Polls, Context, and Time: A Dynamic Bayesian Forecasting Model for US Senate Elections

Yehu Chen[1], Roman Garnett[2], and Jacob M. Montgomery[3]

[1]*Division of Computational and Data Sciences, Washington University in St. Louis.*
[2]*Department of Computer Science and Engineering, Washington University in St. Louis.*
[3]*Department of Political Science, Washington University in St. Louis. Email:* *jacob.montgomery@wustl.edu*

Abstract

We present a hierarchical Dirichlet regression model with Gaussian process priors that enables accurate and well-calibrated forecasts for US Senate elections at varying time horizons. This Bayesian model provides a balance between predictions based on time-dependent opinion polls and those made based on fundamentals. It also provides uncertainty estimates that stem naturally from the induced posteriors trained on historical data. Experiments show that our model achieves high levels of accuracy and a 95% coverage rate for vote share predictions at various forecasting horizons. We validate the model with a retrospective forecast of the 2018 election and a true out-of-sample forecast for 2020. *Keywords:* Forecasts, Bayesian, Gaussian process

NOTE: This manuscript is being prepared for submission to Political Analysis using their recommended LaTeX template. This paper has not been approved or even reviewed by PA or its editorial team.

1 Predicting senate races

In recent years there has been an explosion of interest in election prediction. Primarily, this has been driven media outlets and popular forecasting websites like fivethirtyeight.com. Although the most prominent forecasting efforts have been housed in media organizations, these models often build from, or are inspired by, research in political science.

Broadly speaking, academic election forecasting in the US context can be divided into two approaches. First, there are static models that make a single prediction for a given election (e.g., Fair 1978; Abramowitz 2008; Lewis-Beck and Tien 2008). These models are sometimes referred to as “fundamentals” models and primarily rely on economic indicators, incumbency status, and other factors that shape the general context of an election. To the extent they incorporate polling data at all, they are based on proxies such as presidential approval (Erikson and Wlezien 2008) or snapshots of polls taken well before Election Day (e.g Campbell and Wink 1990).

A smaller body of research has focused on building dynamic forecasting models that change over the course of the campaign as new polling data becomes available. A particular prominent example is Linzer 2013, which introduced a dynamic Bayesian model forecasting the US presidential election results for all fifty states. This model served as a basis for forecasts produced by major media outlets including *The Economist*, *Vox*, and *Daily Kos*. Jackman 2005 presents a somewhat related Bayesian model, although there the goal is to aggregate polls rather than to make a prediction *per se*.

While the US presidential election has received the most attention, a smaller body of research has focused on predicting legislative elections. Most examples in this domain do not seek to predict individual races, but rather the aggregate number of seats that swing to a specific party (e.g., Campbell 2018; Lockerbie 2012). Still, some previous work has attempted to build models of individual US Senate races based on fundamental factors (Klarner and Buchanan 2006; C. Klarner 2008; C. E. Klarner 2012; Hummel and Rothschild 2014). However, to our knowledge there are no published models providing dynamic forecasts of individual senate races in political science.

Political Analysis (2020)

DOI: 10.1017/pan.xxxx.xx

Corresponding author
Jacob Montgomery

Edited by
John Doe

The relative lack of attention to the Senate is largely the result of the intense popular interest in the presidential race. However, it also reflects the fact that predicting individual-level Senate elections is in some ways a harder task. To begin, there is far less polling data for any given Senate election relative to national races, especially early in the cycle. Some states with close races and large media markets may have dozens of polls, but in many others there are very few. In our database of Senate elections starting in 1992, the senate races had a median of eight polls taken more than 14 days before the election. In addition, senate races are relatively low salience to voters, especially early in the season. Even strong challengers can be unfamiliar to voters until the final weeks. As a consequence, public opinion can be far more dynamic as voters learn about their options in the final lead-up to Election Day. In short, senate elections offer fewer polls and what polls exist can be noisy predictors.

A further problem is that local context and other "fundamentals" are often only weakly predictive of candidate performance. Although the general partisan dispositions in each state tends to heavily structure presidential outcomes, election results in Senate elections are far less geographically determined. That is, knowing how a party candidate performed in one election is often a poor predictor of performance in subsequent years. A recent examples would be West Virginia where Democrat Joe Manchin won over 60% of the vote in 2012 and Republican Shelly Capito won over 62% just two years later. These kinds of dramatic partisan swings occur regularly, making “fundamentals” forecasts difficult. Klarner notes that his fundamentals-based model “has never performed well, being off by three seats in 2006 and five seats in 2008. US Senate elections appear to be influenced by race-specific factors that are difficult to include in forecasting models” (C. E. Klarner 2013, p 45). Meanwhile, Hummel and Rothschild (2014) predicted only 83% of races correctly in sample and performed similarly out-of-sample.

In combination, this means that for any single race it is difficult to provide accurate forecasts in the absence of polling. Yet, polling data itself is relatively sparse and subject to significant trends over the course of the election. And, of course, relying on unvarnished polling data can be inaccurate even where it is not just missing, making simple polling averages sub-optimal. Perhaps combining polling data with some information about fundamentals can outperform each in isolation?

On the other hand, there is one very important advantage to working in this setting relative to national elections; *we have far more data*. While presidential elections offer only one observation every four years, the Senate has roughly 33 election outcomes attached to hundreds of polls every two years. In our dataset, which covers only the post-1992 period, we have 432 election results and over 14,744 published polls. This give us some hope that we can train a model that can learn from the past to predict future outcomes and correctly calibrate our uncertainty.

Below, we present a hierarchical Dirichlet regression model in a Bayesian framework that enables us to combine polls and fundamental to accurate forecasts election outcomes at various time horizons. This model provides a structured balance between time-dependent opinion polls and state/candidate-level fundamentals, and allows uncertainty estimates on the forecast vote shares that naturally stems from the induced posterior belief based on historical data. Experiments show that our model can achieve state-of-the-art levels accuracy in winner prediction and 95% coverage rate in vote share prediction for various forecasting horizons. We test the model by holding out the 2018 election and further validate the approach using the 2020 election.

2 Intuition and related work

Before formally introducing the model, we want to focus on the core ideas that inform our approach. First, we suppose that polling for a candidate itself is a noisy measure of a true underlying public opinion $f(t)$ at any given time t. That is, we assume that there is a true underlying support for each candidate that moves smoothly over time and that polling results will imperfectly follow these trends.

While modeling smooth latent public opinion is consistent with previous efforts to aggregate polls (Stoetzer *et al.* 2019; Linzer 2013; Jackman 2005), we adopt a strategy that is more appropriate given the sparseness of polling in many senate elections. Our approach assumes a linear trend in the data with mild nonlinear deviations. This provides a sensible compromise between a simple linear model of public opinion and the trend-free smoothing procedures adopted in Jackman (2005) and Linzer (2013) (see also Stoetzer *et al.* 2019; Walther 2015). Indeed, these other approaches can be viewed as special cases of our more general model where no linear trend is included.

Second, our modeling strategy assumes that latent public opinion is only one predictor of election outcomes. That is, latent public opinion is not assumed to translate directly into election outcomes as in Linzer (2013). Instead, the model learns the degree to which public opinion accurately predicts elections relative to other "fundamental" factors including state-level voting history, candidate quality, and the like. This approach has two advantages. To begin, it allows us to easily train our model at different time horizons such that public opinion is weighted more heavily as the election approaches and polling becomes more predictive. More fundamentally, however, it allows us to explicitly model the inherent uncertainty in election outcomes that cannot be adequately predicted from polling data alone. That is, we assume that even if we knew public support for a candidate perfectly, there would still be uncertainty in the outcome due to turnout and other unmodeled factors. Our aim is to use historical data to calibrate our uncertainty and achieve correct coverage rates at various time horizons in a way that reflects this irreducible uncertainty.

Finally, to the degree possible the model is tuned for the goal of accurately predicting elections. Thus, while polling outcomes are included in the model, the key model parameters are not selected to reduce the error in predicting polls but in predicting elections. Thus, we select hyperparameters intentionally that under-predict individual polling results but provide a better basis for predicting candidate vote share. The result is a parsimonious, but accurate and well calibrated model of elections.[1] In the end, the model takes in only four variables: polling data, Cook's partisan voting index,[2] the partisanship of each candidate, and candidate quality. However, as we show below, it is still able to make accurate predictions for races at various time horizons while maintaining correct coverage.

