Swing Voters

Introduction

This is survey data collected by the thinktank data for progress that represents the registered population of people who 2018 mid-term elections. There are two types of variables of out interest, first, `issue` variables, which has six different issues and people has responded on scale of 1 to 5, 1 being strongly supporting and 5 being strongly opposing. The second variables of interest are `populism` variables, which has three sub-categories. We try to study the `Swing Voters` and what variables determines being a swing voter from this data set. We start with the visual comparison, figure 1, of two types of swing voter, one, who swinged to Republican (to R) and another, who swinged to Democrat (to D)

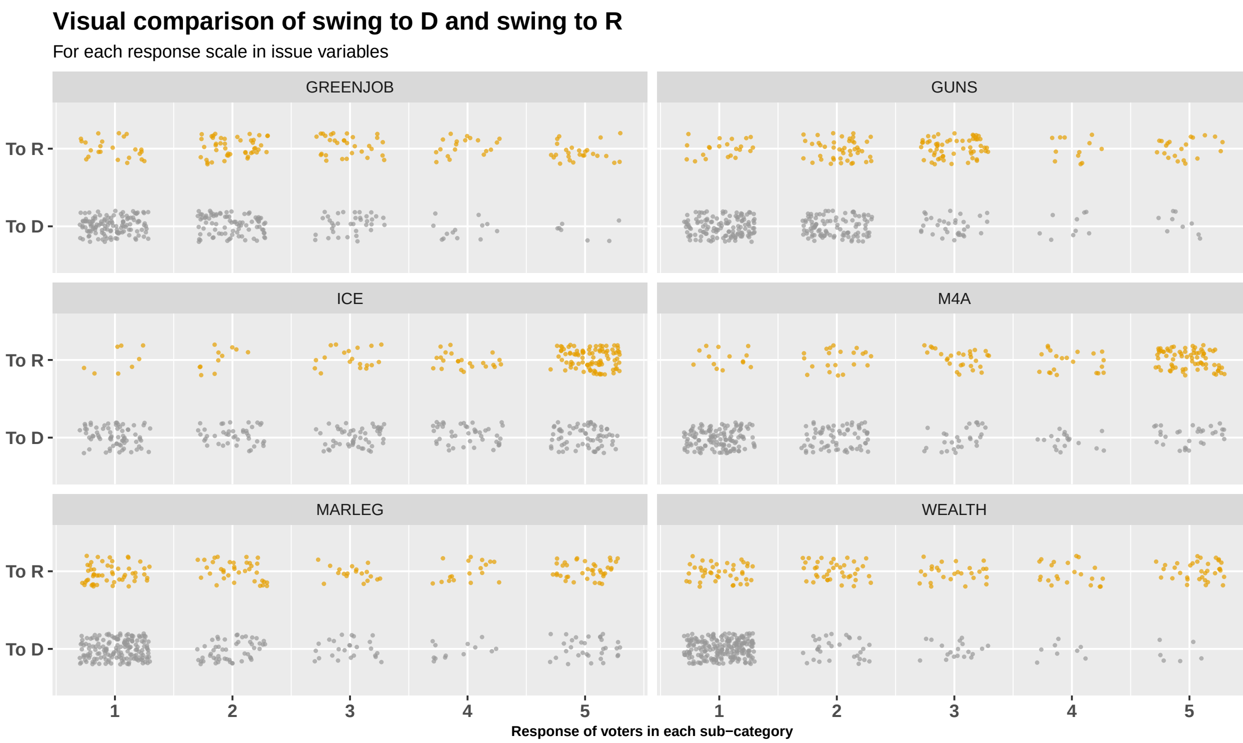

Figure 1: Visual comparison of swing to Democrats and swing to Republican voters

On visual comparison between the number of voters who has switched to Democrat (To D) and those who has switched to Republican (To R), we see that in most of the issue variables there are not much voters who have given scale values of 3 and 4 for either case. This makes one thing clear that swing voters are highly influenced by supporting and opposing various issues. A close observation of the plot in figure 1 shows that there is a larger concerntration of voters for Democrates on issue variables like `Wealth`, `Greenjob`, `Guns` and `M4A` in category of strong support (label 1), and the corresponsing number of voters for Republicans in these variables are very less for strong support category. Based on this, it can be concluded that strong supported of `wealth`, `greenjob`, `guns` and `medicare for all` swing towards Democrats. Similarly, the swing voters to Republicans are those who strongly opposed issues like `ICE` and `M4A.` Republican swingers are more or less equally distributed across all scales for `Wealth.`

