Supplementary Materials:
Dynamic team attack and defense strength in international rugby union

Contents

1. Dataset

The match data were compiled from results listed on Ultimate Rugby (https://www.ultimaterugby.com/). The analysis includes complete results for tier-one men's national teams (e.g., Australia, England, Italy) over the study period, together with results for several major tier-two teams (e.g., Fiji, Japan, United States). Because some lower-tier opponents appear only when playing teams in the core sample, coverage is complete for tier-one teams but partial for some other nations. Table S1 lists the 32 teams appearing in the estimation sample and the number of matches observed for each team.

Table S1. Teams included in the sample

Team	*N* Matches	Team	*N* Matches
Argentina	63	Namibia	10
Australia	67	Netherlands	16
Belgium	14	New Zealand	72
Brazil	5	Poland	3
Canada	27	Portugal	44
Chile	11	Romania	47
England	69	Russia	12
Fiji	40	Samoa	28
France	64	Scotland	63
Georgia	50	South Africa	63
Germany	10	Spain	38
Hong Kong	4	Switzerland	3
Ireland	62	Tonga	28
Italy	61	USA	37
Japan	31	Uruguay	21
Kenya	5	Wales	68

2. Bayesian model parameter summaries

This section reports posterior summaries for the main parameters of the three Bayesian specifications used in the main paper: the weekly random walk (RW), the continuous-time random walk (CTRW), and the two-speed event-gap model (TS). Each model was run with four chains, using 1000 warmup iterations and 1000 post-warmup draws per chain. The target acceptance statistic, `adapt_delta`, was set to 0.95. In all tables, 5% and 95% denote posterior quantiles, $\widehat{R}$ is the potential scale reduction factor, and ESS denotes effective sample size. The parameters σ_α and σ_δ denote attack and defense innovation scales; in the TS model, τ_α and τ_δ denote additional active-appearance scale parameters.

Table S2. RW parameter summaries

Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.149	0.057	3.054	3.243	1.001	3966	3045
κ_a	3.115	0.032	3.064	3.166	1.001	4097	2678
ϕ_h	9.731	0.982	8.173	11.390	1.001	3263	2969
ϕ_a	7.194	0.755	6.014	8.517	1.000	3632	2641
η	0.061	0.056	-0.032	0.152	1.000	4292	2989
σ_α	0.022	0.004	0.016	0.029	1.002	1105	2023
σ_δ	0.012	0.004	0.006	0.019	1.004	748	1126
$\sigma_{\alpha 0}$	0.495	0.090	0.361	0.650	1.006	1556	2177
$\sigma_{\delta 0}$	0.556	0.078	0.439	0.691	1.001	1450	2403

Table S3. CTRW parameter summaries

Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.147	0.054	3.059	3.235	1.000	2659	2616
κ_a	3.114	0.030	3.064	3.164	1.001	3545	3108
ϕ_h	9.389	0.927	7.971	10.982	1.000	3320	3137
ϕ_a	7.066	0.723	5.964	8.317	1.000	2583	2261
η	0.062	0.054	-0.027	0.152	1.000	2846	2954
σ_α	0.020	0.004	0.014	0.027	1.004	872	1891
σ_δ	0.011	0.004	0.004	0.017	1.003	642	530
$\sigma_{\alpha 0}$	0.555	0.083	0.430	0.700	1.000	1073	1662
$\sigma_{\delta 0}$	0.576	0.082	0.456	0.723	1.002	832	1441

Table S4. TS parameter summaries

Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.145	0.057	3.051	3.239	1.002	3619	2776
κ_a	3.116	0.030	3.066	3.165	1.000	5088	3313
ϕ_h	9.494	0.978	8.015	11.206	1.001	3582	2894
ϕ_a	7.124	0.734	5.999	8.398	1.002	3643	2841
η	0.068	0.057	-0.028	0.163	1.001	3574	2906
τ_α	0.028	0.016	0.003	0.053	1.005	845	1812
τ_δ	0.017	0.010	0.002	0.035	1.004	1215	1602
σ_α	0.016	0.007	0.003	0.025	1.005	836	1203
σ_δ	0.008	0.005	0.001	0.016	1.005	922	1527
$\sigma_{\alpha 0}$	0.547	0.079	0.432	0.691	1.002	1470	2306
$\sigma_{\delta 0}$	0.575	0.079	0.456	0.709	1.003	1174	1927

3. Traceplots

The figures below report traceplots for the main parameters of each Bayesian specification.

