

FE630 Portfolio Theory and Applications
Final Project

Final Report due on May 14th, 2024

Prof. Papa Momar Ndiaye

April 23, 2024

1 Overview

The goal of this project is 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 are built, we want to compare the outcomes of the two models while evaluating their sensitivity to 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 March 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/>) 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 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 is built, we will use trend following estimators for the Expected returns. As the quality of those estimators depends on the length of the look-back period, we will typically consider 3 cases: a Long-Term estimator (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_{40}^{90} - 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

$$(\text{Strategy I}) \left\{ \begin{array}{l} \max_{\omega \in \mathbb{R}^n} \rho^T \omega - \lambda \sqrt{\omega^T \Sigma \omega} \\ -0.5 \leq \sum_{i=1}^n \beta_i^m \omega_i \leq 0.5 \\ \sum_{i=1}^n \omega_i = 1, \quad -2 \leq \omega_i \leq 2, \end{array} \right. \quad (1)$$

and

$$(\text{Strategy II}) \left\{ \begin{array}{l} \max_{\omega \in \mathbb{R}^n} \frac{\rho^T \omega}{TEV(\omega)} - \lambda \sqrt{\omega^T \Sigma \omega} \\ -2 \leq \sum_{i=1}^n \beta_i^m \omega_i \leq 2 \\ \sum_{i=1}^n \omega_i = 1, \quad -2 \leq \omega_i \leq 2, \end{array} \right. \quad (2)$$

where

- Σ is the the covariance matrix between the securities returns (computed from the French Fama 3-factor Factor Model),
- $\beta_i^m = \frac{\text{cov}(r_i, r_M)}{\sigma^2(r_M)}$ 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 $\sigma(r_P(\omega) - r_{SPY}) = \sqrt{\omega^T \Sigma \omega - 2\omega^T \text{cov}(r, r_{SPY}) + \sigma_{SPY}^2}$)

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

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

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

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

In equation (2), the 3 coefficients β_i^3 , b_i^s and b_i^v are 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), ρ_{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 Assumptions and Analysis Setup

3.1 Setup

To simplify, we will make the following assumptions for this project.

1. The portfolios will be reallocated (re-optimized every week for period of analysis from March 2007 to end of End of March 2024.
2. Once the allocation models built, consider 3 cases for the construction the inputs:
 - a Long-Term look-back period: 120 data points for estimation of a Sample Covariance and Sample Mean (Scenario LT or S_{120})
 - a Medium-Term look-back period: 90 data points for estimation of a Sample Covariance and Sample Mean (Scenario MT or S_{90})
 - a Short-Term Look-back period: 40 data points for estimation of a Sample Covariance and Sample Mean (Scenario ST or S_{40})
3. Consider 2 possible values for the Target Beta: 0 and 1
4. Consider 2 possible values for the λ : 0.10 and 0.5

3.2 Analysis Periods

- Divide the overall analysis period into 5 sub-periods: before subprime (Period 1), during subprime (Period 2), after the subprime (Period 3), COVID (Period 4) and after COVID (Period 5).
- 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 1st, 2007 to the End of March 2024.

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

3.3 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 60 days of historical data for the covariance and the returns, say S_{60} . In that case you need to estimate the Betas, Covariance and expected returns excluding all information for dates greater or equal to t , so you will be using the data from the 60 previous days for your estimates.
- In the case of weekly rebalancing, assume that you generate a new portfolio every week. Then you have to run a new optimization every 5 days, say at a given sequence of dates $t_1, t_2, t_3, \dots, t_n$. For first date t_1 , use the 60 previous days from historical data to estimate all inputs, run optimization, and store the weights. For the next date, roll the historical data window by 5 days, re-estimate your inputs and generate new weights. Repeat the process until you reach the date t_n .

4 Tools and Data

- (Strategy I) will be implemented using the Quadratic Solver CVXPY in Python and (Strategy II) will be implemented using a nonlinear Optimizer.
- The data can be downloaded in Python from Yahoo finance. We will consider the following set of ETFs of 12 ETF
 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)

- Roughly speaking, you may consider
 1. **Task 1** will be download all the data for the period March 1st, 2007 to end of End of March 2024. and compute the daily returns for each of the securities and multiply the daily returns by 252 to annualize the data.
 2. **Task 2** Build the factor model
 3. **Task 3** Build the Optimizers
 4. **Task 4** Call the Optimizers in the Backtesting Loop
 5. **Task 5** Analyze and compare the performances
 6. **Task 6** Summarize the results in a Report

5 Performance and Risk Reporting for comparing Strategies

Below is a list of Key Indicators to report your Strategies:

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

That information may summarized in a table. For instance comparison with between the strategies and the S&P, a summary table may look like:

	$S_{40}(\beta_T = 0)$	$S_{90}(\beta_T = 1)$	$S_{120}(\beta_T = 0)$	SPY
Mean Return			12	
\vdots				
Max DD			8	

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 benchmark (represented by the SPY)
2. Plot and analyze the distribution of daily Returns.

6 Deliverable for the project

You have to submit the following.

1. A report 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 would you recommend a given estimator and why? (You may want to look at what happens during crisis periods such as subprime or COVID). Keep in mind that you have 2 axes of analysis: sensitivity to the term-structure of estimators (short-term, mid-term and long-term) for covariance and expected returns and sensitivity to the target β_T .
2. The report should contain a clear description of the notations 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 the code in the core of your report.
3. Submit also the code developed for the project 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

$$r_{it}^e = \alpha_i + \beta_i^3(r_{Mt} - r_{ft}) + b_i^s r_{SMBt} + b_i^v r_{HMLt} + \alpha_i + \varepsilon_{it} \quad (5)$$

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

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

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 $R_i = \begin{bmatrix} r_{i1}^e \\ r_{i2}^e \\ \dots \\ r_{iT}^e \end{bmatrix}$, we have the time series regression model for Security i :

$$R_i = \mathbf{1}_T \alpha_i + F \beta_i + \varepsilon_i \text{ for } i = 1, 2, \dots, n \quad (6)$$

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 ($T \times 3$) 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 $R_t = \begin{bmatrix} r_{1t}^e \\ r_{2t}^e \\ \dots \\ r_{nt}^e \end{bmatrix}$, the vector of all Securities excess returns at time t , then we

can write

$$R_t = \alpha + \mathbf{B} \mathbf{f}_t + \varepsilon_t \text{ for } t = 1, 2, \dots, T, \quad (7)$$

where

- $B = \begin{bmatrix} \beta'_1 \\ \beta'_2 \\ \dots \\ \beta'_n \end{bmatrix} = \begin{bmatrix} \beta_1^3 & b_1^s & b_1^v \\ \vdots & \ddots & \vdots \\ \beta_n^3 & b_n^s & b_n^v \end{bmatrix}$ 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 .
- $\mathbb{E}(\varepsilon_t \varepsilon_t' | \mathbf{f}_t) = D = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$

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

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

which implies that

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

and

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