FE630 Portfolio Theory and Applications
Final Project

Prof. Papa Momar Ndiaye

April 22, 2025

1 Overview

The goal of this project to build and compare 2 factor-based long short allocation models with constraints on their betas. The first strategy considers a target Beta in the interval $[-0.5, 0.5]$ while the second has a target Beta in the interval $[-2, +2]$. While the first strategy is more like a Value-at-Risk type of Utility corresponding to Robust Optimization, the second incorporates an Information Ratio term to limit the deviations from a benchmark unless those deviations yield a high return. Once the optimization models built, we want to compare the outcomes of the two models while evaluating their sensitivity to the the length of the estimators for covariance matrix and the expected returns under different market scenarios.

The portfolios will be reallocated (re-optimized every week for period of analysis from March 2007 to end of December 2024. The investment universe is a set of ETFs large enough to represent the World global economy and we will use the French Fama 3-factor model (Momentum, Value and Size).
The factors historical values are freely available for download from Ken French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french) and from yahoo for the ETFs.

The performance and the risk profiles of such strategies may be very sensitive to the target Beta and the market environment. A low Beta meaning a strategy that aims to be de-correlated to the global market represented by the S&P 500, while a high Beta meaning that, having a big appetite for risk, we are aiming to ride or scale up the market risk. In addition to that, such a strategy is likely to to be quite sensitive to the estimators used for the Risk Model and the Alpha Model (for example the length of the look-back period used for estimation risk and expected returns), so it is important to understand the impact of those estimators on the Portfolio's characteristics: realized return, volatility, skewness, VaR/ CVaR and risk to performance ratios such as the Sharpe ratio.

To simplify, we will assume in this project that once the factor model built, we will use trend following estimators for the Expected returns. As the quality of those estimators depend on the length of the look-back period, we will typically consider 3 cases: a Long-Term estimators (LT, look-back 180 days), a Short-Term estimator (ST, 40 to 60 days) and a Mid-Term estimator (MT, 90 days), defining therefore a Term-Structure for the Covariance and Expected Return.

In summary, the behavior of the optimal portfolio built from a specific combination of estimators for Covariance and Expected Return may change with the Market environment, a particular strategy being defined by a specific combination, for example the notation S^{90}_{40} - can be used to say that you are using 40 days for estimation of covariance, 90 days for estimation of Expected Returns. The goal of this project is to understand, analyze and compare the behavior of strategies built using chosen combinations of return/risk estimators during several historical periods subprime (2008) crisis, during the first year of COVID (mid 2020 to mid 2021) and out of crisis periods.

2 Investment Strategy

We consider 2 strategies

$$\textbf{(Strategy I)} \begin{cases} \max_{\omega \in \mathbb{R}^n} \rho^T \omega - \lambda \sqrt{\omega^T \Sigma \omega} \\ -0.5 \leq \sum_{i=1}^{n} \beta_i^m \omega_i \leq 0.5 \\ \sum_{i=1}^{n} \omega_i = 1, \ -2 \leq \omega_i \leq 2, \end{cases} \tag{1}$$

and

$$\textbf{(Strategy II)} \begin{cases} \max_{\omega \in \mathbb{R}^n} \dfrac{\rho^T \omega - r_{SPY}}{TEV(\omega)} - \lambda \sqrt{\omega^T \Sigma \omega} \\ -2 \leq \sum_{i=1}^{n} \beta_i^m \omega_i \leq 2 \\ \sum_{i=1}^{n} \omega_i = 1, \ -2 \leq \omega_i \leq 2, \end{cases} \tag{2}$$

where

- Σ is the the covariance matrix between the securities returns (computed from the French Fama 3-factor Factor Model),
- $\beta_i^m = \dfrac{cov(r_i, r_M)}{\sigma^2(r_M)}$ is the Beta of security S_i as defined in the CAPM Model so that $\beta_P^m = \sum_{i=1}^{n} \beta_i^m \omega_i$ is the Beta of the Portfolio;
- $TEV(\omega) = \sigma(r_P(\omega) - r_{SPY})$ is the Tracking Error Volatility (also explicitly given by the formula $\sigma(r_P(\omega) - r_{SPY}) = \sqrt{\omega^T \Sigma \omega - 2\omega^T cov(r, r_{SPY}) + \sigma_{SPY}^2}$

The French Fama factor models are well documented in the literature but reminded here for reference. For instance, under the 3-factor model, the random return of a security is given by the formula

$$r_i = r_f + \beta_i^3 (r_M - r_f) + b_i^s r_{SMB} + b_i^v r_{HML} + \alpha_i + \varepsilon_i \tag{3}$$

with $\mathbb{E}(\varepsilon_i) = 0$ in such a way that we have in terms of Expected Returns

$$\rho_i = r_f + \beta_i^3(\rho_M - r_f) + b_i^s \rho_{SMB} + b_i^v \rho_{HML} + \alpha_i. \tag{4}$$

In equation (2), the 3 coefficients β_i^3 , b_i^s and b_i^v and estimated by making a linear regression of the time series $y_i = \rho_i - r_f$ against the time series $\rho_M - r_f$ (Momentum Factor), r_{SMB} (Size Factor) and ρ_{HML} (Value Factor)[1]. Note that in general $\beta_i^m \neq \beta_i^3$ and needs to be estimated by a separated regression or computed directly.

