How American Politics Ensures Electoral Accountability in Congress

March 24, 2024

Abstract

An essential component of democracy is the ability to hold legislators accountable via the threat of electoral defeat, a concept that has rarely been quantified directly. Well known massive changes over time in indirect measures — such as incumbency advantage, electoral margins, partisan bias, partisan advantage, split-ticket voting, and others — all seem to imply wide swings in electoral accountability. In contrast, we show that the (precisely calibrated) probability of defeating incumbent US House members has been surprisingly constant and remarkably high for two-thirds of a century. We resolve this paradox with a generative statistical model of the full vote distribution to avoid biases induced by the common practice of studying only central tendencies, and validate it with extensive out-of-sample tests. We show that different states of the partisan battlefield lead in interestingly different ways to the same high probability of incumbent defeat. Many challenges to American democracy remain, but this core feature remains durable.

Words: 9917

1 Introduction

We study *electoral accountability*, a universally recognized definitional requirement for healthy democracies. Although many important versions of this concept have been studied (Canes-Wrone, Brady, and Cogan, 2002; Ansolabehere and Jones, 2010; Hirano and Jr., 2012; Nyhan et al., 2012; Tausanovitch and Warshaw, 2018; Fraga and Hersh, 2018; Ansolabehere and Kuriwaki, 2022; Fouirnaies and Hall, 2022; Iaryczower, Lopez-Moctezuma, and Meirowitz, 2022), its most agreed upon essential component — the ability of citizens to hold legislators accountable via threat of electoral defeat — has rarely been directly quantified (Przeworski et al., 2000). This concept is especially valuable because it leaves to citizens, rather than researchers, the choice of how far and in what way the behavior of elected representatives may acceptably deviate from public opinion.

We formalize electoral accountability as the probability that the winner in an election to the US House of Representatives will be defeated if he or she runs in the next election. We estimate this probability of incumbent defeat with a novel generative statistical model (Section 2) and validate it with extensive out-of-sample tests in 14,710 district-level House elections, 1954–2020 (Section 3). Unlike some approaches ours builds on, estimates from this new model are correctly calibrated (e.g., when the model estimates before an election a 0.2 probability, on average 20% of incumbents actually lose their seats). Our analyses also demonstrate that every component of the model included is necessary to achieve this performance and all others we tried but excluded are unnecessary or harmful.

Estimates from this validated generative model reveal that the probability of defeating an incumbent in the US House has been approximately constant and remarkably high over at least the last two-thirds of a century (Section 4.2). This result is surprising because it appears to contradict much of what we know about American politics from the massive variation over time in more commonly used indirect measures, such as incumbency advantage, vanishing marginals, split-ticket voting, partisan bias, partisan advantage, and others (Section 4.3). For example, how can the probability of incumbent defeat be constant when (as is well known and we show below) incumbency advantage has ranged over time between 2 and 10 percentage points, the percent of marginal districts between 15%

and 45%, and partisan bias from fairness (i.e., partisan symmetry) to as much as 8% extra seats given unfairly to one party or the other? We resolve these and other apparent contradictions by modeling not only the commonly reported mean estimates but changes in the variation around these means, which have often been overlooked, downplayed, or mis-modeled by existing approaches (Section 4.4). For one example, an increase in incumbency advantage from 2 to 10 percentage points has long been regarded as synonymous with better reelection chances, but we show that this is not the case if these averages have confidence intervals of (say) 2 ± 0.5 and 12 ± 30, in which case the probability of defeat may not drop at all.

From a normative perspective, electoral accountability has long been seen as essential for evaluating the health of democracy (Mayhew, 1974; Eskridge, 1987). In contrast, the other measures have been extremely useful for characterizing the ever-changing state of American politics, but they are controversial when used alone as indicators of the health of democracy. For example, few now approve of the current high levels of ideological polarization but, in the 1950s, a prominent American Political Science Association report called for policy reforms to increase polarization (APSA, 1950). Partisan gerrymandering has been reviled in academic and public discourse since the founding of the republic, but it is also often treated as "a self-limiting enterprise" that "can lead to disaster for the legislative majority" (Sandra Day O'Connor in *Davis v. Bandemer*, page 478 U.S. at line 152 (1986); Grofman and Brunell 2005; Erikson 1972; Cox and Katz 2002; Kenny et al. 2023). Vanishing marginals were a major worry in the 1970s, but then the pattern itself vanished in the 1980s, after which incumbency advantage began to soar and worry everyone, until it didn't (Erikson, 1971; Tufte, 1973; Mayhew, 2008; Ferejohn, 1977; Fiorina, 1977; Jacobson, 1987; Jacobson, 2015). Although some levels of each of these measures always strike many commentators (especially minority party members) as indicating that American democracy is in crisis, all are important both in and of themselves and to political scientists trying to understand the current state of American politics. We thus focus on electoral accountability as our measure of the health of American legislative democracy, use our generative model to provide a unified understanding of all the divergent patterns

in these diverse measures, and show how they nevertheless all lead to the same high and constant level of the probability of incumbent defeat.

2 Generative Statistical Models of District-Level Elections

By seeking to represent the full data generation process, generative models attempt to produce a more complete understanding of how congressional elections work than non-model-based approaches, albeit in return for more assumptions which then need to be validated. In our case, generative models led us to explanations and quantities of interest that were not previously apparent to us or others. Of course, different approaches can reinforce each other since, once these quantities are discovered, other methods can then be brought to bear on the same problems.

We summarize the standard generative model used in the literature (Section 2.1) followed by our proposed alternative (Section 2.2). We construct our approach to incorporate more substantive knowledge of elections, to simultaneously analyze more elections, and to attend to more of the known statistical issues than previously possible, all within a single full information Bayesian model. This led us to jointly estimate, integrate over, and represent the uncertainty of 3,567 parameters, including coefficients, missing cell values, uncontested districts, and random effect terms.[1]

2.1 Standard

The outcome variable for modeling US congressional elections is the Democratic proportion of the two-party vote, v_{it} for district i and election (time) t. The standard model in the literature is a linear-normal regression of v_{it} on a vector of K covariates X_{it}, with estimation conducted for each election year t run independently. For most applications in the last quarter-century other than forecasting, an independent normal district-level random

[1] One of the reasons our approach has not been tried before is that it would have been computationally infeasible even a few years ago. With highly tuned computational algorithms we developed on a new server (with 20 cores and 128gb of RAM, and software tuned specially to this hardware), one run of our model on a decade of congressional elections data takes only about twenty minutes, although a full analysis of all our data with calibration and strictly out-of-sample evaluation requires about 48 hours of model run time generating about 44gb of output. Along with this paper, we are making available easy-to-use open-source software that implements all our algorithms and methods, as well as a dataset with numerous measures useful for future researchers.

effect (constant over hypothetical or real elections but varying over districts) is added to the regression to model the political uniqueness of individual districts (Gelman and King, 1994, implemented in JudgeIt software).[2]

The content of the covariates represents decades of work by hundreds of political scientists analyzing thousands of elections. We formalize this knowledge by defining X_{it} to include a lagged vote share ($v_{i,t-1}$), incumbent party (the party that won the previous election, with 1 for Democrat and 0 for Republican), incumbency status (1 if the Democratic candidate is an incumbent, 0 for open seat, and -1 for a Republican incumbent), uncontestedness (1 if a Democrat runs uncontested, 0 if contested, and -1 if Republican runs uncontested), and an indicator for the old confederate states. Many other covariates could be included, but most analyses based on the standard model show that they have relatively minor effects conditional on this set.

We summarize this standard model as

$$\begin{aligned} v_{it} &\sim \mathcal{N}(\mu_{it}, \sigma^2) \\ \mu_{it} &= X_{it}\beta_t + \gamma_i \end{aligned} \tag{1}$$

where β_t is a vector of K linear regression effect parameters, $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$ is an independent normal random effect with variance $\sigma_\gamma^2 > 0$, and σ^2 is the variance of the usual homoskedastic regression's independent normal error term.

2.2 Proposed

We now build on the standard model to develop our proposed approach. We keep the same flexibility in covariate choice within a fully Bayesian framework. First, Section 2.2.1 provides a qualitative description of model components designed to reflect knowledge from the literature on elections that had been excluded from the standard approach. Second, Section 2.2.2 builds a single generative Bayesian model, but for expository purposes focuses in the text only on the simple special case where all elections are contested.

[2]Instead of directly estimating γ_i and modeling multiple elections together, which would have been computationally difficult in the 1990s, JudgeIt analyzes one election at a time, after a preprocessing step to estimate how much variation should be attributed to this random effect.

See Supplementary Appendix 1 for the full likelihood, including extensions to uncontested districts; Supplementary Appendix 2 for a set of "ablation" studies that, by sequentially removing each model component, demonstrates how all components are essential to the performance we achieve; Supplementary Appendix 3, which describes many of the alternative modeling assumptions and extra features we found superfluous or harmful; Supplementary Appendix 4 for computational details; and Supplementary Appendix 5 for model specification details.

We also looked extensively for north/south differences. We found no substantive or statistical differences by running our entire model without the Southern states or by including these states with other parameters varying by region. The model we present below includes an indicator for the south, which favors the Democrats in the 1950s but by the 1960s it reaches zero and stays there.

2.2.1 Novel Model Components

Error terms in statistical models are designed to represent "known unknowns," features that reflect political scientists' knowledge of elections too difficult to code in the covariates. For example, the error term in Equation 1 allows for *district uniqueness* by adding a random term γ_i to model the persistence of this uniqueness for any one district i over the elections within a "redistricting regime" (i.e., all elections for which the district geography remains unchanged), beyond changes accounted for by X_i. For example, Minnesota's 7th Congressional District has long been more Republican than the nation as a whole, favoring Donald Trump in 2016 and 2020 by about 30 percentage points. Yet, Democrat Colin Peterson won this seat from 1991 to 2021 because of his personal brand and unusual political preferences, opposing abortion and supporting the border wall, but (perhaps accounting for how he wins the Democratic nomination) highly progressive economic views.

