

MGT 924 Statistical Foundations

Problem Set 3

Jerry Cao, Grace Gu, Jerry Shen, Angel Zhao

Fall 2026

1. Boosted trees with two risk models

This long-short equity strategy combines boosted-tree and ridge forecasts with two estimates of stock-return covariance. The forecasts identify relative opportunities; risk-based allocation balances expected reward against common and stock-specific risk. In the 1990-2023 competition sample, the locally reproduced gross Sharpe ratio is **3.74**, versus 2.94 for a ridge ensemble benchmark. At an assumed 10 basis points (0.10%) per dollar traded, Sharpe is **3.20**. These are historical estimates; official scoring is pending.

1990-2023 results	New strategy	Ridge benchmark
Gross Sharpe	3.74	2.94
Sharpe after 10 bp costs	3.20	2.42
Annual mean / volatility, actual weights	8.0% / 2.1%	12.3% / 4.2%
Maximum drawdown, actual weights	-2.7%	-4.9%

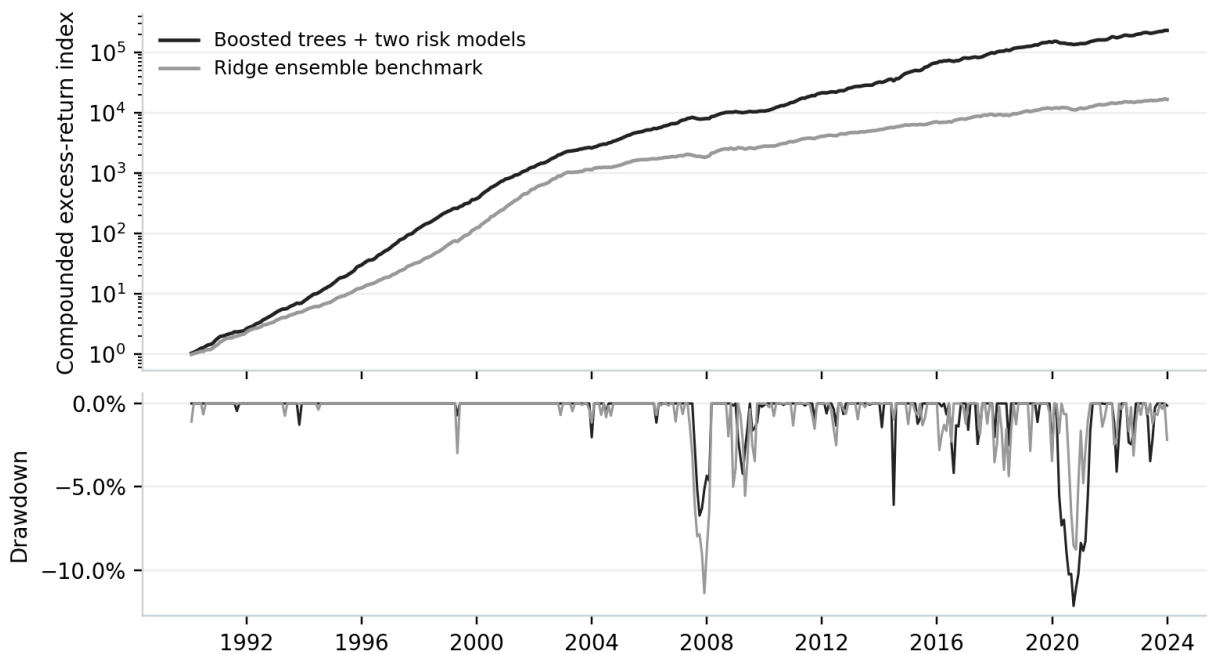


Figure 1. January 1990-December 2023, 408 monthly returns. For this chart only, each gross return series is multiplied by a single constant to give 10% annualized volatility. This makes risk comparable and leaves Sharpe unchanged. The index compounds excess returns from 1; drawdown measures the fall from a prior peak. The table uses actual weights, with different leverage. Annual mean = 12 times monthly mean; volatility = square root of 12 times monthly sample standard deviation; Sharpe = mean / volatility. Source: official CTF data [1] and author calculations.

