AlgoSports23 Final Report (Nested Time-Aware CV)

Leakage prevention decisions:
1) Outer evaluation is expanding-window by Date. Validation blocks are strictly in the future of their fold training blocks.
2) Inner tuning is also expanding-window inside each outer-train.
3) Ratings are fit only on split-train before scoring split-val (no full-season ratings reused inside CV).
4) Rating-to-margin mappings (a + b*diff) are fit only on split-train.
5) Ensemble weights, dynamic regime weights, and stacker are tuned only from inner-CV outputs.
6) Stacker uses inner OOF base predictions to avoid in-sample base-pred leakage.
7) Derby labels are never used anywhere (unavailable).

Runtime safety / determinism:
- Threads capped to 1 (OMP/MKL/OPENBLAS/NUMEXPR): 1/1/1/1
- Budgets: total=420.0s, tuning=180.0s, model_fits=250
- Usage: total_elapsed=6.90s, tuning_elapsed=6.90s, fit_count=250, stop_reason=MAX_MODEL_FITS

Outer / Inner CV Scheme

Outer folds: 5 expanding-window by Date (fallback to 4 if needed).
Inner folds (within outer-train): 3 expanding-window splits (fallback to 2 if early outer fold has too few dates).

ASCII diagram:
Outer fold i: [train .....] [validate next block]
Inner split 1: [train] [val]
Inner split 2: [train .....] [val]
Inner split 3: [train .....] [val]

Combination tuning uses meta-expanding splits over inner OOF rows (fold IDs), preserving time order.

Outer CV Fold Table

fold	n_train	n_val	train_end	val_start	val_end
1	125	193	2025-01-12	2025-01-23	2025-02-19
2	318	205	2025-02-19	2025-03-03	2025-04-02
3	523	151	2025-04-02	2025-04-14	2025-04-30
4	674	102	2025-04-30	2025-05-13	2025-05-29
5	776	164	2025-05-29	2025-05-30	2025-06-30

Rating Systems Implemented

System	Prediction signal	Hyperparams
Elo	pregame Elo diff (+ fixed home adv)	fixed HA=50, K=24, MOV mult on
Elo k-decay	pregame Elo diff with season-indexed K multiplier	decay type / A / G or tau (small grid)
Massey ridge	ridge team-rating margin model	ridge alpha grid
Colley	W/L-only rating diff, then affine map to margin	none + fold-train affine map
Off/Def net ridge	offense/defense point model -> margin	fixed alpha

Single-System Performance (Outer Folds)

name	folds	rmse_mean	rmse_std	mae_mean	mae_std	bias_mean	pred_std_mean	actual_std_mean	dispersion_ratio_mean
p_massey	5	35.997	3.592	28.688	2.553	1.021	30.038	43.112	0.701
p_offdef	5	36.282	3.832	28.905	2.637	-0.228	29.126	43.112	0.680
p_elok	5	37.254	4.209	29.790	3.246	-2.016	26.646	43.112	0.621
p_elo	5	37.659	4.522	30.002	3.545	-2.057	28.790	43.112	0.671
p_colley	5	38.945	4.372	30.925	3.535	1.063	30.008	43.112	0.700

Combination Performance (Outer Folds)

name	folds	rmse_mean	rmse_std	mae_mean	mae_std	bias_mean	pred_std_mean	actual_std_mean	dispersion_ratio_mean
single	5	35.834	3.916	28.349	2.828	0.971	25.280	43.112	0.588
static_simplex	5	36.117	4.239	28.612	3.076	0.082	24.741	43.112	0.576
dynamic_regime	5	36.206	4.411	28.700	3.282	0.181	24.829	43.112	0.578
ridge_stacker	5	36.790	4.515	29.240	3.358	-4.424	22.094	43.112	0.518

