CALL THE MAP

CTMM 2.2

Technical Methodology

The Call the Map Model for the 2026 House, Senate and governor elections

September 2026

ABSTRACT

CTMM 2.2 is a staged statistical forecast of the 435 House, 35 Senate and 36 governor races of the 2026 United States midterm elections. A robust penalised regression on terrain, history and incumbency gives each race a structural centre; a fitted elasticity to the generic ballot adds the national environment; race polls enter through an evidence-weighted aggregate whose influence vanishes without polls and which carries a cycle-level poll-bias posterior. Expert ratings are excluded from every office and historical training fold. Governor centres receive a calibration fitted on the model's own out-of-sample record ($c' = +0.7648 + 1.1543 \cdot c$); Senate centres receive a fitted smooth-tail and inward-poll-drag correction. Reconstructed House priors use a robust presidential-to-House fit before rebuilding their dependent features. House and governor uncertainty is a heteroskedastic core+tail mixture with a nonnegative scale slope beyond 25 points (tail share 0.0580, tail scale 24.44 points); Senate probabilities and intervals use one Student-t CDF fitted to full-model out-of-fold errors. One joint simulation of 100,000 draws with hierarchical national, regional and state dependence, a common poll-bias shock and a per-draw tail indicator produces every published probability, seat distribution and control figure. Validated leave-one-cycle-out over 2018-2024 at six horizons on the executing code, the model's mean absolute errors are 5.64, 5.87 and 5.71 points and its 80% intervals cover 79%, 81% and 80% of results for the House, Senate and governors. Its limitations, including a disclosed Senate centre bias of +2.7 points, are stated.

CONTENTS

SECTION 1

Overview

CTMM 2.2, the Call the Map Model, forecasts the 435 United States House districts, the 35 Senate seats and the 36 governorships on the ballot on November 3, 2026. For every race it publishes a central margin (Democratic minus Republican, in percentage points), a win probability, 50/80/95% intervals and, where the ballot is not a plain two-party pair, the probability that each named candidate wins. From the same set of simulated outcomes it publishes expected seats, seat distributions, the probability that each party controls each chamber and the joint outcomes across chambers.

The model is staged. A structural prior gives each race a starting margin from its terrain, history and incumbency. The national environment moves every race by a fitted elasticity to the generic ballot. Race polls, where they exist, pull the centre toward an evidence-weighted average. Expert ratings are excluded in every office, both in training and current prediction. Separate centre calibrations fitted on out-of-sample results address measured compression in Senate and governor races. Reconstructed House priors retain their fitted values instead of being truncated at 40 points. A fitted error distribution turns each centre into probabilities and intervals. Finally one joint simulation draws every race together, with national, regional and state shocks, so that seat totals and chamber control reflect the dependence between races.

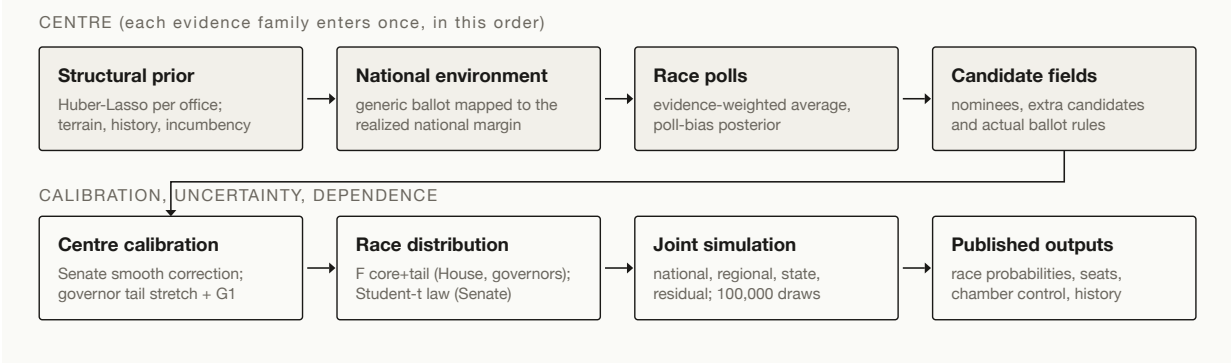

Figure 1: The CTMM 2.2 pipeline. Each evidence family moves the race centre once, in a fixed order; calibration, the race distribution and the joint simulation follow.

Fitted coefficients are estimated within historical training windows, not chosen to produce a desired 2026 race margin. Structural choices, clipping bounds and candidate-transfer ranges are declared assumptions; they are not all fitted parameters. The historical record supporting the fits runs from 2008 to 2024. The validation record covers 2018, 2020, 2022 and 2024 at six forecast origins (60, 35, 21, 14, 7 and 0 days before the election), with the test cycle held out of parameter fitting. Section 13 reports that record; Section 15 states what four cycles cannot settle.

The frozen comparable-race poll-residual overlay from the earlier model is not a CTMM 2.2 input. Its stale donor residuals are not added to this model's own polling update. A September 8 sensitivity comparison found changes below 0.20 points from restoring that overlay, not a solution to safe-seat compression; a newly fitted borrowing model remains research. Historical v1 forecasts and their decompositions remain archived.

Two conventions run through the document. Margins are Democratic minus Republican, so a positive centre favours the Democrat; and a horizon h is the number of days between the forecast origin and the election, 53 days for the board this document accompanies.

SECTION 2

Forecast-origin data

A forecast made on a given day may only use what existed on that day. Race polls require evidence of public availability, not only a field-end date; the generic ballot also uses readings available by the origin. Structural features use the appropriate prior election and geography. Expert ratings are excluded. The same input policy applies to current prediction and historical evaluation.

Structural inputs

Candidate nomination status and input eligibility are distinct. CTMM 2.2 has no blanket nominee-status veto: a poll must meet the actual source, ballot and date rules, but an empty admitted poll set does not mean polls were withheld. Reconstruction retains the geography operative on each date. No expert-rating source supplies a forecast input.

House districts: the district's prior two-party margin on its current lines (uncapped where the result is observed on unchanged lines or officially re-aggregated), the margin two elections back on unchanged lines, the district's 2024 presidential margin relative to the nation, incumbency and its strength, open-seat and defending-party signs, the district's prior over-performance against its state, a prior-uncontested flag, the midterm and White House party indicators, the previous certified national House margin, three American

Community Survey demographics (educational attainment, median household income and median age) and the nine Census divisions. Senate seats: the seat's own prior margin and the state's most recent Senate margin (calendar rule; an appointed seat's prior is the state's last regular election), a prior-uncontested flag, relative presidential lean, incumbency, open-seat sign, defending party, the midterm indicator, the previous national margin and two interactions. Governorships: relative presidential lean, the prior gubernatorial margin by the calendar rule, the midterm indicator, signed incumbency and tenure, and the candidate-experience differential (the highest elected office each nominee has held, 0-3, Democratic minus Republican).

Evidence at the origin

Race polls. The source-admitted poll feed, each reading carrying its pollster, field end date, sample size and sponsor class; a Democratic- or Republican-sponsored poll is flagged partisan. On this board 73 of 435 House races, 28 of 35 Senate races and 30 of 36 governor races carry at least one admitted poll.

Independent of expert ratings. No publisher rating, consensus or rating-derived target enters the House, Senate or governor model. Archived earlier editions retain their original inputs; the current model does not read those ratings.

Generic ballot. The trailing 60-day mean of admitted public generic-ballot readings (Democratic minus Republican). At this origin: +6.2 from 93 polls.

Campaign finance. Leading Democratic and Republican receipts per Senate race are read from FEC filings and gated into a share in [−1, 1]. The finance leg is present in the code and switched off by the backtest in every office (Section 6), so no dollar figure moves a published centre.

Prediction markets, presidential approval and special-election results are not inputs: approval added nothing to a six-cycle environment regression and special elections have no origin-dated historical source; markets are shown on the site for comparison only.

