Remain United
Tactical Voting Project

Methodology

20 May 2019

1. Introduction

The project brief was to estimate the result of the May 2019 European Elections in each electoral region of Great Britain, and to derive suggestions for pro-Remain tactical voters to give the greatest number of seats to the pro-Remain parties, defined as Liberal Democrat, Green, Change UK, SNP and Plaid Cymru.

The technical aspects of the project are divided into two parts:

- Estimation of the election result by region
- Optimization to find the best tactical voting strategy

2. Election Result

To estimate the election result, custom fieldwork was commissioned by pollster ComRes. This consisted of two waves of polling, each surveying around 4,000 respondents. ComRes is a respected pollster and has been operating for over fifteen years, and is a member of the British Polling Council. ComRes surveyed GB adults in a relatively representative way in accordance with its normal practice.

To increase the accuracy of the polling, and to provide region-by-region predictions, the polls was analysed using Regularized Prediction and Post-stratification (RPP). This is a generalization of the existing Multi-level Regression and Post-stratification methodology (MRP) first proposed by Gelman and Little [1].

The advantages of RPP (and MRP) over classic polling are twofold. Firstly, RPP does not require the sample to be strictly or nearly representative. The regression aspects of the method allow fitting of responses to the respondents' demographic and political indicators, as long as there are sufficient numbers of various types of profile, rather than each profile being drawn from a truly representative sample. Instead the representativeness is provided by the post-stratification step.

Secondly, the pooling nature of the RPP methodology allows geographic prediction at more granular geographic levels, such as regional level. The method pools respondents across the whole country, and whilst it takes note of the region of each respondent, it also allows respondents from other regions (with the correct demographic weighting) to influence the regional prediction. This method is similar to that laid out in the papers [1] and [2]. As stated in [2], "MRP has been shown to produce reasonably accurate estimates of state public opinion using as little as a single large national poll – approximately 1,400 survey respondents". Our backtesting, on previous UK general election poll datasets, suggests that a sample size of 4,000 is sufficient to have reasonably accurate results in the UK context.

Although it is possible to use classic polling and consider the sub-sample within each region, these sub-samples can be individually small (around 200 respondents), noisy, and subject to sampling bias. Initial consideration of two polls from YouGov [3] and Survation [4] suggested considerable noisiness and disagreement between the two polls at a regional level, with YouGov showing Change UK as the leading pro-Remain party in most regions, whilst Survation showed the Lib Dems as the leading pro-Remain party in most regions.

These classic polling errors can be mitigated by taking the average over a number of polls from a number of pollsters, though the time-averaging could create an estimate that is slightly out of date if the political situation is quite dynamic. In the case of the European elections this is a concern, since polling for the Brexit party, the Liberal Democrats and Change UK has moved considerably since the start of May 2019.

It must be remembered that neither RPP or classic poll-of-polls will give a prediction which is completely accurate. There will always be random fluctuation between the prediction and the actual result caused by random sampling error, sampling or model bias, respondent distortion, or change in opinion between the polling and 23 May.

It is for this reason that the second step is particularly important in the tactical voting context.

3. Optimizing the best tactical strategy

The second stage is to find the optimal tactical voting strategy in each region. This is based on a model which incorporates two important real-world factors:

- The opinion poll estimates of support for each party in each region have a random error compared with the actual outcome
- The fraction of pro-Remain supporters who will actually vote tactically in an unknown, but important, quantity

To allow for the random error in polling, we create a stochastic model of the actual outcome compared with our known polling estimates. The model used is a multi-variate beta distribution (also called a Dirichlet distribution), which is supported on the probability simplex over the parties. The parameters are chosen so that the polling estimate is taken to be an unbiased estimator for the true outcome. In symbols, if p_{ij} is the estimated vote share for party j in region i, then the actual outcome vote share $X_i = (X_{ij})_j$ is a multi-variate beta distribution, with

$$\mathbb{E}(X_{ij}) = p_{ij}, \text{ and } Var(X_{ij}) = 4p_{ij}(1 - p_{ij})\sigma_{50}^2.$$

The standard deviation of party with 50pc popularity, σ_{50}, was chosen to have a reasonable value of 3pc, which was consistent with the confidence intervals of the regression. That value implies a standard deviation of 1.8pc for a party with 10pc popularity, and a standard deviation of 2.4pc for a party with 20pc popularity.