Per-Outer-Fold Selected Model + Metrics

fold	rmse	mae	bias	pred_std	actual_std	dispersion_ratio	n	chosen_strategy_t	chosen_label
1	38.510	29.894	-0.115	26.087	45.098	0.578	193	single_	single_massey::scale0
2	39.557	31.068	3.435	32.051	44.241	0.724	205	single_	single_massey::scale0
3	32.955	27.009	1.118	22.281	38.591	0.577	151	single_	single_massey::scale1
4	30.479	24.046	-0.827	23.711	40.496	0.586	102	single_	single_massey::scale1
5	37.669	29.727	1.241	22.271	47.133	0.473	164	single_	single_massey::scale1

Outer Fold 1 Inner Family Tuning (Top)

family	label	rmse	mae	bias	pred_std	actual_std	dispersion_ratio	n	inner_folds_used
massey	massey_alpha10.0	44.484	35.786	2.576	14.909	46.171	0.323	103	3
massey	massey_alpha30.0	44.515	35.945	3.753	14.655	46.171	0.317	103	3
massey	massey_alpha100.0	44.555	36.068	5.787	14.824	46.171	0.321	103	3
offdef	offdef_alpha20.0	44.813	35.585	-6.172	14.680	46.171	0.318	103	3
colley	colley	45.653	36.809	8.036	18.167	46.171	0.393	103	3
elo	elo_k20.5_G100_tau50	49.547	40.706	-1.256	37.229	46.171	0.806	103	3
elo	elo_k21.0_G100_tau75	49.835	40.858	-1.288	37.599	46.171	0.814	103	3
elo	elo_k20.25_G50_tau50	49.931	40.962	-1.121	38.113	46.171	0.825	103	3
elo	elo_k20.5_G100_tau50	50.007	40.990	-1.168	38.143	46.171	0.826	103	3
elo	elo_base	50.570	41.380	-0.805	39.657	46.171	0.859	103	3

Outer Fold 1 Inner Combo Tuning (Top)