SECTION 3

Structural foundations

The structural prior is one penalised regression per office of the certified two-party margin on the features of Section 2. Nothing from a later evidence family reaches it: no polls, no generic ballot, no ratings, no finance. Features are median-imputed and standardised on the training rows; the regression is a Lasso whose penalty α is chosen by leave-one-cycle-out mean absolute error inside the training window from the grid {0.05, 0.1, 0.2, 0.4, 0.8, 1.5, 3.0}; and the loss is Huber (δ = 15 points), applied by iteratively reweighted least squares, so a landslide bounds its own leverage without being discarded.

$$c_{0,i} = \mu + \sum_k \beta_k z_{ik}, \quad \beta = \operatorname{argmin} \sum_i w_i (y_i - \mu - \beta \cdot z_i)^2 + \alpha \|\beta\|_1, \quad w_i = \min(1, \delta/|r_i|) \qquad (1)$$

The structural centre of race i: an L1-penalised regression on standardised features z with Huber weights w recomputed from the residuals r at each iteration.

The fitted penalties are α = 0.2 for the House (2,145 district-cycles, 2014-2024), α = 0.8 for the Senate (298 seat-cycles, 2008-2024) and α = 0.1 for governors (216 race-cycles, 2008-2024, odd-year elections included). The Senate list is deliberately compact: about three hundred rows cannot support thirty columns, and a wider list extrapolated a 2018 fold to a sixty-five point margin in New York. Races won by a candidate of neither major party are excluded from the fit (ME 2012, VT 2012, ME 2018, VT 2018, AK 2022, ME 2024, VT 2024 for the Senate; ME-02 2022 for the House; RI 2010 for governors).

The prior-margin feature enters uncapped. The House regression's largest coefficients are the prior margin (+12.31 points per standard deviation) and the relative presidential lean (+14.28); incumbency adds +3.03, and the midterm-against-the-White-House term −2.96. For governors the candidate-experience differential carries +6.19 and presidential lean +6.76. The full coefficient tables are in Appendix B.

Every later layer is fitted on the prior's *leave-one-cycle-out* centres, not its in-sample fit. A layer fitted on in-sample residuals would learn the prior's optimism; one fitted on held-out centres learns its real out-of-sample error, which is the quantity polling and calibration then have to correct.

The Senate state-history feature keeps the latest available regular contest two, four or six years earlier on its uncapped scale in both training and prediction; an uncontested contest stays missing. For MODEL_RECONSTRUCTED House inputs, a robust regression estimates the previous House margin from presidential votes on the same lines. Its full value supplies the main lag; bounded derivatives are recalculated from the replacement parent, retaining their historical training definitions. Observed and partial-official results keep their separate policies. Legacy producer files are preserved, not mistaken for this model's effective prior.

The reconstructed House prior is a + bP, where P = 100(D_pres - R_pres)/(D_pres + R_pres), in two-party margin points. The fitted intercept is −0.5887, slope 1.0679, and Huber epsilon 1.35. Robustness is selected by nested held-state prior-election errors. The coefficient fit uses 38 complete official 2024 reaggregations from North Carolina and Texas, not competitor forecasts. This limited donor geography is a generalization risk. Dated hindcasts use matching dated presidential vote totals; the general matrix's all-candidate margin is not silently substituted.

In the 2022 rolling-origin comparison, covering 402 races at six horizons, the House input replacement lowers overall MAE from 6.944 to 6.705 and safe-seat MAE from 8.473 to 7.927. Close-race MAE rises from 3.588 to 4.280; within the affected close subset it rises from 2.997 to 4.249. CRPS and interval scores improve, while Brier and log loss worsen slightly. With uncertainty fitted only on 2018/2020, 80% coverage rises from 69.61% to 72.47%, still below nominal. This is one election cohort and an explicit tradeoff, not an improvement in every statistic.

SECTION 4

National environment

The structural prior already carries the average environment of its training cycles through its intercept and lag terms. The environment leg adds the *news*: how far the current generic ballot says the coming election differs from that average, scaled by how much each office's structural error has historically moved with the national result.

$$c_{1,i} = c_{0,i} + \beta_o(a + \rho \cdot \mathrm{GB} - \bar{e}_o) \tag{2}$$

The environment update for office o: the generic ballot GB is mapped onto the realized-margin scale by (a, ρ) and compared with the mean realized environment ē of the training cycles; β is the office's residual elasticity.

β_o is the through-the-origin regression of each training cycle's mean out-of-fold prior residual on that cycle's realized national House margin (a result read only for cycles strictly before the forecast cycle), clipped to [0, 1.5]; (a, ρ) is the regression of the realized margin on the election-eve generic ballot over the cycles that have one. The fits are β = 0.3068 (House), 0.5210 (Senate), 0.2996 (governors); a = −4.5215, ρ = 1.3984; ē = −0.0828 for the House and +0.5504 for the statewide offices. ρ above one says the generic ballot has understated realized margins over these cycles; a absorbs its systematic bias.

At this origin the generic ballot of +6.2 maps to an environment estimate of +4.2, and the leg adds +1.30 points to every House district, +1.88 to every Senate race and +1.08 to every governor race. The uncertainty of the environment is not carried here as a parameter: it enters the joint simulation as the national shock of Section 11, whose scale is estimated from the cycle-level residuals that remain after this leg.

SECTION 5

Polling

Polls are aggregated pollster by pollster and then weighted against the centre in proportion to how much evidence they carry. Three constraints are structural and tested: the weight goes to zero as the evidence goes to zero, so an unpolled race keeps its prior; one physical pollster contributes at most one poll's worth of evidence, so repeated releases refresh recency without multiplying influence; and sponsor precision multiplies the absolute evidence, so lowering every poll's precision lowers polling's influence rather than only redistributing it.

$$r_k = 2^{-\mathrm{age}_k/H} \cdot \min\left(1.5, \sqrt{n_k/600}\right) \cdot s_k, \quad e_q = \max_{k \in q} r_k, \quad m_q = \frac{\sum_{k \in q} r_k m_k}{\sum_{k \in q} r_k} \tag{3}$$

Release k of pollster q gets a weight r from its age (half-life H), sample size n and sponsor precision s (1 for a public poll); a pollster's evidence e is its best release and its reading m the weighted mean of its releases, bias-adjusted for partisan sponsorship.

$$E = \sum_q e_q, \quad A = \frac{\sum_q e_q m_q}{E}, \quad w = w_{\max}(h) \cdot f(E), \quad c_{3,i} = (1 - w) c_{2,i} + w(A + b) \tag{4}$$

The race's evidence E and average A; the weight w rises with evidence through f (f(0) = 0) and with proximity through w_max(h); b is the office's poll-bias shift.

f(E) is either E/(E + k) or 1 − exp(−E/k), with k inflated in safer seats, $k_{\mathrm{eff}} = k\,(1 + \kappa|c|/20)$, so a lopsided race needs more evidence to move; $w_{\max}(h) = \mathrm{clip}(w_{\mathrm{eve}} - w_{\mathrm{slope}}\, h/60, 0, 1)$. All of these, and the aggregation constants, are chosen per office by pooled out-of-fold absolute error over a preregistered grid. The executing values are in the table; the half-life of 60 days and the saturation constant k = 0.25 were selected in every office.

Parameter	House	Senate	Governor
Half-life H (days)	60.0	60.0	60.0
Sponsor precision s for a partisan poll	1.00	0.25	1.00
Partisan bias adjustment (points)	2.0	2.0	2.0
Evidence curve f(E)	saturating	exponential	exponential
k	0.25	0.25	0.25
κ (safer seats need more evidence)	2.0	2.0	0.0
w_eve, w_slope per 60 days	1.00, 0.00	1.00, 0.20	1.00, 0.20
Poll-bias shift b (points)	+0.00 (none, by ablation)	−1.57	−2.53
Posterior sd of the shift (points)	—	1.02	0.92

Table 1: Executing poll parameters, selected per office by the backtest inside the training window.

The poll-bias posterior

Polls have missed in the same direction across whole cycles. For each training cycle the election-eve aggregate's mean signed error is computed; the shift b is the mean of a normal-normal posterior with prior N(0,

2^2) over those cycle errors, and its posterior standard deviation is kept. For the Senate the shift is −1.57 ± 1.02 points and for governors −2.53 ± 0.92: polls in these offices have read Democratic candidates too favourably, and the aggregate is shifted accordingly before it is blended. The House has no shift: by ablation it worsened both error and Brier score. The executing interval fit uses full-stack out-of-fold errors, including realized polling errors. The bias posterior is represented as a shared shock in the joint simulation, with national-shock variance adjusted to limit double counting; it is not added again to the new Senate marginal scale.