Figure S1. Traceplots: RW model

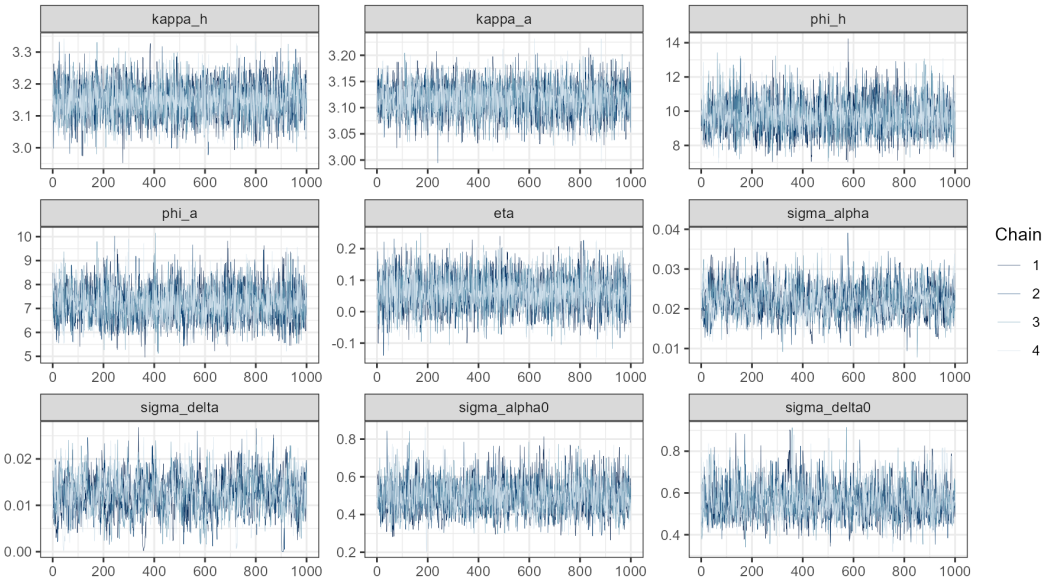


Figure S2. Traceplots: CTRW model

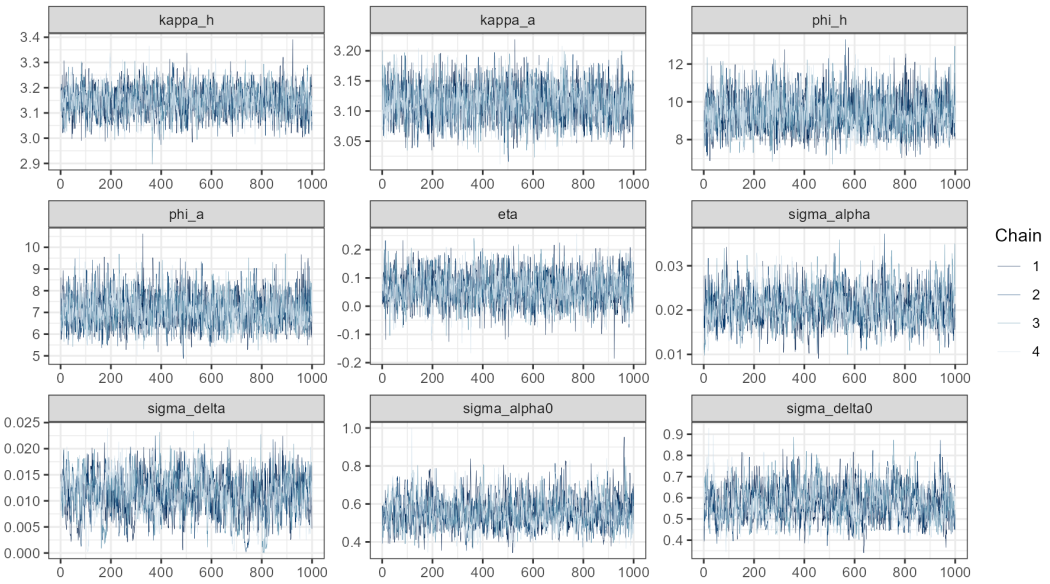
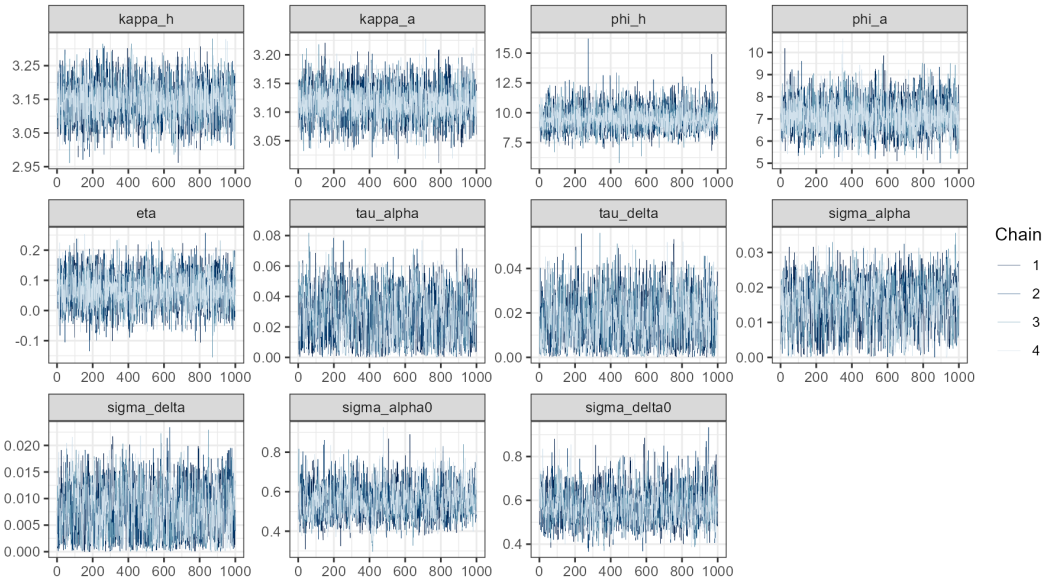


Figure S3. Traceplots: TS model



4. Posterior predictive checks

The following figures report posterior predictive density overlays for home score (left) and away score (right).

