Methods Appendix for Analysis of the Hub's Tax Policy Polling Using Multilevel Regression and Poststratification

January 4, 2018

1 The Regression Model

The tax policy polling by the Hub asked respondents to rate several statements on a scale from strong agree, somewhat agree, somewhat disagree, and strong disagree. We model all four responses using multinomial logistic regression.

Respondents are divided into cells based on demographic characteristics and the state that they live in. Letting i index cells and k index responses, the distribution of responses for the respondents in cell i is modeled with a multinomial distribution

$$y_i \sim Multinomial\left(n_i, p_{[k]i}\right) \tag{1}$$

Where n_i is the number of respondents in cell i. To infer the probabilities for each cell and response $p_{[k]i}$, we use the softmax inverse link function to map from probability to a linear predictor $\eta_{[k]i}$

$$p_{[k]i} = \frac{\exp\left(\eta_{[k]i}\right)}{\sum_j \left(\exp \eta_{[j]i}\right)} = softmax\left(\eta_{[k]i}\right) \tag{2}$$

The linear predictors are then modeled with the following regression equation

$$\eta_{[k]i} = \alpha^0_{[k]} + \alpha^{female}_{[k]} * female_i + \alpha^{college}_{[k]} * college_i + \alpha^{race}_{[k]r(i)} + \alpha^{age}_{[k]j(i)} + \alpha^{race:college}_{[k]r(i),l(i)} + \alpha^{state}_{[k]s(i)} \tag{3}$$

where $female$ and $college$ are binary indicators for gender and whether or not the respondent is a college graduate. The model additionally includes:

- Three race categories white, black, and other race, indexed by r(i)
- Five age categories: 18-29, 30-44, 45-54, 55-64, and 65+, indexed by j(i)
- Six categories for race - education interaction
- Fifty one state categories (DC is counted as a state) indexed by s(i)

There are a total of 3060 cells representing all combinations of these categories. To facilitate partial pooling between cells, we use a hierarchical regression for the state intercepts which are drawn from a normal distribution

$$\alpha^{state}_{[k]s} \sim Normal\left(\beta^{region}_{[k]p(s)} + \beta^{Trump}_{[k]} x^{Trump}_s + \beta^{AvgTax}_{[k]} x^{AvgTax}_s, \sigma^2_{[k]state}\right) \tag{4}$$

There are four regions used: Northeast, South, Midwest, and West. Two state level predictors are also included, 2016 Trump vote share x^{Trump}_s, and the average tax liability from the IRS statistics of income x^{AvgTax}_s. The standard deviation for the group distribution of the state intercepts is estimated from the data, and a half normal prior with unit variance is placed in this parameter. Similarly, hierarchical normal priors are placed on $\alpha^{race}_{[k]r}$, $\alpha^{age}_{[k]r}$, $\alpha^{race:college}_{[k]r}$, and $\beta^{region}_{[k]p}$ i.e.

$$\alpha^x_{[k]} \sim Normal\left(0, \sigma^2_{[k]x}\right) \tag{5}$$

where again half normal priors with unit variances are place in the $\sigma_{[k]x}$. All hierarchical normal models use non-centered parameterizations for more efficient sampling. Vague non hierarchical normal priors are placed on $\alpha^0_{[k]}$, $\alpha^{female}_{[k]}$, $\alpha^{college}_{[k]}$, $\beta^{Trump}_{[k]}$, and $\beta^{AvgTax}_{[k]}$.

The strong approve category is used as a reference category, and as such all coefficients for this category are set to zero. Therefore we have 68 parameters and 12 hyperparameters for each of the three remaining categories, for a total of 204 parameters and 36 hyperparameters. The inference is done with the No U-Turn Sampler, a variant of the Hamiltonian Monte Carlo method, which has been implemented in the open source package PyMC3.

2 Poststratification

Poststratified estimates for each of the 3060 cells are obtained by simulation as follows. For each realization, we draw multinomial probabilities $p_{[k]i}^{post}$ from the posterior distribution of the regression model, and then draw simulated responses for each cell from

$$y_i^{post} \sim Multinomial\left(n_i^{post}, p_{[k]i}^{post}\right) \tag{6}$$

The poststratification counts for each cell n_i^{post} are constructed from the census microdata from the 5 year estimates from the 2015 American Community Survey, and only includes voting age citizens. After drawing y_i^{post} we can aggregate results from different cells together to get state and national level estimates for each response. We can also aggregate responses, for example by combining strong and somewhat approve or by computing strong approve - strong disapprove.

