Supplementary Appendices for: How American Politics Insures Electoral Accountability in Congress

Danny Ebanks* Jonathan N. Katz† Gary King‡

April 4, 2024

*Ph.D. Candidate, California Institute of Technology; DEbanks@Caltech.edu, DannyEbanks.com.
†Kay Sugahara Professor of Social Sciences and Statistics, California Institute of Technology; JKatz.Caltech.edu, JKatz@Caltech.edu
‡Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University; GaryKing.org, King@Harvard.edu.

Contents

1 Likelihood Function **2**
1.1 Allowing for Uncontested Elections 2
1.2 Full Model 3

2 Ablation Studies **5**

3 Alternative Modeling Assumptions **8**

4 Computational Details **9**

5 Model Specification **11**

6 Probability of Defeat for Individual Incumbents **14**

7 Probability of Defeat for Southern and Non-Southern Incumbents **15**

1 Likelihood Function

This appendix provides the full likelihood function for our model, including all the features described in Section 2.2 in the paper, as well as situations where elections are uncontested for both the election at hand and for the lagged vote as a covariate. We begin by extending our model to uncontested elections in Section 1.1 and then bring all the parts together in Section 1.2.

1.1 Allowing for Uncontested Elections

In the standard approach, the vote in uncontested elections is often recoded to fixed values such as $v_{it} = 0.25$ for Democrats running uncontested and $v_{it} = 0.75$ for Republicans running uncontested, or sometimes uncontested elections are deleted entirely. We instead formally distinguish between the observed vote v_{it} and the *effective vote* v^*_{it}, defined as the vote proportion that would be observed if the election had been contested (e.g., King and Gelman, 1991). The effective vote is observed $v^*_{it} = v_{it}$ in contested elections but unobserved if one party runs unopposed. We then impute unobserved values (for uncontested elections) during Bayesian estimation simultaneous with the rest of the model.

To model v^*_{it} when unobserved, we replace the outcome variable v_{it} in Equation 2 with the effective vote, and add a "censoring assumption": candidates who run unopposed would have won even if the election were contested. This assumption is intuitive, probably accounts for why the district was uncontested in the first place, and is a special case of the assumption made by Katz and King (1999). We then replace Equation 2 with

$$v^*_{it} \sim \text{ALT}(\mu_{it}, \phi^2_t, \nu_t), \tag{1}$$

and write the likelihood function for an election district that is fully contested as $\text{ALT}(v_{it} \mid \mu_{it}, \phi^2_t, \nu_t)$, for a district where a Democrat runs uncontested as $\psi_{it} \equiv \int_0^{0.5} \text{ALT}(v^* \mid \mu_{it}, \phi^2_t, \nu_t) dv^*$, and for a district where a Republican runs uncontested as $1 - \psi_{it}$. The integral implements the censoring assumption.

In Figure 1, we show the historical rate of uncontestedness in U.S. Congressional elections, which ranges from 21 percent in 1954 to 4 percent in 1996. Rather than drop

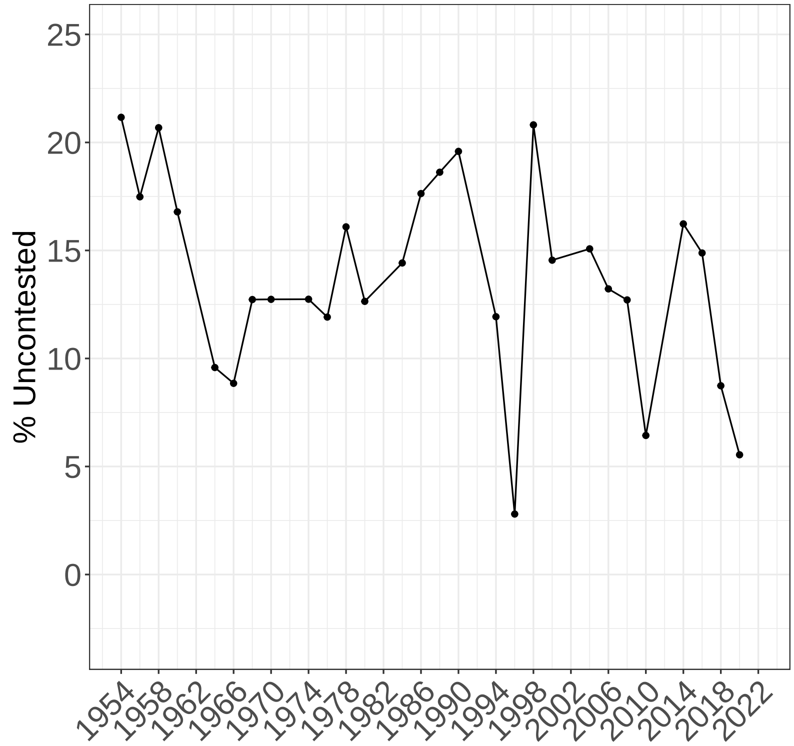


Figure 1: Uncontested Elections over Time

these estimates which compose a nontrivial share of the data in any given election year, we impute predictive vote shares within our fully Bayeisan model, described in Section 1.2.

