

Problem Set 3

MGT 924: Statistical Foundations

Courtney Davis & Muhammad Sheharyar Siddiqui

October 6, 2026

1 Introduction

The goal of this project is to predict excess returns of U.S. stocks and use the predictions to determine portfolio weights while managing risk. The model uses 402 stock characteristics from the JKP Factors dataset, supplemented by two short-horizon features constructed from daily returns: the compounded return and RMS return over the final five trading days of each month, and these were used as inputs to Ridge regression and LightGBM models. The two models were then combined with equal weights to measure the relative attractiveness of stocks, accompanied by a factor risk model that accounts for common industry and style exposures. The factor risk model converted the alpha signals to portfolio weights. Finally, the resulting portfolio was scaled to a predicted 1% monthly volatility target to keep the intended level of risk consistent over time and tested on a held-out period from 1990 to 2023.

2 Feature Selection

All 402 stock characteristics were standardized through a monthly cross-sectional ranking. This takes each characteristic, orders the available stocks from low to high, and replaces the raw value with its position in the ranking for each month. The percentile ranks are centered around zero and range approximately from -0.5 to 0.5. This prevents an outlier from dominating the model based solely on its raw magnitude and also allows for comparability across features. For stocks with missing characteristics, the value assigned is zero so that the model does not assume the stock as unusually high or unusually low on that feature.

While the provided 402 characteristics already include many signals derived from daily data including volatility, beta, skewness, and liquidity, an additional two features were added: a 5-day compounded return and 5-day RMS return. These features account for the most recent return behavior from the final five trading days, because the short-term market behavior may contain useful predictive information. The compounded return captures the direction of recent performance, while the RMS return captures the magnitude of recent price movements.

Model / Mapping	Without Daily Features	With Daily Features	Change
Ridge Factor-MV	5.15	5.66	+0.50
LightGBM Factor-MV	6.50	7.00	+0.51

As seen in the table above, the two additional daily features increase the Sharpe ratio for both the Ridge and LightGBM forecasting models substantially and were therefore retained, resulting in a total of 404 predictors.

3 Ridge Regression

This is the first linear unbiased alpha model used for predicting the excess return next month as a combination of 404 ranked predictors. Ridge is used over OLS because it adds an L2 penalty proportional to the squared coefficient sizes, which causes the coefficients to shrink towards 0.

3.1 Annual lambda selection process

The model tries the following 11 candidates ranging from 10^{-3} to 10^2 in increasing penalty order.

#	Lambda (λ)
1	0.00100
2	0.00316
3	0.01000
$\vdots$	$\vdots$
11	100.000

There is no one permanent lambda. The model decides the lambda fit every December. It takes the most recent 60 months from that specific December of the year and this becomes the validation period. The ridge is then trained on the data before those 60 months and tests which lambda works best in the last 60 months. Thus, making sure it doesn't train on the same 60 months it uses to figure the lambda. The model chooses the lambda that generates the highest Sharpe ratio and the Ridge is then retrained using all available historical information. The Ridge model is used for the coming months until next December.

3.2 Why did Ridge use raw-return target?

The model tested two possibilities: raw-return target and rank target. The rank target model did order stocks better and therefore had a higher information coefficient(IC), but when the model converted predictions into portfolios, raw-return target produced a higher Sharpe ratio of 2.81 compared to 2.6. Thus, raw returns were used as a ridge target.

3.3 Why did Ridge use expanding window?

Training approach	Development Sharpe
Rolling 120 months	2.45
Rolling 240 months	2.54
Expanding window, rank-target comparison	2.60

As evident from the table, expanding the window produced a higher Sharpe ratio, which in essence means that the benefit of including more data outweighs excluding data.

4 LightGBM

This is the second alpha prediction model used to account for non-linear relationships between features. It focuses on building many decision trees sequentially, with each new addition correcting

previous errors. This model produces the within-month rank of next month's stock returns, as it is less sensitive to outliers than exact returns. The base version is computed on the basis of 31 leaves, minimum 2000 observations per leaf, learning rate of 0.05 and feature subsampling.

Variant	Factor-MV DEV Sharpe
Base LightGBM	6.06
Deeper: 127 leaves	6.00
Raw winsorised/demeaned target	5.66
Industry-relative rank target	6.03

The model tested for the following alternatives, but concluded to retain the base version as it was justified with a higher Sharpe ratio.

5 Ensemble

The two ranked forecasts receive equal weights of 50/50, as the correlation between them is 0.67, which will help diversify model error. The combined rank is converted into a standardized Gaussian alpha score, z .

Development ensemble experiment	Factor-MV Sharpe
Ridge + LightGBM	5.86
All-LightGBM combination	6.12
Ridge + all	5.84

Although, LightGBM alone produced a higher Sharpe, the ensemble was retained because it was more stable across development decades.