3 A predictive model of US senate elections

The forecasting model we propose can be conceptualized as having two components as depicted in Figure 1. First, we use candidate-level polling data to predict the latent level of public support for candidate i on Election Day ($t = 0$), which we denote $f_i(0)$. If we are predicting this before the election ($t < 0$), this quantity is predicted based on all polling data up to the current date as well as an informative prior reflecting the general electoral context in that state. Note that the goal is not to create a point prediction, but to estimate a distribution on $f_i(t)$ that reflects our uncertainty about the trajectory of public opinion over the course of the election as well the inherent uncertainty in polling data itself. We refer to this as our *candidate-level model*.

Second, we then use predicted public support as inputs for an *election-level model*[3] with the goal of predicting the proportion of the vote divided among all candidates in a given race (that is, the entire vote share and not only the winner). We model this with a Dirichlet hierarchical model with year-level random effects using a training dataset of election outcomes starting in 1992. Importantly, this Dirichlet regression takes in $f_i(t)$ as an *input* along with contextual factors like the partisan leanings of the state. Thus, we are able to use historical data to estimate the the degree to which

1. The model was built using data from the 1992-2014 period using 2016 as a validation set. All modeling decisions and hyper-parameter selection was done using only this data. We held out 2018 to serve as a test set for this analysis.
2. https://cookpolitical.com/pvi-0
3. Throughout, we use *election* to refer to a race between two or more candidates in a single seat. The overall election cycle (roughly 33 elections in a given year) is referred to as a cycle.

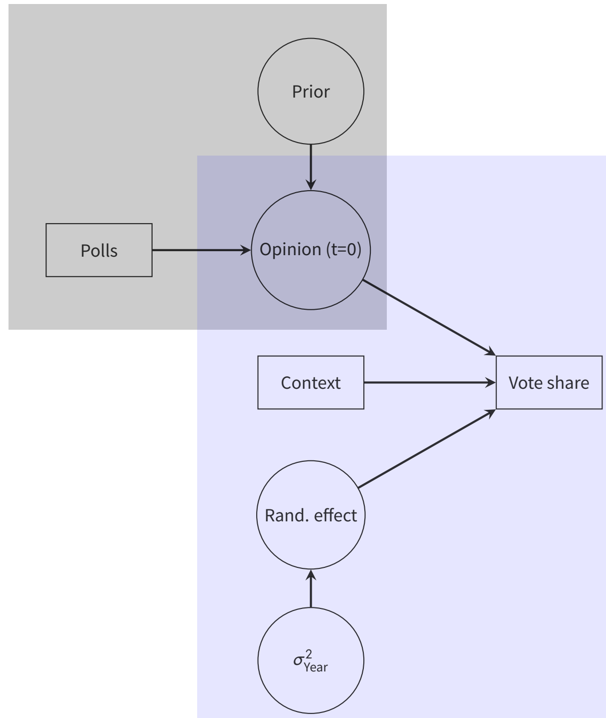


Figure 1. Conceptual outline of the two stage model. The candidate-level model, shaded in gray, predicts public opinion on Election Day. The election-level model, shaded blue, predicts vote share as a function of public opinion and contextual factors.

electoral context, public opinion, or some mix of the two are best able to predict vote shares at different time horizons.

The final output is a prediction for Senate elections that reflects two levels of uncertainty. First, we have uncertainty about where latent public opinion *will be* on Election Day given the polling data we have observed so far. Second, we have uncertainty as to how well public opinion and contextual factors actually predict election outcomes based on historical data.

3.1 Background on Gaussian process regression

Our model for latent public opinion over time is a linear trend with smooth nonlinear deviations from that trend. Here we subsume both components into a single *Gaussian process* (GP) model of the latent opinion. GPs offer a flexible Bayesian framework for nonlinear regression widely adopted in machine learning (Rasmussen and Williams 2006). Many models encountered in the political science literature are special cases of this framework, including Kalman filters (Jackman 2005), many random walks (Linzer 2013), and even standard linear regression with normal priors on the coefficients.

To define a Gaussian process, consider a function $f : \mathcal{X} \to \mathbb{R}$ on some arbitrary domain $\mathcal{X}$; for our model of latent opinion, $\mathcal{X} = (-\infty, 0]$ is the span of times at which we may wish to predict. The defining property of a Gaussian process is that if $\mathbf{x} \subset \mathcal{X}$ is a finite vector of input locations, then the associated function values $f(\mathbf{X})$ has a multivariate normal distribution. The moments of this distribution are provided by a *mean function* $\mu(x) = \mathbb{E}[f(x) \mid x]$ and covariance function $K(x, x') = \text{cov}[f(x), f(x') \mid x, x']$; evaluating these pointwise provides the mean vector and covariance matrix for any desired vector of function values $f(\mathbf{x})$. Modeling with the Gaussian process entails designing the mean and covariance functions to encoding the desired statistical properties of f such as correlations over the domain.

A critical property of GPs is that they enable closed-form inference for regression from observations corrupted by additive Gaussian noise. Let $f \sim \mathcal{GP}(\mu, K)$ have a GP prior and suppose we obtain a vector of observed values $\mathbf{y}$ at locations $\mathbf{X}$, where $y_i = f(x_i) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2)$. Then the posterior belief of f given $D = (\mathbf{X}, \mathbf{y})$ is again a GP with updated mean and covariance function:

$$\mu_{f|D}(x) = \mu(x) + K(x, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1}(\mathbf{y} - \mu(\mathbf{X}))$$
$$K_{f|D}(x, x') = K(x, x') - K(x, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1} K(\mathbf{X}, x').$$

Hence, appealing to the definition above, the posterior predictive distribution of any function value $f(x^*)$ is normal with moments given by the above:

$$f(x^*) \mid D, x^* \sim \mathcal{N}\big(\mu_{f|D}(x^*), K_{f|D}(x^*, x^*)\big).$$

This final point is critical below. When modeling the latent opinion with a Gaussian process, our prediction for latent public opinion on Election Day is a normal distribution that can be directly calculated. This in turn becomes a normal prior for public opinion that is passed directly into the election-level model.

3.2 Projecting public support via Gaussian process regression

We next outline our approach for forecasting voter preferences throughout an election given previous polling results. Our approach is based on Gaussian process regression (§3.1) and entails building independent GP models for each race conditioned on available polling outcomes. The model includes only polls and hyperpriors (discussed below).

Denote by $\mathcal{C}$ the set of all candidates in all races we wish to reason about. We will consider the unknown proportion of voters preferring candidate $i \in \mathcal{C}$ in a given race a function of time, writing $f_i \colon (-\infty, 0] \to [0, 1]$. Here the domain of the function is time (measured in days), where the election is defined to occur at time $t = 0$ days. Let $\mathcal{T}_i$ be the set of times when opinion polls for candidate i were conducted.[4]

We model the trend of voter preferences f_i as a sum of an underlying linear trend $a_i + b_i t$, with smooth nonlinear deviations from this trend, $\eta_i(t)$. We place independent Gaussian priors on the intercept a_i (i.e., the prior mean of the voter preferences on Election Day) and slope b_i of the linear trend, and will place an independent, zero-mean Gaussian process prior on the nonlinear component η_i. The covariance function K determines the correlation of deviations from the linear trend as a function of time and was taken to be taken to be identical across all races. Here we used the Matérn covariance function with $\nu = \frac{3}{2}$, which models isotropic, once-differentiable functions (Rasmussen and Williams 2006). This covariance function has two parameters that we will estimate from training data: a length scale ρ determining the scale of correlations, and an output scale λ determining the pointwise variance of the process. Intuitively, we can think of ρ as determining the "window" of days over which non-linear deviations are estimated and λ as controlling the degree of non-linearity we expect such that higher values lead to more dramatic deviations.

The model can be summarized as:

$$f_i(t) = a_i + b_i t + \eta_i(t) \tag{1}$$
$$a_i \sim \mathcal{N}(\bar{a}_i, \sigma_a^2 = 0.1^2) \tag{2}$$
$$b_i \sim \mathcal{N}(0, \sigma_b^2 = 0.002^2) \tag{3}$$
$$\eta_i(t) \sim \mathcal{GP}(0, K), \tag{4}$$

4. We date polls based on the first day they are in the field.

where the covariance function for the nonlinear deviations is

$$K(t, t'; \rho, \lambda) = \lambda^2 \left(1 + \frac{\sqrt{3}d}{\rho}\right) \exp\left(-\frac{\sqrt{3}d}{\rho}\right); \qquad d = |t - t'|.$$

The priors on the linear parameters are constructed to be broad for the slope (so that over a time period of roughly 100 days the linear trend could plausibly assume any possible value) and vaguely informative for the intercept; we will discuss the intercept mean parameter $\bar{a}_i$ shortly.

