

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

June 11, 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 (N 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.

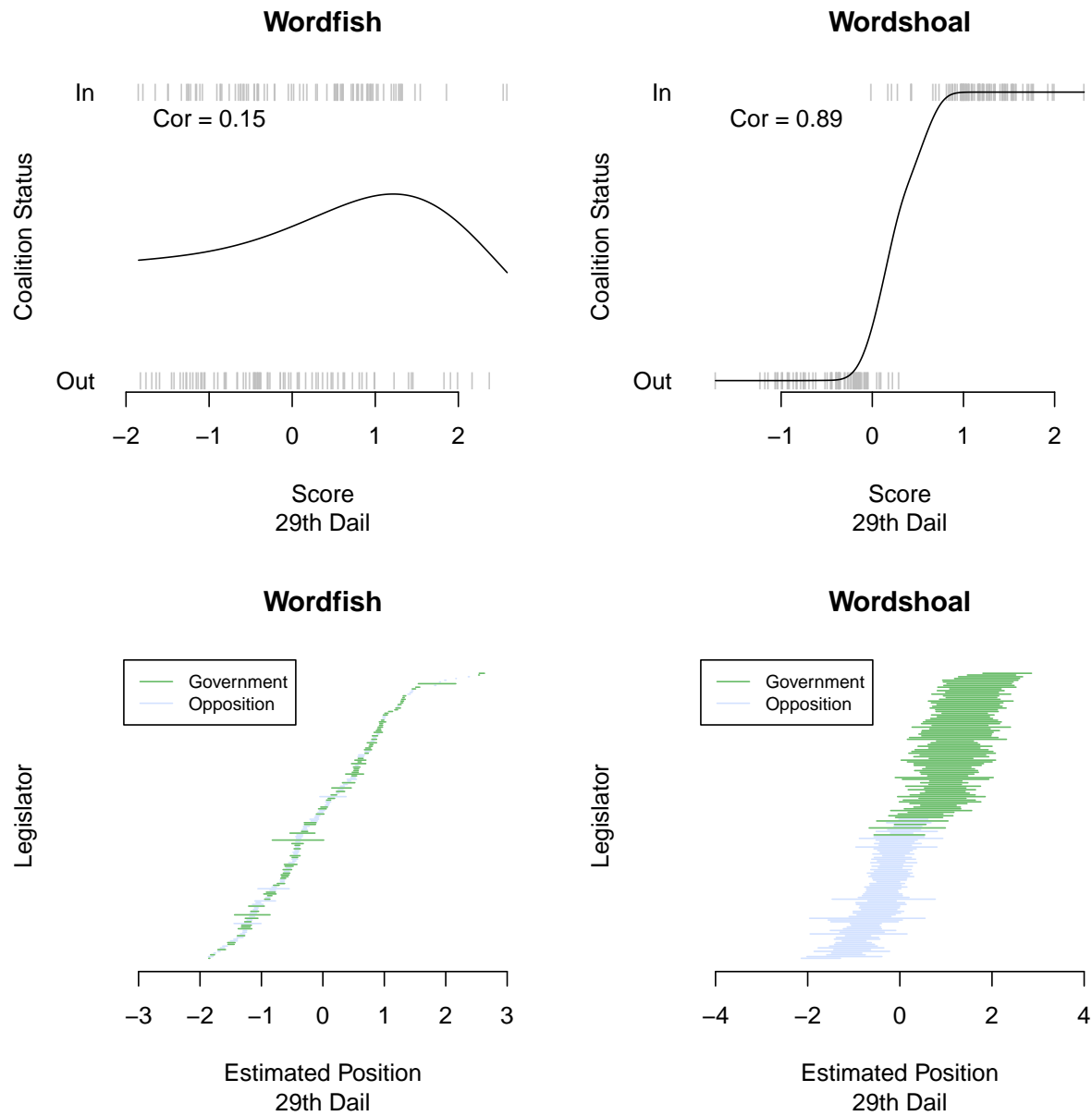


Figure A1: The association between the estimated positions of each legislator and their status as members of the coalition versus opposition, with correlation and local linear smooth, under Wordfish (left) and our approach (right), for the 29th Dáil. In the bottom row, we show the 95% intervals associated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

