

# 1. Executive Summary

## 1. Executive summary

- Train games=940 | Derby games=75
- Nested CV=4 outer folds x 3 inner folds (expanding time-aware)
- Selected model family: Regime-aware simplex stack (Ridge/Huber/HistGB/HistGB-bag) + q50 blend
- Selected Elo variant: elo\_base\_static
- Nested outer RMSE/MAE: 43.363 / 34.420
- Calibration=pooled | Scale=pooled | Regime stack=False
- Runtime safety: fast\_mode=False | budget\_triggered=histgb\_bag\_auto\_reduce, max\_fits

# 1B. Runtime Bottleneck And Fixes

Runtime bottleneck and fixes

- Profile-first instrumentation added for hotspot phase totals, per-family \_predict\_family timing, model fit counts, and top slow events.
- Shared split-level matrix/weight cache keyed by train/val index signatures + feature signature + half-life to eliminate repeated \_xy/time-weight prep.
- Bundled family prediction path prepares X/y/w once per split and fits Ridge/Huber/HistGB/HistGB-bag/q50 sequentially.
- HistGB bagging reduced to 2 models by default (fast mode uses 1), with deterministic seed schedule and budget-aware auto-reduction.
- Candidate scan uses conservative early pruning: coarse score on first inner split, then full evaluation on top candidates only.
- Budget enforcement: total time, scan time, and fit cap stop further scanning/training and continue with best-so-far.
- Budgets triggered: histgb\_bag\_auto\_reduce, max\_fits
- Fast mode=False; optimization toggles enabled=True.
- Reproducibility: seed=23, OMP\_NUM\_THREADS=1, MKL\_NUM\_THREADS=1.
- Artifact validation: predictions rows=75 (expected 75), rankings rows=165 (expected 165), ranks unique 1..165=True.

1C. Fast Profiling Before vs After (timing / fits) (1/1)

kind	name	before_sec	after_sec	delta_sec
phase	_scan_and_select_core_candidates	5.230	5.524	0.294
phase	_core_scan_score	5.226	5.516	0.290
phase	_generate_inner_oof_and_outer_preds	5.540	3.973	-1.567
phase	_predict_family	10.500	9.416	-1.084
phase	_build_split_tables	0.447	0.334	-0.113
phase	_prepare_variant_outer_data	0.550	0.440	-0.110
fit_count	histgb	48.000	51.000	3.000
fit_count	histgb_bag	8.000	6.000	-2.000
fit_count	histgb_q50	8.000	6.000	-2.000
fit_count	huber	8.000	6.000	-2.000
fit_count	ridge	48.000	51.000	3.000

1D. Runtime Focus Phase Timing Summary (1/1)

phase	total_sec	calls
_scan_and_select_core_candidates	29.713	5
_core_scan_score	29.679	1205
_predict_family	28.813	3217
_prepare_variant_outer_data	1.981	20
_build_split_tables	1.597	80
_generate_inner_oof_and_outer_preds	0.081	15

1E. \_predict\_family Aggregated Timing / Fit Counts (1/1)

family	predict_time_sec	predict_calls	fit_count
histgb	28.543	1516	200
ridge	0.266	1516	200
huber	0.002	61	0
histgb_q50	0.001	61	0
histgb_bag	0.001	61	0
histgb_q20	0.000	1	0
histgb_q80	0.000	1	0

## 2-4. Changes / Leakage / Nested CV

### 2. What changed vs previous pipeline

- Nested tuning for post-model layers and key hyperparameters (Ridge alpha, HistGB grid, recency half-life).
- Regime-specific simplex stacking + regime-specific calibration + regime/pool scale correction with fallback.
- Recency features: EMA multi-alpha, last3/5/8, trend, consistency ratio.
- Upgraded Elo variants: dynamic-K, team-specific regularized home effect, inactivity decay.
- HistGB seed bagging and quantile HistGB (median, plus q20/q80 for diagnostics).
- Shift diagnostics (KS + PSI + out-of-range fraction) and OOD-triggered mitigation.

### 3. Data constraints and leakage controls

- All features are pregame-only and built sequentially.
- Static rating models refit inside each train split; no outer-fold leakage in postprocessing.
- Derby predictions are neutral-site (home effects zeroed for derby features).

### 4. Nested CV design (outer/inner schematic)

```
outer_fold=1 train_n=187 val_n=207 train_end=2025-01-24 val=2025-02-04..2025-03-17
outer_fold=2 train_n=394 val_n=210 train_end=2025-03-17 val=2025-03-18..2025-04-28
outer_fold=3 train_n=604 val_n=162 train_end=2025-04-28 val=2025-04-29..2025-05-28
outer_fold=4 train_n=766 val_n=174 train_end=2025-05-28 val=2025-05-29..2025-06-30
```



5. Outer-fold RMSE/MAE table (1/1)

outer_fold	rmse	mae	bias	pred_std	actual_std	resid_std
1	45.161	35.778	0.420	22.528	43.833	45.159
2	43.045	34.280	-7.993	0.000	42.297	42.297
3	39.367	31.303	0.080	0.000	39.366	39.366
4	45.879	36.321	4.009	5.150	45.945	45.703



6. Feature additions and recency ablations (core scan) (1/1)

feature_profile	half_life_days	scan_score_mean	scan_histgb_rmse_mean	scan_ridge_rmse_mean	n_configs
compact_recency		42.668	45.174	44.935	116
no_extra_recency	30.000	42.709	45.218	44.965	118
no_extra_recency		42.709	45.217	44.968	118
no_extra_recency	60.000	42.710	45.219	44.967	118
compact_recency	30.000	42.813	45.371	45.098	103
compact_recency	60.000	42.838	45.410	45.128	100
full_recency		42.900	45.440	45.204	96
full_recency	30.000	42.956	45.471	45.266	96
full_recency	60.000	43.031	45.541	45.341	99

