Using Smartphone Polling to Forecast the German Federal Election 2017

Comparing the performance of digital and traditional polls

Moritz Hemmerlein & Alexander Sacharow

08 May, 2017

Abstract

Using online surveys to forecast election outcomes imposes severe challenges to pollsters. Non-representative samples and likely-voter bias skew gathered information and require adequate statistical adjustment. This paper compares different weighting methods to obtain a forecast from raw data, collected through smartphone polling. To assess the performance of our results, we benchmark them to leadings polling institute's rolling average.

Contents

1   Introduction

The digitalisation is challenging the way polling was done for many decades. Calling people on their landline phones has become difficult as response rates dropped and number of households equipped with landline phones is decreasing (Skibba 2016). In response, pollsters have resorted to other polling-methods ranging from face-to-face interviews to mobile phone calling. However, these methods either face similar difficulties to ensure representativeness or are too expensive for regular polling. In order to tackle these obstacles, pollsters are increasingly using online polls, which are cheap and fast, but can be highly non-representative (e.g. Wang et al. 2015).

The eventual aim of pollsters in online as well as traditional polling is to collect sample data that reflects the view of a population of interest. The major difference between both methods is thatfor various reasons online polls cannot ensure representativeness before the actual poll takes place. For instance, respondents of online surveys are more likely to be from certain demographic groups or share a particular political background depending on the website or app where the survey is conducted. However, such non-representative polls can be statistically adjusted to match the demographic composition of the population.

Additionally, online election polls, like traditional ones, face another problem. Election forecasters are naturally not only interested in the population as such, but in the population of actual voters. By the time a poll is made representative in demographic terms, it is still in question whether it reflects the group of people who actually cast their ballots. This, however, is crucial in order to make an accurate prediction. Traditional polling tries to account for this using likely voter models and could perform fairly well (Gallup 2010, Keeter, Igielnik, and Weisel (2016)). Online surveys will also have to be adjusted to actual voting population in order to provide accurate predictions.

In this paper, we employed different weighting strategies to an online poll, exploiting respondents usage of smartphone apps. We evaluate our methods against the rolling average of polls from leading German polling institutes. While the raw data of the online survey performed fairly poor compared to our benchmark, we show that with proper weighting, we could obtain results that are close to the major polls. Moreover, we compared a rather direct and a indirect method. First, we used information on self-reported vote decision at the German federal election 2013 and data from official exit polls to construct our weights. In a second, more indirect approach, we constructed weights to adjust our demographic clusters to the German census data and subsequently identified so called likely voters to adjust our raw data. As our results show, the direct method performed considerably better than the indirect weighting through census data.

The structure of the paper is as follows: Section two will survey the literature on non-representative polls and presents approaches to employ such data to forecast elections. Subsequently, in chapter three, we present our data and discuss idiosyncratic problems and potential sources of bias. Section four shows the methods we applied to our data. Chapter five is presenting and dicussing our

results and their performance against the benchmark data. Chapter six concludes and gives an outlook on practical obstacles and methodological issues that our approach suffers from.

2   Related Literature

3   Data Sources and Quality

3.1   Data sources and potential biases

3.1.1   The Dalia Research' Europulse

In this paper, we are using data from Dalia Research, an online polling firm which is conducting market and opinion research exclusively through smartphones. To ensure to collect data from a broad variety of target populations, Dalia is using a diverse set of app and website categories such as sports, news, entertainment or games. To control that participants answer the survey seriously, an algorithm analyses the consistency and the response behaviour and computes a "trust score" to every respondent. Dalia praises its methodology as distinctively accounting for potential biases such as interviewer effect, social desirability bias or interviewer data entry errors. (Dalia Research 2016)

Our forecasting project utilizes data of Dalia Research's Europulse Survey which is conducted quarterly in all EU countries. The survey consists of seven waves, but for this project we only use two waves of the survey from December 2016 and March 2017. The first wave is freely available on Kaggle, the second wave was provided to us directly by Dalia Research.[1] Each wave consists of about 11000 individuals, of which roughly 1900 were from Germany which is the fraction of respondents we will focus on in the following analysis. The data is already pre-stratified by Dalia Research based on micro census data for age, gender.[2]

The Europulse data is not particularly collected for election forecasting purposes but contains data on a variety of questions such as online behaviour, media consumption and personal views on political and societal development in the European Union and the respondent's country of origin. Moreover, the survey contains information on the respondent's personal background, demographic data and his or her financial situation. These data can be utilized in order to improve the representativeness of the survey through weighting. This will be explained more detailed in the methodology section.