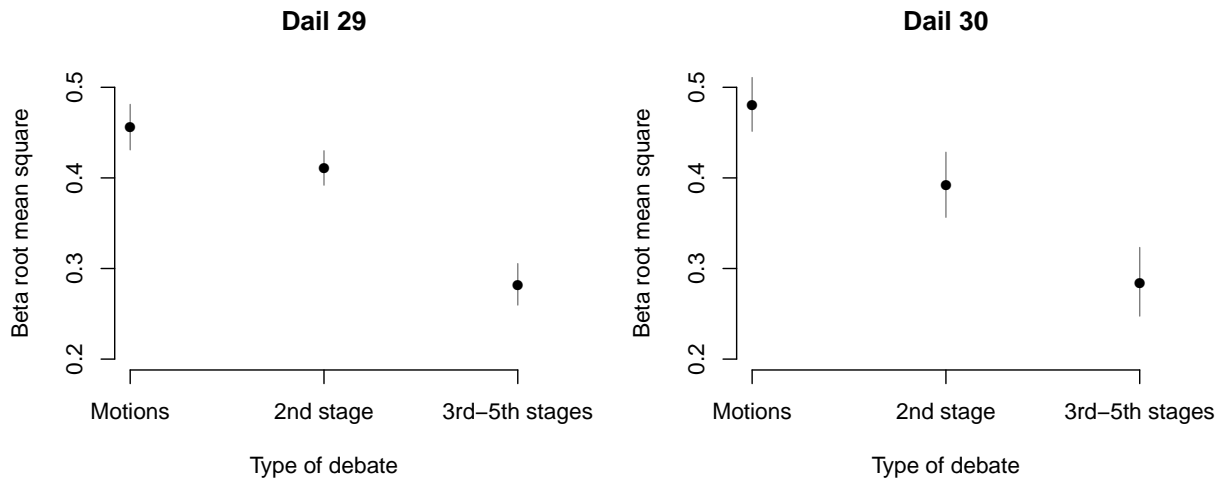


Figure A2: Weighted root mean square of β_j by type of debate for the 29th and 30th Dáil.

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

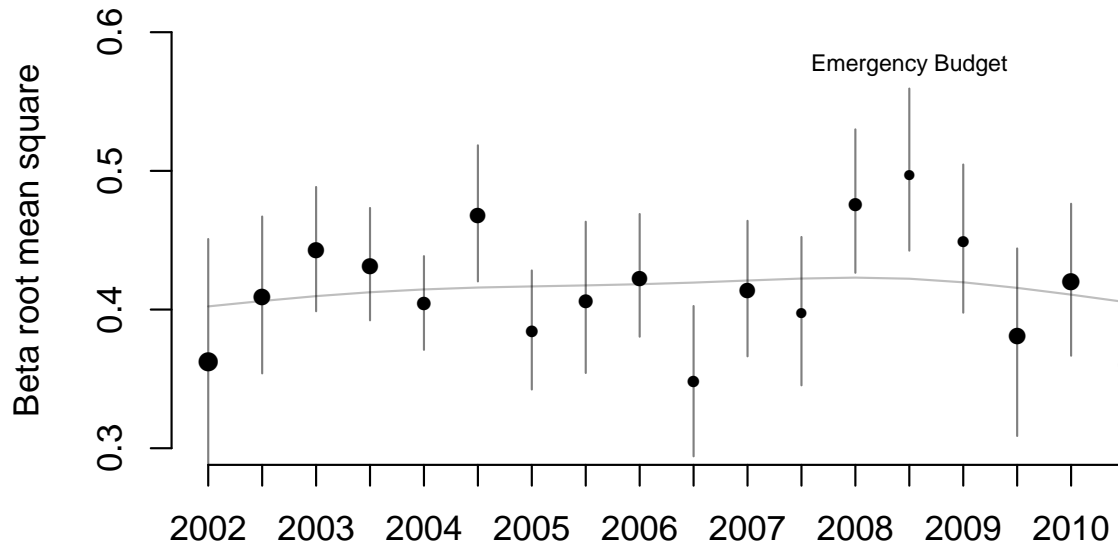


Figure A3: Weighted root mean square of β_j over six-months periods. The size of each point is proportional to the relative number of second-stage debates in each period.

