CESifo Working Papers

Analysing Party Preferences Using Google Trends

Mirko Seithe

Lena Calahorrano

CESifo
Center for Economic Studies & Ifo Institute

Analysing Party Preferences Using Google Trends

Abstract

The formation of party preferences is a complex and not yet fully understood process based on a number of factors. This process, which is of great interest for both social and political science, is usually studied using questionnaire data which has proven to be a very reliable yet often costly and limited approach. Advances in technology and the rise of the internet as a primary information source for many people have created a new approach to keep track of people's interests. The major gateways to the internet's information are the so-called search engines, and Google, arguably the most commonly used search engine, allows scientists to tap the vast source of information generated by its users' search queries. In this paper we describe how this data source can be used to estimate the effect of different issues on party preferences using German voters and the German party system as an example. We find that using data provided by Google Trends can lead to a variety of interesting and occasionally counter-intuitive insights into peoples' party preferences.

Mirko Seithe*
Fraunhofer Institute for Applied
Information Technology FIT
Research Group MIKMOD
Schloss Birlinghoven
Germany – 53754 Sankt Augustin
mirko.seithe@fit.fraunhofer.de

Lena Calahorrano
Fraunhofer Institute for Applied
Information Technology FIT
Research Group MIKMOD
Schloss Birlinghoven
Germany – 53754 Sankt Augustin
lena.calahorrano@fit.fraunhofer.de

*corresponding author

1. Introduction

According to the established spatial models of party competition[1], individual vote choices can be seen as the result of the voter's attempt to choose a party whose intentions align most closely with her own policy preferences. Alternative approaches assume that different political parties emphasise *different* issues and that voters choose the party that they perceive as most competent with regard to a salient issue.[2]

Both approaches imply that changes in party preferences and the resulting voting behaviour may be caused by changes in the perceived relevance of topics, on a microeconomic level (e.g. active labor market policies after becoming unemployed) but also on a macroeconomic level (e.g. environmental protection after a major accident), by changes in the political parties' programmes or by the perceived performance of the governing party in dealing with specific topics.

The established approach to elicit people's interest in certain topics is to use questionnaires. While valid, this approach is both expensive and limited in that it only allows the analysis of a small subset of topics which furthermore has to be specified in advance. In this paper we use relative search volume data provided by the Google Trends service as a proxy for the German voters' interest in selected topics. The Google Trends service provides information on the prevalence of specific search terms, i.e. how often people entered a term into Google's search engine, on a weekly basis for a period of about 5-6 years. We conjecture that this relative search volume is a decent proxy for how important a topic is to the general population.

We find that interest in certain topics does have significant and strong effects on parties' popularity. Most of these effects are in line with the parties' agendas. Furthermore, we find evidence that party preferences are relatively stable and that most shifts take place between the parties who are closest to each other in the political spectrum.

In the upcoming section we provide a brief overview over literature related to this topic which includes both research on party preferences and research using relative search volume data as its primary data source. In section 3 we describe the data used in this paper and present the model we estimate. In section 4 we estimate the relationship between interest in different topics and party preferences. We control for other drivers of party preferences, such as satisfaction with the current government's performance, and discuss two alternative regression approaches. Finally, we summarise the results found in this paper and discuss the advantages and limitations of using search volume

---

[1]See Downs (1957) and for example Rabinowitz and Macdonald (1989).
[2]For a relatively recent overview of publications on *issue salience* see Green and Hobolt (2008).

data in general.

2. Related Literature

Our analysis builds on two strands of literature. The first strand of literature related to our paper examines the *determinants of party preferences or vote choice.* A number of publications focus on *governments' economic performance.*[3] Other recent papers[4] emphasise the importance of *voters' individual characteristics.*

Furthermore, various investigations emphasise the importance of *issue salience.* For instance, Martinez and Gant (1990) show that despite a remarkable stability, non-negligible changes in partisanship occur, and that issue preferences contribute to these changes. Similarly, Green and Hobolt (2008) compare the impact of ideological position on vote choice to that of competence considerations with respect to issues rated as salient by voters. Carsey and Layman (2006) and Dancey and Goren (2010) argue that partisanship and issue attitudes cause changes in each other. However, Dancey and Goren (2010) show that this kind of "dynamic updating" only occurs as media coverage of different issues gives voters new cues on cleavages between parties. Malhotra and Krosnick (2007) also underline the importance of media coverage of different issues (and thus *information supply* on those issues). These studies are closely related to ours since we argue that it is *information demand* that matters for changes in party preferences.

Finally, several studies have previously analysed the behaviour of the German voters in particular, including Kirchgässner (1985) and Geys and Vermeir (2008) who also use German survey data on party preferences and Seithe (2011), who analyses the effect of voters' locus of control on party preferences based on the German Socio-Economic Panel.

The second strand of related literature consists of publications using *relative search volume data for quantitative analysis.* The first attempts to employ search volume data in econometric analysis are Ettredge et al. (2005) and Askitas and Zimmermann (2010), who use it to predict US and German *unemployment rates.* Another recent paper that employs the Google Trends service is Preis, Moat, and Stanley (2013), who show that a *stockmarket trading strategy* based on search volume data can yield significantly higher returns than a random strategy. They argue that 'notable drops in the financial market are preceded by periods of investor concern' (Preis, Moat, and Stanley 2013, p.1), which

---

[3]See Lewis-Beck (1988) or Sheafer (2008), and Başlevent et al. (2009) or Nezi (2012) on economic voting in the light of the European economic crisis.

[4]like Vecchione et al. (2011), Geishecker and Siedler (2012) or Walczak et al. (2012)

may be detected by carefully observing increases in search volume. Choi and Varian (2012) provide a general overview of literature related to Google Trends, provide examples for meaningful applications and argue in favour of *contemporaneous forecasting*, which is the process of exptrapolating a lagging variable using a predictive model based on search volume data.

3. Data Sources and Model

Our model is based on aggregate data for different political parties' popularity and for interest in selected topics accordings to Google trends.

We consider all political parties represented in the German parliament ('Bundestag') in the time between January 2007 and September 2013. Those are the center-right *CDU* and *CSU*, in the following text abbreviated as *CDU*, which led both government coalitions during the entire time frame, the *SPD* (Socio-Democratic Party), which was part of the government coalition until the 2009 elections, the business-friendly *FDP* (Free Democratic Party), which was the CDU's coalition party between 2009 and 2013, the environment-friendly center-left *Grüne* and the left-wing *Linke*.

3.1. Data Sources

3.1.1. Party Popularity ($y_{i,t}$)

In order to measure the parties' popularity at a given time we calculate the average of five different institutes' forecasts on vote intentions.[5] The institutes' forecasts are based on questionnaire data and generated using proprietary algorithms which are intended to correct and improve upon the gathered raw data. These forecasts are typically published on a weekly, biweekly or monthly basis. When gaps in an institute's forecast occured, a linear interpolation of the available neighboring data points was applied.

3.1.2. Interest (S_t)

We use relative search volume data provided by Google Trends as a proxy for people's interest in different issues. Google Trends is a service that allows researchers to download a time series for any search term's relative volume. The relative search volume expresses how often people searched for this specific search term relative to other terms. The

---

[5]Allensbach, Emnid, Forsa, Forschungsgruppe Wahlen and Infratest/Dimap

resulting time series is provided with a weekly resolution and scaled to the interval between 0 and 100 with 100 being assigned to the week with the highest relative search volume.

In order to assemble the search terms analysed in this paper we employed a multi-stage selection process. First, we read the election programmes and web pages of all major parties and identified both the most relevant topics as well as specific words which characterise those topics. Based on this we assembled a list of candidate search terms. This list was then reduced in order to eliminate redundant words as well words which may be ambiguous.[6] Finally, we eliminated words which did not generate sufficient search volume.[7] A complete glossary of the search terms used in this paper including a brief description of their political importance as well as complementary diagrams are provided in the appendix of this paper.

