A Poll-Based Bayesian Hierarchical Model for American Presidential Elections

Brittany Alexander

Statistics PhD Candidate
Texas A&M University

May 11th, 2021

Outline

1. Introduction & Literature Review
2. Model Methodology
3. Results

Table of Contents

1 Introduction & Literature Review

2 Model Methodology

3 Results

Introduction

- Some models are forecasts aiming to predict the election, but others are aggregations to determine the support at the present time
- Election models use polling data exclusively, economic and political data exclusively, or a combination of polls and economic and political examples
- FiveThirtyEight model and The Economist Model are prominent examples of Bayesian Models
- There are rolling average style models (CNN Poll of Polls, RealClearPolitics Average
- Models focus on a variety of election types: Presidential, House, Senate, Gubernatorial

Literature Review

- The Ecomonist model is based on a Dynamic Linear Model (Linzer 2013) and modified by Gelman & Heidemanns
- The Economist model requires a few hours on a professional server and is high dimensional
- Details about the FiveThirtyEight Model are limited, and model has evolved over time
- Bon et al. (2019) details a model that included the level of undecided voters
- Shirani-Mehr et al (2018) is a detailed study of polling error that included a model to aggregate polls
- Quick Non-hierarchial conjugate prior models exist (Alexander & Ellingson 2019, Christensen and Florence (2008) but they severely underestimate uncertainty
- Goal of this model: Try to get close to the accuracy of a complex model while being computationally and statistically efficient

Table of Contents

1 Introduction & Literature Review

2 Model Methodology

3 Results

Election Polling Data

- Quantity and quality of polls varies widely across states and election years
- Competitive and larger states tend to have more polls
- Polling starts before the Presidential primaries and increases in volume over time
- Huffington Post's Pollster aggregated over 5000 state-level Presidential polls from 2008-2016
- The Economist Model also provided polls for 2020

Basic Methodology

- I normalize the polls so that the support for the Democratic and Republican candidate adds to 1.
- I then only predict the Democratic vote share.
- I assign a weight t_{ik} to each poll based on a gamma GLM. The weights are normalized so that the sum of the weights equals the number of polls in the state.
- This model is not a forecast. It is a estimate of the current vote. Close to the election the current vote is highly similar to a election forecast.
- I assume independence among the polling data. If the weighting and likely voter models have correlated error across polls this could lead to an underestimate of uncertainty.

Model Formation

We want to predict the mean for the democratic candidate

- Let p_{ik} be the percent support for the Democratic candidate after normalization from the kth poll in state i. Then $p_{ik} \sim Normal(\mu_i, t_{ik} * \sigma_i)$ All polls are independent. and t_{ik} is an optional weight
- $\sigma_i^2 \sim inv - gamma(1.01, 1.01)$ truncated to (0, 0.0625)
- $\mu \sim MVN(\xi, \Sigma)$. ξ is the vector of the Democratic support in each state from the last election. Σ is a covariance matrix taken from the Economist model or the Economist model is the mean of an Inverse-Wishart model
- (Optional) $\Sigma \sim Inverse - Wishart(54, \Sigma_e)$
- You can restrict polls to certain dates, or use a rolling average

This model has a closed-form Gibbs sampler and can be easily implemented

Table of Contents

1 Introduction & Literature Review

2 Model Methodology

3 Results

Past Accuracy Overview

Here we show the accuracy of the model with no weights, no inv.wishart prior, and all polls within 100 days of the election.

	95% CI Coverage	% of Winners Predicted	Expected % of Winners Predicted	Average Absolute Error
2008	92.16%	92.16%	92.07%	0.0301
2012	100%	100%	93.05%	0.0149
2016	64.71%	86.27%	96.63%	0.0320
2020	98.04%	88.24%	93.36%	0.0233
Average	88.73%	92.16%	92.07%	0.026166667

Comparision to Other models: Absolute Average Error

Here we compare the average absolute error of the model with no weights, no inverse wishart prior, and all polls within 100 days of the election to the Economist and FiveThirtyEight model. Competitive states are states where the margin of the two-party vote is less than 5%.

	This Model	538	Economist
2008	0.0301	0.0205	0.0161
2008 Competitive	0.0239	0.0051	0.0076
2012	0.0149	0.0160	0.0182
2012 Competitive	0.0062	0.0046	0.0087
2016	0.0320	0.0323	0.0364
2016 Competitive	0.0276	0.0120	0.0210
2020	0.0233	0.0235	0.0221
2020 Competitive	0.0287	0.0281	0.0298
Average	0.0251	0.0222	0.0232
Average Competitive	0.0216	0.0144	0.0168

Comparison to Other models: Mean Square Error

Here we compare the average absolute error of the model with no weights, no inv.wishart prior, and all polls within 100 days of the election to the Economist and FiveThirtyEight model (Polls plus if applicable).

	This Model	538	Economist
2008	0.0387	0.0320	0.0229
2012	0.0205	0.0198	0.0225
2016	0.0389	0.0381	0.0470
2020	0.0362	0.0371	0.0350
Average	0.0336	0.0317	0.0318

Conclusion

- This method is a reliable and quick estimator of American Presidential Elections
- This method appears to approximate more complicated models like FiveThirtyEight & The Economist Model in terms of predicted outcomes
- This method tends to be lean on the side of more uncertainty but that can be addressed with different priors
- This model can be easily adapted and is flexible

References

- Alexander, Brittany (2019), "A Bayesian Model for the Prediction of United States Presidential Elections," SIAM Undergraduate Research Online, **12**.
- Bon, J. J., Ballard, T., & Baffour, B. (2019), "Polling bias and undecided voter allocations: US presidential elections, 2004–2016," Journal of the Royal Statistical Society: Series A (Statistics in Society), **182(2)**, 467-493.
- Christensen, W.F., & Florence, L.W.(2008), Predicting Presidential and Other Multistage Election Outcomes Using State-Level Pre-Election Polls," American Statistician, 62:1, 1-10
- Linzer, D. A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. Journal of the American Statistical Association, 108(501), 124-134.
- Lock, K., & Gelman, A. (2010), Bayesian Combination of State Polls and Election Forecasts. Political Analysis," 18(3), 337(348).
- Shirani-Mehr, H., Rothschild, D., Goel, S., Gelman, A. (2018). Disentangling bias and variance in election polls. Journal of the American Statistical Association, 113(522), 607-614.