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

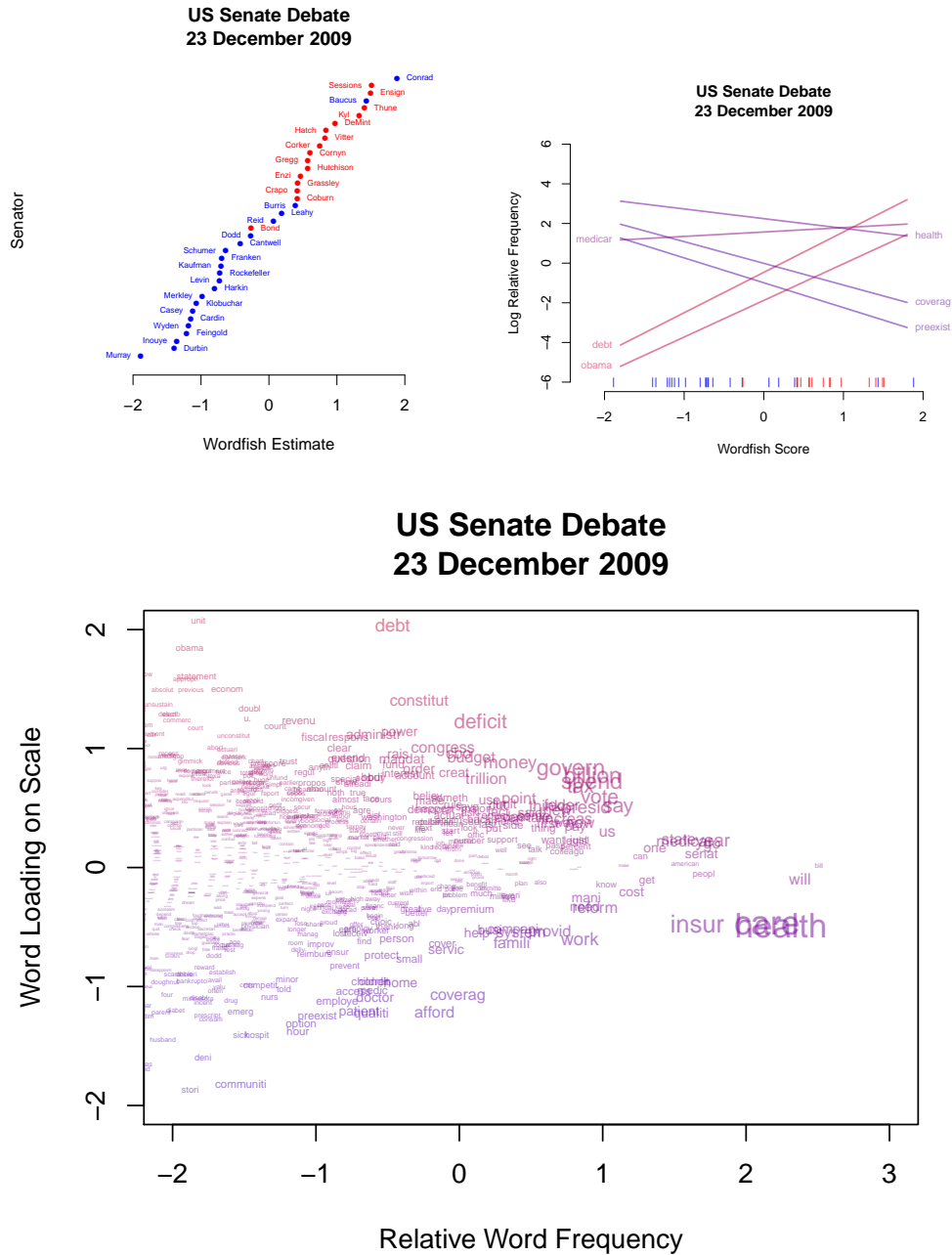


Figure A4: Wordfish estimates from the US Senate debate on 23 December 2009, preceding the vote on cloture on the Affordable Care Act. At top left, the Wordfish scores estimated for senators who spoke in the debate. At top right, the fitted log relative frequencies of several salient words in the debate, 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

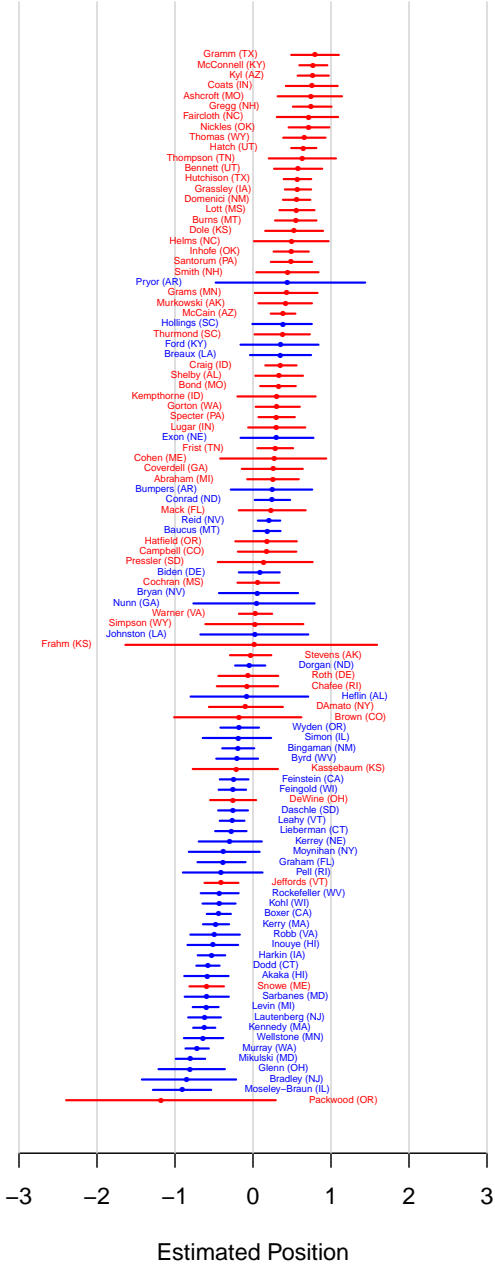
him at the right extreme of the dimension. Most Senators gave set speeches in this debate, rather than directly responding to one another. The fact that Baucus did not do this leads Wordfish to place him in a way that does not reflect his political position on the issue.

