A Bayesian Model for the Prediction of United States Presidential Elections

Abstract

Using a combination of polling data and previous election results, Nate Silver successfully predicted the Electoral College distribution in the presidential election in 2008 with 98% accuracy and in 2012 with 100% accuracy. This study applies a Bayesian analysis of polls, assuming a normal posterior using a conjugate prior. The data were taken from the Huffington Posts Pollster. States were divided into categories based on past results and current demographics. Each category used a different poll source for the prior. This model was originally used to predict the 2016 election, but later it was applied to the poll data for 2008 and 2012. For 2016, the model had 88% accuracy for the 50 states. For 2008 and 2012, the model had the same Electoral College Prediction as Nate Silver. The method of using state and national polls as a prior in election prediction seems promising and further study is needed.

1 Introduction

A presidential election happens in the United States every four years. The winner of the election is decided by the 538 electors in the Electoral College. Under the current system, most of the states choose their electors based on the winning candidates of the state. Most of the states (like Delaware and West Virginia) have voted for the same party in most recent elections. These states are called "red" if they consistently vote Republican and "blue" if they consistently vote Democrat. But in some states, like Florida, the winner is not consistent, and the state occasionally "swings" from one party to another. The key to winning a presidential election is winning as many of these swing states as possible.

The outcomes of presidential elections have been predicted by political scientists and statisticians for decades. Some of the first models used economic and political data to predict the winner of the popular vote. The challenge is to predict the outcome in swing states, which ultimately win a candidate the election. The quality and quantity of state level polling data has increased over time. All states in the 2016 election had at least 2 polls, while in the 2008 and 2012 election, some states had no scientific polls. Poll data are easily accessible through the Huffington Posts Pollster, which provides csv files of polls in individual states. This increase in data has allowed for more accurate predictions of the voting results.

This study was based on methods discussed by [1] for use as an undergraduate research project on presidential elections. While numerous models have been studied as methods to predict presidential elections, the idea of a Bayesian approach using the poll data of other similar states as priors has not been pursued as rigorously. The approach proposed by [1] used previous elections as a prior, but the differences in candidates from different election cycles means that some voters may not vote for the same party as they did in previous elections.

Another problem is the limited information on state level support. As discussed in [4] & [5], estimations of state level support between candidates is difficult due to limited polling information on a state by state level. While the number of polls in individual states has increased over time, there is still a lack of data in smaller states. States like West Virginia may only have one or two polls. To help address this problem, states with similar demographics were used as priors for this model.

The greater availability of polling data has now made it possible for models to use polls to make relatively accurate estimations of the support of the two major candidates in individual states. The csv files provided by Pollster have provided the opportunity for more research in Presidential Election prediction. Because of these advances, it was possible for a model to be created by an undergraduate.

For purposes of simplicity, the data were assumed to be normally distributed. A normal conjugate prior was used for the means, resulting in a normal posterior. This is obviously not ideal considering the non-parametric nature of polls, but this method greatly simplified the process and made it possible for an undergraduate with a limited mathematical statistics background to implement. The goal of this study was to test if the inclusion of current poll data from other areas with a Bayesian analysis would help adjust for limited information in states with little poll data.

2 Methodology

The guidelines and methodology of the experiment were set in September 2016 to allow adequate time to prepare the computer programs necessary to run the analysis. The main focus was on the 2016 election, but a later analysis was performed on the 2008 and 2012 elections to get more data on the accuracy of the method. Pollster, run by the Huffington Post, contains poll results available in csv format for 2012 and 2016. Since Pollster provided the results in an easy to analyze format, they were chosen as the data source for the analysis. For all elections, polls were only considered if they were conducted after July 1st of the election year. For the 2016 election, the polls were pulled on November 4th for red and blue states and November 5th for the swing states for analysis. The deadline for posting predictions was November 5th, 2016 at 11:59 central time. The prediction became final at 8:48 central time. For the analysis of 2012, polls were considered if they occurred after July 1st and were released by 12 pm on the Saturday before the election. In 2008, the release dates of the polls were not available. Polls from 2008 were included in the analysis if they were conducted between July 1st and October 31st. However, for 2008 and certain states in 2012 and 2016, Pollster did not have csvs available and csvs had to be manually created. A few states in 2008 and 2012 only had polls from before July 1st, in these cases those polls were included. Other states had no polls available and data from other states had to be used.