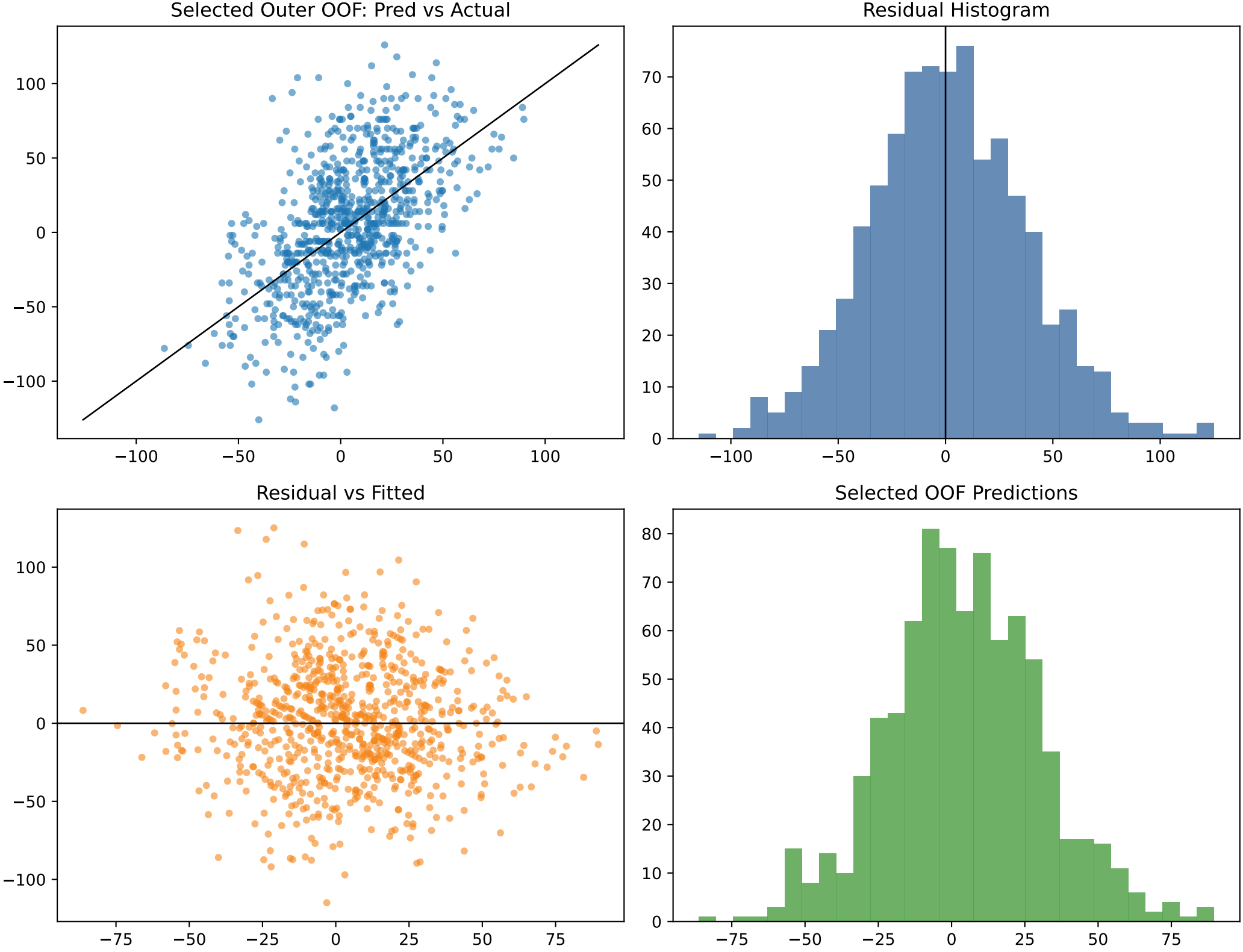
label	strategy_typ	scale_on	family_col	alpha	rmse	mae	bias	pred_std	actual_std	dispersion_rat	n	n_meta_splits
assey::scale0	single	False	p_massey	nan	45.931	36.916	-5.162	15.601	46.481	0.336	61	2
p_elo::scale1	single	True	p_elo	nan	46.624	38.055	-9.603	17.457	46.481	0.376	61	2
olley::scale0	single	False	p_colley	nan	46.628	37.226	2.225	21.523	46.481	0.463	61	2
mplex::scale0	static_simplex	False		nan	46.641	37.451	-5.804	15.796	46.481	0.340	61	2
gime::scale0	dynamic_regime	False		nan	46.715	37.551	-5.680	16.130	46.481	0.347	61	2
_elok::scale1	single	True	p_elok	nan	46.745	38.182	-10.077	16.570	46.481	0.356	61	2
olley::scale1	single	True	p_colley	nan	47.400	37.343	-0.619	18.862	46.481	0.406	61	2
ffdef::scale0	single	False	p_offdef	nan	47.855	38.772	-10.389	14.891	46.481	0.320	61	2
assey::scale1	single	True	p_massey	nan	47.959	38.806	-5.322	18.803	46.481	0.405	61	2
ffdef::scale1	single	True	p_offdef	nan	48.815	40.091	-10.152	20.532	46.481	0.442	61	2
gime::scale1	dynamic_regime	True		nan	48.968	40.316	-3.526	24.447	46.481	0.526	61	2
mplex::scale1	static_simplex	True		nan	49.188	40.588	-3.773	24.409	46.481	0.525	61	2
_elok::scale0	single	False	p_elok	nan	50.948	42.253	-9.107	33.650	46.481	0.724	61	2
p_elo::scale0	single	False	p_elo	nan	51.377	42.525	-8.158	35.453	46.481	0.763	61	2