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

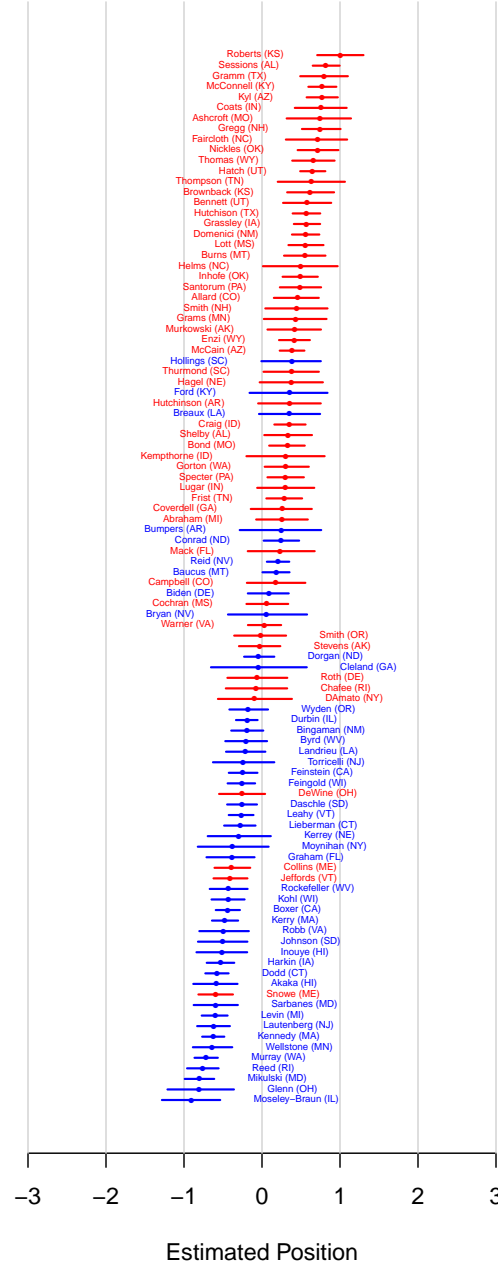
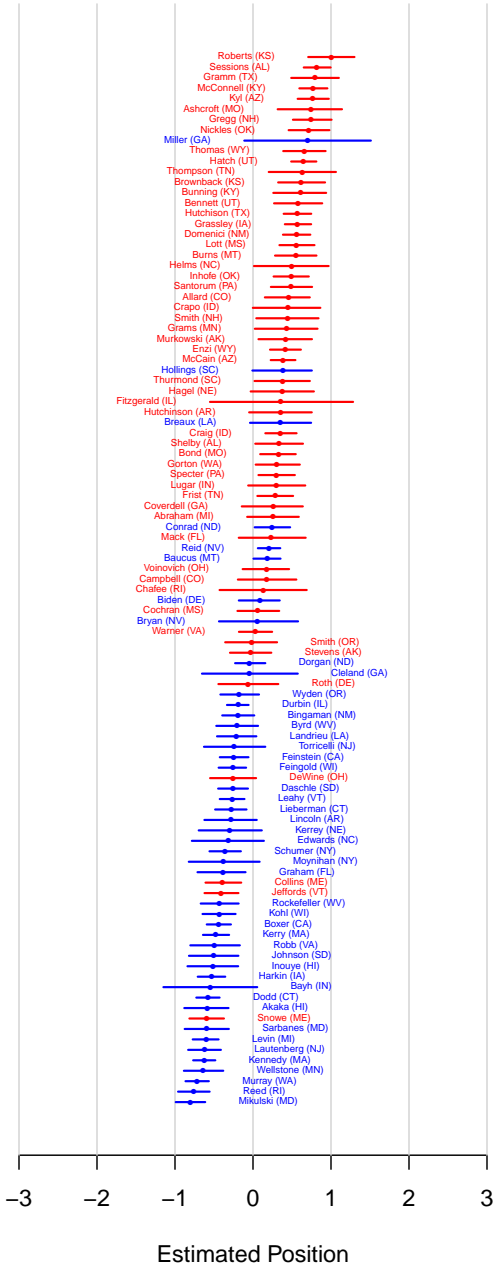


Figure A5: Wordshoal estimates for the 104th and 105th US Senates.

