FE630 - Final Project

Author: Sid Bhatia

Date: May 14th, 2024

Pledge: I pledge my honor that I have abided by the Stevens Honor System.

Professor: Papa Momar Ndiaye

1 Overview

1.1 Goal

The goal of this project to build and compare *two 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 $[-2, +2]$.

The first strategy operates similar to a **Value-at-Risk Utility** corresponding to **Robust Optimization**; the second strategy incorporates an **Information Ratio** term to limit the deviations from a benchmark, provided those deviations yield a 'high return.'

Once the optimization models are built, we want to *compare* the outcomes of the two models while simultaneously evaluating their sensitivity to the *length* of the estimators for the **covariance matrix** in tandem with the **expected returns** under various market regimes/scenarios.

1.2 Reallocation

The portfolios will be *reallocated* or, in other words, 'reoptimized' weekly from the beginning of **March 2007** to the end of **March 2024**. Our *investment universe* encompasses a set of exchange-traded funds (**ETFs**) which is large enough to represent the '**Global World Economy**' (as according to some).

We will utilize the Fama–French Three-Factor Model which incorporates the following factors:

- Momentum
- Value
- Size.

Regarding data accessability, these factors have historical values available for ***free*** from ***Ken French's*** personal website in tandem with Yahoo Finance.

1.3 Performance Evaluation

Naturally, the performance as well as the risk profiles of the aforementioned strategies may be (relatively) sensitive to the *target Beta* and the (current) market environment.

For example, a '**low Beta**' (essentially) means that a strategy is created with the objective or aim to be '**decorrelated**' (no linear relationship between entites) with the 'Global Market,' which, in our case, is represented by the **S&P 500** (i.e., no *systematic relationship*).

A '**high Beta**' is simply the antithesis, or opposite, of what we just discussed. In layman's terms, we have a (higher) appetite for '*risk*' (in this case, let's keep it simple and define our premise as σ or **standard deviation**) and desire to ride or 'scale up' the *market risk* (**systematic risk**).

Moreover, it's imperative that one acknowledges that such a (described) strategy is more probable to be (quite) sensitive to the *estimators* used for the **Risk Model** and the **Alpha Model** (e.g., the length of the *look-back period* utilized); therefore, it is necessary to understand and, most importantly, *comprehend* the impact of said estimators on the **Portfolio's** characteristics:

- (Realized) Return : μ_h
- (Historical) Volatility) : σ_h
- Skewness : $(\mathbb{E}[(\frac{x-\mu}{\sigma})^3]) = \frac{\mu_3}{\sigma_3} = \frac{\kappa_3}{\kappa_2^{3/2}}$
- VaR / Expected Shortfall
- Sharpe Ratio : $S_a = \frac{\mathbb{E}[R_a - R_b]}{\sigma_a} = \frac{\mathbb{E}[R_a - R_b]}{\sqrt{\mathbb{V}(R_a - R_b)}}$

1.4 Simplification

To make it easier, we assume that once the **Factor Model** (FM) has been constructed, we will use trend following estimators for the **Expected Returns**. Since the quality of the estimators depend on the **look-back period**, we define three cases:

- ***Long-Term Estimator (LTE)*** : $\mathrm{LT} \Rightarrow \mathrm{LB} \in \{180 \text{ Days}\}$.
- ***Mid-Term Estimator (MTE)*** : $\mathrm{MT} \Rightarrow \mathrm{LB} \in \{90 \text{ Days}\}$.
- ***Short-Term Estimator (STE)*** : $\mathrm{ST} \Rightarrow \mathrm{LB} \in \{40 \text{ Days}, 60 \text{ Days}\}$.

Specifically, we define a **Term-Structure** for the $\text{Covariance } \boldsymbol{\Sigma}$ and $\text{Expected Return } \boldsymbol{\mu}$.

1.5 Synthesis

To (briefly) summarize, the behavior of a (potential) '*optimal*' portfolio built from a melting pot of *estimators* for **Covariance** and **Expected Return** may vary according to the cadence of the '**Market**' (environment/regime) or an aforementioned strategy.

For example, the (mathematical) notation S_{40}^{90} is just fancy jargon to visually illustrate that we are using **40 days** for the covariance estimation and **90 days** for the expected returns estimations—it's not that deep.

Overall, the goal of this fun, entertaining project is to conceptualize, visualize, understand, analyze, and compare the behavior of our ideas; we want to *see* if we can (actually) make some **$$$**, especially during momentous, historical (time) periods such as the **Subprime Mortgage Crisis** of *2008*, the horrendous commencement of **Coronavirus SARS-CoV-2 Disease** of *2019*, et cetera.

2 (Investment) Strategy

Alrighty, let's get to the fun, juicy portion; shall we?

