My title*

My subtitle if needed

Samarth Rajani

April 17, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

In the global landscape, where each nation sets to maximize its own gain and self interest, the European Union holds a special place. The only political and economic union of its kind, it spans 27 nations in Europe,

However, there is an inherent concern with being party of a larger authority - there is a risk of a misalignment of potential goals. With 26 other states involved, policy across the union becomes a balancing act between furthering an individual nation's goals without compromising another. This unique situation burdens legislative bodies of the nations involved to consistently discuss and negotiate EU-related matters, potentially diverting money, time and manpower away from domestic issues, all of which are important resources. Therefore, the legislative bodies must balance engaging in EU related discourse and tending to the needs of the local public.

The United Kingdom was an influential part of the European Union. Even though the United Kingdom joined the union later than other European giants like France and Germany (15 years later), it spearheaded efforts in making the union one. Most notably, the UK presented a plan that included 282 legislative actions at the European Council in Milan in June 1985, which was the foundation of the single market within the EU. This enabled complete free movement of goods and services within the EU, and without the involvement of the UK the Union may have taken a more protectionist approach(https://www.politico.eu/article/britain-leaves-mark-eu-brexit/). With its economic prowess, UK also commanded areas of foreign policy, free trade and growth among others.

---

*Code and data are available at: LINK.

However, the UK's allegiance to the union did not stand the test of time, succumbing to the dilemma described before. The driving force behind the 'Brexit' in 2020 has been a concern for lack of sovereignty, where the public felt the European Commission was taking away the UK's ability to make decisions for itself. David Frost's speech in 2020 best highlighted this position (https://no10media.blog.gov.uk/2020/02/17/david-frost-lecture-reflections-on-the-revolutions-in-europe/) and detailed how the European Union's institutions felt abstract and "disconnected from natural feeling". After much deliberation since 2016, the UK pulled the trigger in 2020, successfully exiting the Union in dramatic fashion.

Keeping this in mind, it is worth examining how often matter of the European Union dictated discourse in the United Kingdom. Specifically, this paper studies how often each party gave speeches that involved discussions around the European Union right down to the level of individual sentences. Looking at the ParlEE(https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910 dataset, this paper examines how likely is one political party to devote time to discussing EU related matters as compared to domestic ones.

[Paragraph on results]

The remainder of this paper is structured as follows. Section 2….

2 Data

This paper uses the ParlEE plenary speeches data set to conduct its analysis. This dataset specifically has been obtained from version 1, and uses the UK speech data. The dataset has information on speeches from 2009 to 2019.

For their data collection, the authors chose the government body bearing the most responsibility for legislation. For the United Kingdom they chose the House of Commons. As a starting point, the authors used the The Rauh et al ParlSpeech dataset set since it already contained a large collection of raw speech data for many countries of interest, especially the UK. There was some data missing towards the end of the timeframe of interest (towards 2019), so the authors went through the speeches from (https://hansard.parliament.uk/commons) and filled in the missing parts themselves using an API provided to them.

The authors ironed out typical wrinkles with government related datasets - removing boilerplate information, page numbers, eliminating corrupted social characters and introducing consistent naming conventions for party and speaker alike. This made the data uniform across countries, improving readability.

The dataset has variables explained below :

- **Date** : The date on which a specific sentence from the speech has been said is recorded in the 'date' variable. This is recorded in the 'DD/MM/YYYY' format, with duplicated values for multiple sentences spoken in the same speech as well as if there were multiple speeches on the same day.

- **Agenda** : The variable 'agenda' is the title given to a set of speeches given pertaining to the issue at hand in the House of Commons at the time. This is provided by official sources.

- **Speechnumber** : The variable 'speechnumber' gives a unique identifier to all the sentences spoken in the same speech. So if the first speech in the data set has 5 recorded sentences, they will all be allotted the number 1.

- **Sentencenumber** : The variable 'sentencenumber' assigns a unique identifier to every sentence spoken within the same speech. Continuing the example above, every sentence in speech number 1 is allotted a sentence number 1 through 5. This variable resets from speech to speech, such that the first sentence in every speech will always have number 1.

- **Speaker** : The variable 'speaker' indicates the name of the member making the speech. Measures have been taken to eliminate inconsistencies, such as removing any nicknames, variations in spelling or honorary titles and using only one standard naming convention. The only exception is the usage of ministerial titles in the speaker column, since it is worth indicating when the same speaker is speaking as 'the Minister' versus just an individual. This is because that being in that position may lead to influence in the opinion of the speaker as demanded by the given title.

