

Final Manifesto

Raju Adhikari and Devvart Poddar

May 13, 2017

Introduction:

In this paper, we aim to understand the intricate relationship between the political parties and news media based on the topics that political parties and media often touch upon. In a more philosophical sense, different political parties are entities with similar goals but different means of getting there. These means can be construed as the mandates of these parties. For example, the mandate of left-wing parties is to achieve social and economic justice through government intervention and welfare systems, while that of right-wing parties is to achieve social and economic justice through less intervention and allowing the market to function freely. While taking up these different approaches (means), the parties often reach out to the voters through a selection of ideas, topics, and verbiage that is unique to their philosophy. In this paper, we analyze these “contents” of political rhetoric of three main German national parties. We will use the Manifesto Dataset to first see the differences between the three parties in terms of different social and political agendas over the years. Second, we will see which topics do these parties use frequently in order to assess their approach in appealing to the voters. Then, we will combine this data with the topic distribution dataset obtained from newspaper articles to analyze the role of media in facilitating the interaction between political parties and voters.

Data and Methodology

Data

The data for this endeavour is sourced from multiple sources. The primary dataset to analyse the differences in party ideology comes from the Comparative Manifesto Project. The project addresses 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 Spiegel is unique in their free access to their archives as well as their pre-categorization of the news articles. All articles on Der Spiegel are categorized into topics such as politics, sports, culture . . . , with the topic of politics being further categorized 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 in between 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 imperative 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 assumption 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 web scraping framework written in Python. The dataset for news articles post 2000 onward 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 distribution for the parties with the distributions derived from news articles. High correlations between the two would imply a convergence between the different parties on major topics, while low correlations would imply a dichotomy 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 believe 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 believe that the scores may be biased against right and conservative parties in Germany.

Secondly, we believe that it is a difficult task to compare the political scores between media and political parties to explain the convergence 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 to be a gold standard which other parties must concur towards. This is a dangerous notion, both for democracy and free speech. It is a better standard to allow the different parties to differ in their response to an issue, as long as the political parties follow the urgency of a particular issue. Thus by comparing the topic distributions between parties and media, we focus only on the *importance* accorded to each topic by the political parties, allowing them to differentiate the responses according to their political leanings.

Add LDA and lemmatisation

Results

Inter-Party Differences

The following table shows the summary of descriptive statistics for the four following variables: rile, planeco, markeco, welfare. **can we plot this?**

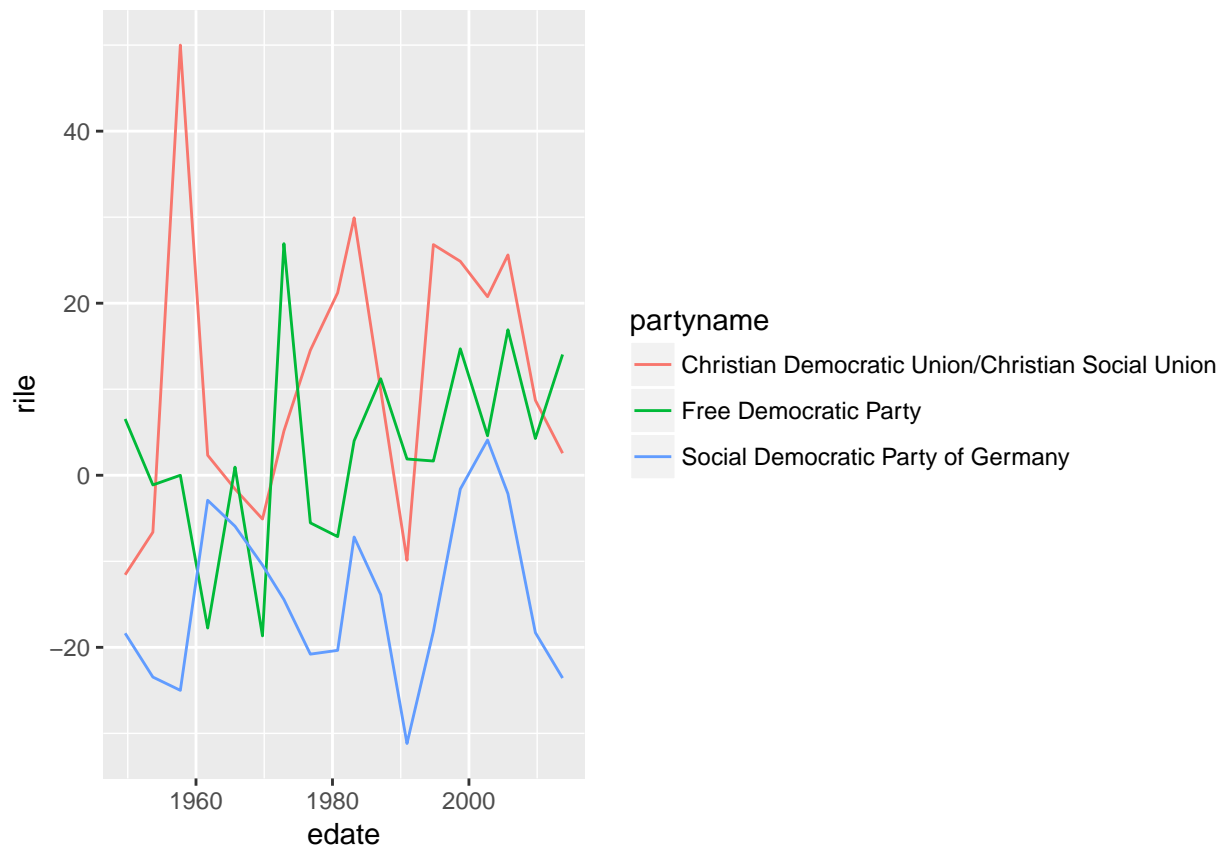
Table 1: Summary of Descriptive Statistics