Senate 106



Senate 107

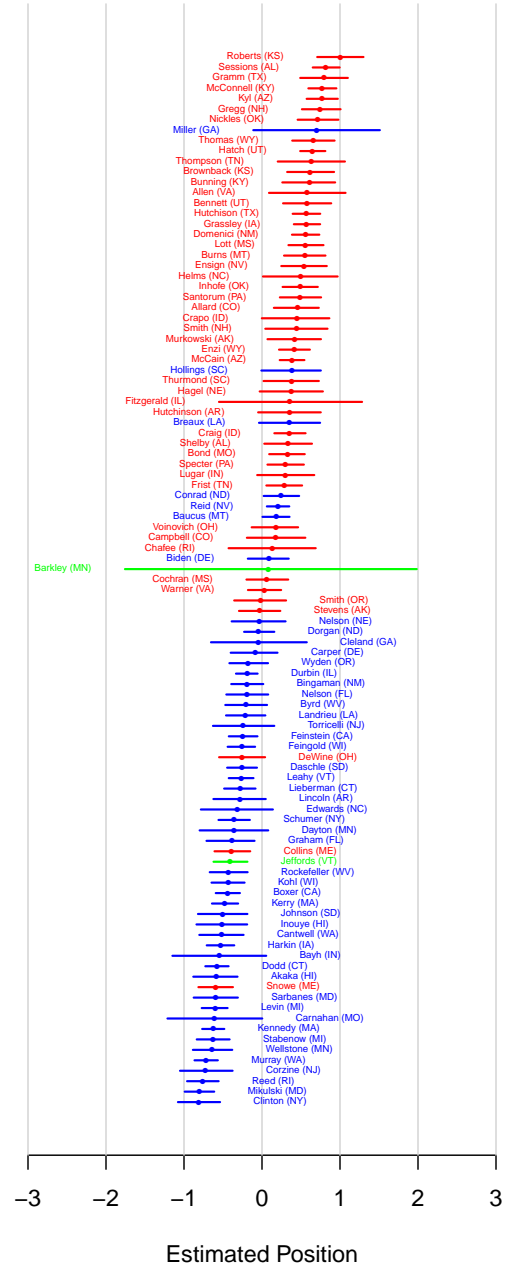
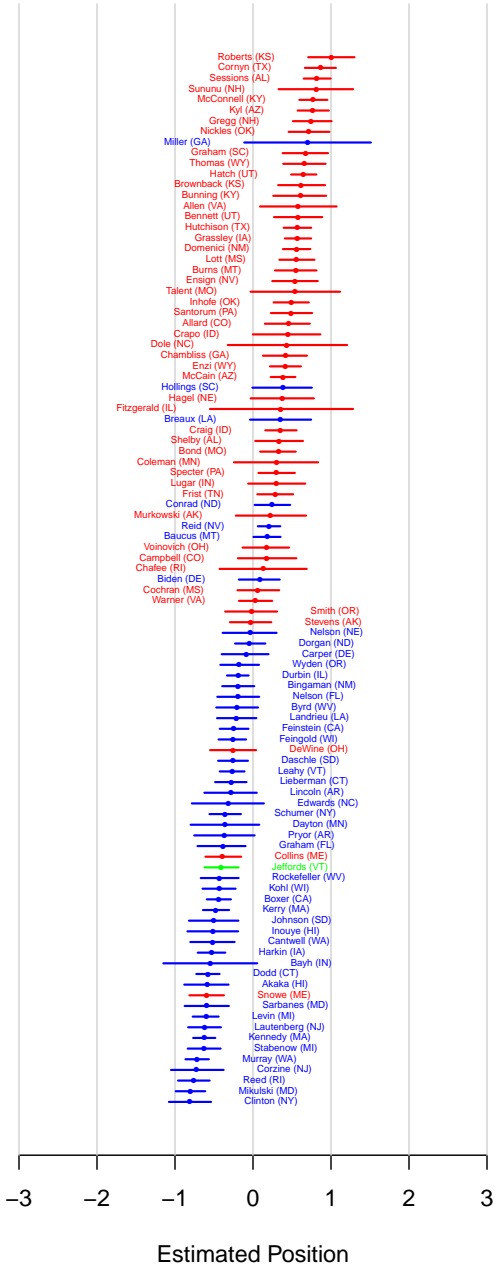


Figure A6: Wordsnoal estimates for the 106th and 107th US Senates.