We now add to this model four other "known unknowns," modeling features that reflect valuable political information well understood by students of elections or observable in the data but rarely modeled formally.

First is *covariate effect stability*: β_t varies relatively little over time. For example, the incumbency advantage might range between two and ten percentage points, with only

rare sudden changes. Similarly, the coefficient on the lagged vote is usually in the range of $[0.6, 0.8]$. We add this feature to the model by (a) modeling all elections within a redistricting regime simultaneously, but with no independence or constancy assumptions, and (b) assuming that each element β_{tk} of vector β_t (corresponding to covariate k, $k = 1, \ldots, K$ and time t) comes from the same distribution $\beta_{tk} \sim \mathcal{N}(\hat{\beta}, \sigma^2_{\beta_k})$, where $\sigma^2_{\beta_k} < \infty$; in contrast, estimating each equation separately and independently, as in the standard approach, is equivalent to setting $\sigma_{\beta_k} \to \infty$.[3] The idea here is to borrow strength for the estimate of each parameter in each year from the estimation of the same parameter in other years, but without the rigidity and potential bias that would come from a more "informative" prior. This will be especially valuable in smaller legislatures, such as many state assemblies and senates and the class up for election in the US Senate.[4]

Explicitly modeling covariate effect stability enables us to simultaneously estimate an entire redistricting decade of elections, which then enables us to include covariates not possible to include in the standard approach. We thus included variables for the midterm penalty, and voluntary and involuntary retirements prior to the election.

Second, in recognition of the fact that all House districts are part of the same national election, held on the same day, we allow for positive cross-district covariances by adding a *national swing* random error term, η_t. This term allows all districts in one election year to be affected in roughly the same way by the same national events, over and above the information in X. For example, the 1994 Republican national congressional campaign strategy (known as the "Contract With America") seemed to be a successful heresthetical maneuver (Riker, 1990; Shepsle, 2003) that moved all the districts in the Republican direction by approximately the same amount. Although we do not know ex ante in which direction this term will swing in any one year, we can estimate the variation caused by national swings, which we know occur regularly, and represent this uncertainty in the model with a common random effect for all districts. The result is the well known "approximate

[3]The notation $\hat{\beta}$ is a shorthand reference to empirical Bayes, meaning that this distribution shrinks different covariate effects in the same redistricting regime toward the estimated mean without favoring one's a prior guess; this is equivalent to a fully Bayesian model with the mean in the null space; see Girosi and King 2008.

[4]We could elaborate this assumption by allowing β_t to trend linearly, as a random walk, or as a function of other covariates, but we find no evidence for these alternative approaches in our data.

uniform partisan swing" pattern common across time periods, electoral systems, and even countries (Katz, King, and Rosenblatt, 2020).

Third, we model *district-level political surprises*, including intentional heresthetical maneuvers and unintentional exogenous political events that affect one district's vote at a point in time differently than others and are not included in X. Consider for example the election in Texas' 22nd district in 2006. Tom Delay was the popular Republican House majority leader from the district, regularly winning election by 35 or more percentage points. During the campaign, he was indicted and abruptly resigned. Worse for his party, the deadline to field a candidate on the ballot line had passed and so his party could only field a write-in candidate late in the campaign. The result was that this overwhelmingly Republican district elected a Democrat over the Republican write-in candidate by over 8 percentage points. Equation 1 already includes the usual normal error term that can be used to model surprises, but (as we show below) a normal distribution indicates that deviations from a prediction this large would happen so infrequently that they should almost never be observed. Of course, as every astute election observer knows, big surprises happen regularly, even if we do not know where or when. As we explain below, we will therefore swap out the normal distribution for one that can more appropriately represent these political surprises, while also keeping predictions within the [0,1] interval.

Finally, we correct for two problems with the commonly used normal distribution. First, although the normal often works well when the average vote outcome is of interest, it fails badly for other aspects of the distribution, such as surprises, the probability of a close election or of one party winning, or for uncertainty estimates. Second, the normal implies that big surprises that we see regularly in politics should almost never occur, meaning that it also gets the concentration around the mean wrong. To fix these problems, we use the additive logistic Student t (ALT) distribution, which (a) constrains the vote proportion to the [0,1] interval, (b) has appropriately fatter tails to represent surprises, and also (c) has more of the remaining density concentrated near the mean. The result is that point predictions are more informative than the normal at the same time as it accounts better for surprises at the tails of the distribution.[5]

[5]Roughly, the ALT is the implied distribution on v (and so restricted to the [0,1] interval) when the t

2.2.2 The Model, with Fully Contested Elections

We now combine all the features described above in one model, reusing the notation (and redefining symbols) from Section 2.1. For expository simplicity, we start with all districts contested (see Supplementary Appendix 1 for the complete version). Thus, let

$$v_{it} \sim \text{ALT}(\mu_{it}, \phi_t^2, \nu_t), \tag{2}$$

$$\mu_{it} = X_{it}\beta_t + \gamma_i + \eta_t \tag{3}$$

where the variance is decomposed by the ALT for additional flexibility into scale ϕ and degrees of freedom ν_t parameters (as $\nu_t \to \infty$, the ALT approximates the additive logistic normal). The systematic component for the conditional expected value includes three independent random effect terms for covariate effects, district uniqueness, and national swing, respectively,

$$\beta_{tk} \sim \mathcal{N}(\hat{\beta}_k, \sigma^2_{\beta_k}), \quad \gamma_i \sim \mathcal{N}(0, \sigma^2_\gamma), \quad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta),$$

for $k = 1, \ldots, K$ covariates, $i = 1, \ldots, n$ observations, $t = 1, \ldots, T$ elections, and diffuse priors chosen for estimation convenience (see Supplementary Appendix 4).[6]

As with the standard approach, some covariates one might put in this model vary over i and t (e.g., the lagged vote, $v_{i,t-1}$), some vary only over i (e.g., the confederate states indicator), and some vary only over t (e.g., midterm penalty). A random effect can also be excluded, which can be useful when little information exists such as for covariates of the last type when T is small.

3 Evaluation

We now evaluate both the standard linear-normal approach and our proposed "additive logistic t model with contemporaneous correlations," or LogisTiCC for short. We do this by

distribution is applied to the (unbounded) logistic transformation of the vote $\ln v_{it}/(1 - v_{it})$. For technical details, and extensive evaluations in multiparty elections, see Katz and King (1999).

[6]For intuition, consider the vote on the logistic scale by letting $y_{it} \equiv \ln[v_{it}/(1 - v_{it})] = \mu_{it} + \omega_{it} = X_{it}\beta_t + \gamma_i + \eta_t + \omega_{it}$, with error term $\omega_{it} \equiv \ln[v_{it}/(1-v_{it})] - \mu_{it}$, which Equation 2 implies is t distributed. This enables us to see, first, that the national swing term η_t induces a positive covariance between any two districts i and i' ($i \neq i'$) for each election year t: $\text{Cov}(y_{it}, y_{i't}|X_{it}, X_{i't}, \beta_t) = \sigma^2_\eta > 0$. And second, the random district uniqueness term γ_i induces a positive covariance for election outcomes in any one district i at two times t and t' (within the same redistricting decade), over and above differences due to X: $\text{Cov}(y_{it}, y_{it'}|X_{it}, X_{it'}, \beta_t, \beta_{t'}) = \sigma^2_\gamma > 0$.

summarizing the models' statistical properties (Section 3.1), comparing the probabilities of rare events from each approach to actual elections (Section 3.2), evaluating the calibration of model estimates (Section 3.3), and studying each model's confidence interval coverage (Section 3.4).

For our empirical analyses, we analyze 28 US Congressional elections from 1954 to 2020, including a total of 14,710 district-level contests, with forecasts limited to the 10,778 contests that exclude the first year of each redistricting decade. This large dataset enables us to conduct numerous rigorous evaluations (cf. Grimmer, Knox, and Westwood, 2022), all of which we do out of sample (so that no data from the election being predicted is used during calibration or estimation). For each analysis, we use a one-step-ahead forecast whenever possible and a leave-one-out forecast otherwise (we also run both whenever possible, but have never found a material difference)

3.1 Statistical Properties

As political scientists have long understood, the linear-normal model can reveal important information about elections when its specification is close to correct. The standard modeling approach is not formally a limiting special case of the LogisTiCC, although it can be thought of as an approximation in some situations. For one, as with all potentially misspecified models, point estimates of the mean from the linear-normal model will choose the distribution closest to the true data generation process (in the sense of the Kullback-Leibler information criterion; see White 1996) even if the data comes from the LogisTiCC. In addition, if the linear specification is correct, both the normal and the LogisTiCC models will produce similar (and approximately consistent) estimates of (the same) β coefficients.

Nevertheless, given the covariance structure of the proposed model, point estimates from the normal will be inefficient relative to the LogisTiCC and standard errors of β and other quantities will be incorrect. More importantly, quantities of interest other than β and simple means, such as the probability of a candidate winning an election, can be badly biased under the normal even when consistent under the LogisTiCC.

A key problem with the normal model is its incorrect independence assumptions, lead-

ing to substantial false precision in its uncertainty estimates (confidence intervals and standard errors that are too small). In contrast, the LogisTiCC allows for dependence among elections held in the same district at different times and among elections held in different districts on the same day.

3.2 Rare Event Probabilities

We begin by studying the largest political surprises, involving the most difficult parts of the data to model appropriately. We do this by comparing the estimated probability of extraordinarily rare events under each model. As a metric for this particularly hard inference, we use the notion of *moral certitude* from the Enlightenment, which is the idea that events with probabilities smaller than 1 in 10,000 should be disregarded. (Because demographers of the time observed that the probability of a healthy person dying within a day was smaller than 1 in 10,000, and does not seem to affect anyone else's decisions, people act as if they are "morally certain" that any events a probability smaller than 0.0001 will never occur and thus can be treated as 0; see Kavanagh 1990; Buffon 1777.) Updating this (quaint but useful) idea, we make predictions under each model for all elections in our dataset and count the number of elections for which the vote proportion observed out-of-sample appears outside a 99.99% (i.e., $1 - 1/10,000$) forecast confidence interval (for simplicity, we use "credible intervals" and "confidence intervals" interchangeably throughout). If this interval is correct, we should observe only about 1 in 10,000 elections falling outside the interval.

