Investment Theory
FIN 5320

Prof. Lorenzo Naranjo
Fall 2026

Group Project

Instructions

This group assignment is due no later than Wednesday 9/30 at 11:59pm CST. You may choose your own group members, but each group must have exactly three students. If your group has fewer than three, I will randomly assign additional students to fill it. You can do the assignment either in Microsoft Excel (recommended) or Python.

- If you do your assignment in Excel, please submit your spreadsheet showing your work and a pdf document with your report. I expect your spreadsheet and report to be formatted professionally.
- If you do your assignment in Python, please convert your Jupyter notebook to HTML and submit the resulting HTML file together with a pdf document with your report. Using ChatGPT to get the Python code to run is okay, but copy-and-pasting full paragraphs in your report is not.

Your grade on this assignment depends exclusively on what you write in your report, and how you format the document. Your report must use titles, proper capitalization, paragraphs explaining your work, properly formatted tables and figures, equations if necessary, and full sentences. Your report should read and look like a professional report generated by a large investment bank for an important client.

I will check all assignments for plagiarism using Turnitin.

Objective

As a group of portfolio managers at an asset management firm, you have been tasked with constructing a portfolio of sector ETFs with a higher Sharpe ratio than the market. You will compare two ways of choosing the weights: maximizing the Sharpe ratio, which relies on estimated expected returns, and minimizing the variance, which does not. Solving each problem with and without short sales gives four portfolios. You will build them with data up to August 2025 and then test how they would have performed over the following year.

Data and Sector Views

We will be using the following **eight sector ETFs** from the Select Sector SPDR family: Technology (XLK), Financials (XLF), Consumer Discretionary (XLY), Communication Services (XLC), Health Care (XLV), Industrials (XLI), Consumer Staples (XLP), and Energy (XLE).

All the data you need is in the file `project-data.csv` on Canvas. It contains month-end prices from August 2020 to August 2026 for the eight sector ETFs and for the SPDR S&P 500 ETF Trust (SPY), which is your market proxy. The prices are adjusted for dividends and splits, so their percentage changes are total returns. The column `IRX` is the 13-week Treasury bill rate, your risk-free rate. It is an annualized rate in percent, so convert it into a monthly rate compounded monthly, and use the rate at the end of each month as the risk-free rate for the following month.

All estimation uses the 60 monthly returns from September 2020 to August 2025. The twelve months from September 2025 to August 2026 are reserved for testing your portfolios, so do not use them until the out-of-sample section.

Before running any numbers, put yourselves at the end of August 2025 and say, for each sector, whether you would have expected it to outperform or underperform the market over the following year and why, for example because of the interest rate outlook, the stage of the business cycle, commodity prices, valuation ratios, or recent news events. Use only information available at that time. These views do not enter the optimization, but you will use them later to judge the choices the optimizer makes.

Single-Index Regressions

For each ETF, run the OLS regression

$$R_i = \alpha_i + \beta_i R_m + e_i, \tag{1}$$

where $R_i = r_i - r_f$ and $R_m = r_m - r_f$ are the monthly excess returns of ETF i and of the market.

I do not need to see the output of each regression. Instead, generate one single table that displays the relevant information for all ETFs at once:

- Alpha (per year)
- Beta
- R-square
- Volatility of excess returns (per year)
- Sector-specific volatility (per year)

To compute the sector-specific volatility, use $\sigma^2(e_i) = \sigma_i^2 - \beta_i^2 \sigma_m^2$, which requires the volatility of excess market returns as well. For each ETF, also generate a scatter plot that includes the regression line, and display all your plots tightly in a four-row two-column configuration.

Maximizing the Sharpe Ratio

Estimate the expected excess return of each ETF with its annualized sample average over the $T = 60$ months,

$$\bar{\mu}_i = 12 \times \frac{1}{T} \sum_{t=1}^{T} R_{i,t},$$

and estimate risk with the sample covariance matrix of the monthly excess returns, multiplied by 12 to annualize it. Call $\hat{\sigma}_{ij}$ the entry of this matrix for ETFs i and j. The expected excess

return and the volatility of a portfolio with weights $w_1, \ldots, w_n$ of the $n = 8$ ETFs are then

$$\bar{\mu}_P = \sum_{i=1}^{n} w_i \bar{\mu}_i \quad \text{and} \quad \sigma_P = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \hat{\sigma}_{ij}}.$$

In Excel, build the covariance matrix with `=COVARIANCE.S()` and compute σ_P^2 with `=SUMPRODUCT(w, MMULT(S, w))`, where `w` is the column of weights and `S` is the covariance matrix. This gives the same number as the annualized variance of the portfolio's monthly excess returns, but it makes the role of the covariance matrix explicit.

Solve two versions of the problem. The unconstrained version only requires the weights to add up to one, so short sales are allowed:

$$\max_{\{w_1,\ldots,w_n\}} \frac{\bar{\mu}_P}{\sigma_P} \quad \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1.$$

The short-sale constrained version also requires $w_i \geq 0$ for every asset i. If you use Excel Solver, uncheck "Make Unconstrained Variables Non-Negative" when you solve the unconstrained versions of this problem and the next one, since that option silently rules out short sales.

