FORECASTING THE 2020 U.S. ELECTIONS: A COMPARTMENTAL MODELING APPROACH

Alexandria Volkening, William He, Christopher Lee, Daniel F. Linder, Mason A. Porter, Grzegorz A. Rempala

SIAM Workshop on Network Science 2020
July 9–10 · Toronto

Summary

Election forecasting is an exciting, high-stakes problem with many sources of subjectivity and uncertainty. Here we develop transparent, poll-based methods for forecasting elections from a complex-systems perspective, and we use our model to forecast the 2020 presidential election.

Introduction

Election forecasting is a challenging problem, and, given the same polling data, forecasters predict different outcomes and report a range of certainty levels in their forecasts. In the U.S., election outcomes are related in states that have similar characteristics (e.g., if an analyst is wrong in their forecast of Ohio, they are likely also wrong in their forecast of Pennsylvania [1]), and this makes the process of forecasting elections richer and more difficult. Our work focuses on producing transparent, data-driven forecasts from a mathematical-modeling perspective ahead of upcoming elections. We are particularly interested in better understanding not just how states are related symmetrically, but also how states may influence each other in directed ways (Figure 1b).

With the goal of better understanding the election-forecasting process broadly and state–state relationships in particular, we developed a compartmental model to forecast the 2018 midterms (Figure 1a). Our eve-of-election forecasts for the 2018 governor and Senate races performed as well as those of popular analyst FiveThirtyEight [1] (Figure 2). Going back in time to produce forecasts of the governor, Senate, and presidential races between 2012 and 2016, we find that our compartmental-modeling approach consistently performs at a similar level as popular forecasters. In this talk, we now use our model to forecast the 2020 presidential election. By simulating our model based only on early polling data for past elections, we also study how early our past presidential forecasts have crystallized and raise questions for future research.

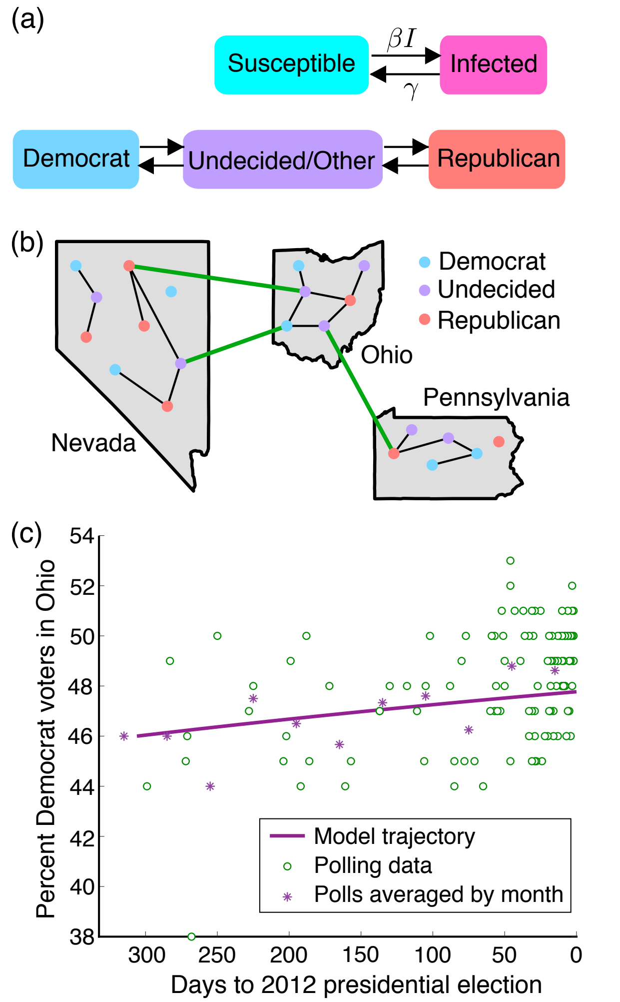

Figure 1: Overview of our approach to forecasting elections. (a) We use a compartmental model to frame (b) directed relationships between voters in different states. (c) We base our model parameters on the timeline of polling data in each state. Images adapted from [2].