SECTION 6

Candidate information

Candidate information reaches the centre through three routes. Incumbency, its strength and tenure sit in the structural prior of every office. For governors the prior also carries the candidate-experience differential: the highest elected office each nominee has previously held on a four-step ladder, Democratic minus Republican, the one candidate-quality term that improved the out-of-fold record. Nothing else about a candidate, no fundraising, no ideology score, no scandal flag, moves a published centre.

$$c_{2,i} = c_{1,i} + \gamma_o \cdot m_i, \quad \gamma_o = 0 \ \text{ in every office} \tag{5}$$

The finance update. m is the gated share of leading receipts in [−1, 1]; γ is a ridge-through-origin coefficient on the out-of-fold residual after the environment leg.

The finance leg is executed and evaluated in every fold. It is switched off in every office by the backtest's own ablation: for the House the coefficient was negative and the ablation difference indistinguishable from zero; governors have no as-of-origin finance source; for the Senate removing it improved the Brier score with a confidence interval excluding zero and cost a mean absolute error inside noise. It stays in the code as a refusable, tested component, and the stage c_2 is identical to c_1.

Candidate *fields*, who is actually on the ballot and under what rule, are a separate matter and are handled in the simulation (Section 8).

SECTION 7

Model independence

Expert ratings were removed from all three offices by the owner's September 9, 2026 decision. The expert-free path excludes rating loaders and targets, then refits downstream centre calibration, error distributions and dependence using expert-free historical predictions. This is not a manual partisan shift or a claim that every validation score improves.

$$c_{4,i} = c_{3,i} \tag{6}$$

The former expert stage is the identity: the post-poll centre passes through unchanged.

$$w_{\text{expert},i} = 0 \tag{7}$$

Every race in every office has zero expert weight. No rating target is loaded.

The historical expert-enabled model remains archived for comparison. The removal is an independence and input-policy choice; its costs and benefits are evaluated on election outcomes, not on whether current forecasts move toward another forecaster.

SECTION 8

Candidate fields and election mechanics

Most races are a Democrat against a Republican and the simulated margin decides the winner. Where the ballot is different, the same field draw decides candidate winners and party seats. Same-party concentration and ranked-choice transfer parameters remain declared ranges integrated over simulation batches; their historical evidence is limited. The extra-candidate logit scale is now fitted separately to historical poll-to-election error, as described in Appendix C. It is not adjusted to produce a desired win chance in a current race.

- **Same-party general elections** (California's top-two, where two candidates of one party advance). The party is fixed; the two candidates split the party's vote by a Dirichlet draw centred on the primary shares, with a concentration drawn log-uniformly on [8, 60].
- **Extra candidates** (an independent or third-party candidate beside the pair). A polled extra takes a logit-normal share around its polled mean, read from the race's own field polls (9 races at this edition: AK-AL, CA-06, CA-07, CA-11, CA-40, GOV-AK, GOV-ME, SEN-MS, SEN-MT). A poll is admitted when it offers every candidate of the current field that this race's polls can measure — a candidate no pollster has ever named cannot be required, and one who has withdrawn is dropped from the denominator rather than counted against those still running. Field tables never enter the Democrat-versus-Republican margin. An unpolled extra takes a multiple of the race's admitted 'other' share, the multiple drawn uniformly on [0.30, 0.90]. An extra sharing a pivot's party is subtracted from THAT pivot; a cross-cutting candidate is split between the two in a proportion drawn uniformly on [0.25, 0.75]. Where a race's own current-field polls measure both pivots, the two-party centre is recentred on that measurement — dispersion untouched — and the winner is the plurality of the resulting shares.
- **Ranked-choice and top-four counts** (Alaska, Maine). Rounds eliminate the lowest candidate and transfer its votes with affinities drawn per batch: same party [0.45, 0.75], the opposite major party [0.05, 0.30], non-major [0.15, 0.45]; a share of each transferred pool, drawn on [0.05, 0.30], exhausts each round. The count ends when a candidate holds a majority of the continuing vote.
- **Runoffs.** A race whose rule sends the top two to a later runoff is simulated on its two-candidate coordinate; extras present in the first round transfer as above.
- **Independents.** A race whose second pivot is an independent is simulated on its D-versus-R coordinate. Nebraska's Senate race, where the independent Dan Osborn is the alternative to the Republican, is not credited to Democratic control: the Senate control probability is reported both with and without it.

Field specifications come from the governed candidate-field records: certified top-two pairs, ranked-choice and top-four rules, extra candidates with any polled share, and the race's admitted 'other' share. On this board 22 races carry a field specification beyond the plain pair.

After subtracting an extra candidate's votes, a pivot share cannot fall below zero. The nonnegative candidate shares are then rescaled together to the available ballot total, 100% minus the remaining 'other' share. This common factor preserves plurality ordering and ensures each simulated ballot totals 100%; ranked-choice transfers start from that consistent ballot.

Candidate high/low bands are vote-share percentages. History bands are margins in percentage points. In a complete two-candidate ballot a candidate's share is (100 + margin)/2, so its share interval is half as wide as the margin interval. That shortcut does not apply to multiparty shares or ranked-choice rounds.

SECTION 9

Centre calibration

After the evidence updates, governors receive the tail and G1 calibrations below. Senate races receive a smooth, party-symmetric correction for measured compression. House centres are not globally stretched: a fitted presidential-to-House estimator replaces reconstructed prior inputs before their dependent structural features are built.

Senate: a smooth tail and inward-polling correction

$$h(c) = \operatorname{sign}(c) T \left[\log\left(1 + \exp\left(\frac{|c| - K}{T}\right)\right) - \log\left(1 + \exp\left(-\frac{K}{T}\right)\right)\right] \tag{8}$$

c is the calibrated stack margin before this additional correction. The subtraction makes the curve continuous at zero.

$$d(c,u,v) = \operatorname{sign}(c)\left[1 - \exp\left(-\frac{|c|}{W}\right)\right]^2 \max(0, \operatorname{sign}(c)(u - v)); \quad c' = c + b_h h(c) + b_d d(c,u,v) \tag{9}$$

u and v are the same race's centres before and after polling. The drag correction is zero if polls did not pull the margin inward; original poll percentages remain unchanged.

Nonnegative coefficients between zero and one are fitted by bounded least squares, giving equal total weight to historical result bands below 5, 5–15, 15–25, 25–40 and 40-plus points. Inner leave-one-cycle-out selection tries K = 5, 10, 15 or 20; T = 2 or 5; W = 8, 16 or 32. It minimizes error in safe seats (absolute result at least 25), allowing less than one additional point of overall or close-race MAE versus the existing stack. This explicit trade-off was authorized on September 8. The zero correction remains available. No competing forecast or desired 2026 margin enters fitting. Senate uncertainty is refitted after the correction.

Senate parameter	Executing value
threshold	5.0000
temperature	2.0000
width	8.0000
tail	0.2088
drag	0.2837

Table 2: Parameters fitted from the historical training panel, not hand-set race adjustments.

The governor tail calibration

$$c_{5,i} = c_{4,i} + (b_t - 1) \cdot \operatorname{sign}(c_{4,i}) \cdot \max(0, |c_{4,i}| - 10), \quad b_t = 1.1319 \tag{10}$$

The governor tail calibration: an expansion beyond ±10 points, selected by nested leave-one-cycle-out from a family that also contained the identity, a global slope, a monotone hinge spline and an anchored spline.

The earlier centre-family selector remains unchanged: House and Senate chose identity there, while governors selected the tail expansion with b_t = 1.1319. The separate Senate correction above follows that stage and uses its explicitly stated tail-focused objective.

G1: the governor slope and intercept

$$c_i' = a + b \cdot c_{5,i}, \quad a = +0.7648, \quad b = 1.1543 \tag{11}$$

The published governor centre. One global slope and intercept, fitted by Huber iteratively reweighted least squares (δ = 15) on the executing model's out-of-fold final centres 2018-2024 at all six horizons against certified results.