The *tactical fraction* is defined to be the fraction of supporters of the pro-Remain parties who will actually vote tactically in accordance with the advice. This can be any number between zero (no-one votes tactically) and 100pc (everyone votes tactically). In practice, the tactical fraction is a unknown parameter, but may exist somewhere in the reference range of 0pc to 60pc.

Lastly a universe of possible tactical strategies is defined. The tactical strategies are defined by two vectors: the "from-vector" and the "to-vector". The from-vector describes from which parties tactical voters will defect away, and the to-vector describes to which parties tactical voters will defect towards. The from-vector includes possibilities such as: From Green only; From Lib and Change UK; From All; and so on. The to-vector includes possibilities such as: To Green only; To Green and Lib equally split; to Green, Lib and Change UK equally split. In Scotland and Wales the to-vector also includes possibilities involving the nationalist parties. Overall around 50 possible strategies were defined using all combinations of possible from- and to- vectors.

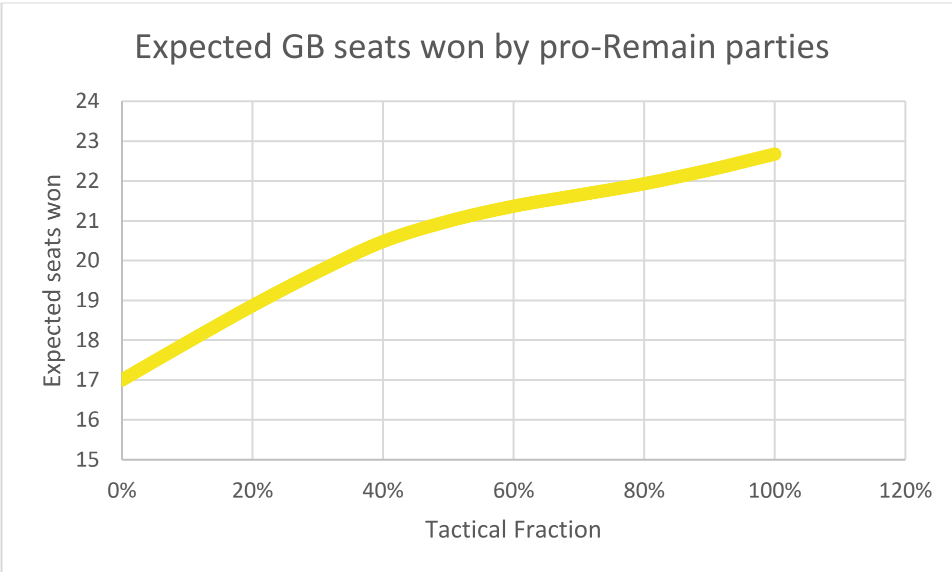
Given all these definitions, the calculation is performed. Over a large number of random simulations, the random actual support for each party in each region is simulated, using the multi-variate beta distribution described above. For each possible value of the tactical fraction, and for each of the possible strategies, the adjusted vote share after tactical voting affects is then calculated. The vote shares are then fed into the D'Hondt PR seat calculator to calculate the number of seats for each party in each region, for each random simulation, for each tactical fraction, and for each possible strategy.

The average result is then taken over all the random simulations to get the expected number of seats won by each party in each region, for each tactical fraction and for each possible strategy. As this is an expected number, it can be fractional.

