Methodology

The Big Picture

We use the correlation between the presidential vote on the one hand, and state legislative and congressional votes on the other, to predict how new districts will likely vote and so how biased a plan will be. Our correlations come from the last 10 years of elections, and factor in both any extra advantage incumbents might have as well as how much each state's results might differ from others. We also allow our predictions to be imperfect by quantifying how much our method missed the actual outcomes of past elections, including the degree to which partisan tides have changed party performance from one election to the next. This enables us to generate the most accurate, data-driven, and transparent prediction we can.

The Details

We use a Bayesian hierarchical model of district-level election returns, run for all state legislatures and congressional delegations on the elections from 2012 through 2020. Formally, the model is:

$$y_i \sim \mathcal{N}(\boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{X_i}\boldsymbol{\beta_{s(i)}} + \boldsymbol{X_i}\boldsymbol{\beta_{c(i)}}, \sigma_y^2)$$

$$\begin{pmatrix} \beta_{0s} \\ \beta_{1s} \\ \beta_{2s} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_{0s}} & \rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}} & \rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}} \\ \rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}} & \sigma^2_{\beta_{1s}} & \rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}} \\ \rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}} & \rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}} & \sigma^2_{\beta_{2s}} \end{pmatrix}\right)$$

$$\begin{pmatrix} \beta_{0c} \\ \beta_{1c} \\ \beta_{2c} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_{0c}} & \rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}} & \rho\sigma_{\beta_{0c}}\sigma_{\beta_{2c}} \\ \rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}} & \sigma^2_{\beta_{1c}} & \rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}} \\ \rho\sigma_{\beta_{0c}}\sigma_{\beta_{2c}} & \rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}} & \sigma^2_{\beta_{2c}} \end{pmatrix}\right)$$

where

- i indexes district level elections
- s indexes states, with $s(i)$ denoting the state of district election i
- c indexes election cycles, with $c(i)$ denoting the election cycle of district election i
- y_i is the Democratic share of the two-party vote in district election i
- $\boldsymbol{X_i}$ is a matrix of covariate values for district election i
- $\boldsymbol{\beta}$ is a matrix of population-level intercept and slopes corresponding to covariates $\boldsymbol{X}$
- $\boldsymbol{\beta_{s(i)}}$ and $\boldsymbol{\beta_{c(i)}}$ are matrices of coefficients for the state and election cycle, respectively, of district election i
- σ_y is the residual population-level error term

The model allows the slope for all our covariates—as well as the corresponding intercept—to vary across both states and election cycles. Chambers accounted for minimal variation in an ANOVA test, so state legislative and congressional results were modeled together as emerging from a common distribution. The model includes two covariates: 1) the two-party district-level Democratic presidential vote share, centered around its global mean (0.494); and 2) the incumbency status in district election i, coded -1 for Republican, 0 for open, and 1 for Democratic. We do not have the 2020 presidential vote for estimating new plans in two states—Kentucky and South Dakota—so we used the 2016 presidential vote in the model for those states. In the small number of state-cycle combinations that were missing presidential vote we used the presidential vote for the same district in the next presidential election (or the previous presidential election where the next one was not available).

When generating predictions, PlanScore draws 1000 samples from the posterior distribution of model parameters, and uses them to calculate means and probabilities. We also add in the offsets for the 2020 presidential election cycle, and then also add in samples from the covariance matrix of cycle random effects to allowing the uncertainty of predicting for an unknown election cycle to propagate into our predictions. This has the effect of predicting for an election like 2020 in most respects, but with error bounds that encompass the full range of partisan tides that occurred over the last decade.

Table 1: PlanScore prediction model results

	Estimate	95% Credible Interval
POPULATION-LEVEL		
Intercept (β_0)	0.50	$[0.46, 0.53]$
Presidential vote (β_1)	0.78	$[0.62, 0.93]$
Incumbency (β_2)	0.05	$[0.03, 0.07]$
STATE-LEVEL		
Standard Deviations		
Intercept ($\sigma_{\beta_{0s}}$)	0.02	$[0.02, 0.02]$
Presidential vote ($\sigma_{\beta_{1s}}$)	0.11	$[0.09, 0.14]$
Incumbency ($\sigma_{\beta_{2s}}$)	0.02	$[0.01, 0.02]$
Correlations		
Intercept - Pres. vote ($\rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}}$)	-0.53	$[-0.71, -0.29]$
Intercept - Incumbency ($\rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}}$)	0.29	$[0.00, 0.54]$
Pres. vote - Incumbency ($\rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}}$)	-0.73	$[-0.85, -0.56]$
CYCLE-LEVEL		
Standard Deviations		
Intercept ($\sigma_{\beta_{0c}}$)	0.03	$[0.01, 0.09]$
Presidential vote ($\sigma_{\beta_{1c}}$)	0.15	$[0.07, 0.37]$
Incumbency ($\sigma_{\beta_{2c}}$)	0.02	$[0.01, 0.06]$
Correlations		
Intercept - Pres. vote ($\rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}}$)	-0.13	$[-0.81, 0.66]$
Intercept - Incumbency ($\rho\sigma_{\beta_{0c}}\sigma_{\beta_{2s}}$)	-0.20	$[-0.85, 0.60]$
Pres. vote - Incumbency ($\rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}}$)	-0.57	$[-0.96, 0.39]$

Note: Model estimated in brms for R. Model based on 4 MCMC chains run for 4000 iterations each with a 2000 iteration warm-up. All model parameters converged well with $\hat{R} < 1.0$.