QCA

15 May 2017

Data and Methodology

Data

The data for this endevours is sourced from multiple sources. The primary dataset to analyse the differences in party ideology comes from the Comparative Manifesto Project. The project adresses the collection and the comparative content analysis of parties' manifestos, covering over 1000 parties from 1945 onward in over 50 countries on five continents. The dataset is limited to the major political parties in Germany for our analysis. Furthermore, due to the lack of historical archives to media articles, as well as issues with storing and analysing the huge corpora of text, the data is further limited to elections post 2000 for comparison with trends in media.

Dataset for media is collected from scraping Der Spiegel, a daily left-leaning German news publisher. Der Speigel is unique in their free access to their archives as well as their pre-categorisation of the news articles. All articles on Der Spiegel are categorised into topics such as politics, sports, culture . . . , with the topic of politics being further categorised into domestic and international political news. The pre-processing of data into broad categories, allows us to focus on collecting only the relevant dataset towards our analysis, without the risk of bias. Thus, we scrape all article which relate to domestic politics from 2000 onward.

The dataset on news is further limited to only articles which were published 1 year before each election. Thus for the 2002 elections, articles published inbetween the periods September 2001 and 2002 were scraped, and so forth. This is led by the assumption of *limited memory*; voters and individuals are constrained by their limited memory when making decisions. Studies have noted the consumer choice to be bounded by small term memory and processing skills [@James1986, @Dick2017]. In a similar vein, voters are limited to recent memory when making a choice. Similarly, individuals do not focus on historical events, but on the events that occur closer to elections. Thus it would be imperetive to assume that the events that are closer to elections, and the media coverage of these events, should drive the narrative of political parties as well. It is however important to note that the use of 1 year before each election is an assumprion and may be violated. There are no studies which look at the limits of long-term memory when it concerns towards political choice, and the use of different data intervals are left to future research.

The website is scraped using Scrapy, a webscraping framework written in Python. The dataset for news articles post 2000 onward is encompasses nearly 70000 observations, though they are limited to nearly 14000 for our analysis. The data as well as the scraping codes are available on Github.

Methodology

The analysis can be broken into two steps; the first step will look at analysing the differences withing the priorities for each major party in Germany. We will use the CMP dataset to work on the differences between the parties using both a measure of right-left slant of the topic, as well as the importance of that topic for the party, measured using the distribution of the different topics.

The next step would be to compare the individual distribtion for the parties with the distributions derived from news articles. High correlations between the two would imply a convergance between the different parties on major topics, while low correlations would imply a dichotoy between the issues of the hour and the response to the parties to those issues.

A brief divergence must be made here to defend the methods followed by the paper. We use a topic model to differentiate the different topics in news across the different years, but we do not provide them with a left

or a right leaning political score. There are two reasons for this approach; Firstly there is reason to belive that the scores might be biased for our dataset. As explained earlier, Der Spiegel is a left leaning daily news publisher in Germany. As such it is not difficult to beleive that the scores may be biased against right and conservative parties i Germany.

Secondly, we believe that it is a difficult task to compare the political scores between media and political parties to explain the convergance or dichotomy between the parties and the general public. The act of comparison presumes a standard to which one or the other party must reach, in our case we presume that the media is a *vox populi*, and political parties will be driven towards addressing those voices. However by using political scores such as rile, we further assume the slant, and opinions, of the media as a gold standard to which other parties must concur. This is a dangerous notion, both for democracy and free speech.

Results

Inter-Party Differences

Party and Media

Conclusion and Further Work

References