Figure 1 gives a count of these extraordinarily rare events (on the vertical axis) by election year (on the horizontal axis) and for the normal model (in gold) and the LogisTiCC (in black). As can be seen, the data dramatically violate the normal model's predictions in a disturbingly large number of elections. In the entire dataset of 10,778 elections, we would expect to see only about *one* 1-in-10,000 event, but this claim is wrong by a factor of more than fifty, in that surprise events the model is morally certain will not occur actually happened in 58 elections (and as many as 10 of the 435 elections in a single year) (see also Gelman, Carlin, et al., 1995, Ch. 8). The figure also annotates some points with the exact probability that we would expect to see these results under the model. These

forecasts are embarrassingly bad. The late Richard McKelvey was fond of arguing that a fix for over-claiming in empirical work would be to require anyone reporting a p-value to take a bet with the implied odds (i.e., the reciprocal of the p-value) against someone finding evidence to the contrary. Using this logic, a one-dollar bet against the linear-normal model's claimed level of certainty would give an equal chance of winning quadrillions of times more money than exists in circulation in all the world's currencies.

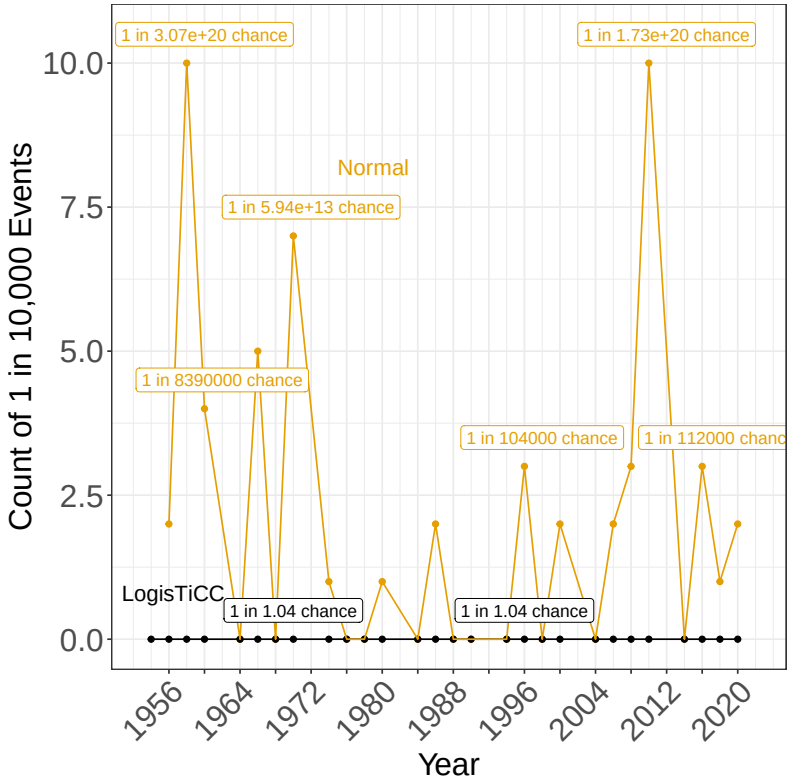


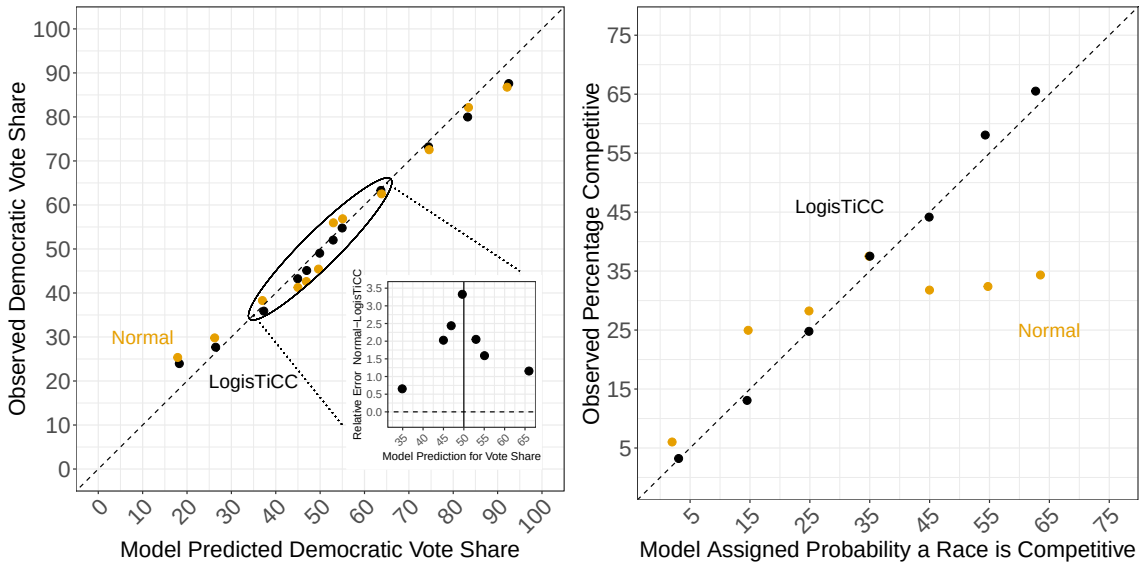
Figure 1: Moral Certitude: Count of elections outside a 99.99 percent credible interval for each election year (with selected points labeled with the probability each model gives for seeing this many 1-in-10,000 events). Separate calculations appear for the normal model (in gold) and our proposed LogisTiCC model (in black).

In stark contrast, the black line in Figure 1 shows that none of the 10,778 out-of-sample observed election results is a surprise to the proposed LogisTiCC model. All 27 election years have zero events. Thus, for this measure of extraordinarily unlikely events, the out-of-sample performance of our proposed model vastly exceeds that of the standard approach.

3.3 Calibration

We now show that this result is general, in that estimates from our model are well *calibrated*. By "calibrated" we mean that vote proportion predictions are approximately unbiased and, for probabilities, when the model predicts that a certain event will occur with, e.g., a 30% probability that event will actually occur in about 3 of every 10 elections. In the process, we also show that the normal model does well for averages but fails for other quantities and that the LogisTiCC does well for all quantities.

We begin with the best case for the normal model, which is for the Democratic vote share, a simple average. To do this, imagine scatterplots for each model of predicted vote shares (horizontally) by observed vote shares (vertically). Then, to evaluate unbiasedness nonparametrically (while simultaneously avoiding graphical clutter with 10,778 data points), we sort predictions for each model into bins, such as $[0, 0.1]$, $(0.1, 0.2]$, $(0.2, 0.3]$, with finer bins in the middle of the scale where vote predictions are more common. Within each bin, we plot the average out-of-sample vote prediction (horizontally) by the average of the observed values for the same elections (vertically). The resulting plot appears in Figure 2(a) for the normal (in gold) and the LogisTiCC (in black).



(a) District vote

(b) Competitive race probability

Figure 2: Calibration: Predicted out-of-sample probabilities (horizontally) by observed frequencies (vertically).

Figure 2(a) reveals excellent out-of-sample predictions for both the normal and LogisTiCC models, as can be seen by both gold and black dots appearing near the 45-degree line, although the differences do not appear large. For all but one dot, LogisTiCC predictions are closer to the 45-degree line, indicating lower error rates. Somewhat more seriously, the normal, but not the LogisTiCC, makes systematically biased errors in the region we highlighted near the majority threshold, where the errors in the gold dots switch from being too small for predictions of Republican victories (i.e., below 0.5) to too large for predictions of Democratic victories. Because the 50% threshold is so important in elections, we add an inset graph that plots (vertically) the number of percentage points by which normal absolute prediction errors exceed the LogisTiCC prediction errors. As can be seen, the absolute errors of the normal exceed the LogisTiCC by 3.3 percentage points right at the winning threshold of 50%, and declining for predictions farther away. For these highly competitive elections, this politically meaningful error rate can easily lead to predicting the wrong winner.

The normal model thus does reasonably well for these averages, although the LogisTiCC has lower error rates and, unlike the normal, no evidence of systematically biased predictions. We now move from this best case for the normal to quantities other than the mean, which depend on aspects of the distribution we need to get right to accurately evaluate electoral accountability. Thus, Figure 2(b) plots a calibration graph for the (out-of-sample) probability of a competitive outcome (which we define as $v_{it} \in [0.45, 0.55]$). To create this graph, we sort all district-level probabilities for each model into bins, $[0, 0.1]$, $(0.1, 0.2]$, $(0.2, 0.3]$,..., and plot a dot (horizontally) based on the average estimated probability in a bin and the proportion of elections in the bin observed to be competitive (vertically). Dots for a perfectly calibrated model should again fall approximately on the 45-degree line.

In Figure 2(b), the dots computed from the LogisTiCC model (in black) are all quite close to the 45-degree line, with no systematic error pattern, and hence well calibrated. In contrast, those from the normal (in gold) are all farther from the line than the LogisTiCC and substantially deviate from the 45-degree line of equality as the predicted probability

of a competitive election gets higher. More seriously, the normal model fails most dramatically in elections predicted to be the most politically important, the competitive ones. To be specific, elections that the normal model gives a 45% or 55% or even 65% probability of being competitive are actually competitive less than 35% of the time. Clearly, the normal does not work for predicting the probability of a competitive election and so would be misleading for evaluating electoral accountability.

3.4 Coverage

We now study, in three ways, the properties of credible intervals computed from the standard and proposed models.

