

# Measuring Partisanship in Congressional Committee and Floor Speeches

Pamela Ban

Harvard University, Political Economy and Government  
pban@fas.harvard.edu

## Positions of Committees vs. Parties on the Floor

- Congressional scholars have claimed that committees are “preference outliers” in policy positions compared to the chamber because members seek assignments that match their special or district-based interests.
- Empirical studies give this theory mixed support and have tested it by using measures based on roll call votes. However, using roll call vote measures neglect information before a floor vote – especially within committees – that may be valuable in revealing true policy positions.
- **Do committees have more partisan policy positions than the floor?** Using committee and floor speech data, I extract the relative positions of House committees and the floor on the policy area specific to each committee.
- Comparing positions from committee speech reveals varying committee–floor partisanship gaps from committee to committee.

## Committee Transcripts and the Congressional Record

- Raw data: Scraped transcripts of House committee hearings, along with the Congressional Record.
- Parse every committee’s hearings into a single text per committee. Parse each party’s House member floor speeches into a single text by party, to serve as reference texts.
- Using topics identified by a PLDA model for every floor speech (Bonica 2016), map House floor speeches to committee by policy area.
- Preprocessing steps for each text:
  - Remove stop-words and stem (nltk python package).
  - Remove proper nouns and politically uninformative procedural phrases (technical language of procedure and legislation).
  - Remove speech from witnesses (committee texts).
  - Generate counts of all two- and three-word phrases.

## Measuring Partisanship with DW-NOMINATE Scores

- Previous studies have used aggregated roll call vote measures or interest group scores to test whether committees are preference outliers.
- However, not every committee member participates equally in committee – varying amounts of committee participation by members may underlie any skews in the preferences of the committees. I account for this by weighting the average DW-NOMINATE score of a committee by members’ participation rates in committee.
- **Generate participation rates of committee members:** Using committee transcripts, calculate the relative number of times each member spoke within each committee.
- **Generate weighted score of each committee:** Use committee members’ DW-NOMINATE scores and their relative participation rates as weights to calculate a weighted average for each committee.

## Measuring Partisanship from Speeches

**Identify politically informative phrases:** Following Gentzkow and Shapiro (2010), let  $f_{pld}$  and  $f_{plr}$  be the total number of times phrase  $p$  of length  $l = 2$  or  $3$  is used by Democrats and Republicans, and let  $f_{\sim pld}$  and  $f_{\sim plr}$  be the total occurrences of length- $l$  phrases that are not phrase  $p$  spoken by Democrats and Republicans. The Pearson  $\chi^2$  statistic for each phrase is:

$$\chi_{pl}^2 = \frac{(f_{plr}f_{\sim pld} - f_{pld}f_{\sim plr})^2}{(f_{plr} + f_{pld})(f_{plr} + f_{\sim plr})(f_{pld} + f_{\sim pld})(f_{\sim plr} + f_{\sim pld})}$$

Select the top 1,000 two-word phrases and 1,000 three-word phrases with the highest  $\chi_{pl}^2$  values, for a final set of 2,000 phrases.

**Score committee speech using floor speech as the reference text**, following Laver, Benoit, and Garry (2003):

1. **Generate scores from two reference texts:** Use floor Democrats and floor Republicans as the reference texts.  $F_{pr}$  is the relative frequency of each phrase  $p$  (in the selected set above) used in reference text  $r$ . The probability that seeing phrase  $p$  means we are reading text  $r$  is  $P_{pr} = \frac{F_{pr}}{\sum_r F_{pr}}$ . For each phrase  $p$ , the expected position on dimension  $d$  when seeing phrase  $p$  is  $S_{pd} = \sum_r (P_{pr} \cdot A_{rd})$ , where  $A_{rd}$  is the reference text’s *a priori* position. For the two *a priori* positions, I use the two mean DW-NOMINATE scores of floor Democrats and floor Republicans.

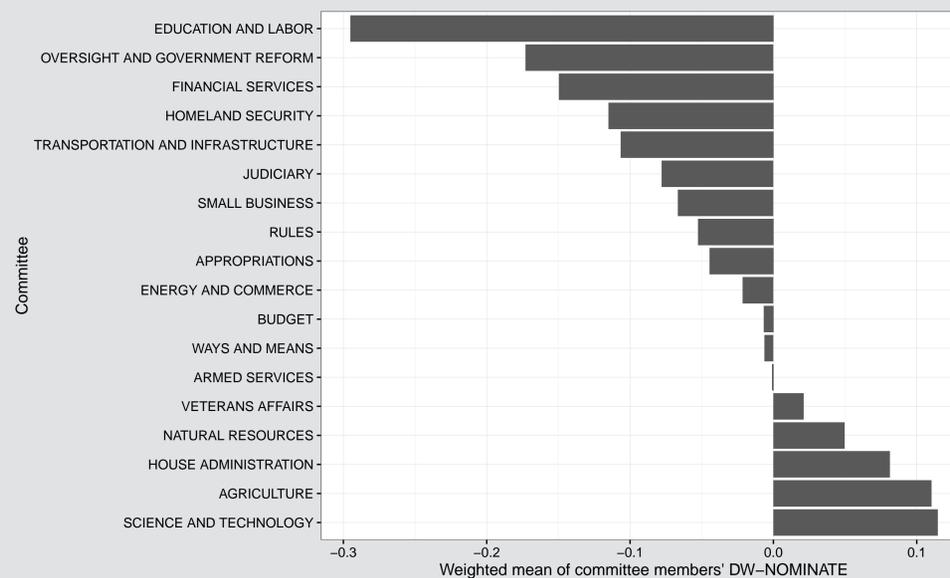
2. **Score each committee’s text:** Compute  $F_{pC}$ , the relative frequency of each phrase in each committee’s text. The score of a committee’s text is:

$$S_{Cd} = \sum_p (F_{pC} \cdot S_{pd})$$

3. **Transform committee text scores:** Preserve the mean and relative positions of the committee text scores and set the variance equal to the variance of the floor texts:  
 $S_{Cd}^* = (S_{Cd} - \bar{S}_{Cd}) \left( \frac{SD_{rd}}{SD_{Cd}} \right) + \bar{S}_{Cd}$ , where  $\bar{S}_{Cd}$  is the average score of the committee texts and  $SD$  is standard deviation.

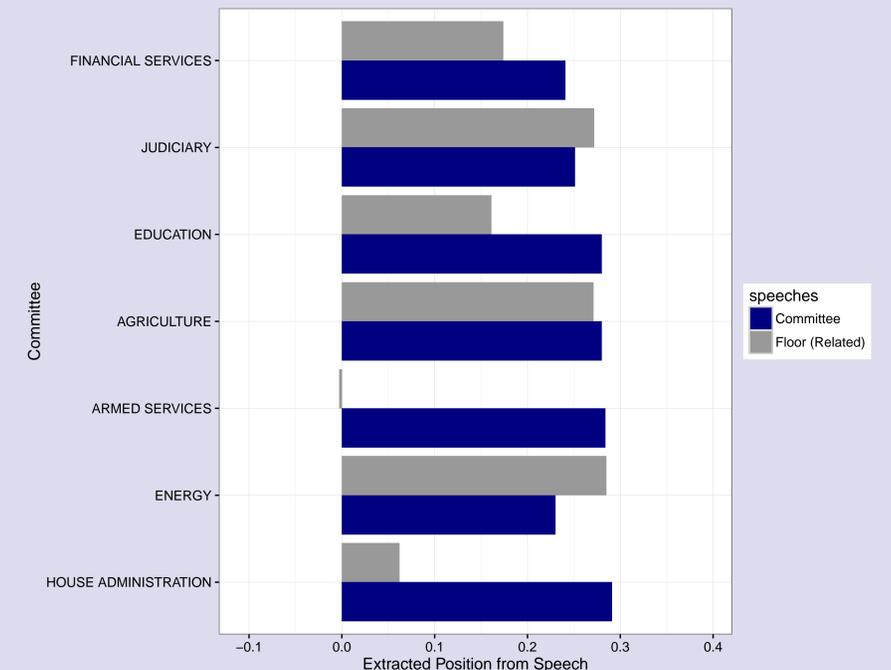
## DW-NOMINATE Scores Show Varying Partisanship

Committee DW-NOMINATE scores, weighted by members’ participation (110th Congress). Floor Democrats and Republicans had means of -0.347 and 0.647, respectively.



## Committee Speech vs. Related Floor Speech

This figure shows the results of extracting committee positions from speeches, for the 110th Congress. The floor positions shown are the positions extracted from the floor speeches on topics related to that committee’s policy area. The two reference texts are the set of all floor speeches by Democrats and Republicans.



## Current Observations and Next Steps

- Committee speech scores differ from the related floor speech scores. The difference varies by committee.
- Do members speak differently in committee than on the floor? Examine differences between committee speech and floor speech for individual members and factors that may drive any differences.
- Has polarization grown within committees (policy-making) as it has in Congress (party grand-standing?) Compare the gaps between parties within committee speeches and within floor speeches.
- Further work on filtering floor speeches by topic in order to use all the committee and floor speech data available.
- Incorporate comparisons across time and the Senate.

## References

- Bonica, A. 2016. “A Data-Driven Voter Guide for U.S. Elections.”
- Cox, G. and M. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.
- Gentzkow, M. and J. Shapiro. 2010. “What Drives Media Slant? Evidence from U.S. Daily Newspapers.” *Econometrica* 78(1): 35-71.
- Krehbiel, K. 1990. “Are Congressional Committees Composed of Preference Outliers?” *American Political Science Review* 84: 149-163.
- Laver, M., K. Benoit, and J. Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97(2): 311-331.
- Lowe, W. 2008. “Understanding Wordscores.” *Political Analysis* 16: 356-371.
- Weingast, B. and W. Marshall. 1988. “The Industrial Organization of Congress; or, Why Legislatures, Like Firms, are not Organized as Markets.” *Journal of Political Economy* 96: 132-163.