	rile	planeco	markeco	welfare
nobs	54.00	54.00	54.00	54.00
NAs	0.00	0.00	0.00	0.00
Minimum	-31.18	0.73	0.00	2.90
Maximum	50.00	15.65	22.22	20.99

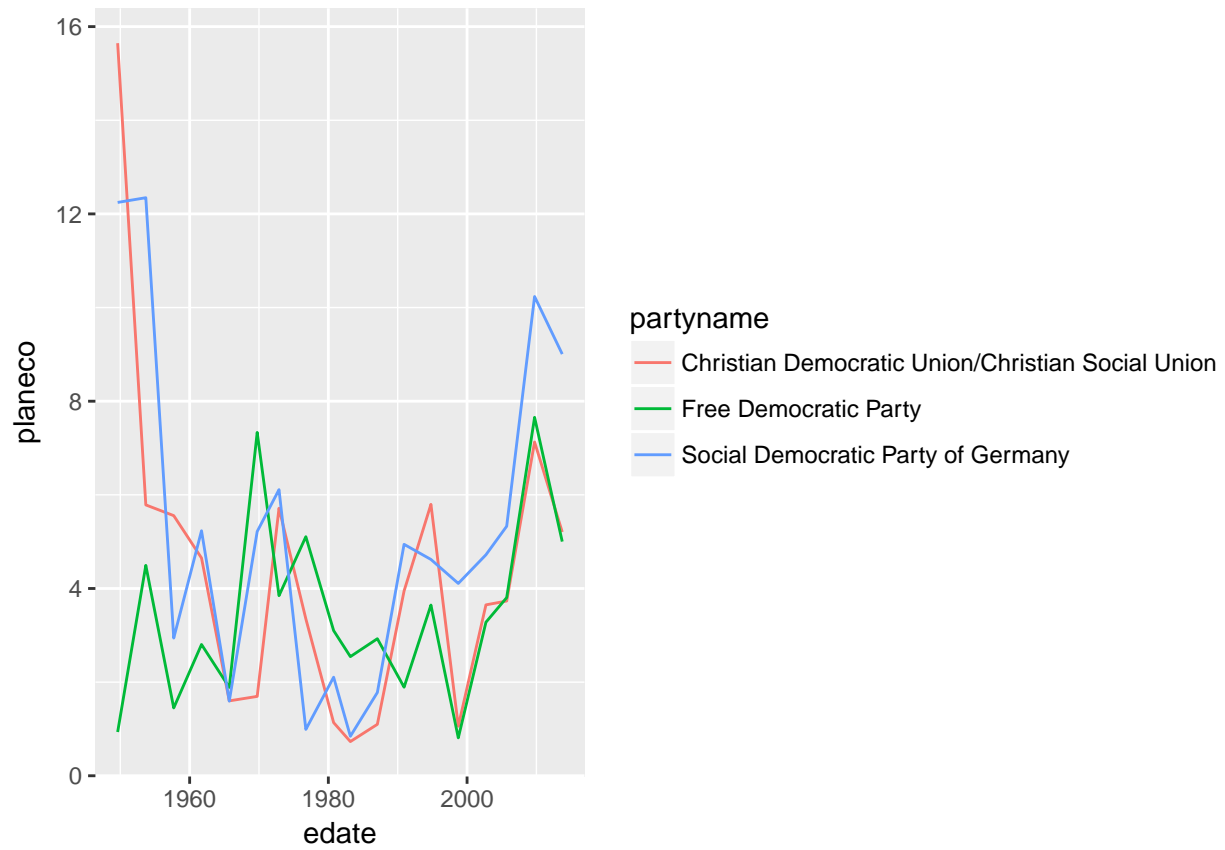
	rile	planeco	markeco	welfare
1. Quartile	-11.28	1.89	2.37	6.48
3. Quartile	9.60	5.30	8.83	13.43
Mean	0.21	4.34	5.91	10.34
Median	-0.56	3.83	5.06	9.79
Sum	11.53	234.37	319.34	558.57
SE Mean	2.26	0.42	0.61	0.66
LCL Mean	-4.31	3.49	4.68	9.02
UCL Mean	4.74	5.19	7.15	11.67
Variance	275.01	9.73	20.37	23.63
Stdev	16.58	3.12	4.51	4.86
Skewness	0.48	1.46	1.15	0.44
Kurtosis	0.04	2.44	1.65	-0.90