Out of sample the stack had read safe governor races too close: the signed error of races decided by 15 points or more ran −5.73 points toward the prediction, and safe-race error was 7.57 points against 2.95 in competitive races. G1 is the smallest correction that repairs most of it: with the same folds the safe signed error becomes −2.98, safe-race error 6.87, overall error 5.873 to 5.712 and root mean squared error 7.50 to 7.28. The cost is real and accepted: competitive-race error rises from 2.95 to 3.34, because a global slope moves a close race by about a point too. Nested leave-one-cycle-out (the calibration refit with each cycle held out) is what these figures report, so the improvement is not the fit admiring itself.

The fit is computed for every office and applied to governors alone. For the House it is nearly the identity (a = −0.1558, b = 0.9716). For the Senate it is a = −2.6362, b = 0.8922: the intercept is the Senate's historical Democratic centre bias, discussed in Section 15, and it is not applied; the separate smooth correction addresses compression without a global partisan offset.

On this board the largest G1 movement is HI (+31.5 to +37.1); Colorado's governor race moves from +17.4 to +20.9 with a win probability of 98.6% and an 80% interval of +10.5 to +31.2. Every published governor record carries both the pre-calibration and the published centre.

SECTION 10

Race uncertainty

A race's uncertainty describes forecast error *e* = result − centre. The House and governors retain a two-component normal mixture, with a race-specific core and a pooled tail. The Senate retains a Student-t family, now fitted on full-model out-of-fold errors. Its probability and every interval level use the same CDF. Candidate-field winners and party seats are resolved in the joint simulation; an analytical D-versus-R coordinate cannot give a Democratic win chance to a ballot with no Democrat.

F: the core+tail mixture (House, governors)

$$e \sim (1-\pi)N(0, s_{\text{core}}^2) + \pi N(0, s_{\text{tail}}^2), \quad \pi = 0.0580, \quad s_{\text{tail}} = 24.44 \tag{12}$$

The error distribution. With probability π the race is a tail race whose error has the pooled tail scale; otherwise its error has the race's own core scale.

$$\begin{aligned} g(X) = \beta_1 \mathbb{1}[\text{House}] + \beta_2 \mathbb{1}[\text{Senate}] + \beta_3 \mathbb{1}[\text{Governor}] \\ + \beta_4 h/60 + \beta_5 \log(1 + \min(E, 6)) + \beta_6 \min(|c|, 25)/25 + \beta_7 \mathbb{1}[\text{open}] \end{aligned} \tag{13}$$

The original seven core predictors. The following term prevents their centre response from flattening beyond 25 points.

$$\log s_{\text{core}} = g(X) + \beta_8 \max\left(\frac{|c|}{25} - 1, 0\right), \quad \beta_8 \geq 0 \tag{14}$$

All eight coefficients, mixture weight and tail scale are fitted jointly. The extra coefficient cannot make the core shrink farther into the tail. The final scale remains clipped to [1,40].

Term	β	Reading
1[House]	+1.6713	House core scale at the election eve, unpolled, even race, incumbent: exp(β) = 5.32 points
1[Senate]	+1.8884	the Senate level enters the fit through pooling only; the Senate's published distribution is its own scale law
1[Governor]	+1.8336	governor core scale under the same conditions: 6.26 points
h/60	+0.0441	× 1.045 per 60 days of horizon
log(1 + min(E, 6))	−0.0903	six pollster-equivalents of evidence narrow the core by a factor of 0.84
min(|c|, 25)/25	+0.0148	a 25-point centre widens the core by a factor of 1.01
1[open seat]	+0.1380	an open seat widens the core by a factor of 1.15
max(|c|/25 − 1, 0)	+0.2889	each 25 points beyond a 25-point centre multiplies the core by 1.34

Table 3: The executing F parameters, maximum likelihood on 10,314 out-of-fold residuals of all three offices, 2018-2024, governor residuals taken after G1, with a weak prior centring π near 10%.

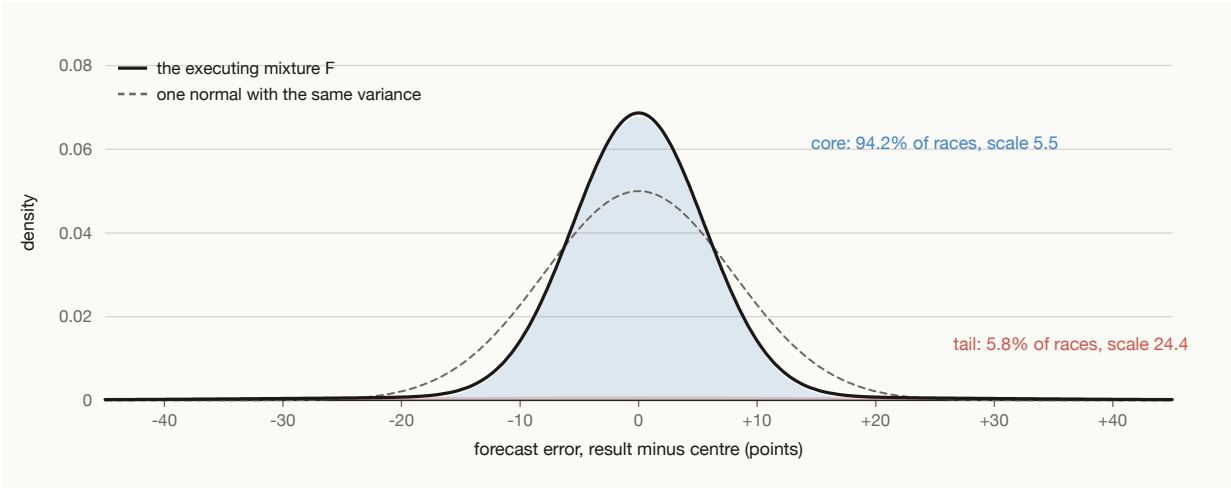

Figure 2: The core+tail error distribution F for an unpolled House race 53 days out with a centre of 5 points (core scale 5.55). The blue core carries ordinary error; the red tail, 5.8% of races, carries rare large misses. A single normal with the same variance (dashed) is too wide in the middle and too thin in the tails.

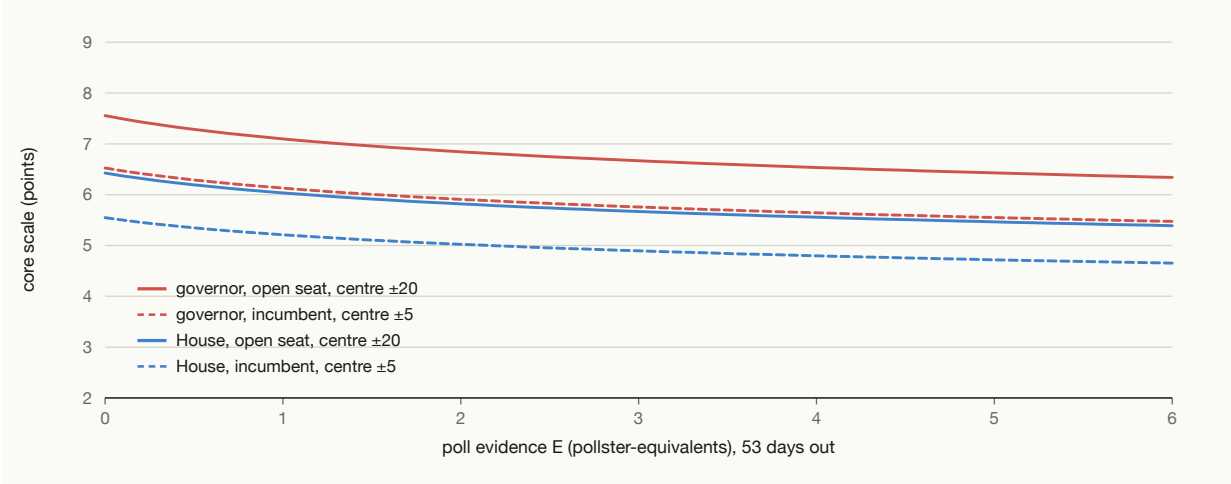

Figure 3: The heteroskedastic core scale as a function of poll evidence, 53 days out. More polling narrows the core; an open seat and a larger margin widen it; governors are wider than House districts.

$$P(D \text{ wins}) = 1 - F_e(-c'), \quad [q_{\text{lo}}, q_{\text{hi}}] = c' + F_e^{-1}((1-L)/2), c' + F_e^{-1}((1+L)/2) \tag{15}$$

Probability and intervals from the mixture: the probability is the mixture mass above zero; the level-L interval's ends are mixture quantiles found by bisection on the mixture CDF.