## 6. Core scan results (top rows) (1/4)

candidate_key	elo_variant	feature_profile	half_life_days	ridge_alpha	histgb_idx	scan_score	scan_histgb_score	scan_histgb_score	scan_ridge_score	scan_ridge_score	red_std_gap	scan_splits_used	scan_stage	outer_fold
extra_recency elo_base_stat	extra_recency elo_base_stat	extra_recency elo_base_stat	60.000	16.000	2	37.138	39.567	31.126	39.985	32.325	-27.637	1	coarse	1
extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.138	39.567	31.126	39.985	32.325	-27.637	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.138	39.567	31.126	39.985	32.325	-27.637	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.138	39.567	31.126	39.985	32.325	-27.637	1	coarse	1
extra_recency elo_base_stat	extra_recency elo_base_stat	extra_recency elo_base_stat	30.000	16.000	2	37.146	39.579	31.133	39.978	32.313	-27.704	1	coarse	1
extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.146	39.579	31.133	39.978	32.313	-27.704	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.146	39.579	31.133	39.978	32.313	-27.704	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.146	39.579	31.133	39.978	32.313	-27.704	1	coarse	1
compact_recency elo_base_stat	compact_recency elo_base_stat	compact_recency elo_base_stat	60.000	16.000	2	37.172	39.567	31.126	40.081	32.418	-27.637	1	coarse	1
compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.172	39.567	31.126	40.081	32.418	-27.637	1	coarse	1
compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.172	39.567	31.126	40.081	32.418	-27.637	1	coarse	1
no_compact_recency elo_dynamic elo_base_stat	no_compact_recency elo_dynamic elo_base_stat	no_compact_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.172	39.567	31.126	40.081	32.418	-27.637	1	coarse	1
extra_recency elo_base_stat	extra_recency elo_base_stat	extra_recency elo_base_stat		16.000	2	37.176	39.626	31.180	39.992	32.337	-27.698	1	coarse	1
extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat	extra_recency elo_dynamic elo_base_stat		16.000	2	37.176	39.626	31.180	39.992	32.337	-27.698	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat		16.000	2	37.176	39.626	31.180	39.992	32.337	-27.698	1	coarse	1
no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat	no_extra_recency elo_dynamic elo_base_stat		16.000	2	37.176	39.626	31.180	39.992	32.337	-27.698	1	coarse	1
compact_recency elo_base_stat	compact_recency elo_base_stat	compact_recency elo_base_stat	30.000	16.000	2	37.180	39.579	31.133	40.075	32.407	-27.704	1	coarse	1
compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.180	39.579	31.133	40.075	32.407	-27.704	1	coarse	1
compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	compact_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.180	39.579	31.133	40.075	32.407	-27.704	1	coarse	1
no_compact_recency elo_dynamic elo_base_stat	no_compact_recency elo_dynamic elo_base_stat	no_compact_recency elo_dynamic elo_base_stat	30.000	16.000	2	37.180	39.579	31.133	40.075	32.407	-27.704	1	coarse	1
full_recency elo_base_stat	full_recency elo_base_stat	full_recency elo_base_stat	60.000	16.000	2	37.208	39.567	31.126	40.185	32.505	-27.637	1	coarse	1
full_recency elo_dynamic elo_base_stat	full_recency elo_dynamic elo_base_stat	full_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.208	39.567	31.126	40.185	32.505	-27.637	1	coarse	1
no_full_recency elo_dynamic elo_base_stat	no_full_recency elo_dynamic elo_base_stat	no_full_recency elo_dynamic elo_base_stat	60.000	16.000	2	37.208	39.567	31.126	40.185	32.505	-27.637	1	coarse	1
compact_recency elo_base_stat	compact_recency elo_base_stat	compact_recency elo_base_stat		16.000	2	37.210	39.626	31.180	40.087	32.429	-27.698	1	coarse	1