First, for each model, we compute a 95% out-of-sample credible interval around every individual district's vote share and tally up the percentage of districts that interval captures. Our results appear in Figure 3, with time on the horizontal axis and the percent coverage on the vertical axis (again with normal in gold and LogisTiCC in black). A properly calibrated model should capture 95% of districts which, aside from estimation error, should be at the flat black line near the top of the figure. This is approximately the case for the LogisTiCC, which has well-calibrated intervals; on average 95.1% of elections are captured in the 95% confidence intervals. In contrast, the normal interval substantially deviates from capturing 95% of the elections in all but a few years. Worse, in almost every case, the normal is overconfident, meaning that uncertainty estimates computed from this commonly used model should not be trusted.

At the bottom right of the figure, we also added the average confidence interval length for two regions of the graph. For competitive elections ($v_{it} \in [45, 55]\%$), the LogisTiCC interval is somewhat larger than the normal, whereas for less competitive elections $v_{it} > 80\%$ the LogisTiCC interval is smaller. For both, the 50% LogisTiCC interval (in darker black) shows high concentration, and the 95% interval (the entire length of the black line) captures long tails.

Second, we evaluate our distributional assumption (a compound error term with random effects and an additive logistic t distribution). To do this, we use methods of "conformal inference" that offer guarantees of accurate distribution-free finite sample coverage

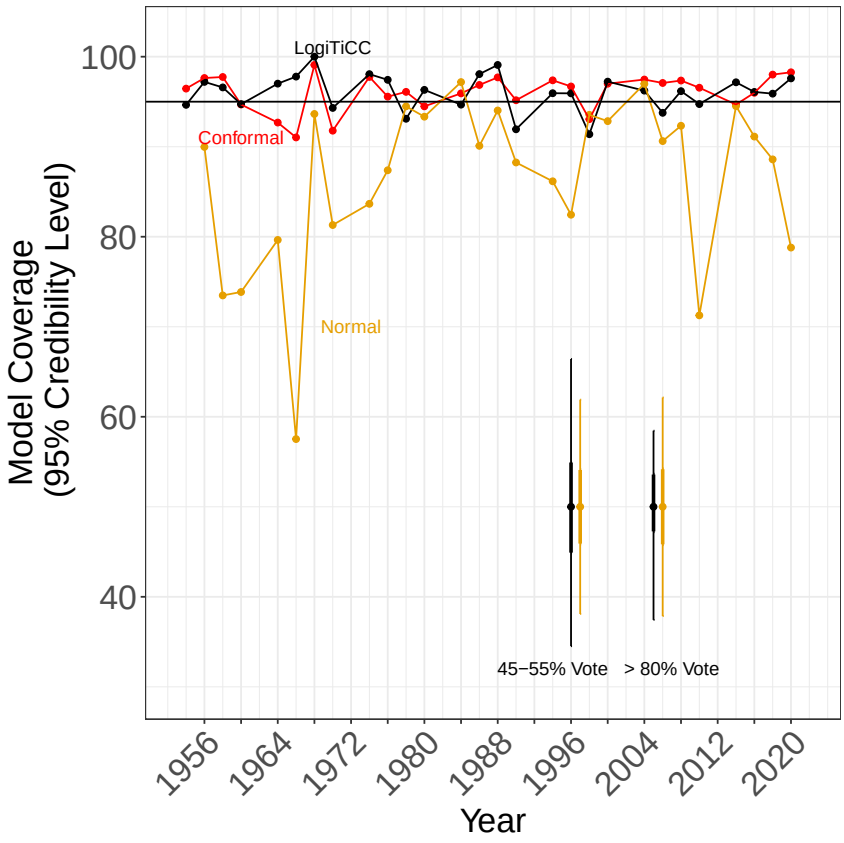


Figure 3: Confidence Interval Coverage the 95 Percent Level, with average 95% and 50% confidence interval length (at the bottom right)

even under model misspecification, for any predictive model, and so we use it to check for misspecification in our model (Vovk, Gammerman, and Shafer, 2005). (Intuitively, the method works by computing confidence intervals based on errors from previous years' forecasts, assuming primarily that the data generation process is exchangeable conditional on the covariates.) In Figure 3 we add conformal confidence intervals (in red). We first confirm that the conformal intervals have accurate coverage, as designed, which we can see as the red line varies around the flat 95% line across the years. More relevant for our purposes is the comparison between the fit of the red and black lines to the 95% line. This comparison indicates that the LogisTiCC has approximately the same high-quality coverage as these distribution-free intervals. These results thus support the veracity of our LogisTiCC distributional assumptions and as a valid generative model of US congressional elections, with an advantage over the conformal approach of being able to compute estimates of any quantity of interest.

Finally, we plot in Figure 4 a time series of the Democratic proportion of the vote of the median seat in the House of Representatives (see the red stars). This is one of the most consequential quantities in US politics because whichever party controls the median seat controls the House. For each year and model, we omit a year from the dataset and compute a point forecast and 95% out-of-sample credible interval around it. These results appear in gold for the normal and black for the LogisTiCC.

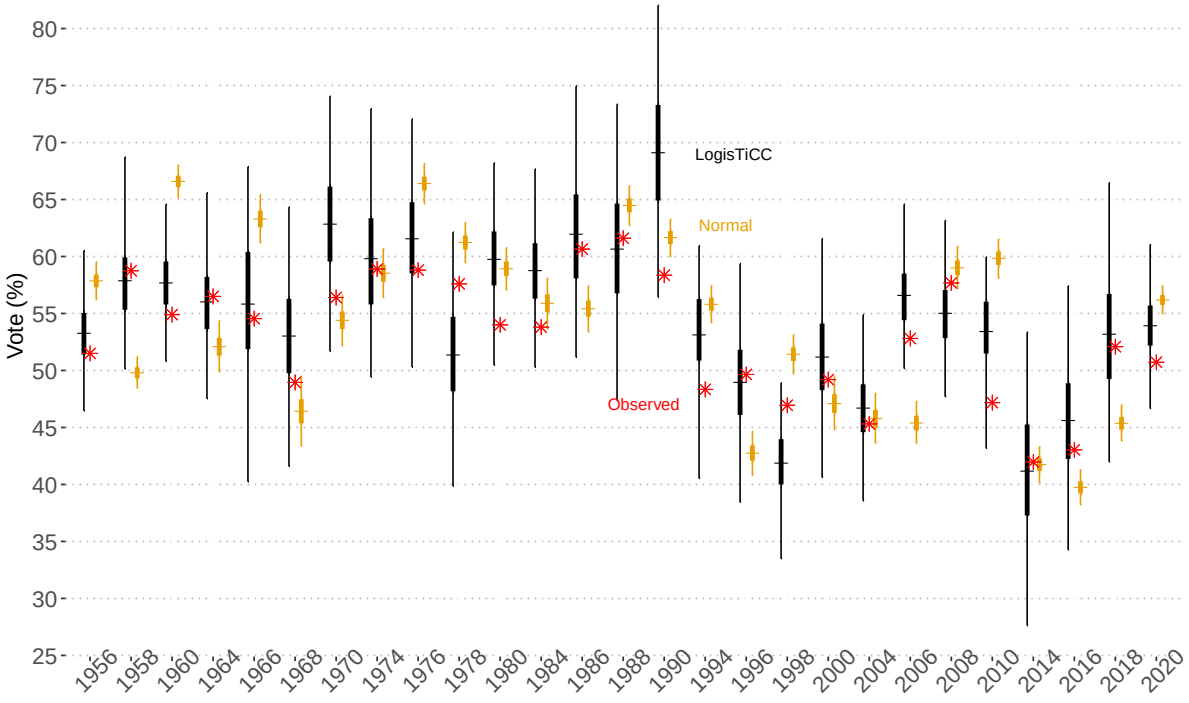


Figure 4: Expected Vote Share of the Median House Seat (95 Percent Credible Interval)

The absence of random effects in the normal model reflects its unrealistic independence assumptions which, for computing the vote of the median seat, makes its nominal confidence intervals drastically overconfident (see how much smaller are the gold than black intervals). The effect can also be seen because, in these out-of-sample tests, we would expect a well-calibrated model to miss only about 1.3 elections of 27, but the normal intervals fail to capture the election result for 20 of 27 elections. In contrast, LogisTiCC's predictive confidence interval does not miss any.

4 Empirical Results

The advantage of an accurate generative model is the big-picture understanding it can provide. It can be used to compute valid estimates of diverse quantities of interest and also help us understand large- and small-scale patterns over thousands of districts and elections. In Section 4.1, we provide our formal definition of the probability of incumbent defeat, along with supporting behavioral assumptions and empirical evidence. In Section 4.2, we show that the probability of incumbent defeat has been high and approximately constant for decades. Then, in Section 4.3, we estimate a variety of other commonly used quantities and show how these are useful for characterizing the changing state of American legislative politics but apparently contradict the constant probability of defeat result. And finally, in Section 4.4, we reconcile the stability we see in Section 4.2 with the change and diversity from Section 4.3 by looking beyond means to the full distribution of the vote.