A series of simplifying assumptions were made about the election to make the analysis easier. It was assumed that, over the campaign, people who reported themselves as decided would not change their minds, and therefore, all polls within the timeline of analysis could be treated equally. It was decided to analyze the data parametrically instead of using non-parametric methods because of the differences in polling size and methods. A parametric approach would also be far easier to compute and implement, and given the time constraints of an undergraduate project, this was the best option at the time. The results were then normalized after all calculations so that the total of the probabilities equaled 1. The level of undecided voters in the polls included in the analysis did not change much over time, but there were slightly more undecided voters in the early polls. Normalizing before the analysis would require more calculations and computation time, and the prediction for 2016 had to be conducted in less than 36 hours. A change could have been made in the analysis for 2008 and 2012 to include methods like normalization before the analysis. This change was not done because the goal was to test the performance of the model under similar circumstances.

As discussed previously, this model utilizes a Bayesian approach using the poll data of other similar states as priors. The model predicted the final percentages of the votes within each state. The python programs calculated the mean and variance for the polls from the prior state, and the mean and variance of the polls from the state being predicted.

Formula 1: Prediction Calculation

Given the sample mean $\bar{x}_0$ and sample variance s_0^2 of the polls of the state used as the prior, and the sample mean $\bar{x}$, number of polls n, and sample variance s^2 of the polls from the state being analyzed, the posterior distribution is a Normal distribution with updated mean and variance parameters as calculated as follows:

$$\mu_{new} = \frac{s^2}{n * s_0^2 + s^2}\bar{x}_0 + \frac{n * s_0^2}{n * s_0^2 + s^2}\bar{x}, \qquad \hat{\sigma}_n^2 = (\frac{1}{s_0^2} + \frac{n}{s^2})^{-1}$$

Included below are the tables of the inputs and outputs of the model for the states of Colorado and Wyoming in 2016. The table expresses variability in standard deviations instead of the variance because standard deviation has the same units as the mean. Only the posterior column is standardized for the sums to equal one. This was done because it was a much faster calculation. Since the polls are treated as independent and identically distributed (i.i.d) random variables, the posterior standard deviation is of the form $\frac{\sigma}{\sqrt{n}}$. This is why Colorado which had 36 polls, has a much smaller standard deviation than Wyoming which had 4 polls. This caused a great underestimation of the variability of the predictions made by the model.

Table 1: Sample Prediction for Colorado in 2016

Candidate	Prior	Data	Posterior
Clinton Mean	0.431	0.444	0.471
Clinton SD	0.03	0.027	0.005
Trump Mean	0.397	0.391	0.415
Trump SD	0.029	0.035	0.006
Other Mean	0.039	0.059	0.059
Other SD	0.017	0.052	0.008
Johnson Mean	0.073	0.048	0.054
Johnson SD	0.023	0.046	0.008
Number of Polls	219	36	

Table 2: Sample Prediction for Wyoming in 2016

Candidate	Prior	Data	Posterior
Clinton Mean	0.346	0.253	0.315
Clinton SD	0.025	0.046	0.017
Trump Mean	0.511	0.608	0.624
Trump SD	0.03	0.053	0.02
Other Mean	0	0	0.001
Other SD	0.001	0.069	0
Johnson Mean	0.041	N/A	0.06
Johnson SD	0.029	N/A	0.029
Number of Polls	7	4	

As shown in the tables above, the posterior gets closer in distribution to the data as the number of observations (polls) increases. Since the polls were treated as identically distributed independent random variables, the standard deviation decreases with each new observation.

The true mean and variance of the polls are unknown and point estimators $\bar{x}$ and σ^2 were used to approximate the mean and variance. The Gaussian conjugate prior is simpler than the binomial conjugate prior. Polls are probably not independent and that is one of the assumptions of the Gaussian conjugate prior. The number of polls in the state varied widely, but was rarely above 30, and ideally one would not invoke the Central Limit Theorem in a case like this. From a theoretical perspective this model is not the ideal approach. However, this model is a very simple calculation and the predictions needed to be made quickly.

The five different priors were: national polls for swing states, Texas for southern red states, Nebraska for midwestern red states, California for western blue states, and New York for northern blue states. The prior data usually had more polls than the state being predicted. The prior data often had higher-quality polls than other states in that category. Using more than five categories may have worked better, but the limited choices made choosing a prior easier. The states that served as priors were assumed to have more information by election day than the state they would be applied to. There were a few states that were not clear which prior should be used. For red states, the decision on whether a state was in the midwest or south was primarily based on which state was more similar culturally and demographically, Texas or Nebraska. The addition of another category for the southeast red states and the Midwest blue states may have worked better, but the small amount of choices in priors made decisions easier. Included below is a map of the United States color coded to reflect the prior used for that state.