## 6. Core scan results (top rows) (2/4)

candidate_key	elo_variant	feature_profile	half_life_days	ridge_alpha	histgb_idx	scan_score	scan_histgb_score	scan_histgb_std	scan_ridge_rms	scan_ridge_max	red_std_gap	scan_splits_used	scan_stage	outer_fold
act_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency		16.000	2	37.210	39.626	31.180	40.087	32.429	-27.698	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency		16.000	2	37.210	39.626	31.180	40.087	32.429	-27.698	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency		16.000	2	37.210	39.626	31.180	40.087	32.429	-27.698	1	coarse	1
full_recency hl=elo_base_stat full_recency	elo_base_stat	full_recency	30.000	16.000	2	37.216	39.579	31.133	40.179	32.494	-27.704	1	coarse	1
full_recency hl=elo_dynamic full_recency	elo_dynamic	full_recency	30.000	16.000	2	37.216	39.579	31.133	40.179	32.494	-27.704	1	coarse	1
na_full_recency hl=elo_dynamic full_recency	elo_dynamic	full_recency	30.000	16.000	2	37.216	39.579	31.133	40.179	32.494	-27.704	1	coarse	1
full_recency hl=elo_base_stat full_recency	elo_base_stat	full_recency		16.000	2	37.246	39.626	31.180	40.191	32.515	-27.698	1	coarse	1
full_recency hl=elo_dynamic full_recency	elo_dynamic	full_recency		16.000	2	37.246	39.626	31.180	40.191	32.515	-27.698	1	coarse	1
full_recency hl=elo_dynamic full_recency	elo_dynamic	full_recency		16.000	2	37.246	39.626	31.180	40.191	32.515	-27.698	1	coarse	1
extra_recency hl=elo_base_stat extra_recency	elo_base_stat	extra_recency	60.000	4.000	2	37.573	39.567	31.126	41.227	33.554	-27.637	1	coarse	1
extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	60.000	4.000	2	37.573	39.567	31.126	41.227	33.554	-27.637	1	coarse	1
no_extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	60.000	4.000	2	37.573	39.567	31.126	41.227	33.554	-27.637	1	coarse	1
no_extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	60.000	4.000	2	37.573	39.567	31.126	41.227	33.554	-27.637	1	coarse	1
extra_recency hl=elo_base_stat extra_recency	elo_base_stat	extra_recency	30.000	4.000	2	37.580	39.579	31.133	41.219	33.545	-27.704	1	coarse	1
extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	30.000	4.000	2	37.580	39.579	31.133	41.219	33.545	-27.704	1	coarse	1
no_extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	30.000	4.000	2	37.580	39.579	31.133	41.219	33.545	-27.704	1	coarse	1
no_extra_recency hl=elo_dynamic extra_recency	elo_dynamic	extra_recency	30.000	4.000	2	37.580	39.579	31.133	41.219	33.545	-27.704	1	coarse	1
compact_recency hl=elo_base_stat compact_recency	elo_base_stat	compact_recency	60.000	4.000	2	37.587	39.567	31.126	41.268	33.582	-27.637	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency	60.000	4.000	2	37.587	39.567	31.126	41.268	33.582	-27.637	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency	60.000	4.000	2	37.587	39.567	31.126	41.268	33.582	-27.637	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency	60.000	4.000	2	37.587	39.567	31.126	41.268	33.582	-27.637	1	coarse	1
compact_recency hl=elo_base_stat compact_recency	elo_base_stat	compact_recency	30.000	4.000	2	37.595	39.579	31.133	41.260	33.574	-27.704	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency	30.000	4.000	2	37.595	39.579	31.133	41.260	33.574	-27.704	1	coarse	1
compact_recency hl=elo_dynamic compact_recency	elo_dynamic	compact_recency	30.000	4.000	2	37.595	39.579	31.133	41.260	33.574	-27.704	1	coarse	1

## 6. Core scan results (top rows) (3/4)

candidate_key	elo_variant	feature_profile	half_life_days	ridge_alpha	histgb_idx	scan_score	scan_histgb_score	scan_histgb_score	scan_ridge_score	scan_ridge_score	mean_std_gap	scan_splits_used	scan_stage	outer_fold
compact_dynamic_k_tea	compact_recer		30.000	4.000	2	37.595	39.579	31.133	41.260	33.574	-27.704	1	coarse	1
full_recer hl=elo_base_stat	full_recer		60.000	4.000	2	37.604	39.567	31.126	41.317	33.597	-27.637	1	coarse	1
full_recer hl=elo_dynamic_full_recer			60.000	4.000	2	37.604	39.567	31.126	41.317	33.597	-27.637	1	coarse	1
full_recer hl=elo_dynamic_k_tea	full_recer		60.000	4.000	2	37.604	39.567	31.126	41.317	33.597	-27.637	1	coarse	1
extra_recer hl=elo_base_stat	extra_recer			4.000	2	37.612	39.626	31.180	41.236	33.563	-27.698	1	coarse	1
extra_recer hl=elo_dynamic_full_recer				4.000	2	37.612	39.626	31.180	41.236	33.563	-27.698	1	coarse	1
extra_recer hl=elo_dynamic_k_tea	extra_recer			4.000	2	37.612	39.626	31.180	41.236	33.563	-27.698	1	coarse	1
no_extra_recer hl=elo_dynamic_k_tea	extra_recer			4.000	2	37.612	39.626	31.180	41.236	33.563	-27.698	1	coarse	1
full_recer hl=elo_base_stat	full_recer		30.000	4.000	2	37.612	39.579	31.133	41.310	33.590	-27.704	1	coarse	1
full_recer hl=elo_dynamic_full_recer			30.000	4.000	2	37.612	39.579	31.133	41.310	33.590	-27.704	1	coarse	1
full_recer hl=elo_dynamic_k_tea	full_recer		30.000	4.000	2	37.612	39.579	31.133	41.310	33.590	-27.704	1	coarse	1
compact_recer hl=elo_base_stat	compact_recer			4.000	2	37.626	39.626	31.180	41.276	33.591	-27.698	1	coarse	1
compact_recer hl=elo_dynamic_full_recer				4.000	2	37.626	39.626	31.180	41.276	33.591	-27.698	1	coarse	1
compact_recer hl=elo_dynamic_k_tea	compact_recer			4.000	2	37.626	39.626	31.180	41.276	33.591	-27.698	1	coarse	1
compact_recer hl=elo_base_stat	compact_recer			4.000	2	37.626	39.626	31.180	41.276	33.591	-27.698	1	coarse	1
full_recer hl=elo_base_stat	full_recer			4.000	2	37.642	39.626	31.180	41.324	33.605	-27.698	1	coarse	1
full_recer hl=elo_dynamic_full_recer				4.000	2	37.642	39.626	31.180	41.324	33.605	-27.698	1	coarse	1
full_recer hl=elo_dynamic_k_tea	full_recer			4.000	2	37.642	39.626	31.180	41.324	33.605	-27.698	1	coarse	1
extra_recer hl=elo_base_stat	extra_recer			16.000	1	37.830	40.953	32.421	39.992	32.337	-26.179	1	coarse	1
extra_recer hl=elo_dynamic_full_recer				16.000	1	37.830	40.953	32.421	39.992	32.337	-26.179	1	coarse	1
extra_recer hl=elo_dynamic_k_tea	extra_recer			16.000	1	37.830	40.953	32.421	39.992	32.337	-26.179	1	coarse	1
extra_recer hl=elo_base_stat	extra_recer		30.000	16.000	1	37.853	41.002	32.493	39.978	32.313	-26.200	1	coarse	1
extra_recer hl=elo_dynamic_full_recer			30.000	16.000	1	37.853	41.002	32.493	39.978	32.313	-26.200	1	coarse	1