Figure S4. Posterior predictive checks: RW model

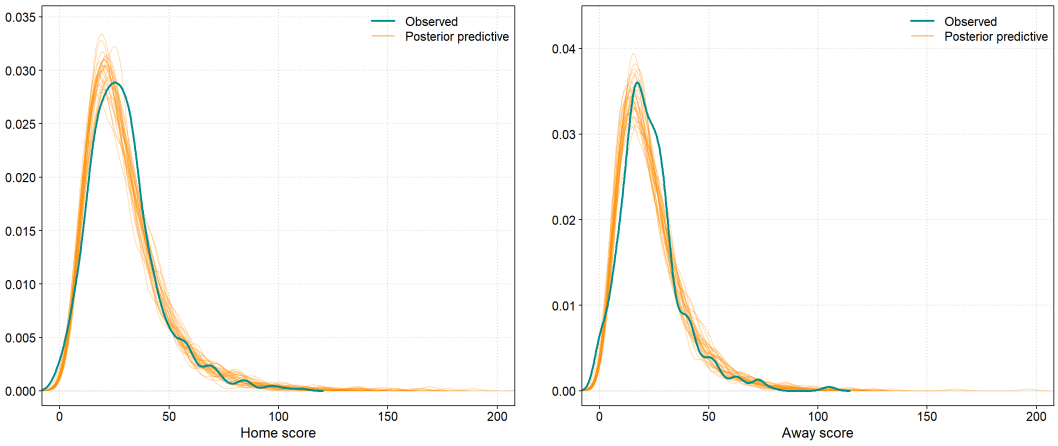


Figure S5. Posterior predictive checks: CTRW model

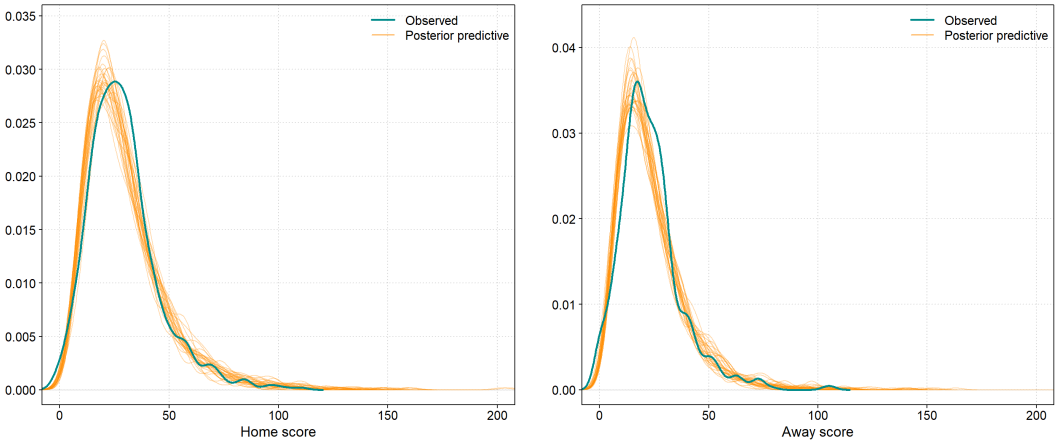
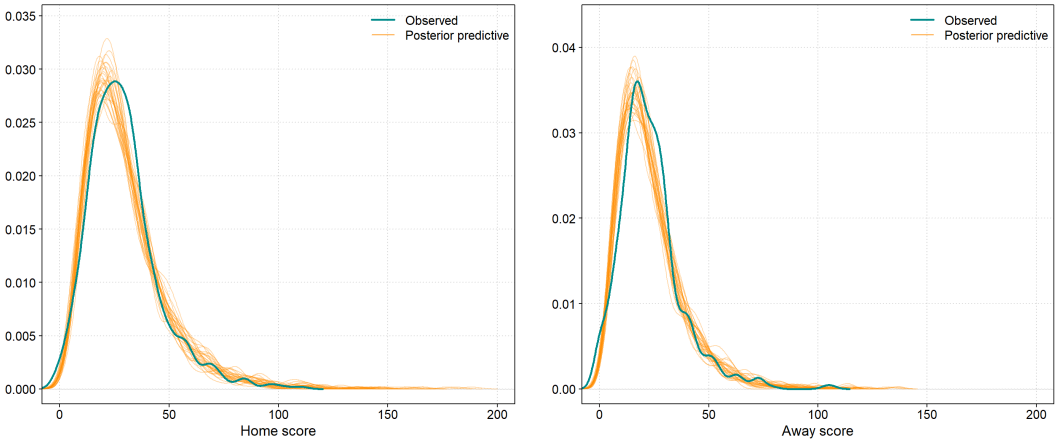


Figure S6. Posterior predictive checks: TS model



5. Autoregressive specifications

We develop and fit an autoregressive AR(1) version of the baseline random walk model:

$$\alpha_{k,t} \sim \mathcal{N}\Big(\rho_\alpha\, \alpha_{k,t-1},\, \sigma_\alpha^2\Big), \tag{1}$$

$$\delta_{k,t} \sim \mathcal{N}\Big(\rho_\delta\, \delta_{k,t-1},\, \sigma_\delta^2\Big), \tag{2}$$

for $t \geq 2$. Initial states are drawn from $\alpha_{k,1} \sim \mathcal{N}(0, \sigma_{\alpha 0}^2)$, and $\delta_{k,1} \sim \mathcal{N}(0, \sigma_{\delta 0}^2)$. This formulation implies persistence with mean reversion: shocks to team quality carry forward over time, but decay geometrically at rates governed by ρ_α and ρ_δ.