4.1 Definitions and Behavioral Assumptions

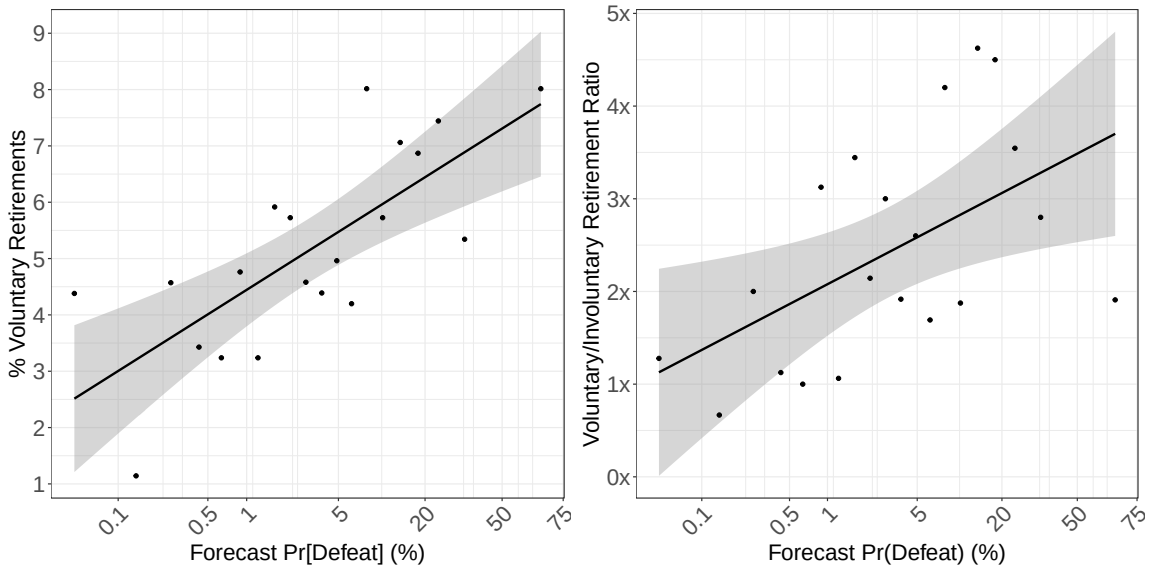
We define the *probability of incumbent defeat*, for candidates who win a seat in the House, as the probability of that same person being defeated for office if he or she runs in the next election held two years later. Losing in either the primary or general election counts as a defeat. This definition requires counterfactual estimation if the incumbent winds up missing the electoral process because of death, jail, or voluntarily choosing not to run for reelection.[7]

Although we estimate the probability of electoral defeat at the end of the primary season, it is useful as a measure of electoral accountability if incumbents are aware of it earlier in their term and act accordingly. We can show this in two ways. First, incumbents respond to higher probabilities of defeat by raising more campaign funds (Erikson and Palfrey, 2000). And second, we now show that incumbents are indeed aware of the proba-

[7] We find that the easier-to-estimate non-counterfactual probability of a candidate who wins one election not holding office after the subsequent election is about two percentage points higher than the estimates we give below for our quantity of interest. Although this is a useful reality check on our partly counterfactual estimates, involuntary retirements makes this quantity not immediately relevant for studying electoral accountability.

bility of defeat early enough to act on it with an increased probability of early retirement.

Figure 5 plots our estimate of the probability of incumbent defeat (horizontally) by two measures of voluntary retirements in the two panels (vertically). The horizontal axis is plotted on the logit scale for graphical clarity, with election probabilities binned, with about 250 elections per bin.



(a) Voluntary Retirements

(b) Voluntary/Involuntary Retirement Ratio

Figure 5: Probability of Defeat Predicts Voluntary Retirements. The horizontal axis is plotted on the logit scale for graphical clarity.

In Panel (a), the vertical axis is the percentage of voluntary retirements in each bin (defined by a range in the probability of incumbent defeat). The results are unambiguous: As the probability of defeat increases over its observed range on the horizontal axis, voluntary retirements more than triple. Incumbents thus seem aware of the information in our probability of defeat estimates, and it changes their behavior. To be clear, this conclusion relies on the assumption that involuntary retirements (death, serious illness, jail), not represented directly here, occur independently of voluntary retirements. Although this is a plausible assumption, we provide in Panel (b) another way of looking at the data that explicitly accounts for both types of retirements. It does this by plotting on the vertical axis the ratio of voluntary to involuntary retirements. Again, the relationship is strong: At the lowest end of the range of the probability of incumbent defeat, incumbents retire voluntarily about as much as they do involuntarily but, as the probability of defeat in-

creases, the ratio of voluntary to involuntary retirements increases dramatically, and at the high end about 3.5 times as many incumbents retire voluntarily as they do involuntarily. Again, it seems that incumbents are well aware of the information in our probability of defeat estimates and appear to use it to make retirement decisions.[8]

Finally, we use our conclusions from Figure 5 to clarify our behavioral assumptions, that is, the elements of our model (in Section 2.2.2) that can reasonably be treated as if they were known by House incumbents at different points in time. Thus, we assume, over a two-year congressional session, that incumbents know anything researchers know and explicitly code in their models, including the values of the covariates and the statistical specification. In addition, each incumbent i has special knowledge of γ_i, which records their electoral strengths or weaknesses. Although the precise national swing η_t is not known to anyone until shortly before the next election, incumbents learn more and more about it over the congressional session, and they know its variance, σ_η^2, from the outset.

4.2 The Probability of Incumbent Defeat

We now study trends in the probability of incumbent defeat. We generate all estimates broken down by the four categories defined by presidential/midterm election years and in-party/out-party members, which political scientists have long known predict the probability of defeat. We present these estimates in two steps, each with a separate figure.

We begin in Figure 6 by computing the average probability of incumbent defeat (on the vertical axis), for each year (on the horizontal axis), within the four (color-coded) categories. The results reflect the well-known pattern that, in midterm elections, in-party incumbents (in gold) are at most risk of defeat and out-party incumbents (in black) are at the least risk; incumbents during presidential elections from both parties (in red for in-party and green for out-party members) are at medium risk.

Figure 6 also reveals the striking result that the mean (out-of-sample, ex ante) probability of incumbent defeat is high and approximately constant over the entire 64-year

[8]Technically, involuntary retirements pose a statistical problem known as "censoring by death", which prevents us from knowing whether incumbents exiting involuntarily would have retired voluntarily if they had been in office. More formal methods of analysis here would require accurate involuntary retirement predictions, which is unlikely almost by definition.

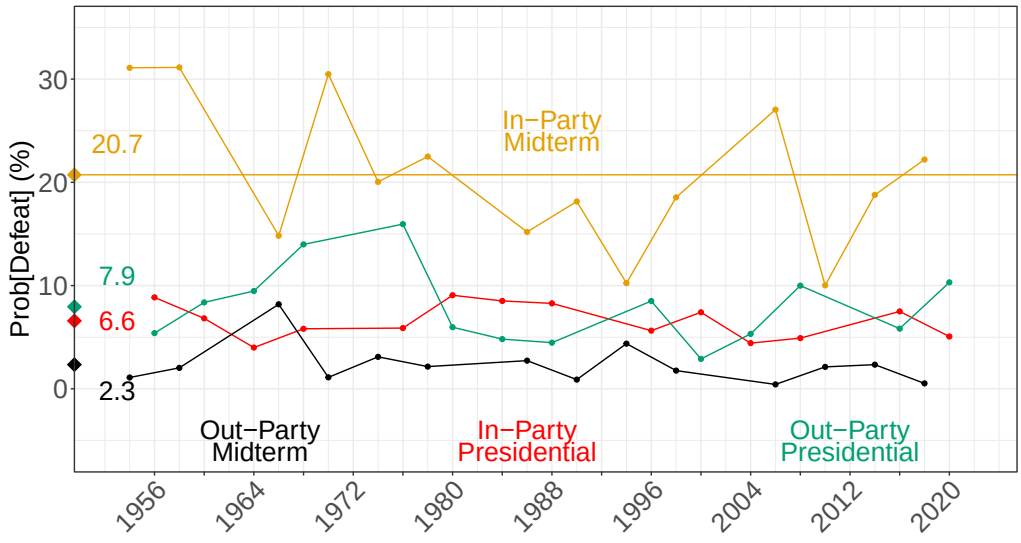


Figure 6: Mean Probability of Incumbent Defeat

period. Statistical tests (not shown) confirm what is apparent from the figure, that each of the four time series do not trend at all. Since each time series is approximately constant, we summarize them all by their averages, which appear on the left side of the graph. We see that the probability of incumbent defeat for in-party members running in midterm elections is a remarkable 20.7%, meaning that the odds of losing their seat is higher than 1 in 5. The probability of defeat for out-party members during presidential elections is 7.9% and in-party members is 6.6%. Even out-party members during the midterm have about a 2.3% probability of defeat. Overall, averaging over members and election years, the average probability of an incumbent being defeated in any election is a substantial 10.9%. For those who are or hope to be tenured professors, think of how much more we might pay attention to the chair of our departments, review committees, and students if every four or eight years we knew that one-fifth of everyone on our hallway would be summarily fired. Our laurels wouldn't be very restful. Although these results indicate that being a member of the House of Representatives is not a comfortable, secure position, they are good news for the incentives American democracy provides to its elected legislators to be responsive to their constituents.

Unfortunately for House members, but fortunately for American democracy, elections may provide even more random terror and potentially more electoral accountability than

Figure 6 indicates. To see this, Figure 7 zooms in to in party members during the midterm. The average in party midterm line from Figure 6 (in gold) also appears in this figure (in dark black), to which we have added one thin line for each incumbent traced out over every midterm election in which he or she competed (with different colors to ensure the end of one person's career is distinct from the start of a new person).