- **Party** : The 'party' variable records the alignment of every speaker. The parties recorded are APNI, Conservative Party, DUP, GPEW, Labour Party, LibDem, other, Plaid Cymru, SDLP, SNP, UKIP and UUP.

- **Text** : The 'text' variable records the raw text of sentences part of a speech given in parliament. This has already been cleaned as stated previously, and therefore is ready to be studied across the other variables like party, agenda, etc.

- **Parliament**The 'parliament' variable stores the name of the legislative body. For the UK and for the purposes of this paper, this variable is 'UK- HouseofCommons' everywhere.

- **iso3country** : Similarly, the variable 'iso3country' records the country in question, which for this paper is the United Kingdom throughout, recorded as 'GBR'.
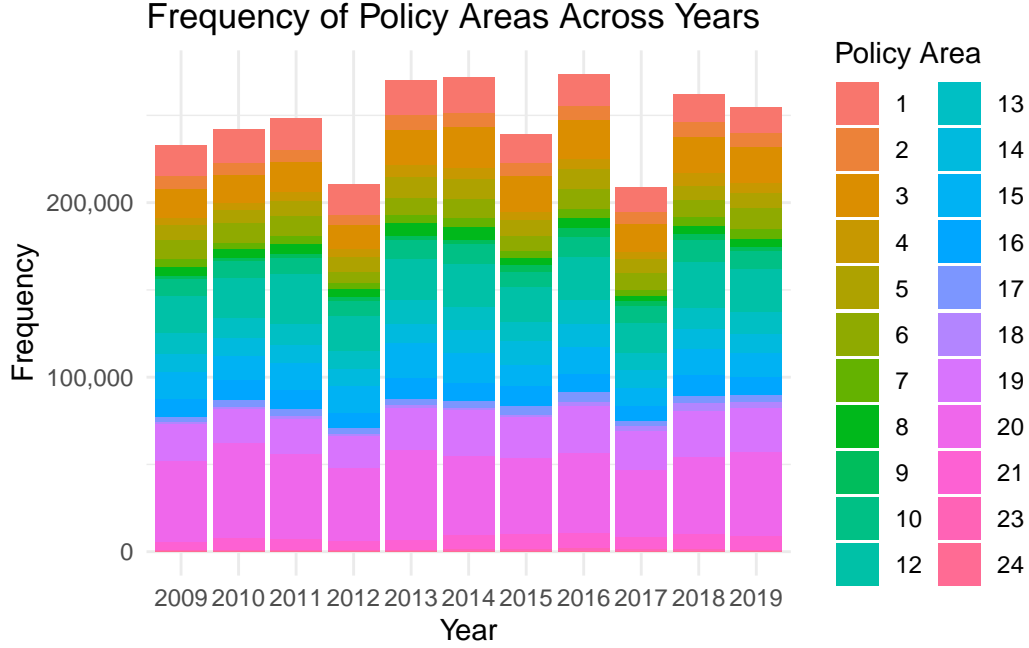
The authors augmented the collected data so as to make it enable analysis beyond what was already possible. These variables are described below:

- **EU** : The variable 'eu' is a dummy variable introduced by the authors that determines whether the given text or sentence discusses anything related to the European Union. The dummy is '1' if it does discuss the European Union and '0' if it doesn't.

- **Policyarea** The variable 'policyarea' assigns a number in accordance with the specific policy topic of a text. The numbers used are from the CAP major categories coding scheme.

The CAP coding scheme is shown in the table below :

```
           Policy.Area CAP.number
1          Macroeconomics          1
2          Civil Rights            2
3                 Health           3
4             Agriculture          4
5                 Labour           5
6              Education           6
7            Environment           7
8                 Energy           8
9             Immigration          9
10         Transportation         10
11           Law & Crime          12
12         Social Welfare         13
13                Housing         14
14      Domestic Commerce         15
15                Defence         16
16             Technology         17
17          Foreign Trade         18
18 International Affairs          19
19 Government Operations          20
20           Public Lands         21
21                Culture         23
22               Fisheries        New
```

The following graph shows the occurrence of the given policy types over the years:

Frequency of Policy Areas Across Years

Frequency

200,000

100,000

0

2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

Year

Policy Area

1 13
2 14
3 15
4 16
5 17
6 18
7 19
8 20
9 21
10 23
12 24

3 Model

We use a linear logistic regression model in order to estimate how likely each individual party is to talk about European Union related affairs. Using the 'eu' binary values, the regression is run using the Alliance Party of Northern Ireland (APNI) as the predictor category

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algorithm

Figure 1: **?(caption)**

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.