- Schwarz, Daniel, Denise Traber and Kenneth Benoit. Forthcoming. "Estimating intra-party preferences: comparing speeches to votes." *Political Science Research and Methods* .
- Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.
- Spirling, Arthur and Iain McLean. 2007. "UK OC OK? Interpreting Optimal Classification Scores for the U.K. House of Commons." *Political Analysis* 15(1):85–96.
- Strøm, Kaare, Wolfgang C. Müller and Torbjörn Bergman, eds. 2008. *Cabinets and Coalition Bargaining: The Democratic Life Cycle in Western Europe*. Oxford: Oxford University Press.
- Thies, Michael F. 2001. "Keeping tabs on partners: the logic of delegation in coalition governments." *American Journal of Political Science* 45(3):580–598.
- VanDoren, Peter M. 1990. "Can We Learn the Causes Of Congressional Decisions From Roll-Call Data?" *Legislative Studies Quarterly* 15(3):311–340.
- Weeks, Liam. 2010. Parties and the Party System. In *Politics in the Republic of Ireland*, ed. John Coakley and Michael Gallagher. Routledge chapter 5, pp. 137–167.
- Zucco, Cesar Jr. and Benjamin E. Lauderdale. 2011. "Distinguishing Between Influences on Brazilian Legislative Behavior." *Legislative Studies Quarterly* 36(3):363–96.