Walk-Forward Ridge Factor Timing with Meta-Selection

JKP Characteristics Trading Factor (CTF) challenge submission · Lorcan Doherty, Queen's University Belfast · October 2026

Files: `model.py` (exposes `main(chars, features, daily_ret)`), `requirements.txt`, `weights.csv`, this document.

Summary. A long-short US equity strategy built on linear-rank portfolios of the 402 CTF characteristics. A fixed menu of 51 factor-allocation "configurations" is simulated walk-forward every month. The menu holds 36 top-N factor-selection rules plus 15 ridge-regularised mean-variance portfolios over *all* factors. A meta layer picks the configurations to hold from their own out-of-sample records to date. Every decision at month-end t uses only information realised by the end of t. All design choices were made on 1960–1989 data, which the challenge does not score, and the 1990–2023 results were computed once, after the design was locked. 1990–2023 gross Sharpe ratio: **2.16**, against **1.64** for the course baseline from which this model was developed.

1. Data and timing convention

Inputs are the official CTF tables (1,369,324 stock-months, Dec 1951 – Nov 2023). `daily_ret` is not used. The target `ret_exc_lead1m` on the row for month t is the excess return over month $t+1$, so it is **known only at the end of $t+1$**. It is never used as a feature. Convention: *a factor return stored at row t is realised over t+1; a decision at month-end rp may use rows up to the previous month-end (rp − 1), which is realised at the end of rp.*

2. The four stages

Stage 1 — Factor portfolio returns

For each characteristic k and month t, stocks are percentile-ranked with $w_i = \text{rank}_i - 0.5$, and $\text{LS}_{k,t} = \Sigma\, w_i\, \text{ret_exc_lead1m}_{i,t} \,/\, \Sigma|w_i|$. At least 20 stocks are required, and factors need at least 60 valid months. The result is an 864 × 402 matrix. **Guard:** ranks use month-t characteristics only; the forward return enters only row t.

Stage 2 — Walk-forward simulation of 51 configurations (monthly)

- **36 selection configs** (as in the baseline): window {10y, 20y, expanding} × N {5, 10} × selection {Sharpe, ClusterSharpe} × weighting {EW, RankSharpe, long-only shrunk MeanVar}.
- **15 ridge configs (new):** window {10y, 20y, expanding} × shrinkage $a \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. Over every valid factor, $w \propto [(1-a)\Sigma + a\cdot\text{tr}(\Sigma)/n\cdot I]^{-1}\,\mu$, normalised to $\Sigma|w| = 1$. Weights are long-short, so factors with negative premia are held short. This is a regularised tangency portfolio in the spirit of Kozak, Nagel & Santosh (2020).

Each month rp, every configuration estimates from LS rows up to $rp - 1$ only (windows are exactly the last 120 or 240 known rows). It then earns the realised LS returns of its factors for that month, which builds its own out-of-sample (OOS) record. **Guard:** the estimation slice ends at the previous month-end, and factor eligibility is counted inside that slice.

Stage 3 — Meta-selection

Each month, the 3 configurations with the best expanding-window Sharpe on their own OOS returns up to $rp - 1$ are blended equally. Only configurations with at least 120 months of OOS history qualify. Before any does (1953–1963), all are blended equally. **Guard:** the methodology live at any date is chosen only from earlier performance. Of the meta-selected months, ridge configurations make up 99% of choices before 1990. After 1990 the meta layer shifted toward lighter shrinkage (25%), choosing that walk-forward rather than by hand.

Stage 4 — Stock weights (monthly)

$w_{i,t} = (1/3)\,\Sigma_c\,\Sigma_k\, fw_{c,k}\,(\text{rank}_{i,k,t} - 0.5) \,/\, \Sigma_j|\text{rank}_{j,k,t} - 0.5|$, using month-$t$ characteristics only, held over $t+1$. This exactly replicates the factor-level returns evaluated in Stage 3 (V3).