A more clear observation can be made from weighted plot (figure 2) below, which clearly shows that there are two most important scale values that drives the swing voters i.e. category 1 and 5.If a voter either strongly oppose of strongly support some issue variables then they are likely to be swing voters, though this might

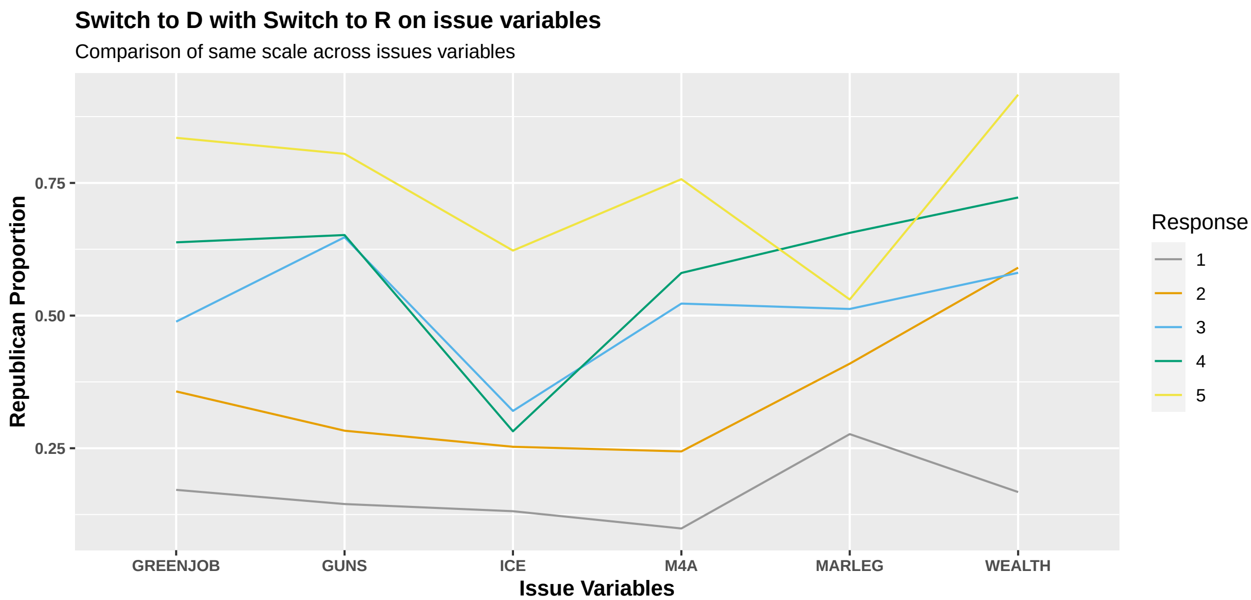

Figure 2: Proportion of swing voters for Republican party for each sub−category of support

not be always true. Swing voters who have responded on different issue variables in category 3 and 4 as well, the plot shows that this is not a decisive factor for any of the issue variables, rather we observe that strong supporter of `greenjob`, `wealth`, `medicare for all` and `wealth` are democrats voters and those who oppose these issues are Republican supporters. Issue variables like `Ice` and `MARLEG` are not strongly opposed by even those republican voters who has opposed other issues. This makes `Marleg` and `ice` as not that important issue variables.