The above prior choices induce the following joint prior over the voter preference, as shown in (5). Notice that our model provides an automatic marginalization over the linear slope parameters, since the covariance function in our GP model has absorbed the hyperparameters controlling the prior distribution of the linear function parameters.

$$f_i(t) \sim \mathcal{N}(\mu_i(t), V(t, t')) \tag{5}$$

$$\mu_i(t) = \bar{a}_i \tag{6}$$

$$V(t, t') = \sigma_a^2 + tt'\sigma_b^2 + K(t, t') \tag{7}$$

Our goal is to infer the latent voter preference trend from opinion poll outcomes, which are by their nature noisy. Our approach is to model the observed poll outcomes as binomially distributed, then approximate each binomial with a Gaussian for mathematical convenience. This will allow closed-form exact inference, yielding a posterior GP belief about underlying voter trends conditioned on available data.

For a candidate $i \in \mathcal{C}$ with S_i conducted opinion polls, let $\mathcal{D}_i = \{t_{is}, n_{is}, x_{is}\}, (s = 1, \ldots, S_i)$ denote the outcomes of all available polls involving that candidate. Here t_{is} is the time of the poll, n_{is} is the sample size of the poll, and x_{is} is the number of polled people expressing support for the candidate. Dropping subscripts momentarily, consider one such polling outcome $(t, x, n) \in \mathcal{D}$. We make the natural assumption that the number of supporters x is binomially distributed given the sample size n and the true (unknown) voter support f at time t:

$$x \sim \text{Binomial}\big(n, f(t)\big) \tag{8}$$

Unfortunately it is not possible to condition a Gaussian process exactly on observations with a binomial likelihood. However, sample sizes for election polls tend to be large enough (often in the hundreds) that we can safely make a Gaussian approximation to the likelihood by moment matching. Here we also explicitly consider an additional general noise term σ^2, which designates another level of noise stemming from the polling data. Let $p = x/n$ to be the observed proportion of support in a poll, so (8) could be approximated with

$$p \sim \mathcal{N}\big(f, \hat{p}(1 - \hat{p})/n + \sigma^2\big) \tag{9}$$

where we have substituted the estimated $\hat{p}$ for the true unknown proportion $f(t)$ in the variance (in our case, $\hat{p} = p$ *after observation*). This likelihood is now conjugate to our GP prior on f and allows exact inference.

Let us define the vector $\mathbf{p}$ to entail a set of polling outcomes at observed at times $\mathbf{t}$: $p_s = x_s/n_s$, and further define $\mathbf{B}$ to be a $S \times S$ diagonal matrix with $B_i i = p_i(1 - p_i)/n_i + \sigma^2$; this is the approximate noise variance for each of these measurements appearing in (9). As described previously, the posterior predictive distribution of the voter preference at any time t is:

$$f(t) \mid D \sim \mathcal{N}\big(\mu_{f\mid\mathcal{D}}(t), K_{f\mid\mathcal{D}}(t,t)\big) \tag{10}$$

$$\mu_{f\mid\mathcal{D}_i}(t) = \mu(t) + V(t,\mathbf{t})\big(\mathbf{V}+\mathbf{B}\big)^{-1}(\mathbf{p}-\boldsymbol{\mu}) \tag{11}$$

$$K_{f\mid\mathcal{D}}(t,t') = V(t,t') - V(t,\mathbf{t})\big(\mathbf{V}+\mathbf{B}\big)^{-1}V(\mathbf{t},t'), \tag{12}$$

where $\boldsymbol{\mu}$ and $\mathbf{V}$ are the prior mean and covariance of $f(\mathbf{t})$, respectively. Although we may make forecasts for any time t, we are especially interested in public opinion on Election Day, $f(t=0)$. This will also be normal following 10. For notational convenience below, we will again use subscripts for candidates and write $f_i(0) \sim \mathcal{N}(\mu_{f_i}, K_{f_i})$.

The candidate-level model is completed by choosing prior parameters for the intercepts $\{\bar{a}_i\}$ and the set of shared hyperparameters $\boldsymbol{\omega} = (\rho, \lambda, \sigma^2)$. Here σ^2 represents the level of unmodeled noise remaining in the polling data, λ controls the degree to which the time trend deviates from linearity, and ρ represents the "bandwidth" of these smoothing window for these non-linear deviations.

We chose informative, but wide priors for $\{\bar{a}_i\}$ so that projections could be made in races with few or zero polls but that polling data would quickly swamp the prior when plentiful. Since the standard deviation for the prior is set at 0.1, any vote share within ±30 points of the prior should be well supported. To set the values, we ran a simple regression in the training set with normalized vote share as the dependent variable and using just party, lagged partisan vote index (PVI), and level or prior experience.[5] While certainly not an accurate model by itself, it proved to be an adequate prior.[6]

For $\boldsymbol{\omega}$, we adopt a leave-one-year-out (loyo) cross-validation approach using the training period from 1992-2016. First, we define the search region of output scale and shared noise both to be $[0, 0.05]$. We search length scale with a minimum of 7 days and a maximum of 56 days.[7] Empirically, we select prospect $\boldsymbol{\omega}$'s in the validation procedure from a low-discrepancy Sobol sequence (Sobol 1979) in the search region, since it covers the space more efficiently than a grid. We fit the *complete* model, including the election-level model, for each of 100 values of $\boldsymbol{\omega}$ at each time horizon (τ) leaving out each year in turn.

For example, for choosing the hyperparameters for the model predicting four weeks in advance of the election, we used all of the polling data up to day $t = -28$. We then fit the GP models and trained the election-level model described below leaving out each cycle in turn. We then generate out-of-sample predictions for vote shares and choose the parameter setting that maximized the loyo log-likelihood averaged across all cycles.

The chosen hyperparameters for each time horizon are shown in Table 1, and example candidate-level models for one candidate (Kamala Harris in 2016) are shown at various time horizons in Figure 2. As illustrated by the figure, this approach to parameter selection yields models that begin as linear far from Election Day but become increasingly non-linear as τ approaches zero. Note also that the uncertainty in $f_i(0)$ narrows considerably in the final run up to Election Day.

