

World view



By Andrew Gelman

Why forecast an election that's too close to call?

Predictive models don't make the news, but they have a crucial role in democracy.

Four years ago, *The Economist* magazine asked me to construct a model for forecasting the results of the US presidential election. My colleagues and I did a pretty good job of capturing the uncertainty – we predicted that Joe Biden would receive between 259 and 415 votes, and he gained 306.

This year, we're back at it, for an even closer race (see page 1017). We currently estimate that the two candidates, Kamala Harris and Donald Trump, have roughly equal chances of winning.

Our model, like other electoral forecasts, uses state and national polls, along with political and economic data from past elections. Combining these data sets yields correlated uncertainties about each state's election outcome, which are added up to forecast the candidates' totals.

I think the main value of forecasts is not in the predictions themselves, but in how they portray uncertainty and the stability of the race over time.

Daily polls attract attention, but are easy to overreact to. Electoral forecasts – interpreted appropriately – can help us all to keep our heads in an environment of information overload. After all, one of the most important roles of science is to temper enthusiasm for outlandish claims, whether about miracle cures or perpetual motion.

This year, the numbers coming out of our model are not going to grab headlines. Not much has changed in the past few months. Harris's win probability moving between 45% and 55% is hard to distinguish from noise.

On the basis of forecast uncertainties, I've estimated that, as a rule of thumb, a 10% change in a candidate's probability of winning roughly corresponds to a 0.4 percentage point swing in the national vote.

Four-tenths of a percentage point is not nothing – and in a close election it can be decisive. But it's beyond the precision forecasters can expect to get from any poll, or even any aggregate of polls, because the margin of error in most polls is around three percentage points. And biases in polls could even double that margin.

It's impossible to know which forecaster is the most successful, except in extreme cases. Rating forecasters on the basis of their track record of predicting the winner reveals little. The differences between the results are too small and elections happen too infrequently for researchers to statistically identify which predictor is better.

For example, in the 2016 US election, the poll-aggregating website FiveThirtyEight predicted that Trump had a 30% chance of winning, whereas the newspaper *The New York Times* gave him a 15% chance. Trump won, so that 30%

Forecasting models don't stop the storytelling, but I think they make the stories more sophisticated and more politically accurate."

prediction looks better than the 15% estimate – but it was just one roll of the dice. Indeed, if you forecast an event to have a 15% chance of happening, you'd expect it to occur about one time in six.

The problem is that, in statistics, frequent events are what allow researchers to judge whether models are better or worse – in sports betting or weather predictions, for instance, forecasters get daily data and have decades of past records that can be used for calibration. Events that happen every two to four years do not allow for such assessments.

We can use past performance to remove extremely overconfident forecasts, such as those that gave Hillary Clinton a 99% chance of winning in 2016, but it could take hundreds of years of elections for scientists to be able to distinguish between forecasts that stay within reasonable bounds.

Why, then, do I make forecasts? First, political science. The fact that US presidential elections are predictable, to within a few percentage points, helps scientists to understand US politics. This predictability affects how politicians and journalists think about elections, the economy and the balance between parties.

Second, as baseball analyst Bill James supposedly said, the alternative to good statistics is not 'no statistics', it's 'bad statistics'. Although data-based forecasts don't provide the predictive accuracy that would allow forecasters to call the election early, they do give useful boundaries for the contours of the race – however blurry those might be.

In the absence of prediction models, political observers would be inclined to spin a story around each campaign event and every poll. Forecasting models don't stop the storytelling, but I think they make the stories more sophisticated and more politically accurate.

Why, then, is news coverage of the election so dominated by the race, and not the politics? I have a theory.

If a voter is a politically engaged follower of the news, they probably already know who they will vote for. They won't be hugely motivated to learn more about the candidates' positions – but they are interested in who is going to win. This spurs news outlets to commission and report on polls, which in turn promotes probabilistic forecasts such as ours.

Primary elections, in which candidates are selected, are another story. Voters have several options to choose from in a single party, and these candidates are likely to have similar political positions. Even the strongest partisans are motivated to learn more about the candidates and where they stand on specific issues.

Right before a presidential election, it's rational for the media to cater to the majority who already know where they stand, rather than to those who are open to persuasion – and probably less interested in politics, anyway. In the end, elections will always be uncertain, because it is up to the individual to decide how to vote, and whether to vote at all.

Andrew Gelman

is professor of statistics and political science at Columbia University, New York City.
e-mail: gelman@stat.columbia.edu

The author declares competing interests; see go.nature.com/3byzzbd for details.