In practice, the AR parameters $(\rho_\alpha, \rho_\delta)$ are estimated to be very close to 1 (see Table S5), implying an absence of mean reversion in latent attack and defense strength over the weekly time frame of the analysis. This is consistent with the enduring strength of the main rugby union nations (e.g., England, New Zealand) over the 100+ year history of the sport.

Table S5. AR parameter summaries

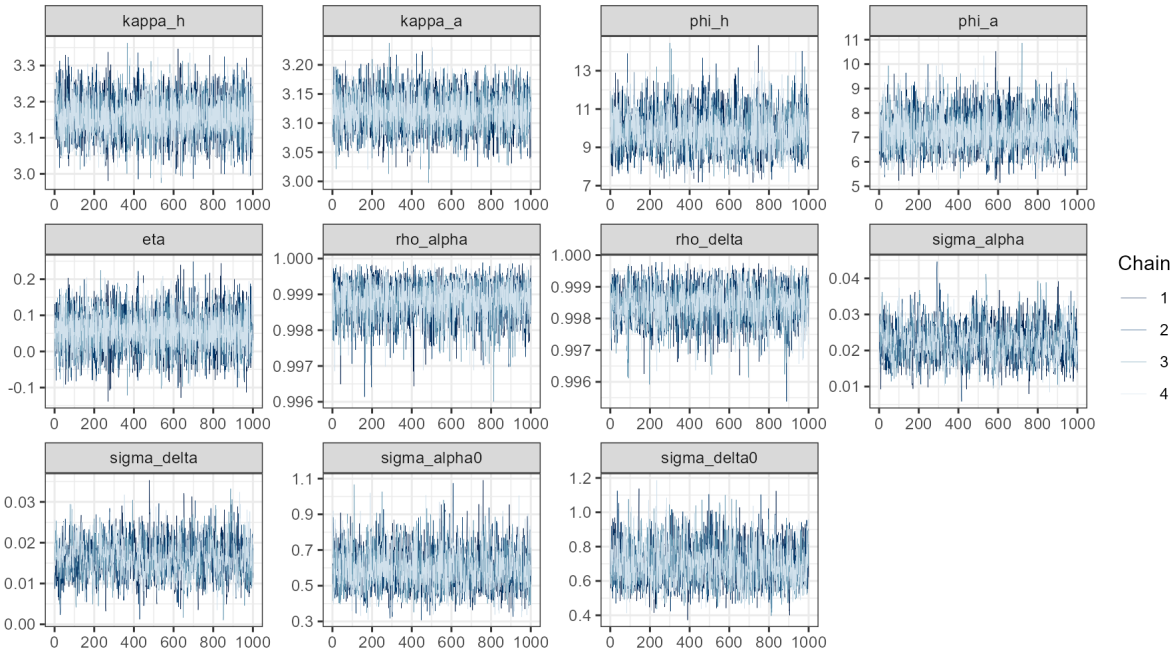
Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.158	0.058	3.062	3.254	0.999	3708	2854
κ_a	3.120	0.033	3.066	3.173	1.000	4430	2680
ϕ_h	9.805	1.010	8.262	11.595	1.003	2949	3032
ϕ_a	7.243	0.754	6.078	8.564	1.000	3517	3128
η	0.057	0.056	-0.035	0.149	1.001	4406	2968
ρ_α	0.999	0.001	0.998	1.000	1.000	3903	2827
ρ_δ	0.998	0.001	0.997	0.999	1.000	2234	2384
σ_α	0.022	0.004	0.015	0.029	1.003	908	1808
σ_δ	0.016	0.004	0.009	0.023	1.003	835	1122
$\sigma_{\alpha 0}$	0.595	0.106	0.435	0.779	1.003	2086	2450
$\sigma_{\delta 0}$	0.691	0.113	0.524	0.887	1.003	1488	2615

Table S6. Out-of-sample predictive performance, AR(1) model

Model	MJLPD	CRPS_H	CRPS_A	CRPS_{PD}	RMSE_{PD}	Brier	Time	Max $\hat{R}$	Min ESS
AR(1)	-7.76	6.53	6.69	10.29	19.76	0.144	13.1	1.014	227
ECES	—	—	—	—	20.55	0.147	<0.1	—	—
WRRA	—	—	—	—	22.47	0.170	<0.1	—	—

Notes: MJLPD = mean joint log predictive density; CRPS_H = CRPS for home score; CRPS_A = CRPS for away score; CRPS_{PD} = CRPS for point difference; RMSE_{PD} = RMSE for point difference; Brier = Brier score for home-win probability; Time = median computation time in minutes across evaluation weeks; Min ESS = minimum bulk ESS. ECES = Error correction exponential smoother; WRRA = World Rugby rankings algorithm.