The mixture's variance for the illustrated race is 63.6 points squared, but its shape is not a normal's: a race with a 19-point centre keeps a win probability near 99% rather than certainty, because the tail component has real mass twenty points from the centre. Every published House and governor record declares its family, core scale, tail scale and tail share, and a release gate recomputes the probability and the 80% interval from the declared parameters and refuses the edition on any mismatch.

The Senate scale law

$$\log \sigma_i = a_0 + a_1 h/60 + a_2 \log(1 + \min(E, 6)) + a_3 \min(|c|/25, 1) + a_4 \max(|c|/25 - 1, 0), \quad a_4 \geq 0 \tag{16}$$

Quantile regression on log absolute full-stack out-of-fold error, targeting its 80th percentile minus log(normal q90). The fitted horizon slope controls how quickly uncertainty declines; the nonnegative continuation prevents tail saturation. Sigma is bounded to [2,35].

$$P(D \text{ wins}) = T_{\nu(c/(\sigma z))}, \quad \text{half-width}_L = \sigma z T_\nu^{-1}((1+L)/2) \tag{17}$$

One Student-t probability law, including at a tied centre. All interval endpoints are quantiles of that same CDF; there is no separate probability-only recalibration.

The executing Senate values are a = (+2.3272, +0.0127, −0.2819, −0.1598, +0.4760), ν = 8, z = 0.9493, b_p = 1.0000 (identity, not applied), and $z_{0.50}$ = 0.671, $z_{0.80}$ = 1.326, $z_{0.95}$ = 2.189. Held-cycle standardized residuals supply their 80th absolute-error percentile; z preserves that percentile while ν is selected by likelihood from {4, 6, 8, 12, 30}. Training uses the full-stack 2012–2024 out-of-fold record and excludes the evaluated cycle. Historical residuals already include realized poll errors; no separate polling-error variance is added again to this marginal scale.

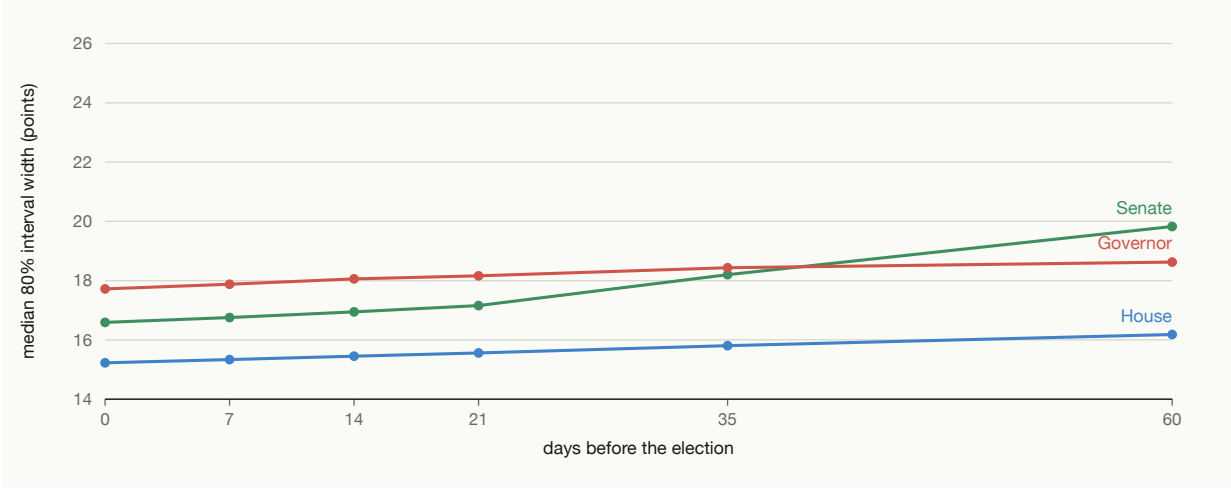

Figure 4: Median width of the published 80% interval by forecast horizon in the out-of-fold record, 2018-2024. The Senate's Student-t law widens with the horizon; the House and governor mixture is nearly flat because most of its width is the tail.

SECTION 11

Correlated simulation

Races do not miss independently. A national mood that the polls did not see moves every race in the same direction, regions move together, and districts in one state share their state's swing. The joint simulation draws

every race's error as the sum of four components, so that seat totals and chamber control carry the dependence a reader would expect from a wave.

$$e_i = s_{N,o} Z_{N,o} + s_{R,o} Z_{\text{region}(i),o} + s_{S,i} Z_{\text{state}(i)} + s_{\text{res},i} Z_i,$$
$$Z_{N,o} = \sqrt{\lambda_o} Z_{\text{common}} + \sqrt{1-\lambda_o} Z_o \tag{18}$$

The error of race i in office o: a national shock (shared across offices through λ), a regional shock and a state shock (both shared across offices, each office loading them at its own scale), and a race-specific residual, all standard normal draws.

$$s^2_{S,i} = \rho_o \left(s_i^2 - s^2_{N,o} - s^2_{R,o} \right),$$
$$s^2_{\text{res},i} = (1-\rho_o) \left(s_i^2 - s^2_{N,o} - s^2_{R,o} \right) \tag{19}$$

The state and residual scales are sized so that the four components sum to the race's own core scale s_i: the shared components are the dependence structure, and they never add variance a race was not published with.

The scales are estimated from the model's own out-of-fold election-eve residuals at the 60-day horizon, 2018-2024, with the election cycle as the independent unit: $s_{N,o}$ is the standard deviation of cycle-mean residuals (3.56 House, 3.58 Senate, 4.01 governors), $s_{R,o}$ the root mean square of region-by-cycle means after the cycle mean (1.37, 1.89, 3.35), ρ_o the same-state residual correlation (0.04, 0.80, 0.00), and λ_o the share of the national variance common to all offices, from the shrunk correlations of cycle-mean residuals between offices (1.00, 0.85, 0.91).

Four cycles cannot pin a standard deviation down, and the simulation does not pretend they can. Each of the 2000 batches of the simulation draws its own parameter vector: the national and regional scales from their posteriors under a cross-office hierarchical lognormal prior (the House national scale's 10th, 50th and 90th percentiles are 2.8, 3.7 and 4.9 points), and ρ and λ from Fisher-z posteriors with the cycle count as the effective sample. The published seat distribution therefore integrates over parameter uncertainty; a fixed-parameter sensitivity grid is reported beside it.

The poll-bias common shock and the tail indicator

The Senate and governor poll-bias posteriors of Section 5 are drawn once per batch in one shared direction, each office taking its own Student-t quantile of it (sd 1.02 with 8 degrees of freedom for the Senate; 0.92 with 10 for governors) and added to every race in proportion to its poll weight. Because the national scale was estimated from residuals that already contain the realized shift error, its variance is reduced by the common shock's share, floored at a quarter of its value, so nothing is counted twice.

For House and governor races the tail indicator of F is drawn per race per draw with probability $\pi = 0.0580$. The shared national, regional and state components are unchanged, being sized from the core scale; the tail's extra variance enters only the race-specific residual, so that conditional on the indicator the race's marginal is exactly $N(0, s_{\text{core}}^2)$ or $N(0, s_{\text{tail}}^2)$ and the unconditional marginal is the published mixture.

The simulation runs 100,000 joint draws. A reconciliation over ordinary House and governor races (456 races, 200,000 draws) finds a maximum difference of 0.0025 between simulated and standalone win probabilities (Monte Carlo standard error 0.0011 at even odds) and 0.11 points in the 80% quantiles. Senate dependence begins with correlated normal scores; their normal-CDF ranks are transformed through the fitted Student-t inverse CDF. The resulting Senate marginal uses the same scale and tail shape as its published probabilities and intervals, including an exactly tied race. Candidate-field rules subsequently determine candidate winners.

