Voting intentions versus participation in voting

Rustamdjan Hakimov (University of Lausanne & WZB), Max Schaub (WZB)

Preregistration plan submitted to EGAP on 6. December 2018

Introduction

When scholars analyze voting behavior at the individual level, they usually have to rely on self-reports of past and intended voting. However, it is widely known that these responses suffer from recall-bias: that respondents over-report their past votes, and overstate the strength of their voting intentions. But by how much? We use a unique dataset to answer this question. We conduct a large-scale (n=2,500) survey in Armenia prior to a crucial parliamentary election. This, together with unique policy of obligatory publishing of the names of all voters, allows us to answer these fundamental questions.

Population and sample

We aim to recruit a representative sample of the Armenian working-age population. The interviews are conducted by a reputable research institute. Sampling follows a stratified multi-level clustered random sampling design.

Our starting population is the voting-age population according to the voter register, which has about 2.5 million entries. Our primary sampling unit/clusters is the election precinct (there are a total of 1,959 election precincts in Armenia). To ensure a good geographic spread, we formed nine strata: one for the capital, and one each for rural/urban zones in four geographical areas (NE, NW, SE, SW) of the country. Within each stratum, we determined the number of PSUs to

be sampled in accordance with the relative number of voters in this stratum relative to the total. We then randomly selected a number of PSUs. The probability for a PSU of being selected was set to the relative number of voters relative to their total number in a stratum. Within each selected PSU/cluster, 12 individuals are recruited, aiming for a total sample of 2,496 observations.

The recruitment at the PSU-level follows a random-walk protocol. Starting with a randomly chosen number of steps and direction, enumerators recruit respondents in every 4th dwelling. Within-household selection is conducted by entering the identifying information of all household members aged 18 and above into the CAPI device, and have the device randomly pick one household member. In case the household or selected household member is absent, three contact attempts are made to conduct the interview. After a failed third attempt, the house to the left of the selected household is chosen.

Measures

In the survey, we recorded individuals voting intentions and we also asked respondents whether they participated in voting in April 2017.

1. A lot of people don't vote these days for one reason or another. Did you vote in the last national election in April 2017? [dummy_voted]
    a. Yes
    b. No
2. Do you intend to vote on the next Parliamentary elections (to be scheduled)? [vote_intention]
    a. Yes
    b. No
    c. I do not know

Armenian law makes it compulsory that the names of all voters and their date of birth are published after election day. We use this information and match it with the responses to our survey. This allows us to identify actual voting behavior of respondents. Based on respondents'

names and voting precincts, we access match official data on the names and age of voters to our survey data, to generate two variables: actual_participation2017 which is equal one in case of submitted wrote, and zero otherwise; actual_participation2018 which is equal one in case of submitted wrote, and zero otherwise.

Research questions

1. Do people over-report voting intentions and past voting behavior. If yes, is there a difference between the misreporting of voting intentions and past behavior?

2. Is "do not know" option in the question about voting intentions is an excuse to not the say "known "no"? We hypothesize that the share of nonvoters among those who answered "do not know" will be as high as the share of nonvoter who answers "not going to vote".

The first question is interesting per se, independent of the direction of effect and the presence of the difference between misreporting about intentions and past behavior. We construct the variable "misreporting 2017" and "misreporting_2018" which equal to the difference between the response in the survey and the corresponding actual_participation variables. We can answer Question 1 by analyzing the distribution of the misreporting variables, and test the second part of question 1 using simple proportion tests of, and Fisher exact tests, with treatment being the year of elections.

The hypothesis in question 2 is also directly tested through Fisher tests of participation_2018 with treatment variable being the answer "no" or "do not know" in voting_intentions survey question. Question 2 makes a methodological contribution of the importance of giving the "do not know "option and guidance to its interpretation.