---

[1]The data for the first wave was collected between 5th and 15th of December 2016. The second wave was conducted between 13th to 27th of March 2017.

[2]Dalia Research also provides weights from the micro census in their data. However, since we do not know the methodology used to obtain these weights we will employ our own weighting algorithm for Census adjustment.

The election-related variables collected in the Europulse survey are similar to traditional vote intention polling questions. First of all, the respondents are asked if and for which party they will vote in the upcoming election and for which party they voted in the previous election. Moreover, they are asked to rank political parties and describe the degree of certainty to cast a ballot for a particular party. Important for our analysis is a variable containing self-reported voting decision at the past German federal election, since we exploit this data compute survey weights that adjust the data to demographic composition of the past election in 2013.

3.1.2   Raw forecasts and data quality

As with all surveys, the methodology of Dalia Research has several sources of potential biases. In the case of the Europulse survey they are primarily from flawed measurement and representation of the surveyed population, as listed in figure 1 (Groves et al. 2009).
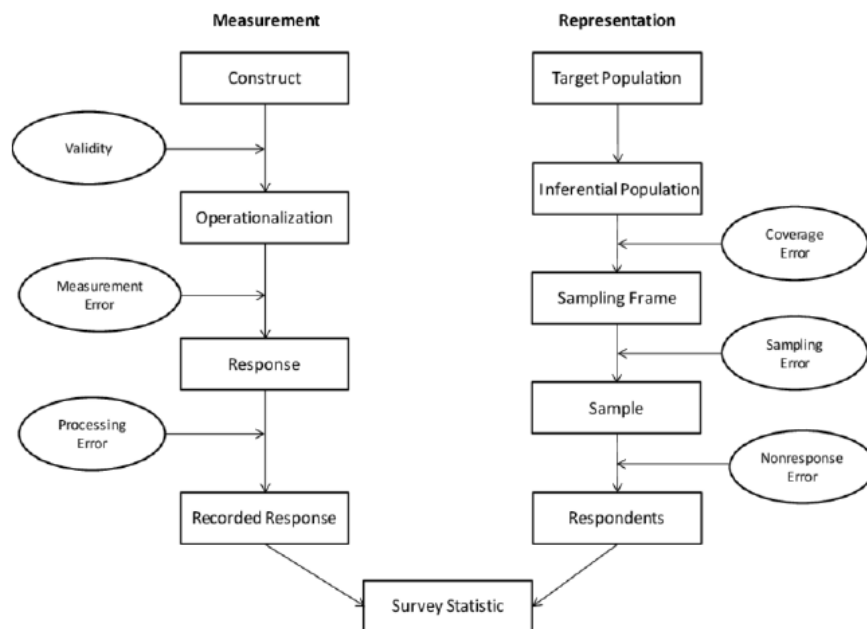

Figure 1: Potential sources of survey error

With regard to measuring vote intention for a particular party, the Europulse survey uses a similar approach as traditional surveys: it asks respondents directly which party they intent to vote for at the upcoming election. Whether such questions measure correctly the actual voting behaviour at the election day is questionable but the approach does not differ significantly from traditional polling methods.[3] However, online surveys, such as Europulse, can reasonably claim to avoid

---

[3]The extact question used for instance by *infratest dimap*, a major german polling firm, is: "What party would you vote for if there would be a Federal election next Sunday." (dimap 2017) This wording is slightly different than in the Dalia questionaire but evaluating the impact of the formulation of questions shall not be content of this study.

