Grappling with uncertainty in forecasting the 2024 U.S. presidential election

Andrew Gelman[†,*], Ben Goodrich[‡], Geonhee Han[◇]

[†] Department of Statistics and Department of Political Science, Columbia University, New York

[‡] Department of Political Science, Columbia University, New York

[◇] Graduate School of Arts and Sciences, Columbia University, New York

3 Oct 2024

ABSTRACT. We discuss acknowledged and unacknowledged sources of uncertainty in the Economist magazine's state-by-state election forecast.

Four years ago we worked with *The Economist* magazine to produce a state-by-state election forecast, combining national polls, state polls, economic and political "fundamentals," and a hierarchical Bayesian model allowing for correlation among states, variation over time, and sampling and nonsamplng error of surveys. The model, built off the hierarchical Bayesian time-series models of Lock and Gelman (2010) and Linzer (2013), was described in this journal by Heidemanns et al. (2020), with further discussion of communication in Gelman et al. (2020). We fit the model in Stan (Carpenter et al., 2017), and our forecast updated daily as polls came in during the summer and fall. With some hiccups, it performed reasonably well, albeit with some concerns regarding the quantification of uncertainty (Gelman, 2020), issues that have arisen with poll-based forecasts more generally (Gelman, 2021).

This year, we accepted the invitation of Dan Rosenheck of *The Economist* to help with their 2024 forecast. The starting point was the code from 2020, to which which we considered various improvements, including: (i) improving the fundamentals-based model to better account for the declining importance of the economy as a predictive factor in an increasingly polarized electorate; (ii) more carefully estimating the state-level correlations of polling errors and time trends in opinions; (iii) accounting for more nonsampling error in polling. As before, we checked our model by fitting it to data from past presidential campaigns, along with existing polls from 2024 after Joe Biden withdrew from consideration for the Democratic nomination, to check that it produced inferences that seemed reasonable given our current political understanding. We also performed some forward checking, considering different hypothetical polling scenarios for the rest of the campaign and checking that the resulting inferences made sense—that they were not too stable but did not swing too widely. We want our model to be responsive to trends without overreacting to each poll.

[*]gelman@stat.columbia.edu

It might seem silly to check a model by comparing its inferences to reasonable expectations—if we knew what to expect, what is the purpose of the model at all?—but there are two reasons why this procedure seems reasonable to us. First, we are forecasting a multivariate outcome —50 state elections plus the District of Columbia —and it require a lot of care to construct a full forecast with all its correlations. Second, we are constructing a sort of robot—a forecast that should be able to update itself over time as new polls and economic and political information arrive—so our checking is not just on the current forecast probabilities but also on how they develop over time. For example, if a new poll comes in from Ohio showing a stronger-than-expected support for the Democratic candidate, how much should this shift the forecast in Ohio and in other states, and how does that map to the probability of each candidate winning?

When we wrote the first draft of this article, in early July when it looked as if Biden would be the Democratic nominee, our model gave the Republican candidate an expected 51% share of the national two-party vote and a 3/4 probability of winning the Electoral College. At the time of this writing at the end of September, Kamala Harris is predicted to win 52% of the two-party vote but with a roughly even chance of winning the Electoral College majority (Economist, 2024). With the current state of public opinion and the expected relative distribution of votes among the states, it makes sense that the Democrats are expected to need more than half the vote to have the Electoral College edge; the exact magnitude of this edge is unknown, as it depends on future state-by-state election outcomes. This geographic bias varies from election to election and at times has favored the Democrats. The forecast probability expresses an appropriate uncertainty given the closeness of the polls and the possibility of large polling errors and national swings between now and November.

Here are a few possible failures that we anticipated with our forecast going forward:

• *What if one candidate or another takes a solid lead in the national polls?* This would result the candidate's predicted national vote share—and, through the correlations in the model, individual state vote shares—going up, and as our model is set up, a swing of just a few percentage points would result in a probability of 90% or more of winning–but then what if later there is a big swing in the other direction, leading to that candidate's win probability going below 10%? A month before the election, this seems highly unlikely, but it was a legitimate concern when we were setting up our model in the spring. A probabilistic forecast should be a martingale–that is, if the forecast at time t of a certain future event has a probability of $X(t)$, then $E(X(t+s))$, given all information available at time t, should be equal to $X(t)$. So a swing in predicted probability from 90% to 10%, while possible, should be very unlikely, and a forecasting procedure that regularly shows such swings has problems (Taleb, 2017). We do not expect this to happen, but it could! Polling has been very steady during the past several election campaigns, but large swings were common in decades past (Gelman & King, 1993). The most relevant parameter in our model is the standard deviation of the random walk of national vote preference over time. When implementing our model for *The Economist*, we set this scale to a value that seemed high enough to allow for plausible changes during the half year leading up to the election while still allowing informative inferences during those early months. But large enough variation over time could break this model and yield overconfident predictions.