In the graphs below, we can see the change in the above four variables across the timeperiod from 1949 to 2013 for both political parties for each election year.

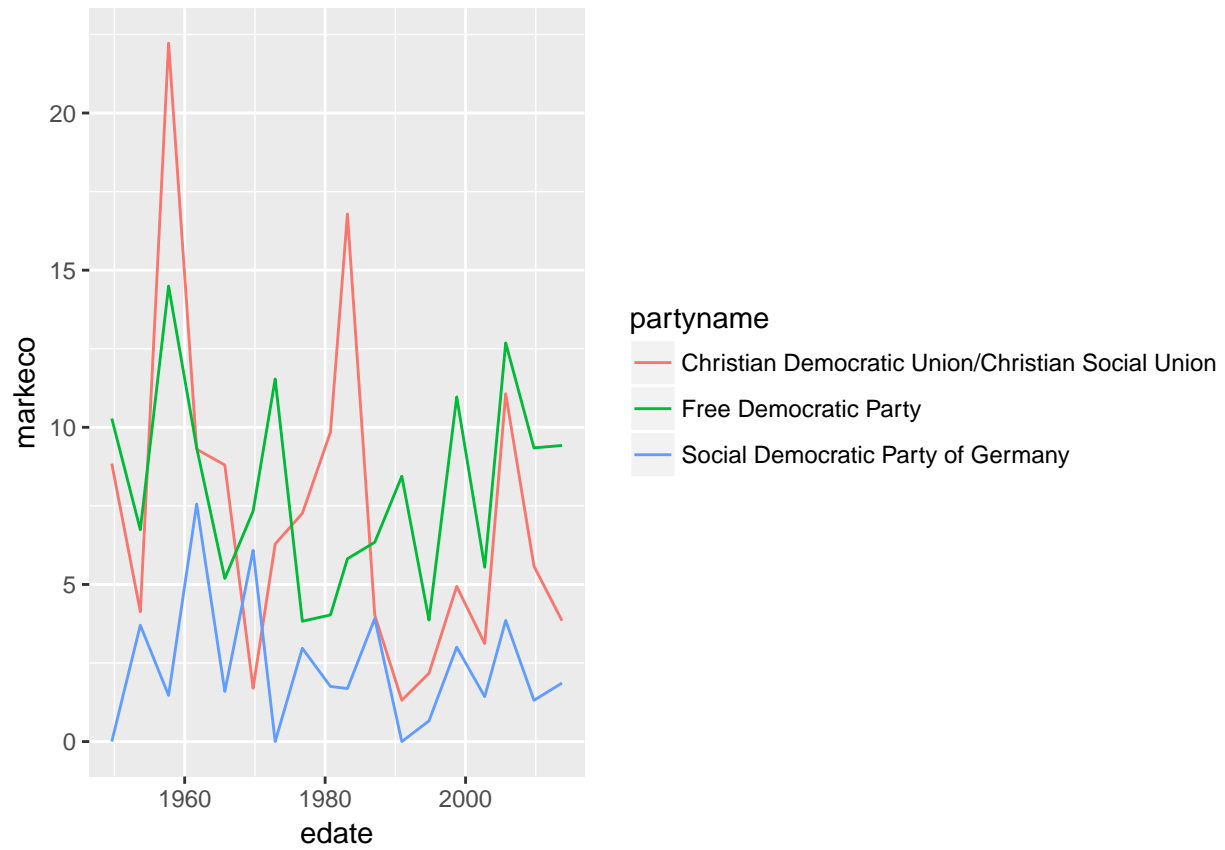
Rile indicates the position of the party in the left-right index, left meaning liberal and right meaning conservative.