To account for missing data due to uncontestedness, we jointly estimate a multivariate model that predicts the uncontested vote share and missing lagged uncontested vote share. To this end, we assume that missing vote share is a censored variable where an uncontested incumbent is constrained to always win. That is, we know uncontested vote share data are not missing at random.

In Figure 2, we show that our predictions are bimodal around modes centered at 25 and 75 percent vote shares. These predictions are in line for historical estimates of uncontested vote shares (Gelman and King, 1994).

1.2 Full Model

To write the full likelihood function, define an uncontestedness indicator U_{it} as 1 if the Democrat runs uncontested, 0 if contested, and -1 if the Republican runs uncontested

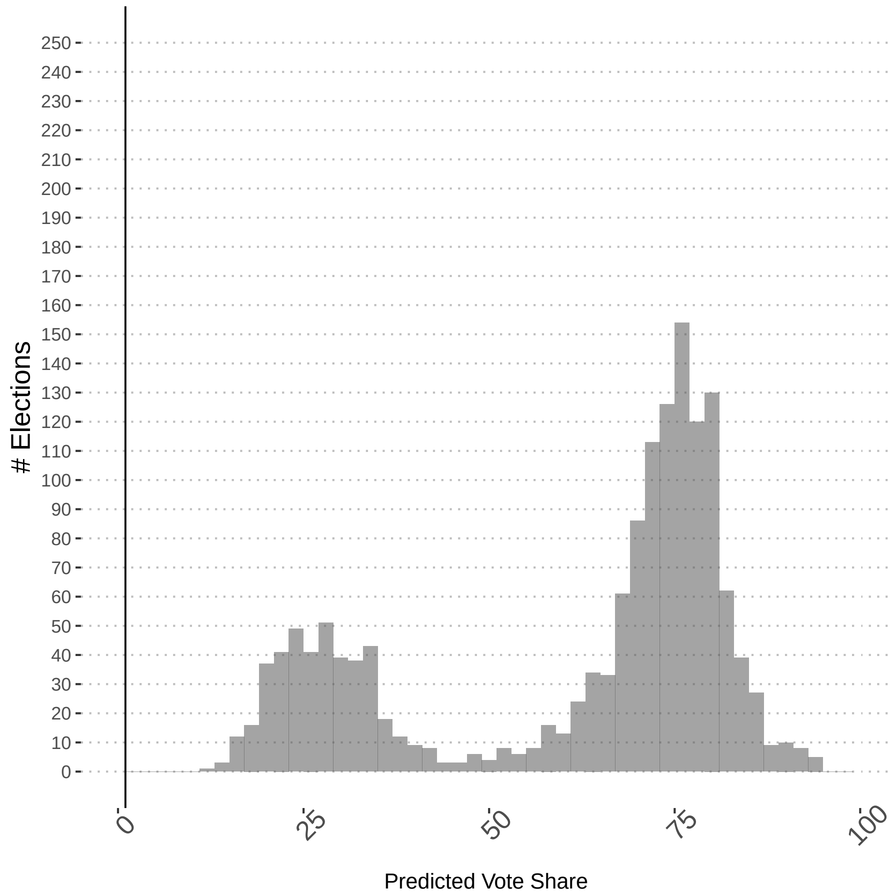


Figure 2: Histogram of Predicted Values for Uncontested Elections

in district i and time t. Then partition elections into four sets depending on whether the current election i, t and its lag $i, t-1$ are contested or uncontested. Denote CC as the set of all elections for which $U_{it} = 0$ and $U_{i,t-1} = 0$; UC as the set of elections for which $U_{it} \neq 0$ and $U_{i,t-1} = 0$; CU as the set of elections where $U_{i,t} = 0$ and $U_{i,t-1} \neq 0$; and UU as the set of elections for which $U_{it} \neq 0$ and $U_{i,t-1} \neq 0$. Then the likelihood function factors into four parts corresponding to these sets:

$$L = \left(\prod_{i,t \in \{\text{CC}\}} L_{it}^{\text{CC}} \right) \left(\prod_{i,t \in \{\text{UC}\}} L_{it}^{\text{UC}} \right) \left(\prod_{i,t \in \{\text{CU}\}} L_{it}^{\text{CU}} \right) \left(\prod_{i,t \in \{\text{UU}\}} L_{it}^{\text{UU}} \right) \tag{2}$$

each of which we now define.