• *What about third parties?* Following our practice in previous elections, we model preferences for the Democrat and the Republican, ignoring other candidates, which has seemed reasonable given that no third-party nominee has won any states since 1968. For a while, though, Robert Kennedy, Jr. appeared to be a strong alternative to Biden and Trump, which could affect our forecast directly if Kennedy were to win any states and indirectly to the extent that changes in his support were to go unevenly to the major-party candidates. Presumably other minor parties will not matter much,

at least not compared to 2016, when the Libertarian and Green candidates did not win many votes despite widespread discontent with the options of Clinton and Trump.

• *Actuarial concerns.* Biden and Trump are both around 80 years old, with a nontrivial risk of death or disability before election day. What happens if one or the other candidate needs to be replaced? Even before the first presidential debate, this was a vigorously-discussed topic, with pundits arguing that both parties were hobbled by weak candidates; see Gelman (2024). We did not have anything on this on our model, implicitly assuming that any replacement candidate would do about as well as the existing nominees. Ever since Rosenstone (1983), there has been a consensus in political science that candidates do not matter so much for presidential voting, except that there is a slight advantage to political moderation. Given that most prominent alternatives within their parties are no more politically moderate than Biden or Trump, it seemed safe to not worry about specific candidate effects. That said, since Biden was replaced by Harris on the Democratic ticket, we observed changes in the polls beyond what might be expected from our default time-series model. Thus, the model did not use any Biden-Trump polls.

• *Concerns specific to 2024.* This is the first presidential election where either major-party candidate has been convicted of a felony, and the first since 1984 where there have been serious concerns about either candidate's mental deterioration. Pundits have also noted the unusual disconnect between relatively strong economic performance and the president's low approval ratings. Another noteworthy feature, with effects already apparent in the 2022 midterm elections, has been a series of controversial Supreme Court decisions on issues ranging from abortion to presidential immunity. On the other hand, other recent campaigns have had historically unique features: the 2020 election was complicated by covid, early voting, two already elderly candidates, and justified concerns that one of these candidates would not accept the election outcome; and the three elections before that had the first African-American, Mormon, and female nominees, all of which might seem commonplace today, but at the time many people polled expressed resistance to voting for candidates with these attributes. This is not to say that it is a bad idea to adjust for what we can, just that we would hope our existing error terms to capture some of the unexpected. The Supreme Court issue is related to concerns about partisan balance, another tricky feature this year, with both houses of congress up for grabs.

• *Polling errors.* These were major concerns in 2016 and 2020. What about 2024? It is hard to say with certainty. Our model allows for systematic errors at the national and state level, but they all have prior expectation of zero. A study of state-level polling errors since 2000 found a positive correlation among successive elections—that is, if state polls are biased toward the Republicans or Democrats one year, they are likely to have a similar bias in the next election (Heidemanns, 2022). Our model does not include this autocorrelation (because we assume that pollsters are trying to correct for such biases), so we may be leaving some information on the table. We hope that a reasonable range of possible polling bias is included in our predictive uncertainties.

Traditionally, the general election campaign is said to begin on Labor Day, after the two parties' nominating conventions. This year, neither party's candidates faced serious primary challenges, the two candidates appeared to be set in the spring, and observers were anticipating a long slog through November. Recently we have seen three shocks—Trump's felony conviction and subsequent erratic performance in campaign events, concerns about Biden's age culminating in his withdrawal from the race, and his replacement by Harris—and the summer brought us a new and potentially volatile race. In the modern era of extreme political polarization, we expect our state and national

forecasts to still be reasonable, but ultimately they are conditional on model assumptions, hence the importance of transparency in methods and data.

Contributions. All three authors contributed to the statistical modeling and the writing.

Appendix A. Model combining state polls, national polls, and prior information

A.1. **The 2024 Economist model.** The model begins with a fundamentals-based forecast, a regression model predicting the incumbent party's share of the two-party vote given economic conditions, presidential popularity, and a measure of political polarization. We turn this into a state-level forecast by adding an estimate of each state's "lean" relative to the national average. We use these fundamentals-based forecasts as a prior expectation and uncertainty to form a multivariate normal prior distribution for the election outcomes.

We then include the information from polls. Let y_i be the Democratic candidate's share of the two-party vote in poll i. We assume

$$y_i \sim \text{binomial}(n_i, p_i),$$

where p_i represents the expected response for the particular survey. We add nonsampling error to p_i later in the model.