3 Investment Universe and Analysis Setup

3.1 Investment Universe

We will consider the following set of ETFs that you can download from Yahoo finance for the period from March 2007 to December 2024.

1. CurrencyShares Euro Trust (FXE)
2. iShares MSCI Japan Index (EWJ)
3. SPDR GOLD Trust (GLD)
4. Powershares NASDAQ-100 Trust (QQQ)
5. SPDR S&P 500 (SPY)
6. iShares Lehman Short Treasury Bond (SHV)
7. PowerShares DB Agriculture Fund (DBA)
8. United States Oil Fund LP (USO)
9. SPDR S&P Biotech (XBI)
10. iShares S&P Latin America 40 Index (ILF)
11. iShares MSCI Pacific ex-Japan Index Fund (EPP)
12. SPDR DJ Euro Stoxx 50 (FEZ)

[1] ρ_M for Market hence Momentum, ρ_{SMB} for Small minus Big and ρ_{HML} for High minus Low

3.2 Benchmark

The benchmark will be the Market Portfolio S&P 500 (SPY ETF)

3.3 Analysis Periods and BackTesting

- Divide the overall analysis period into 5 sub-periods: before, during and after the subprime crisis and covid.
- Run separate backtests for each sub-period when comparing strategies and assessing the impact of the term structure, for example, compare S_{40}^{180} to S_{40}^{90}.
- Run also a comparison over the whole period from March 1^{st}, 2007 to end of December 2024.
- Analyze the sensitivity to λ. Consider 3 different values: 0.1, 0.5 and 1.
- For your backtesting, rebalance your portfolios once a week.

3.4 Important remark for BackTesting

A backtesting is non anticipative. That means that if you testing a strategy at given date t, you should only use past data as inputs for your optimization.

- For example, assume that you you decide to backtest a strategy with 40 days of historical data for the covariance, 90 days for the returns. In that case you need to estimate the Betas, Covariance and expected returns excluding all information for dates greater or equal to t.
- In the case of weekly rebalancing, assume that you generate a new portfolio every week. Then you will have a run a new optimization every 5 days, say at a given sequence of dates $t_1, t_2, t_3, \ldots, t_n$. For first date t_i, use historical data to estimate all inputs, run optimization, and store the weights. For the next date, roll the historical data window, re-estimate your inputs and generate new weights. Repeat the process until you reach the date t_n.

3.5 Performance and Risk Reporting for comparing Strategies

Use a Performance Analytics Module in R, Matlab or Python as much as possible for the Risk and Performance Reporting. Below is the list of Key Indicators to report your Optimal Strategies. All daily indicators will be annualized assuming that each year has 250 business days. For example, the Reporting for a Strategy over a given period (example: from 03/01/1997 to 12/30/2008) will be provided by a summarizing Table with the following lines

- Cumulated PnL or Return
- Daily Mean Arithmetic / Geometric Return, Daily Min Return
- Max 10 days Drawdown
- Volatility
- Sharpe Ratio
- Skewness, Kurtosis
- Modified VaR, CVaR

In addition to that table:

1. Plot the evolution the graph of cumulated daily Profit and Loss (PnL) assuming that you invest \$100 at the first allocation date in Portfolio and \$100 in the S&P 500 (when you benchmark your strategies against the Market, the SPY is representing the S&P500 Index).
2. Plot and analyze the distribution of daily Returns.
3. A summarizing Table with the following lines for comparison with the underlying

   For comparison with between the strategies and the S&P, a summary table may look like:

	S_{60}^{90}	S_{120}^{30}	S_{180}^{90}	SPY
Mean Return			12	
⋮				
Max DD			8	

3.6 Tools

- Data can be downloaded Python using Pandas and yfinance.
- The strategy (I) will be implemented using a the convex optimizer CVXPY and the strategy (II) will be implemented using a nonlinear optimizer.

4 Submission of the Final Report

You have to submit the following.