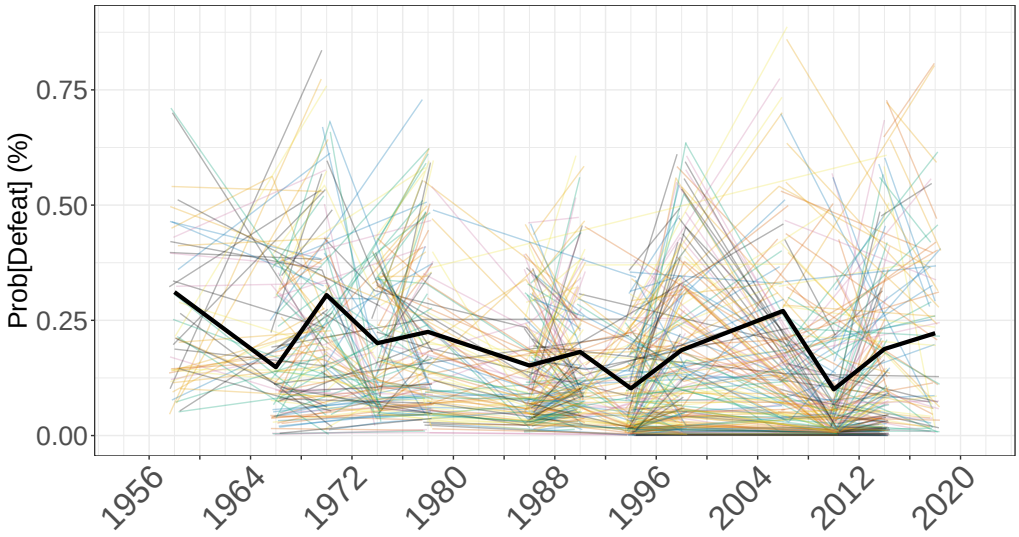


Figure 7: Individual-Level Probability of Incumbent Defeat

This graph shows that an incumbent cannot even count on the high average probability of defeat from Figure 6 to guide their worries about losing their job, because they may well wind up by a luck of the draw far from the mean with even higher (or lower) probability. The average probability of incumbent defeat has been high and constant over many years, but the actual probability that any one incumbent will be defeated in the next election is effectively a random draw from a high variance distribution with the same mean. Indeed, the interquartile range of individual incumbent probabilities is a remarkable 63.6 percentage points. Furthermore, not only is the variance around the average high, but the variance for any one incumbent is also high and apparently unpredictable over time, which can be seen by the jumble of "pick up sticks" in the graph jumping up and down. (The corresponding figure for out-party midterm and in- and out-party presidential elections, which we do not present here, look like this graph, centered on their average lines from Figure 6.) The results from both analyses demonstrate in different ways that the proba-

bility that an incumbent will need to find another job is usually high, highly variable, and highly unpredictable (cf. Little and Meng, 2024). See Supplementary Appendix 6 for an alternative visualization.

Finally, we find strong evidence for stable probabilities of defeat even during the realignment of the South. Figure 8 illustrates that incumbents in the North and South faced similar probabilities of defeat following the passage of the Civil Rights in 1964. Even before that, probabilities of incumbent defeat differed only by 3 percentage points on average through the mid-1950s, 7 percent in the South and 10 percent in the remaining states. The similarity between the South and the rest of the country is even more striking when we disaggregate by region in the same way as Figure 6, which we do in Supplementary Appendix 7, which shows that House seats belonging to both in-party and out-party incumbents in the South and the rest of the country both exhibit stable mean probabilities of defeat during midterm and presidential elections over the postwar period.

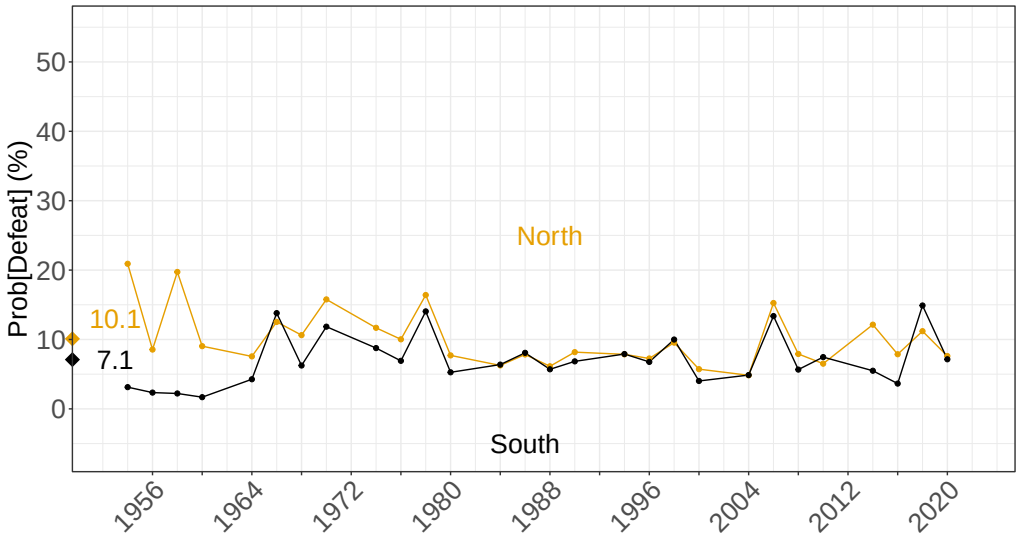


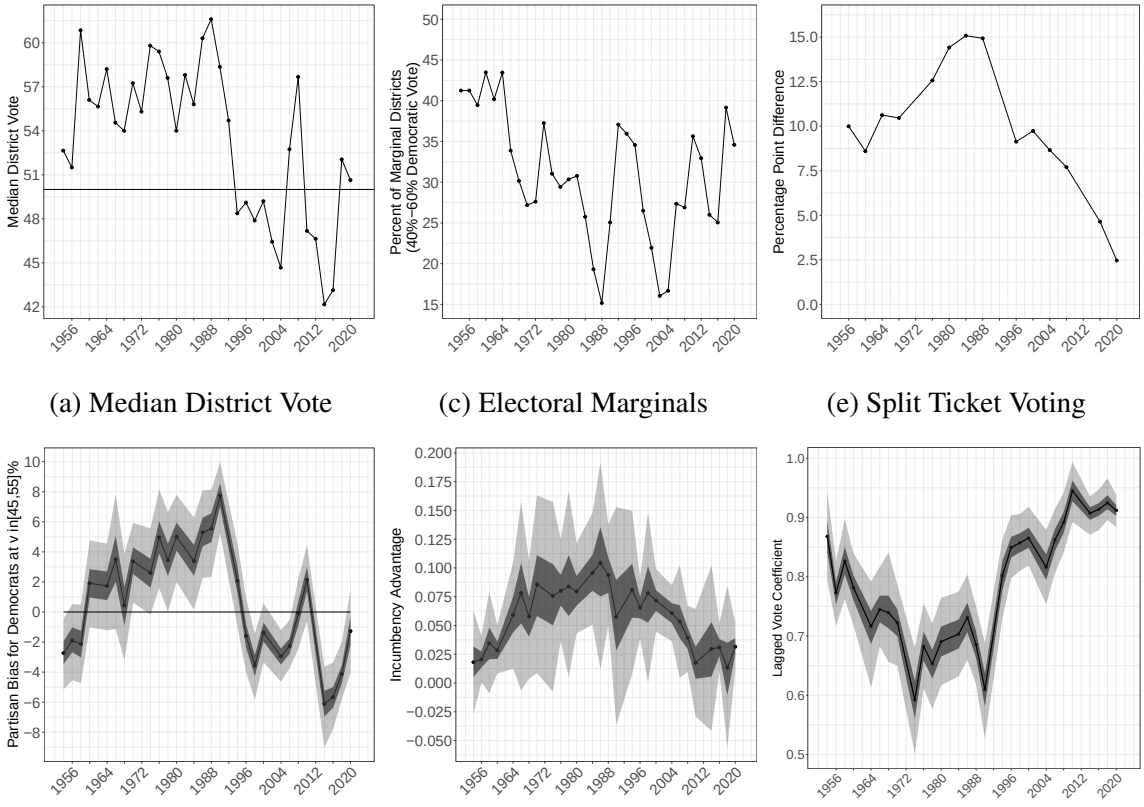
Figure 8: Probability of Incumbent Defeat by Region

4.3 The Changing State of American Politics

We now summarize some major changes in American politics over the last two-thirds of a century and show how each seems to contradict the finding from Section 4.2 that the probability of incumbent defeat is high and unchanging. We highlight the puzzles here

and save their resolution for the next section.

Figure 9 gives six time series (with the election year on the horizontal axis and a measure of an aspect of American politics on the vertical axis). The first row (Panels a, c, and e) are simple summaries of raw data, whereas the statistics in the second row (Panels b, d, and f) are computed from our model, although because each is a different type of average similar results would have come from normal model estimates too. Figures are organized into columns based on substantive relationships.



(a) Median District Vote (c) Electoral Marginals (e) Split Ticket Voting

(b) Partisan Bias (d) Incumbency Advantage (f) Lagged Vote

Figure 9: Changes in American Politics, apparently contradicting the constant probability of incumbent defeat. Panels in the first row are computed from raw data, whereas those in the second row are computed from our model (each with 50% and 95% confidence intervals).

For orientation, Panel (a) gives the familiar median House district vote, revealing the well-known Democratic advantage from the 1950s until the 1994 "Contract With America" election, after which party strength has been closer to even but more variable. This time series is puzzling because we would expect the probability of defeat to drop during

periods of one party's dominance and rise during competitive periods, but we know from section 4.2 that this is not the case.

Figure 9(b) gives the trend in partisan bias, which is the deviation from partisan symmetry (where the electoral system fairly translates votes to seats in the same way for each party; Katz, King, and Rosenblatt 2020). Partisan bias can result from gerrymandering or population changes with fixed district lines. The trend indicates that the electoral system was approximately fair to the two parties in the 1950s, before gerrymandering was common; the system then became increasingly biased towards the Democrats following Baker v. Carr (1962), possibly because gerrymandering became common when the Democrats happened to control most state legislatures. This trend continued until the early 1990s, topping out at a remarkable 8% extra House seats unfairly allocated to the Democrats, at which point a Republican effort to contest and win more state legislatures had an effect. The result was a wave of gerrymandering in the mid-1990s that helped Republicans by reducing Democratic bias; this brought the electoral system closer to partisan symmetry, but with considerable variation over time. The puzzle here is that we would have expected incumbents to be at more risk of defeat when one party consistently receives many seats unfairly, but we know from Section 4.2 that this is not the case for either party.

Panel (c) gives the percent of marginal districts ($v_{it} \in [40, 60]\%$), revealing that the marginals "vanished" (i.e., declined) from the 1950s until the late 1980s, at which point we only see high volatility without a clear trend. This pattern is puzzling because the literature has long assumed that incumbent reelection chances are lower in marginal districts, but we know from Section 4.2 that this is untrue, at least in the aggregate.