Senate 108



Senate 109

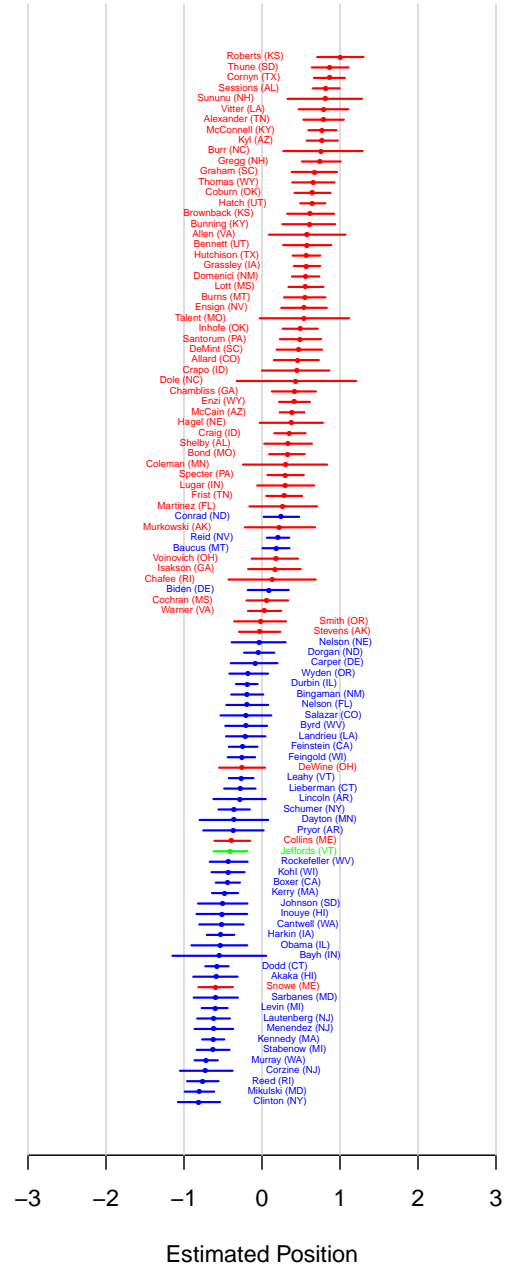
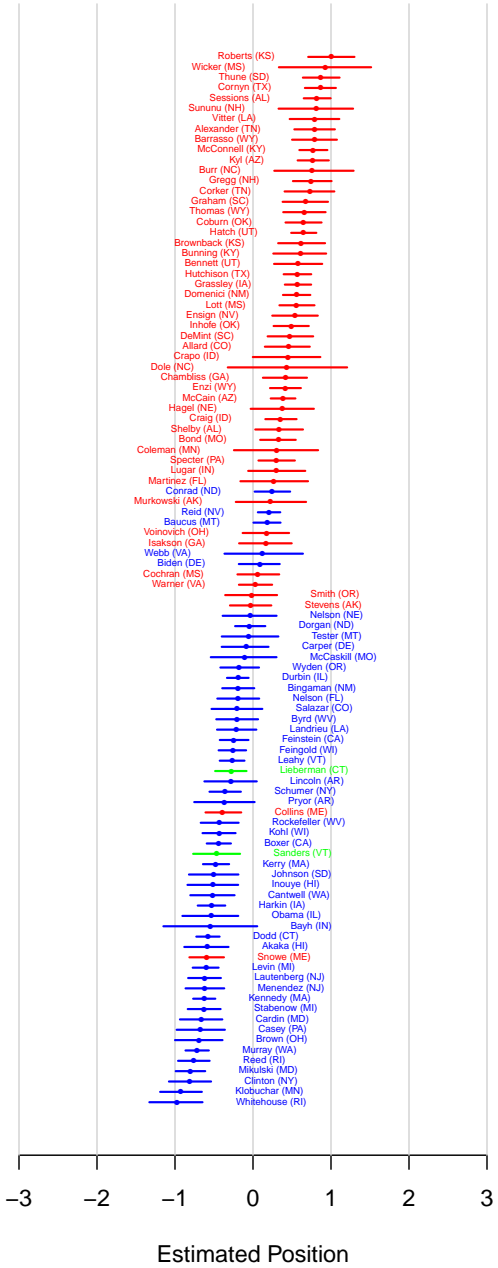


Figure A7: Wordshoal estimates for the 108th and 109th US Senates.

Senate 110



Senate 111

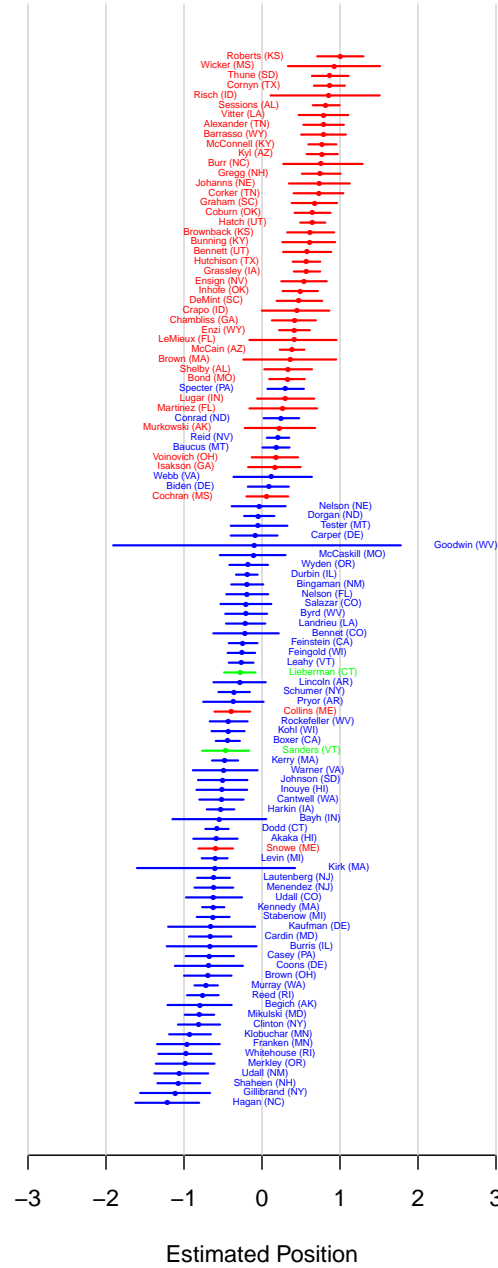
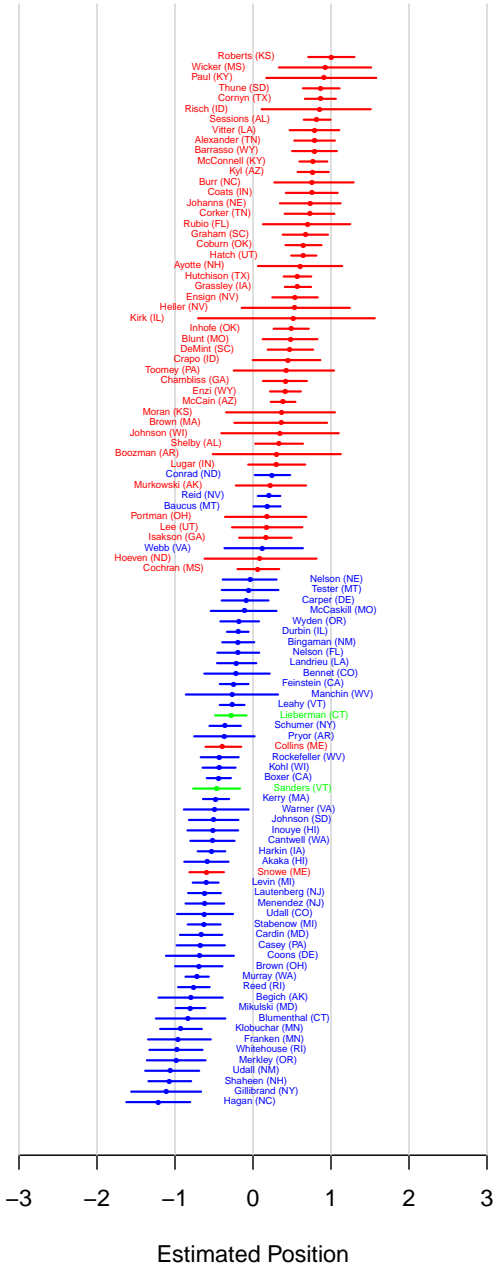