The approach of using Google Trends data has both drawbacks and advantages over traditional data sources like panel studies and questionnaires. First, it is apparent that the subset of people using Google to inform themselves about current issues is representative for neither the electorate nor the general population. Instead, it is probably biased towards younger and more tech-savvy people. Second, the event of a person searching for a topic merely tells us that she is interested in it. It does not by itself contain information on her actual opinion on the topic.

However, using data from the Google Trends service allows the retrospective analysis of trends in people's interest. Furthermore, it provides a vast amount of data at almost no cost. A more in-depth discussion of when Google Trends data is useful is provided in the conclusion of this paper.

3.1.3. Control Variables (X_t, h_t)

We control for the macroeconomic environment by including the *unemployment rate* and for the general political climate by using *satisfaction with the ruling government* as measured by the Infratest/Dimap institute. This time series started in November 2009, which is within our time frame. In order to account for this mismatch we generate a dummy variable which is set to 1 when the time series is not available, to ensure that the satisfaction coefficient is unbiased.

We also try to account for the fact that interest in politics all together may be weak

---

[6]Ambiguous words turned out to be a common problem. The search term "Euro", for instance, refers to both the football cup as well as the currency.

[7]While Google Trends does not provide information about the absolute number of queries on which the time series is based, a lack of volume can be easily identified as a gap in the data set.

when more pressing concerns are at hand. Therefore we include the Google Trends time series for *football* and *vacation.*

Finally, in order to control for medium-term persistent time trends not captured by the other variables in the model we include the average party popularity of the previous 18 weeks:

$$h(i,t) := \frac{1}{18} \sum_{j=1}^{18} y_{i,t-j} \tag{1}$$

3.2. Model

Our main equation based on the variables presented above is given by

$$y_{i,t} = \alpha_i + \beta_i S_t + \gamma_i X_t + \lambda_i h(i,t) + u_{i,t} \tag{2}$$

where t denotes the time (in weeks) and i refers to one of the k parties.[8]

3.2.1. Relevant Restrictions

In order for the model to yield reasonable results, a number of restrictions need to be fulfilled. These restrictions, which are based on Kirchgässner (1985), reflect the requirement that the popularity ratings of all parties need to add up to 100% regardless of the realisations of the explanatory variables:

$$\forall t \in T : \sum_{i=1}^{k} y_{i,t} = 100 \tag{3}$$

$$\forall j \in J_S : \sum_{i=1}^{k} \beta_{i,j} = 0 \quad \text{and} \quad \forall j \in J_X : \sum_{i=1}^{k} \gamma_{i,j} = 0 \tag{4}$$

$$\sum_{i=1}^{k} \alpha_i = 100 \quad \text{and} \quad \sum_{i=1}^{k} \lambda_i = 0 \tag{5}$$

$$\forall t \in T : \sum_{i=1}^{k} u_{i,t} = 0 \tag{6}$$

---

[8]In this paper that includes CDU/CSU, SPD, Grüne, Linke and FDP. When including the "other parties" we arrive at $k = 6$.

There are two ways of ensuring that these restrictions are fulfilled. The first approach is to estimate all k equations simultaneously using a regression method that explicitly accounts for these restrictions. The second approach is to estimate only $k - 1$ equations and define the kth party's popularity as the remainder:

$$y_{k,t} = 100 - \sum_{i=1}^{k-1} y_{i,t} \tag{7}$$

It can be shown, using straightforward algebra, that this approach of estimating the coefficients for party k ensures that none of the restrictions are violated. We therefore follow the second approach and drop the kth equation, which corresponds to "other parties".

4. Results

In this section we describe the methods used to estimate the model presented in the previous section and present our results.

There are two challenges in estimating the model presented above in an efficient way. First, there is a strong autocorrelation in the parties' popularities which biases the standard errors of the estimated coefficients. Second, there may be unobserved effects like major events that influence all parties' equations at once, leading to contemporaneous correlation between the different equation's error terms.

Our main approach is to estimate[9] all $k - 1$ equations separately using ordinary least squares (OLS) estimation and to address the autocorrelation problem by recalculating the estimated coefficients' standard errors using the more robust Newey/West heteroskedasticity- and autocorrelation-consistent (HAC) covariance matrices.[10]

While alternative regression approaches like seemingly-unrelated regressions (SUR) seem promising due to the prevalence of contemporaneous correlation, it would arguably not lead to a significant increase in efficiency in this case because the set of explanatory variables is largely identical for all $k - 1$ regressions.

So far we have assumed that party preferences are influenced by current issues and not vice versa. However, it is not entirely obvious that this is always the case since

---

[9]The following estimations and outputs are produced using R (R Core Team 2013) and the packages knitr (Xie 2013) and texreg (Leifeld 2013). For data processing and system estimation we use the packages zoo (Zeileis and Grothendieck 2005) and systemfit (Henningsen and Hamann 2007).

[10]Newey/west covariance matrix estimation uses Zeileis 2004.

some issues may become popular due to specific parties' rise in popularity. The rise of environment-friendly parties, for example, may make environment-related issues more popular. In order to examine this possible endogeneity problem we repeat the main regression using lagged versions of our explanatory variables instead. We find that our results are largely unchanged by this variation. The corresponding results are provided in section C of the appendix.

It is important to consider that the estimated coefficients that are reported in the upcoming regression tables are based on *relative* search volume. Comparing the coefficients of different search terms is therefore likely to result in misleading and erroneous conclusions. However, the coefficients can be used to interpret an individual search term's effect over time.

4.1. Search Terms

We focus on interpreting search terms which have a significant impact on party popularity in our main regressions.

Concerning *economic policy*, interest in search terms prominent in the CDU election programs like 'Ländlicher Raum' (rural areas) and 'Marktwirtschaft' (market economy) enhance the CDU's popularity while reducing the popularity of the SPD. However, the CDU does not benefit from interest in all matters of economy policy. Instead, interest in 'Subventionen' (state subsidies) and 'Mittelstand' (medium-sized businesses) seems to make the Grüne more popular, primarily at the expense of the SPD. Meanwhile, voters consent to the Linke is negatively affected by interest in the subject of 'Bürokratie' (bureaucracy). The introduction of the 'Schuldenbremse' (debt brake) seems to have affected mainly the SPD and the Grüne – the former gained popularity at the expense of the latter.

The Linke benefits from interest in *social policy* issues. Increases in the search volume data for 'Gerechtigkeit' (justice) have a significant positive impact on the popularity of the Linke, at the expense of the SPD. 'HartzIV' (unemployment benefits) also significantly and negatively affects the SPD, which introduced these labour market reforms between 2003 and 2005. The impact on the Grüne, junior coalition partner of the SPD during that time, is insignificant, however. In contrast, the SPD does benefit from interest in the terms 'Mindestlohn' (minimum wage), prominent in its party programs. The peaks in the search term 'Praxisgebühr' (fee for medical consultations) around the time of the abolishment of this fee, approved by all political parties represented in parliament, also benefited the SPD. Although responsible for its introduction in the first