The first component of the likelihood, for when election i, t and $i, t-1$ are both contested, is by far the most prevalent for the US congress. The likelihood for observation i, t is then simply

$$L_{it}^{\text{CC}} = \text{ALT}(v_{it} \mid \mu_{it}, \phi_t^2, \nu_t). \tag{3}$$

The second component of the likelihood accounts for which party is running uncontested at time t:

$$L_{it}^{\mathrm{UC}} = \mathbf{1}(U_{it} = 1)\psi_{it} + \mathbf{1}(U_{it} = -1)(1 - \psi_{it}), \tag{4}$$

where our censoring assumption from Section 1.1 implies that $\psi_{it} \equiv \int_0^{0.5} \mathrm{ALT}(v^* \mid \mu_{it}, \phi_t^2, \nu_t) dv^*$, given the indicator function defined as $\mathbf{1}(a) = 1$ if a is true and 0 otherwise, for any statement a.

To write the third component, where the lagged value of the effective vote is unobserved (because it is uncontested), we require a prior distribution for how this variable is distributed. The posterior will be computed from the entire model, but to begin we need an assumption about this prior. One option is to let $v_{i,t-1}^*$ be a censored ALT when unobserved (and equal to v_{it} when observed) but this creates a substantial computational burden with little substantive benefit. Instead, we find we can represent almost all relevant information by assuming that, when unobserved, $v_{i,t-1}^* \sim \mathcal{N}(Z_{i,t-1}\alpha_t, \sigma_v^2)$, with $Z_{i,t-1}$ a vector of covariates such as lagged presidential vote in a congressional district and incumbency status. Then this component of the likelihood is

$$L_{it}^{\mathrm{CU}} = \int_{-\infty}^{\infty} \mathrm{ALT}(v_{it} \mid \mu_{i,t}, \phi_t^2, \nu_t) \cdot \mathcal{N}(v^* \mid Z_{i,t-1}\alpha_t, \sigma_v^2) dv^*, \tag{5}$$

where the unobserved lagged effective vote v^* is included in X and so contributes to μ_{it}.

For the final component of the likelihood, we use features of all three previous components, so that

$$L_{it}^{\mathrm{UU}} = \mathbf{1}(U_{it} = 1)\psi_{it}' + \mathbf{1}(U_{it} = -1)(1 - \psi_{it}'), \tag{6}$$

where

$$\psi' = \int_{-\infty}^{\infty} \int_0^{0.5} \mathrm{ALT}(v \mid \mu_{i,t}, \phi_t^2, \nu_t) dv \cdot \mathcal{N}(v^* \mid Z_{i,t-1}\alpha_t, \sigma_v^2) dv^*.$$

2 Ablation Studies

We introduce four modeling innovations for our generatively accurate model: a national trend, coefficient stability, local uniqueness, and electoral surprises (see Section 2.2 in the paper). In this section, we conduct "ablation studies," where each model component is

