

# Multilevel Logistic Model on the proportion of voters for Trump and Biden

Benjamin Zhang, Daniel Leung, Diane Kim, Pablo Mercado

11/2/2020

## Model

Here we are interested in predicting the popular vote outcome of the 2020 American federal election using the survey data by Democracy Fund Voter Study Group to create a model. Then we will apply a postratification of the model with the census data by IPUMS USA. In the following sub-sections I will describe the model specifics and the post-stratification calculation.

## Model Specifics

To determine the likelihood of an individual to vote for a specific candidate in the upcoming election, voter survey data from June 25th 2020 (democracy fund voter study group) was used to generate two multilevel logistic regression models, one for Trump and one for Biden. First, individual responses were stratified by state, and then three survey results: family income, race and education, were specifically chosen as independent variables for the models. The first of these variables, family income, has been separated into 5 different income categories: 0 - 15k: living in poverty, 15k - 35k: low income, 35k - 70k: middle class, 70 - 150k: upper middle class, and 150k+ : upper class. Responder's race has been categorized under two categories, caucasian and non-caucasian. If the survey result of an individual's race is "mixed", they will also be treated as non-caucasian. Finally, the degree of education received is divided into four categories according to the highest academic degree achieved by an individual: no highschool diploma, highschool diploma, undergraduate diploma, and graduate diploma. Each of these categorical variables were then implemented as dummy variables within the logistic regression model. The following is a given equation for the probability of an individual to vote trump.

$$\hat{y} = \log\left(\frac{p_{vote}}{1 - p_{vote}}\right) = \beta_0 + \beta_1 x_{income} + \beta_2 x_{race} + \beta_3 x_{education} + \epsilon$$
$$\beta_0 = r_0 + r_1 w_{state} + u$$

This is the general model that we will apply to both Trump and Biden. Where  $y$  represents the proportion of voters who will vote for Donald Trump or Joe Biden. Similarly,  $\beta_0$  represents the intercept of the model, and is the probability of voting for Donald Trump or Joe Biden at having a household income of 0 - 15k: living in poverty, caucasian, and no highschool diploma. Additionally,  $\beta_1$  represents the slope of the model for each income bracket,  $\beta_2$  represents the slope of the model for race, and  $\beta_3$  represents the slope of the model for each education bracket.  $\beta_0$  is dependent on each state since this is a multilevel regression. Where  $r_0$  is the intercept and  $r_1$  is the slope of model depending on the state.  $\epsilon$  and  $u$  are the errors.

## Post-Stratification

In order to estimate the proportion of voters who will vote for Donald Trump and Joe Biden I need to perform a post-stratification analysis. We postratified using the results from the voter study group using the census data, with individuals being separated into different cells based on state, family income, race, and education. By using our logistic regression model, we calculate a predicted voting outcome within our survey

data based on those different cells from our census data. Next, each cell’s predicted estimate is weighted by its respective population from the census data. So the results from this post-stratifying can be used to predict how people in each cell would vote based on their family income, race, and amount of education. These variables were used in the poststratification because the logistic model indicated that they have some influence on voter outcome.

$$\hat{y}^{PS} = \frac{\sum N_j \hat{y}_j}{\sum N_j}$$

## Results

Table 1: Trump

Coefficients	P-value	intercept
educationGraduate.degree.or.more	0.7823966	-0.0349708
educationHighschool.diploma	0.4250661	-0.0773167
educationUndergraduate.degree	0.1691428	-0.1484771
household_income.150k...upper.class	0.0000000	0.9529140
household_income.15k...35k..low.income	0.0059672	0.2654526
household_income.35k...70k..middle.class	0.0000522	0.3815911
household_income.70...150k..upper.middle.class	0.0000000	0.5566683
race_ethnicityNot.Caucasian	0.0000000	-1.2878431
X.Intercept.	0.0000079	-0.4985759

In our logistic regression model, several weight parameters were more influential and significant than others. Specifically in our Trump logistic model, the weight values of household income and race were extremely significant ( $P < 0.01$ ). Increases in household income values appeared to positive correlate with the probability of an individual voting Trump. Furthermore, the model revealed the greatest influence on a person’s voter outcome was their race/ethnicity. Individuals who weren’t caucasian and came from an ethnic minority group were less inclined to vote Trump. Surprisingly none of the weight parameters for the education variables were statistically significant. Suggesting that the likelihood of an individual to vote Trump is irrespective of degree of education.