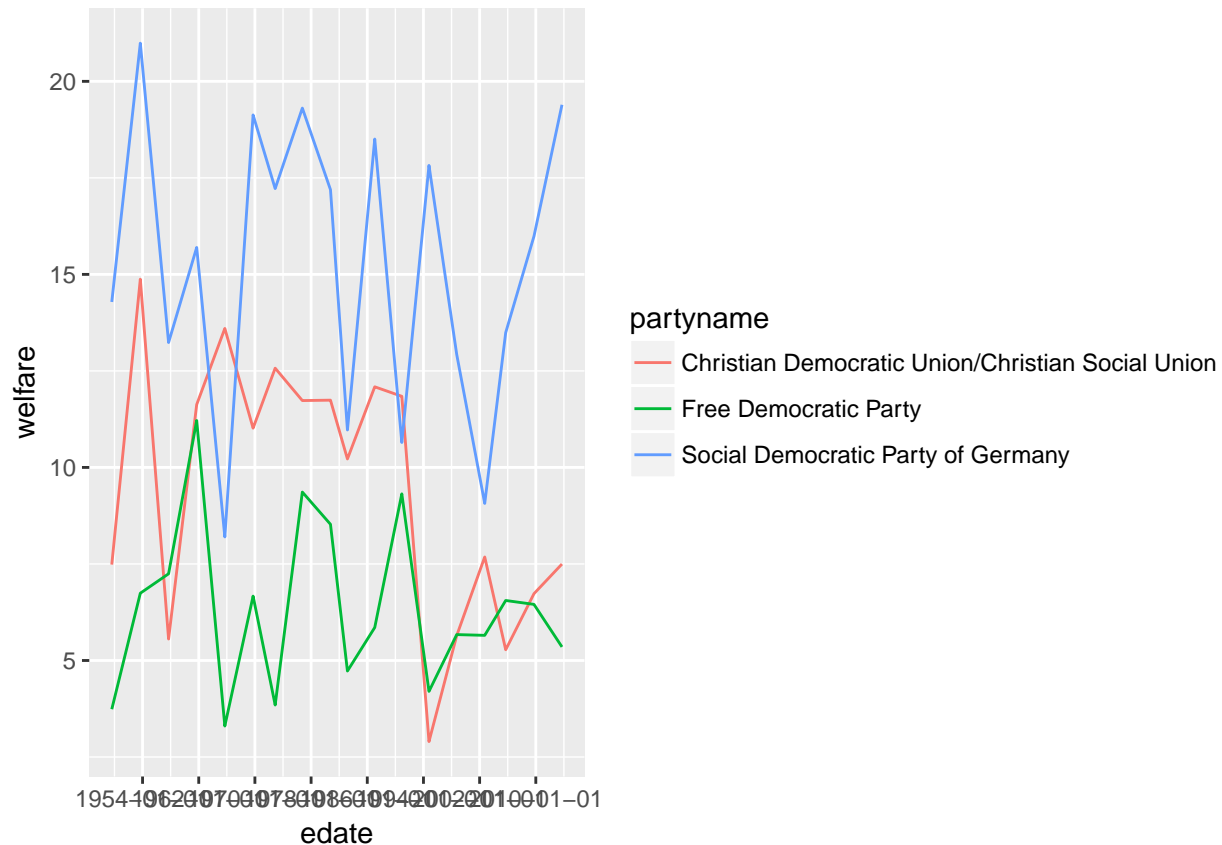
The variable planeco captures three important aspects: Market Regulation, Economing Planning, and Controlled Economy. Higher planeco score means that the government calls for increased consumer protection, increases competition, reduces monopoly, defends small businesses from big companies, takes an active role in economic planning, and introduces price control and minimum wage policies.