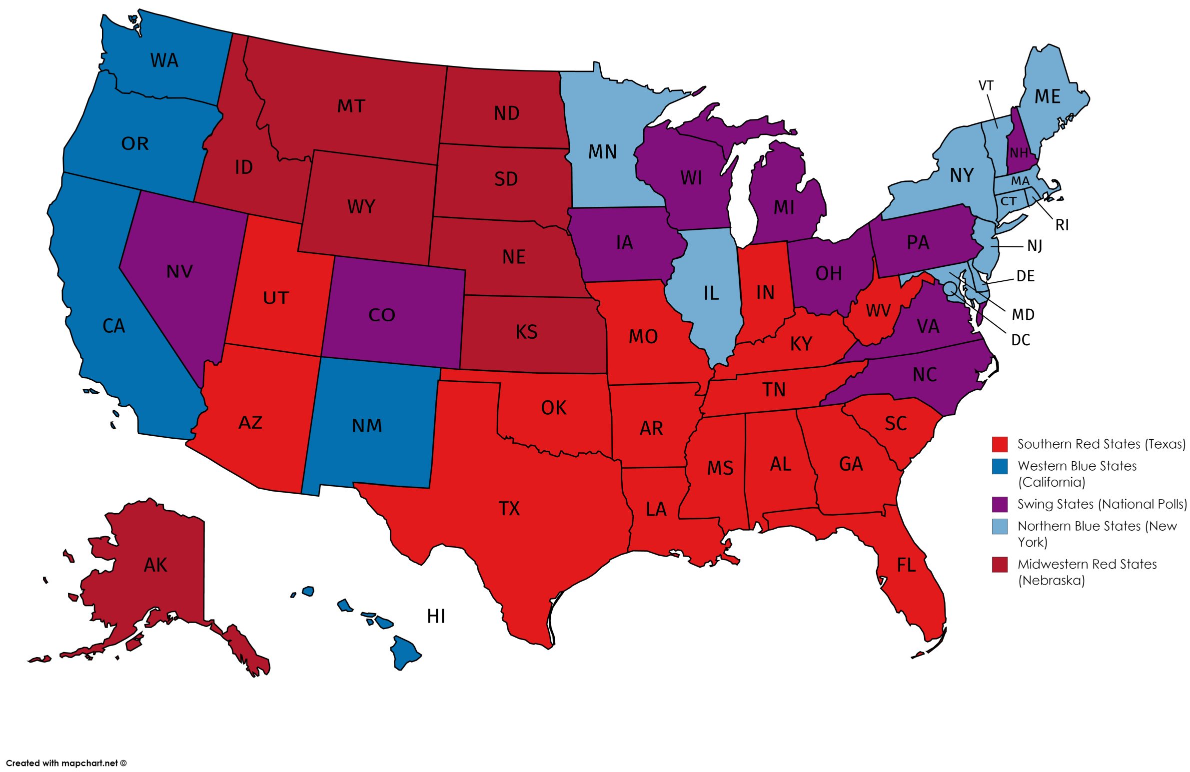

Figure 1: Map of the Priors Used for Each State (Created with mapchart.net)

States were chosen as swing states if the presumed winner was not clear in advance, or if the state did not have a clear leading party that controlled most of the government offices in that state. A few swing states were temporarily considered swing states for a single election cycle if that state did not appear to be voting as usual. The following thirteen states were considered swing states for at least one election in the analysis: Arizona (2016 only), Colorado (2008, 2012, 2016), Florida (2008, 2012, 2016), Iowa (2008, 2012, 2016), Indiana (2008 only), Ohio (2008, 2012, 2016), Nevada (2008, 2012, 2016) New Hampshire (2008, 2012, 2016), North Carolina (2008, 2012, 2016), Michigan (2008, 2012, 2016), Pennsylvania (2008, 2012, 2016), Utah (2016 Only), Wisconsin (2008, 2012, 2016).

In Indiana in 2008, President Obama got a surprising amount of support considering the strong red voting history of that state. In 2016 in Arizona, the increased registration of Hispanics and young voters coupled with close polls lead to the decision to consider Arizona as a swing state. Arizona still used Texas as its prior, but Arizona got more attention because of its potential swing state status. In 2016 in Utah, an independent candidate, Evan McMullin got as high as 30% in some polls in October, and it was decided that the strong support for an independent candidate could mean that Trump (the presumed winner) may not win in Utah. This led to the decision to make Utah a swing state in 2016. Although Utah was considered a swing state, it did use Texas as its prior since Utah was not expected to vote like the popular vote. It is important to note that there is not a clearly defined universal definition of a swing state, so this decision is somewhat subjective.