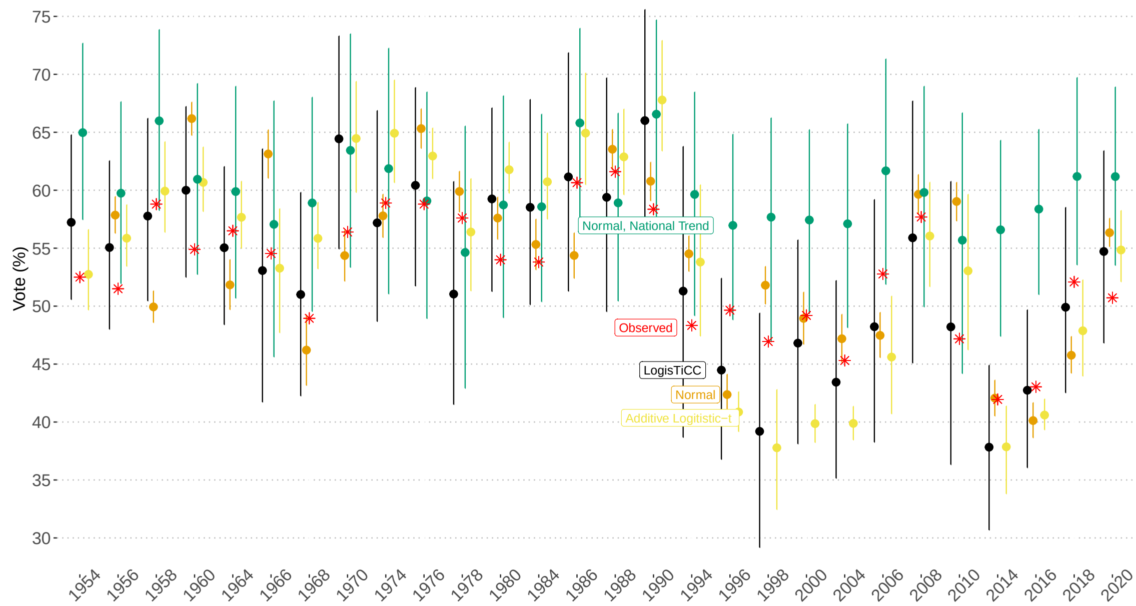


Figure 3: Comparison of Model Calibration under Ablation

sequentially removed to show how the model degrades. The results here demonstrate that each separate model component is essential to achieve the performance we report.

The linear-normal model treats the data as having 435 independent district-level observations for each election year. In reality, congressional elections data have high levels and sophisticated patterns of dependence among voting outcomes across districts. In Figure 3, we replicate the calibration exercise from Figure 4 in the paper, which reports the model predictions and observed values for the median congressional seat in the given election year. We report results for three ablated models. We give the normal model with none of the modeling innovations (in gold); a model with neither a national trend assumption nor coefficient stability, but with an additive logistic student-T (ALT) assumption on the error term (in yellow); and a model with normal errors, but with a national trend and coefficient stability (in green).

We would expect a well-calibrated model to contain the true value of the median seat's vote share about ~ 95 percent of the time. To that end, we see that the normal (with none of our innovations) fares poorly, correctly containing the true value for the median seat for only 25% of the elections. If we switch to the ALT specification, we achieve a 40% accuracy rate, which is still inadequate, but better than normal alone. When we assume normal errors with a national trend and coefficient stability, we achieve 64% accuracy. Under the ablated models, we find that the coefficient stability and national trend alone allow the model to achieve about 60 percent accuracy in our calibration, while the ALT error assumption achieves 40 percent accuracy. Only the inclusion of all our modeling assumptions allowed us to achieve 100 percent accuracy.

In Figure 4, we reproduce Figure 4 from the paper with additional information. As in the original, the linear-normal model (in gold), which assumes independence, has confidence intervals that are extremely overconfident, and the LogisTiCC (in black) has accurately calibrated intervals. To these results, we add a version of our LogisTiCC that zeros out the parameters that model the dependence among elections. These include the national swing parameter σ_η and also our covariate stability parameter $\sigma_\beta > 0$ which, after transforming to the vote scale, also allows for some dependence across districts. In this model, we retain local uniqueness.