Report the composition, expected excess return, standard deviation, Sharpe ratio, alpha and information ratio of both portfolios. To get the alpha and the sector-specific volatility of a portfolio, run the regression in (1) on its monthly excess returns, $R_{P,t} = \sum_{i=1}^{n} w_i R_{i,t}$. Which sectors survive in the constrained portfolio, and are they the ones you expected to outperform? Relate the optimizer's choices to the alphas, betas and correlations of your ETFs. Because OLS fits the averages exactly, $\bar{\mu}_i = \alpha_i + \beta_i \bar{\mu}_m$, so the alphas are what separates the expected returns of ETFs with similar betas.

Minimizing the Variance

The portfolios above treat historical averages as expected returns, but averages estimated from five years of monthly data are very noisy. The optimizer takes them at face value and loads up on the ETFs whose averages happen to be highest in your sample, which is why Michaud (1989)

calls mean-variance optimizers "estimation-error maximizers." The minimum variance portfolio avoids this problem because it uses only the covariance matrix.

Find the minimum variance portfolio of the eight sector ETFs, again in two versions. The unconstrained version solves

$$\min_{\{w_1,\ldots,w_n\}} \sigma_P \quad \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1,$$

and the short-sale constrained version also requires $w_i \geq 0$ for every asset i.

Report the same statistics as before and present all four portfolios side by side in a single table. Which sectors survive in the constrained portfolio, and how do they compare with the ones in the constrained maximum Sharpe ratio portfolio?

What the Short-Sale Constraint Does

Removing a constraint can never make the objective worse in the sample used to solve the problem, so in your sample each unconstrained portfolio beats its constrained version on its own objective. Their weights, however, are usually extreme, with large long and short positions in sectors that move closely together. The optimizer is exploiting small differences in the estimated inputs, and many of those differences are just sampling error.

Jagannathan and Ma (2003) show that, for the minimum variance portfolio, ruling out short sales has the same effect as shrinking the estimated covariances of the assets the optimizer would like to short. Those are the assets with the largest estimated covariances, which are the estimates most likely to be inflated by sampling error, so the constraint undoes part of that error. This is why the constraint can help even when the true minimum variance portfolio includes short positions.

The constraint also works like a Lasso regression (Tibshirani 1996). Because the weights add up to one, ruling out short sales is the same as requiring the absolute values of the weights to add up to one. The Lasso imposes this kind of bound on the coefficients of a regression, which sets some coefficients exactly to zero and keeps the rest from taking extreme values. This is why some sectors drop out of your constrained portfolios while the others hold moderate weights.

Compare the unconstrained and constrained versions of each portfolio. How extreme are the unconstrained weights, and which sectors do they short? Which sectors does the constraint remove? Relate what you see to the covariance shrinkage and Lasso interpretations of the constraint.

Out-of-Sample Performance

A portfolio that looks good on the data used to build it can still disappoint afterward, because the optimizer fits the noise in the sample as well as the signal. DeMiguel et al. (2009) find that none of the optimized portfolios they study consistently beats simply giving every asset the same weight. To see how your four portfolios would have done in practice, hold their weights fixed and compute the monthly return of each portfolio, $r_{P,t} = \sum_{i=1}^{n} w_i r_{i,t}$, from September 2025 to August 2026. This amounts to rebalancing back to your weights at the end of every month. Use SPY as a passive benchmark.

Plot the growth of $1 invested at the end of August 2025 in each of your four portfolios and in SPY. Then report, for each portfolio and for SPY, the annualized average excess return, the annualized standard deviation and the Sharpe ratio over the twelve out-of-sample months. For the four sector portfolios, also run the regression in (1) on these months and report the realized alpha and information ratio. Present these numbers in a single table next to the in-sample numbers, keeping in mind that the in-sample numbers come from the data the optimizer used to choose the weights and are therefore biased in its favor.

Do the in-sample rankings survive out of sample? Did the maximum Sharpe ratio portfolios keep their advantage, and did the minimum variance portfolios deliver the low volatility they promised? Did the short-sale constraint help or hurt each approach? Relate what you find to the noise in your estimated average returns and to the argument of Jagannathan and Ma (2003), and keep in mind that a single year of returns can favor any portfolio by luck.

Conclusion

Based on all your results, including the out-of-sample performance, write a convincing paragraph recommending the best approach to form your portfolio. Weigh the benefits and costs of relying on estimated expected returns rather than only on the covariance matrix, and of allowing short sales rather than ruling them out.

References

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. "Optimal Versus Naive Diversification: How Inefficient Is the $1/N$ Portfolio Strategy?" *Review of Financial Studies* 22 (5): 1915–53.

Jagannathan, Ravi, and Tongshu Ma. 2003. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance* 58 (4): 1651–83.

Michaud, Richard O. 1989. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal* 45 (1): 31–42.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.