The use of poll data from sources outside the state was chosen because most states did not have polls from a variety of agencies. However, there were states that with better polling that were sufficient approximations of the electorate in the state they were being applied to. For swing states, the popular vote usually is a good estimation of the votes in swing states. This choice of prior provided a much larger data set. There were over 200 polls used in the prior for swing states for the 2016 election. The number of polls in the prior data was not a factor in the calculation, but larger data sets do tend to be more stable and accurate than smaller data sets.

Since the normal conjugate prior approaches the distribution of the data as the number of observations increases, the prior did not need to be an exact approximation if the state being analyzed had sufficient poll data, like in the case of Texas. Some of the larger states like Georgia and Illinois have almost as much poll data as the state used as the prior, and the posterior distribution is similar to the distribution of poll data from that state. As seen in Table 1 and Table 2, the prior distribution had a stronger influence in Wyoming which had 4 polls than in Colorado which had 36 polls.

It was also assumed based on the early poll data that Gary Johnson would get at least five percent of the vote in 2016, and therefore should have a predicted mean for all states. If a candidate gets 5% of the vote nationally their party is considered a major party. A minor candidate is a candidate that did not get 5% of the national vote on election day. The inconsistency in the inclusion of other candidates in polls from all election years made estimating support for other candidates difficult. For 2008 and 2012, an analysis of minor candidates was not included because the combined votes for third party candidates were less than five percent. Unlike 2016, there was not a consistent pattern of polls above 5% for any third party or independent candidate. Like 2016, the data for 2008 and 2012 did not have consistent information on the support of minor candidates. If an analysis on the minor candidates was conducted, it probably would not have been more accurate because most polls had no data on the minor candidates meaning the mean of other polls would be around 0. Not predicting the support of minor candidates may not be ideal, but it was practical.

Johnson did not get 5% of the vote in 2016. Over time his poll numbers began to slowly decrease. Although Johnson was treated like a major candidate in the analysis, he is considered a minor candidate in the rest of this paper. Since it was assumed that Johnson would be a major candidate, a method to approximate his support was necessary. However, the poll data on Johnson was limited and a measurement called the "Other-Johnson Factor" was created. The Other-Johnson Factor measured the percentage support of non-major party candidates (not Clinton or Trump) that was for Johnson. If the prior state had data on Johnson (i.e. California) then the Other-Johnson Factor was based on the prior state.

Formula 2: Other-Johnson Factor
Let J be the average poll support for Johnson, and O be the average poll support for other candidates besides Trump, Clinton, and Johnson. Then

$$\text{Other-Johnson Factor} = \frac{J}{J+O}$$

In most cases, the Other-Johnson Factor was based on national polls. For the swing states, only national polls with Johnson data were included. In October, Evan McMullin, an independent candidate, appeared in some polls in Utah at around 20% to 30%. At this point, it was decided that

Evan McMullin should be included in the analysis of Utah using the same method as the Other-Johnson Factor. Evan McMullin did not have enough support in any other state to be considered as a major candidate in that state. In the case of the western blue states, where the Other-Johnson Factor was based on the poll data of California, the presence of polls without minor candidates lowered the average significantly and caused an underestimation of Johnson in the western states. In Nebraska, there was a very limited amount of information on other candidates, and this made synthesis of other support difficult in the midwest. In 2016, if other support was not specified it was calculated by subtracting the sum of Trump's support, Clinton's support, and undecided voters from 1. A poll respondent can only support Trump, Clinton, or another candidate, or be undecided, making the total probability of these numbers 1.

The predictions of the election were not the only element of this study. Another key element was the blog. A blogger account was used to provide both a public record of prediction and a forum for discussion on the roles of mathematics and statistics in the campaign. In addition to the model and blog, the events of the election were closely monitored. Debates and press conferences were watched. News articles from multiple organizations were studied. The goal was to understand as much about the process of the Presidential Election as possible. Books on political science and behavioral economics were read to try to understand human behavior during an election. A post was made weekly on Friday about the events of the week and their possible effects on the election. These commentary posts helped track the events of the election and provided a source for outreach to the general public about the role of statistics and mathematics in the election and everyday life. No previous posts were ever edited later in the process, regardless if they had any predictions or not. Revisions were noted in the comments. Transparency was a key factor in the process and personal political opinions of the author were disclosed.

The inclusion of multiple elements in this study was done to create a comprehensive approach to the analysis of an ongoing election. If this study were done primarily retrospectively, and over a longer time it would have been easier to try multiple advanced methods. But a prospective approach on the 2016 election provided a chance to try a new idea and follow the process in real time. The analysis of the elections of both 2008 and 2012 did not have the opportunity for opinions about the behavior of the states to be created before the results were known and were more subject to possible bias in the choosing of priors.