Figure A8: Wordshoal estimates for the 110th and 111th US Senates.

Senate 112



Senate 113

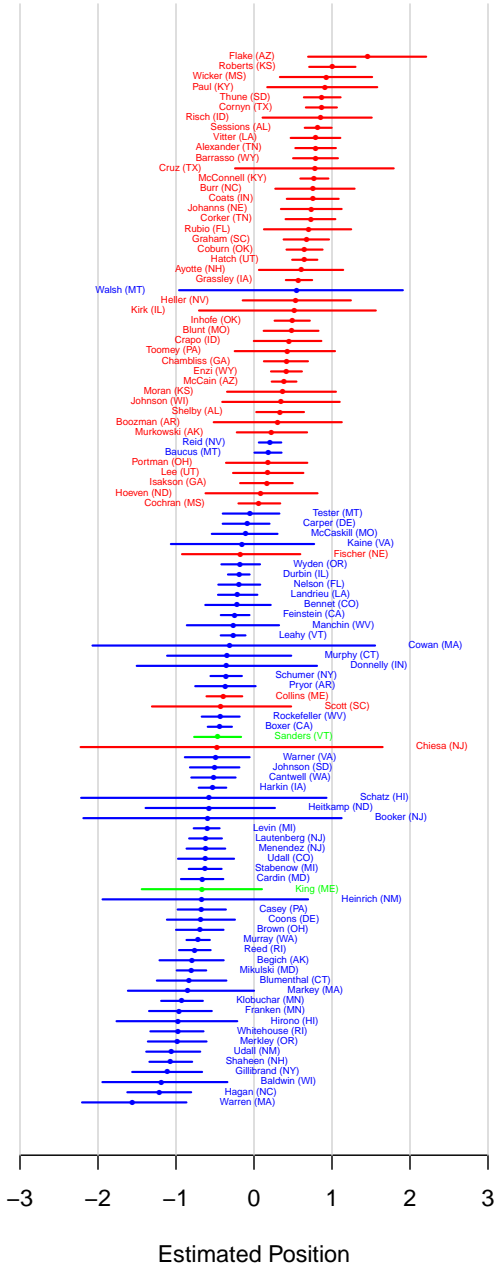


Figure A9: Wordshoal estimates for the 112th and 113th US Senates.

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.

