

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

Election Forecasting Project

*Moritz Hemmerlein & Alexander Sacharow*

*01 April, 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 weighting. This paper sets out to discuss approaches of weighting-design to adjust a non-representative online poll at hand on the upcoming german federal election.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Related Literature</b>	<b>2</b>
<b>3</b>	<b>Data and Potential Biases</b>	<b>3</b>
<b>4</b>	<b>Weighting Approach</b>	<b>4</b>
4.1	Approach number one . . . . .	4
4.2	Approach number two . . . . .	5
<b>5</b>	<b>Data Overview</b>	<b>5</b>
<b>6</b>	<b>Results</b>	<b>5</b>
<b>7</b>	<b>Conclusion</b>	<b>5</b>
<b>8</b>	<b>References</b>	<b>5</b>

## 1 Introduction

The digitalization is challenging the way polling was done for many decades. Calling people on their landline phones has become difficult as response rates drop and households equipped

with such 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 the obstacles, pollsters have begun to experiment with non-representative polling methods (W. Wang et al. 2015).

The major aim of pollsters in using non-representative surveys is not genuinely different from traditional goals of polls. The task is to gather responses which are representative for the population with regards to demographic and other criteria. The major difference is that representativeness cannot be ensured before the actual poll takes place, but is introduced ex-post. 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 post-stratified in order to reflect the replies from the true population.

In our paper we want to explore and analyse different approaches to post-stratification for non-representative election polls of the German federal election 2017. We work with individual level data from mobile-phone app users. This raw data will then be post-stratified with different statistics about voters in Germany. Ultimately, we want to identify the most promising approach in order to generate an election forecast. As the ultimate election will only be after the end of the research, we can compare our post-stratified election forecast only with other forecasts of the 2017 federal election.

The paper is structured as follows: Chapter two will survey the literature on online polls, their representativeness and approaches to employ such data to forecast elections. Proceedingly, we present our data and discuss several issues of representativeness and measurement error that need to be accounted for. In chapter four we present two (or three) methods want to apply to weight our data at hand. Chapter five is presenting and discussing our preliminary results. Chapter six concludes and gives an outlook on practical obstacles and methodological issues that our approach suffers from.

## 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 unreliable. First, response rates have declined heavily (Keeter et al. 2006; Holbrook, Krosnick, and Pfent 2007). XXX reported that the response rate is down to X% in XXXX, compared to X% in XXXX or even X% in XXXX. 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 realize its biased sample (Squire 1988). But the problem still exists, as non-representativeness was for example a major problem when forecasting 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 than online polls. However, they used only simple post-stratification methods.

W. Wang et al. (2015) in contrast are much more optimistic about the possibilities 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.

The difference between the two studies illustrates that the 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?

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

The forecasting project utilizes data from the Europulse Survey which is conducted quarterly by [Dalia Research](#) 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 census data for age and gender.

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) microcensus data.

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

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

## 4 Weighting Approach

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 Zensus 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 discuss several weighting methods. And we want to make a forecast of the Bundestagswahl. blablablubb

### 4.1 Approach number one

For stratifying the data we orient ourselves 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. empty clusters or clusters with low number of observations. Implications: If empty, there is a real problem. If the number of observation is low, e.g. below 20, the weights 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 possible after the election

## 4.2 Approach number two

## 5 Data Overview

how representative our data already is

1. raw data forecast. Compared to other forecasts 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

## 7 Conclusion

Summary of the core finding

Further implications

## 8 References

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 Japac, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 499–528. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:[10.1002/9780470173404.ch23](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1093/poq/nfr020).