3. Development protocol (no tuning on the scored period)

The CTF scores months from Dec 1989 on. All candidate changes were evaluated with a harness (`06-DevelopPre1990.py`) that is passed LS data only up to Nov 1989 and reports only pre-1990 statistics. Ideas came from the literature and the template's extension points.

Variant (menu)	Configs	Sharpe 1964–89	1964–76	1977–89	Decision
Baseline menu, 6-monthly	36	2.53	1.97	3.25	reference
+ signed |Sharpe| selection (short negative-premium factors)	72	2.67	2.08	3.44	superseded
+ N = 20	54	2.54	1.99	3.26	rejected
+ all-factor ridge	42	3.17	2.52	3.95	kept
+ everything (signed, N=20, ridge)	114	2.30	1.73	3.95	rejected (meta noise)
ridge menu + monthly rebalancing	42	3.35	2.70	4.13	kept
+ 5 shrinkage levels	51	3.35	2.70	4.16	kept
+ volatility targeting	51	3.20	2.52	4.02	rejected
Final (plus exact month-end dates)	**51**	**3.42**	**2.78**	**4.22**	**locked**

"Exact month-end dates": the baseline computed rp − 1 month with a calendar offset that maps 30 June to 30 May, skipping the 31 May row (conservative, never look-ahead). With monthly rebalancing this would affect every 30-day month, so month-end arithmetic is used.

4. Results (computed once, after locking)

Annualised Sharpe	1960–89	1990–2003	2004–13	2014–23	1990–2023
Baseline, gross	2.49	3.01	0.65	0.66	1.64
This model, gross	**2.63**	**3.61**	**1.43**	**0.99**	**2.16**
Baseline, net of 10 bp per side	2.11	2.78	0.29	0.32	1.37
This model, net of 10 bp per side	2.33	3.37	1.13	0.72	1.93
Baseline, net of 25 bp per side	1.52	2.44	–0.25	–0.19	0.97
This model, net of 25 bp per side	1.85	3.01	0.69	0.32	1.60

Net returns use $\text{net}_t = \text{gross}_t - 2 \cdot \text{turnover}_t \cdot \text{cost}$. Average one-way turnover is 0.36 of the gross book per month (baseline: 0.42). Gross leverage $\Sigma|w|$ averages 0.16 because ridge weights net across factors. Sharpe ratios are invariant to that scale. Performance is weaker after 2003, consistent with anomaly decay after publication, but it is positive in every period.

5. Verification

Check	Outcome
V1 Shift (LS one month staler)	Pass: 1990–2023 Sharpe 2.16 → 2.03 (2.63 → 2.56 pre-1990); mild degradation, no collapse.
V2 Truncation (cut at Jan 1990)	Pass: 433 monthly rebalances with identical factor selections, factor weights and meta-choices. All 479,455 stock weights match to the 10-digit CSV precision (max difference 5e-13).
V3 Replication	Pass: stock-level = factor-level return to 2e-18 (same universe).
V4 Sanity	Pass: max period Sharpe 3.61 (< 4); volatility ≈ 3.6% a year at unit gross leverage.
V5 Determinism	Pass: byte-identical `weights.csv` from two runs in two separate Python 3.13 environments.
Validation mode	Pass: 123 months with 46–2,095 stocks per month; valid output with no missing values.

6. Reproducibility

Python 3.13; numpy 2.5.4, pandas 2.2.3, scipy 1.18.1, pyarrow 26.0.0 (pinned). The full run takes about 4 minutes on an 8-core laptop (peak resident memory about 5.5 GB). No GPU, network, file I/O or randomness inside `main()`. `weights.csv` holds all 1,361,641 rows from `main()` (885,698 in `ctff_test` months), with `w` written to 10 significant digits (relative rounding ≤ 5e-10) to stay under 50 MB.

References: Jensen, Kelly & Pedersen (2023), *Journal of Finance*; Kozak, Nagel & Santosh (2020), "Shrinking the Cross-Section", *Journal of Financial Economics*; Brandt, Santa-Clara & Valkanov (2009), *Review of Financial Studies*; Moreira & Muir (2017), "Volatility-Managed Portfolios", *Journal of Finance*.