The variable `markeco` represents the policies in favor of free market economy.



The variable welfare captures equality and welfare state expansion.



Part IV: Alternate left-right measures:

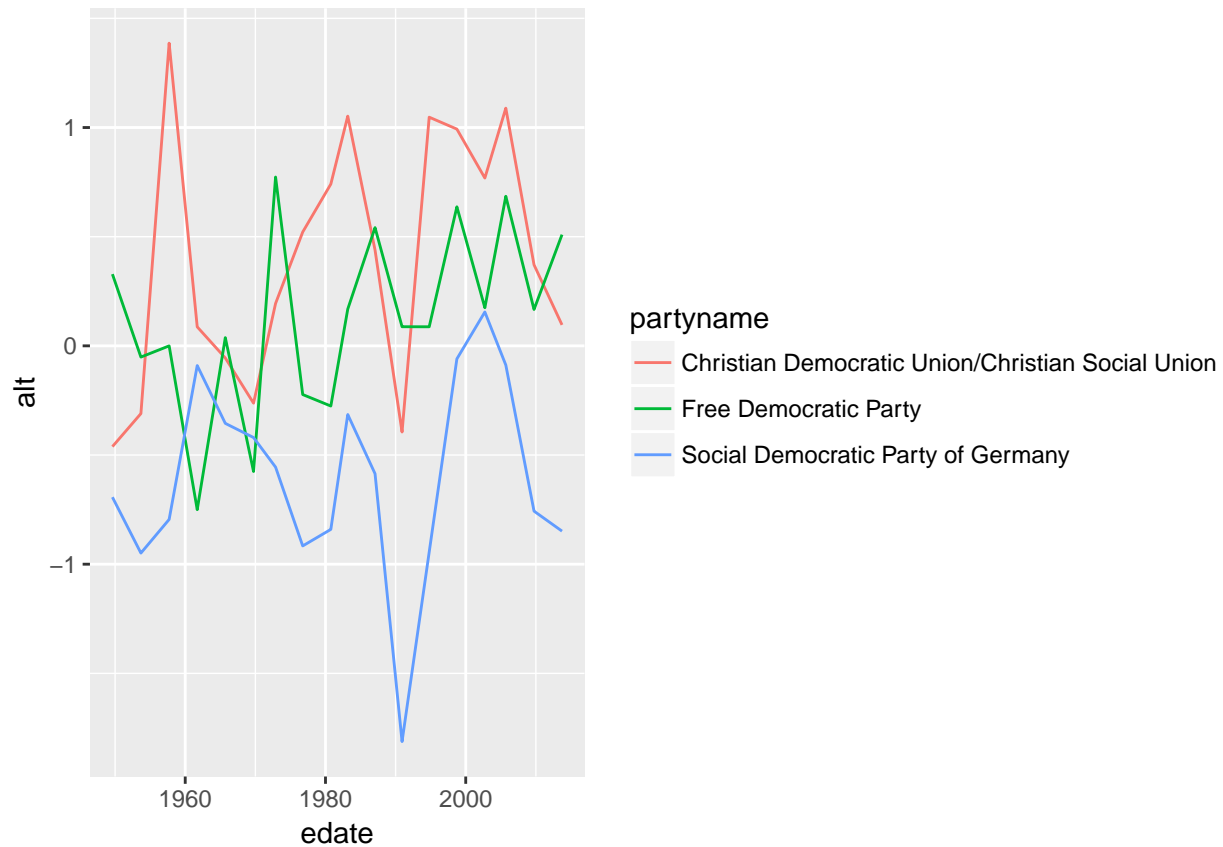
One alternative way for rile as proposed by Lowe et al. 2010 suggests taking the log of the ratio of Right and Left. For this we combine as all the variables that make up Right and Left separately using the Codebook, then take a ratio of right to left and perform a log on it.