SECTION 12

Chamber control

A draw is a full election night. The House total is the number of the 435 simulated districts the Democrat wins; a majority is 218 seats. The Senate total adds the 34 Democratic-aligned seats not on the ballot to the simulated wins; control needs 51. The governor count is the number of the 36 contests the Democrat wins, of which 19 is a majority of those on the ballot.

$$\begin{gathered}\text{seats}_d = \text{fixed}_o + \sum_{i \in o} \mathbb{1}[D \ \text{ wins } \ i \ \text{ in draw } \ d], \\ P(\text{control}) = \text{mean}_d \mathbb{1}[\text{seats}_d \geq M_o], \quad E[\text{seats}] = \text{fixed}_o + \sum_{i \in o} P_i\end{gathered} \tag{20}$$

Chamber outcomes from the joint draws. Because every public quantity is a functional of one draw set, the expected seat count from the draws equals the sum of the race probabilities up to Monte Carlo error by construction.

On this board the House expectation is 229.5 Democratic seats (draw mean and marginal sum differ by 0.00), with a 73.8% probability of a Democratic majority and a 90% seat range of 204 to 261; the Senate expectation is 50.3 Democratic-aligned seats, control 48.0% (47.0% without crediting Nebraska); governors 19.5 of 36. Cross-chamber outcomes (both chambers to one party, split control) are read from the same draws, and the tipping-point and seat-distribution products are regenerated from the persisted draws by the exporter, never from a second simulation.

SECTION 13

Historical validation

The validation protocol is nested rolling-origin, leave-one-cycle-out: for each of the four full-tier cycles every fitted layer (prior penalty, environment, poll curve, centre calibration, G1, F and the Senate law) is refit without the evaluated cycle, the held-out cycle is scored at all six horizons with the evidence available at each origin, and the scores are pooled. Expert ratings are absent throughout. The numbers below are those of the executing code, not of a research replica. Older cycles (2012-2016) are not pooled with the full tier.

Metric (out of fold, 2018-2024, six horizons)	House	Senate	Governor
Race-horizon rows	8,970	786	558
Mean absolute error of the centre (points)	5.642	5.869	5.712
Competitive races, |result| < 5	4.713	3.192	3.344
Safe races, |result| ≥ 15	5.930	6.954	6.870
Signed error of safe races toward the prediction	−0.252	+1.010	−2.982
Root mean squared error	8.528	8.084	7.276
Mean signed error (D − R, points)	+0.038	+2.746	−0.261
CRPS	4.259	4.244	4.080
Weighted interval score	3.400	3.289	3.134
Brier score	0.0414	0.0543	0.0462
Log loss	0.1396	0.1768	0.1590
Calibration slope (logit of outcome on logit of forecast)	0.99	1.34	2.16
50% interval coverage	0.509	0.538	0.457
80% interval coverage	0.789	0.805	0.803
95% interval coverage	0.936	0.962	0.950
Median 80% width, 60 days (points)	16.2	19.8	18.6
Median 80% width, 35 days	15.8	18.2	18.4
Median 80% width, election eve	15.2	16.6	17.7

Table 4: Historical validation of the executing model. Every parameter is refit with the test cycle held out; the rows are race-horizon pairs. The full out-of-fold record — every race, cycle and horizon with its centre, win probability, result and fold provenance — is published at /methodology/ctmm22-out-of-fold.csv; filter on one horizon to read a cycle without stacking horizons.

Three things to hold against the table. The House and Senate probability slopes are near one (0.99 and 1.34); the governor slope of 2.16 means governor probabilities remain under-confident, an artefact of the office's small folds (35 races in 2018, 11 races in 2020, 36 races in 2022, 11 races in 2024). The Senate's mean signed error of +2.75 points is the Democratic centre bias of Section 15. And 80% coverage is 0.79, 0.81 and 0.80: the House and governor intervals are honest to within two points of their nominal level, the variation across margin bands matters as much as the overall coverage.

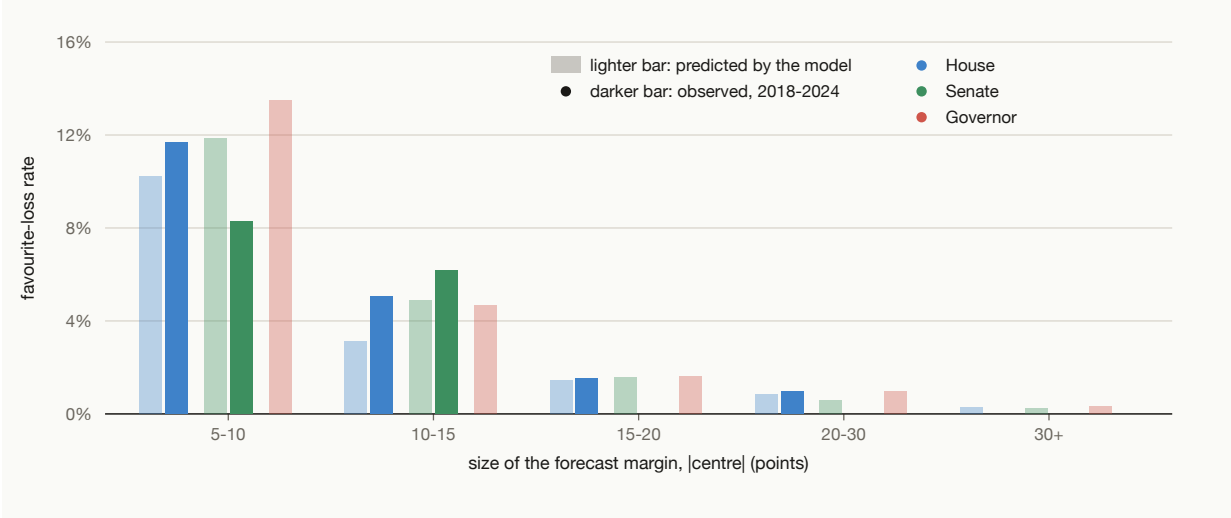

Figure 5: Favourite-loss rates the model assigned (lighter) and the rates observed (darker) in the out-of-fold record, by the size of the forecast margin. Governor favourites of ten points or more did not lose in 2018-2024. One Senate favourite did: the 2020 Georgia special election, which the stack centred at about R+13 at every horizon and Warnock won by 11.9 points — that single race, counted once per horizon, is the 6.2% observed in the 10-15-point Senate bin (six of 97 race-horizon rows). The model keeps a small probability that a favourite can lose, and the record shows why.

|centre| bin	House n	House predicted	House observed	Senate n	Senate predicted	Senate observed	Governor n	Governor predicted	Governor observed
5-10	840	10.2%	11.7%	121	11.8%	8.3%	104	13.5%	0.0%
10-15	848	3.1%	5.1%	97	4.9%	6.2%	95	4.7%	0.0%
15-20	858	1.5%	1.5%	81	1.6%	0.0%	54	1.6%	0.0%
20-30	1777	0.8%	1.0%	162	0.6%	0.0%	103	1.0%	0.0%
30+	3652	0.3%	0.0%	184	0.2%	0.0%	97	0.3%	0.0%

Table 5: Favourite-loss rates assigned by the model and observed in the out-of-fold record, by the size of the forecast margin.

Cycle	House MAE	House CRPS	House cov.80	Senate MAE	Senate CRPS	Senate cov.80	Governor MAE	Governor CRPS	Governor cov.80
2018	5.61	4.18	0.80	6.57	4.66	0.78	5.79	4.15	0.78
2020	5.58	4.10	0.80	7.75	5.63	0.63	7.24	5.15	0.53
2022	7.22	5.36	0.64	5.51	3.90	0.88	4.93	3.55	0.91
2024	4.03	3.31	0.93	3.54	2.72	0.94	6.50	4.49	0.80

Table 6: Validation by held-out cycle. Four cycles is the whole full-tier record; the spread across them is the honest uncertainty of every pooled figure.

Days out	House CRPS	House Brier	House cov.80	Senate CRPS	Senate Brier	Senate cov.80	Governor CRPS	Governor Brier	Governor cov.80
60	4.375	0.0445	0.78	4.486	0.0539	0.82	4.947	0.0508	0.77
35	4.272	0.0404	0.79	4.334	0.0543	0.79	3.940	0.0438	0.81
21	4.249	0.0418	0.79	4.219	0.0564	0.80	4.103	0.0409	0.78
14	4.237	0.0412	0.79	4.281	0.0555	0.79	4.048	0.0494	0.80
7	4.220	0.0407	0.79	4.096	0.0530	0.82	3.817	0.0470	0.83
0	4.200	0.0395	0.79	4.050	0.0527	0.81	3.623	0.0451	0.83

Table 7: Validation by forecast horizon.