8

	CDU.CSU	SPD	Grüne	FDP	Linke
(Intercept)	6.475 (3.426)*	5.531 (2.066)***	4.457 (1.904)**	1.578 (1.256)	0.210 (1.283)
Abtreibung	0.019 (0.007)***	0.018 (0.008)**	−0.026 (0.007)***	−0.012 (0.006)*	−0.005 (0.005)
Arbeitsbedingungen	0.005 (0.007)	−0.009 (0.006)	0.015 (0.009)*	−0.004 (0.004)	0.002 (0.005)
Atomausstieg	−0.006 (0.010)	−0.027 (0.008)***	0.046 (0.014)***	0.000 (0.005)	0.000 (0.004)
Ausländeranteil	−0.007 (0.005)	−0.008 (0.007)	0.005 (0.006)	0.008 (0.004)**	0.008 (0.004)**
Bundeswehr	0.017 (0.012)	0.016 (0.009)*	−0.042 (0.014)***	−0.014 (0.010)	0.005 (0.008)
Bürokratie	0.000 (0.003)	0.007 (0.004)*	0.003 (0.003)	0.003 (0.003)	−0.012 (0.004)***
Datenschutz	−0.023 (0.008)***	−0.003 (0.008)	0.006 (0.010)	0.004 (0.007)	0.007 (0.007)
DeutscheMark	−0.005 (0.006)	0.007 (0.004)*	−0.011 (0.006)*	0.005 (0.005)	0.003 (0.005)
Emanzipation	0.002 (0.006)	−0.007 (0.005)	0.010 (0.006)*	0.000 (0.004)	−0.003 (0.003)
Energiewende	−0.013 (0.006)**	0.016 (0.007)**	0.012 (0.006)**	0.001 (0.006)	0.001 (0.006)
Gentechnik	−0.016 (0.008)**	−0.001 (0.008)	0.021 (0.010)**	0.013 (0.008)*	−0.008 (0.006)
Gerechtigkeit	0.005 (0.009)	−0.019 (0.008)**	−0.002 (0.007)	−0.002 (0.004)	0.017 (0.006)***
HartzIV	−0.001 (0.007)	−0.017 (0.008)**	0.002 (0.006)	0.012 (0.010)	0.005 (0.006)
Innovationen	−0.014 (0.007)*	−0.002 (0.008)	−0.006 (0.007)	0.004 (0.004)	0.005 (0.005)
Kinderbetreuung	0.002 (0.006)	0.006 (0.008)	−0.006 (0.009)	−0.001 (0.004)	−0.001 (0.005)
Klimawandel	0.006 (0.009)	0.013 (0.006)**	−0.003 (0.008)	−0.019 (0.009)**	0.008 (0.006)
Korruption	0.012 (0.005)**	−0.008 (0.007)	0.002 (0.006)	0.000 (0.004)	0.000 (0.005)
Lebensqualität	−0.007 (0.008)	0.002 (0.006)	−0.016 (0.005)***	0.001 (0.007)	0.008 (0.005)*
LändlicherRaum	0.009 (0.004)**	−0.006 (0.003)*	0.005 (0.005)	−0.003 (0.003)	0.001 (0.002)
Marktwirtschaft	0.006 (0.006)	−0.001 (0.008)	−0.008 (0.010)	0.000 (0.004)	−0.003 (0.004)
Mietpreise	0.008 (0.005)*	−0.003 (0.005)	0.011 (0.006)*	−0.003 (0.003)	−0.004 (0.004)
Mindestlohn	0.000 (0.005)	0.009 (0.005)*	−0.006 (0.005)	−0.004 (0.003)	0.003 (0.004)
Mittelstand	−0.004 (0.007)	−0.015 (0.006)**	0.022 (0.010)**	−0.006 (0.004)	0.000 (0.005)
Neuverschuldung	−0.003 (0.004)	0.001 (0.004)	−0.004 (0.004)	−0.002 (0.002)	−0.002 (0.002)
Praxisgebühr	0.018 (0.008)**	0.009 (0.006)*	0.007 (0.005)	−0.014 (0.007)**	0.009 (0.005)*
Rassismus	−0.004 (0.006)	0.005 (0.006)	−0.003 (0.007)	0.005 (0.005)	−0.003 (0.004)
Schuldenbremse	−0.010 (0.007)	0.017 (0.007)**	−0.018 (0.007)**	0.006 (0.005)	−0.003 (0.003)
Solidarität	0.002 (0.006)	0.004 (0.005)	−0.006 (0.004)	0.006 (0.004)	−0.001 (0.003)
Subventionen	−0.004 (0.007)	−0.033 (0.011)***	0.026 (0.007)***	0.006 (0.008)	−0.001 (0.004)
Tierschutz	0.006 (0.008)	0.017 (0.008)**	−0.008 (0.008)	−0.003 (0.005)	−0.004 (0.005)
Urheberrecht	−0.001 (0.011)	0.003 (0.008)	−0.019 (0.007)**	0.002 (0.007)	−0.004 (0.006)
Verbraucherschutz	0.002 (0.008)	0.017 (0.010)	−0.005 (0.008)	0.000 (0.009)	−0.008 (0.007)
Verfassungsschutz	0.005 (0.006)	0.003 (0.008)	−0.017 (0.009)*	−0.003 (0.004)	0.014 (0.004)***
Wohlstand	−0.002 (0.004)	0.004 (0.003)	0.004 (0.005)	−0.004 (0.003)	0.000 (0.002)
Unemployment rate	0.087 (0.216)	0.064 (0.217)	0.161 (0.223)	−0.051 (0.155)	0.054 (0.123)
Satisfaction w. Govt.	1.538 (0.558)***	−1.659 (0.442)***	−1.536 (0.403)***	0.745 (0.279)***	−0.057 (0.227)
Satisf. w. Govt. (missing)	−0.761 (0.480)	−0.253 (0.371)	−1.022 (0.381)***	1.316 (0.389)***	0.881 (0.300)***
Fussball (soccer)	−0.020 (0.008)**	0.010 (0.010)	−0.018 (0.008)**	0.001 (0.008)	0.004 (0.007)
Urlaub (vacation)	−0.001 (0.007)	−0.009 (0.011)	0.002 (0.010)	0.005 (0.005)	0.004 (0.005)
h(i,t)	0.793 (0.081)***	0.717 (0.057)***	0.778 (0.050)***	0.844 (0.061)***	0.870 (0.042)***
R^2	0.900	0.859	0.948	0.976	0.938
Adj. R^2	0.886	0.839	0.941	0.973	0.929
Num. obs.	334	334	334	334	334

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Standard errors and p-values are corrected using the heteroskedasticity and autocorrelation consistent (HAC) Newey/West covariance matrix.

Table 1: Main Regression (OLS w. NW/SE)

place, large fractions of the SPD soon had turned against it. The abolishment of the fee was accompanied by a deal between the government coalition parties CDU and FDP, which apparently helped the CDU but hurt the FDP, although it had been in favour of abolition in the first place. Increasing interest in 'Mietpreise' (rents for housing) significantly benefits the Grüne, who made this subject a central part of their 2013 election campaign but also the CDU.

Popularity of the Grüne is also significantly affected by various search terms relating to rather *technical issues*. Interest in 'Atomausstieg' (phasing out nuclear energy), 'Energiewende' (energy turnaround) and 'Gentechnik' (genetic engineering) all have a positive impact on the Grüne, as would be expected from their party's program. Quantitatively, interest in phasing out nuclear energy seems to have a strong and significant effect for the Grüne: A 10 point increase in the search volume for 'Atomausstieg' (scaled between 0 and 100) is accompanied by a gain in popularity of almost 5 percentage points. This increase in popularity is reflected by a decrease in popularity of the SPD of around 3 percentage points. Meanwhile, interest in the energy turnaround benefited the SPD even more strongly than the Grüne, whereas interest in genetic engineering also benefited the FDP, which is more open-minded in this regard, at the expense of the CDU. This may reflect interest in the subject from both opponents and supporters.

On the contrary, interest in the 'Bundeswehr' (German armed forces) has a strong negative impact on the Grüne, traditionally a pacifist party. Both the CDU, responsible for the latest reforms concerning the armed forces, and the SPD benefit, whereas the FDP is negatively affected. However, only the impact on popularity of the Grüne is significant in all regressions. Interest in 'Datenschutz' (data protection) has a significant negative impact on the CDU, held responsible by voters for cooperating with the US National Security Agency at their expense. Interest in 'Urheberrecht' (copyright) and 'Verfassungsschutz' (the German national secret service) apparently hurt the Grüne.