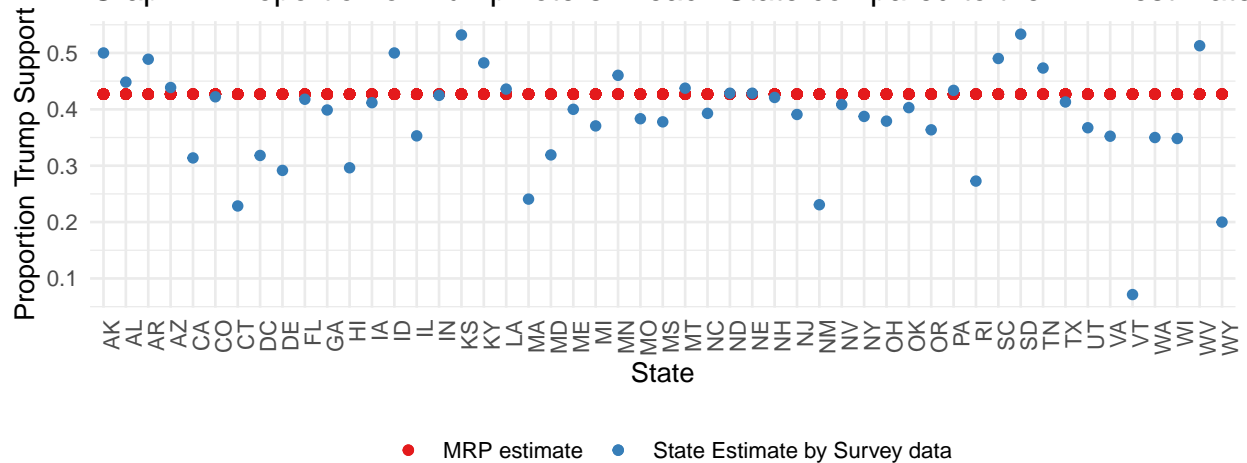
Table 2: Biden

Coefficients	P-value	intercept
educationGraduate.degree.or.more	0.0000010	0.6051512
educationHighschool.diploma	0.0024669	0.2838843
educationUndergraduate.degree	0.0000001	0.5623622
household_income.150k...upper.class	0.0054092	-0.3422176
household_income.15k...35k..low.income	0.3340284	0.0859753
household_income.35k...70k..middle.class	0.1774289	0.1176280
household_income.70...150k..upper.middle.class	0.8615833	0.0160141
race_ethnicityNot.Caucasian	0.0000000	0.8410260
X.Intercept.	0.0000000	-0.9284131

Running our logistic regression model on Biden voting outcomes, our model revealed that a different set of variables seemed to contribute to a voter’s decision to vote for Biden. In this model, all of the weight parameter values for race and degree of education were statistically significant ( $P < 0.01$ ). Household income values only seemed to play a role when an individual’s household makes more than \$150 000 annually. According to our model, the greatest influence on whether an individual votes Biden seems to be race, Individuals who are

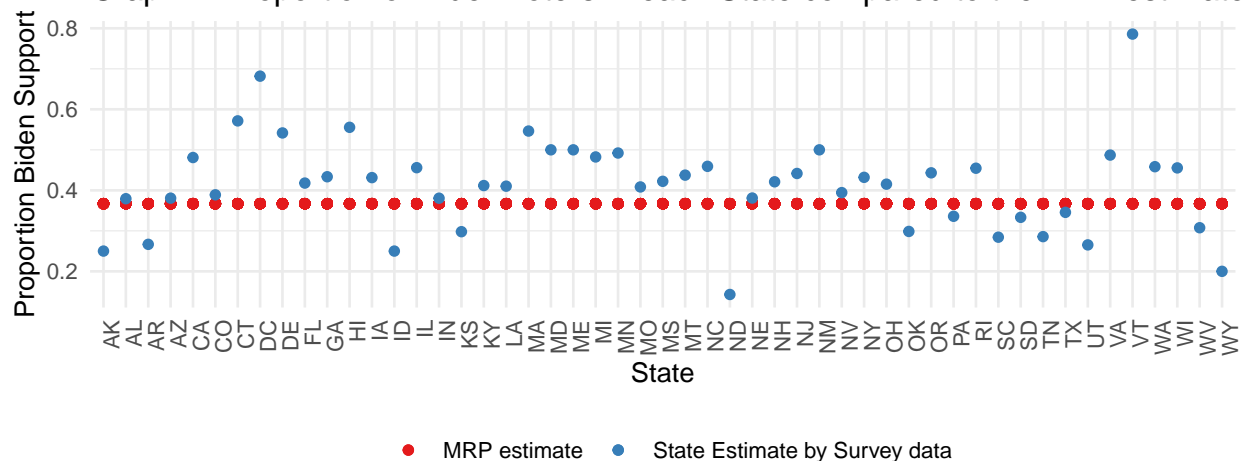
ethnically non-caucasian are more inclined to vote for Biden. Furthermore, the probability for an individual to vote Biden positively correlates with degree of education. Therefore, the subset of individuals who are non-white, have a graduate degree, and live in a middle-class family have the highest probability of voting for Biden.

Graph 1: Proportion of Trump voters in each State compared to the MRP estimate



The figure above represents the percentage of Trump supporters within each state recorded by the recent survey data. The red dots representing the popularity vote MRP estimate and the blue dots being the collected voting survey results from June 25th 2020. Trump support being highest in South Dakota and Kansas.