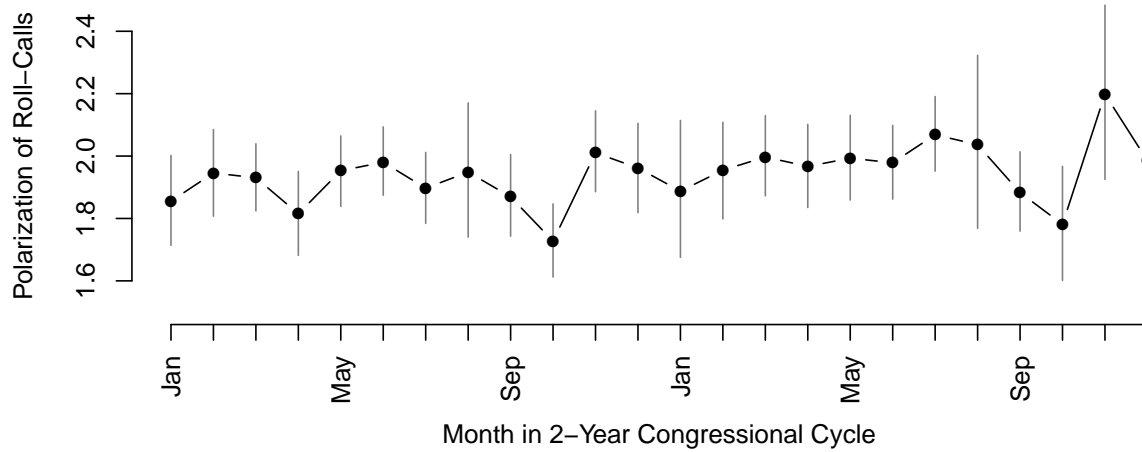
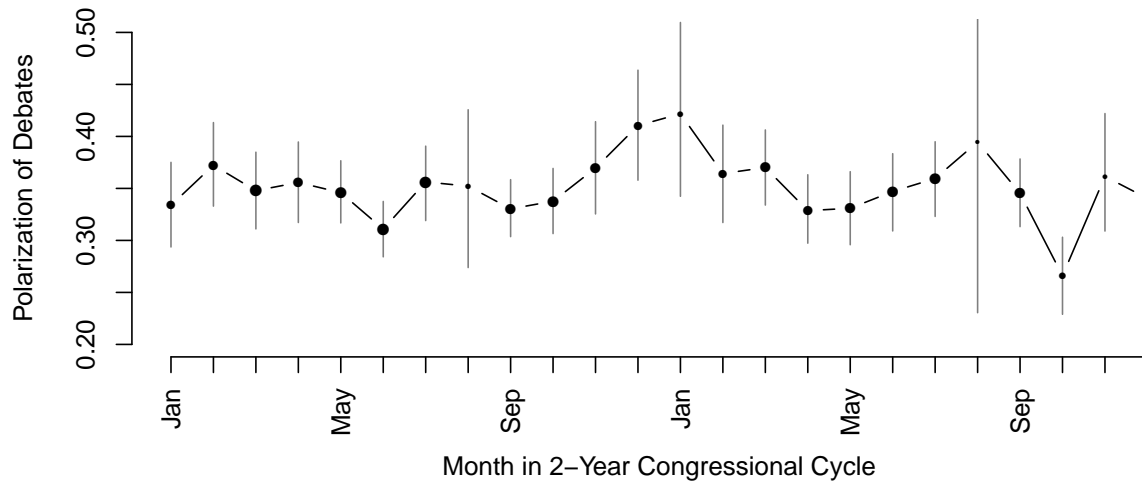


Figure A10: Average magnitude of debate loadings over the 24 months of a Congress, across the 104th to 113th Congress. The size of each point is proportional to the number of debates.

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

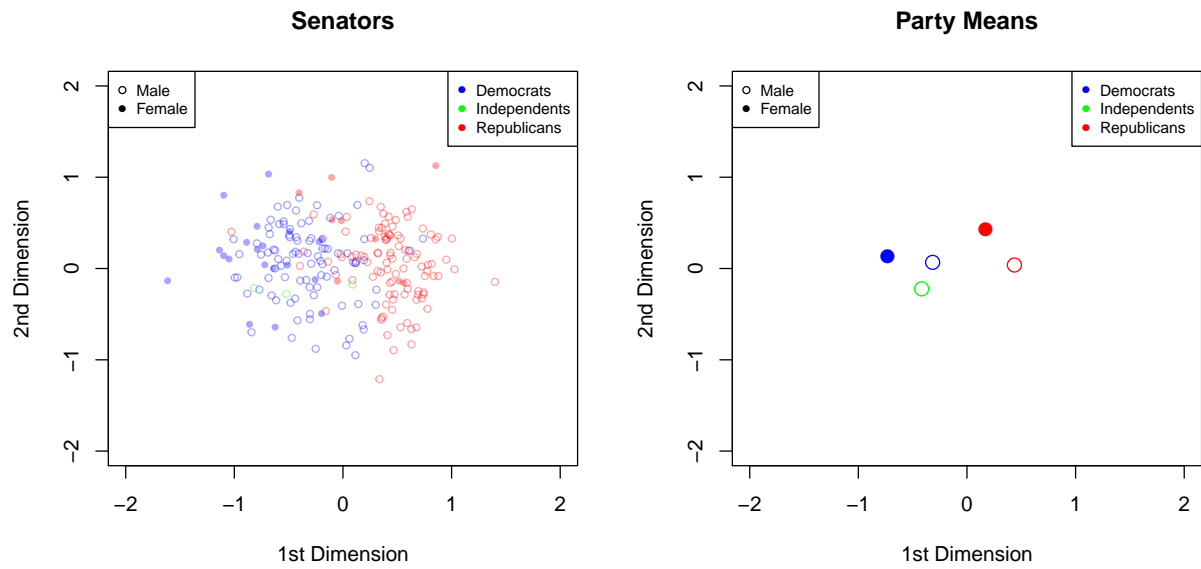


Figure A11: 2D Wordshoal speech scores for male versus female senators, by party.

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.