Incumbency advantage in Panel (d) estimates the expected vote increase for a party that nominates the incumbent as compared to the best non-incumbent willing to run (Gelman and King, 1990). This measure has varied considerably over the last six and a half decades, beginning at about 2 percentage points in the 1950s, smoothly increasing until about 10 percentage points in the late 1980s, and then smoothly declining to the present 2 points. This result is puzzling because incumbency advantage has been interpreted as (and calculated for the purpose of providing) an indirect measure of the probability of

incumbent defeat; yet, as we know from Section 4.2, the probability of incumbent defeat has been approximately constant, directly contradicting the pattern in this graph.

The graph in Panel (e) measures the prevalence of districts that split their votes for a party between the president and the house, as measured by the average absolute difference between the two percentages. Split ticket voting has varied considerably, from low points in the 1950s at 10% and currently at 2.5%, and as high as 15% in the less partisan 1980s. The puzzle here is that most would expect high levels of split ticket voting — with the House vote being less influenced by the vicissitudes of highly visible presidential campaigns and incumbents building independent reelection organizations that do not rely on their party — to reduce the probability of incumbent defeat; as Section 4.2 shows, however, this is not what happened.

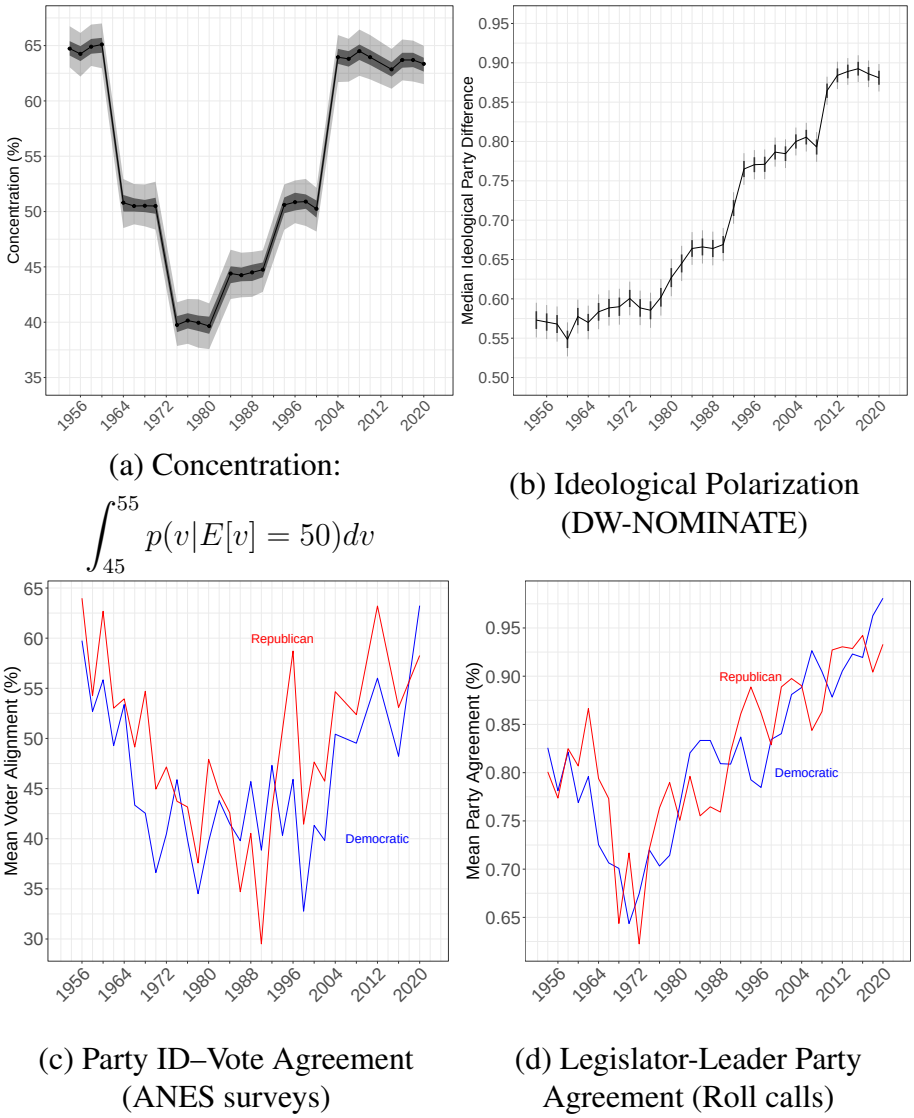
Finally, Figure 9(f) gives a time series of the stability of the vote, measured by the coefficient on the vote from the previous election (conditional on the other covariates, on the logit scale) as it predicts the current vote. The changes in this figure are as dramatic as all the others, with high levels of stability at the start and end of the series, and quite low levels in between. This figure is puzzling because we would expect the probability of incumbent defeat to be lower when vote totals are sticky from one time to the next, but again Section 4.2 shows that this too is incorrect.

4.4 How Changing Politics Leads to an Unchanging Probability of Defeat

Each of the six measures portrayed in Figure 9 documents massive and meaningful change in an important aspect of American politics. Each would also seem to imply that electoral accountability, as measured by the probability of incumbent defeat, varies dramatically over time. However, our single validated generative model — which can reproduce every pattern discussed in Sections 4.2 and 4.3 — clearly demonstrates otherwise. Thus, in three steps, we now resolve this puzzle.

First, all six of the commonly used measures in Figure 9 provide valuable information about the central tendency of their respective concept, but obviously ignore the rest of the vote distribution. As it happens, understanding the variation around the central tendency

completely changes the interpretation of each of these quantities, especially their implications for electoral accountability. To see this, consider the time series plot of the vote *concentration* in Figure 10(a), which we define as the proportion of a vote distribution falling between 45% and 55% in the situation where the expected value is 50%.



(a) Concentration: $\int_{45}^{55} p(v|E[v] = 50)dv$

(b) Ideological Polarization (DW-NOMINATE)

(c) Party ID–Vote Agreement (ANES surveys)

(d) Legislator-Leader Party Agreement (Roll calls)

Figure 10: How the Full Vote Distribution Matters

Figure 10(a) reveals remarkably high variation in vote concentration, from about 65% in the 1950s and 2020s to as low as 40% in between. This variation demonstrates the inadequacy of using the central tendency as a summary of the full distribution: an average can mean fundamentally different things at different times. Only with the full distribution we can resolve the puzzles highlighted in Section 4.3. For a stark example, consider

the tremendous variation in incumbency advantage in Figure 9(e), ranging between 2% and 10%. Although this indicator has been interpreted as almost synonymous with risk to incumbents, it only estimates the expected vote advantage, which turns out to have been high in precisely the election years when the vote concentration was low, and was low when concentration was high. The probability of incumbent defeat is the integral to one side of 50%, which is affected by both the expected value and the variation in the vote around it; empirically, the U shape of the vote concentration graph thus cancels out the upside down U shape of incumbency advantage graph, leaving the probability of incumbent defeat approximately constant.

A similar story of variation canceling out changes in central tendency applies to all the other panels in Figure 9: Complete understanding of the probability of incumbent defeat requires the full distribution of vote predictions, rather than only central tendencies. The approximately U or upside down U shapes of electoral margins, split ticket voting, lagged vote, and to some extent partisan bias are canceled out by voter concentration. Concentration, however, is only one simple summary of the full distribution empirically fit by the generative LogisTiCC model. All aspects of that model, including central tendency, variation, skewness, and other features of the distribution are incorporated into our calculation of the probability of incumbent defeat.

Second, high voter concentration in recent years will come as a surprise to no one, but nearly as high voter concentration in the 1950s may be more jarring. The surprise is that it is easy to confuse ideological and partisan alignment. Indeed, the electoral system was far less ideologically polarized in the 1950s, which the striking, almost linear, increase in average DW-NOMINATE ideology scores in Figure 10(b) confirms. However, the 1950s was nevertheless a time of high levels of partisan alignment, both among the public (see the percent agreement between Party identification and the vote in Panel (c)) and among House members (see the percent agreement between legislators and their party leader in Panel (d)). Parties were highly aligned in the 1950s, but based on social groups and parental socialization, not ideology.[9]

[9]Note the asymmetry in our presentation: For party differences, we give results among voters (c) and legislators (d), but for ideological differences, we only present differences among legislators (b). Why do

And finally, one reason for the strong time series patterns in voter concentration is that the national swing turns out to be far more important than the variation due to district uniqueness (even after accounting for the covariates). To see this, we compute the ratio of the standard deviation of the vote (on the logit scale) due to variations in national swing relative to district uniqueness: $\sigma_\eta/\sigma_\gamma = 0.2/0.036 = 5.6$ (a ratio we find to be largely stable over time). Campaign observers have long known that exogenous events and heresthetical maneuvers by individual congressional candidates in their district campaigns can be important, but this result shows that exogenous national events and national-level heresthetical maneuvers are more than five times as consequential as the sum of all the individual district campaigns. All politics may well be local in its effect, but national level political issues have a far bigger effect both nationally and locally than local issues (thus confirming Hopkins 2018 and Caughey and Warshaw 2022: Sec. 3.3).

5 Concluding Remarks

Although electoral accountability is one of the few widely recognized and uncontroversial components of a healthy democracy, it has rarely been quantified and studied at scale. We measure this concept by the extent to which individual members of the US House of Representatives have a serious chance of being defeated in a campaign for reelection if they choose to run again. This is a minimal, essential component of a healthy democracy because it allows the voters, rather than political scientists, to define what characteristics of the candidates they approve or disapprove of. Surprisingly, our results reveal that this quantity has been approximately constant and at a high level for over two-thirds of a century.

