Non-Representative Polling
Mobile User Polling Data for the German Federal Election 2017

Election Forecasting Project

Moritz Hemmerlein & Alexander Sacharow

11 April, 2017

Abstract

Using online surveys to forecast election outcomes imposes severe challenges to pollsters. Non-representativ samples and likely-voter bias skew gathered information and require adequate adjustment. This paper sets out to discuss different approaches to adjust a non-representative online polls on the upcoming german federal election and how they can be applied to a poll conducted with mobile phone app users.

Contents

1 Introduction

The digitialization is challenging the way polling was done for many decades. Calling people on their landline phones has become difficult as response rates dropped and households equiped with landline phones are getting less and less (Skibba 2016). In response, polling institutions have resorted to other methods for polling 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 have began to experiment with non-representative polling methods (W. Wang et al. 2015).

The eventual aim of pollsters in non-representative 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 that with non-representative polls representativeness cannot be ensured before the actual poll takes place. For example, in the case of online surveys, respondents are more likely to be from certain age groups or with a particular political background depending on the website where the survey is conducted. The non-representative poll can, however, be statistically adjusted (post-stratification) to match the demographic composition of the population.

Additionaly, non-representative, like traditional election polls, face a another problem. Pollsters 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 cast their ballots. This is, however, crucial in order to make an accurate prediction. Traditional polling tries to account for this using likely voter models and could perform fairly well. [Reference: Gallup paper, likely voter method, post-stratification using exit polls] Non-representative surveys will also have to be adjusted to actual voting population in order to provide accurate predictions.

In the paper at hand we want to explore and analyse how different approaches to adjust non-representative polls perform. Our first approach will be a two step procedure where the polling data is first made representative of the population and then likely voter methods are used to resemble the probable voters population. Our second approach attempts to combine both steps into one by post-stratifying the non-representative poll with exit polls and voting statistics from previous elections. For this we work with individual level data from mobile-phone app users who were surveyed on their vote intention in the German federal election 2017. As the ultimate election will only be after the end of research, we use other forecasts of the 2017 federal election as a benchmark.

The structure of the paper is as follows: Chapter two will survey the literature on non-reprentative polls and approaches to employ such data to forecast elections. Subsequently, we present our data and discuss possible problems with this data set. In chapter four we present two approaches we want to apply to adjust our data at hand.

2 Related Literature

Traditional polling and in particular election polling has relied heavily on telephone surveys for the last decades. To ensure representativeness the standard was randomized digit dialing (RDD). The selection of random respondents was intended to eliminate the sample bias of the survey. However, for several reasons this approach has become unreliables. First, response rates have declined heavily (Keeter et al. 2006; Holbrook, Krosnick, and Pfent 2007). ("Assessing the Representativeness of Public Opinion Surveys" 2012) reported that in the U.S. the response rate was down to 9% in 2012, compared to 36% in 1997. This is mainly a result of technical changes like the possible identification of caller. This induces a non-response bias to polls which cannot be handled with classical approaches anymore. Second, more and more people don't get landlines telephones after moving to new places or just give them up as mobile phone and other means of communication have increasingly become popular. This induces a sample bias which the RDD approach intended to eliminate. How problematic this can be was famously illustrated by the Literature Digest poll in 1936, which failed to realized its biased sample (Squire 1988). But the problem still exists, as non-representatives was for example a major problem when forecast the UK General elections in 2015 (Mellon and Prosser 2015).

Can non-representative polls fix this? A study by Yeager et al. (2011) is rather pessimistic. They have compared telephone and online polls and argued that despite the problems of phone surveys they are still more accurate then online polls. However, their criticism mainly highlights the implication of non-probabilities samples which is a problem for many online survey. If the problem non-probability samples is addressed in online survey, they can perform as accurate as phone surveys.

W. Wang et al. (2015) in contrast are much more optimist about the possiblities of non-representative polling. They used polling results from XBox users which were highly unrepresentative of the population to forecast the 2012 U.S. presidential elections. By employing a sophisticated multi-stage approach to stratify and calibrate the data they were able to generate accurate forecasts of the elections. Hence, methods for stratifying and calibrating the forecasts are the key to meaningful forecasts based on non-representative polling.