Stratum (by result)	House n	House MAE	House cov.80	Senate n	Senate MAE	Senate cov.80	Governor n	Governor MAE	Governor cov.80
competitive<5	924	4.71	0.81	120	3.19	0.94	84	3.34	0.92
5-15	1800	5.12	0.80	234	5.24	0.76	192	5.05	0.82
15-25	1830	5.03	0.79	210	6.13	0.73	132	6.00	0.79
25-35	1698	5.47	0.78	120	4.24	0.94	96	7.14	0.72
35+	2718	6.82	0.78	102	11.85	0.73	54	8.51	0.76
unpolled	6611	5.85	0.79	124	9.92	0.80	39	9.41	0.67

Table 8: Validation by stratum of the certified result and for unpolled races.

The race distribution, by ablation

The distribution is judged by margin scores (CRPS and weighted interval score), winner scores (Brier and log loss), coverage and the shape of its bands across safer and closer races. The 2.2 non-saturating extension explicitly trades small aggregate score losses against better-calibrated House tail and middle-range widths: compared with the seven-term mixture, held-cycle House CRPS worsens 0.0105 and WIS 0.0115, while 80% coverage above a 40-point predicted margin improves from 70.4% to 75.2%. This is not an improvement in every score or every cycle. The table compares the executing distribution with the retained Student-t reference law.

Metric	House, F	House, reference law	Governor, F	Governor, reference law
CRPS	4.259	4.292	4.080	4.097
Weighted interval score	3.400	3.449	3.134	3.137
Brier score	0.0414	0.0422	0.0462	0.0498
Log loss	0.1396	0.1460	0.1590	0.1674
80% coverage	0.789	0.851	0.803	0.803
Median 80% width, 60 days	16.2	21.6	18.6	26.9
Median 80% width, eve	15.2	19.8	17.7	14.6

Table 9: Ablation of the race distribution: the executing core+tail mixture F against the Student-t sigma law that the Senate uses, refit for the House and governors in the same folds. Centres are identical in both columns of each office except that the governor's reference law is scored around its own pre-G1 centre.

SECTION 14

Hindcast and issued history

The site's history charts show two kinds of point. From 2026-08-01 to the day before launch the series is a *reconstruction*: today's CTMM 2.2 method scored at each historical date, not a forecast published then and not an out-of-sample backtest. The replay starts with dated repository captures and supplements polling records collected later only when a publication date, archive-entry date or earlier observed capture establishes availability by the origin. Fieldwork must also have ended by that date. A field end alone is not evidence that a poll was already public. Unknown availability excludes an individual poll, not the whole day's reconstruction. Expert ratings are excluded from the reconstruction, as they are from the current forecast. House district/prior rows come from the historical matrix rather than today's boundaries. Capture dates can lag a real-world announcement; per-date input manifests record those vintages and limitations. Every reconstructed point is labelled reconstructed = true and issued = false and drawn faded.

From the launch marker, 2026-09-09, issued points come from forecasts that reached publication; the current page appends its own reading without splicing in earlier-model history. Inputs are checked several times a day, and a change in model-relevant evidence can produce a new live forecast at any of those checks. Separately, one permanent daily snapshot preserves that scheduled archival reading and its content hashes. Daily archiving is not a limit on live updates, and neither operation requires rebuilding the website. Historical snapshots are immutable; corrections create new records rather than rewriting an issued forecast. The reconstruction contains 40 daily points through 2026-09-09.

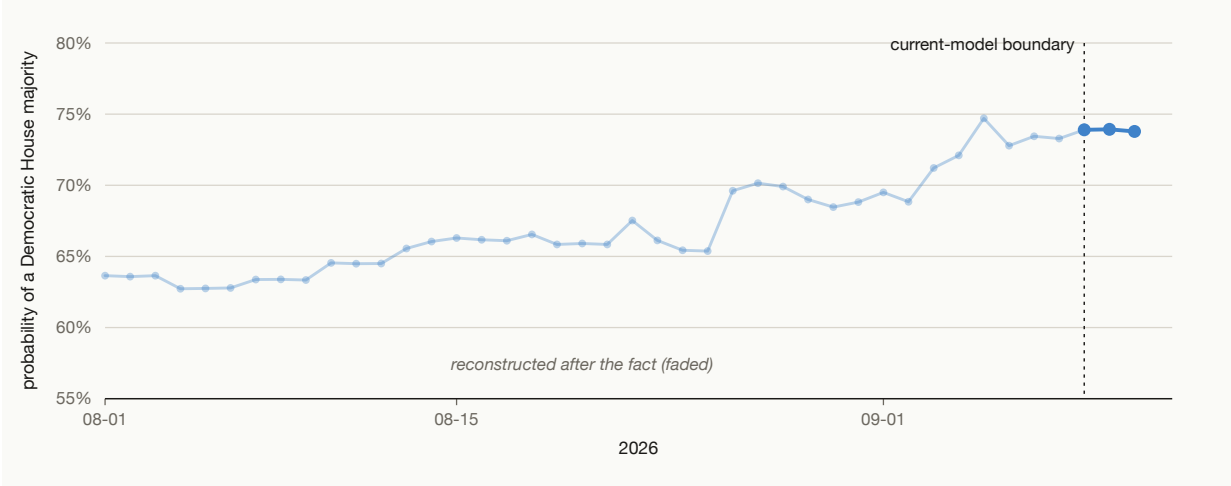

Figure 6: House control: faded reconstructions from 2026-08-01; solid current-model readings from 2026-09-09, including this guide's edition. A local preview does not establish publication.

A reader comparing two points that straddle the launch marker is comparing a reconstruction with an issued forecast; the chart marks the seam and the site's movement surfaces say so. Movement across a model-version date is recorded in the model-revision register as a change of method, not as news about the race.

SECTION 15

Limitations

- **Few comparable cycles.** The full validation record is four elections. Every pooled figure in Section 13 hides a spread across them (2020 was a bad Senate year for polls; 2022 a bad House year), and the dependence scales of Section 11 are estimated from four cycle means. The simulation integrates over that uncertainty; it cannot remove it.
- **Residual Senate bias remains.** Out of fold, 2018-2024, the Senate's mean signed error is +2.75 points (positive means too Democratic). The smooth correction targets safe-seat compression, not a desired partisan result. A global Senate G1 intercept and slope are not applied. Improvement in tail error does not establish that every aspect of calibration improved.
- **Governor probabilities are under-confident.** The calibration slope of 2.16 says a stated 80% has historically won more often than 80% of the time. The folds are small (governor races come 11 to 36 at a time) and a slope estimated on them would be an unstable correction; the under-confidence is left in.

- **Safe-seat error remains.** The signed safe-seat error is −0.3 points for the House, +1.0 for the Senate, and −3.0 for governors; negative values indicate compression. A smoother central margin does not by itself establish accurate uncertainty bands in the most lopsided races.
- **Unusual fields rest on declared ranges.** Same-party pairs, extra candidates, ranked-choice transfers and exhaustion have no historical panel large enough to fit. Their parameters are declared ranges, the probabilities integrate over them and sensitivity intervals are published, but a range is a statement of ignorance, not a measurement.
- **Rare mechanics are approximated.** Runoffs are simulated on their two-candidate coordinate; independents on the D-versus-R coordinate; an unpolled extra candidate takes a share from the race's 'other' vote. In a race like Alaska's top-four count the person the simulation favours can differ from the sign of the two-party coordinate, and the site shows both rather than reconciling them by hand.
- **Expert ratings are excluded.** No third-party rating or consensus target enters current forecasts. Earlier expert-enabled editions remain archived, not rewritten.
- **Pre-launch history is reconstructed.** Points before the launch marker were computed afterwards on today's poll bank. They are the best available account of what the model would have said, and they are not forecasts anyone could have read on those dates.
- **Refused features stay refused.** Presidential approval, special elections, poll trend and disagreement, pollster house effects, latent-environment models, splines and interactions in the priors, statewide indices and latent candidate quality were each tested on the historical record and did not improve it out of fold. They are documented as refusals and are not in the model.