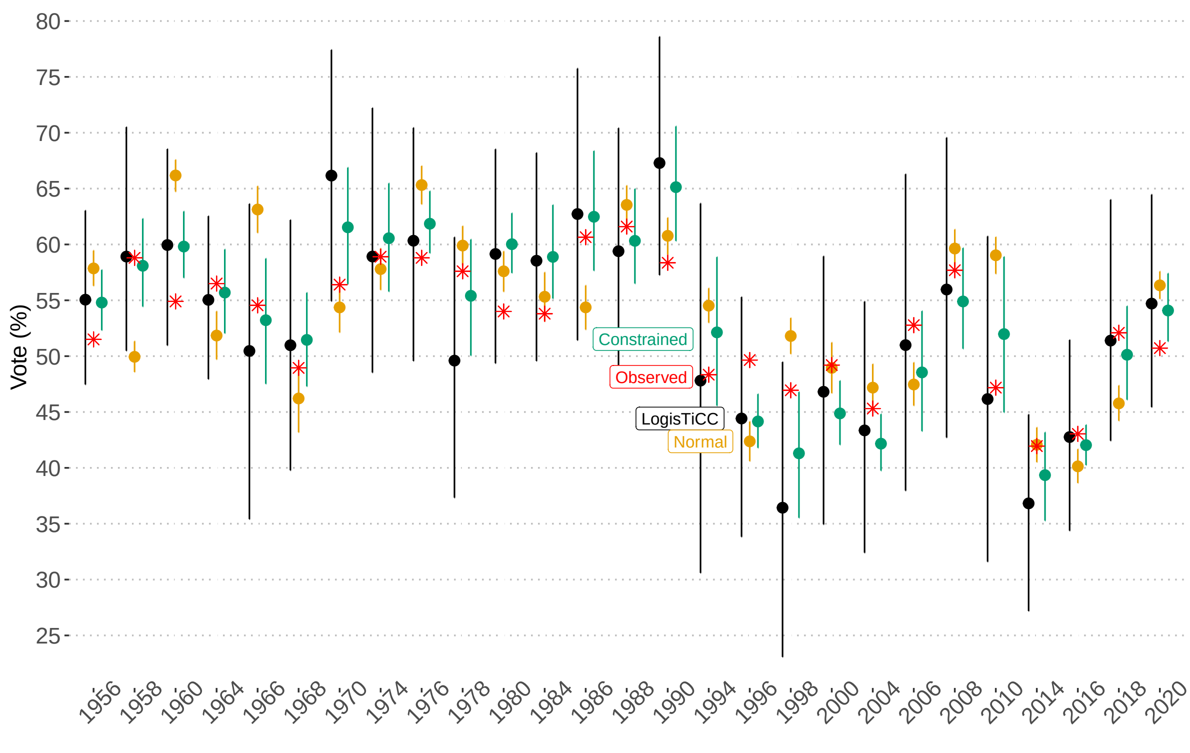


Figure 4: Expected Vote Share of the Median House Seat (95 Percent Credible Interval)

Thus, we add to Figure 4, in green, estimates from the LogisTiCC model constrained to give predictions with zero cross-district independence, while retaining local uniqueness. While this set of assumptions reduces model's overconfidence relative to the normal somewhat, the model is still much too overconfident. Only when we allow our full ALT error structure with cross-district correlations are the out-of-sample model predictions from the LogisTiCC well-calibrated to the historical data (in black). Indeed, under the linear-normal error structure, the incumbent party will never lose control of the House of Representatives. Under the ALT without cross-district correlation, the uncertainty gets larger so that the incumbent party is sometimes forecast to lose an election, but clearly not as often as it should. By introducing cross-district correlation, our forecasts are well-calibrated.

3 Alternative Modeling Assumptions

We tried to eliminate any feature of our model not required for accurate out-of-sample validation and accurate uncertainty intervals (see the previous section), to include additional features that would improve performance, and to consider alternative specifications that might be easier to understand.

As we have shown in the main text, the linear-normal model is poorly calibrated for congressional elections. Additionally, we fit a linear-normal Student-t, which failed because it lacked the flexibility and asymmetry in the tails provided by the additive logistic t (ALT). The Additive Logistic Normal failed because it could not properly capture the levels of concentration (nearly 60 percent in the 1980s) exhibited in Figure 10(a), nor did it accurately capture surprises with appropriate tails. Fitting an independent ALT, that is without contemporaneous correlations, is not well-calibrated because it misses the correlations due to year-to-year swings in the national trend or dependence due to the stability of coefficient estimates.

We also tried other flexible distributions. We tried the Beta distribution, which models the unit interval directly, but produces poorly calibrated results because it, like the IID normal, does not capture appropriate levels of concentration or tail behavior. We also tried mixture distributions and errors which, while flexible, wound up being highly model dependent, poorly identified, and computationally fragile.

We also attempted to find alternative correlation structures, besides time mixed effects and district random effects on the logit scale, such as regional mixed effects. Besides districts in the south and outside the south, there was little predictable regional variation. Districts in the North, West, and Southwest do not seem to systematically vary, conditional on other covariates. Our covariates include an indicator for districts in the South that varies over time to capture what appear to be the most important systematic effects. In terms of covariate selection, we made choices for easy comparison to the literature. Our general model structure, like the normal, can easily accommodate other indicators if discovered by future scholars to be relevant.

4 Computational Details

The standard approach is usually estimated with a linear regression for forecasting (i.e., dropping γ_i) or, for other quantities of interest, via an approximate two-step procedure designed to avoid computational challenges that were difficult in the 1990s (see Gelman and King, 1994).