Figure S7. Traceplots: AR model



6. Prior sensitivity

As a prior-sensitivity check, we re-estimate the two-speed and autoregressive models under a broader prior specification. For both models, the scoring intercept priors are widened from $\mathcal{N}(0, 1)$ to $\mathcal{N}(0, 1.5)$, the home-advantage prior from $\mathcal{N}(0, 0.5^2)$ to $\mathcal{N}(0, 0.75^2)$, the initial-state scale priors from $\mathcal{N}^+(0, 0.5^2)$ to $\mathcal{N}^+(0, 0.75^2)$, the innovation-scale priors from $\mathcal{N}^+(0, 0.2^2)$ to $\mathcal{N}^+(0, 0.35^2)$, and the negative-binomial dispersion $\phi_\cdot$ priors from Gamma$(2, 0.5)$ to Gamma$(2, 0.2)$. Since the Gamma distribution is specified using the shape–rate parameterization, this changes the prior mean for the dispersion parameters from 4 to 10, placing more prior mass on weaker overdispersion while still allowing strongly overdispersed scores. For the autoregressive model, we also replace $\rho_\alpha, \rho_\delta \sim$ Beta$(6, 3)$ with $\rho_\alpha, \rho_\delta \sim$ Beta$(2, 2)$, weakening the prior preference for high persistence and allowing a wider range of mean-reversion rates.

The results in Tables S7 and S8 show that the main parameter estimates are robust to these alternative priors. The largest changes occur for the negative-binomial dispersion parameters, as expected given the broader prior on ϕ_h and ϕ_a, but these changes are modest and do not alter the substantive interpretation. The autoregressive parameters remain very close to one under the broader Beta$(2, 2)$ prior, reinforcing the conclusion that the AR specification behaves much like a random walk in this application.

Table S7. AR sensitivity analysis parameter summaries

Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.160	0.057	3.067	3.250	1.000	3769	2903
κ_a	3.121	0.032	3.069	3.175	1.001	4298	3446
ϕ_h	10.094	1.060	8.433	11.902	1.001	2406	2712
ϕ_a	7.400	0.785	6.199	8.772	1.000	2740	2636
η	0.054	0.056	-0.036	0.144	1.000	4209	3062
ρ_α	0.999	0.000	0.998	1.000	1.002	3484	2734
ρ_δ	0.999	0.001	0.998	1.000	1.000	2481	2744
σ_α	0.022	0.004	0.015	0.030	1.002	824	1694
σ_δ	0.015	0.004	0.008	0.023	1.005	778	1479
$\sigma_\alpha 0$	0.595	0.106	0.437	0.782	1.004	1366	2129
$\sigma_\delta 0$	0.686	0.114	0.516	0.887	1.002	1767	2632

Table S8. TS sensitivity analysis parameter summaries

Parameter	Mean	SD	5%	95%	$\widehat{R}$	ESS_{bulk}	ESS_{tail}
κ_h	3.148	0.054	3.060	3.236	1.001	3686	3110
κ_a	3.117	0.029	3.070	3.165	1.002	4698	2923
ϕ_h	9.843	1.035	8.294	11.666	1.001	2989	2852
ϕ_a	7.296	0.755	6.169	8.609	1.001	3166	2897
η	0.064	0.054	-0.026	0.152	1.001	4112	3214
τ_α	0.027	0.016	0.002	0.054	1.011	613	1496
τ_δ	0.018	0.011	0.002	0.037	1.007	1018	1746
σ_α	0.016	0.006	0.004	0.025	1.006	759	1385
σ_δ	0.009	0.005	0.001	0.016	1.003	1082	1529
$\sigma_{\alpha 0}$	0.554	0.086	0.431	0.708	1.001	1344	2110
$\sigma_{\delta 0}$	0.583	0.084	0.461	0.737	1.001	904	1414

7. Error correction exponential smoother benchmark

As a benchmark, we implemented a simple exponential error-correction smoother (ECES) for match margin. This model represents each team by a single latent scalar rating rather than separate attack and defense strengths. Let $r_{i,t}$ denote the rating of team i immediately before match t. For a match between home team i and away team j, the benchmark predicts point difference as

$$\mu_t = r_{i,t} - r_{j,t} + \eta H_t,$$

where H_t is an indicator for non-neutral matches and η is a fixed home-advantage term. The observed margin is

$$m_t = y_{t,h} - y_{t,a},$$

with $y_{t,h}$ and $y_{t,a}$ denoting home and away scores.

After each match, both teams' ratings are updated symmetrically according to the forecast error,

$$e_t = m_t - \mu_t.$$

The home and away ratings are then revised as

$$r_{i,t+1} = r_{i,t} + \alpha e_t, \qquad r_{j,t+1} = r_{j,t} - \alpha e_t,$$

where $\alpha \in (0, 1)$ is a smoothing or learning-rate parameter. This is an error-correction formulation because ratings are shifted in proportion to the latest prediction error. It is also an exponential smoother, since repeated substitution implies that current ratings are weighted averages of past information with geometrically declining weights on older results. Higher values of α make the benchmark more responsive to recent matches, while lower values imply slower-moving ratings.

The ECES benchmark was tuned on the 2024 calendar year. Specifically, all matches before 1 January 2024 were used to initialize team ratings, after which the model was run sequentially through 2024 under candidate values of α and η. The chosen parameter pair minimized one-step-ahead out-of-sample error in 2024, with point-difference RMSE as the primary criterion and home-win Brier score as a secondary criterion. The selected values were then fixed for the 2025 out-of-sample evaluation.