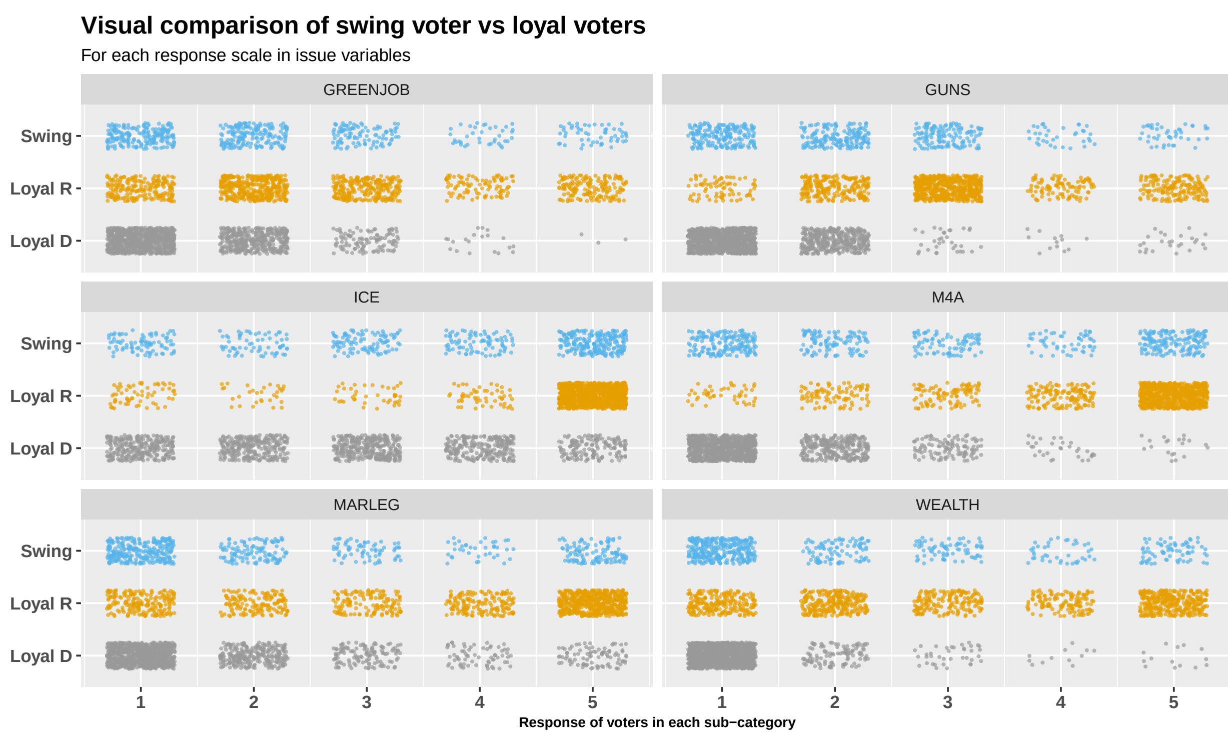

Figure 3: Visual comparison of swing voter with loyal Republicans and loyal Democrats on various issues

Having explored the pattern in swing voters, we now try to observe what drives loyal voters to stick to there party. Again, we start with the visual comparison for the loyal voters and swing voters, the plot in figure 3 shows the visual comparison of number of swing voters with loyal republicans and loyal democrats. The plot reveals that in some issue variables like `Greenjob`, `Ice`, `marleg` and `M4A` the swing voters follow the pattern of loyal republicans i.e. the number of swing voters are also concentrated highly in those categories in which republicans have higher number. In case of `Guns` issue variables, the swing voters are close to the behavior of loyal Democrats and also in Wealth issue variable, the swing voters are most closely behave like democrats loyalists. One important things to observe it loyal democrats are strong supporter of all these issues except `ICE`, where there are almost equally distributed across all categories. But the same pattern is not seen for loyal Republicans, Republicans strongly oppose `ICE` and `M4A` issues but most of them are on scale 3 for Guns and more or less equally distributed for rest of the variables.