Because of improvements in computation and Bayesian modeling, we estimate our LogisTiCC model via a fully Bayesian specification of Equation 2, beginning with the likelihood in Equation 2. We implement the model in "brms," open-source software that calls on Stan, which uses Hamiltonian Markov Chains (HMC) sampling to draw from the posterior distribution of a mixed-effects model (Bürkner, 2018). In practice, we draw 50,000 samples of the posterior distribution from the Bayesian mixed-effects representation (and 200,000 for the final run). When lagged congressional vote share is a covariate, we drop the first election of each redistricting decade to fit the model. Our Bayesian methods are computationally demanding but efficient, which enables us to analyze large legislatures, and does not require asymptotic assumptions, which is especially important for legislatures like the small U.S. Senate class up for election in any one year, small national legislatures, or the many small state houses. We are also able to simulate quantities of interest directly from the full joint posterior distribution of the predicted values and parameters, which means researchers can easily calculate any relevant quantity of interest, along with accurate and calibrated uncertainty estimates.

In order to achieve valid calibrated uncertainty estimates, we use conservative search parameters for Stan's HMC sampler. We set a delta step of 0.999, set a maximum tree depth of 11, draw $16,000$ samples with a warm up of $6,000$ iterations on 20 chains run in parallel, for a total of $200,000$ posterior samples per parameter. All Markov Chains successfully converged, with no divergent transitions, Rhats of 1 across all parameters, well-mixed chains, and no breaches of maximum tree depth.

We employ weakly informative priors for estimation convenience. In our case, because we have an average of about $1,500$ elections per decade, we do not require regularization to identify model parameters, although our weakly informative priors reduce computational time for HMC convergence. Priors are useful for speeding computation but, in our data, the choice of hyperprior parameter values does not have a material effect on empirical results. In fact, even under robustness tests with tight priors far away from their implied values, core results are largely unchanged. The one exception is that a strong, tight prior on the midterm penalty simultaneously increases the penalty and lowers the national trend. That is because with only 3 years of data per decade per leave-one-year-out run, the difference between the national trend and midterm penalty is sometimes weakly identified. We impose a empirically backed assumption on midterm penalty — that it is negative with a wide variance over the range. The specific values we use are $\sigma_\beta, \sigma_\omega, \sigma_{tk}, \sigma_i, \sim$ Exponential(0.2) , $\nu \sim \Gamma(3, 0.5)$ and Midterm Penalty $\sim$ Normal$(-0.5, 0.2)$.

In Figure 5, we show the prior and posterior histograms for the coefficient on our predictor of the "normal" vote. This figure shows that our weakly informative prior is diffuse, while the coefficient posterior is tightly estimated around its mean, confirming that our model estimates are mostly a function of the data rather than priors. We have also found that small changes in the priors have little substantive consequences for our estimates.

Statistical results would likely be less robust to the choice of these parameters in legislators smaller than the US House. In applications with small legislatures, researchers

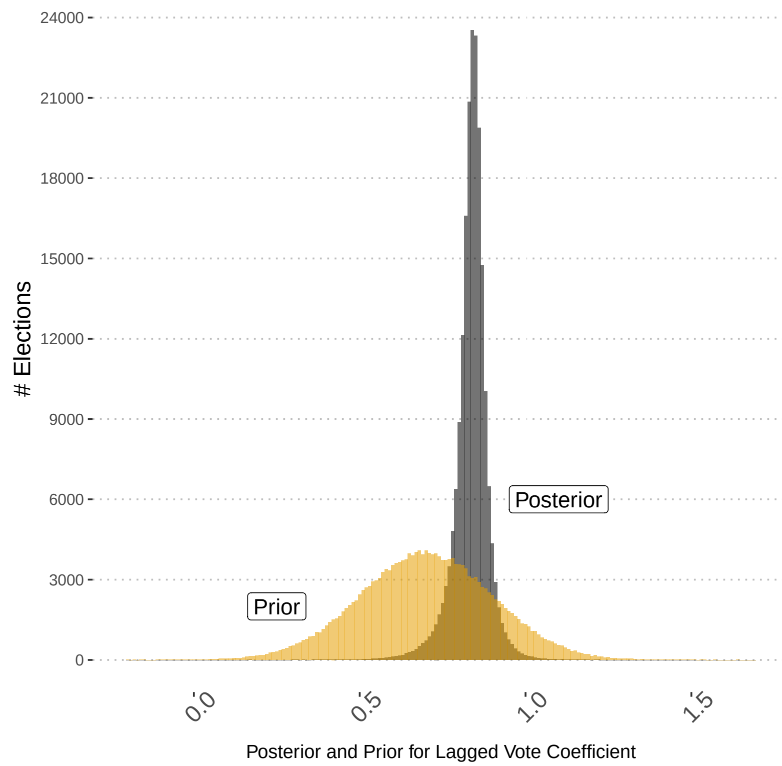


Figure 5: Posterior vs. Prior Densities

should carefully consider the impacts of both prior specification and sampler behavior to guarantee statistically valid inferences from the HMC chains.