Right (R) = per104 + per201 + per203 + per305 + per401 + per402 + per407 + per414 + per505 + per601 + per603 + per605 + per606

Left (L) = per103 + per105 + per106 + per107 + per403 + per404 + per406 + per412 + per413 + per504 + per506 + per701 + per202

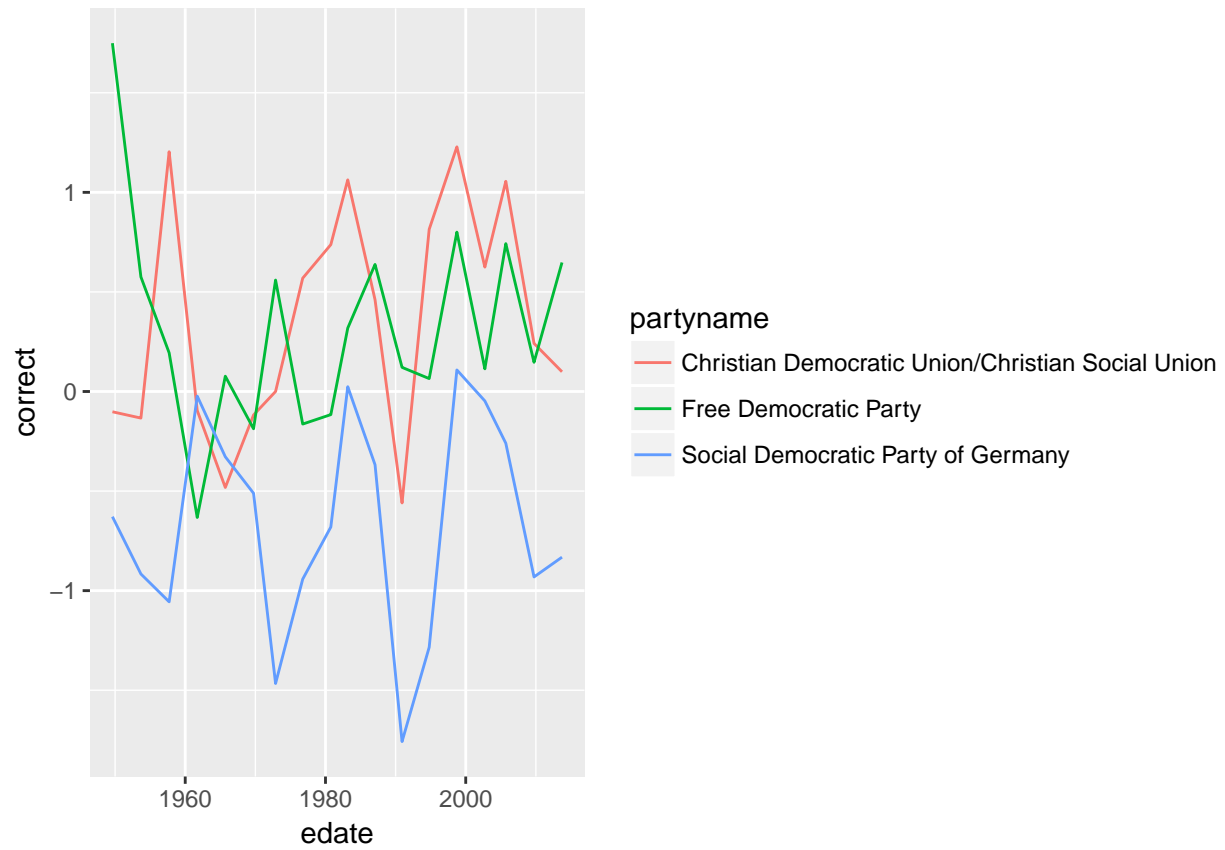
Alternative Rile (alt) = $\log(R/L)$

The graph below does not show any significant change in comparison to normal rile calculated above in section II.



Part V: Most Unreliable categories:

Mikhaylov et al. 2011 underscore the presence of several unreliable categories. For the left side of the rile scale they point out 202 and 701 as extremely unreliable and for the right side they point out 605 and 305 as extremely unreliable due to their extremely low kappa coefficient. After the removal of the unreliable categories, the rile graph below shows some significant changes.



Party and Media