6. Core scan results (top rows) (4/4)

candidate_key	elo_variant	feature_profile	half_life_days	ridge_alpha	histgb_idx	scan_score	scan_histgb_score	scan_histgb_score	scan_ridge_score	scan_ridge_score	compared_std_gap	scan_splits_used	scan_stage	outer_fold
no_extra_recency	dynamic_k_team	extra_recency	30.000	16.000	1	37.853	41.002	32.493	39.978	32.313	-26.200	1	coarse	1
no_extra_recency	dynamic_k_team	extra_recency	30.000	16.000	1	37.853	41.002	32.493	39.978	32.313	-26.200	1	coarse	1
act_recency hilo_base_stat		compact_recency		16.000	1	37.864	40.953	32.421	40.087	32.429	-26.179	1	coarse	1
act_recency hilo_dynamic		compact_recency		16.000	1	37.864	40.953	32.421	40.087	32.429	-26.179	1	coarse	1
compact_recency dynamic_k_team		compact_recency		16.000	1	37.864	40.953	32.421	40.087	32.429	-26.179	1	coarse	1
compact_recency dynamic_k_team		compact_recency		16.000	1	37.864	40.953	32.421	40.087	32.429	-26.179	1	coarse	1
extra_recency hilo_base_stat		extra_recency	60.000	16.000	1	37.882	41.057	32.513	39.985	32.325	-26.122	1	coarse	1
extra_recency hilo_dynamic		extra_recency	60.000	16.000	1	37.882	41.057	32.513	39.985	32.325	-26.122	1	coarse	1

7. Elo variant comparisons (1/1)

elo_variant	scan_score_mean	scan_histgb_rmse_mean	scan_ridge_rmse_mean	n_configs
elo_base_static	42.617	45.079	44.868	298
elo_dynamic_k_teamha	42.803	45.377	45.093	222
elo_dynamic_k	42.814	45.387	45.105	223
elo_dynamic_k_teamha_decay	43.052	45.561	45.351	221

7. HistGB hyperparameter sweep summary (1/1)

histgb_idx	scan_score_mean	scan_histgb_rmse_mean	n	params
2	42.706	45.013	37	{'learning_rate': 0.0337, 'max_depth': 4, 'max_leaf_nodes': 63, 'min_samples_leaf': 6}
1	42.626	45.084	10	{'learning_rate': 0.0531, 'max_depth': 4, 'max_leaf_nodes': 31, 'min_samples_leaf': 8}
0	43.089	45.911	10	{'learning_rate': 0.0430, 'max_depth': 3, 'max_leaf_nodes': 31, 'min_samples_leaf': 10}

8. Regime-specific stacking / calibration results (1/1)

use_regime_stack	calibration_mode	scale_mode	outer_rmse_mean	outer_mae_mean	outer_dispersion_ratio_mean	regime_gain_vs_pooled_mae	n
False	pooled	pooled	41.206	32.791	0.000	0.000	6
True	regime	pooled	45.161	35.778	0.514	1.975	3
True	regime	regime	45.879	36.321	0.112	0.383	3



9. Dispersion and scale correction analysis (1/1)

outer_fold	stage	pred_std	actual_std	rmse
1	raw_mean	0.000	43.833	44.393
1	calibrated	32.182	43.833	49.012
1	scaled	22.528	43.833	45.161
1	final	22.528	43.833	45.161
2	raw_mean	0.000	42.297	42.400
2	calibrated	0.000	42.297	43.045
2	scaled	0.000	42.297	43.045
2	final	0.000	42.297	43.045
3	raw_mean	0.000	39.366	39.370
3	calibrated	0.000	39.366	39.438
3	scaled	0.000	39.366	39.367
3	final	0.000	39.366	39.367
4	raw_mean	0.000	45.945	45.954
4	calibrated	7.358	45.945	46.567
4	scaled	5.150	45.945	45.879
4	final	5.150	45.945	45.879

10. Shift diagnostics: Train recent vs Derby (1/1)

feature	train_recent_mean	derby_mean	train_recent_std	derby_std	ks_stat	ks_pvalue	psi_derby	outside_train_1_9	ood_flag
games_played_min	9.122	11.067	1.095	0.660	0.730	0.000	8.252	0.227	True
massey_diff	5.870	2.613	9.415	7.140	0.223	0.006	0.506	0.000	True
elo_diff_pre	60.795	-1.682	116.795	106.947	0.283	0.000	0.348	0.053	True
offdef_margin_neutral	0.581	-0.123	12.300	9.369	0.121	0.338	0.219	0.000	True
elo_neutral_diff_pre	5.795	-1.682	116.795	106.947	0.119	0.359	0.211	0.000	True
same_conf_flag	0.697	0.013	0.459	0.115	0.684	0.000	0.000	0.000	True
offdef_net_diff	0.534	-0.146	8.951	13.180	0.104	0.526	0.169	0.147	False
margin_last5_vs_season	-1.434	1.554	16.927	19.606	0.105	0.514	0.133	0.067	False
consistency_ratio_derby	-0.016	0.050	0.329	0.347	0.122	0.335	0.061	0.013	False
volatility_sum	79.972	80.179	8.181	7.712	0.079	0.831	0.060	0.053	False

## 10. Shift Mitigation Notes

10. Shift mitigation rules

- Detected 6 OOD-like feature shifts (KS/PSI/out-of-range checks).
- Applied softer scaling (blend toward 1.0) and wider winsor bounds under OOD risk.