(a) Internet usage by age
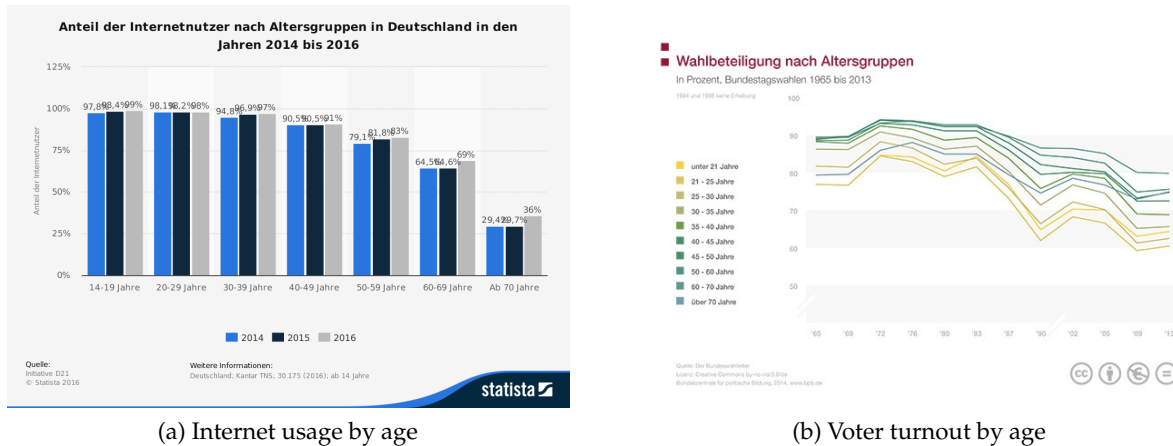

(b) Voter turnout by age

Figure 2: Internet usage and turnout by age

some sources of measurement error such as social desirability bias or interviewer bias. Since such surveys are often anonymous, the social pressure on the respondent is presumably negligible, and interaction with the interviewer does not bias the response.

Regarding representativeness, the Europulse survey has some limitations in comparison to face-to-face or RDD interviews. First of all, Europulse' framework does not select participants at random but offers visitors of certain websites or app-users the opportunity to participate. Hence, this approach carries the risk of self-selection. Moreover, representativeness would require that the Europulse survey in principle should be available to the entire population. However, not everyone is using smartphones and even if they are, they might not use the applications Dalia targets for its surveys. The users Dalia Research actually tends to reach are likely to be much younger and technology oriented than the general population would be (see figure). This is in particular a problem for election polling, as older voters are systematically underrepresented in such online methodologies while turnout amongst the elderly tends to be higher compared to the young population.

Dalia Research tries to account for these problems. First, as we have mentioned, they try to increase the diversity of their respondents by presenting their surveys on various apps and platforms targeting different user groups. Second, they pre- and post-stratify their data. Pre-stratification is done on the basis of age and gender, using self-reported demographic information. For each age and gender strata they target a certain number of respondents so that their sample resembles roughly the German population. In a second step, they use data from the German Census and compute weights for cluster of age, gender, education and whether the respondent lives in an urban or rural area. The methodology how Dalia obtained these weights remained not transparent to us. For that reason we used our own approach for post-stratification that specifically addressed our purpose of election forecasting. Finally, Dalia tries to avoid selection bias since they to not inform respondents that they are answering an election related survey before they actually start it and in

general Dalia Research has high completion rates, hence participants are only rarely dropping out after they started a survey.

Despite these efforts, it is questionable whether the data can be regarded as representative. Below we present a simple frequency table of the seld-reported past election turnout as well as the intended vote at the next election for the most recent wave of Europulse. The tables show that the results are significantly diverging from other forecasts (see www.wahlumfragen.org).
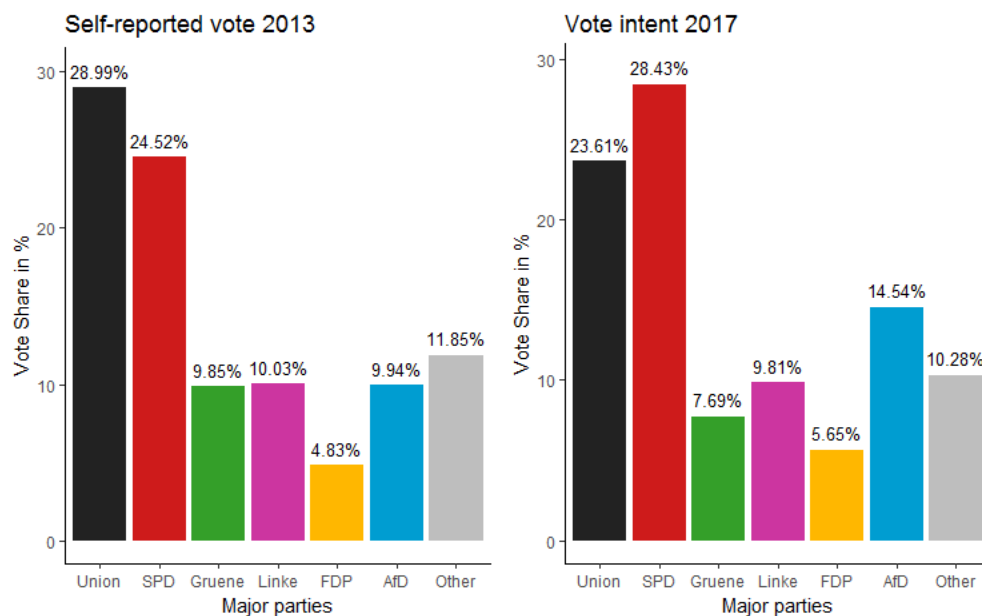