The strategy selects its forecast, risk-model combination and volatility overlay on the 120 months before the competition sample. The sample itself was already known during this review. Holdings use only information available at formation, but these results do not constitute a fresh holdout test.

2. Data and return forecasts

Use the same information available at each formation date

The three official CTF tables contain 1,369,324 stock-months, 402 characteristics and 103,382,883 daily returns. All listed characteristics enter automatically. Portfolios formed from December 1989 through November 2023 earn returns in January 1990 through December 2023. The supplied return-date field verifies this one-month alignment. Monthly return labels enter training only once their returns have been realized. Daily risk estimates end at the formation date. [1]

Put predictors and the training target on stable scales

Within each month, convert each characteristic to its percentile rank, averaging ties. Store these ranks in 16-bit precision. Trees retain missing values and learn their routing; ridge replaces them by rank 0.5 and subtracts 0.5 from every rank. The target is arctan of the next-month return after subtracting its cross-sectional mean and dividing by its standard deviation. This focuses learning on relative performance and limits the influence of extreme returns, following Geertsema [3].

Combine nonlinear prediction with a stable linear estimate

Histogram gradient boosting fits 500 trees with learning rate 0.02, up to 255 leaves per tree and 255 histogram bins. Each tree corrects errors left by earlier trees. The usual minimum of 2,000 observations per leaf limits overfitting; no early stopping or L2 penalty is used. With a smaller training sample of n rows, the minimum becomes $\min(2,000, \max(20, \text{floor}(n/50)))$. Random seed 42 and fixed input order make the calculation reproducible. [2]

Ridge regression provides a linear counterpart. With centered ranks X , transformed targets y and n training rows, its estimate is $\beta = (X'X/n + (\lambda/12)I)^{-1}X'y/n$. The factor 1/12 scales the penalty by the approximate variance of percentile ranks. The penalty λ is selected from 0.001, 0.01, 0.1, 1 and 10 using mean monthly rank correlation in the latest 60 training months, or one-third of the window if shorter. The remaining earlier months fit each candidate; the selected penalty is then refit on the full window.

Both models refit every 12 months on the previous 240 months, after at least 36 months of history. If needed, retain at most the latest 600,000 training rows. Standardize each model's forecast within the current month. The blend is 70% tree score and 30% ridge score, standardized again. The alternative is the tree score alone. The resulting score q ranks relative performance; it is not a calibrated expected return μ .

Keep model selection separate from the competition sample

The code compares 12 fixed combinations on the 120 pre-test formation months: two forecasts, three portfolio combinations and two overlay choices. No constants are changed in response to competition-period results. Subsequent refits use newly realized history, while the chosen combination stays fixed.

Define the comparison model

The benchmark averages six ridge regressions: three on ranked characteristics and three adding 128 fixed random nonlinear features. The penalties are 0.1, 1 and 10, with exponentially weighted training and a 120-month half-life. It allocates using characteristic-factor risk, with gross exposure capped at 2 and stock weights at 2%. Its settings and results were frozen before this comparison.

3. Estimate risk in two ways

Model A: shrink the sample covariance

Use up to 756 daily excess returns for each current stock. In the full sample, require at least 250 observations; shorter inputs use a smaller threshold. Replace missing daily returns by zero, then demean each series. Ledoit-Wolf shrinkage pulls the sample covariance toward a scaled identity matrix, with the shrinkage intensity estimated from the data [4]. Multiply daily covariance by 21 to obtain a monthly estimate. This captures broad co-movement while reducing sensitivity to a singular sample covariance.

Model B: industries and characteristic clusters

Construct exposures to the 12 Fama-French industries from the supplied SIC codes [5]. Re-estimate style clusters every 12 months using the previous 60 months of ranked characteristics. Retain characteristics with at least 50% observed values and positive variation. Compute a normalized cross-product matrix from ranks centered at 0.5, with missing centered values set to zero. This is a rank-correlation approximation. Average-linkage hierarchical clustering groups the characteristics into up to 13 clusters using distance 1 minus absolute correlation. This adapts the clustering idea in Jensen, Kelly and Pedersen [6].