6 Portfolio Construction

Forecasting identifies attractive stocks but does not highlight the position size to take. Several methods were tested below for converting predictions to weights. Rank L/S longed for high-ranked stocks and shorted low-ranked stocks. Industry neutral adjusted the weights such that the model doesn't place large bets on specific industries. Alpha variance doesn't account for covariance, thus performs worst. MV Industry model accounts for expected alpha, risk, and covariance. Full factor MV model builds on the above and controls for common style-factor exposures. Ex-ante volatility scaling scaled the final portfolio to 1% predicted monthly volatility target, increasing the Sharpe from 5.86 to 5.96.

Mapping	Ridge	LightGBM	Ensemble
Rank L/S	2.81	2.90	3.09
Industry-neutral	3.79	4.20	4.28
Alpha / idio variance	2.46	2.33	2.67
MV, industry risk model	4.05	4.65	4.64
MV, full factor model	5.15	6.06	5.86
+ ex-ante vol scaling	5.24	6.12	5.96

Thus, mean-variance portfolio construction with a full factor risk model consisting of 12 industry factors and 30 clustered style factors was used, followed by ex-ante volatility scaling.

7 Risk Model

The purpose of this model is to prevent the model from taking large bets on common styles or industries. The stocks are classified into 12 Fama-French industries using SIC codes. Rather than using 402 characteristics, they were clustered into 30, with each representing a common style exposure. $K=30$ was selected as it performed better in development. The final stock-covariance matrix is as follows

$$\Sigma = BFB' + D$$

$\Sigma^{-1}z$ is used to turn predictions into risk-adjusted weights.

$$w \propto \Sigma^{-1}z$$

As a result, the stocks with the strongest predicted alpha were preferred and positions that created excessive common risk were reduced. This resulted in the largest improvement in Sharpe.

8 Alpha Signal complementing Final Portfolio weights

The Ridge-LightGBM ensemble produces a final alpha score, z . If z is positive, the stock is relatively attractive, while negative values indicate relatively unattractive stocks. In essence, the ensemble tells which stock to own, and the risk model specifies how much to own. The final portfolio is scaled to target 1% predicted monthly volatility, which alters the overall leverage of the portfolio and not the preference of stocks ranking.

9 Validation Approach

- I. Initial training data: 1951-1969 - Enough past data to estimate models
- II. Development Model selection: 1970-1989 - Choose features, models, construct portfolios
- III. Test period: 1990-2023 - Evaluate how the model performs on holdout
- IV. Inner validation at each annual refit also happens for hyperparameter selection

10 Development Results

LightGBM performed better than Ridge, especially with the addition of two daily features. LightGBM+daily produced the highest Sharpe of 7, but Ensemble + daily with Sharpe of 6.86 is retained for greater stability and model diversification.

Alpha model	Rank IC	Rank L/S SR	Factor-MV SR
Ridge	0.087	2.81	5.56
Ridge + daily	0.095	3.09	6.11
LightGBM	0.109	2.90	6.50
LightGBM + daily	0.119	3.32	7.00
Ensemble	0.106	3.09	6.43
Ensemble + daily - final	0.115	3.42	6.86

11 Final Holdout Results

For the holdout period 1990-2023, the final model achieved the following statistics with a Sharpe of 3.12. The same alpha using simple rank L/S achieved a Sharpe ratio of 1.55, while the course baseline achieved 1.36. This concludes that factor-based portfolio construction contributes to final performance.

Metric	Result
Final Sharpe	3.12
Annualized mean return	31.5%
Annualized volatility	10.1%
Max drawdown at 10% vol	8.9%
Same alpha, simple rank L/S	1.55
Course baseline	1.36

12 Reproducibility and No Look-Ahead Checks

Reproducibility was tested by re-running the final model and comparing the newly generated weights with the original weights. While LightGBM may have had small numerical differences due to calculations across multiple processor threads, the largest difference in the final weights was about $5e-16$, which is effectively zero. Additionally, truncation tests were run to confirm that future data did not impact the formation of earlier portfolio weights. The portfolio weights remained the same even while later data was added, validating that there was no look-ahead bias in the model.

13 Limitations

There are several limitations that need to be acknowledged when evaluating this model's performance. Notably, the model does not account for transaction costs, so the 59.5% monthly average turnover may reflect higher returns and Sharpe ratios than what could actually be achieved in practice. Also, the portfolio was scaled to a predicted 1% volatility target, which ended up being lower than the realized volatility of approximately 2.9% for the testing period of 1990-2023. This means that the portfolio experienced substantially more risk than the model predicted. Finally, the model became less effective over time, with the Sharpe ratio falling from roughly 6.6 in the 1990s to 0.9 from 2020–2023. This may reflect market anomalies becoming more well-known over time and exploited, potentially reducing their profitability and, consequently, the model's effectiveness in more recent periods.