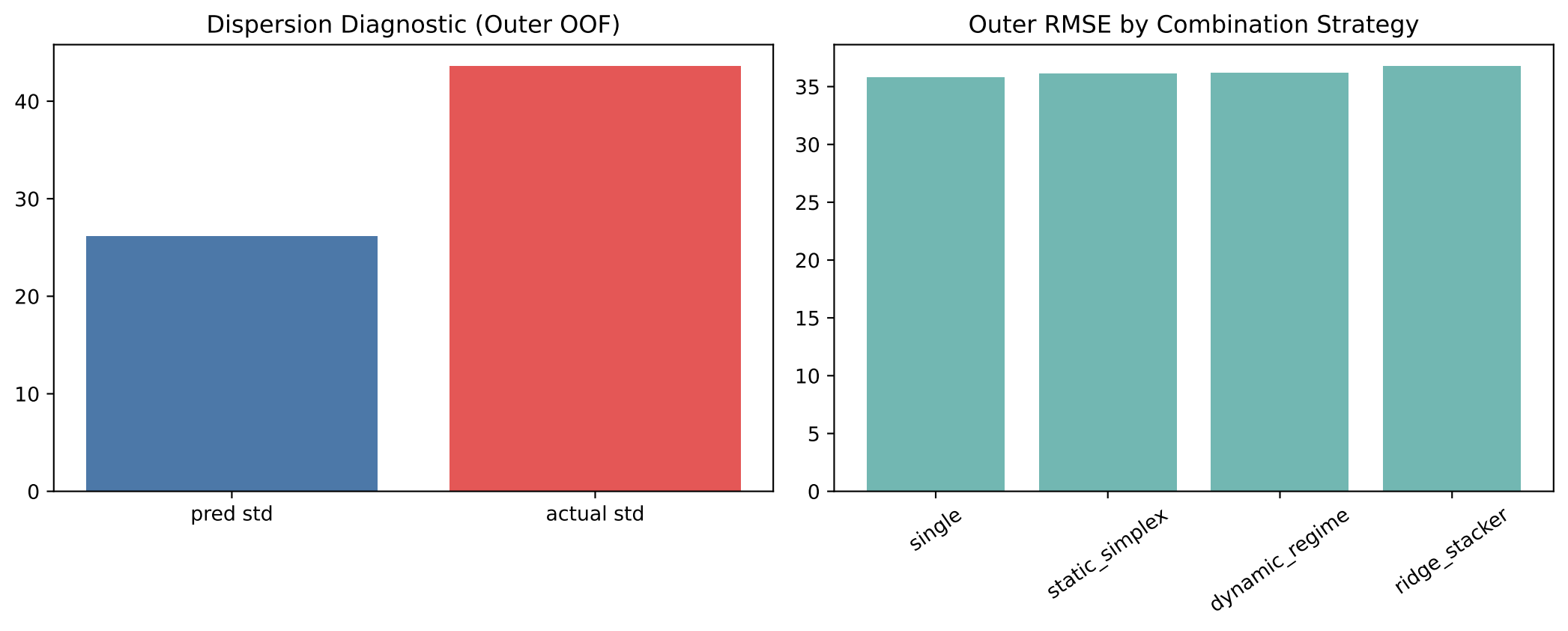
Outer Fold 2 Inner Family Tuning (Top)

family	label	rmse	mae	bias	pred_std	actual_std	dispersion_ratio	n	inner_folds_used
massey	massey_alpha10.0	40.709	32.122	0.639	24.563	45.245	0.543	284	3
massey	massey_alpha100.0	40.750	32.242	2.069	22.679	45.245	0.501	284	3
massey	massey_alpha30.0	40.787	32.226	0.999	23.260	45.245	0.514	284	3
offdef	offdef_alpha20.0	42.111	32.868	-4.541	24.225	45.245	0.535	284	3
colley	colley	42.590	34.041	1.207	24.436	45.245	0.540	284	3
_exponential_k2	elo_k32.0_G100_tau75	44.384	35.366	-4.597	29.823	45.245	0.659	284	3
elok_2_linear_k2	elo_k20.5_G100_tau50	44.407	35.463	-4.454	30.085	45.245	0.665	284	3
_exponential_k2	elo_k20.5_G100_tau50	44.661	35.666	-4.487	30.750	45.245	0.680	284	3
elok_1_linear_k2	elo_k20.25_G50_tau50	44.911	35.917	-4.371	31.400	45.245	0.694	284	3
elo	elo_base	45.394	36.343	-4.418	32.540	45.245	0.719	284	3