Within each cluster, align signs with the first principal component, average the signed centered ranks, and standardize that exposure within the month. Together with industry indicators, these columns form exposure matrix B . The factor-risk design draws on Jensen et al. [7]. Daily cross-sectional regressions use exposures from the last month-end strictly before each day, with ridge penalty 0.0001 times the number of observed stocks. They estimate factor returns and stock-specific residuals.

The factor estimate is $\hat{\Sigma}_B = B\hat{\Omega}B' + D$: Ω is factor-return covariance and diagonal D contains stock-specific variance. Estimate factor covariance from up to 1,260 days, using exponentially weighted correlations with a 504-day half-life and variances with an 84-day half-life. Combine these correlations and standard deviations, then multiply by 21. The diagonal uses residual squared returns over 252 days with an 84-day half-life, also multiplied by 21. Normally require 200 residual observations; use smaller thresholds in short samples, fill unreliable estimates with the eligible median, and floor positive variances at their fifth percentile.

Why combine the estimates?

The sample model can capture risks omitted from the chosen factors, but estimates many noisy relationships. The factor model estimates fewer relationships and uses economically interpretable exposures, but can miss risks outside its factors. The strategy can use either portfolio or average the two portfolios equally. It averages holdings, not covariance matrices. Clusters can be very unequal in size, so 13 clusters need not represent 13 equally important sources of risk.

4. From forecasts to holdings

Allocate using regularized covariance

Let π denote portfolio weights and let a hat mark an estimate. For each risk model, add η times the identity matrix to its covariance, where η is average stock variance. This further limits sensitivity to covariance-estimation error. Following Geertsema [3], use relative score q to form the dollar-neutral mean-variance direction:

$$\hat{\Sigma}_\eta = \hat{\Sigma} + \eta I, \quad \eta = \text{tr}(\hat{\Sigma})/N$$

$$\tilde{\pi} = \hat{\Sigma}_\eta^{-1} q - \hat{\Sigma}_\eta^{-1} \mathbf{1} \frac{\mathbf{1}' \hat{\Sigma}_\eta^{-1} q}{\mathbf{1}' \hat{\Sigma}_\eta^{-1} \mathbf{1}}$$

Scale each direction to unit gross exposure, the sum of absolute weights. For the combined portfolio, average the two directions. Regress the resulting weights on a constant and estimated market beta, retain the residuals, and rescale to unit gross. Betas use 252 trailing daily returns and the equal-weight current-stock market proxy. Following Keshri [8], this final Euclidean projection achieves dollar and estimated-beta neutrality; it is not the exact joint covariance-constrained optimum. If a risk model is unavailable, use the other; if both fail, use centered forecast ranks.

Adjust exposure when predicted risk changes

The selected combination is **the tree-ridge blend with both risk portfolios and the volatility overlay**. Once 12 prior risk estimates exist, the overlay multiplies weights by median past predicted volatility divided by current predicted volatility, using model B. Clip the multiplier to 0.5-2.0. Thus exposure changes with risk; it is not a fixed volatility target. Observed gross exposure ranges from 0.50 to 1.42; the largest absolute stock weight is 0.53%. There is no explicit single-stock cap.

Deduct trading costs from investable capital

Cost per dollar traded	Annual mean	Sharpe	Max drawdown
0 bp	8.0%	3.74	-2.7%
5 bp	7.4%	3.47	-3.0%
10 bp	6.8%	3.20	-3.3%
25 bp	5.1%	2.38	-4.3%
50 bp	2.2%	1.02	-23.4%

Monthly purchases plus sales average 96.4% of capital, including holdings drift. For cost rate c , pretrade weights h and target weights π relative to post-fee capital, solve for the remaining capital fraction a :

$$a = 1 - c \sum_i |a\pi_i - h_i|, \quad r_{\text{net}} = a(1 + r_{\text{gross}}) - 1$$

Fees apply at entry and rebalancing. The calculation approximates the risk-free rate as zero and excludes final liquidation, stock borrowing, financing and market impact. These are hypothetical proportional-cost scenarios. The optimizer itself does not penalize turnover.