8. Regime weights (outer\_fold\_1\_selected) (1/1)

level	regime	n	w_pred_ridge	w_pred_huber	w_pred_histgb	w_pred_histgb_bag
pooled	ALL	165	0.250	0.250	0.250	0.250
full	moderate cross lowinfo highv	76	0.250	0.250	0.250	0.250
simple	moderate cross lowinfo	82	0.250	0.250	0.250	0.250
gap	moderate	114	0.250	0.250	0.250	0.250
gap	close	31	0.250	0.250	0.250	0.250

8. Regime weights (outer\_fold\_2\_selected) (1/1)

level	regime	n	w_pred_ridge	w_pred_huber	w_pred_histgb	w_pred_histgb_bag
pooled	ALL	330	0.250	0.250	0.250	0.250

8. Regime weights (outer\_fold\_3\_selected) (1/1)

level	regime	n	w_pred_ridge	w_pred_huber	w_pred_histgb	w_pred_histgb_bag
pooled	ALL	479	0.250	0.250	0.250	0.250

8. Regime weights (outer\_fold\_4\_selected) (1/1)

level	regime	n	w_pred_ridge	w_pred_huber	w_pred_histgb	w_pred_histgb_bag
pooled	ALL	579	0.250	0.250	0.250	0.250
full	moderate same lowinfo highv	48	0.250	0.250	0.250	0.250
simple	mismatch same lowinfo	71	0.250	0.250	0.250	0.250
simple	moderate same lowinfo	78	0.250	0.250	0.250	0.250
simple	close same lowinfo	68	0.250	0.250	0.250	0.250
simple	moderate same midinfo	64	0.250	0.250	0.250	0.250
simple	close same midinfo	47	0.250	0.250	0.250	0.250
simple	mismatch same midinfo	68	0.250	0.250	0.250	0.250
gap	mismatch	213	0.250	0.250	0.250	0.250
gap	moderate	196	0.250	0.250	0.250	0.250
gap	close	170	0.250	0.250	0.250	0.250

8. Calibration params (outer\_fold\_1\_selected) (1/1)

level	regime	n	intercept	slope
pooled	ALL	165	24.281	0.300
full	moderate cross lowinfo highvol	76	46.929	0.300
simple	moderate cross lowinfo	82	44.233	0.300
simple	moderate same lowinfo	32	13.495	0.300
gap	moderate	114	31.564	0.300
gap	close	31	-60.514	2.200



8. Calibration params (outer\_fold\_2\_selected) (1/1)

level	regime	n	intercept	slope
pooled	ALL	330	-2.391	0.446

8. Calibration params (outer\_fold\_3\_selected) (1/1)

level	regime	n	intercept	slope
pooled	ALL	479	5.904	0.300

8. Calibration params (outer\_fold\_4\_selected) (1/1)

level	regime	n	intercept	slope
pooled	ALL	579	15.389	0.300
full	moderate same lowinfo highvol	48	17.440	0.300
full	close same lowinfo lowvol	38	-44.919	2.200
full	mismatch same lowinfo lowvol	38	59.503	0.300
full	mismatch same midinfo lowvol	37	-92.569	2.200
simple	mismatch same lowinfo	71	38.118	0.300
simple	moderate same lowinfo	78	30.123	0.300
simple	close same lowinfo	68	-26.803	2.017
simple	moderate same midinfo	64	12.338	0.300
simple	close same midinfo	47	-37.337	2.200
simple	mismatch same midinfo	68	-53.074	2.200
gap	mismatch	213	12.160	1.090
gap	moderate	196	16.036	0.300
gap	close	170	0.280	0.300

9. Scale params (outer\_fold\_1\_selected) (1/1)

level	regime	n	scale
pooled	ALL	165	0.700

9. Scale params (outer\_fold\_2\_selected) (1/1)

level	regime	n	scale
pooled	ALL	330	1.000

9. Scale params (outer\_fold\_3\_selected) (1/1)

level	regime	n	scale
pooled	ALL	479	0.700

9. Scale params (outer\_fold\_4\_selected) (1/1)

level	regime	n	scale
pooled	ALL	579	0.700
full	moderate same lowinfo highvol	48	0.700
full	close same lowinfo lowvol	38	0.700
full	mismatch same lowinfo lowvol	38	0.700
full	mismatch same midinfo lowvol	37	0.700
simple	mismatch same lowinfo	71	0.700
simple	moderate same lowinfo	78	0.700
simple	close same lowinfo	68	0.700
simple	moderate same midinfo	64	0.700
simple	close same midinfo	47	0.700
simple	mismatch same midinfo	68	0.700
gap	mismatch	213	0.700
gap	moderate	196	0.700
gap	close	170	0.700

## 11. Residual Diagnostics (Summary)

### 11. Residual diagnostics

- OOF final RMSE=43.557, MAE=34.523, Bias=-1.170
- Residual std=43.541, skew=-0.006, kurtosis=-0.104
- Mean-stack RMSE=43.180
- Quantile median (q50) RMSE=43.180
- Mean+median blend RMSE=43.180
- Calibrated RMSE=44.864
- Tail error threshold (top 10%)=70.832