3 Results

The 2016 election proved to be difficult to predict. Most models did not predict a Trump win. Florida, Michigan, North Carolina, Pennsylvania, and Wisconsin were frequently missed in the predictions of multiple companies. The tested model called 88% of the states, missing Florida, Ohio, Michigan, North Carolina, Pennsylvania, and Wisconsin. Since all polls were weighted equally regardless of the date of the poll, the model did not adapt well to trends over time. As Election Day approached, Trump began to have a small lead in the polls in Ohio. But since the tested model weighted all polls equally, the polls with Trump leads did not raise his average enough for the model to predict a Trump win. The five frequently missed states were close with margins less than 1% in Pennsylvania, Michigan, Florida, and Wisconsin. The exact reasons for the errors in the polls and models are unknown at this time. Polling and election prediction models should be examined in light of this election. It is possible that the 2016 election was an outlier that may have been unpredictable. As discussed previously the model was applied to the 2008 and 2012 elections, using poll data with the same poll inclusion criteria as the 2016 model.

Below is a table comparing the accuracy the various models (including the tested model) of predicting the winning candidate in each of the fifty states plus Washington DC (if available).

Table 3: Percentages of Winners Called Correctly

Race	RCP[1]	PEC[2]	538[3]	PW[4]	Tested Model
2008 Accuracy	0.96078	0.98039	0.98039	N/A	0.98039
2012 Accuracy	0.98039	0.98039	1	0.98039	1
2016 Accuracy	0.92157	0.90196	0.90196	0.90196	0.88235
Average Accuracy	0.95425	0.95425	0.96078	0.94118	0.95425

[1] Real Clear Politics: realclearpolitics.com
[2] Princeton Election Consortium: election.princeton.edu/
[3] Five Thirty Eight Polls Plus Model: fivethirtyeight.com
[4] Predict Wise: predictwise.com/

All models achieve highly similar accuracy in predicting the winners in each state. In 2008 and 2012, the tested model had the same predictions of the winning candidates as Five Thirty Eight. All models achieve highly similar accuracy in predicting the winners in each state. Measurements such as Brier scores or root mean standard error help to provide more clarity in the true accuracy of predictions by measuring the accuracy of the probabilistic and mean predictions. The tested model predicted the distribution of votes among the major candidates, to create a prediction for the voting percentages. To compare the accuracy of mean prediction the root mean standard errors (RMSE), were calculated for 2008, 2012 and 2016 for the tested model, Five Thirty Eight's model and the Real Clear Politics Average for all swing states. The Real Clear Politics Average did not always sum to 1, but the numbers were not normalized before the RMSE analysis. The RMSE was calculated based on the error in percentage points where an RMSE of 2 would represent 2%. States were weighted equally, and Washington D.C. was treated like a state for the purpose of this calculation.

The root mean square error helps provide a comparison the accuracy of a models prediction of the final vote percentages. If the model normalizes probabilities to 1, the sum of overestimations equals the sum of underestimations.

Formula 3: Root Mean Square Error Calculation
The RMSE was calculated in an excel document with the following formula:

$$RMSE_{model} = \sqrt{\frac{\sum_{i=1}^{n_{states}} \left(\frac{|ra_i - rp_i| + |da_i - dp_i| + |la_i - lp_i| + |oa_i - op_i|}{2} \right)^2}{n_{states}}}$$

Where r_{a_i}, d_{a_i}, l_{a_i} o_{a_i} are the actual votes for the Republican, Democrat, Libertarian (2016 only), and other candidates in state i, respectively, and r_{p_i}, d_{p_i}, l_{p_i}, o_{p_i} are the predicted votes for the Republican, Democrat, Libertarian (2016 only), and other candidates in state i, respectively. Note that the terms l_{a_i} & l_{p_i} are not used in 2008 and 2012 where the Libertarian party was not considered to be a major party by predictors.

Referring to the sample prediction from above: the predicted outcome of the tested model for Colorado in 2016 was 47.1% for Hillary Clinton, 41.5% for Donald Trump, 5.5% for Gary Johnson and 5.9% for other candidates. The actual voting result was 48.2% for Hillary Clinton, 43.3% for Donald Trump, 5.2% for Gary Johnson and 3.3% for minor candidates. This would give the prediction for Colorado in 2016 a RMSE of 1.703. The RMSE was calculated for each individual state for the three models being compared. Then, the RMSE of all of the predictions by the particular model were calculated for each election year. Since predicting the exact voting results is more important in swing states where the winner is less predictable, an analysis on the accuracy of swing state prediction was also conducted. Arizona was not considered a swing state in the RMSE calculated for 2016 because the assumption the race was closer than usual was wrong. Utah was also not included in the swing state RMSE because it Trump won Utah with a large margin. In all years, Real Clear Politics did not have a complete data set on the states so only the swing states were analyzed. This is not the only way to measure the accuracy of mean prediction. For example, the RMSE can be calculated for the relative margins between candidates. This method was chosen because it can be used to provide an estimate of the margin of error of these models.

