

# Homework 1, MATH 5398

Zongyi Liu

Wed, Oct 1, 2025

## 1 Preparations

### 1.1 Data Sources

Firstly we are asked to retrieve data from three sources:

- S&P 500 Historical Components & Changes
- S&P 500 component stocks from 1996-01-01 to the most current date in WRDS-Fundamentals Quarterly
- daily data for S&P 500 component stocks from 1996-01-01 to the most current date from WRDS-Security Daily

### 1.2 Data Processing

Then we use the provided `Step2_preprocess_fundamental_data.py` to process two files, where we generated a final result in `final_ratios.csv`, which contains headers like PE (Price-to-Earnings Ratio), PS (Price-to-Sales Ratio), PB (Price-to-Book Ratio), OPM (Operating Margin), NPM (Net Profit Margin), ROA (Return On Assets), ROE (Return on Equity), EPS (Earnings Per Share), BPS (Book Per Share), DPS (Dividend Per Share), etc.

In using FinRL Trading model, we follow the dynamic stock recommendation framework proposed by Yang, Liu, and Wu (2018) [1], which formulates stock selection as a rolling, cross-sectional machine learning ranking problem rather than static factor-based prediction.

There are also 11 excel files inside the folder, showing 11 different sectors of stocks provided in the S&P 500.

## 2 Mean–Variance Portfolio Construction

### 2.1 Theoretical Framework

Consider  $n$  risky assets with expected returns  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^\top$  and covariance matrix  $\Sigma \in \mathbb{R}^{n \times n}$ . Let  $\boldsymbol{w} = (w_1, w_2, \dots, w_n)^\top$  denote the portfolio weights, satisfying  $\sum_{i=1}^n w_i = 1$ .

Then[2]:

$$\begin{aligned} \text{Expected return: } \quad \mu_p &= \boldsymbol{w}^\top \boldsymbol{\mu}, \\ \text{Portfolio variance: } \quad \sigma_p^2 &= \boldsymbol{w}^\top \Sigma \boldsymbol{w}. \end{aligned}$$

### 2.2 Optimization Problems

#### Minimum-Variance Portfolio

Given a target return  $r^*$ , the investor solves:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^\top \Sigma \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^\top \boldsymbol{\mu} = r^*, \\ & \boldsymbol{w}^\top \mathbf{1} = 1. \end{aligned}$$

#### Tangency Portfolio

Assuming a risk-free rate  $r_f$ :

$$\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^\top (\boldsymbol{\mu} - r_f \mathbf{1})}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}}.$$

## Key Formulas

$$E[R_p] = \mathbf{w}^\top \boldsymbol{\mu}, \quad \sigma_p^2 = \mathbf{w}^\top \Sigma \mathbf{w}.$$

$$\text{Sharpe ratio: } S_p = \frac{E[R_p] - r_f}{\sigma_p}.$$

## Efficient Frontier

The efficient frontier represents all optimal portfolios that minimize variance for a given expected return:

$$\mathcal{F} = \{(\sigma_p, \mu_p) : \exists \mathbf{w} \text{ s.t. } \mu_p = \mathbf{w}^\top \boldsymbol{\mu}, \sigma_p^2 = \mathbf{w}^\top \Sigma \mathbf{w}\}.$$

## 2.3 Remarks

- If short selling is prohibited, impose  $w_i \geq 0$ .
- Adding a risk-free asset allows the construction of the Capital Market Line (CML).
- The tangency portfolio gives the highest Sharpe ratio and serves as the optimal risky portfolio.

Here we used the `minimize` in `scipy` to calculate the MVP.

## 3 Backtesting

Then I used the `backtest.py` file to do the back testing from the year 2018 as the assignment suggested.