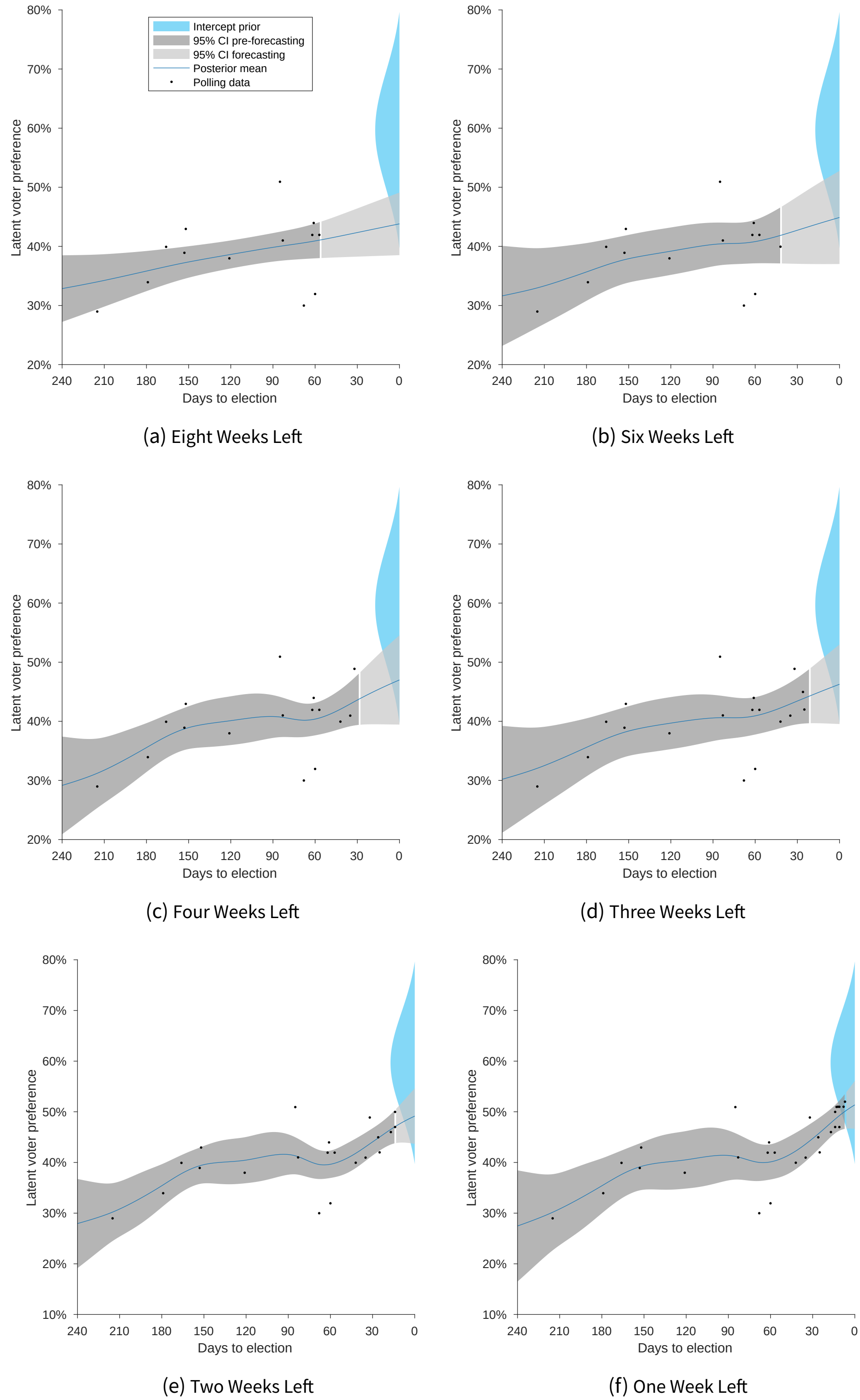
5. Letting w represent standard regression coefficinets, this model was just

$$\text{vote share}_i \sim w_0 + w_1 \text{experienced}_i + w_2 \text{party}_i + w_3 pvi_i \times \text{party}_i$$

where *party* was coded as 1 for Republicans, -1 for Democrats, and 0 otherwise. We lag PVI to the previous presidential cycle and experienced is a binary indicator for whether or not the candidate has ever previously held elected office. The entire cycle is left out of the training data when constructing these priors. Indeed, the construction of the priors themselves could be included in the cross-validation strategy we use for $\boldsymbol{\omega}$ below.

6. The few exceptions where the prior proved to be wildly off were for third-party candidates. Future efforts to forecast might create a separate prior structure for third-party candidates.

7. These ranges were determined from earlier exploratory work.



(a) Eight Weeks Left

(b) Six Weeks Left

(c) Four Weeks Left

(d) Three Weeks Left

(e) Two Weeks Left

(f) One Week Left

Figure 2. Voter Preference Estimate for Kamala Harris in 2016 California race at various time horizons. Points represent individual polls, the blue distribution is the prior, the dark gray region is the estimated latent trend, and the light-gray region is the projected latent trajectory.

Table 1. Learned Model Hyperparameters.

Time horizon (τ)	Length Scale (ρ)	Output Scale (λ)	Noise std (σ)
0	38.3	4.60%	2.89%
7	49.1	4.70%	4.60%
14	45.2	3.90%	3.28%
21	54.4	2.90%	4.84%
28	52.1	3.20%	3.67%
42	39.9	1.90%	4.92%
56	44.8	0.70%	3.47%

3.3 Election-level model

The goal of the candidate-level model (§3.2) is to project forward at any time horizon a predictive distribution of latent public support on Election Day, $f_i(0)$. The election-level model in this section takes $f_i(0)$ as an input and combines it with additional contextual factors to generate a predictive distribution. Our method is based on Dirichlet regression, which generalizes beta distribution to multivariate cases that allows prediction of the election vote shares of multiple candidates. We model the concentration parameters in the Dirichlet distribution with a linear regression model depending on voter preference belief and contextual factors or "fundamentals." Since exact inference is impossible, we use Markov Chain Monte Carlo (MCMC) sampling method to perform inference.

Consider arbitrary race j with m_j candidates and a specific candidate i. We assume a simple linear model that maps the voter preference $f_i(0)$ to the underlying concentration parameters α_i of a Dirichlet distribution. Although there are many possible covariates we could include, we found that very few actually improved out-of-sample predictive performance. We therefore include only the lagged partisan voting index (PVI) generated by Cooks political report, and an indicator for the experience of the candidate where a one indicates the candidate has held elected office before and it is coded zero otherwise. We also include a year-level random effect to accommodate unmodeled electoral "swings" associated with specific election cycles. PVI and the year random effects are reverse coded by party.

Collect $\boldsymbol{\alpha}_j = (\alpha_{1j} + \tilde{\alpha}, \ldots, \alpha_{m_j j} + \tilde{\alpha})$ from all candidates in the race ($\tilde{\alpha} \geq 0$). The base parameter $\tilde{\alpha}$ here is introduced for two reasons. First, it can reduce variance of samples and thus stabilize the MCMC sampling. Second, $\tilde{\alpha}$ encodes the prior belief on how equally the vote shares should be distributed without any additional information. We assume that the actual vote share vector $\mathbf{y}_j = (y_{1j}, \ldots, y_{m_j j})^\top$ is distributed with a Dirichlet distribution $\mathbf{y} \sim \mathrm{Dir}(\mathbf{y}; \boldsymbol{\alpha}_j)$. However, α_{ij} is also a linear function of $f_i(0)$ and observed fundamentals. Due to uncertainty of voter preferences, we also need to integrate over the distribution of $f_i(0)$; in our case, the distribution of $f_i(0)$ is a truncated Gaussian. In total, we assume the election outcomes follow the following data generating process:

$$f_i(0) \sim \mathcal{N}(\mu_{f_i}, K_{f_i}), \quad 0 \leq f_i \leq 1 \tag{13}$$

$$\alpha_{ij} = \theta_1 f_i(0) + \theta_2 \mathrm{party}_{ij} \times \mathrm{pvi}_j + \theta_3 \mathrm{experience}_{ij} + \mathrm{party}_{ij} \times \gamma_{\mathrm{year}}, \quad \alpha_{ij} \geq 0 \tag{14}$$

$$\gamma_{\mathrm{year}} \sim \mathcal{N}(0, \sigma^2_{\mathrm{year}}) \tag{15}$$

$$(y_1, \ldots, y_m)^\top \sim \mathrm{Dirichlet}(\tilde{\alpha} + \alpha_1, \ldots, \tilde{\alpha} + \alpha_m), \quad \tilde{\alpha} \geq 0 \tag{16}$$

Here we allow party to be equal to 1 for Democratic candidates, −1 for Republican candidates, and 0 for independents.[8] This allows for PVI and the year random effects to have opposite effects by

8. We code third party candidates who regularly caucus with one party as belonging to that party. So, for instance, Sen.

party.

The model is completed by placing proper but vague priors across all parameters. Recall that the outcome is vote share $y_{ij} \in [0,1]$. The priors for the θ parameters are set to be wide based on the scale of the relevant variable. Specifically, we set independent truncated Gaussian priors on the θ coefficients and the year-level random effects.

$$\theta_1 \sim \mathcal{N}(0, 100^2); \theta_1 > 0$$
$$\theta_2 \sim \mathcal{N}(0, 10^2); \theta_2 > 0$$
$$\theta_3 \sim \mathcal{N}(0, 10^2); \theta_3 > 0$$
$$\sigma^2_{\text{year}} \sim \text{Gamma}(1, 0.5);$$

We can combine all of these parameters together in $\boldsymbol{\Theta} = \left(\{\theta\}, \{\gamma_{\text{year}}\}, \sigma^2_{\text{year}}, \tilde{\alpha}\right)$ and let $\mathbf{z}_j$ be the vector of contextual factors for election j. We obtain the posterior $p(\boldsymbol{\Theta} \mid \{\mathbf{y}_j\}, \{\mathbf{z}_j\}, \{f_i(0)\})$ using MCMC estimation. Specifically, we use no-U-turn sampling in the stan programming language (Hoffman and Gelman 2014; Carpenter *et al.* 2017).