Graph 2: Proportion of Biden voters in each State compared to the MRP estimate



The figure above represents the percentage of Biden supporters within each state recorded by the recent survey data. The red dots representing the popularity vote MRP estimate for Biden and the blue dots are values from the voting survey results from June 25th 2020. Support for Biden being greatest in both Vermont and Washing DC. The model however seems to be slightly under predicting Biden's voting percentage.

```
## # A tibble: 1 x 2
##   alp_predict_trump alp_predict_biden
##   <dbl>             <dbl>
## 1      0.427         0.367
```

Our model predicts that Trump will win the 2020 election. Specifically, 42.7% of the population has decided

on voting Trump and 36.7% has decided on voting Biden. These values however don't add up to 100%, and thus it seems to suggest that the resulting 20.6% of voters are still undecided.

## Discussion

The results of our multilevel regression study concluded that Trump would be the winner of the 2020 election. However a great number of individuals are still undecided and recent campaigning efforts may change these results.

The greatest impact on whether an individual votes for Trump or Biden seems to be race, with white people favouring Trump and non-white people favouring Biden. Part of the reasoning for this difference seems to be the nature of Trump's campaigning. Trump has a tendency to be prejudiced against non-whites, with many of his slogans from his 2016 campaign such as "We need to build a wall" being blatantly discriminatory towards immigrants and minority groups. As a result, these influences may be why non-white voters have been less approving of Trump, and are more likely to vote for Biden.

Household income is also a significant factor when it comes to voting for Trump. As noted earlier in the model section, there is a positive correlation with income where richer people are more likely to vote for Trump. In contrast, with Biden, the only income value that is significant is whether someone has a family income of more than \$150,000; and if they do, then that has a negative impact on whether they'll vote for him. Possible reasonings for this discrepancy is due to the nature of republican belief and policies. Republicans are more conservative with money spending and funding. As a result they tend to introduce policies that are more lenient on taxation and thus, individuals with higher income, tend to prefer having more "right-winged".

Lastly, people who are more educated are more likely to vote for Biden. This could be due to Trump having a history of disregarding experts in their fields of people in higher level education. This could also be representative of his campaigning style, mainly revolving around using general emotion of his supporters to win them over, since people with higher levels of education may be more resistant to these techniques.

## Weaknesses

The sample and census data only gives us probability of whether the individual is voting for Trump or not and not if they are voting for Biden or neither. This may be helpful in determining who wins the next US election, however there are still some limitations in regards to our variables.

A major concern of our model is the use of family income as a representation of financial state. Income is important, but we do not get the full picture. For example, we do not know if the individual's "family" consists of a single individual, or someone who is married but with no children, or married with no children, etc. With this in mind, how the family income would be divided depends on the kind of family. For example, someone with children would have to pour a lot of their income towards childcare compared to someone who doesn't have children. In addition, the region where they live in is something to consider. While we know the state they live in, we do not know what sort of neighborhood they live in; for instance, whether they live in the city or suburban areas. Costs for households can vary heavily between the two areas. The choice in one's household could be influenced based on how many individuals are in a family, with bigger families requiring more of their income to go towards bigger living spaces.

## Next Steps

This calculation is based on popular vote but the winner of the presidential elections are not determined in this method and it is possible to win the popular vote but lose the election. The winner is determined by getting the majority of electoral college votes where there are 538 electors that vote based on the popular vote. Electors are divided among states but they are not divided equally which is why the voting outcome can be different from the popular vote. Our report used popular vote to determine voting outcome instead of electoral college votes so it may not be accurate. Since we didn't know much about the electoral colleges when writing this report, we couldn't use them to predict the voting outcome in this way. However in the future, using electoral college votes would be a more accurate method instead of the one that we used.

## References

- Blum, Benjamin. Understanding the Electoral College and How the U.S. President Is Really Elected | CBC News. 26 Oct. 2020, [www.cbc.ca/news/world/electoral-college-explainer-1.5768507](http://www.cbc.ca/news/world/electoral-college-explainer-1.5768507).
- Ruggles, Steven, et al. "U.S. CENSUS DATA FOR SOCIAL, ECONOMIC, AND HEALTH RESEARCH." IPUMS USA, Minneapolis, MN: IPUMS, 2020 (version 10.0), [doi.org/10.18128/D010.V10.0](https://doi.org/10.18128/D010.V10.0).
- Snider, Susannah. "Where Do I Fall in the American Economic Class System?" U.S. News & World Report, U.S. News & World Report L. P, 29 Oct. 2019, [money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system](https://money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system).
- Tausanovitch, Chris, and Lynn Vavreck. "Insights into the Beliefs and Behaviors of American Voters." Democracy Fund Voter Study Group, Democracy Fund + UCLA Nationscape, 10 Oct. 2019 (version 20200814), [www.voterstudygroup.org/downloads?key=41f75fff-0967-42e0-8e3c-d30e598bb539](http://www.voterstudygroup.org/downloads?key=41f75fff-0967-42e0-8e3c-d30e598bb539).
- "U.S. Census Bureau QuickFacts: United States." Census Bureau QuickFacts: United States, U.S. Department of Commerce, 2019, [www.census.gov/quickfacts/fact/table/US/PST045219](https://www.census.gov/quickfacts/fact/table/US/PST045219).
- Duffin, Erin. "Percentage distribution of household income in the U.S. in 2019", Statista. 9 Oct 2020. <https://www.statista.com/statistics/203183/percentage-distribution-of-household-income-in-the-us/>