For comparability with the Bayesian models, the ECES benchmark was evaluated in the same weekly rolling-origin framework. For each held-out week in 2025, ratings were reconstructed using all matches strictly before that week, then held fixed while predicting the matches within the evaluation week. Ratings were updated only after all matches in the held-out week had been completed. Thus, as in the Bayesian comparison, no information from earlier matches within the same evaluation week was used to predict later matches in that week.

The ECES model directly produces a point-difference prediction μ_t, from which we computed RMSE for points difference. To obtain home-win probabilities for Brier-score evaluation, we used a normal approximation analogous to that applied in the World Rugby benchmark. Specifically, we assumed

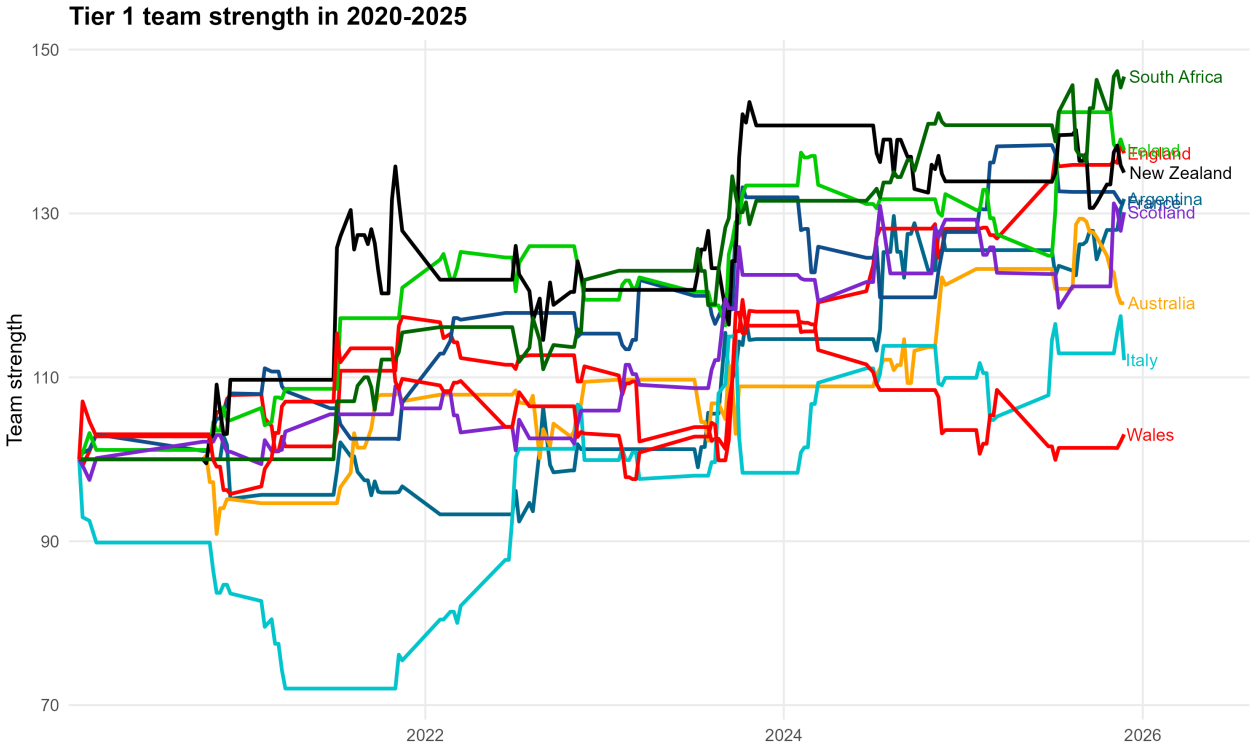
$$m_t \sim \mathcal{N}(\mu_t, \sigma^2),$$

where σ was estimated from residuals in the training data. Under this approximation, the implied probability of a home win is

$$P(\text{home win}_t) = \Phi\left(\frac{\mu_t}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. These probabilities were then used to compute Brier scores for the weekly out-of-sample evaluation.

Figure S8. ECES team strength



Team strength estimates from the exponential error-correction smoother, centered so that the average team-week value equals 100.

8. World Rugby rankings benchmark

As an additional benchmark, we evaluated the official World Rugby men's ranking algorithm. The World Rugby system is a points-exchange rating method in which each team carries a scalar rating that changes only when that team plays a match. Pre-match ratings are adjusted for home advantage by treating the home team as three rating points stronger, except at neutral venues. Rating exchanges then depend on the match result, the adjusted pre-match rating difference between the two teams, whether the margin of victory exceeds 15 points, and whether the match is a Rugby World Cup fixture, in which case exchanges are doubled. A key feature of the official system is that teams cannot gain points for losing or lose points for winning. World Rugby also specifies that new teams enter the system with an initial rating of 30 points.

For comparability with the official system, we initialized teams in our sample at their published World Rugby men's ratings on 6 January 2020. Ratings were then updated match by match using the official exchange rules over the 2020–2025 sample. Ratings in this benchmark remain fixed between matches: if a team does not play, its rating does not move.

The World Rugby algorithm does not itself produce a full predictive distribution for scorelines, nor does it directly output expected point differences or win probabilities. To use it in the out-of-sample evaluation, we therefore treated the adjusted pre-match rating difference, $(R_{home} + 3 \cdot H) - R_{away}$, as the benchmark prediction of points difference, where H is an indicator for non-neutral matches. This yielded a deterministic prediction for match margin and allowed us to compute RMSE for points difference directly.

