

Election Prediction Report

Ashley Murray, Hunter Gregory, Matty Pahren, Nathan O'Hara, Scott Heng

10/29/2020

Introduction

Forecasting elections is of great interest, as politicians might election outcome models to adjust campaign strategies, economists might use it to predict how markets will react, and citizens might use it to decide whether to vote. The 2020 election is one of great importance, and it has seemingly gained even more attention than past elections.

In this report, we seek to model predicting the outcomes of several key races, specifically the presidential race, the US senate races, with special attention to the North Carolina senate race, and the North Carolina House race. First, we will describe our methods for all of the models, including a model of who votes in North Carolina, which aims to better the predictive power of our NC senate and House models. Next, we will discuss the results of our models. After that, we will walk through some of the limitations of our models. Finally, the appendices will include more detailed information about our modeling procedures and the data sources used.

Data

Considering the different predictions this study is aiming to make, we draw data from multiple sources. Our datasets include Fivethirtyeight senate polling data, Fivethirtyeight presidential polling data, North Carolina Voter Data from the North Carolina voter registration database and National Census Data. Based on the question, we will draw information on different datasets, which will be explained more thoroughly at each model further in the report.

By considering the presidential polls data, see can observe from Fig 1, we can observe how the probabilities of being elected for both candidates over time. We see that Biden has consistently had higher probabilities of being elected than Trump since April 2020, while both candidates experience similar degrees of increase in percentage during their 'Convention Bounce'. Biden appears to consistently have a mean percentage of being elected around 50% and up, and has had a recent surge during October. On the contrary, Trump's percentage estimates have largely been below 45% with a significant drop towards October. From the exploratory data analysis on national polls we can already see that Biden is projected to have an advantage over President Trump.

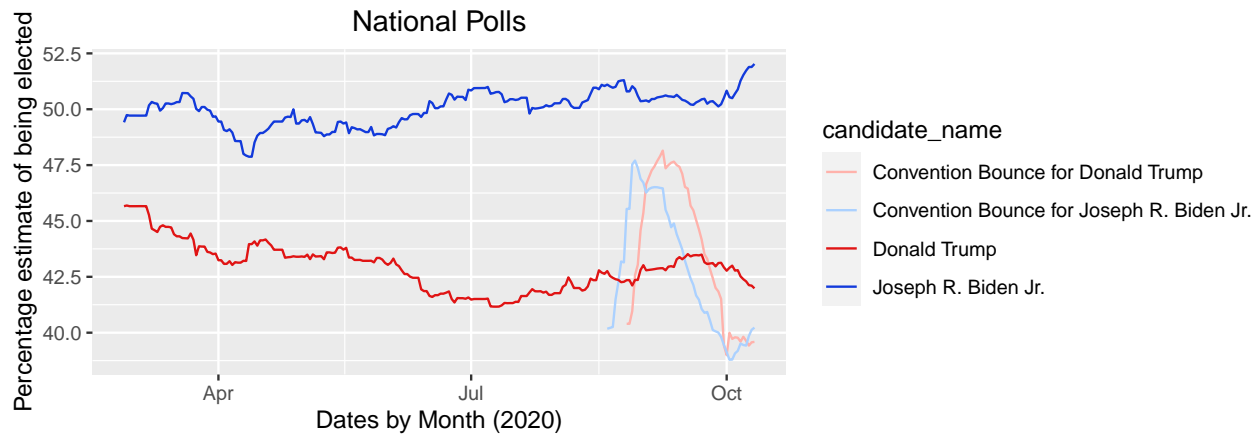


Fig 1.1 National Polls over time for Biden and Trump

However, this does not tell the full picture, as the United States Presidential Elections adopt an Electoral College system, which sometimes can allow a candidate to become President if that candidate did not receive the popular vote nationally, but secured support in key states that contain a large number of electoral votes. These states have a larger influence on the course of the elections and therefore are highly contested between the two candidates. Figure 1.2 shows the polls for both candidates by state over time, revealing which candidates have advantages over the other in certain states and which states require more attention in order to gain the people's support. Fig 1.3 shows the polls in swing states- States that have relatively equal support for either party and therefore can heavily influence the outcome of the election. For the majority of swing states, Biden similarly is favored statistically to be elected, with the exception of Iowa and Ohio.

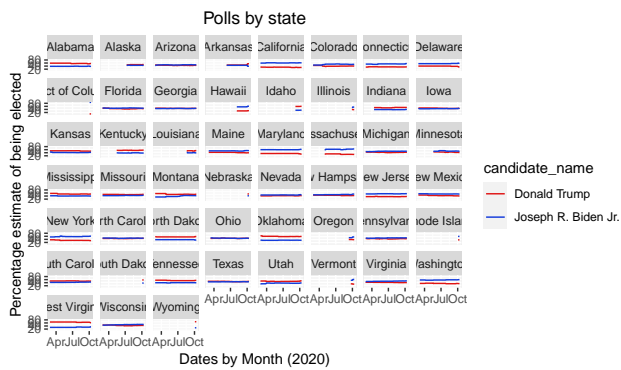


Fig 1.2 State Polls over time for Biden and Trump

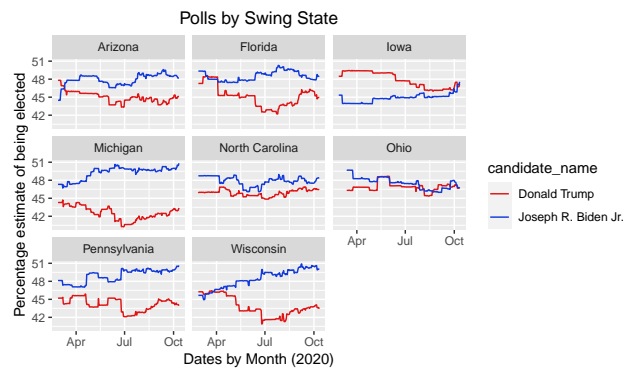


Fig 1.3 Swing State Polls over time for Biden and Trump specifically in the swing states (Democratic or Republican)

For senate data, we look at the senate races and their probabilities for being elected (Democrat or Republican). Over the years, we can observe that there are many close senate races as seen in Fig 1.4. In Fig 1.5, we can also see the percentages for being elected into the senate for individual candidates based on their party affiliation. We can observe that overall Democratic candidates tend to have high percentages for being elected into the senate over Republican candidates, and this distinction becomes more prominent closer to the election date.