5. Evidence and investor expectations

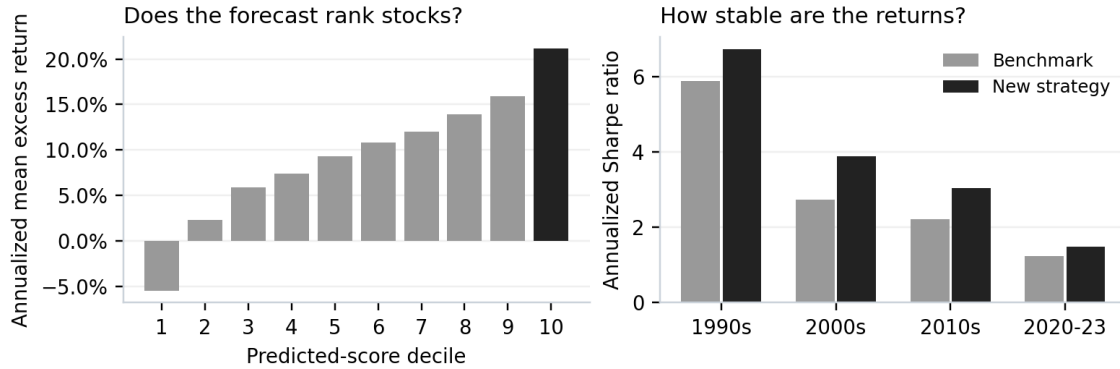


Figure 2. Left: average excess returns of stocks sorted into ten groups by forecast score. Right: Sharpe by return-year subperiod, before costs. These are descriptive comparisons within the already known competition sample.

What was selected, and what was verified?

The fixed 12-way selection chose the tree-ridge blend, both risk portfolios and the overlay using January 1980-December 1989 validation returns. Its validation Sharpe is 6.62. On the 119 months shared with the ridge benchmark's development record, Sharpe is 6.78 versus 5.92. Selection inflates the winning validation score. The known competition sample did not choose among the 12 combinations.

Average monthly forecast rank correlation is 0.064. A 12-month moving-block bootstrap with 2,000 draws gives a 95% competition-sample Sharpe interval of [3.12, 4.50]; the paired gain over the ridge benchmark has interval [0.44, 1.14]. Newey-West inference [9] with six lags gives a mean-return t-statistic of 12.59. These calculations allow time dependence under a sufficiently stable process. They do not adjust for all research choices or reuse of the sample.

The full-data entry point was rerun and its returns independently reconstructed. Separate checks verify covariance arithmetic, clustering, allocation and costs. Tests on 123-month panels cover shuffled inputs, missing values, large integer IDs, and removal or alteration of future information. Python 3.13 and exact dependencies are supplied. The submitted source reads no files or external data. Official security review and scoring remain pending.

What should an investor expect?

Gross Sharpe is 1.48 in 2020-2023, compared with 3.74 over the full sample. Neutrality to an estimated internal market proxy does not remove every risk; covariance estimates, clusters and forecasts can fail when relationships change. Shorting and trading thousands of stocks may cost more than these scenarios assume. Prospective evidence and stock-specific execution costs are needed before treating historical Sharpe as an investment expectation.

References

- [1] JKP Global Factor Data. *Common Task Framework: Rules and official WRDS dataset* (2026). [Source](#).
- [2] Scikit-learn developers. *HistGradientBoostingRegressor*, version 1.7 documentation (implementation: 1.7.2). [Source](#).
- [3] Geertsema, P. (2026). *LightGBM with penalised VCV (MV - GMV)*. Public CTF model, August 30. [Source](#).
- [4] Ledoit, O., and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2), 365-411. [Source](#).
- [5] French, K. R. *12 Industry Portfolios: Industry Definitions*. Dartmouth Data Library. [Source](#).
- [6] Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *Journal of Finance* 78(5), 2465-2518. [Source](#).
- [7] Jensen, T. I., B. Kelly, S. Malamud, and L. H. Pedersen (2026). Machine learning and the implementable efficient frontier. *Review of Financial Studies*. [Source](#).
- [8] Keshri, A. (2026). *XGB+MLP + two risk models (beta neutral)*. Public CTF model, September. [Source](#).
- [9] Newey, W. K., and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703-708. [Source](#).

Online sources accessed October 6, 2026. AI assistance: Claude and OpenAI Codex supported implementation, testing and drafting.