5 Model Specification

The unit of analysis for fitted models is district i in election year t, with the Democratic proportion of the vote as the outcome variable. For all calibration computations and estimates of descriptive quantities, we use the same specification with the same priors and the same computational details. For all analyses, we drop years ending in 2, when nearly all districts are usually redrawn. We drop districts that are redrawn mid-cycle and elections where a top-two primary resulted in two candidates from the same party. We calibrated the model decade-by-decade, dropping one election year at a time. We then forecast the dropped year from the remaining data. That is, we looped through every decade and fit the model 28 times, dropping one year at a time. All calibrations and probability of defeat were computed in this fashion. We tried alternative estimation routines, such as using the previous decade to estimate empirical priors and gradually including more data as we forecast further into the decade. These strategies always yielded similar results, while being more prior-reliant and more computationally demanding.

For descriptive statistics such as concentration, partisan bias, coefficient estimates on lagged vote, and marginal incumbency advantage, we also estimated a full in-sample model where the data were fit to all four years of available data per decade. In both the leave-one-election-out and in-sample models, we report a forecast integrating over the temporal uncertainty induced by the national trend. We integrate over the fundamental uncertainty assumed by our functional form.

To compute the probability of incumbent defeat, we run a counterfactual forecast on the leave-one-election-out estimates that assumes every incumbent at time $t-1$ runs again at time t and compute the integral over $[0, 50)$. If the incumbent loses their primary, we treat their probability of defeat as 100%.

We report all variables and how each is coded in Table 1. For variables whose coefficients are estimated at the decade level, we mark a check under Decade in the table. For variables whose coefficients we estimate as an annual deviation, we check Annual. Coefficients for the midterm penalty, lagged uncontested, and involuntary retirements require multiple years of data and are estimable at the decade level, and thus possible under our LogisTiCC modeling framework, but not under the normal approach run one election at a time. The coefficient for the midterm penalty is only identified for multiple years of data with at least one midterm and one presidential election. Voluntary retirements and lagged uncontested do not have enough data at the annual level to reliably estimate coefficients, so we estimate them only at the decade level. The incumbency and retirement variables are mutually exclusive and would be collinear if we included voluntary Democratic retirements, which we drop as the base variable. The national trend and local uniqueness are random intercepts for year and congressional district, respectively.

Table 1: Definition of Variables and Coding

Variable	Coding	Decade	Annual
Lagged Democratic Vote	Lag of Democratic Vote. N/A for redistricted districts (including all districts in the first year after redistricting.)	✓	✓
Incumbent Republican	1 if the incumbent is Republican and elects to run in the current election. 0 otherwise	✓	✓
Incumbent Democrat	1 if the incumbent is Democrat and elects to run in the current election. 0 otherwise	✓	✓
Voluntarily Republican Retirement	1 if the incumbent voluntarily chooses not to run and is a Republican. (Retires, runs for higher office). 0 otherwise	✓	✓
Involuntary Retirement	1 if the incumbent dies in office, resigns due to health reasons, is expelled, is convicted, resigns due to scandal, wins higher office before completing their term, loses their primary election, or is appointed to another office. 0 otherwise	✓	
South	1 if a district is in the former Confederacy. 0 otherwise.	✓	
Uncontested	1 if the district only has a Democrat running for office. -1 if the district only has a Republican running. 0 otherwise	✓	✓
Lagged Uncontested	Lag of the Uncontested Indicator	✓	
Midterm Penalty	1 if the incumbent is a Democrat and in the president's party during a midterm. -1 if the incumbent is a Republican and in the president's party during a midterm. 0 otherwise	✓	
National Trend	Intercept for each year		✓
Local Uniqueness	Intercept for each district	✓	

6 Probability of Defeat for Individual Incumbents

We now supplement Figure 7 with Figure 6, to provide further information about individual incumbents' probability of defeat. We plot this probability for individual districts (vertically) by the lag for the same districts (horizontally). We do this on the logit scale for graphic clarity, with axis labels on the probability scale for interpretability, and colors for each of the four groups, following the color scheme from Figure 6 in the paper. Each dot is one district with election years included where the lag is possible within a single redistricting decade.