With this posterior, the final predictive distribution of future election outcomes with new $\{\mathbf{f}^*_i(0)\}, \{\mathbf{z}^*_j\}$ will be defined by (13), (14),(15) and (16) marginalized by the posteriors:

$$p(\mathbf{y}_j \mid \{\mathbf{z}^*_j\}, \{f_i(0)^*\}, \{\mathbf{y}_j\}, \{\mathbf{z}_j\}, \{f_i(0)\}) = \int p(\mathbf{y}_j \mid \{\mathbf{z}^*_j\}, \{f_i(0)^*\}, \boldsymbol{\Theta}) p(\boldsymbol{\Theta} \mid \{\mathbf{y}_j\}, \{\mathbf{z}_j\}, \{f_i(0)\}) d\boldsymbol{\Theta} \quad (17)$$

A final issue is how to handle the dynamic nature of our forecasting task. While we have the complete set of polls for elections in our training set, when making real-time forecasts we have only the polls up to the current date. Training the model on the complete set of polls (all the way up to Election Day) is likely to lead higher weight being assigned to polling data and poor predictive performance at more remote time horizons. For instance, the coefficients for the Dirichlet regression component in the election model may put too much confidence on the polling. As noted above, this same issue applies to hyperparameter selection for the candidate-level model.

To address this concern, we train the complete model at various time horizons denoted by τ. For any threshold, we discard all data where $|t| < \tau$. Thus, when $\tau = 28$, we ignore all polls in the training data closer than 28 days to the election. This again helps calibrate the model for the levels of accuracy we can expect at various horizons. Table 2 provides the posterior summaries for these parameters at horizons ranging from $\tau = 0$ to $\tau = 56$ (eight weeks before the election). As expected, the θ parameter associated with $f_i(0)$ increases as Election Day approaches while the fundamental parameter become relatively less important.

Figure 3 shows the posterior prediction for Senator Kamala Harris in 2016 for various time horizons. Note that the outcome (marked with the vertical blue line) is near the center of the posterior for all horizons, but that the prediction becomes more concentrated as Election Day nears. This reflects both more certainty in the estimate of $f_i(0)$ and changing weights in the Dirichlet regression.

4 Empirical evaluation

In this section, we investigate our model using historical polling data and vote shares in U.S. Senate elections from 1992 to 2018. Throughout our model building process, we held out the 2018 election as a test case and it was not involved in any hyperparameter tuning, variable selection, or other stages. Therefore, we can assess the model's predictive performance using the 1992-2016 period

Bernie Sanders is coded as a Democrat.

Table 2. Summaries of Dirichlet Regression models at different horizons for 2018 model

Horizon	Parameter	Mean	SD	2.5%	97.5%
$\tau = 0$	$\tilde{\alpha}$	20.1	0.096	12.6	28.5
	Voter preference	505	1.47	410	607
	Party:pvi	0.78	0.003	0.53	1.05
	Experience	1.96	0.033	0.09	5.37
	σ_{year}	5.39	0.033	3.17	8.47
$\tau = 7$	$\tilde{\alpha}$	19.4	0.118	10.8	29.5
	Voter preference	531	1.37	422	644
	Party:pvi	0.74	0.004	0.45	1.05
	Experience	2.20	0.041	0.07	6.14
	σ_{year}	5.40	0.035	3.03	8.71
$\tau = 14$	$\tilde{\alpha}$	16.2	0.102	8.54	24.3
	Voter preference	444	1.37	346	556
	Party:pvi	0.53	0.003	0.30	0.79
	Experience	2.58	0.043	0.13	6.55
	σ_{year}	4.08	0.046	1.77	6.9
$\tau = 21$	$\tilde{\alpha}$	10.2	0.104	2.75	18.4
	Voter preference	420	1.41	327	529
	Party:pvi	0.32	0.003	0.08	0.56
	Experience	2.15	0.039	0.09	5.79
	σ_{year}	3.71	0.037	1.46	6.41
$\tau = 28$	$\tilde{\alpha}$	10.5	0.104	3.33	18.5
	Voter preference	378	1.58	285	489
	Party:pvi	0.26	0.003	0.05	0.51
	Experience	2.28	0.036	0.13	5.82
	σ_{year}	2.96	0.044	0.59	5.80
$\tau = 42$	$\tilde{\alpha}$	5.48	0.083	0.51	12.1
	Voter preference	346	1.47	255	459
	Party:pvi	0.26	0.003	0.06	0.49
	Experience	2.73	0.041	0.14	6.75
	σ_{year}	3.16	0.036	0.96	5.89
$\tau = 56$	$\tilde{\alpha}$	6.03	0.062	1.29	11.1
	Voter preference	206	0.879	150	271
	Party:pvi	0.24	0.002	0.09	0.40
	Experience	2.42	0.031	0.23	5.19
	σ_{year}	2.71	0.024	1.24	4.70

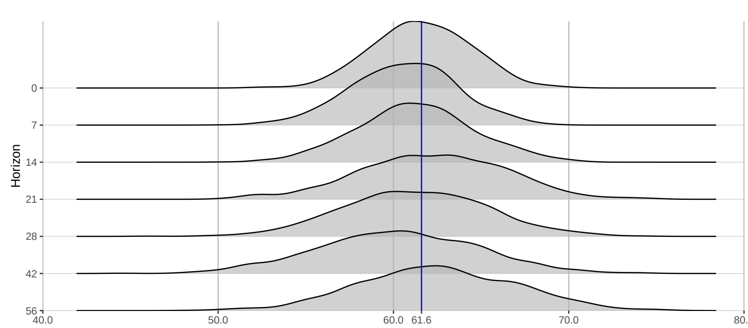


Figure 3. Posterior predictive densities for Sen. Kamala Harris in the 2016 California Election at different time horizons. The blue vertical line indicates the final vote share. Note that the posterior narrows noticeably as the time horizon shrinks.

but also approximate its true out-of-sample performance using the 2018 data. In the next Section, we report predictions for 2020 actually made in advance of Election Day.

4.1 Data and evaluation criteria

We obtained opinion polls and election outcomes of all senate election races from 1992 to 2018 from www.fivethirtyeight.com and from *CNN*. On average 16 polls were conducted for each race, although some races such as the 2016 Florida election had over 80. Most of the surveys are conducted two weeks to four months prior to election, with a median number of respondents of 635. Over 470 entities have conducted these polls, but several active pollsters contribute over half of them, including Rasmussen Reports, Mason-Dixon Polling, Public Policy Polling, SurveyUSA, YouGov, SurveyMonkey, Quinnipiac, and Zogby Interactive. To guarantee credibility, we eliminated polls sponsored by parties or candidates since unfavorable polls from these sources are not released.

We also acquired the partisan voting indices in every election cycle for each election state. The Cook Partisan Voting Index (PVI) is a measurement of how strongly a United States congressional district or state leans toward the Democratic or Republican Party, compared to the nation as a whole. For example, PVI for California in 2018 is 10.76, indicating a strong preference for Democratic candidates, while PVI for the pro-republican state Texas in 2018 is -7.02. For each candidate, we coded partisan affiliation and past experience (whether or not the held office). Where not provided in the *CNN* data, we coded these manually using ballotpedia.com.

To evaluate performance, we examine both the forecasting precision and the validity of our model. Hence, we consider the following measures: the averaged root-mean-squared-error (RMSE) between the expectation of the Dirichlet posterior samples and actual vote shares, the prediction accuracy of winners for election defined by higher winning probability (we calculate the winning probability of each candidate as the proportion of samples with the highest vote shares in Dirichlet posteriors), the coverage rate of actual vote shares in 95% credible intervals for Dirichlet posteriors, the averaged multinomial predictive likelihood and the averaged log-scaled Dirichlet predictive likelihood (LL). RMSE and prediction accuracy focus on the precision of the forecasting ability, while coverage rate focuses on the validity of the claimed credible intervals. The two likelihood measures serve as out-of-sample evaluation criteria for both vote share (Dirichlet) and final outcome (multinomial) that also reflect the uncertainty in the full posterior.