APPENDIX

A. Executing parameters

Layer	Parameter	House	Senate	Governor
Structural prior	Lasso penalty α; loss	0.2; Huber δ=15	0.8; Huber δ=15	0.1; Huber δ=15
	Training rows; cycles	2,145; 2014-2024	298; 2008-2024	216; 2008-2024
Environment	β; ē	0.3068; −0.0828	0.5210; +0.5504	0.2996; +0.5504
	a; ρ (shared)	−4.5215; 1.3984	same	same
Finance	γ	0 (off)	0 (off)	0 (off)
Polls	H; s; bias adj.	60.0; 1.00; 2.0	60.0; 0.25; 2.0	60.0; 1.00; 2.0
	f; k; κ; w_eve; w_slope	saturating; 0.25; 2.0; 1.0; 0.0	exponential; 0.25; 2.0; 1.0; 0.2	exponential; 0.25; 0.0; 1.0; 0.2
	bias shift b ± sd	none	−1.57 ± 1.02	−2.53 ± 0.92
Experts	cap; w0; w1; d; φ; target	0.0; 0.0; 0.0; 0.0; 0.0; band	0.0; 0.0; 0.0; 0.0; 0.0; band	0.0; 0.0; 0.0; 0.0; 0.0; band
Centre calibration	tail family; b_t	identity	smooth tail + inward-poll drag (Section 9)	tail_linear; 1.1319
	G1 a; b (applied)	—	—	+0.7648; 1.1543
Distribution	family	F core+tail	Student-t scale law	F core+tail
	π; s_tail	0.0580; 24.44	—	0.0580; 24.44
	F β	+1.6713, +1.8884, +1.8336, +0.0441, −0.0903, +0.0148, +0.1380, +0.2889	—	same
	Senate a; ν; z; b_p	—	(+2.3272, +0.0127, −0.2819, −0.1598, +0.4760); 8; 0.9493; 1.0000	—
	Senate z_0.50; z_0.80; z_0.95	—	0.671; 1.326; 2.189	—
Dependence	s_N; s_R; ρ; λ	3.56; 1.37; 0.04; 1.00	3.58; 1.89; 0.80; 0.85	4.01; 3.35; 0.00; 0.91
	common poll-bias shock sd; df	—	1.02; 8	0.92; 10
Simulation	draws; batches	100,000; 2000	same	same
Chamber	fixed seats; majority	0; 218	34; 51	0; 19

Table 10: Every fitted or declared number the executing model uses, by layer and office.

APPENDIX

B. Structural prior coefficients

Feature (standardised)	Coefficient (points per standard deviation)
hs_margin_lag1	+12.311
hs_margin_lag2	+0.121
hs_pres_lean_cd_rel	+14.277
hs_inc_running	+3.033
hs_inc_strength	0 (removed by the penalty)
hs_open_signed	0 (removed by the penalty)
hs_defender_party	0 (removed by the penalty)
hs_overperf_lag1	+1.590
hs_prev_uncontested	0 (removed by the penalty)
midterm_wh	−2.961
wh_party	0 (removed by the penalty)
env_house_natl_prev	−0.376
ix_hs_overperf_lag1__x__midterm_wh	+1.784
ix_hs_margin_lag1__x__hs_inc_running	0 (removed by the penalty)
tf_hs_margin_lag1_abs	0 (removed by the penalty)
tf_hs_margin_lag1_sq	0 (removed by the penalty)
demo_bachelors_2017	+1.024
demo_median_household_income_2017	+0.064
demo_median_age_2017	−0.061
region_div1	+0.101
region_div2	0 (removed by the penalty)
region_div3	0 (removed by the penalty)
region_div4	0 (removed by the penalty)
region_div5	0 (removed by the penalty)
region_div6	0 (removed by the penalty)
region_div7	0 (removed by the penalty)
region_div8	+0.042
region_div9	0 (removed by the penalty)

Table 11: House structural prior: Huber-Lasso coefficients, penalty α = 0.2, 2,145 training races, cycles 2014-2024.

Feature (standardised)	Coefficient (points per standard deviation)
sen_seat_margin_lag1	+4.647
sen_state_margin_last	+2.663
sen_seat_uncontested_lag1	0 (removed by the penalty)
pres_lean_rel	+10.770
inc_running	+5.356
open_seat_signed	0 (removed by the penalty)
defender_party	0 (removed by the penalty)
midterm_wh	−2.262
env_house_natl_prev	0 (removed by the penalty)
ix_inc_running__x__midterm_wh	0 (removed by the penalty)
ix_pres_lean_rel__x__midterm_wh	+0.920

Table 12: Senate structural prior: Huber-Lasso coefficients, penalty α = 0.8, 298 training races, cycles 2008-2024.

Feature (standardised)	Coefficient (points per standard deviation)
gov_pres_lean_rel	+6.762
gov_seat_margin_lag1	+3.291
gov_midterm_wh	−2.979
gov_inc_running_signed	0 (removed by the penalty)
gov_inc_tenure_signed	+3.592
gov_exp_diff	+6.189

Table 13: Governor structural prior: Huber-Lasso coefficients, penalty α = 0.1, 216 training races, cycles 2008-2024.

APPENDIX

C. Candidate-field priors and fitted share uncertainty

Assumption	Range (drawn per batch)
Same-party Dirichlet concentration	log-uniform on [8, 60]
Unpolled extra candidate, multiple of the 'other' share	uniform on [0.30, 0.90]
Ranked-choice transfer affinity, same party	uniform on [0.45, 0.75]
Transfer affinity, opposite major party	uniform on [0.05, 0.30]
Transfer affinity, non-major	uniform on [0.15, 0.45]
Ballot exhaustion per round	uniform on [0.05, 0.30]
Split of an extra candidate's vote between the two pivots	uniform on [0.25, 0.75]
Extra-candidate share logit scale	0.7206, fitted to historical poll-to-result errors

Table 14: Only the extra-share scale is fitted here. Other field priors remain declared assumptions, not empirical estimates.

$$\tau = Q_{0.80,w}(|\mathrm{logit}(y/100) - \mathrm{logit}(p/100)|)/\Phi^{-1}(0.90) \tag{21}$$

The fitted logit scale: each candidate has equal total weight across horizons. y is the election share; p is the dated poll aggregate. Fusion ballot lines are combined and reallocated ranked-choice questions are excluded.

The historical fit has 686 horizon observations for 140 Senate candidates across four cycles; only eight candidates have aggregate shares at least 8%. Applying this constant to House/governor extras and unpolled candidates is an explicit transfer assumption. Held-cycle coverage of the 80% share interval improves from 93.9% at the old 1.1 to 78.6%; share CRPS improves from 1.1326 to 0.9504, while log-density score worsens from 1.3637 to 1.4211. This variance change does not correct the observed tendency of polls to overstate minor candidates.

APPENDIX

D. Edition and code revision

This document was generated on 2026-09-13 from the executing artifacts of the CTMM 2.2 lane from a checkout based on code revision 3b4f2401593d; the board it accompanies is edition ctm-2.2-20260911-41a79ca1a423

(origin 2026-09-11, 53 days before the election). Scheduled refreshes re-run the same code on new inputs; the methodology is frozen from activation, and no parameter in this document changes except through the governed mechanisms of the release specification (source corrections, verified bugs, election-rule or candidate changes, new evidence through the unchanged pipeline).

APPENDIX

E. Glossary

Term	Meaning
Centre	The forecast margin, Democratic minus Republican, in percentage points.
Horizon h	Days between the forecast origin and the election.
Out of fold / leave-one-cycle-out	Fitted with the scored election held out entirely; the only kind of historical figure this document reports.
Poll evidence E	Pollster-equivalents of admitted polling, after recency, sample-size and sponsor weights; one poll's worth per pollster at most.
CRPS	Continuous ranked probability score: the expected absolute error of the whole forecast distribution, lower is better.
WIS	Weighted interval score over the 50/80/95% intervals, lower is better.
Brier score / log loss	Proper scoring rules for the win probability, lower is better.
Coverage	The share of results that fell inside the stated interval; an 80% interval should cover about 80%.
Calibration slope	The slope of the logit of the outcome on the logit of the forecast; 1 is calibrated, above 1 under-confident.
Core scale, tail scale, π	The parameters of the F mixture: the ordinary-error scale, the rare-miss scale and the share of rare misses.
Reconstructed / issued	A history point computed after the fact from that date's inputs, versus one that was actually published on that date.

Table 15: Terms used in this document.