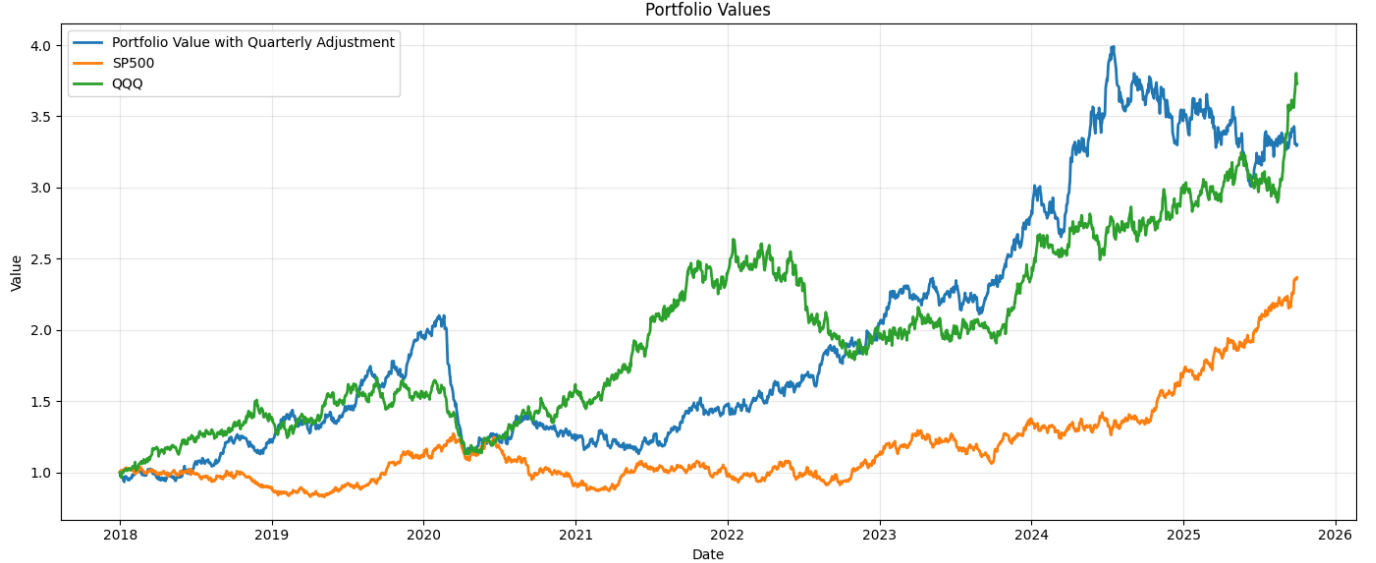


Table 1: Performance Metrics of Portfolio and Market Benchmarks (2018–2025)

Metric	Portfolio	S&P 500	QQQ
Cumulative Return (%)	1289.42	136.86	272.96
Annual Return (%)	13.21	13.77	18.51
Annual Volatility (%)	16.48	16.74	19.95
Max Drawdown (%)	-27.95	-24.77	-32.58
Sharpe Ratio	0.8017	0.6389	0.8053
Win Rate (%)	55.12	–	–
Information Ratio	0.2984	–	–

As shown in table, the portfolio significantly outperforms the S&P 500 and QQQ in cumulative terms, achieving a return of 1289.42% over the 2018–2025 period. Despite this strong long-term growth, the portfolio maintains an

annual return (13.21%) comparable to the S&P 500 while exhibiting lower annual volatility (16.48%), resulting in a high Sharpe ratio of 0.8017.

The portfolios maximum drawdown of  $-27.95\%$  lies between that of the S&P 500 and QQQ, indicating a balanced exposure to growth and downside risk. Moreover, a win rate of 55.12% and a positive information ratio of 0.2984 suggest that excess returns are generated in a stable and systematic manner. Overall, the results highlight a portfolio that delivers strong risk-adjusted performance while controlling volatility and drawdowns.

## References

- [1] Hongyang Yang, Xiao-Yang Liu, and Qingwei Wu. A practical machine learning approach for dynamic stock recommendation, 2018. SSRN 3302088.
- [2] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.