

# Non-representative polling

## Mobile user polling data for the German Federal Election

Election Forecasting Project

*Moritz Hemmerlein & Alexander Sacharow*

*April 12, 2017*

### 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</b>	<b>2</b>
<b>4</b>	<b>Empirical Strategy</b>	<b>3</b>
<b>5</b>	<b>Data Overview</b>	<b>3</b>
<b>6</b>	<b>Results</b>	<b>3</b>
<b>7</b>	<b>Conclusion</b>	<b>3</b>
<b>8</b>	<b>References</b>	<b>4</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 (CITE). In response polling institutions have started to use other methods for polling ranging from face-to-face interviews to mobile phone calling. But they proven difficult to ensure representativeness or are expensive. To tackle this problem, pollsters began to experiment with non-representative polling methods (Wang).

In our forecasting project we want to use non-representative polling data from mobile-phone app users.

1. Surveys are in hardly any case representative; weighting is part of every survey practice; populations consist of different strata; those can be chosen beforehand (stratification) or afterwards (post-stratification)
2. Dalia data, forecast on 2017 election; needs to be weighted; we will do that using exit polls (alternatively census data?)

How

Results

Structure 3. Chapter 2 surveys the literature and discusses alternative approaches

## 2 Related Literature

Papers on polling

Papers on non-representative polling

Methods on correcting polls

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

## 3 Data

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.

(how did we get the data...)

## 4 Empirical Strategy

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.

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

## 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 census / 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 statistics
3. Election forecast weighted with census

Compare the three different weighted approaches

## 7 Conclusion

Summary of the core finding

Further implications

## 8 References