Figure 3: Self-reported vote 2013 and 2017 (Europulse March 2017)

The large divergence from can be due to a variety of flaws that bias Europulse. First of all, even if the sample represents the German population in terms of age groups and gender, it can be question whether the old female or male users are truly representative of their age group. Moreover, not revealing the survey content might sort out political motivated respondents, but interest in political matters is likely correlated to general willingness to respond to surveys. Hence, in order to utilize the data to election forecasting further weighting and post-stratification will be necessary.

3.2   Official election statistics and Census

The data used for post-stratification and weighting come from several sources. Our primary secondary data is obtained from the German official election statistics. (Der Bundeswahlleiter 2017) It contains combined information of turnout and voting decision by gender and age. Age is clustered in groups of 18-25, 26-35, 36-45, 46-60 and 60 plus. This data is the closest estimate of the actual voting population available and is a common measure to account for likely voter bias.

To adjust our sample to demographic composition of the German population we use data of the

German Zensus 2011. The data is freely available and we obtained combined frequencies foras age, gender, education and confession. The Census claims to reflect the actual demographic distribution of the German population and hence is suitable in order to post-stratify our sample to demographic criteria. Both, election statistics and Census is the basis on which we adjust our raw data to obtain representative results.

Finally, in order to set our forecasts into context we will use polling data from other institutes. The daily newspaper *Süddeutsche Zeitung* has recently started to modernize their approach on polling. Leaning on the famous approach of the American statistician Nate Silver [SOURCE POLLING THE POLLS] they computed a rolling average of all major German polling institutes forecasts. [SOURCE] To meet full transparency criteria they published their data (from wahlrecht.de) and their R-code on GitHub. We utilized their efforts and use the rolling average as benchmark to assess how our estimates perform.

Finally, as well as the Dalia data, the weighting data suffers from a variety of problems. Most important, the exit poll data we use to compute our weights is not clustered for voters of the *Alternative für Deutschland* (AfD). In order to compute the vote share among groups of age and gender we took the share of the total election results that the AfD got among all parties that are summarized among *Others*. The weak spot of this approach is that it assumes a equal distribution of AfD voters across all groups of age and gender. While this assumption is very unlikely it is the best approximation we can get with the data at hand.

Moreover, the Census as well as the election statistics do not provide combined frequencies of more than three dimensions of demographic variables. This makes it impossible to weight the data on a large set of criteria. Yet, since our Dalia dataset is to small to compute more than three-dimensional cluster anyways that has no severe impact on our methdology.

4   Methodology

5   Results

6   Conclusion

References

Dalia Research. 2016. "Dalia Research Methodology." https://daliaresearch.com/wp-content/uploads/2016/08/Methodology-PDF-1.pdf.

Der Bundeswahlleiter. 2017. "Ergebnisse Der Representativen Wahlstatistik." https://www.

bundeswahlleiter.de/bundestagswahlen/2013/ergebnisse/repraesentative-wahlstatistik.html.

dimap. 2017. "Allgemeine Informationen Zur Sonntagsfrage." infratest dimap. https://www.infratest-dimap.de/service/faqs/faqs-zur-sonntagsfrage/.

Gallup. 2010. "Understanding Gallup's Likely Voter Models." http://www.gallup.com/poll/143372/understanding-gallup-likely-voter-models.aspx?version=print.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd ed. New Jersey: John Wiley & Sons Inc.

Keeter, Scott, Ruth Igielnik, and Rachel Weisel. 2016. "Can Likely Voter Models Be Improved? Evidence from the 2014 U.S. House Elections." Pew Research Center.

Skibba, Ramin. 2016. "The Polling Crisis: How to Tell What People Really Think." *Nature* 538 (7625): 304–6. doi:10.1038/538304a.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91. doi:10.1016/j.ijforecast.2014.06.001.