4.2 Baselines

We compare the performance of our combined GP and Dirichlet regression (GP+DR) model against three benchmarks. First, we compare our model to the dynamic Bayesian model in Linzer 2013. This model was developed for predicting state-level results in presidential elections, but we adjusted it for predicting Senate races. Intuitively, this model is a dynamic Bayesian random walk (BRW) model similar to the non-linear component in the model described above except that latent public opinion is assumed to be a random walk. We use the same informative prior for $\{\bar{a}\}$ as used above and use the tuning parameters and basic estimation procedures as described in Linzer 2013.[9]

Second, we consider a baseline Dirichlet Regression model that uses a Bayesian linear regression model to forecast voter preferences. We refer the second baseline as LM+DR. To ensure a fair comparison, we choose the same priors for the linear coefficients as those in GP priors. We also chose the σ^2 hyperparameter using the same cross-validation approach described above. This model is, in essence, the same as what we report above without allowing for deviations from linearity. Finally, we examine the performance of the GP model in isolation excluding the Dirichlet regression portion of the hierarchy.

9. The major deviation from the original model implementation is we do not include a national over time trend since the senate races are far more independent than the state-level presidential races.

Table 3. Predictive accuracy in the 1992-2016 Period

		Days until Eleciton Day (τ)						
	Model	56	42	28	21	14	7	0
RMSE	GP+DR	0.082	0.074	0.068	0.065	0.059	0.056	0.053
	GP	0.102	0.096	0.090	0.087	0.082	0.078	0.075
	LM+DR	0.082	0.073	0.069	0.065	0.060	0.055	0.055
	BRW	0.085	0.081	0.078	0.076	0.072	0.065	0.060
95% Coverage	GP+DR	0.922	0.921	0.927	0.942	0.940	0.954	0.942
	GP	0.800	0.869	0.861	0.859	0.818	0.823	0.746
	LM+DR	0.926	0.918	0.926	0.926	0.930	0.939	0.941
	BRW	0.548	0.523	0.504	0.510	0.514	0.514	0.527
Predictive accuracy	GP+DR	0.895	0.895	0.919	0.915	0.920	0.933	0.946
	GP	0.879	0.897	0.913	0.915	0.923	0.936	0.936
	LM +DR	0.881	0.898	0.910	0.913	0.925	0.936	0.934
	BRW	0.798	0.814	0.827	0.828	0.847	0.859	0.898
APLL Vote Share	GP+DR	1.397	1.489	1.635	1.700	1.786	1.897	1.947
	GP	0.548	1.125	1.170	1.185	1.138	1.214	0.940
	LM+DR	1.397	1.510	1.617	1.661	1.759	1.854	1.877
	BRW	−2.171	−1.606	−1.513	−1.126	−1.048	−0.377	−0.143
APLL Winner	GP+DR	−0.134	−0.115	−0.101	−0.097	−0.091	−0.081	−0.075
	GP	−0.169	−0.117	− 0.101	−0.097	−0.092	−0.078	−0.072
	LM+DR	−0.135	−0.117	−0.103	−0.101	−0.091	−0.081	−0.079
	BRW	−0.371	−0.337	−0.342	−0.289	−0.252	−0.234	−0.172

4.3 Results

First, we present results from a leave-one-year-out cross validation exercise where each election cycle from 1992-2016 was held out. However, since we followed an identical procedure when choosing our hyperparameters above, there may still be some risk of overfitting. We therefore also present results for the 2018 election separately which serves as a stronger out-of-sample test.

We simulate a real forecasting scenario and examine the forecasting ability at various horizon τ's. Specifically, we consider horizons of four months, three months, six weeks, four weeks, two weeks, one week and Election Day, where $\tau = 56, 42, 28, 21, 14, 7, 0$. As noted above, Table 1 summarizes parameters learnt for the candidate-level model used throughout this exercise.

Table 3 shows the results for the leave-one-year-out cross validation exercise for the 1992-2016 period. The results show that GP+DR model on average outperforms the other baselines across metrics. The closest competitor is actually the LM+DR model, which performs quite well in terms of coverage and accuracy. This is explained in part by the fact that the GP model itself is mostly linear at distant horizons and when there is little polling data. However, the non-linear component in the GP does provide measurable improvements over the linear version in the final lead up to the election when the hyperparameters most enable non-linear deviations (see Figure 2 above).

Table 4 shows these results for the 2018 election cycle. These results again show that the GP+DR model is accurate and well calibrated at all time horizons. The nearest competitor is the LM+DR model, which again becomes less comparable as Election Day nears and the nonlinearities become more pronounced in the candidate-level models. Note that the GP+DR model correctly predicted all but one election at $\tau = 0$ (Arizona) and all but two elections at the $\tau = 14$ horizon (Arizona and Nevada).

Figure 4 shows the predictions, 95% predictive confidence intervals, and outcomes for the 2018 senate elections with $\tau = 7$. The results show that all election outcomes fell within the 95% credible range an that on average the forecast tracked the actual election outcomes very closely. Moreover,

Table 4. Predictive accuracy in 2018

	Model	Days until Eleciton Day (τ) 56	42	28	21	14	7	0
RMSE	GP+DR	0.068	0.063	0.055	0.054	0.050	0.047	0.042
	GP	0.082	0.081	0.067	0.066	0.063	0.059	0.054
	LM+DR	0.067	0.063	0.055	0.054	0.050	0.046	0.043
	BRW	0.066	0.065	0.061	0.061	0.058	0.053	0.050
95% Coverage	GP+DR	0.972	0.972	1.000	1.000	1.000	1.000	1.000
	GP	0.916	0.986	1.000	0.986	0.986	0.958	0.944
	LM+DR	0.972	0.972	1.000	1.000	1.000	1.000	1.000
	BRW	0.653	0.681	0.653	0.681	0.681	0.736	0.653
Predictive accuracy	GP+DR	0.942	0.914	0.914	0.914	0.942	0.942	0.971
	GP	0.885	0.914	0.914	0.914	0.942	0.857	0.914
	LM +DR	0.914	0.942	0.942	0.914	0.914	0.885	0.914
	BRW	0.829	0.829	0.857	0.857	0.857	0.857	0.857
APLL Vote Share	GP+DR	1.729	1.854	2.061	2.070	2.155	2.288	2.374
	GP	1.541	1.663	1.797	1.811	1.899	1.999	2.032
	LM+DR	1.750	1.830	2.022	2.033	2.125	2.270	2.350
	BRW	−1.371	−1.331	−2.455	−0.564	−2.442	−0.250	−0.228
APLL Winner	GP+DR	−0.108	−0.099	−0.082	−0.080	−0.067	−0.072	−0.072
	GP	−0.108	−0.095	−0.077	−0.083	−0.062	−0.072	−0.068
	LM+DR	−0.109	−0.100	−0.083	−0.078	−0.071	−0.070	−0.069
	BRW	−0.511	−0.503	−0.503	−0.472	−0.464	−0.444	−0.458

the elections where the model is incorrect at a seven-day range are also among the closest contests in that cycle (Arizona and Nevada). Finally, the width of the credible interval can vary significantly depending on the number and recency of polls for that election. For instance, the credible intervals for Wyoming are very large reflecting the fact that we had only one poll for that race with a sample size of 858. This contrasts markedly with, for instance, Missouri where dozens of polls were reported in the final months.

5 Predicting the 2020 Election

Finally, we turn to the task of predicting the 2020 senate elections. For this cycle, we again acquired all data from the fivethirtyeight.com website. Following the procedures outlined above, we exclude all partisan polls and and mark each poll based on the first day it was fielded. No third-party candidate was been polled with sufficient regularity for inclusion in our model. We exclude the Louisiana senate race and the special election in Georgia which are scheduled for special elections after November 3rd.[10] We used the same hyperparameters as shown in Table 1, but refit the Dirichlet regression using the complete 1992-2018 training period.

Figure 5 shows the results for several candidate-level models at the six-week and two-week horizons. Note that the model tends towards linearity at a farther time horizon or when the polling data is more sparse. Note also, that the prior is suggestive but does not dominate the polling data even when the horizon is still six weeks out.

The predictive densities for each candidate are shown in Figure 6. At a two week time horizon, the model predicted that the Democrats were favored to take control of the US Senate, but that

10. We do this because our training data does not include election results from any runoff or special election that took place on date that was not the national Election Day. Our position is that it is simply unwise to extrapolate to elections for which we do not have adequate training data. One possible extension to this model would be to better accommodate runoff and elections that take place outside of November.