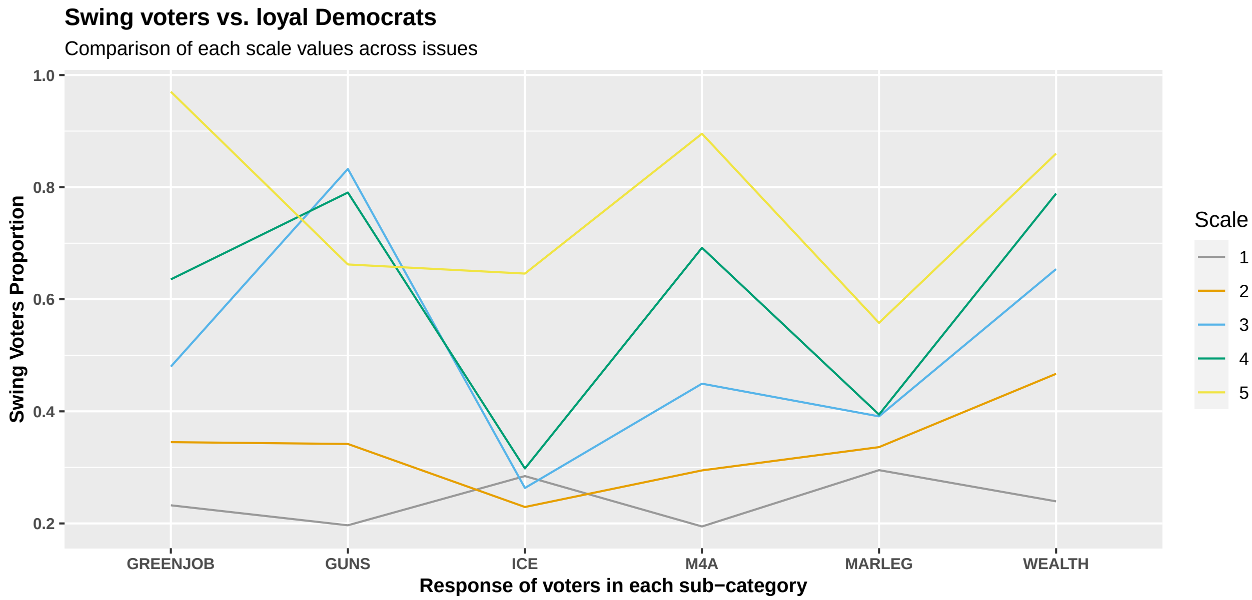

Figure 4: Proportion of swing voters with loyal Democrats for each sub−category of support

A more one to one comparison can be mode from weighted plot in figure 4, which shows the proportion of swing voters as compared with loyal democrats for each category of different issue variables. The plot shows that swing voter as compared to loyal democrats are more likely to strongly oppose `greenjob`, `wealth` and `Medicare` for all issue variables which has especially high support among democrats. The proportion of swing voters who strongly oppose `Guns`, `Ice` and `Marleg` are relatively less as on other issues. There is not such an strong opposition by swing voters on Guns issues but there is strong support for this among Democrats.

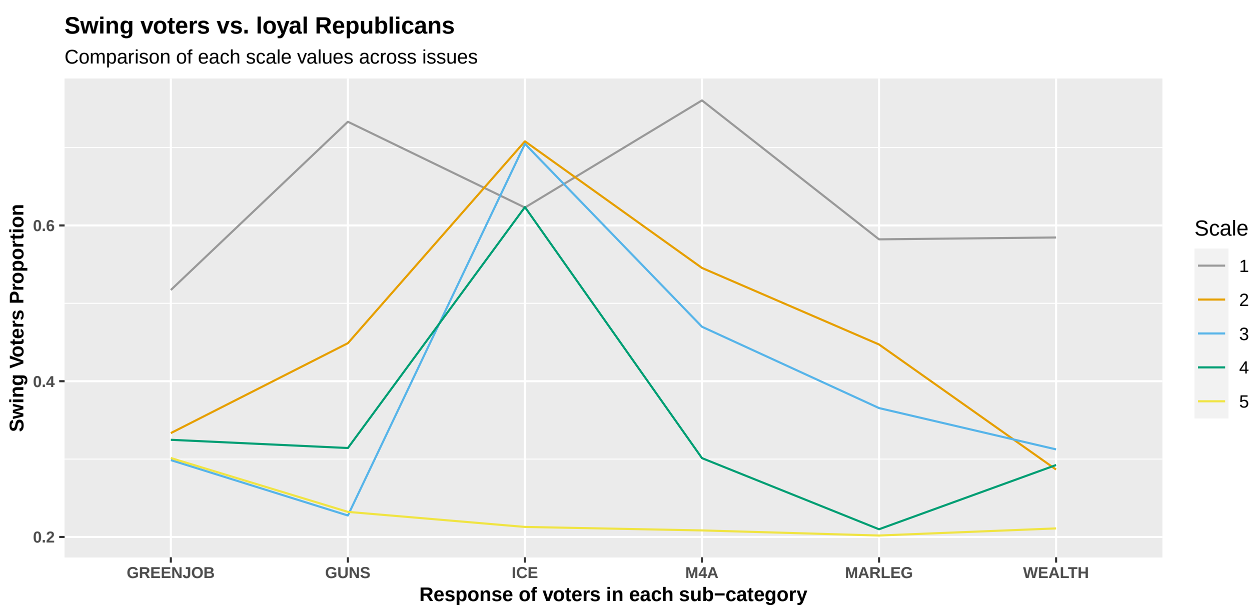

Figure 5: Proportion of swing voters with loyal Republicans for each sub−category of support

