Methodology

The vote share inputs for calculating the metrics come from 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 includes two covariates: 1) the two-party district-level Democratic presidential vote share, averaged across 2012 and 2016 and centered around its global mean; 2) the incumbency status in district election i, coded -1 for Republican, 0 for open, and 1 for Democratic. The model allows the slope for each–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.

When generating predictions, PlanScore assumes an average election year for the 2012-2020 period ($\boldsymbol{\beta_c} = 0$), but otherwise draws from the posterior distribution of model parameters for means and probabilities.

Table 1: PlanScore prediction model results

	Estimate	95% Credible Interval
POPULATION-LEVEL		
Intercept (β_0)	0.50	$[0.47, 0.53]$
Presidential vote (β_1)	0.83	$[0.66, 0.98]$
Incumbency (β_2)	0.05	$[0.02, 0.07]$
STATE-LEVEL		
Standard Deviations		
Intercept ($\sigma_{\beta_{0s}}$)	0.02	$[0.02, 0.03]$
Presidential vote ($\sigma_{\beta_{1s}}$)	0.10	$[0.08, 0.13]$
Incumbency ($\sigma_{\beta_{2s}}$)	0.02	$[0.01, 0.02]$
Correlations		
Intercept - Pres. vote ($\rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}}$)	-0.43	$[-0.65, -0.16]$
Intercept - Incumbency ($\rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}}$)	0.04	$[-0.25, 0.34]$
Pres. vote - Incumbency ($\rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}}$)	-0.69	$[-0.83, -0.49]$
CYCLE-LEVEL		
Standard Deviations		
Intercept ($\sigma_{\beta_{0c}}$)	0.03	$[0.01, 0.08]$
Presidential vote ($\sigma_{\beta_{1c}}$)	0.16	$[0.07, 0.38]$
Incumbency ($\sigma_{\beta_{2c}}$)	0.02	$[0.01, 0.06]$
Correlations		
Intercept - Pres. vote ($\rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}}$)	-0.16	$[-0.80, 0.64]$
Intercept - Incumbency ($\rho\sigma_{\beta_{0c}}\sigma_{\beta_{2s}}$)	-0.16	$[-0.83, 0.64]$
Pres. vote - Incumbency ($\rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}}$)	-0.61	$[-0.97, 0.27]$

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 $\hat{R} < 1.01$.