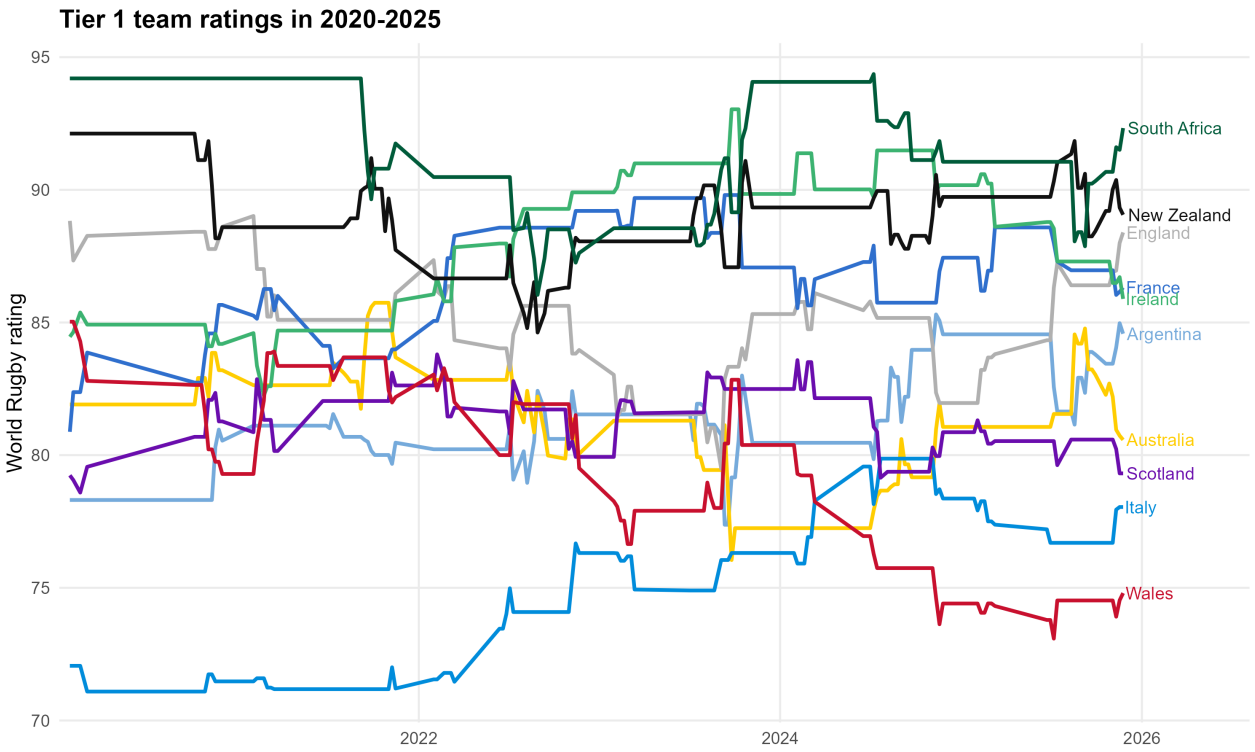
For comparability with the Bayesian models, the World Rugby benchmark was also evaluated in the same weekly rolling-origin framework. For each held-out week in 2025, ratings were reconstructed using all matches strictly before that week and then held fixed while predicting the matches within the evaluation week. Thus, no information from earlier matches within the same evaluation week was used to predict later matches in that week.

To obtain home-win probabilities for Brier score calculation, we imposed a simple normal approximation. Specifically, we assumed that the observed points difference was normally distributed around the benchmark mean implied by the adjusted rating difference, with variance estimated from historical residuals in the 2020–2024 training period. If μ_g denotes the benchmark point-difference prediction for held-out match g and σ denotes the estimated residual standard deviation, then the implied home-win probability is

$$P(\text{home win}_g) = \Phi\left(\frac{\mu_g}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. These probabilities were then used to compute Brier scores in the same weekly out-of-sample framework used for the Bayesian models. This procedure preserves the official ranking algorithm as the core benchmark while allowing it to be compared to the Bayesian models on the common metrics of points-difference RMSE and home-win Brier score.