1. post stratification
2. raking
3. multiple level regression + post stratification
4. random sampling?

But even if the problems of non-representative

" 'Likely voter' modelling is notoriously the secret-sauce aspect of polling," says Kennedy Courtney Kennedy, director of survey research at the Pew Research Center in Washington DC (from Skibba (2016))

Typical biases in polling (most important: non-response bias; sample-selection bias) Which biases do we expect in the data

3 Data and Potential Biases

3.1 Europuls Survey

In this paper we are using data from Dalia Research, an online polling firm who is conducting market and opinion research through smartphones exclusively. 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 how serious participants answer the survey an algorithms 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. [Source Dalia Methodology PDF]

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. Each wave consists of about 11000 individuals, of which roughly 1900 were from Germany. The data is already pre-stratified by Dalia Research based on micro zensus data for age and gender.

The Europuls 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 country of origin. Moreover, the survey contains information on the respondents 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 Chapter four.

The important variables collected in Europuls in order to utilize the data for election forecasting are likewise to traditional polling questions. First of all 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 on the level of agreement with several political parties and the degree of certainty to cast a ballot, overall and for a particular party.

3.2 Potential Sources of Bias

As with all surveys there are several sources of potential bias included in the methodology used by Dalia Research. Several of such issues that come from flawed measurement or representation of the surveyed population are listed in the figure below.

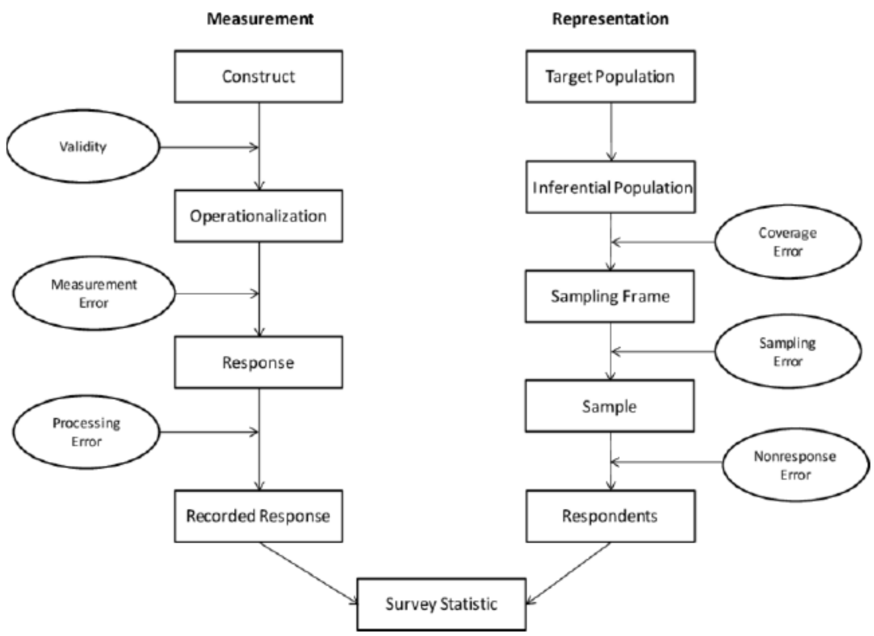

Figure 1: Potential sources of survey error [@Groves.2009]

With regard to measuring voters intention to vote for a particular party Europuls uses a similar approach as traditional surveys. As such it asks directly which party a respondent intents to vote for at the upcoming election. Whether such questions measure correctly the actual voting behaviour at the election day is questionable but does not differ from other polling methods of validity. Moreover, online surveys such as Europuls can reasonably claim to avoid other sources of measurement error such as social desirability bias or interviewer bias. Since such surveys are anonymous social pressure on the respondent is smaller and interaction with the interviewer can not bias the respond.

Regarding representativeness online surveys such as Europuls are likely to have more

need to count NAs for relevant variables?