In summary, changing priorities among voters, at least among internet users, do affect the relative popularity of the different political parties. A lot of the shifts occur between left-wing or right-wing parties, rather than from one extreme to the other. This observation confirms previous findings that political preferences are rather stable. This finding is also confirmed by the strong significantly positive impact of past party popularity.

4.2. Control Variables

Increases in the unemployment rate tend to increase accordance with the CDU, probably considered most apt to counter rising unemployment. A lack of interest as captured by

	CDU.CSU	SPD	Grüne	FDP	Linke
(Intercept)	8.086 (4.747)*	1.284 (2.346)	3.132 (4.374)	1.718 (1.277)	0.711 (1.825)
Unemployment rate	0.171 (0.200)	0.273 (0.355)	0.012 (0.318)	−0.215 (0.177)	−0.022 (0.170)
Satisfaction w. Govt.	1.740 (0.566)***	−0.463 (0.905)	−1.330 (0.737)*	0.398 (0.281)	−0.096 (0.200)
Satisf. w. Govt. (missing)	−0.366 (0.405)	−0.719 (0.527)	−1.081 (0.839)	1.314 (0.284)***	0.710 (0.415)*
Fussball (soccer)	−0.027 (0.008)***	0.023 (0.019)	−0.012 (0.011)	0.000 (0.012)	0.003 (0.011)
Urlaub (vacation)	0.001 (0.010)	−0.003 (0.019)	0.003 (0.016)	0.003 (0.005)	0.002 (0.008)
h(i,t)	0.762 (0.123)***	0.871 (0.132)***	0.803 (0.227)***	0.888 (0.054)***	0.887 (0.051)***
R^2	0.866	0.785	0.915	0.968	0.917
Adj. R^2	0.863	0.781	0.914	0.968	0.916
Num. obs.	334	334	334	334	334

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Standard errors and p-values are corrected using the heteroskedasticity and autocorrelation consistent (HAC) Newey/West covariance matrix.

Table 2: Regression without Search Terms (OLS w. NW/SE)

the search volume for 'Fußball' (football) negatively affects the CDU and the Grüne and does not have a significantly positive effect on any party. However, the same does not hold for the search volume for 'Urlaub' (vacation), which does not significantly affect any party's popularity.

Satisfaction with the government's performance strongly increases the popularity of the CDU, which has led both governments within the time frame in question, but also the popularity of the FDP, which was the CDU's coalition partner since the end of 2009, at the expense of the other parties' popularity.

4.3. Overall improvement of R^2

In order to resolve whether including the search terms leads to an increased quality of our model, we compare the adjusted R^2 values of estimations with and without them. Table 2 shows the results of an estimation without search terms. Comparing these results to those in table 1, we find that the adjusted R^2 does increase for all parties when adding search terms as explanatory variables. This increase is especially pronounced for the SPD, for which the value jumps from 0.781 to 0.839, which may indicate that this party's popularity was strongly influenced by hot topics in recent years.

5. Conclusion

In this paper we used search volume data provided by the Google Trends service to analyse the popularity of German parties.

Our first goal was to examine how well search volume data is suited for this kind of analysis. The obvious advantage of using data provided by the Google Trends service

is that it allows researchers to obtain *a vast amount of time series at virtually no cost.* Gathering comparable time series using questionnaire data requires either conducting a multitude of questionnaires, which would be very expensive, or using already existing questionnaires like the German Socio Economic Panel, which limits researchers to use questions which are already included. Another advantage is that Google Trends provides *retrospective data for arbitrary search terms.* This allowed us for example to measure people's interest in nuclear energy both after and *before* the Fukushima incident.

An obvious caveat of search volume data is the *lack of representativity* since Google's user base is likely skewed towards younger and more tech-savvy people. Furthermore, a lot of thought has to be given to the search terms. *Ambiguous search terms* can lead to misleading conclusions and deducing a user's opinion based on search terms alone can be problematic. Many search terms' time series are also rather *noisy* due to a lack of search volume. Finally, interpretation of *relative* search volume can be problematic as a decrease in one term's volume can be caused merely by the overwhelming prevalence of other topics. In summary, search term data has to be selected, used and interpreted carefully.

Our results regarding party preferences include both expected and surprising results. The CDU's popularity benefits from interest in *market economy and rural areas* which are traditional cornerstones of its conservative policy. The SPD profits from *environmental topics* but suffers from interest in the unpopular *Hartz IV-reforms* it implemented several years ago. The Grüne benefit strongly from the renewed interest in phasing out *nuclear energy* after the recent incidents. One surprising result is that interest in the issue of *social justice*, which has always been a cornerstone of the SPD, seems to hurt it in favour of the Linke, which seems to have acquired ownership of this issue. Another surprising result is that interest in the supposedly unpopular *Praxisgebühr* seems to hurt the FDP even though it negotiated hard in order to enforce its abolishment.

References

Askitas, N. and K. F. Zimmermann (2010). Google econometrics and unemployment forecasting. *SSRN 899*.

Başlevent, C., H. Kirmanoğlu, and B. Şenatalar (2009). Party preferences and economic voting in turkey (now that the crisis is over). *Party Politics 15*, 377–391.

Carsey, T. M. and G. C. Layman (2006). Changing sides or changing minds? party identification and policy preferences in the american electorate. *American Journal of Political Science 50*, 464–477.

Choi, H. and H. Varian (2012). Predicting the present with google trends. *Eonomic Record 88*, 2–9.

Dancey, L. and P. Goren (2010). Party identification, issue attitudes, and the dynamics of political debate. *American Journal of Political Science 54*, 686–699.

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy 65, no. 2*, 135–150.

Ettredge, M., J. Gerdes, and G. Karuga (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM 48*, 87–92.

Geishecker, I. and T. Siedler (2012). Job loss fears and (extremist) party identification: First evidence from panel data. SOEPpapers on multidisciplinary panel data research 511.

Geys, B. and J. Vermeir (2008). The political cost of taxation: new evidence from german popularity ratings. Discussion papers wzb no. sp ii 2008-06.

Green, J. and S. B. Hobolt (2008). Owning the issue agenda: Party strategies and vote choices in british elections. *Electoral Studies 27*, 460–476.

Henningsen, A. and J. D. Hamann (2007). systemfit: A package for estimating systems of simultaneous equations in r. *Journal of Statistical Software 23*(4), 1–40.

IfM (2013). *KMU-Definition des Institut für Mittelstandsforschung Bonn (IfM)*. http://www.ifm-bonn.org/mittelstandsdefinition/definition-kmu-des-ifm-bonn/.

Kirchgässner, G. (1985). Rationality, causality, and the relation between economic conditions and the popularity of parties. an empirical investigation for the federal republic of germany, 1971-1982. *European Economic Review 28*, 243–268.

Leifeld, P. (2013). *texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables.* University of Konstanz. R package version 1.28.

Lewis-Beck, M. S. (1988). *Economics and Elections: The Major Western Democracies.* University of Michigan Press.

Malhotra, N. and J. A. Krosnick (2007). Retrospective and prospective performance assessments during the 2004 election campaign: Tests of mediation and news media priming. *Political Behavior 29*, 249–278.

Martinez, M. D. and M. M. Gant (1990). Partisan issue preferences and partisan change. *Political Behavior 12*, 243–264.

Nezi, R. (2012). Economic voting under the economic crisis: Evidence from greece. *Electoral Studies 31*, 498–505.

Preis, T., H. S. Moat, and H. E. Stanley (2013). Quantifying trading behavior in financial markets using google trends. *Scientific Reports 3*, 1–6.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rabinowitz, G. and S. E. Macdonald (1989). A directional theory of issue voting. *American Political Science Review 83*, 93–121.

Seithe, M. (2011). *Essays in Behavioural Economics.* Universitäts- und Landesbibliothek Bonn.