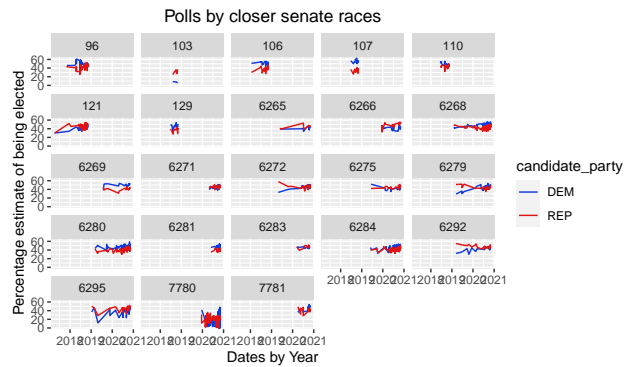


Fig 1.4 Polls by closer senate races

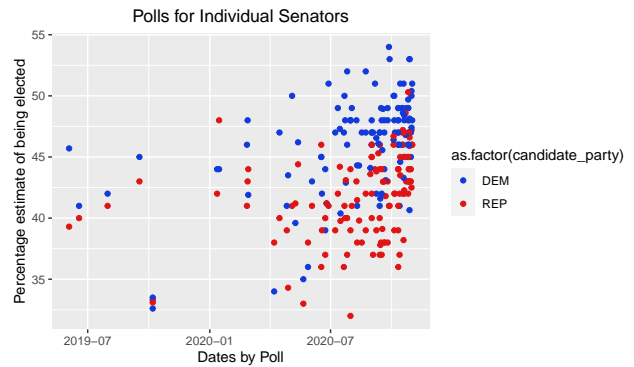


Fig 1.5 Polls for individual Senate races labelled by part

Methods

Who Votes Model

Electoral College Model

US Senate Model

North Carolina Senate Model

North Carolina House Model

Results

Electoral College Model Results

US Senate Model Results

North Carolina Senate Model Results

North Carolina House Model Results

Discussion

Appendix A: Models Used

Who Votes Model

Electoral College Model

Describe Linzer model in more detail? The purpose of this model is to predict the outcome of the presidential election by looking at electoral college votes. The probability of Trump winning is calculated by looking at the percentage of times in our 1500 *check number* simulations Trump gets 270 or more electoral college votes. In modeling the overall outcome of the presidential election, this model also predicts how each state's electoral college vote will go. Each state's electoral college outcome is modeled by looking at polling data

from fivethirtyeight for that state in order to predict what percent of the popular vote both Trump and Biden receive. Whichever candidate receives a larger percentage of the popular vote receives all of the electoral votes for that state. In reality, Maine and Nebraska can have split electoral college votes instead of the all-or-nothing approach other states take, but we chose to also model these states as all-or-nothing for simplicity sake, and many other election modelers take this same approach *possibly add citation?*.

add Beta estimates/ JAGS output for all 50 states

US Senate Model

Describe Linzer model in more detail?

$$\begin{aligned}
 y_k &\sim \text{Binom}(\pi_{i[k]j[k]}, n_k) \\
 \text{logit}(\pi_{ij}) &= \beta_{ij} + \delta_j \\
 \text{for } j > 1 : \beta_{ij} &\sim N(\beta_{i,j-1}, \sigma_\beta^2) \\
 \delta_j &\sim N(\delta_{j-1}, \sigma_\delta^2) \\
 \text{for } j = 1 : \beta_{i1} &\sim N(\text{logit}(h_i), s_i^2) \\
 \delta_1 &= 0
 \end{aligned}$$

Similar to our electoral college model, this model looks at senate race polling data from fivethirtyeight in order to predict the outcome of all current US senate races. The data was filtered to only include general election polling data, so in other words all jungle primary or runoff election polls were removed. Additionally, the data was filtered to only include polls less than a year out from the election. Lastly, polling data for third-parties was filtered out with one exception. Ricky Dale Harrington was an independent, but he was recoded to be a Democrat in our dataset. We did this because the few independents currently in Congress lately tend to vote with Democrats, and he was the only Independent in the dataset that was predicted to get a majority of votes in one of these senate races. After filtering, we were left with data for all states where senate elections were being held except for Arkansas, Rhode Island, South Dakota, West Virginia, Wyoming, Louisiana, and the second Georgia Race. However, most of these races were predicted to be blowout races, with Arkansas, South Dakota, West Virginia, Wyoming, and Louisiana deemed as safely Republican and Rhode Island deemed safe Democrat. Since this is the case, we counted these states as a corresponding Republican or Democratic win in each of our simulations. Georgia was the only race we were missing data for that seemed close, but we chose to use the Georgia predictions generated from the other Georgia race we did have data on, since senators are elected by the whole state and it is likely that the same people are voting for both senators. After making predictions for each individual race, we aggregate the results and look at the percentage of times where Republicans win more than 21 seats, where 21 is the number of seats Republicans need to win in order to have a 50/50 Republican/Democrat split in the senate.

add Beta estimates/ JAGS output

North Carolina Senate Model

North Carolina House Model

Appendix B: Data Sources