# Referee Report for "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech"

Nan Ge

December 14, 2018

## 1 Summary

In light of concerns about extreme political polarization in the US, there have been many efforts to measure ideological uniformity in the US political system. One critical question is how does today's level of partisanship compare with past records. It is important to see the dynamics of political partisanship in order to understand its behavior in a broader context. On the one hand, the pattern of political partisanship is a joint force of political bargaining, balancing strategy, national interests, etc. On the other hand, the impact of political partisanship could dissolve deep into the public discourse and policy making. It is hard to identify the causal relationship among these intertwined factors, but studying the trend could convey a consequential message.

Gentzkow, Shapiro and Taddy present a study where they attempt to use data on the text of speeches in the US Congress to measure the dynamics of partisanship in US political language. Overall, they deliver convincing evidence that partisanship in US political language is quickly increasing in recent decades, contrast to the message in previous studies, like Jensen et al. (2012). They define political partisanship as the "ease with which an observer could infer a congressperson's party" (Gentzkow, 2018, pp. 3) based on their rhetoric, and used machine learning algorithms to measure the level of partisanship. Next, they unpack the measurement into within-topic and between-topic partisanship. The analysis shows that there is no significant divergence in the topics Republicans and Democrats chose to talk about. Rather, most of the observed partisanship could be explained by differences in congressmen's choice of wording when talking about a given topic (Gentzkow, 2018, pp. 20). In the discussion, the authors further link the trend in political partisanship to major reforms in the US congress history. Although this is not causal inference, the dynamics could give us a hint about how congress reasoning interact with the public interests.

Their major contribution is to address the estimation difficulties with recent advances in machine learning. Attempt to measure political language partisanship falls into a broader research topic known as "measuring choices made by different groups" (Gentzkow, 2018, pp. 2). Particularly, it belongs to a subset of this topic where "the number of possible choices is larger relative to the number of actual choices observed" (Gentzkow, 2018, pp.2). In the context of political language, a congressman's vocabulary choice set is much larger than the vocabulary in his final speech. In scenarios like this, the choices we observed is only a finite, and most likely biased sample from the full distribution. Bias in sampling could lead to false dispersion between different choices and result in severe bias in the measurement. Plus, the high dimension of the choice set makes estimation computationally difficult. The authors propose an accurate and computationally feasible approach to address this bias. Furthermore, they present evidence of severe bias in the conventional maximum likelihood estimator (MLE).

The study uses a leave-out estimator and a penalized estimator to gauge the estimation. The leave-out estimator uses complement samples to estimate the frequency $q_t$ and the posterior $\rho_t$ [1] . The idea is to make the two error terms orthogonal to each other, thus mitigating the bias. We will talk more about their implementation of the penalized estimator below. For both the leave-out estimator and the penalized estimator, they performed inference in 100 random subsamples and reported the average confidence interval. I also like the first validation test in the paper. They randomly assigned each speaker to Democratic or Republican and performed the same estimation. The difference between actual observation and random assignment intuitively speaks for their model's estimation power.

# 2 Commentary

This is not the first paper to estimate political polarization with the help of machine learning methods. For example, Peterson and Spirling (2017) look at the same research question in a British context, using various supervised learning algorithms, including a stochastic gradient descent (SGD) classifier, and logistic regression with an L2 penalty (Peterson and Spirling, 2018, pp. 5). The greatest feature of this paper, however, is to combine parametric methods (apparently, the authors prefer not to call it structural model) with classic supervised learning algorithms. They first build a model featuring the choice set probabilities and posterior belief of the neutral observer. To implement the penalized estimator, a L1 penalty is imposed on the parameter for the effect of party affiliation on vocabulary choice. Then they approximate the parameters in the model with regard to the minimand. In their defense, this approach enables us to use "a generative model of the data and to measure group differences using objects that have a well-defined meaning in the context of the model" (Gentzkow, 2018, pp. 5). However, the paper doesn't report the estimation results, nor does it talk about the interpretation of these parameters. It seems to me that they are not realizing the full potential of the parametric method.

In the introduction, there is a brief review on other estimation approaches taken to address the finite sample bias, such as benchmarking against random allocation, bootstrap bias corrections, and estimating mixture models. It's unclear to the reader why this paper's proposal has an edge over other resolutions[2], and why they chose L1 penalty instead of other

[1] According to the authors, $q_t^P(\mathbf{x}_{it})$ is the vector of choice probabilities, $\rho_{jt}$ is the posterior belief that an observer with a neutral prior assigns to a speaker being Republican if the speaker chooses phrase j in session t and has characteristics x. $\rho_{jt}(\mathbf{x}) = \frac{q_{jt}^R(\mathbf{x})}{q_{jt}^R(\mathbf{x})+q_{jt}^D(\mathbf{x})}$

[2] The authors do compare their study with Peterson and Spirling (2018). The latter fails on the randomization validation test.

supervised learning algorithms. It would be helpful to elaborate the methodology review and talk about each method's contribution and potential flaws.