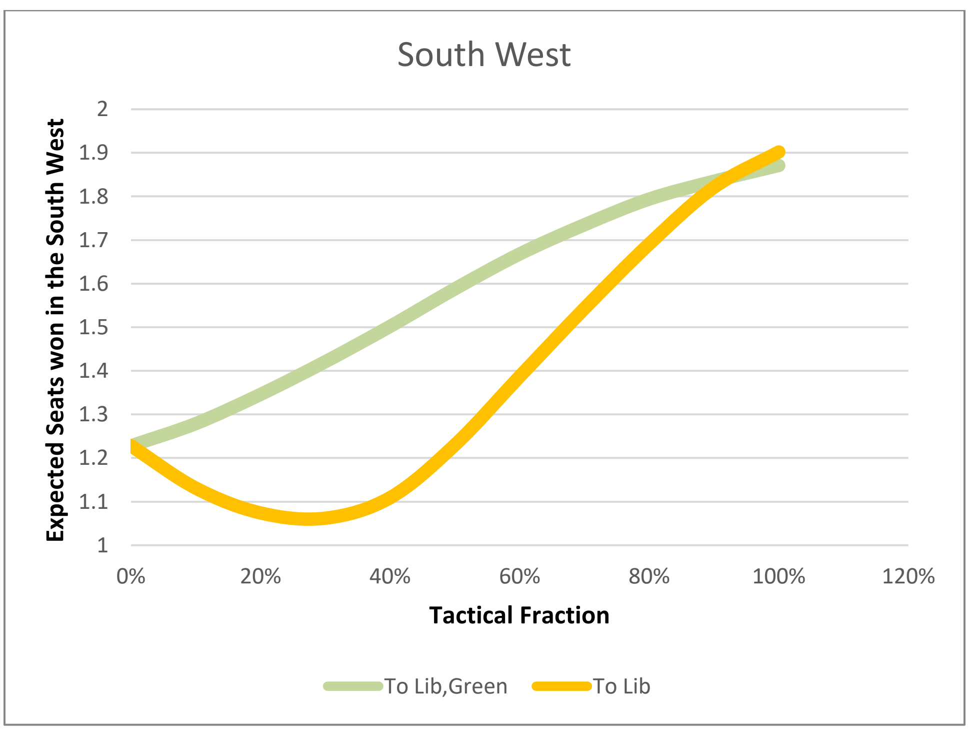
A score for each strategy is defined as the average of the seats won by pro-Remain parties over the values of the tactical fraction from 0pc to 60pc.

The strategies are then sorted by score, and the strategy in each region with the highest score is identified as the optimal strategy.

The graph shows the sum of the optimal strategies for each GB region, giving the total performance over all regions. It shows a steady increase in pro-Remain seats won as a fraction of how many people vote tactically. More tactical voting leads to more seats won.



Since the likely tactical fraction is not more than 60pc, this allows optimal strategies which would be sub-optimal if the tactical fraction were near 100pc. For example, the optimal strategy in the South West is for tactical voters to support either the Lib Dems or Greens. Although this under-performs the alternative strategy "vote Lib Dem only" if the tactical fraction is very near 100pc, it is better than that strategy if the tactical fraction is less than 90pc. The optimization automatically selects this strategy as the best for this region.



The alternative to using a stochastic model is to use a deterministic model where the polling estimates are taken as completely accurate, or at least accurate to around 0.5%. There are three main problems with this approach:

- It is not realistic to "know" the election outcome to that degree of precision
- Inference is then made about precisely how many votes a party can lose while just retaining a seat. But if the estimate is slightly wrong then the seat is lost needlessly.
- It privileges one particular scenario when comparing strategies leading to false comparisons.

A classic example of the last problem is to evaluate a strategy on the basis of one single estimate of party vote share, while ignoring all the other possible vote shares that might happen. It is the average performance of a strategy that matters more.

4. Conclusions

For estimating the prediction election result by region, a modern regression-based method was used. This helps correct for unrepresentative sampling in the fieldwork and also should give more accurate regional-based predictions than using the regional sub-samples.

For identifying the optimal strategy in each region, we used an optimization over a stochastic model of outcome vote shares, and over possible values of the tactical fraction, and over many possible strategies. This gives an optimal strategy which is robust against errors in the polling estimates, and over a range of different amounts of tactical voting.

In summary, modern scientific and mathematical tools have been used to determine the best advice for tactical voters at the European elections.

References

[1] "Poststratification into many Categories using Hierarchical Logistic Regression", Andrew Gelman and Thomas C Little, Survey Methodology Dec 1997. URL: http://www.stat.columbia.edu/~gelman/research/published/poststrat3.pdf

[2] "Estimating State Public Opinion with Multi-level Regression and Poststratification using *R*", Jonathan Kastellec, Jeffrey Lax and Justin Phillips, October 2016. URL: http://www.princeton.edu/~jkastell/MRP_primer/mrp_primer.pdf

[3] YouGov survey of 5,412 respondents between 23 and 26 April 2019, for Hope not Hate. URL: https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/gx95u8qvbi/HopeNotHate_190426_EUElections_bpc_w.pdf

[4] Survation survey of 1,999 respondents between 17 and 25 April 2019. URL: https://www.survation.com/wp-content/uploads/2019/04/European-Parliament-Election-Voting-Intention-UK.xlsx