Figure S9. World Rugby rating points

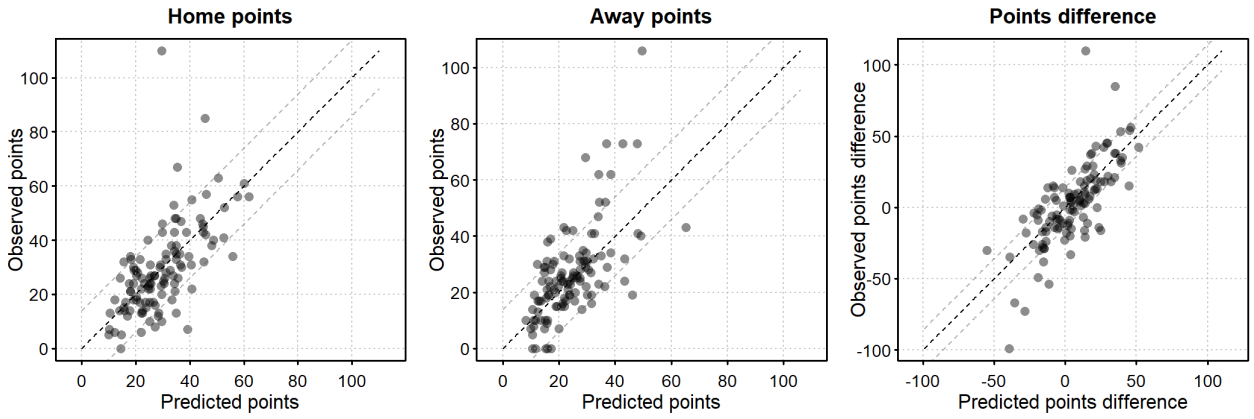


Team rating points produced by the World Rugby algorithm. Our reproduced ratings differ slightly from the official World Rugby rankings because the official system is based on the full population of eligible international matches, whereas our dataset is restricted to 32 teams and is only complete for tier-one sides.

9. Observed vs predicted scores for all models

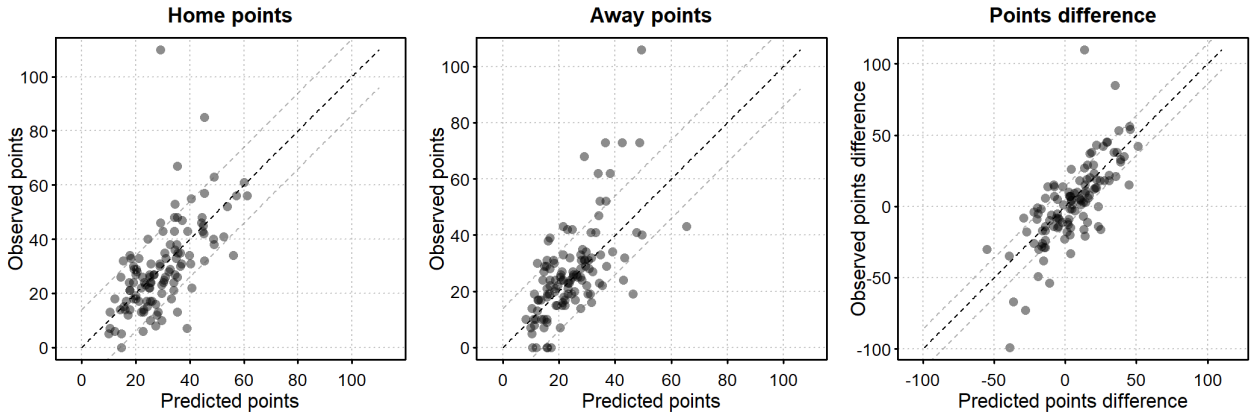
Figures S10–S12 show observed and predicted home points, away points, and point differences for the 116 matches in the 2025 held-out sample. The same plotting scale is used within each panel type, and the diagonal reference lines indicate prediction errors of −14, 0, and +14 points. The ±14 point bands correspond to two converted tries, providing a rugby-specific benchmark for the size of the prediction errors.

Figure S10. Observed vs predicted scores, RW model



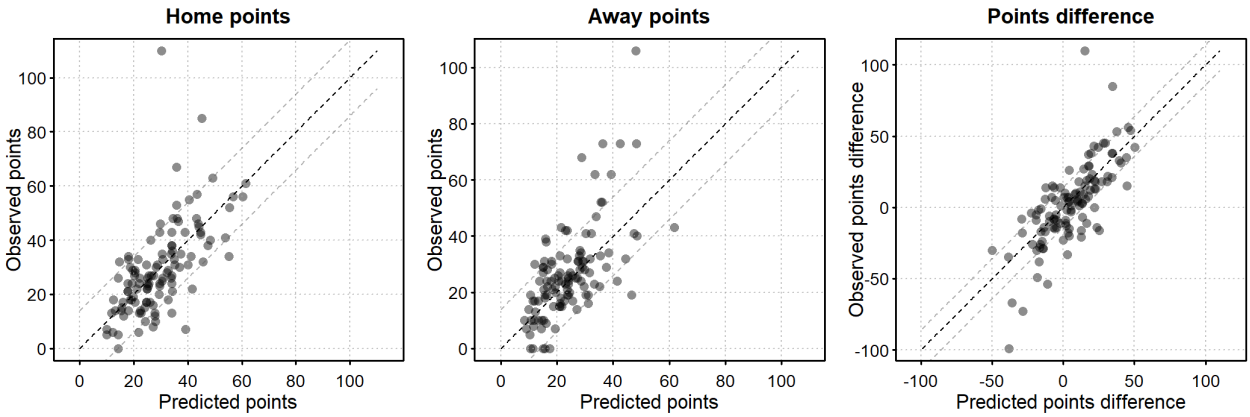
Observed home points, away points, and point differences for the matches in the held-out sample (y-axes), plotted against predicted values from the random-walk model (x-axes).

Figure S11. Observed vs predicted scores, CTRW model



Observed home points, away points, and point differences for the matches in the held-out sample (y-axes), plotted against predicted values from the continuous-time random-walk model (x-axes).

Figure S12. Observed vs predicted scores, TS model



Observed home points, away points, and point differences for the matches in the held-out sample (y-axes), plotted against predicted values from the two-speed model (x-axes).