Note, the Hub's survey was of registered voters, while we have constructed our post-stratification data from the universe of all voting eligible adults. Despite this, we find that our national level predicted responses are consistent with the estimates produced using the survey weights used by the Hub. For example, below we have plotted the MRP predicted distribution and mean along with the estimates using the Hub's weights for a question on cutting the corporate tax rate. The two methodologies lead to comparable results (the MRP predicted mean for overall approval and the Hub estimate plot on top of one another and are therefore hard to make out on the graph).

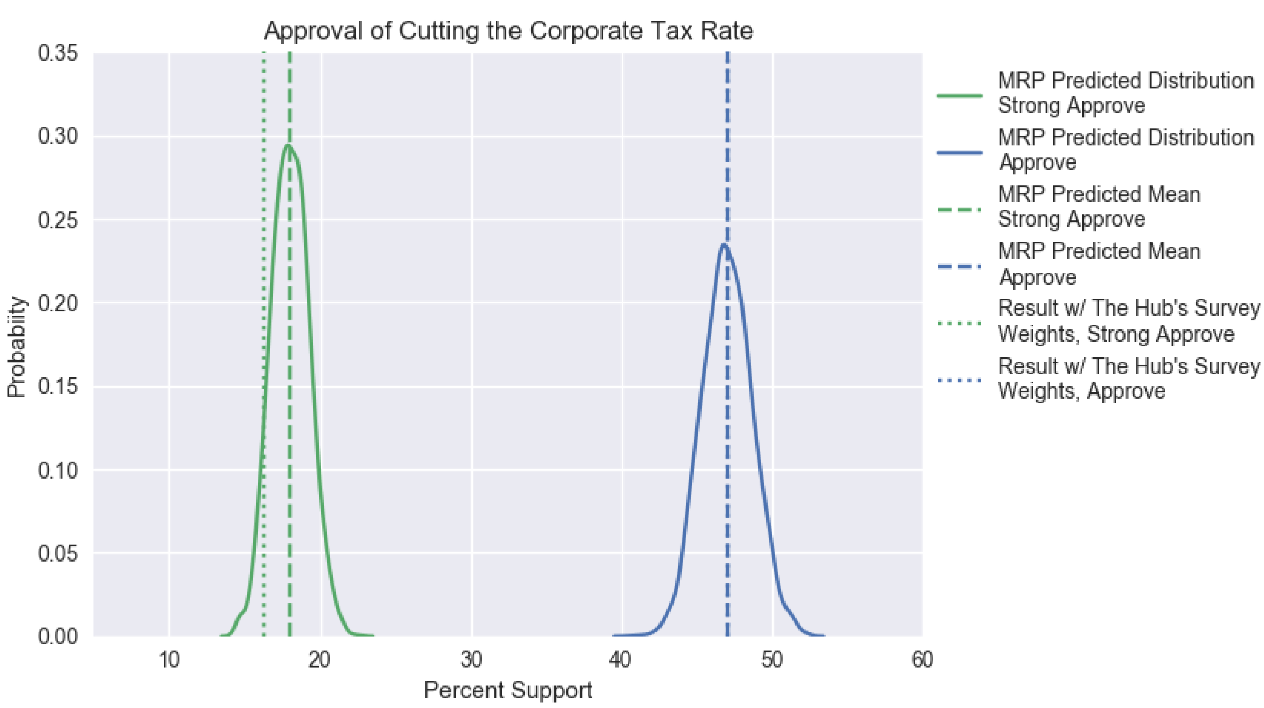

Figure 1: Comparison of National Level Estimates Using MRP and the Hub's Survey Weights

3 Example State Level Results

Here we'll show some example state level results using the corporate tax cut question. State level trends for overall strong plus somewhat approve are generally mild, but trends in strong approval are more interesting. In figure 2 we have the predicted mean estimates for strong approve and strong plus somewhat approve. Uncertainty in state level estimates is relatively high; the average size of the state level 95% predicted interval is 11% and 19% for strong approve and overall approve respectively. Estimates are shown for the overall population in each state, as well as for college educated whites. This group is generally more fiscally conservative than other demographic groups, and we find that this group is slightly more likely to approve and strongly approve of corporate cuts than the overall population. However, strong approval is generally low, even among this more fiscally conservative group. Therefore, the notion of some sort of severe electoral backlash against Democratic legislators for voting against the tax cuts and jobs act does not seem credible.