1. A final report can be a Word, Latex File or PPT slides presenting your findings and conclusions about the impact of the estimators on the behavior of your strategy, and also what kind of estimators would recommend to use, when and why (before, during and after the crisis). So to repeat again, a global period of backtest from March 1^{st}, 2007 to end of December 2024 with sub-periods (before, during and after crisis periods and the axis of analysis is the sensitivity to the term-structure of estimators (short-term, mid-term and long-term) for covariance and expected returns.

2. The report should contain a clear description of the notations, models and strategies you have analyzed, the graphs and summarizing tables supporting your quantitative and qualitative analysis. You can include a brief description of the computational engine you have built but do not include any code or Rmarkdown output in the core of your report.

3. Submit also the code developed for the project (R, Matlab, Python or other) and all supporting graphs, tables and simulation results in a Zip file. The code should ready to run when unzipped and with minimal directions to the evaluators. The submitted code will be tested for comparison and it is a requirement to build your code in a modular and clearly documented manner..

Appendix

A Practical aspects

For estimation of the parameters of the factor model, you can use a cross sectional regression model by gathering all the individual securities model in a single "big" factor model. If you assume, that you have 3 factors, then the model at time t for each asset is given by

$$r^e_{it} = \alpha_i + \beta^3_i(r_{Mt} - r_{ft}) + b^s_i r_{SMBt} + b^v_i r_{HMLt} + \alpha_i + \varepsilon_{it} \tag{5}$$

with $r^e_{it} = r_{it} - r_{ft}$, and moreover the ε_{it} are independent of the factors and satisfy

$$cov(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma^2_i \text{ when } i = j, \text{ and } t = s \\ 0 \text{ otherwise.} \end{cases}$$

A.1 Time Series Model for a given Security

If we consider T observations of the excess return of Security S_i stacked in a column vector $R_i = \begin{bmatrix} r_{i1}^e \\ r_{i2}^e \\ \cdots \\ r_{iT}^e \end{bmatrix}$, we have the time series regression model for Security i:

$$R_i = \mathbf{1}_T \alpha_i + F\beta_i + \varepsilon_i \text{ for } i = 1, 2, \ldots n \tag{6}$$

where

- $\beta_i = \begin{bmatrix} \beta_i^3 \\ b_i^s \\ b_i^v \end{bmatrix}$ is the (3 by 1) vector of Betas
- $\mathbf{F} = \begin{bmatrix} \mathbf{f}_1' \\ \vdots \\ \mathbf{f}_T' \end{bmatrix} = \begin{bmatrix} r_{M1} - r_{f1} & r_{SMB1} & r_{HML1} \\ \vdots & \ddots & \vdots \\ r_{MT} - r_{fT} & r_{SMBT} & r_{HMLT} \end{bmatrix}$ is the is the (T x 3) matrix of observations of the factors.
- the residual term ε_i is a (T by 1) vector satisfying $\mathbb{E}(\varepsilon_i \varepsilon_i') = \sigma_i^2 \mathbb{I}_T$

The previous model is well-suited for a regression to estimate the coefficients of the model using data for the securities and the factors.

A.2 Cross Sectional Model

Alternatively, we can use a cross sectional formulation that can be useful for risk analysis including the derivation of the covariance matrix of the returns. If we define $R_t = \begin{bmatrix} r_{1t}^e \\ r_{2t}^e \\ \cdots \\ r_{nt}^e \end{bmatrix}$, the vector of all Securities excess returns at time t, then we can write

$$R_t = \alpha + \mathbf{B}\mathbf{f}_t + \varepsilon_t \text{ for } t = 1, 2, \ldots T, \tag{7}$$

where

- $B \begin{bmatrix} \beta_1' \\ \beta_2' \\ \cdots \\ \beta_n' \end{bmatrix} = \begin{bmatrix} \beta_1^3 & b_1^s & b_1^v \\ \vdots & \ddots & \vdots \\ \beta_n^3 & b_n^s & b_n^v \end{bmatrix}$ is a N by 3 matrix,

- $\mathbf{f}_t = \begin{bmatrix} r_{Mt} - r_{f1} \\ r_{SMBt} \\ r_{HMLt} \end{bmatrix}$ is the vector of factor returns at time t.
- $\mathbb{E}(\varepsilon_t \varepsilon_t^{'} | \mathbf{f}_t) = D = diag(\sigma_1^2, \ldots, \sigma_n^2)$

The cross sectional model implies that if Ω_f is the covariance of the factors, then

$$cov(R_t) = \mathbf{B} \Omega_f \mathbf{B}^{'} + D \tag{8}$$

which implies that

$$cov(R_{it}) = \beta_i \Omega_f \beta_i + \sigma_I^2 \tag{9}$$

and

$$cov(R_{it}, R_{jt}) = \beta_i \Omega_f \beta_j \tag{10}$$