Sheafer, T. (2008). The media and economic voting in israel. *International Journal of Public Opinion Research 20*, 33–51.

Vecchione, M., H. Schoen, J. Gonzalez Castro, J. Cieciuch, V. Pavlopoulos, and G. Caprara (2011). Personality correlates of party preference: The big five in five big european countries. *Personality and Individual Differences 51*, 737–742.

Walczak, A., W. van der Brug, and C. de Vries (2012). Long- and short-term determinants of party preferences: Inter-generational differences in western and east central europe. *Electoral Studies 31*, 273–284.

Xie, Y. (2013). *knitr: A general-purpose package for dynamic report generation in R.* R package version 1.2.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software 11(10)*, 1–17.

Zeileis, A. and G. Grothendieck (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software 14* (6), 1–27.

Appendix

A. Glossary of Search Terms Used
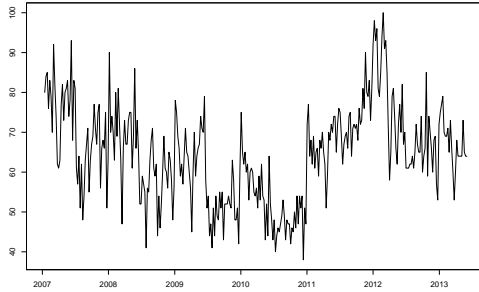
Abtreibung (Abortion)

Abortion has not been a prominent issue in social debates since the 1990s. Opposition to abortion is an issue, however, in extreme right wing's parties' platforms.
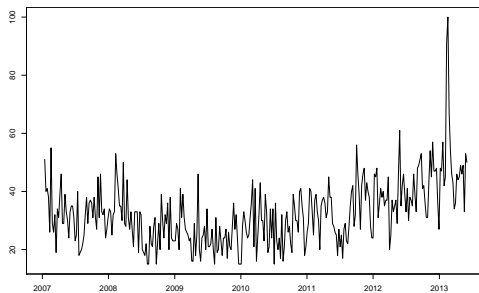
Arbeitsbedingungen (Working conditions)

Recent scandals concerning working conditions in Germany include the inappropriate surveillance of employees as well as draconian punishment for minor misdoings.
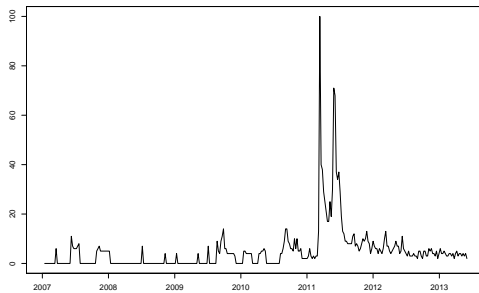
Atomausstieg (Phasing out nuclear energy)

Since the incident at the Fukushima Daiichi plant in early 2011 doubts about the safety of nuclear power have grown and many Germans now prefer to gradually close down existing nuclear power plants.
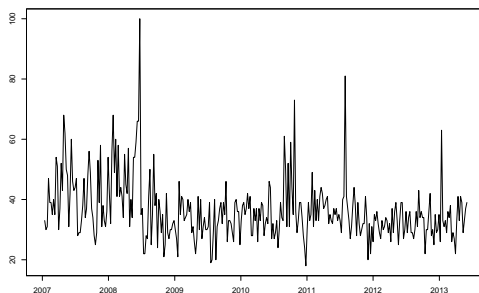
Ausländeranteil (Ratio of foreigners)

Foreigners living in Germany are often blamed for social problems including crime or unemployment by extreme right wing parties. The term is typically used in the context of city districts.
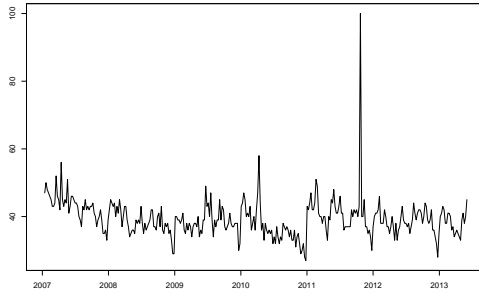
Bundeswehr (German armed forces)

The foundation of the German armed forces after World War II and its deployment abroad after the end of the Cold War were heavily debated between left and right wing parties. Recent reforms led to the suspension of the mandatory military draft since July 2011.
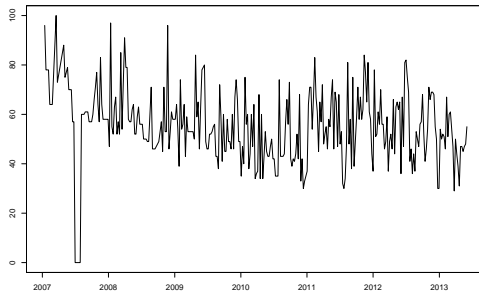
Bürokratie (Bureaucracy)

Reducing administrative burdens has been part of the German and European Union political agenda since the mid-2000s.
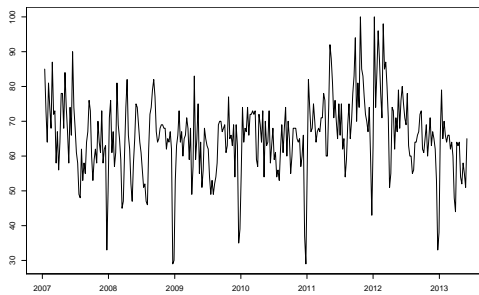
Datenschutz (Data privacy)

Data protection was widely discussed in the 2013 election campaigns after revelations on the US National Security Agency's activities in the European Union.
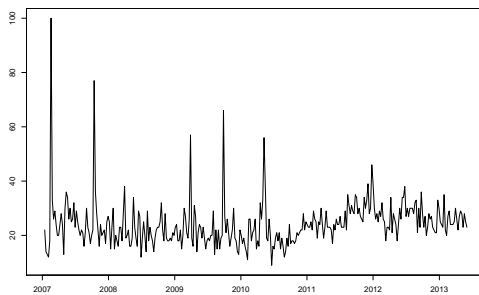
DeutscheMark (Prior German currency)

The majority of Germans prefers the Euro over the previous Deutsche Mark and is intent of sticking to it. However, the recent events regarding the Euro crisis have led to scepticism regarding the Euro.
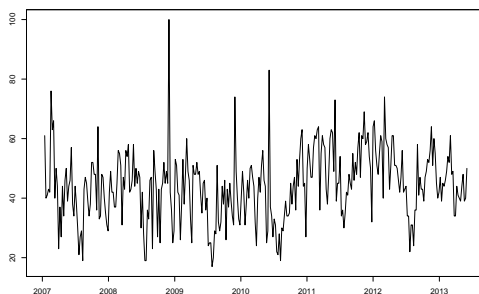
Emanzipation (Emancipation)

In the context of women's rights, currently debated issues in Germany are quotas for women in management, equal pay and child care availability.