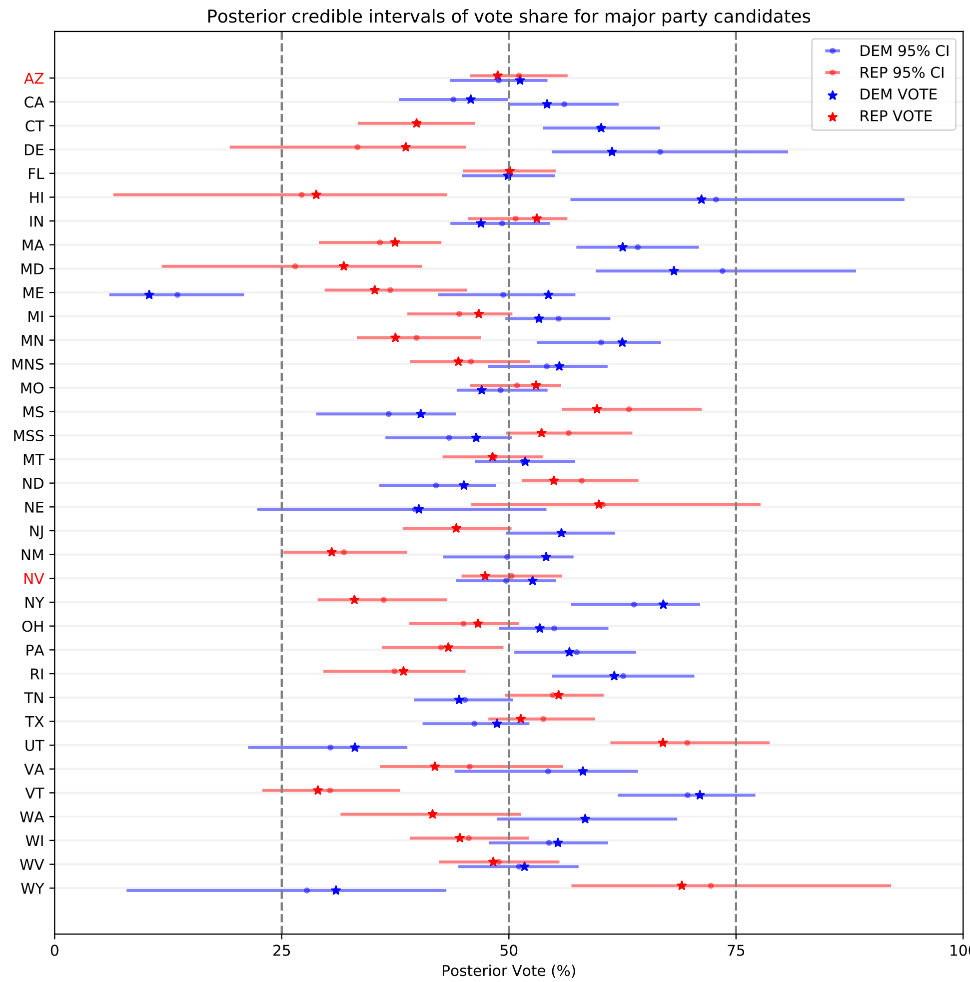
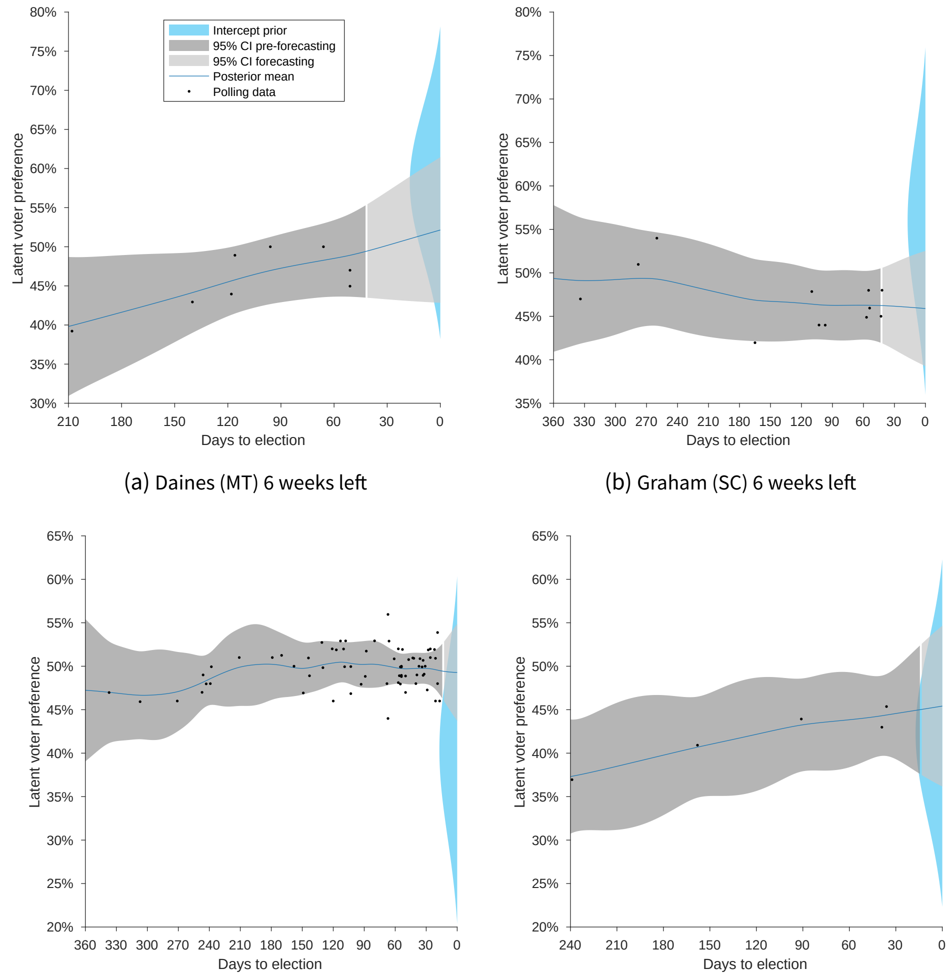


Figure 4. Forecast for 2018 at one week time horizon for major party candidates. Stars indicate actual vote share, while points and confidence intervals reflect posterior mans and 95% credible predictions. The California election had two Democrats and we coded Sen. Angus King of Maine as Democrat. Red font for state names indicates an incorrect prediction. This model also included a forecast for Libertarian candidate Gary Johnson in New Mexico (posterior median 0.185, 95% CI [0.117, 0.247], outcome 0.154).



(a) Daines (MT) 6 weeks left

(b) Graham (SC) 6 weeks left

(c) Kelly (AZ) 2 weeks left

(d) Boiler (KS) 2 weeks left

Figure 5. Example 2020 candidate-level models showing trajectories of latent public opinion