Perhaps counter intuitively, we find that the highest levels of strong approval for corporate tax cuts are in bluer, northeastern states (see figure 3). The association here is likely due to larger populations of well educated individuals, but it is also possible that Republican voters living in high tax blue states are particularly tax averse. However, it bears repeating that even among the most supportive groups, strong support for corporate tax cuts is low.

Lastly, we'll examine West Virginia in a bit more detail. Democratic Senator Joe Manchin appeared to be open to voting for the Tax Cuts and Jobs act, but at least on corporate tax cuts, we estimate that his constituents are among the least receptive in the nation. In figure 4 we show the predicted distribution for strong approve - strong disapprove of corporate tax cuts in West Virginia. Uncertainty is high, but our best estimate is that among those with strong opinions, corporate cuts are underwater. This casts serious doubt that a conservative democrat like Manchin will be punished by voters for a no vote on the TCJA.

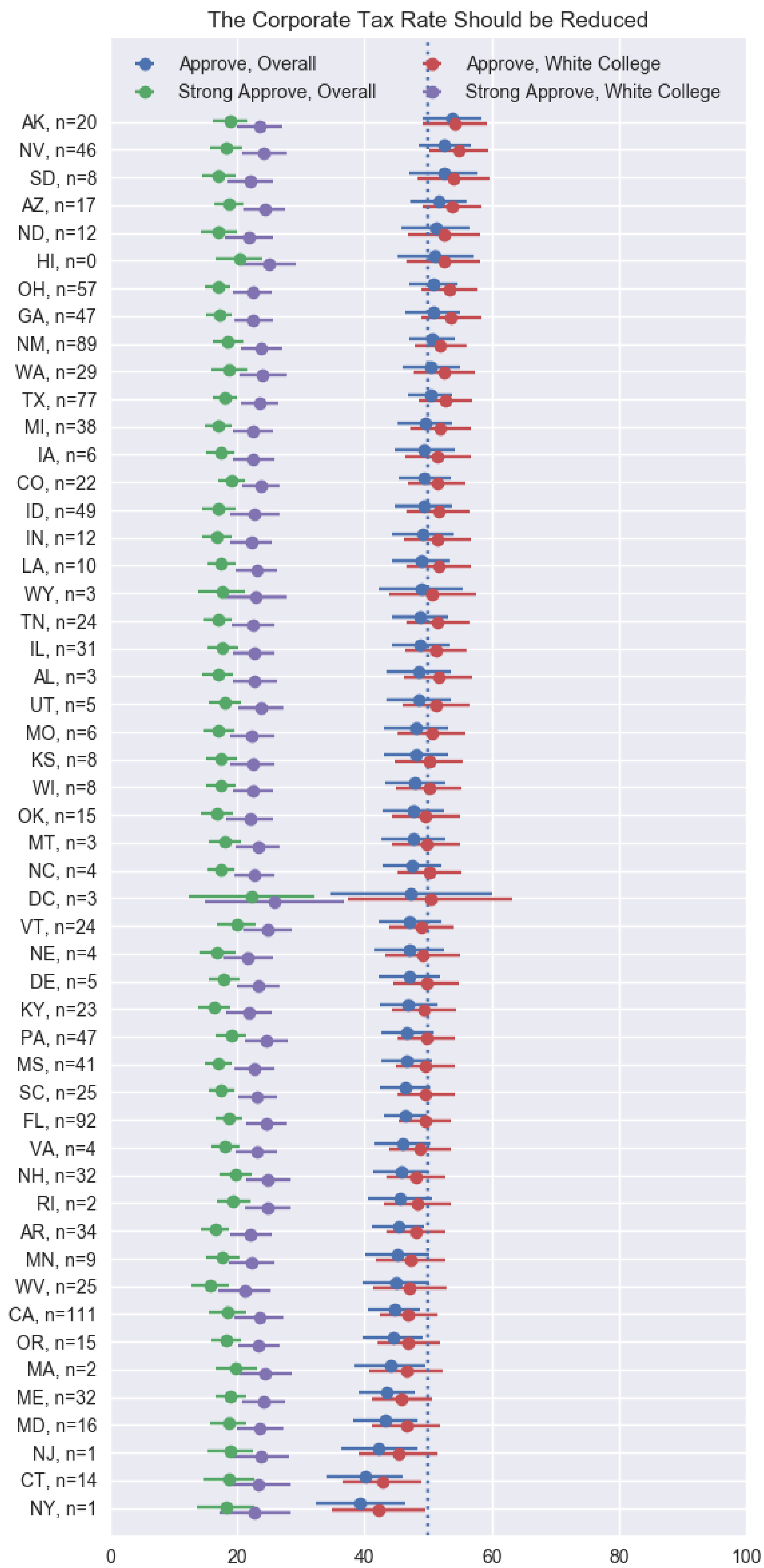

Figure 2: MRP results at the state level, error bars represent the standard deviation. The number of respondents from each state is shown next to the state's abbreviation on the left

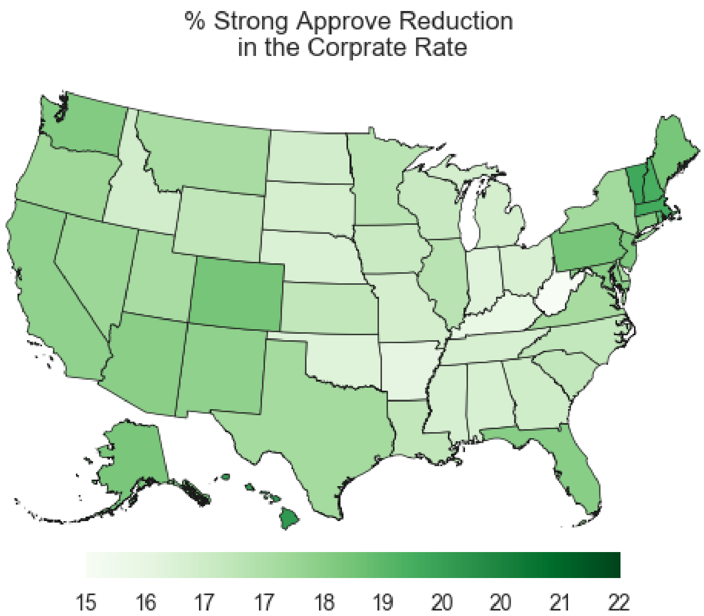

Figure 3: Mean estimate for strong approval of corporate tax cuts

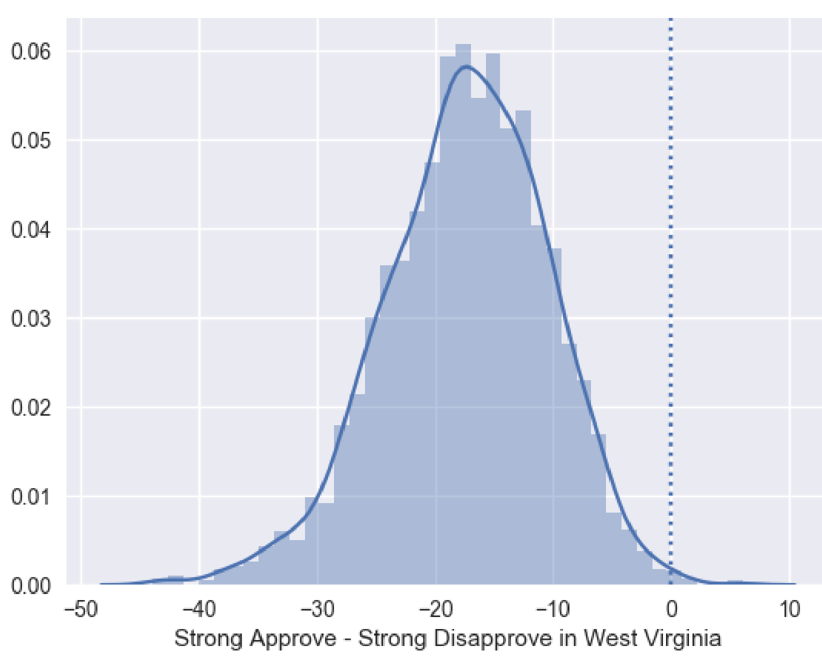

Figure 4: Predicted distribution for West Virginia