MRP polling specifications and methodology

Sample

5,317 adults aged 18+. These were in two groups:

- 2,019 respondents who were nationally representative of the GB population (the 'NatRep' portion of the sample)
- 3,298 respondents in 100 constituencies of interest for the modelling purposes of MRP
    - 1,768 were in the 44 constituencies which were won by the Conservatives in 2019 but were previously held by Labour in 2017 in the North and Midlands ('Red Wall' seats), an average of 40 respondents per constituency
    - 1,580 were in an additional 56 constituencies of interest, predominantly other seats that were marginal in 2019 plus other Conservative gains from Labour, an average of 28 respondents per constituency

Fieldwork dates

Fieldwork was carried out between 29th May-5th June.

MRP methodology

Brief explanation of how MRP works

MRP (Multilevel Regression and Poststratification) is a statistical technique that enables estimating opinions at constituency level. It has three steps: first, learn what predicts opinion using survey data, e.g. those who voted Remain are more likely to support an extension of the transition period, using regression analysis. Second, build a detailed model of the population at constituency level, using statistics on demographics and past vote, so we can say e.g. how many women aged 65+ who voted to Leave there are in each constituency. Third, we combine the two previous steps, to say what proportion of those women aged 65+ who voted to Leave hold an opinion (from step 1) multiplied by the number of them (from step 2) to get the population holding an opinion. This gives us a detailed breakdown of opinions by demographics and past voting behaviour which we aggregate to a constituency level.

Inputs

focaldata's MRP model uses a range of individual and constituency level variables. Individual variables selected for this model were age, gender, education, region (NUTS1), constituency, plus previous election votes (2019, 2017, 2016 referendum). These were selected due to their predictive power in our analysis of each of the questions we applied MRP to. Additionally, we had an interaction term combining 2019 General Election vote and 2016 EU Referendum votes, to capture more accurately the different patterns of e.g. 2019 Conservative Remain voters and 2019 Conservative Leave voters. For more details on variable selection, see 'Detailed comments on MRP methodology' below.

For constituency variables, we include a significant number of variables, which include (but are not limited to) population density, % long term unemployed, % leave 2016, GE2017 and GE2019 vote share, deprivation index and EU parliament 2019 vote share. All data is sourced from the Office for National Statistics (Annual Population Survey and Census) where possible, plus the Electoral Commission for election data, and estimated by focaldata otherwise.

Model specification

We use a bayesian exploded logit model, which is fit using Hamiltonian Monte Carlo with the open-source software Stan. The models are trained on the Google Cloud Platform.
As the questions modelled were not vote intention, there was no turnout model applied, so results are representative of nationwide opinion (rather than the opinions of those who turn out to vote).

Detailed comments on MRP methodology

Questions

There were three questions that we modelled using MRP:

1. If both the UK and the EU agreed that an extension to the transition period would help them deal with Coronavirus, which of these options would you prefer?
    a. Support an extension to the transition period
    b. Oppose an extension to the transition period

2. Do you think the cost of daily essentials will get better or worse if the UK leaves the transition period without a trade deal?
    a. Better
    b. Worse

3. The Conservative campaign Manifesto said that the Government would pursue "....a new free trade agreement with the EU. [and that]...this will be a new relationship based on free trade and friendly cooperation..." -- how important is it that the Gov keep this promise?

a. Very important
b. Important
c. Somewhat important
d. Not important

Variables for each question

These questions had different variables that were predictive. Questions 1 and 2 broke down on familiar lines; our exploratory analysis found the most predictive variables were all of the past vote questions (in order: 2019 vote most predictive, then 2016 referendum, then 2017 vote); this pattern is the same as in vote intention questions. For question 1, gender, age and education were all predictive. Region was less predictive, but still worth including in the model as part of modelling constituencies more effectively. For question 2, all the demographics were less predictive than in question 1, but still worth including in the model. Most notably, age was the weakest demographic predictor in question 2, whereas it was one of the strongest demographic predictors (with gender) in question 1.

For question 3, all variables were substantially less predictive, because there was widespread agreement that the government keeping their promise was important. Nevertheless, previous vote choice was most predictive in the same order (2019 vote most predictive, then 2016 and 2017). Age was the most important demographic predictor, with region, education and gender not being particularly predictive.

Given the similar patterns of which variables were predictive, we used the same variables in all of the models; the best case for exceptions to this were removing gender and education in question 3, but we decided it was best to keep the models consistent across questions (and the inclusion of non-predictive variables increases run-time and cost, but does not decrease model performance, due to regularisation applied that prevents overfitting).

In addition to adding variables individually, we also added an interaction variable between 2019 vote and 2016 vote. These interaction variables are most useful when there are non-linear effects among different groups, but are difficult to accurately use without large samples. Given our sample size, we added one interaction between the two most predictive variables, 2019 vote and 2016 vote. This matches what we did during the 2019 General Election (where, when we had a larger sample size, we also added other interaction variables).

Sampling methodology

This survey has been conducted using an online interview administered by focaldata. Our platform collects data from our commercial suppliers, such as traditional online panels and numerous programmatic sampling platforms, which allow us to find respondents to a range of panels through software. We then use Machine Learning to filter out bad respondents and get as representative a sample as possible. Users fill out the surveys in real-time across mobile, desktop, and tablet devices on the focaldata platform.

The data was weighted to be representative of the GB population for the NatRep portion of the sample. focaldata contacted members of the panel that match the demographic profiles of the country, in particular age, gender, region and socioeconomic grade (using the NRS definition). It then weighted the raw data to match the known population of Great Britain. A second weighting was applied to use the whole sample and make it representative, which was done by weighting to age, gender and socioeconomic grade (but not region, so that the red wall sample was not substantially downweighted).

About focaldata

focaldata is an AI-driven polling company based in London. It has conducted MRP for a range of commercial organisations and campaigns including Propercorn, Hanbury Strategy, Hope not Hate, M&C Saatchi, AbinBev and Best for Britain.

It was the main MRP provider for the Conservative Party for the 2019 General Election, although focaldata is a politically neutral organisation.

focaldata is a member of the British Polling Council (BPC) and abides by its rules. focaldata is also a member of the MRS

Further enquires

justin@focaldata.com