Table 4: Root Square Mean Error for Various Models

Race	TM[1]	TM SS[2]	RCP[3] SS[2]	538[4]	538 SS[2]
2008 All Candidates	3.54740	3.14788	4.23389	3.19332	1.66958
2008 2-Party	2.89669	2.57051	3.63513	3.03050	1.47846
2012 All Candidates	3.25139	1.94492	2.33511	2.38019	1.29790
2012 2-Party	2.37053	1.17163	1.61076	1.98642	0.93420
2016 All Candidates	6.82013	3.99585	3.32952	5.37952	3.56511
2016 2-Party	3.99534	3.14325	2.04295	3.81296	2.31948
All Candidate Average	4.52876	3.01755	3.299507	3.65101	2.17753
2-Party Average	3.07768	2.29513	2.42961	2.94329	1.57738

[1] SS: Swing State
[2] Real Clear Politics: realclearpolitics.com
[3] Five Thirty Eight Polls Plus Model: fivethirtyeight.com

An estimate of a prediction for only the two major candidates was made by taking the estimates for the Republican and Democrat, and then proportionally normalizing that support to equal one. Since the vast majority of votes goes to either of the two major parties, predicting the performance of those candidates relative to one another is important. All of these errors are large enough for it to be reasonable to expect models to not always call the winner of the election.

Predicting the support of third-party and independent candidates was difficult and was often underestimated by the model. National polls may include the minor candidates in their questions about support, but the state polls often did not ask about the minor candidates. The support for minor candidates is often small, and in most states the winner is clear. This lack of information on minor candidates did not affect the model's ability to call a winner, but most of the errors in prediction were related to an underestimation of minor candidates. Since some polls do not include minor candidates this lowered the average of the support of those candidates, which sometimes caused an underestimation of minor candidates. Johnson got under five percent of votes nationally, and Johnson was only included in the analysis because of an assumption he would get at least five percent of the national vote on Election Day. The Other-Johnson Factor did not work very well because in some national polls Johnson was the only minor candidate included. This meant that in the final prediction the Other-Johnson Factor was over .9999, thus implying Johnson would get 99.99% of all the votes for the minor candidates. However, Johnson did not receive 99% of the votes for minor candidates and the other minor candidates were underestimated. In some states the only candidates on the ballot were Trump, Clinton, and Johnson so this underestimation did not matter much, but in other states, particularly those where Jill Stein (the Green Party candidate) was on the ballot, this underestimation lowered the accuracy. It may be better to ignore minor candidates and instead predict the levels of support in a two-party race. A two candidate prediction approach appears to be the most accurate way to predict and compare the support of the two major candidates.

In states with limited poll information, the errors were larger, but these states often had a clear leader so polls were not needed to be conducted as often as closer states. For example, in West Virginia, the Republican candidate had over a five percent lead in all three elections studied. West Virginia is known to be a red state, but the tested models prediction of West Virginia in 2012 was seven percent off. However, the model still called the winner so this error did not affect the accuracy of predicting the Electoral College. But in swing states where the winner may be decided by a percentage within the past root square error of the means for swing states, accuracy matters more. The inaccuracies in predicting the winner of the 2016 election mattered more than in 2012, where the swing states were not as closely decided. Only one state in 2012 was decided by less than one percent, but in 2016 this happened in four states. Given the unusually large size of the errors in poll-based models for the 2016 election, it is possible that the poll data itself could have been bad. The 2016 election may have been unpredictable given the poll data, but that does not mean the models can not be improved. The Five Thirty Eight model did the best at predicting means probably because it used methods to address non-response and polling agency bias. Included below is a table of selected two party predictions for 2016 with actual two party results and two party RMSEs for individual states. A two party analysis better showcases the accuracy of a model, because it ignores minor candidates, which are both difficult to predict and are only a small portion of the vote. The table includes all swing states, the red and blue state with the highest error (West Virginia & Hawaii), the red and blue state with the lowest error (Texas & Massachusetts), and two close partisan states (Arizona & Minnesota). Wyoming is also included since it was used as an example prediction.