The weighted plot in figure 5 shows the comparison of swing voters with loyal Republicans. It is quite visible from the plot that a larger section of loyal republicans strongly oppose on the issues like `Ice`, `M4A`, `Marleg` and `Wealth`, but that is not the trend which is followed by swing voters. Swing voters strongly support `Guns` and `M4A` issue variables. There is also a very proportion of swing voters who neither strongly support or oppose the issue of `Ice` as compared to loyal republicans. Overall, it can be said that, swing voters think more like Democrats on some issues and more like Republicans on some issues. These patterns can be easily seen from figure 3, 4 and 5.

Predicting a swing voter

For both the models, using only issue variables as the predictor and using only populism variable as predictor, we fitted a GLM model with family `quasibinomial` and used the `weight_DFP` as the weight parameter. We did try a number of combinations for variable interactions but nothing seems to improve the model, so I choose my final model without any interaction terms (there wasn't any point using more degree of freedom for no improvement in model). Since, we are only interested in knowing the explanatory power of the model, we tested the model on same data with following outcomes.

Model with only issue variables as predictor:
:- On calculating the accuracy of prediction with direct classification as `swing` and `non-swing` voters, the model gives around 80.92 percent of accuracy, but this accuracy is deceiving as this model could only classify 2-3 voters as swing voters and classified rest of the voters are non-swing voters. This certainly creates doubts for this type of evaluation for this model.
:- then we tried to assess the difference in Probabilities of being a swing voter and being a non-swing voter by averaging the fitted values in two original labels of being swing and non-swing voters. The outcome highlights the weak predictive power of the model.
:- The model predicts a 0.2083 probability (on average) for non-swing voters and 0.2296 probability (on average) for being a swing voter. This small difference in the probability values shows that the model has very weak explanatory power. Still we tried to capture an average pattern of this model for each variables, which contributed in predicting a swing voter. This is shown in figure 6 below.