Some of the model's assumptions also need to be more carefully justified. As the authors have acknowledged, the model is restrictive. Their baseline model does not account for personal wording habits. Online appendix incorporates into speaker characteristic a speaker random effect, but the assumption for a random effect model requires that the unobserved attribute does not correlate with any of the other covariates, which is not likely if we want to interpret the random effect as habit. The vector of personal attributes $x_{it}$ in the baseline model consists of indicators for state, chamber, gender, Census region, and whether the party is in the majority for the entirely of the session (Gentzkow, 2018, pp. 13). Four out of five of these characteristics don't change across sessions, which yields very little variation in the personal characteristics variables. Some salient indexes, such as ethnicity and education were excluded from the baseline model. The authors claimed to have considered specifications of $x_{it}$ with different sets of observable characteristics, but they didn't report the results.

This paper also gives an innovative definition for partisanship. They conceptualized the average partisanship as how well "an observer with a neutral prior" could identify the speaker's true party after "hearing the speaker utter a single phrase". A higher measure suggests that the phrase is more informative about a congressman's party affiliation, thus more partisan. To the best of my knowledge, this definition is first seen in this paper. The authors mentioned in Part 4 that the conventional measurement of Euclidian distance would lead to serious bias in finite samples (Gentzkow, 2018, pp. 12). Nevertheless, it would be better if they could talk more about the considerations behind this definition. Specifically, it would be helpful to talk about how this new definition helps to mitigate the finite sample bias, and its explanatory power compared with other notations.

As regard to the results, it would be helpful to add the authors' interpretation about the magnitude of partisanship increase in their work and in previous work. So far, the figures mainly illustrate the time trends. However, the variations in estimation are relatively small. For example, the preferred penalized estimator suggests that the difference between the peak and bottom of average partisanship is less than 0.01 over the entire time span. It might be unclear to the readers what the 0.01 increase implies. If the estimation is only informative about "the relative level of polarization" as in Peterson and Spirling (2018), the authors may want to put up a similar notice before dwelling into the results.

This paper is an improvement over the authors' 2017 working paper: Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech. Looking at the topic, it seems that the authors were trying to generalize their methodology to a broader topic. However, the authors failed to clarify which features of their estimation scheme is applicable to similar research topics. The authors address the broader literature clearly, though some of the citations are redundant. For example, the bulk citations in footnote 5 are very remotely related to the topic and methodology of this study. The authors were very discrete in interpreting the surge in partisanship in recent years. It has been a hot topic since the Trump administration. Explorations have been done in social media like Gentzkow et al. (2016), demographics like Boxell et al. (2017), geographical like distributionAllcott and Gentzkow (2017) and fake news like Rodden (2010). Although it might be out of the picture of this paper, the authors might want to point readers to these interesting threads.

There are some additional minor issues requiring revision. First, it would be helpful to add figure and table titles for those in the online appendix. It takes a lot effort to figure out which figure or table the authors were referring to. Second, Figure 3 shows the informativeness of speech by number of phrases and congress sessions. It would be clearer to talk about how the figure is produced in the text, instead of in the footnote of the figure.

# 3 Potential Extensions

As an extension of the paper, it would be interesting to take the method and result of this paper, and study the interaction between polarization in lanugage and polarization in power. The extension comes from the authors' comment that language is "a striking exception" (Gentzkow, 2018, pp. 25). It appears to me that language is more of a strategic move than a truthful confession of the speaker's stand. Therefore, I want to study how a congressman's choice of rhetoric changes with regard to his/her opponent's choice of rhetoric. However, there are two intertwining strengths. First, the politician might stay close to the opponent to win the discussion. Second, the politician might differ himself/herself from the opponent to win over votes. Which force is stronger depend on several factors. Here are some of my speculations. First, the relative power of the party in the congress would affect whether the congressman need to directly debate with this/her opponent. It is easier for a majority party to lead the debate, so the first force would be weaker for its congressman. Second, the distance between the congressman's political ideology and that of the other speakers would affect the direction of strategic move.

To identify the mechanism, we could regress each speaker's score of partisanship to our variables of interest. So far, we have two variables, whether the speaker's party was the majority in the congress, and the opponents' political ideology. The latter variable is defined as the average Common Space DW-NOMINATE score (representing political ideology) of the other speakers in the same session. Alternatively, we could follow this paper's method. Add the variables into the minimand of the penalized estimator, and update the observer's posterior. The model could be interpreted as leaking additional information to the observer. Note that in order to compare the effect of the two forces, we could only add one variable at a time.

# References

**Allcott, Hunt and Matthew Gentzkow**, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–36.

**Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro**, "Greater Internet use is not associated with faster growth in political polarization among US demographic groups," *Proceedings of the National Academy of Sciences*, 2017, p. 201706588.

**Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy**, "Measuring polarization in high-dimensional data: Method and application to congressional speech," Technical Report, National Bureau of Economic Research 2016.

**Jensen, Jacob, Suresh Naidu, Ethan Kaplan, Laurence Wilse-Samson, David Gergen, Michael Zuckerman, and Arthur Spirling**, "Political polarization and the dynamics of political language: Evidence from 130 years of partisan speech [with comments and discussion]," *Brookings Papers on Economic Activity*, 2012, pp. 1–81.

**Peterson, Andrew and Arthur Spirling**, "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems," *Political Analysis*, 2018, *26* (1), 120–128.

**Rodden, Jonathan**, "The geographic distribution of political preferences," *Annual Review of Political Science*, 2010, *13*, 321–340.