Model

Because it provides a general, well-studied way of accounting for asymmetric, directed relationships, we base our work on compartmental modeling of disease transmission. Political dynamics and biological disease spread are certainly different, but this general, multidisciplinary approach provides a useful framework for us. We consider two "contagions" – Democrat and Republican voting intensions, and we rethink "susceptible" individuals as undecided voters. Our model works at the level of states, and we track the fraction of Democrat (I_D^i), Republican (I_R^i), and undecided voters (S^i) in each state i as below:

$$dI_D^i(t) = \left(\underbrace{-\gamma_D^i I_D^i}_{\text{Dem. loss}} + \underbrace{\sum_{j=1}^{M} \beta_D^{ij} \frac{N^j}{N} S^i I_D^j}_{\text{Dem. "infection"}} \right) dt + \underbrace{\sigma dW_D^i(t)}_{\text{uncertainty}} \tag{1}$$

$$dI_R^i(t) = \left(\underbrace{-\gamma_R^i I_R^i}_{\text{Rep. loss}} + \underbrace{\sum_{j=1}^{M} \beta_R^{ij} \frac{N^j}{N} S^i I_R^j}_{\text{Rep. "infection"}} \right) dt + \underbrace{\sigma dW_R^i(t)}_{\text{uncertainty}}, \tag{2}$$

where we use that $S^i + I_D^i + I_R^i = 1$ to reduce our equations from three to two. The parameters β_D^{ij} (respectively, β_R^{ij}) capture directed relationships between undecided voters in state i and Democrat (respectively, Republican) voters in state j.

To fit our model parameters, we rely on the timeline of polling data in the months leading up to each election (Figure 1c). To produce early forecasts, we use only a portion of the polling data – for example, to forecast the 2020 elections in July, we will use the polling data currently available for 2020 for a generic Democrat vs. Trump race. We include uncertainty in our forecasts by correlating noise on state demographics.

Results

We posted our election forecasts [2] online on arXiv before the 2018 midterms, and we have also tested our model on past governor, Senate, and presidential races since 2012. Across these elections, we find that our model performs similarly to popular forecasters on the eve of the election. Our work also highlights how accounting for uncertainty in different ways can strongly affect forecasts and highlights the importance of correlating noise in election models.

By fitting our model parameters using progressively smaller amounts of polling data, we are now using our model to study how early our forecasts crystallize across past elections. As we show in Figure 2, we find, for example, that our categorizations of each swing state as

	Model 8 July	Model 7 Aug.	Model 6 Sept.	Model 6 Oct.	Model 3 Nov.
Arizona	90.0%	84.6%	71.8%	70.0%	65.3%
Florida	79.3%	78.2%	75.1%	66.1%	59.4%
Indiana	55.8%	88.5%	88.9%	80.8%	75.1%
Minnesota*		99.1%	97.9%	97.3%	95.3%
Missouri	65.8%	61.4%	67.9%	63.4%	57.1%
Montana	91.3%	83.3%	79.9%	77.0%	83.3%
Nevada	66.7%	58.5%	54.2%	56.6%	51.8%
New Jersey	67.2%	51.6%	71.0%	78.4%	77.5%
North Dakota	72.5%	76.1%	77.1%	88.0%	89.3%
Ohio	99.8%	99.7%	99.6%	99.5%	99.1%
Tennessee	87.8%	91.9%	82.0%	69.0%	56.4%
Texas	92.0%	90.4%	84.7%	85.9%	86.1%
West Virginia	90.8%	93.1%	93.0%	93.4%	93.4%
Wisconsin	94.0%	97.9%	94.6%	96.2%	95.6%

Solid Rep. (≥95%)
Likely Rep. (≥75%)
Lean Rep. (≥60%)
Toss-Up (<60%)
Lean Dem. (≥60%)
Likely Dem. (≥75%)
Solid Dem. (≥95%)

Figure 2: Sample forecasts: our 2018 Senate forecasts by month leading up to the election. Notably, we find that our state categorizations (e.g., as Republican or Democrat) for swing states change by only 2 states after July and are consistent from August onward. We indicate the outcome of each election by the color of the state name in the first column (using traditional party colors).

Democrat or Republican change little from July onward for the 2018 Senate races. This study will help us better understand what our confidence level should be in our early 2020 forecasts, and we are excited to share our initial forecasts for the 2020 presidential elections in July.

References

[1] N. Silver. FiveThirtyEight: A User's Guide To FiveThirtyEight's 2016 General Election Forecast. https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/, 2016. Accessed: 31-10-2018.

[2] A. Volkening, D. F. Linder, M. A. Porter, and G. A. Rempala. Forecasting elections using compartmental models of infections, 2018. arXiv:1811.01831. Preprint, posted 5 November 2018. In review.