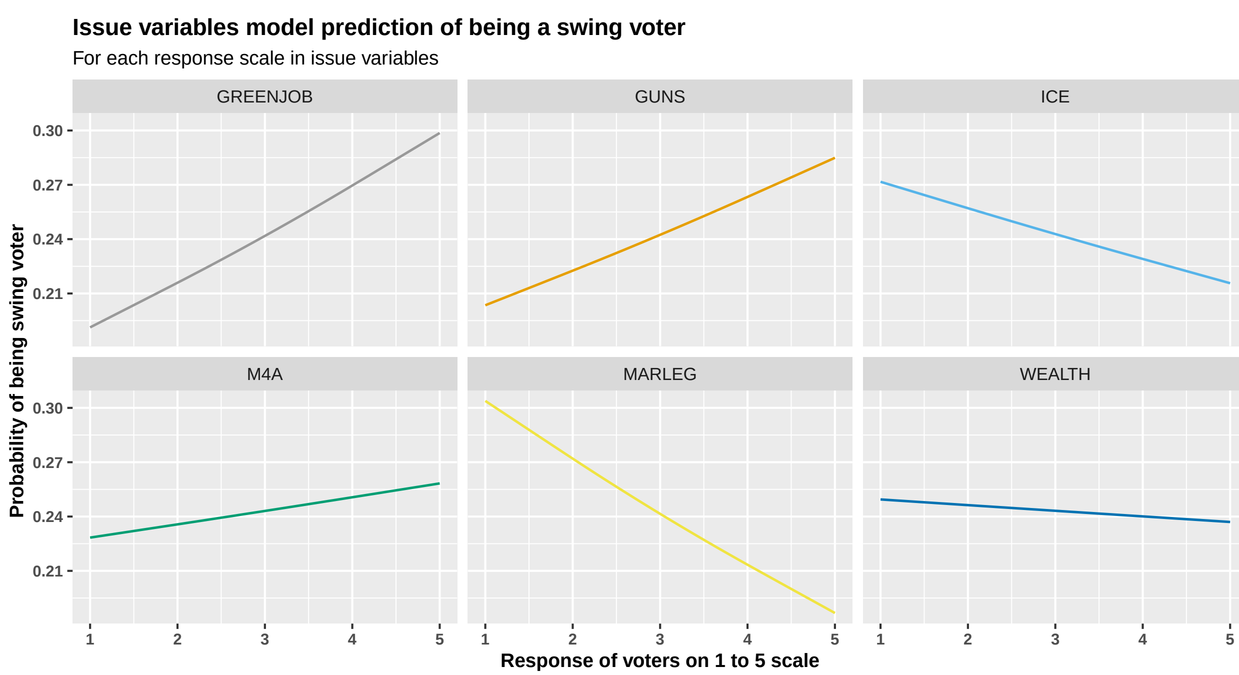

Figure 6: A general pattern of model prediction done on a grid.

Model with issue variables

swing.weight.logit1 = glm(response.variable ~ M4A + GREENJOB + GUNS + ICE + MARLEG + WEALTH, family = quasibinomial, weights = weight_DFP, data = DFP.issue)

This plot is by no mean accurate but this highlights the weakness of this model and some trends with different issue variables. We can observe that, there is no any variable that clearly predict a voter being swing voter, though it captures the general trend of each issue variables. It can be said that voters who opposes `greejob`, `guns` and `wealth` and supports `ice` and `marleg` are more likely to be a swing voter. A general trend is more or less captured here but it would not be advised to use this model for drawing any conclusion as this model has very less explanatory power.

Model with only populism variables as predictor:
:- On calculating the accuracy of prediction with direct classification as `swing` and `non-swing` voters, the model gives around 80.66 percent of accuracy, but this accuracy is deceiving as this model could only classify a very-limited voters as swing voters and classified rest of the voters are non-swing voters. This certainly creates doubts for this type of evaluation for this model.
:- then we tried to assess the difference in Probabilities of being a swing voter and being a non-swing voter by averaging the fitted values in two original labels of being swing and non-swing voters. The outcome highlights the weak predictive power of the model.