## 11. Tail error diagnostics (top 10% abs-error) (1/2)

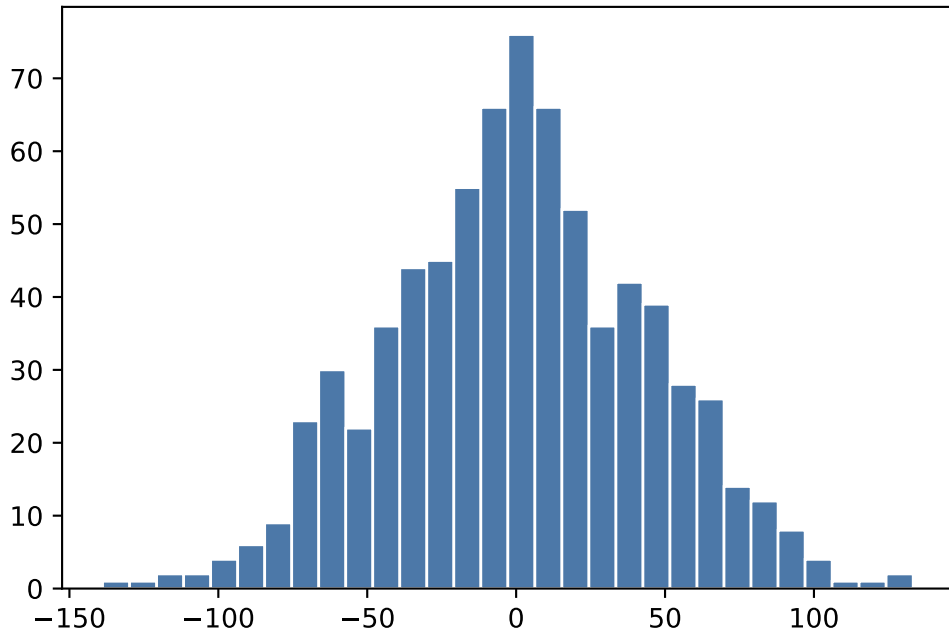
[illegible]

## 11. Tail error diagnostics (top 10% abs-error) (2/2)

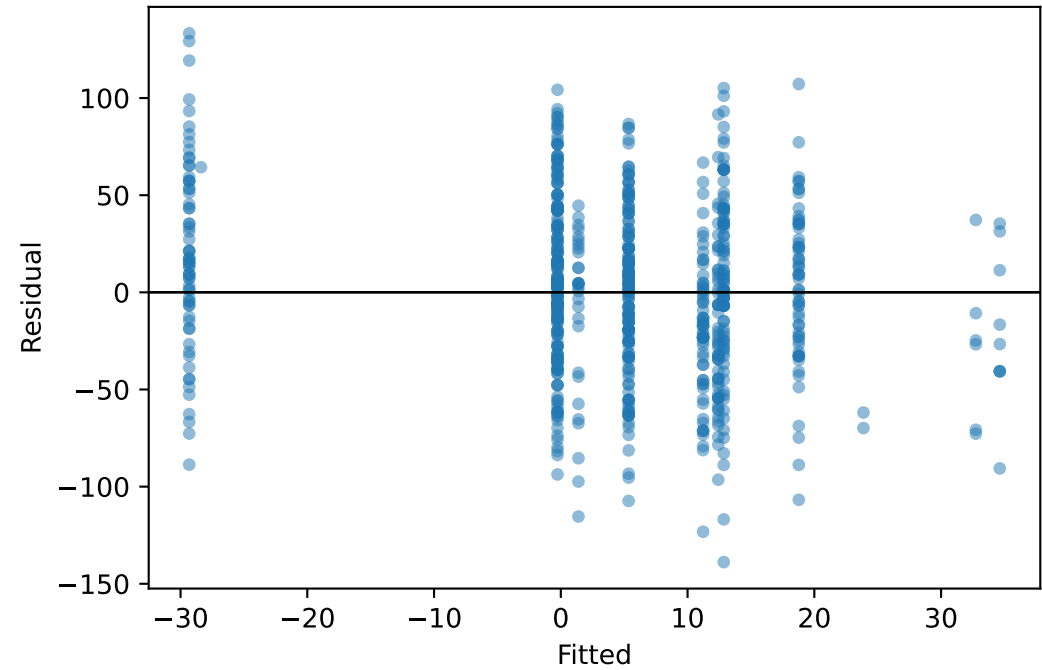
[illegible]

