

Bayesian Hierarchical Models for the Prediction of American Elections



Brittany Alexander, Texas A&M University

Abstract

We propose a series of hierarchical Bayesian models to predict the outcome of American elections using only topline polling. The models are applied to recent federal elections with a focus on the past three Presidential elections. The models are structured to allow the pooling of information between states deemed similar. The models are similarly structured but they have different distributional assumptions (i.e. Gaussian) and likelihood structures. They are based on previous models that were found to be approximately 93% as accurate as the FiveThirtyEight Polls Plus model in terms of RMSE in predicting the 2008, 2012, and 2016 presidential elections and made nearly identical predictions of the winners of states as FiveThirtyEight. These models also were more accurate than averaging the polls. The methodology is implemented in the forthcoming bayesurvey R package and will be implemented to predict the 2020 Presidential and senate elections..

New Methodology

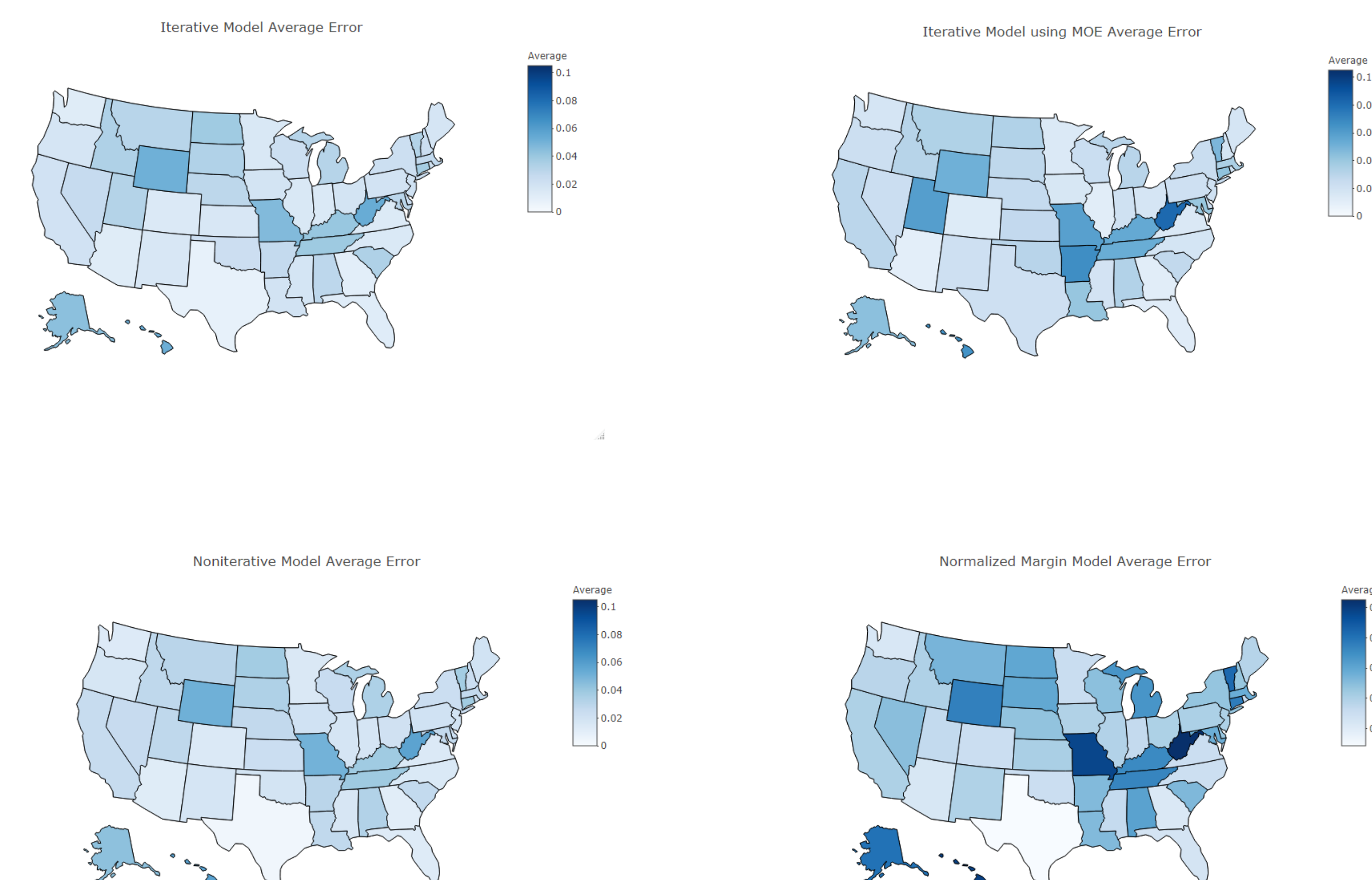
- Includes a new model that predicts the margin (difference in Democratic and Republican support)
- A Rolling average model using the last 2-25 polls were test
- The best performing rolling average models used about 8-12 polls and was rounded to 10 for simplicity
- Fit models using only the last 60 days of polling or using all polls
- Included iterative model that used provided margin of error to estimate variance instead of standard error of a proportion
- 60 days is when polls error starts decreasing over time
- These models are now implemented in the bayesurvey R package (available on Github as beta version) and can be almost instantaneously fit
- DC is removed from average error calculation because it is an outlier that in recent elections has always had significantly support for the Democratic than the vote in any other states
- A state is called competitive is it's margin is between -5 and 5.

Average Error of Models

	Iterative Gaussian Model	Iterative Gaussian with Variance from MOE Model	Noniterative Gaussian (Candidates) Model	Noniterative Gaussian (Margin) Model	Noniterative Gaussian (Margin) Model Normalized	Average of All Polls
All Polls	0.026	0.031	0.026	0.055	0.049	0.025
Rolling Average	0.025	0.030	0.026	0.054	0.048	0.025
Last 60 days	0.027	0.029	0.026	0.053	0.048	0.025
All Polls Competitive States	0.014	0.015	0.015	0.027	0.029	0.013
Rolling Average Competitive States	0.017	0.017	0.019	0.036	0.038	0.013
Last 60 days Competitive States	0.012	0.014	0.013	0.023	0.025	0.013

Previous Work

- Alexander & Ellingson (2019) created a series of Bayesian conjugate prior models for predicting the two party support for the 2008, 2012, and 2016 American Presidential election
- Data source was Huffington Post's Pollster
- Bayesian modelling uses prior information and that prior information comes from clustering states in those models
- States are clustered into groups based on the average margin from the past four elections
- The cutoffs for groups are -.2, -.1, -0.025, 0.025, .1, .2 with states in the group where the average margin was between the two points
- Alexander & Ellingson (2019) only considered polls after July 1st This projects uses the Iterative Gaussian (updates with every poll) and noniterative gaussian model (fit once)
- Poll and election data was proportionally normalized so that the sum of the Democratic and Republican support was equal to 1
- Iterative model updates for every poll, noniterative averages the polls and updates once



Discussion

- Restricting the poll data to the last days of the election or the last few polls provides only a minor improvement
- These models can be reliable predictors of the final Electoral College outcome but they will not always get a very close race (margin (< .03)) correct
- Better Clustering definitions should be tested
- More cutoff dates for including polls should be tested
- Combining a rolling average with a cutoff date should be tested to see if it improves error
- The use of previous election results could be a potential new source of prior information that could be combined with the clusters
- Bayesian Methods show promise to predict elections
- These models have an tendency to underestimate the uncertainty and historical model performance should be considered when evaluating the uncertainty in the 2020 election