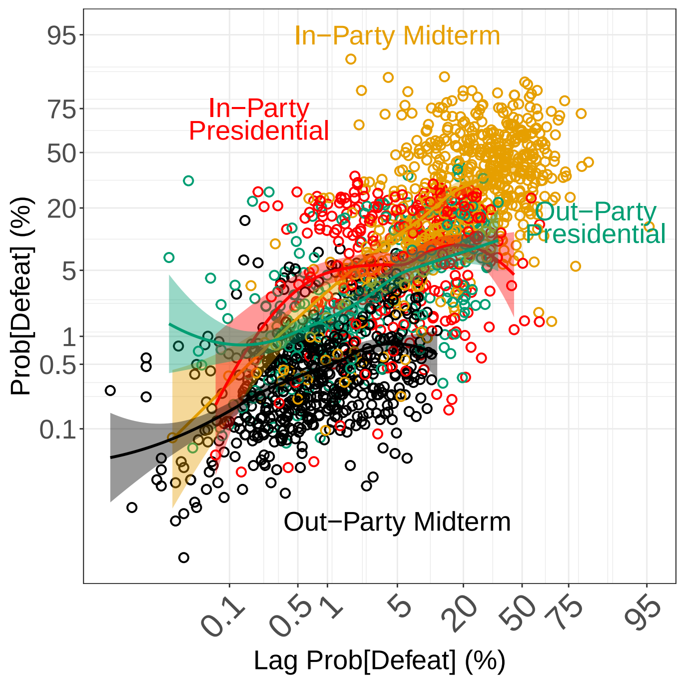
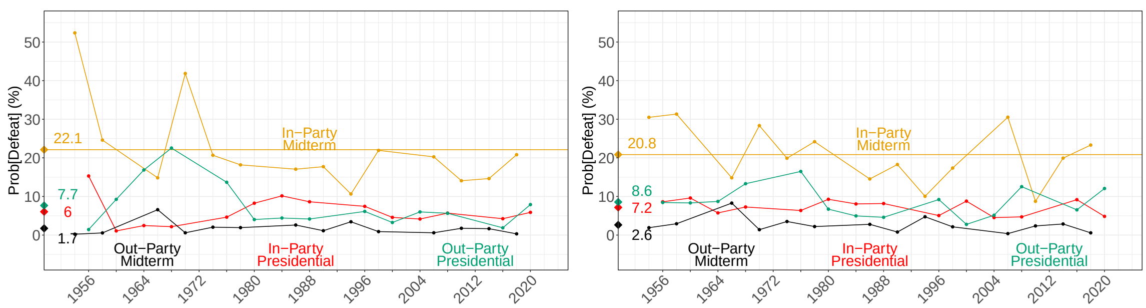


Figure 6: Probability of Incumbent Defeat by its Lag (logit Scale)

The overall pattern of the four categories is replicated in this figure: The higher probability of defeat for in party incumbents can be seen by the gold dots being closer to the top right in the figure. We can also see that relatively few districts at the lower left have small probabilities of defeat for both elections. But most remarkably is the wide scatter, indicating high unpredictability for individual incumbents. For example, conditional on a lagged probability of 20% for in party incumbents during the midterm, the probability of defeat in the next election can range from under 1 percent to over 85%. This means what may feel safe at one point in time can be dangerous at another.

7 Probability of Defeat for Southern and Non-Southern Incumbents

We supplement paper Figure 8 with the disaggregation by region in the style of Figure 6 in the main paper. We show remarkable similarlty between regions in both the means of the probability defeat (22.1% in the South for in-party incumbents in a midterm compared with 20.8% in the rest of the country) and stability over the same period.



(a) South

(b) Rest of the Country

Figure 7: Probability of Defeat Predicts by Election Type and Region

References

Bürkner, Paul-Christian (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms". In: *The R Journal* 10.1, pp. 395–411. URL: `https://doi.org/10.32614/RJ-2018-017`.

Gelman, Andrew and Gary King (May 1994). "A Unified Method of Evaluating Electoral Systems and Redistricting Plans". In: *American Journal of Political Science* 38.2, pp. 514–554. URL: `j.mp/unifiedEc`.

Katz, Jonathan N. and Gary King (Mar. 1999). "A Statistical Model for Multiparty Electoral Data". In: *American Political Science Review* 93.1, pp. 15–32. URL: `bit.ly/mtypty`.

King, Gary and Andrew Gelman (Feb. 1991). "Systemic Consequences of Incumbency Advantage in the U.S. House". In: *American Journal of Political Science* 35.1, pp. 110–138. URL: `bit.ly/SystCs`.