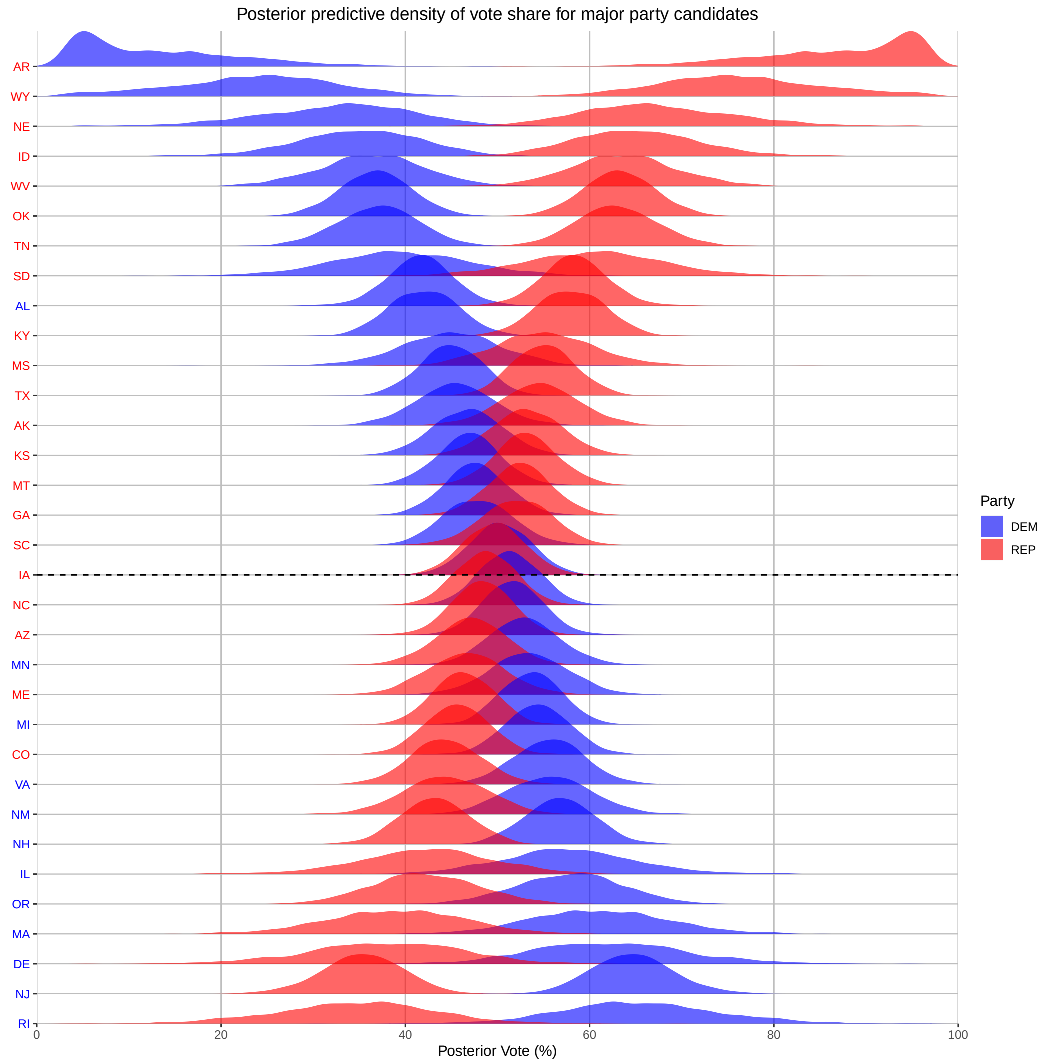


Figure 6. Predicted vote share densities for 2020 at two-week time horizon for major party candidates. States are arranged in order of increasing probability of Democrats winning. Colors of state names indicate pre-election partisan control, while the dashed black line indicates the point at which the Democrats would achieve a 51-seat Senate majority. The plot excludes Louisiana and the Georgia special election.

many elections would be close. The closest states are predicted to be MS, AK, TX, MT, SC, KS, GA, NC, IA, AZ, ME, MI, VA, CO, and MN (states here are ordered by the degree to which they favor the Democratic candidates).[11] At a two week window, the closest races were expected to be North Carolina and Iowa, which conforms largely with other forecasts available at this time.

6 Conclusion

In this paper, we offer a novel approach to dynamic election forecasting that combines both poll-based and fundamentals-based forecasting. Although the model itself is somewhat complex, in the end it includes only a few variables: polling, PVI, candidate experience, and party. The novelty here is not in what factors go into the model, but how they are combined to create accurate, well-calibrated predictions.

Our approach contains two basic stages. The first stage is to treat polling data as a probabilistic

11. One noticeable difference in our model from those in media organizations is that we do not have VA as a "safe" Democratic seat. This simply reflects the lack of non-partisan polling for his race. We expect that the model will update significantly once more polls are released.

representation of latent public support for a candidate, where this latent support has a linear and non-linear trend. By fitting a model to this trend, we can accurately predict forward to where public opinion will be on Election Day. Second, we then incorporate predictions about this latent position into a Dirichlet regression that uses historical data and a few simple features about the election to estimate the degree to which polling can be used to predict elections based on historical data. A final innovation is that we train the data completely at different time horizons to ensure that our final predictions reflect an appropriate level of uncertainty.

While we believe that this model improves upon other Senate forecasting models in the literature, it could be improved in several ways. First, we might better extend it to handle unusual cases like runoff elections or special elections (e.g., the 2020 Georgia special election). We could in theory also extend the model to account for "house effects" of various polling firms or weight more accurate firms more highly in the candidate-level model. Likewise, we could try alternative variables to include in the construction of the candidate-level prior parameter or in the election-level model. However, adding such complications should be done with caution as they may lead overfitting. Many variables (e.g., money raised or incumbency status) should be reflected in the polling data. Once we have conditioned on latent public support, the list of accurate predictors of outcomes is much smaller.

A further shortcoming of our model is that our model does not allow online-update: forecasters have to learn from scratch customized hyperparameters for every new horizon. In Table 1, the learnt length scale and noise standard deviations are somewhat constant across horizons, while the learnt output scales shrink at earlier horizon. When computation capability is limited, practitioners may use the same optimal hyperparameters across horizons and warp the output scale according to the forecasting horizon.

A final extension would be to adjust the model to handle elections at different levels. This model would be relatively straightforward to extend to, for example, gubernatorial races. However, more significant adjustments may be needed for lower (e.g., races for the U.S. House of Representatives) or higher (presidential) elections. Lower-level races are unusual in that there is even less polling data available for most races, which may require heavier reliance on contextual factors or election-cycle factors such as "momentum" for a specific party. Meanwhile, presidential races usually offer many more polls but the historical training data is necessarily very sparse and the outcomes (state-level results) are much more correlated across states. Researchers wishing to extend this basic approach to those settings should think carefully about how to construct the election-level and candidate-level models to account for these important differences.

Supplementary Material

(This is dummy text) For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.xxxx.xx.

References

Abramowitz, A. I. 2008. "Forecasting the 2008 Presidential Election with the Time-for-Change Model." *PS: Political Science & Politics* 41 (4): 691–695.

Campbell, J. E. 2018. "The seats-in-trouble forecasts of the 2018 midterm congressional elections." *PS: Political Science & Politics* 51 (S1): 12–16.

Campbell, J. E., and K. A. Wink. 1990. "Trial-heat Forecasts of the Presidential Vote." *American Politics Research* 18 (3): 251–269.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. "Stan: A probabilistic programming language." *Journal of Statistical Software* 76 (1).

Erikson, R. S., and C. Wlezien. 2008. "Leading Economic Indicators, the Polls, and the Presidential Vote." *PS: Political Science & Politics* 41 (4): 703–707.

Fair, R. C. 1978. "The Effect of Economic Events on Votes For President." *The Review of Economics and Statistics* 60 (2): 159–173.

Hoffman, M. D., and A. Gelman. 2014. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15 (1): 1593–1623.

Hummel, P., and D. Rothschild. 2014. "Fundamental models for forecasting elections at the state level." *Electoral Studies* 35:123–139.

Jackman, S. 2005. "Pooling the polls over an election campaign." *Australian Journal of Political Science* 40 (4): 499–517.

Klarner, C. 2008. "Forecasting the 2008 us house, senate and presidential elections at the district and state level." *PS: Political Science and Politics* 41 (4): 723–728.

Klarner, C. E. 2012. "State-level forecasts of the 2012 federal and gubernatorial elections." *PS, Political Science & Politics* 45 (4): 655.

———. 2013. "2012 presidential, US House, and US Senate forecasts." *PS, Political Science & Politics* 46 (1): 44.

Klarner, C., and S. Buchanan. 2006. "Forecasting the 2006 Elections for the United States Senate." *PS: Political Science and Politics* 39 (4): 849–855.

Lewis-Beck, M. S., and C. Tien. 2008. "The Job of President and the Jobs Model Forecast: Obama for '08?" *PS: Political Science & Politics* 41 (4): 687–690.

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

Lockerbie, B. 2012. "Economic expectations and election outcomes: The Presidency and the House in 2012." *PS, Political Science & Politics* 45 (4): 644.

Rasmussen, C. E., and C. K. Williams. 2006. *Gaussian processes for machine learning.* MIT Press.

Sobol, I. M. 1979. "On the systematic search in a hypercube." *SIAM Journal on Numerical Analysis* 16 (5): 790–793.

Stoetzer, L. F., M. Neunhoeffer, T. Gschwend, S. Munzert, and S. Sternberg. 2019. "Forecasting elections in multiparty systems: a Bayesian approach combining polls and fundamentals." *Political Analysis* 27 (2): 255–262.

Walther, D. 2015. "Picking the winner (s): Forecasting elections in multiparty systems." *Electoral Studies* 40:1–13.