## 11. Residual diagnostics plots

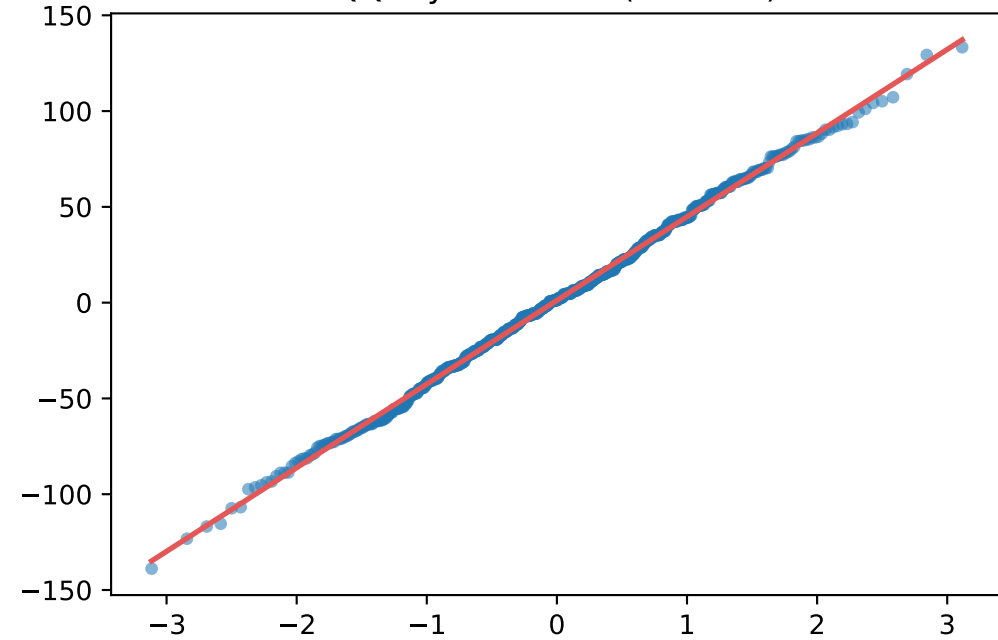
Residual histogram



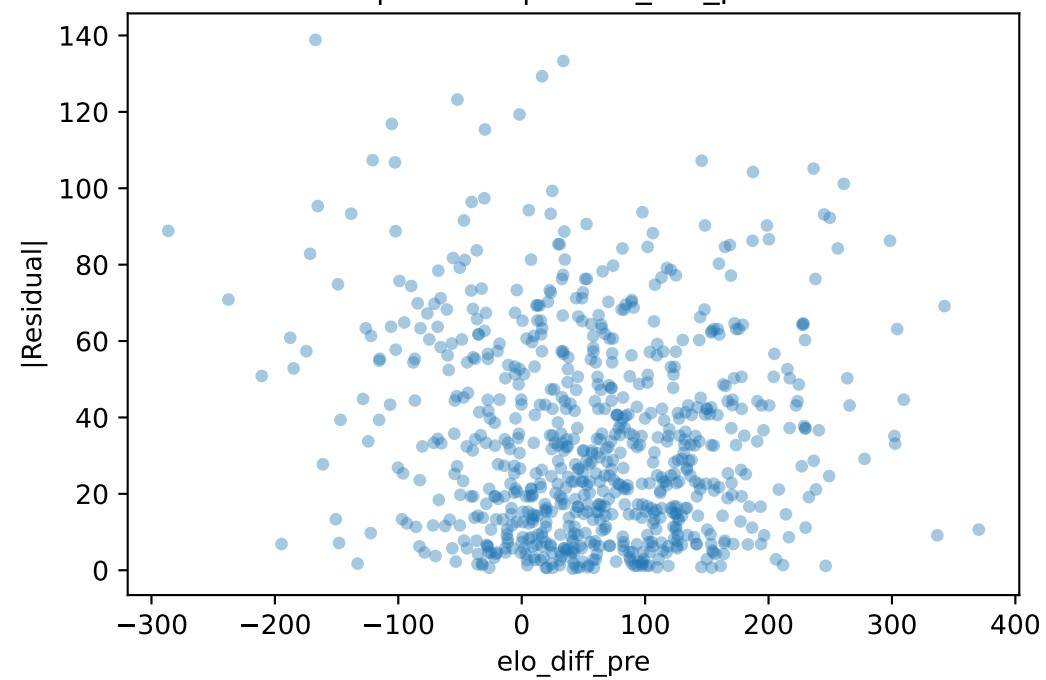
Residual vs fitted



QQ-style residual ( $r=0.999$ )



$|\text{Residual}|$  vs elo\_diff\_pre



## 12-14. Final Decisions / Limitations / Next Steps

### 12. Final derby prediction distribution

- count=75, mean=8.566, std=0.000
- min=8.566, max=8.566
- quantiles 1/5/25/50/75/95/99 = 8.566, 8.566, 8.566, 8.566, 8.566, 8.566, 8.566

### 13. Final decisions and rationale

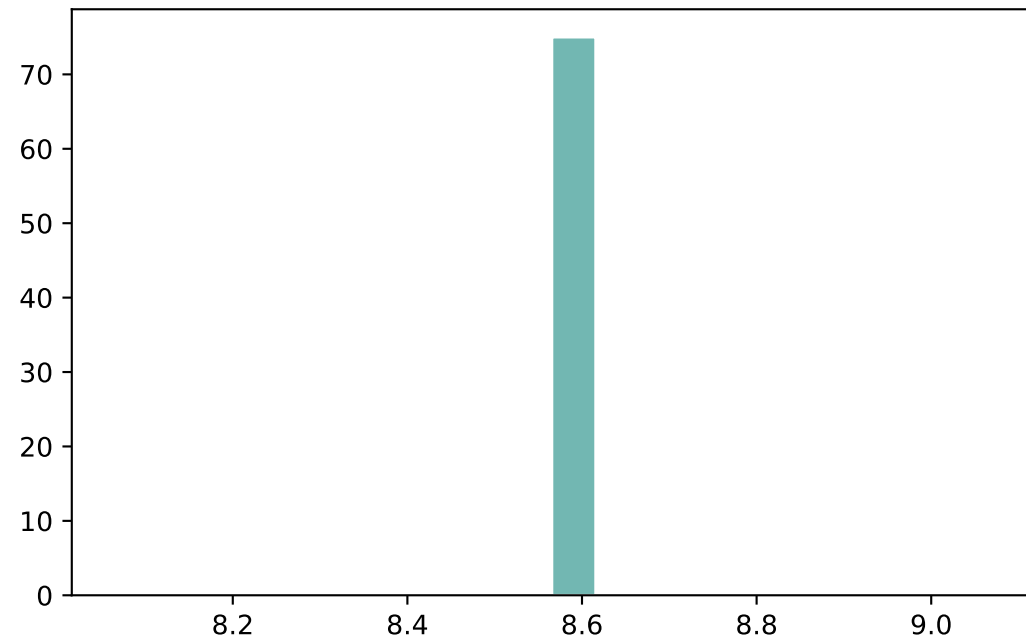
- Final core candidate: elo\_base\_static|no\_extra\_recency|hl=none|ridge=4|hgb=0
- Final postprocess candidate: stack=pool|cal=pooled|scale=pooled|q50=0.00|win=0.000
- Recency profile=no\_extra\_recency | half\_life=None
- Shift mitigation applied=True
- Dispersion guard applied=False