17
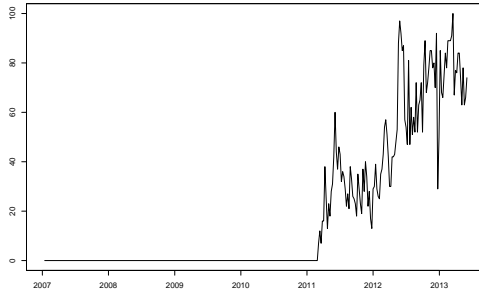
Energiewende (Energy turnaround)

The term 'Energiewende' describes the general process of phasing out energy generation from fossil fuel sources in favor of renewable energy. It is often associated with rising energy prices as this turnaround is strongly subsidised by energy consumers.
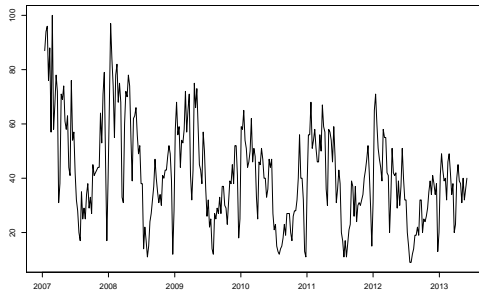
Gentechnik (Genetic engineering)

Genetically engineered crops are often regarded to be a potential environmental threat as the possible spread of engineered genes can be neither predicted nor prevented. Furthermore some people are anxious regarding some companies' attemps to monopolise specific plants, seeds and genetic sequences.
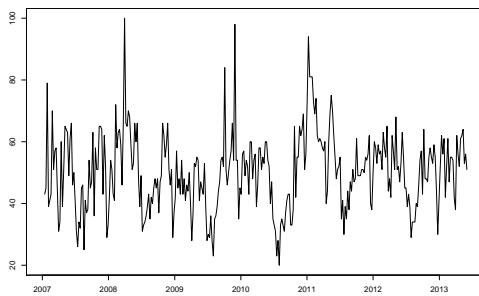
Gerechtigkeit (Justice)

The social debate on justice often involves 'just' salaries and old age pensions and recently focused on the relation between top managers' and workers' salaries.
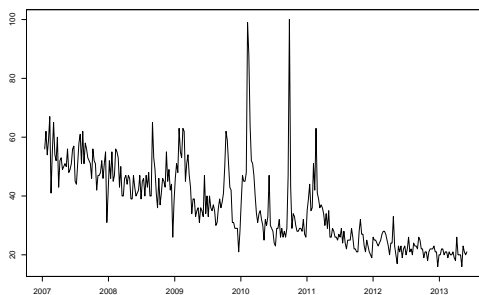
HartzIV (Unemployment benefits)

'Hartz IV' is the name associated with a set of labor market reforms introduced in 2003-2005. The reforms were intended to improve German job market competitiveness and include the reduction of unemployment benefits.
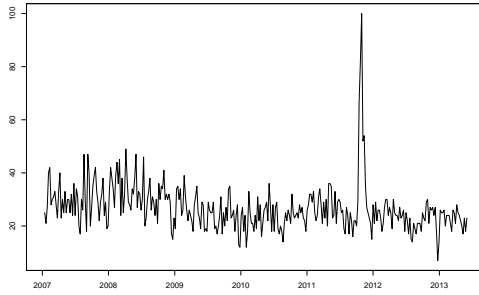
Innovationen (Innovation)

Innovation can refer to both technical and process innovations. Innovation is acknowledged to be a primary driver for economic growth in Germany by most parties. The Grüne often stresses the importance of environment-friendly process innovations like reducing energy consumption or waste in a production process.
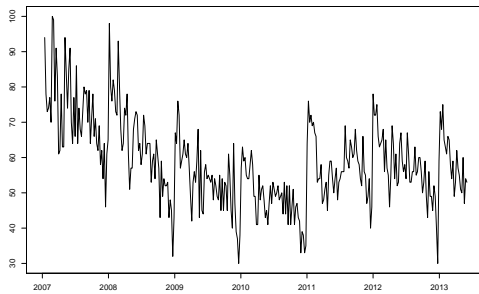

Kinderbetreuung (Child care)

Many young Germans name a lack of child care places as the main reason for not realizing their desire to have children. Whereas child care availability has traditionally been a prominent issue for left-wing parties, the 2009-2013 government, composed of the CDU and FDP, forced an accelerated expansion of child care availability.
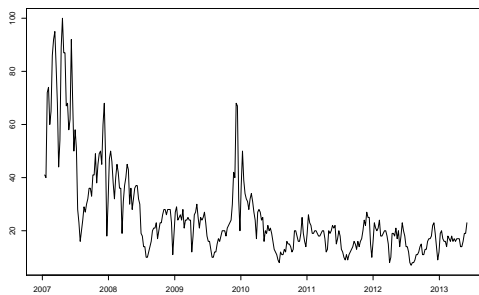
Klimawandel (Climate change)

The existance of the climate change phenomenon and the need to prevent further global warming is accepted as a reality by most Germans. However, parties differ in their willingness to sacrifice economic wealth for climate protection. Proponents of sacrificing wealth now suggest that investing in climate-friendlier technology will bring competitive advantages in the long run.
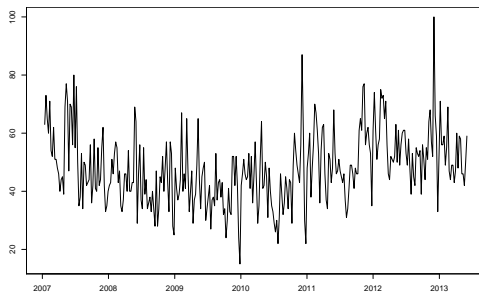

Korruption (Corruption)

While corruption is arguably not a strong issue in Germany, there are occasional cases in which officials or politicians are accused of misusing their powers in exchange for personal favours, including accusations that caused Christian Wulff, the German president prior to Joachim Gauck, to resign.

19
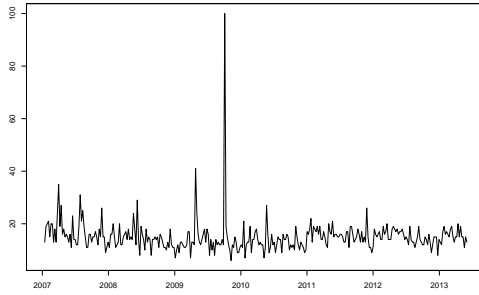
Lebensqualität (Life quality)

Recent political agendas include life quality as a necessary complement to prosperity or as the goal to pursue instead of prosperity.
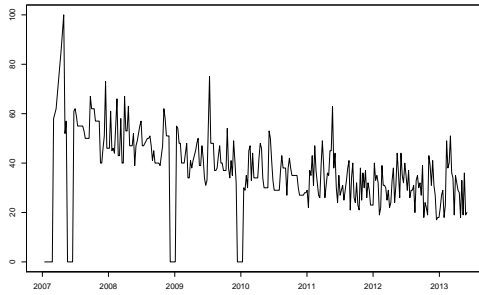

LändlicherRaum (Rural areas)

Rural areas, especially in Eastern Germany, have been facing the challenge of shrinking populations. There is an ongoing political debate on how to counter this challenge.
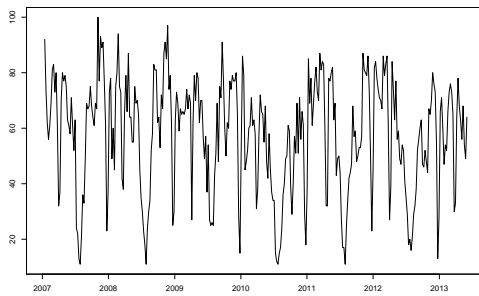

Marktwirtschaft (Market economy)

Whereas primarily the CDU claims that the principle of a market economy with social components has secured prosperity in Germany for decades, there has been some discussion on its justifications following the start of the financial crisis in 2007.
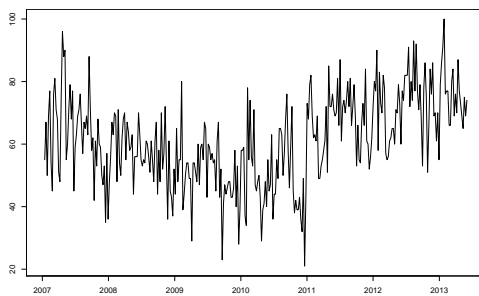

Mietpreise (Rents for housing)

Rising house prices and rents for housing in the light of the Euro crisis and the ongoing urbanisation process have given rise to a debate on how to provide adequate housing at affordable rents.
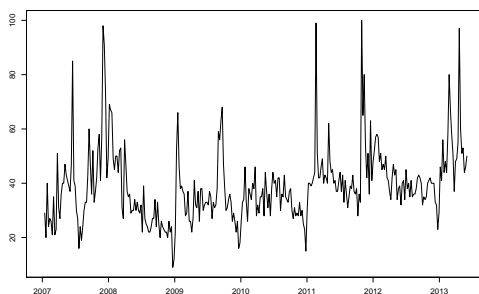

Mindestlohn (Minimum wage)

While there are several branch specific minimum wages in Germany, there is no general minimum wage. Low-income jobs are often subsidised by the state.