We index dates by $t = -T, \ldots, 0$, representing the days before the election, and state by $s = 1, \ldots, S = 51$, with $s = 0$ representing national polls, so that each poll i has a date $t[i]$ and state $s[i]$. The expected response p_i combines a state-specific (or national-level) time-varying component $\theta_{t[i],s[i]}$, and poll characteristic adjustments α_i:

$$p_i = \text{logit}^{-1}(\theta_{t[i],s[i]} + \alpha_i + \varepsilon_i).$$

The dynamic component is modeled as,

$$\theta_{t,s} = \begin{cases} \mu_{t,s} + \varepsilon_s & \text{if } s \geq 1 \\ \sum_{s=1}^{S} w_s(\mu_{t,s} + \varepsilon_s) & \text{if } s = 0 \end{cases}.$$

The time series, $\mu_t = (\mu_{t,1}, \ldots, \mu_{t,S})$ is modeled as a correlated S-variate random walk:

$$\mu_{t+1} - \mu_t \sim \text{MVN}_S(0, \Sigma^{(\mu)}), \quad \text{for } t = -T, \ldots, -1$$

$$\mu_0 \sim \text{MVN}_S(m, V),$$

where m is the S-variate fundamentals-based forecast, and $w = (w_1, \ldots, w_S)$ is a vector of weights that sum to 1, where w_s is proportional to the number of votes cast in state s in the previous presidential election. The state-specific error ε_s is assumed to be correlated over s to account for correlated polling biases over states, with

$$(\varepsilon_1, \ldots, \varepsilon_S) \sim \text{MVN}_S(0, \Sigma).$$

The term α_i adjusts for poll characteristics:

$$\alpha_i = \alpha_{r[i]}^{(r)} + \alpha_{m[i]}^{(m)} + \alpha_{p[i]}^{(p)} + \alpha_{l[i]}^{(l)}.$$

These terms are designed to adjust for:

- $\alpha_r \overset{iid}{\sim} \text{normal}(0, v^{(r)})$: polling population (e.g. likely voters, registered voters, or all adults),
- $\alpha_m \overset{iid}{\sim} \text{normal}(0, v^{(m)})$: polling mode (e.g. automated phone, live phone, internet, etc.),
- $\alpha_p \overset{iid}{\sim} \text{normal}(0, v^{(p)})$: polling organization,
- $\alpha_l^{(l)}$: poll partisan lean, with

$$
\alpha_l^{(l)} = \begin{cases} \varepsilon_l & \text{(Democratic sponsored)} \\ 0 & \text{(Nonpartisan)} \\ -\varepsilon_l & \text{(Republican sponsored)} \end{cases}, \qquad \varepsilon_l \sim \text{exponential}(\lambda).
$$

The term ε_i accounts for poll-specific errors:

$$
\varepsilon_i \sim \text{normal}(0, v_{s[i]}),
$$

where the variances $(v^{(\cdot)}, v_s)$ are hyperparameters.

As described by Heidemanns et al. (2020), the model produces a forecast of the latent support in favor of one of the two major parties as a byproduct of inferring the latent multivariate random walk $\mu_{1:T}$. State-level forecasts are produced as the marginal posterior of

$$
(\mu_{1,s}, \ \ldots, \ \mu_{T,s})
$$

for a given state s. The national popular vote forecast is simply a weighted average of the state forecasts.

A.2. **Some ideas for model improvement.** We discuss some improvements to the model that we considered which could make sense to implement in future election cycles. One difficulty was that the decision of how to model polling errors has a direct impact on the forecast probability of each candidate winning the election, thus it can be contentious to change the model in real time.

Perhaps the most important change is the way in which a forecast model should be evaluated. Since the current cycle's results will not be known until weeks after the November election, *The Economist*'s model has been calibrated based on how well it has predicted *past* elections. We would prefer to evaluate models based on how well they are expected to predict *future* polls in the current cycle. Over the past few years, we have collaborated to estimate the expected log predictive density (ELPD) for future data using Pareto smoothed importance sampling (PSIS) (Vehtari et al., 2024). A model with a higher ELPD tends to be better, although the predictions from different models (that are applied to the same outcomes) can be weighted to yield a better ELPD than any constituent model (Yao et al., 2018).

However, there are two difficulties with the PSIS estimator of ELPD. First, the outcome variable must be identical, so it is not straightforward to compare a model that treats the outcome variable as being discrete counts with a model that considers the outcome to be continuous proportions. Perhaps a normal approximation to a discrete likelihood could be applied with a continuity correction to facilitate such comparisons, but to date, this approach has not been evaluated in an ELPD context. Second, the PSIS estimator assumes that each past observation could be dropped without having a major effect on the posterior distribution. This assumption will be violated for a small percentage but a large number of polls, which introduces a bias in the ELPD estimator and its standard error, and if it is severe enough, can imply that the expectation of the estimator does not exist. Recently, this assumption has been relaxed using mixtures rather than PSIS (Silva & Zanella, 2023).