### 14. Limitations + next steps

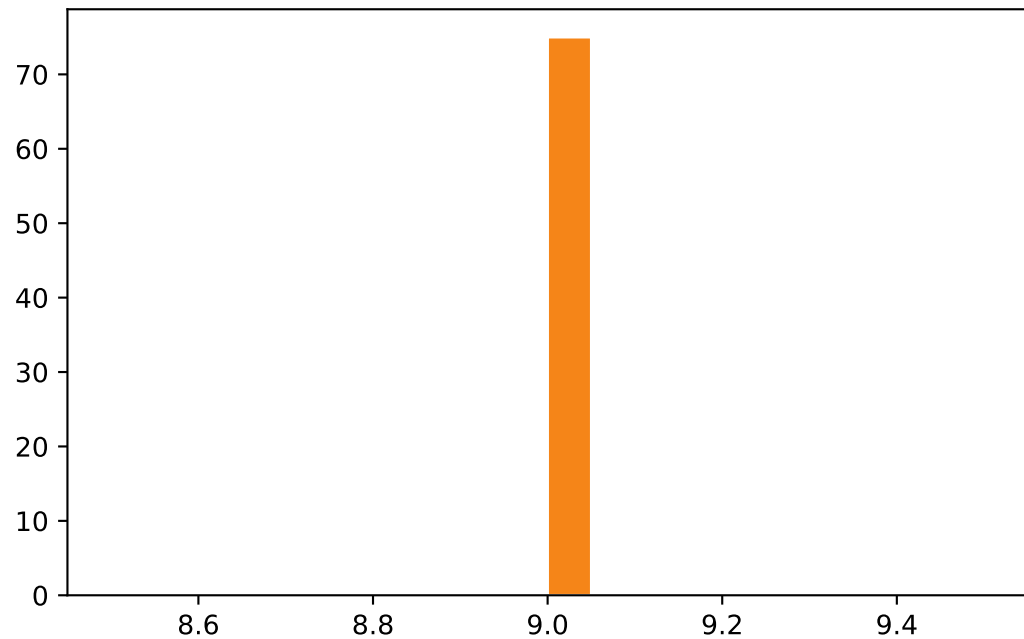
- Derby labels hidden: final selection relies on nested train-only proxies.
- No injuries/lineups/travel features available.
- Future work: monotone piecewise regime calibration and richer Bayesian uncertainty.

## 12. Final derby prediction distribution

Float predictions



Rounded predictions



Appendix A. First 10 rows of predictions.csv (1/1)

GamelD	Date	Team1_Conf	Team1_ID	Team1	Team2_Conf	Team2_ID	Team2	Team1_WinMargin
1941	7/4/2025	Purple	68	Idaho	Yellow	119	Houston	9
1942	7/4/2025	White	105	Helena	Yellow	120	Iowa	9
1943	7/4/2025	Red	82	Jackson	Purple	67	Concord	9
1944	7/4/2025	White	108	Montpelier	Yellow	129	West Virginia	9
1945	7/4/2025	Yellow	121	Lansing	White	113	South Dakota	9
1946	7/4/2025	Crimson	21	Lincoln	Yellow	118	Florida	9
1947	7/4/2025	Yellow	117	Columbia	Crimson	23	Nashville	9
1948	7/4/2025	Orange	63	New Jersey	Yellow	126	Pierre	9
1949	7/4/2025	Purple	77	Wilmington	Orange	58	Honolulu	9
1950	7/4/2025	Crimson	22	Michigan	Green	49	Natchez	9

## Appendix B. Top 20 rankings by Rank (1/1)

TeamID	Team	Rank
45	Boise City	1
94	Kentucky	2
2	Baton Rouge	3
40	Pennsylvania	4
26	Phoenix	5
153	Tokyo	6
36	Massachusetts	7
90	Big Sur	8
37	Minnesota	9
137	Milan	10
92	Illinois	11
88	Port St. Lucie	12
116	California	13
149	Seoul	14
84	Mississippi	15
131	Belgrade	16
19	Georgia	17
54	Boise	18
46	Boston	19
22	Michigan	20

# Appendix C. Selected hyperparameters (1/1)

name	value
elo_variant	elo_base_static
feature_profile	no_extra_recency
half_life_days	none
ridge_alpha	4.0
histgb_idx	0
histgb_params	{'learning_rate': 0.04, 'max_depth': 3, 'max_leaf_nodes': 31, 'min_samples_leaf': 10, 'l2_regularization': 0.3, 'max_features': 0.5}
postprocess	stack=pool cal=pooled scale=pooled q50=0.00 win=0.000
winsor_bounds	(-116.61, 113.305)
dispersion_guard_factor	1.0
seed	23
OMP_NUM_THREADS	1
MKL_NUM_THREADS	1