20
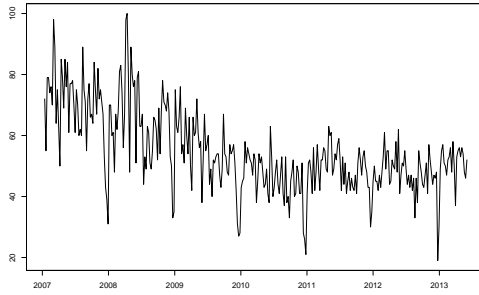
Mittelstand (Medium-sized businesses)
Medium-sized businesses are often regarded as
the backbone of the German economy. A busi-
ness is medium-sized if it has between 10 and 500
employees and up to 50 million euros in annual
revenue according to IfM 2013.
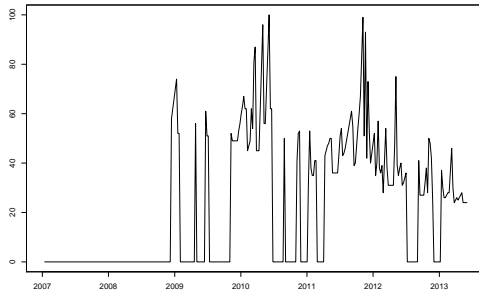
Neuverschuldung (New borrowing)
Many Germans regard high national debts in
some European Union countries as the main
cause of the Euro crisis. Consequently they de-
mand that both Germany and other Euro coun-
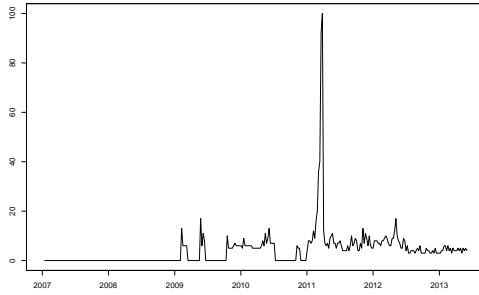tries stop 'spending more than they earn'.
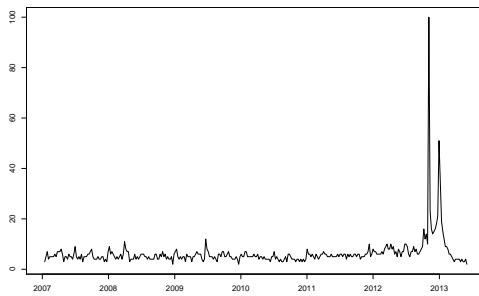
Praxisgebühr (Fee for medical consultations)
The 'Praxisgebühr' is a fee for medical consulta-
tions introduced in 2003 and in effect until the
end of 2012. Persons with compulsory health in-
surance were obliged to pay 10 euros upon their
first visit to a physician each quarter.
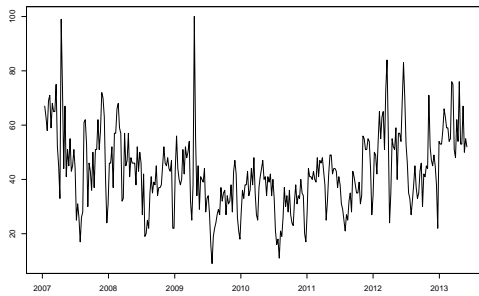
Rassismus (Racism)
How to counter racism is one of the main themes
of the political agenda of the 'Grüne'. After a
series of racist murders committed by an extreme
right-wing group became public in 2011, German
secret services were accused of biases which led
them to not investigate racist motives for the
individual crimes.

21

Schuldenbremse (Debt brake)

Many Germans regard high national debts in some countries as the main cause of the Euro crisis. Consequently they demand that both Germany and other Euro countries stop 'spending more than they earn'. Binding rules concerning the budget deficit have been in place in Germany since 2011. Transitional rules run out in 2020.
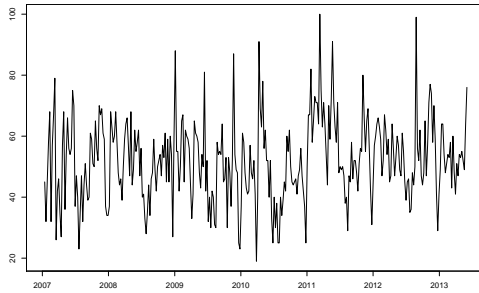
Solidarität (Solidarity)

Whereas solidarity and financial redistribution within the German society have traditionally been on the agenda of left wing parties, the term "solidarity" is also used frequently in the context of financial programs intended to stabilise the Euro countries in crisis. Furthermore, after German reunification a 'solidarity tax' was introduced to financially support the East German regions and is currently still in place.
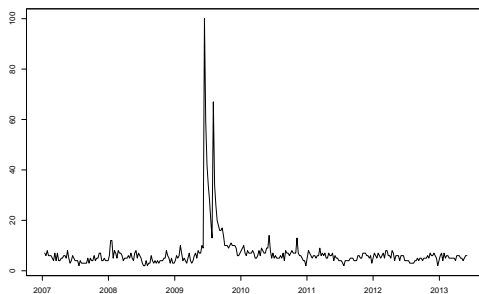
Subventionen (Subsidies)

State subsidiaries for agriculture and other business are criticised by many Germans to be both unfair and ineffective. Especially subsidies distributed via the European Union are often perceived to be arbitrary and economically unsound.
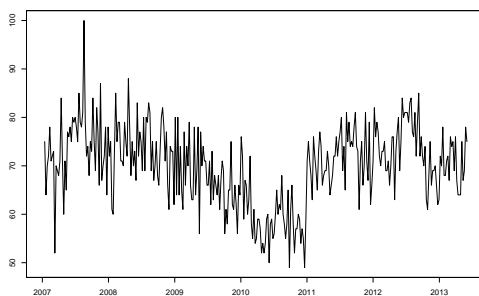
Tierschutz (Animal protection)

Animal protection has recently been discussed with respect to zoos and circuses, but also with respect to farm animals.
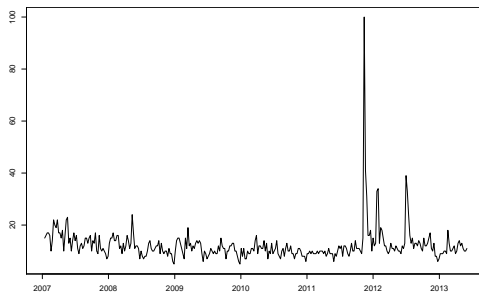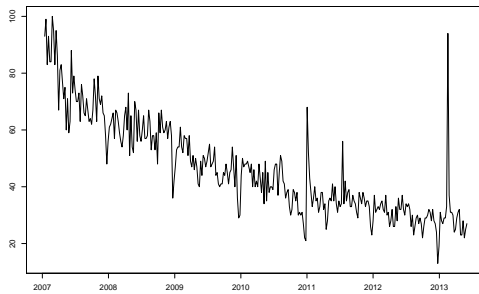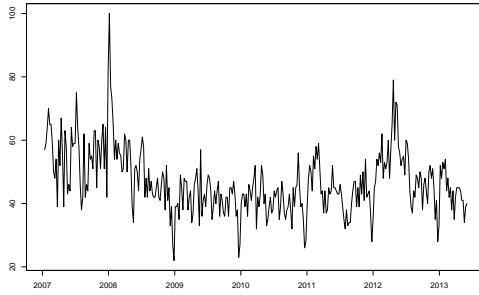
Urheberrecht (Copyright)

Several reforms concerning copyright legislation regarding digital media have been passed since 2007. A surrounding controversy is fuelled by illegal movie sharing sites on the one end of the spectrum and cases of children being fined four-digit sums for sharing a picture of their favourite singer on the other end.