Table 5: Selected Two Party Predictions, Results, and RMSEs for 2016

State	Trump Actual	Trump Predicted	Clinton Actual	Clinton Predicted	2 Party RMSE
West Virginia	0.721	0.62	0.279	0.38	10.1
Texas	0.547	0.54	0.453	0.46	0.7
Arizona	0.519	0.508	0.481	0.492	1.1
Hawaii	0.312	0.354	0.688	0.646	4.2
Massachusetts	0.353	0.352	0.647	0.648	0.1
Minnesota	0.492	0.449	0.508	0.551	4.3
Wyoming	0.757	0.664	0.243	0.336	9.3
Colorado	0.473	0.468	0.527	0.532	0.5
Florida	0.506	0.485	0.494	0.515	2.1
Iowa	0.551	0.507	0.449	0.493	4.4
Michigan	0.501	0.461	0.499	0.539	4
Nevada	0.487	0.492	0.513	0.508	0.5
New Hampshire	0.498	0.468	0.502	0.532	3
North Carolina	0.519	0.489	0.481	0.511	3
Ohio	0.543	0.497	0.457	0.503	4.6
Pennsylvania	0.504	0.466	0.496	0.534	3.8
Virginia	0.471	0.455	0.529	0.545	1.6
Wisconsin	0.504	0.469	0.496	0.531	3.5

There is a great deal of variability in the accuracy of predictions. The table includes all swing states, the red and blue state with the highest error (West Virginia & Hawaii), the red and blue state with the lowest error (Texas & Massachusetts), and two close partisan states (Arizona & Minnesota). Wyoming is also included since it was used as an example prediction. Trump was usually underestimated by the model because polls regularly underestimated Trump. This consistent underestimation of Trump is probably the cause of the higher modeling error in 2016.

Considering that the tested model was simplistic, the addition of additional considerations like adjusting for bias and normalization before analysis may have increased accuracy. Further study is needed to try the use of data from inside the election process to both address non-response and polling agency bias. Considering the limitations of a short-term undergraduate research project, the tested model was relatively successful. The tested model had comparable success in predicting winners. Further study on this approach is needed to more definitively determine the viability of this method.

4 Discussion

It is important to interpret this study in context. As an undergraduate research project, there were knowledge and time limitations with this model. The project started in August of 2016, with the goal to have a fully functional model by late October. The model needed to predict the outcome for all 50 states and Washington D.C. quickly to allow time for the predictions and an commentary on the predictions to be posted by the November 5th deadline. The simplifications and assumptions made the study possible. The model may be relatively simplistic compared to models discussed in [1], [2], [3], and [4], but it performed well at predicting a two-party race. Polls are not completely independent in practice and opinions change over the campaign. This version of Gaussian conjugate prior assumes that the observations are independent random variables. The Gaussian conjugate worked quite well in predicting the winner, but it greatly underestimated the variability of polls. While the tested method was not designed to primarily create a probabilistic representation of winners in the states, the credible intervals were narrow and frequently failed. Looking forward to future elections, the primary author plans to try more complicated models that are currently unfeasible for her at this time. A possible model would be a binomial multilevel time series Bayesian model, with adjustments for poll bias to predict two party support. The nature of American Presidential elections is that they only happen once every four years. This makes perfecting a model difficult because, while a model can be tested retrospectively at any time, it may not be able to be applied to a future election for as many as four years.

The opinions of the American electorate are constantly changing and the 2016 election exposed the limitations of models that worked for previous elections. By the Central Limit Theorem, as the number of polls approaches infinity the distribution normalizes. This normalization may not be happening fast enough in the context of a presidential campaign. Other distributions may work better for poll analysis. It is important that predictors understand that, while they may be able to predict the winner, estimates of the mean and variability are usually inaccurate in simplistic models like that of Real Clear Politics. While the errors in predicting voting results always existed, 2016 was the only year where the errors caused predictors to inaccurately predict the winner of the election. The problem was not necessarily that the error in predicting means was significantly larger, but rather that the error affected the prediction of the winner.

The major drawback of the Five Thirty Eight model is that it is proprietary, and therefore, not available to the public. While it seems to be more statistically sound compared to other models, its actual methodology is unknown. The Princeton Election Consortium model is open source, but it is based on assumptions that seem not to hold. The lack of multiple good open-source models for presidential elections makes the academic study of polling analysis by statisticians important work. There are numerous studies based on a fundamental model using data from other sources, such as the economy, as discussed in [2]. Fundamental models have been successful at predicting the winner of the popular vote. Most of the time, the winner of the popular vote wins the Electoral College, but there are cases like 2000 or 2016 where the winner of the popular vote does not win the presidency. Polls are probably improving in both their methods and volume. Even across the elections analyzed in this study the number of polls began to increase over time. The difficulties in creating a representative poll means the poll data is subject to non-response bias. Polls may have flaws, but the integrations of multiple approaches to address possible biases can improve accuracy.