Puzzlingly, however, over this time period we have seen dramatic changes in many important and well-known characteristics of electoral politics, such as the incumbency advantage, split-ticket voting, partisan bias, etc., each of which suggests that electoral

we exclude a graph for ideological differences among voters? The reason is that ideology is an idea invented by philosophers and used by political scientists; it was relatively unknown among voters until recently (Lee, 2009). In fact, questions about ideology were not even asked in the American National Election Survey until 1972 and even then prefaced with an explanation: "You may have recently heard a lot of talk about left/right…". Ideology is a normative idea that academics impose on voters, not one that voters have always chosen to describe themselves.

accountability varies widely. To resolve this puzzle, we go beyond the measures of central tendency commonly used for these quantities and build a fully generative model of House elections, validated with extensive out-of-sample forecasts. This model thus includes the central tendency, variation around this tendency, and all other features of the vote distribution. When combined, the full distribution from our generative model produces the constant and high incumbent defeat probabilities that we observe.

Previously used generative models of district-level election results have enabled political scientists to learn a great deal about American legislative democracy. But the observable implications of some aspects of these models often fail spectacularly in ways that these models indicate should almost never happen. We build on these models and add statistical and computational features known to be substantively important but not previously available. Our generative model is generic in that it can be used, with the appropriate assumptions and covariates when necessary, to estimate almost any quantity of interest in the literature, and many others, all with calibrated probabilities and honest uncertainty intervals, but in the context of a full generative model so that we are not mislead.

Further growth in computational power may one day enable feasible estimation of joint generative models with an even richer substantive portrait of the electoral system, such as modeling at the precinct-level to include redistricting periods, or encompassing other elections such as for the US Senate, president, and state legislatures. With a continual focus on rigorous out-of-sample validation, and larger more accurate generative models, it may even eventually be possible to estimate these simultaneously with data from other sectors of society such as the economy, demography, public policy, and public health, or potentially even data from other countries.

Finally, although scholars seem to agree that electoral accountability is an essential component of a healthy democracy, it would be valuable to add precision to this idea by studying it in context with other desirable characteristics of democracies. Clearly, the probability of defeat must be high enough to motivate legislators to be responsive to their constituents, but we may not want it to be so high that it disincentivizes high-quality candidates from running for office in the first place, reduces the average policy

expertise of incumbents, or causes them to give up any hope of reelection and thus ignore their constituents altogether. Working through these trade-offs via some combination of empirical experimentation, normative scholarship, or formal modeling would be another valuable direction for future research.

References

Ansolabehere, Stephen and Philip Edward Jones (2010). "Constituents' Responses to Congressional Roll-Call Voting". In: *American Journal of Political Science* 54.3, pp. 583–597. DOI: https://doi.org/10.1111/j.1540-5907.2010.00448.x.

Ansolabehere, Stephen and Shiro Kuriwaki (2022). "Congressional Representation: Accountability from the Constituent's Perspective". In: *American Journal of Political Science* 66.1, pp. 123–139. DOI: https://doi.org/10.1111/ajps.12607.

APSA (1950). *Toward a More Responsible Two-Party System. A Report of the Committee on Political Parties of the American Political Science Association*.

Buffon, George Louis Leclerc de (1777). "Essai d'arithmétique morale". In: *Euvres philosophiques*.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan (2002). "Out of Step, Out of Office: Electoral Accountability and House Members' Voting". In: *American Political Science Review* 96.1, pp. 127–140. DOI: 10.1017/S0003055402004276.

Caughey, Devin and Christopher Warshaw (2022). *Dynamic Democracy: Public Opinion, Elections, and Policymaking in the American States*. University of Chicago Press.

Cox, Gary W. and Jonathan N. Katz (2002). *Elbridge Gerry's salamander: The electoral consequences of the reapportionment revolution*. Cambridge University Press.

Erikson, Robert S (1972). "Malapportionment, gerrymandering, and party fortunes in congressional elections". In: *American Political Science Review* 66.4, pp. 1234–1245.

— (1971). "The Advantage of Incumbency in Congressional Elections". In: *Polity* 3, pp. 395–405.

Erikson, Robert S. and Thomas R. Palfrey (2000). "Equilibria in Campaign Spending Games: Theory and Data". In: *American Political Science Review* 94.3, pp. 595–609.

Eskridge, William N. (1987). "Dynamic Statutory Interpretation". In: *University of Pennsylvania Law Review* 135.6, pp. 1479–1555. ISSN: 00419907. URL: http://www.jstor.org/stable/3312014 (visited on 10/30/2023).

Ferejohn, John A. (1977). "On the Decline of Competition in Congressional Elections". In: *American Political Science Review* 71, pp. 166–176.

Fiorina, Morris P. (1977). "The Case of the Vanishing Marginals: The Bureaucracy Did It". In: *The American Political Science Review* 71.1, pp. 177–181. URL: http://www.jstor.org/stable/1956961 (visited on 10/29/2023).

Fouirnaies, Alexander and Andrew B. Hall (2022). "How Do Electoral Incentives Affect Legislator Behavior? Evidence from U.S. State Legislatures". In: *American Political Science Review* 116.2, pp. 662–676. DOI: 10.1017/S0003055421001064.

Fraga, Bernard L and Eitan D Hersh (2018). "Are Americans stuck in uncompetitive enclaves? An appraisal of US electoral competition". In: *Quarterly Journal of Political Science* 13.3, pp. 291–311.

Gelman, Andrew, J.B. Carlin, H.S. Stern, and D.B. Rubin (1995). *Bayesian Data Analysis*. Chapman and Hall.

Gelman, Andrew and Gary King (Nov. 1990). "Estimating Incumbency Advantage Without Bias". In: *American Journal of Political Science* 34.4, pp. 1142–1164. URL: https://tinyurl.com/yymdaj5r.

— (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.

Girosi, Federico and Gary King (2008). *Demographic Forecasting*. Princeton: Princeton University Press. URL: j.mp/dsmooth.

Grimmer, Justin, Dean Knox, and Sean Westwood (2022). "Assessing the Reliability of Probabilistic US Presidential Election Forecasts May Take Decades". In.

Grofman, Bernard and Thomas L Brunell (2005). "The art of the dummymander: The impact of recent redistrictings on the partisan makeup of southern House seats". In: *Redistricting in the new millennium*, pp. 183–99.

Hirano, Shigeo and James M. Snyder Jr. (2012). "What Happens to Incumbents in Scandals?" In: *Quarterly Journal of Political Science* 7.4, pp. 447–456. ISSN: 1554-0626. DOI: 10.1561/100.00012039.

Hopkins, Daniel J (2018). *The increasingly United States: How and why American political behavior nationalized*. University of Chicago Press.

Iaryczower, Matias, Gabriel Lopez-Moctezuma, and Adam Meirowitz (2022). "Career Concerns and the Dynamics of Electoral Accountability". In: *American Journal of Political Science* n/a.n/a. DOI: https://doi.org/10.1111/ajps.12740.

Jacobson, Gary C. (1987). "The Marginals Never Vanished: Incumbency and Competition in Elections to the U.S. House of Representatives, 1952-82". In: *American Journal of Political Science* 31.1, pp. 126–141. ISSN: 00925853, 15405907. URL: http://www.jstor.org/stable/2111327 (visited on 10/29/2023).

— (2015). "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections". In: *The Journal of Politics* 77.3, pp. 861–873. ISSN: 00223816, 14682508. URL: http://www.jstor.org/stable/10.1086/681670 (visited on 10/29/2023).

Katz, Jonathan N, Gary King, and Elizabeth Rosenblatt (2020). "Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies". In: *American Political Science Review* 114.1, pp. 164–178. URL: GaryKing.org/symmetry.

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.

Kavanagh, Thomas M (1990). "Chance and Probability in the Enlightenment". In: *French Forum*. Vol. 15. 1, pp. 5–24.

Kenny, Christopher T., Cory McCartan, Tyler Simko, Shiro Kuriwaki, and Kosuke Imai (2023). "Widespread partisan gerrymandering mostly cancels nationally, but reduces

electoral competition". In: *Proceedings of the National Academy of Sciences* 120.25, e2217322120. DOI: 10.1073/pnas.2217322120.

Lee, Frances E (2009). *Beyond ideology: Politics, principles, and partisanship in the US Senate*. University of Chicago Press.

Little, Andrew T. and Anne Meng (2024). "Measuring Democratic Backsliding". In: *PS: Political Science and Politics*. DOI: http://dx.doi.org/10.2139/ssrn.4327307.

Mayhew, David R (1974). *Congress: The electoral connection*. Yale university press.

— (2008). "Congressional Elections: The Case of the Vanishing Marginals". In: *How the American Government Works*. New Haven: Yale University Press, pp. 54–72. URL: https://doi.org/10.12987/9780300151763-003.

Nyhan, Brendan, Eric McGhee, John Sides, Seth Masket, and Steven Greene (2012). "One Vote Out of Step? The Effects of Salient Roll Call Votes in the 2010 Election". In: *American Politics Research* 40.5, pp. 844–879. DOI: 10.1177/1532673X11433768.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi (2000). *Democracy and Development*. Cambridge University Press. URL: https://EconPapers.repec.org/RePEc:cup:cbooks:9780521790321.

Riker, William H (1990). "Heresthetic and rhetoric in the spatial model". In: *Advances in the spatial theory of voting* 46, p. 50.

Shepsle, Kenneth A (2003). "Losers in politics (and how they sometimes become winners): William Riker's heresthetic". In: *Perspectives on politics* 1.2, pp. 307–315.

Tausanovitch, Chris and Christopher Warshaw (2018). "Does the ideological proximity between candidates and voters affect voting in US house elections?" In: *Political Behavior* 40, pp. 223–245.

Tufte, Edward R (1973). "The relationship between seats and votes in two-party systems". In: *American Political Science Review* 67.2, pp. 540–554.

Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer (2005). *Algorithmic learning in a random world*. Springer Science & Business Media.

White, Halbert (1996). *Estimation, Inference, and Specification Analysis*. New York: Cambridge University Press.