3.3 Data for Post-stratification and Weighting

In order to post-stratify the data, we want to use different sources and compare their effect on the forecast: (1) exit poll data from Forschungsgruppe Wahlen or intratest dimap Institute (what), (2) representative election statistics and (3) microzensus data.

General: include how did we get the data...

(dicuss biases along the framework of potential sources of bias measurement and representativeness)

4 Methodology

1. Weighting
2. Simple use of self-reported turnout? (Gallup paper: distinguish deterministic and probabilistic methods)

The general solution to non-representativeness is weighting. With online surveys you need to do a little more, instead of doing a stratification-design to draw from certain strata you need to define the strata afterwards and weight your responses accordingly (post-stratification). This can be done with Census data that gives you the frequencies of the strata in the population. Then, when you made your survey representative of the population, in a second step you do what all polls have to do. Account for self-selection and non-response bias and actual-voter bias: -> Therefore, weight according to likelihood of actual voting and party affiliation of certain strata. (Here I want to point out the two steps!)

Now: Our actual project. In our forecasting project we want to use non-representative polling data from mobile-phone app users. (Obtained from Dalia research, citation blabla) So we have a raw survey and we want to weight it. We want to dicuss severall weighting methods. And we want to make a forecast of the Bundestagswahl. blalablablubb

4.1 Approach number one

For stratifying the data we orient ourself at the work of XXX. The basic idea is to compute clusters of voters along several demographic categories and use their past votes to compute weights.

First, if possible we use post-stratification (see Lumpley, ch. 7?) to compute weights for subgroups in the sample. Post-stratification tries to make a sample representative of the actual population by ensuring the relative size of subgroup resembles the relative size of the same subgroup in the 'true' population. For forecasting the 'true' population is not known, as it is a question of who will actually turn out to vote.

How we plan to make poll representative

Compare different stratification approaches

Likely problems we will encounter: 1. empthy clusters or clusters with low number of observations. Implications: If empthy, there is a real problem. If the number of observation is low, e.g. below 20, the weigths will will amplify the impact of this small group in the total forecasting result.

Benchmark -> other publically available polls. This is straight-forward, but also problematic as it might induce a herding effect. The final evaluation is only possibe after the election

4.2 Approach number two

5 Data Overview

how representative our data already is

1. raw data forecast. Compared to other forecastes the data under represents the CDU as well as the SPD. (Verify)
2. Show distribution of respondents on different demographic clusters and compare to zensus / exit polls / election statistics
3. Raw (voted last election)

6 Results

What the result is of making it representative

1. Election forecast Weighted with exit polls
2. Election forecast weighted with election statics
3. Election forecast weighted with zensus

Compare the three different weighten approaches

Comment Moritz: I think we have to weight the data with Zensus data in any case at least for gender and age as long we don't want to use the Dalia weights; then we can either use election statistics or exit poll data (that we don't have)

My approach would be:

1. Weighting with Zensus (accounts for self-selection of the survey)
2. Different mixes of weighting with election statistics with education (accounts for likely voter bias)

7 Conclusion

Summary of the core finding

Further implications

8 References

"Assessing the Representativeness of Public Opinion Surveys." 2012. Pew Research Center. http://www.people-press.org/files/legacy-pdf/Assessing%20the%20Representativeness%20of%20Public%20Opinion%20Surveys.pdf.

Holbrook, Allyson L., Jon A. Krosnick, and Alison Pfent. 2007. "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." In *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 499–528. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470173404.ch23.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National Rdd Telephone Survey." *The Public Opinion Quarterly* 70 (5): 759–79. http://www.jstor.org/stable/4124225.

Mellon, Jonathan, and Chris Prosser. 2015. "Investigating the Great British Polling Miss: Evidence from the British Election Study." *SSRN Electronic Journal*. doi:10.2139/ssrn.2631165.

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

Squire, Peverill. 1988. "Why the 1936 Literary Digest Poll Failed." *The Public Opinion Quarterly* 52 (1): 125–33. http://www.jstor.org/stable/2749114.

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.

Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang. 2011. "Comparing the Accuracy of Rdd Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." *Public Opinion Quarterly* 75 (4): 709–47. doi:10.1093/poq/nfr020.