A binomial likelihood for polls is too restrictive. Either a beta-binomial likelihood for the count of the number of people in a poll who support the Democratic candidate or a normal likelihood for the proportion of such people (among respondents who support either the Democratic or Republican candidate) would be preferable because both add a parameter that would account for design effects as well as nonsampling errors, which past research suggest are as large as sampling errors in election polls (Houshmand Shirani-Mehr & Gelman, 2018).

In 2024, the forecast is conditional on a point estimate of the correlation matrix across the states, which was updated from the 2020 version using individual-level polling data from early in the cycle. We would prefer to estimate the correlation matrices along with the other parameters in the model. There are difficulties with this approach as well. Most of the states are rarely, if ever, polled during a cycle, and national-level polls are not constructed to be representative at a state level (an exception is the Cooperative Election Study, but that is not released until well after the election). Thus, not much information is available during the campaign to update the correlation matrices among most states. However, there are many polls in swing states whose cross-state correlations have a small effect on the predicted vote shares but an enormous effect on the predicted electoral votes: the aspects of the correlation matrix that are most important for the most important predictive goal are those for which the most information is available.

The model we have implemented of time-varying trends may be viewed as a *bottom-up* approach, where the S-variate random walks are aggregated to produce nation-level predictions. An alternative, *top-down* approach, would specify a national-level random walk as a primitive that each of the S states deviate from, which we believe would allow us to more effectively parameterize the two separate sources of covariation in the trends: a uniform national swing and correlations of state-specific effects.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Economist. (2024). US election 2024 [Accessed: 2024-07-03 and 2024-09-26]. https://www.economist.com/interactive/us-2024-election/prediction-model/president

Gelman, A. (2020). Concerns with our Economist election forecast. *Statistical Modeling, Causal Inference, and Social Science*. https://statmodeling.stat.columbia.edu/2020/10/28/concerns-with-our-economist-election-forecast/

Gelman, A. (2021). Failure and success in political polling and election forecasting. *Statistics and Public Policy*, *8*(1), 67–72. https://doi.org/10.1080/2330443X.2021.1971126

Gelman, A. (2024). How would the election turn out if Biden or Trump were replaced by a different candidate? *Statistical Modeling, Causal Inference, and Social Science*. https://statmodeling.stat.columbia.edu/2024/06/12/how-would-the-election-turn-out-if-biden-or-trump-were-not-running/

Gelman, A., Hullman, J., Wlezien, C., & Elliott Morris, G. (2020). Information, incentives, and goals in election forecasts. *Judgment and Decision Making*, *15*(5), 863–880.

Gelman, A., Katz, J., & King, G. (2004). Empirically evaluating the electoral college. In A. N. Crigler, M. R. Just, & E. J. McCaffery (Eds.), *Rethinking the Vote: The Politics and Prospects of American Electoral Reform* (pp. 75–88). Oxford University Press.

Gelman, A., & King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, *23*(4), 409–451. https://doi.org/10.1017/S0007123400006682

Heidemanns, M. (2022). Prediction and error: Forecast aggregation and adjustment. *Ph.D. thesis, Department of Political Science, Columbia University*.

Heidemanns, M., Gelman, A., & Morris, G. E. (2020). An updated dynamic Bayesian forecasting model for the US presidential election. *Harvard Data Science Review*, *2*(4). https://doi.org/10.1162/99608f92.fc62f1e1

Houshmand Shirani-Mehr, S. G., David Rothschild, & Gelman, A. (2018). Disentangling bias and variance in election polls. *Journal of the American Statistical Association*, *113*(522), 607–614. https://doi.org/10.1080/01621459.2018.1448823

Linzer, D. A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, *108*(501), 124–134. https://doi.org/10.1080/01621459.2012.737735

Lock, K., & Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, *18*(3), 337–348. http://www.jstor.org/stable/25792016

Rosenstone, S. J. (1983). *Forecasting Presidential Elections*. Yale University Press. Retrieved September 27, 2024, from http://www.jstor.org/stable/j.ctt1xp3vfx

Silva, L. A., & Zanella, G. (2023). Robust leave-one-out cross-validation for high-dimensional Bayesian models. *Journal of the American Statistical Association*, *119*(547), 2369–2381. https://doi.org/10.1080/01621459.2023.2257893

Taleb, N. N. (2017). Election predictions as martingales: An arbitrage approach. *Quantitative Finance*, *18*(1), 1–5. https://doi.org/10.1080/14697688.2017.1395230

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2024). Pareto smoothed importance sampling. *Journal of Machine Learning Research*, *25*(72), 1–58. http://jmlr.org/papers/v25/19-556.html

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917–1003. https://doi.org/10.1214/17-BA1091