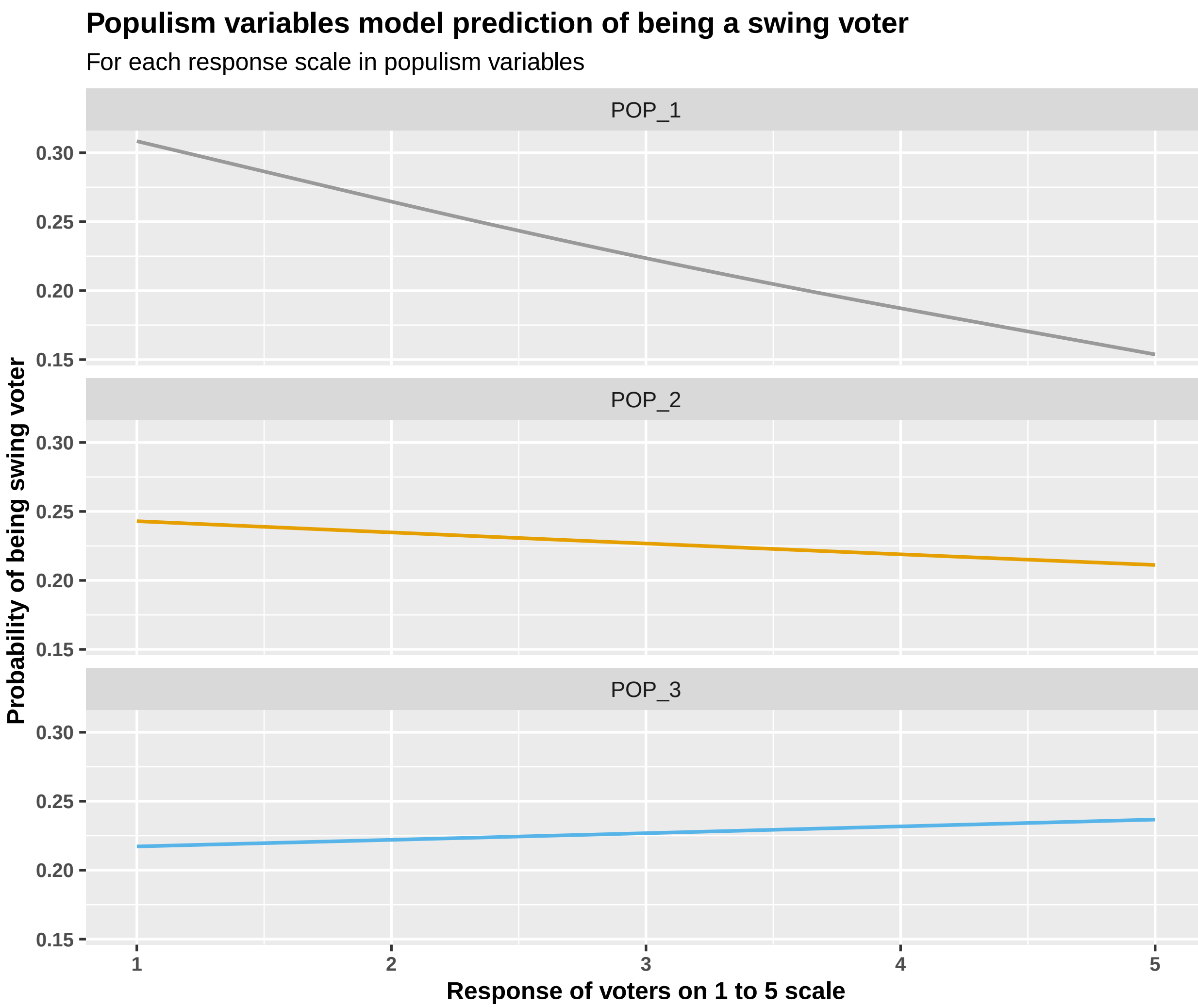

Figure 7: A general pattern of model prediction done on a grid.

:- The model predicts a 0.2114 probability (on average) for non-swing voters and 0.2320 probability (on

average) for being a swing voter. This small difference in the probability values shows that the model has very weak explanatory power. Still we tried to capture an average pattern of this model for each variables, which contributed in predicting a swing voter. This is shown in figure 7 below.
y

model with populism variable

swing.weight.logit.pop1 = glm(response.variable ~ POP_1 + POP_2 + POP_3, family = quasibinomial, weights = weight_DFP, data = DFP.populism)

This plot also only captures the general trend with populism variables but it don't clearly shows what affect more being a swing voter. This is also not sure accurate model and hence, shouldn't be used for any predictive purposes. As explained above,this model as well do not have much explanatory power but it shows a general trend that those who support `pop_1` and `pop_2` variable and opposes `pop_3` variable has higher chances of being a swing voter than the rest of the voters, conversely, it can also be said that those voters who strongly oppose `pop_1` and `pop_2` variables and strongly support `pop_3` variables have slightly higher chance of being non-swing voters.

After exploring these model separately, we combined the both the model and tried to access the explanatory power of a model based on both `issue` and `populism` variables. In this, model as well, we didn't use interaction between any of the variables. The outcome of the model is as follows:
:- On calculating the accuracy of prediction with direct classification as `swing` and `non-swing` voters, the model gives around 81 percent of accuracy, which might be 1 percent better than the other two models but this accuracy is deceiving as this model could only classify a very-limited voters correctly as swing-voters.
:- then we tried to assess the difference in Probabilities of being a swing voter and being a non-swing voter by averaging the fitted values in two original labels of being swing and non-swing voters. The gap between the average probability of being a swing voter and a non-swing voter widened in this model a bit.
:- The model predicts a 0.20 probability (on average) for non-swing voters and 0.24 probability (on average) for being a swing voter.
though the difference has widened between the average probability of swing and non-swing voters but this model as well do not have good explanatory power. A tabular comparison is shown below for all three model for their average probability values of predicting a voter to we swing and non-swing voters.

complete model with issue variable and populism variable

complete.swing.weight.logit.pop = glm(response.variable ~ POP_1 + POP_2 + POP_3 + M4A + GREENJOB + GUNS + ICE + MARLEG + WEALTH , family = quasibinomial, weights = weight_DFP, data = complete.DFP)

Voter	IssueModel(AvgProb)	PopulismModel(AvgProb)	CompleteModel(AvgProb)
Non-swing	0.2083	0.2114	0.2
Swing	0.2296	0.2320	0.24

I also tried to capture the general trend of this model with each variables. this is shown in figure 8 below:

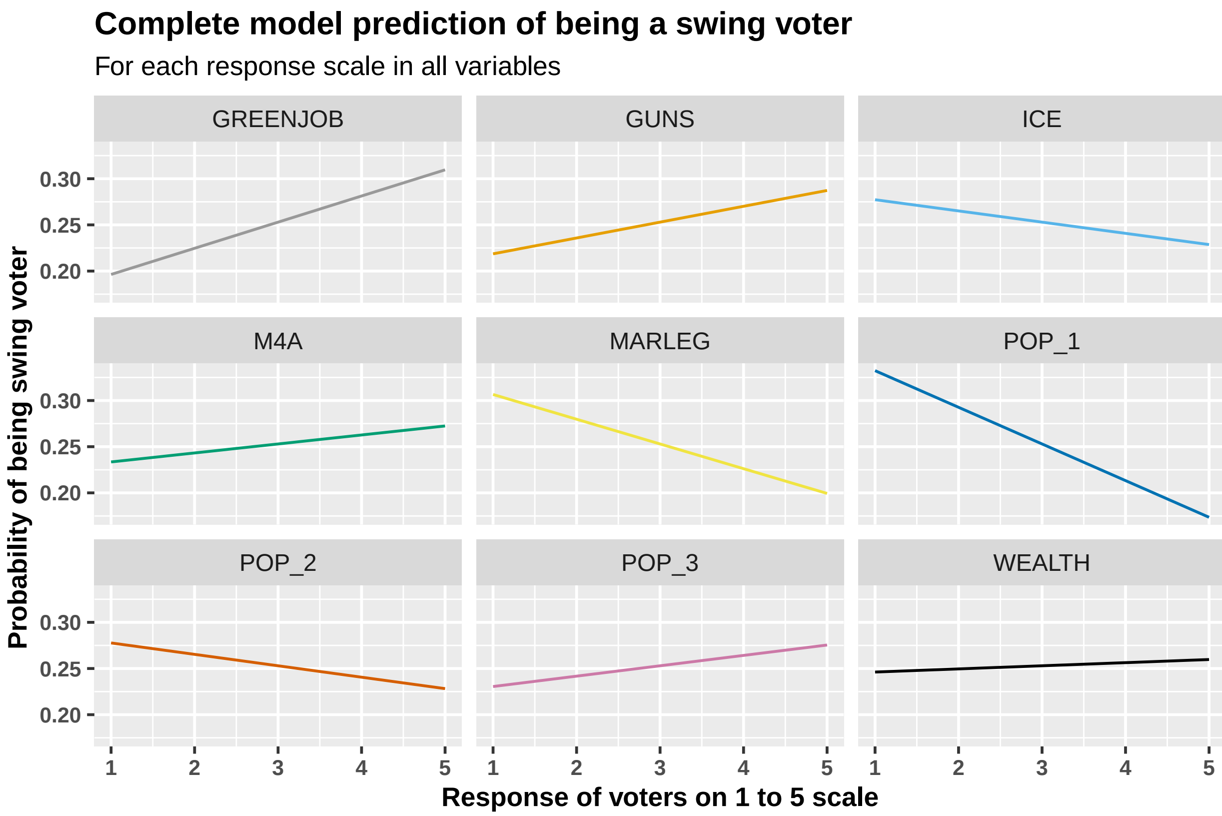

Figure 8: A general pattern of model prediction done on a grid.

The plot in figure 8 shows the general trend of being a swing voter for each explanatory variables. All these patterns are more or less consistent with the trend from the individual models that we made. Still, if we have to guess that what factors drives being a swing voters, then we can say that overall, any single variable doesn't give much information about being a swing voter but a combination of factors like, a voter who opposes `greenjob`, `guns`, `pop_3` and at the same time strongly supports `ice`, `marleg`, `pop_1` and `pop_2` are more likely to be a swing voter and vice-versa for non-swing voters. But we have already concluded that this model have weak explanatory power, so it doesn't sufficiently predict being a swing or non-swing voter. If we compare the point difference in prediction probability, it is only 4% as shown in the table 1, which is not enough to conclusively predict swing or non-swing voters.

If we compare the three models, first, model based only on issue variables, second, model based only on populism variables, and third, model based on both populism and issue variables, the third model i.e. the model based on both populism and issue variable is the best model. This model gives the highest difference between the average probabilities of swing voters and non-swing voters as shown in table 1 above. Other models don't not capture the difference.