Outer Fold 2 Inner Combo Tuning (Top)

	label	strategy_typ	scale_on	family_col	alpha	rmse	mae	bias	pred_std	actual_std	dispersion_rat	n	n_meta_splits
assey::scale0		single	False	p_massey	nan	38.383	30.043	1.303	27.585	45.098	0.612	193	2
ffdef::scale0		single	False	p_offdef	nan	39.328	30.568	-1.733	26.494	45.098	0.587	193	2
assey::scale1		single	True	p_massey	nan	40.039	31.927	0.893	24.561	45.098	0.545	193	2
gime::scale0		dynamic_regime	False		nan	40.099	31.641	0.647	25.984	45.098	0.576	193	2
nplex::scale0		static_simplex	False		nan	40.252	31.954	0.037	26.437	45.098	0.586	193	2
ffdef::scale1		single	True	p_offdef	nan	40.323	31.606	-1.664	19.500	45.098	0.432	193	2
gime::scale1		dynamic_regime	True		nan	40.961	32.796	1.113	26.983	45.098	0.598	193	2
_elok::scale1		single	True	p_elok	nan	40.967	32.202	-3.001	14.766	45.098	0.327	193	2
p_elo::scale1		single	True	p_elo	nan	41.065	32.332	-2.888	16.306	45.098	0.362	193	2
nplex::scale1		static_simplex	True		nan	41.110	32.804	0.288	24.740	45.098	0.549	193	2
_elok::scale0		single	False	p_elok	nan	41.173	32.848	-3.669	27.414	45.098	0.608	193	2
olley::scale1		single	True	p_colley	nan	41.515	33.051	1.000	24.252	45.098	0.538	193	2
olley::scale0		single	False	p_colley	nan	41.716	33.051	1.285	27.832	45.098	0.617	193	2
p_elo::scale0		single	False	p_elo	nan	42.539	34.103	-3.456	31.040	45.098	0.688	193	2


Selected Outer OOF: Pred vs Actual
Residual Histogram
Residual vs Fitted
Selected OOF Predictions


Dispersion Diagnostic (Outer OOF)
0
10
20
30
40
pred std
actual std
Outer RMSE by Combination Strategy
0
5
10
15
20
25
30
35
single
static_simplex
dynamic_regime
ridge_stacker

Final Selection, Derby Summary, Validations

Selected final approach (based on outer CV):
- strategy_type: single
- force_scale_policy: True

Final full-train tuning (time-aware, no derby labels):
- combo_choice: {"label": "single::p_elok::scale1", "strategy_type": "single", "scale_on": true, "family_col": "p_elok", "alpha": null, "rmse": 36.26705607310889, "mae": 28.833515960302954, "bias": -0.12243001702471933, "pred_std": 15.403647301977209, "actual_std": 42.599845454527504, "dispersion_ratio": 0.3615892766188454, "n": 417, "n_meta_splits": 2}
- selected families: {"elo": "elo_base", "elok": "elok_2_linear_A0.5_G100_tau50"}

Derby prediction distribution (raw before rounding):
{   "mean": -3.383975180871826,   "std": 18.054486675821867,   "min": -43.39010623826734,   "q05": -36.74341323794989,   "q25": -13.195042201193198,   "median": -2.627404137788435,   "q75": 6.92492647738138, "q95": 28.007210428891362,   "max": 31.865579125533603 }

Output validations:
- predictions.csv rows=75 no_missing=True numeric=True
- rankings.xlsx rows=165 rank_set_valid=True
- final_report.pdf exists=True size_bytes=111012