Verbraucherschutz (Consumer protection)

In recent years, consumer protection has been an issue mainly with respect to food, bank services and contracts concluded on the internet. Another recent topic is the alleged 'planned obsolescence' of electronic products, i.e. products being intended to break in order for more products to be sold.

Verfassungsschutz (Protection of the constitution)

The 'Verfassungsschutz' is a German secret service recently criticised for not having been able to uncover a series of racist murders that had started in 2000 and for its alleged support of breaches of data protection by the US National Security Agency.

23

B. Search Term Selection Process

(Forthcoming)

C. Estimation with lagged explanatory variables

	CDU.CSU	SPD	Grüne	FDP	Linke
(Intercept)	4.904 (3.812)	5.470 (2.101)***	5.596 (2.062)***	1.367 (1.396)	−0.421 (1.239)
Abtreibung	0.015 (0.007)**	0.018 (0.008)**	−0.020 (0.009)**	−0.008 (0.006)	−0.008 (0.005)
Arbeitsbedingungen	0.006 (0.007)	−0.010 (0.005)*	0.013 (0.009)	−0.003 (0.005)	0.000 (0.005)
Atomausstieg	−0.008 (0.009)	−0.031 (0.009)***	0.049 (0.017)***	−0.002 (0.005)	0.001 (0.005)
Ausländeranteil	−0.008 (0.005)	−0.008 (0.007)	0.004 (0.006)	0.008 (0.003)**	0.010 (0.004)**
Bundeswehr	0.007 (0.012)	0.015 (0.010)	−0.032 (0.016)*	−0.010 (0.010)	0.012 (0.008)
Bürokratie	−0.002 (0.003)	0.010 (0.004)***	0.000 (0.003)	0.005 (0.002)**	−0.011 (0.003)***
Datenschutz	−0.014 (0.007)**	−0.005 (0.009)	0.001 (0.010)	0.001 (0.007)	0.006 (0.008)
DeutscheMark	−0.006 (0.006)	0.010 (0.005)*	−0.014 (0.006)**	0.002 (0.004)	0.005 (0.005)
Emanzipation	0.006 (0.005)	−0.007 (0.005)	0.009 (0.007)	−0.002 (0.004)	−0.002 (0.003)
Energiewende	−0.009 (0.007)	0.018 (0.007)**	0.012 (0.008)	−0.001 (0.005)	0.003 (0.004)
Gentechnik	−0.020 (0.008)**	−0.002 (0.008)	0.022 (0.011)**	0.013 (0.008)*	−0.004 (0.006)
Gerechtigkeit	0.009 (0.008)	−0.017 (0.007)**	−0.004 (0.008)	0.000 (0.004)	0.013 (0.005)**
HartzIV	0.001 (0.006)	−0.017 (0.010)*	0.004 (0.009)	0.012 (0.008)	0.004 (0.004)
Innovationen	0.000 (0.007)	−0.004 (0.008)	−0.002 (0.008)	0.002 (0.005)	0.000 (0.005)
Kinderbetreuung	0.001 (0.007)	0.002 (0.008)	−0.002 (0.010)	−0.003 (0.005)	−0.002 (0.005)
Klimawandel	0.012 (0.010)	0.014 (0.006)**	−0.004 (0.008)	−0.022 (0.009)**	0.008 (0.006)
Korruption	0.011 (0.005)**	−0.005 (0.007)	0.000 (0.006)	−0.001 (0.004)	0.006 (0.004)
Lebensqualität	−0.007 (0.008)	−0.001 (0.007)	−0.013 (0.006)**	0.004 (0.008)	0.008 (0.004)**
LändlicherRaum	0.010 (0.004)**	−0.004 (0.004)	0.007 (0.004)	−0.003 (0.003)	0.000 (0.003)
Marktwirtschaft	0.003 (0.006)	0.001 (0.008)	−0.008 (0.011)	0.003 (0.005)	−0.005 (0.004)
Mietpreise	0.007 (0.005)	−0.006 (0.005)	0.011 (0.005)**	−0.002 (0.003)	−0.002 (0.003)
Mindestlohn	−0.006 (0.006)	0.008 (0.006)	−0.007 (0.006)	−0.002 (0.003)	0.008 (0.005)
Mittelstand	−0.009 (0.007)	−0.011 (0.007)	0.017 (0.011)	−0.004 (0.004)	0.002 (0.004)
Neuverschuldung	−0.004 (0.003)	0.003 (0.004)	−0.006 (0.005)	−0.001 (0.003)	−0.001 (0.002)
Praxisgebühr	0.019 (0.007)***	0.004 (0.006)	0.006 (0.006)	−0.012 (0.007)	0.006 (0.005)
Rassismus	0.000 (0.006)	0.003 (0.006)	−0.001 (0.007)	−0.001 (0.004)	−0.004 (0.004)
Schuldenbremse	−0.016 (0.006)***	0.011 (0.008)	0.005 (0.008)	0.002 (0.005)	−0.007 (0.003)**
Solidarität	0.001 (0.005)	0.005 (0.005)	−0.006 (0.005)	0.006 (0.004)	−0.001 (0.003)
Subventionen	0.000 (0.007)	−0.036 (0.011)***	0.024 (0.009)***	0.003 (0.010)	0.003 (0.005)
Tierschutz	0.011 (0.006)*	0.013 (0.008)*	−0.011 (0.009)	−0.005 (0.005)	−0.003 (0.005)
Urheberrecht	−0.005 (0.010)	0.007 (0.008)	−0.021 (0.007)***	0.002 (0.009)	−0.002 (0.006)
Verbraucherschutz	0.000 (0.007)	0.013 (0.010)	−0.003 (0.009)	0.002 (0.010)	−0.008 (0.007)
Verfassungsschutz	0.008 (0.006)	−0.003 (0.006)	−0.014 (0.008)*	−0.003 (0.004)	0.013 (0.007)**
Wohlstand	0.000 (0.004)	0.003 (0.003)	0.004 (0.006)	−0.007 (0.003)**	0.001 (0.002)
Unemployment rate	0.157 (0.248)	0.067 (0.222)	0.088 (0.223)	−0.017 (0.174)	0.010 (0.137)
Satisfaction w. Govt.	1.323 (0.518)**	−1.729 (0.465)***	−1.546 (0.469)***	0.812 (0.270)***	−0.125 (0.221)
Satisf. w. Govt. (missing)	−0.717 (0.486)	−0.159 (0.355)	−1.119 (0.408)***	1.263 (0.393)***	0.894 (0.313)***
Fussball (soccer)	−0.023 (0.007)***	0.009 (0.011)	−0.017 (0.009)*	0.004 (0.011)	0.002 (0.008)
Urlaub (vacation)	0.003 (0.007)	−0.006 (0.009)	−0.001 (0.010)	0.006 (0.004)	0.003 (0.005)
h(i,t)	0.814 (0.088)***	0.723 (0.065)***	0.750 (0.062)***	0.833 (0.064)***	0.908 (0.041)***
R^2	0.902	0.856	0.948	0.976	0.939
Adj. R^2	0.889	0.837	0.941	0.973	0.931
Num. obs.	333	333	333	333	333

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Standard errors and p-values are corrected using heteroskedasticity and an autocorrelation consistent (HAC) covariance matrix (Newey/West). The explanatory variables are lagged by one period.

Table 3: Estimation with Newey West standard errors and lagged variables.