The model did not do well at predicting minor candidates. For the analysis of 2008 and 2012, other candidates were ignored because there was not enough information from the poll data to make a meaningful prediction. In 2016, the Other-Johnson Factor proved to be unsuccessful for states using California as the prior. Californias poll data initially had Johnson as an option regularly, but polls with Johnson decreased over the campaign. It appears that the Five Thirty Eight model may have some way to predict minor candidate support outside of using state-level poll data. The Five Thirty Eight model did much better at predicting minor candidates than the tested model. However, minor candidates are only a small proportion of the overall vote. Since poll data does not always include minor candidates, it may be more practical to predict two party support and relative margins. A two-party approach could use the binomial distribution which is much easier to work with than a multinomial distribution.

A comparison of the root mean square error of two-party predictions for both the 2012 and 2016 elections for Five Thirty Eight and the tested model shows that the tested model was 95.329% as accurate than the Five Thirty Eight model. Considering the relative simplicity of the model, the model performed very well overall. It is important to note that while the tested model is almost as accurate as the Five Thirty Eight at predicting a two-party race, the tested model is not good at predicting the support of minor candidates.

The choice of national polls as the prior may have affected the accuracy in the swing states. While the model had good performance overall compared to other models, the model's performance in swing states was relatively poor. The use of national polls to represent swing states felt natural. It made more sense to use national polls for the swing states since they vote relatively close to the popular vote than any other state. A similar Bayesian model could not be found, so it was unknown how this model would perform. It was known that national polls would not be a perfect fit. After a more thorough examination of the assumption that swing states vote like the national vote, it was concluded that national polls should not be used as the prior. Swing states are the most important states to predict, and the method failed in the area that mattered the most. The tested model performed better than the Real Clear Politics average in swing states, which shows the model is possibly better than simple poll averaging. However the model performed very well in other states, and shows promise as a method for determining relative support in partisan states.

A possible application of this method would be Senate and House elections. Senate and House elections may not have a lot of polling information. The number of polls varies widely based on the particular race. Some areas get more polls than others, and an approach like the one this model used could help improve the prediction of Senate and House races. Priors would have to be chosen to reflect multiple factors, including incumbency advantage, demographics, and the political lean of that area.

The limitations of poll-based election prediction is that there is currently a low number of elections with easily accessible poll data. There have been promising results with a method that uses a fundamental model to predict the winner of states, as discussed by [3] and used by PredictWise. The fundamental approach has produced similar results to poll-based methods for the past three elections, but it is difficult to tell if a fundamental based model is truly a better method of predicting state-level support. The definition of accuracy for Presidential election prediction is difficult to distinctly define because accuracy can be determined by Brier scores, root mean square error, or correct prediction of the Electoral College. It is difficult to distinguish which models are more accurate than others. It does appear, though that the Real Clear Politics average is not a good method to predict exact voting results.

The answer to predicting elections and estimating public opinion may not be simple, but is a worthy topic to study. Every election is different. The changes in both the candidates and the state of the nation mean that older data and older methods may not work in the future. If methods to more accurately predict voting results could be found they could be possibly applied to nomination processes and some foreign elections, which are hard to accurately predict because of the higher level of precision needed to predict a winner given the possibly larger number of candidates. An application to estimation of public opinions as suggested in [5] may exist. The prediction of elections may not be trivial and the methods discovered could solve other problems in political science and beyond. In light of the 2016 election, political science statisticians must begin to examine the assumptions and distributions used in models and try to determine a more accurate method to predict elections. If a solid approach is discovered, it may not work indefinitely and could have to be reexamined for every unique election cycle. But presidential election prediction is an important field to study given its potential for both statistical outreach and applications to other problems.

5 References

[1] Christensen, W.F., & Florence, L.W., Predicting Presidential and Other Multistage Election Outcomes Using State-Level Pre-Election Polls. Amer. Stat., 62:1(2008), pp. 1-10
[2] Gelman, A., & King, G. Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?, British Journal of Political Science 23:04(1993), pg. 409-451
[3] Hummel, P., & Rothschild, D., Fundamental Models for forecasting elections at the state level, Electoral Studies, 35(2014), pp. 123-139
[4] Lock, K., & Gelman, A. (2010). Bayesian Combination of State Polls and Election Forecasts. Political Analysis, 18(3), 337-348.
[5] Park, D., Gelman, A., & Bafumi, J., Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls, Political Analysis, 12(2004), pp. 375,385