2.1 (Mathematical) Strategic Formulation

Let's make things interesting—spicy, one may say.

Consider two strats [(clipping) of 'strategies,' as embodied in *Morphology*)]:

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

and

$$\text{(Strategy II)} \quad \begin{cases} \max\limits_{\omega \in \mathbb{R}^n} \; \frac{\rho^T \omega}{\text{TEV}(\omega)} - \lambda \sqrt{\omega^T \Sigma \omega} \\ -2 \leq \sum_{i=1}^n \beta_i^m \omega_i \leq 2 \\ \sum_{i=1}^n \omega_i = 1, \quad -2 \leq \omega_i \leq 2, \end{cases} \tag{2}$$

where we define the hieroglyphics used above:

- Σ is the covariance matrix between the securities returns (as computed from the **FF3FM**);
- $\beta_i^m = \frac{\text{Cov}(r_i, r_M)}{\sigma^2(r_M)}$ is the Beta) (not to be confused with the colloquial slang usage) of some security) S_i as defined by the CAPM Model such that $\beta_P^m = \sum_{i=1}^n \beta_i^m \omega_i$ is the **Portfolio Beta**;

- $\mathrm{TEV}(\omega) = \sigma(r_P(\omega) - r_{\mathrm{SPY}})$ is the '**Tracking Error Volatility**', which (if you're *really nerdy*) you can derive it as such:

$$\sigma(r_P(\omega) - r_{\mathrm{SPY}}) = \sqrt{\omega^\intercal \Sigma \omega - 2\omega^\intercal \mathrm{Cov}(r, r_{\mathrm{SPY}}) + \sigma_{\mathrm{SPY}}^2} \tag{3}$$

Oh yeah, I should probably define what '**FF3FM**' means; that would (probably) be helpful.

2.2 Fama–French Three-Factor Model

So, to echo the previous sentiment, we should (almost surely) explain what is this *funky* model we kept referencing:

$$r_i = r_f + \beta_i^3(r_M - r_f) + b_i^s r_{\mathrm{SMB}} + b_i^v r_{\mathrm{HML}} + \alpha_i + \epsilon_i \tag{4}$$

Sorry for writing (or, to be *really technical*, typesetting) more hieroglyphics. We gotta keep going for a bit—stay with me!

If we assume our white noise/error terms, on 'average', have a (numerical) value of 0 (i.e., $\mathbb{E}[\epsilon_i] = 0$), we can derive a new goofy equation:

$$\rho_i = r_f + \beta_i^3(\rho_M - r_f) + b_i^s \rho_{\mathrm{SMB}} + b_i^v \rho_{\mathrm{HML}} + \alpha_i \tag{5}$$

In the new cursive script defined above, the 3 coefficients β_i^3, b_i^s, and b_i^v are estimated by making a linear regression, or, in 'plain English', drawing a line of best fit of the time series $y_i = \rho_i - r_f$ against the other cool time series $\rho_M - r_f$ (**Momentum Factor**), r_{SMB} (**Size Factor**), and ρ_{HML} (**Value Factor**).

I feel like I'm forgetting something $\ldots$

Oh yeah! There's an extra (nerdy) thingy we gotta verify: (generally), $\beta_i^m \neq \beta_i^3$ and needs to be estimated by a separate regression or directly computed.

2.3 'Plain' English Formulation

Whew. Let's a take breather, shall we?

I get it; that was a *mouthful*, to say the least.

But, let's try and *digest* that in a slower, easier fashion.

Overall, we are exploring two *different investment strategies*, each with its own set of rules and objectives; let's dive right into them.

2.3.1 Strategy I Breakdown

1. **Objective**: Maximize returns while considering risk (i.e., make as much

$ as humanly possible without it being (bi)polar)

2. **Constraints**:
   - The portfolio's beta (a measure of its *volatility* relative to the market; i.e., how *silly* and *spread out* it is relative to the 'market') must be between -0.5 and 0.5.
   - The sum of the weights assigned to each asset in the portfolio must equal 1 (i.e., ***we gotta put our money to work!*** As such, let's buy a bunch of stuff that can make us money but, also, let's (try) not to violate the Laws of Probability Theory).
   - Each individual weight can range from -2 to 2 (i.e., we can be like *certain individuals* from WallStreetBets and put all our eggs in one basket or, like a more prudent investor, do anything *but that*).

2.3.2 Strategy II Breakdown

1. **Objective**: Maximize returns relative to the portfolio's **tracking error volatility** (**TEV**), which measures how much the portfolio's returns deviate from a benchmark (e.g., the S&P 500 or 'big boy stock market').
2. **Constraints**:
   - The portfolio's beta (a measure of its *volatility* relative to the market; i.e., how *wild* and *crazy* it gets compared to the 'market') must be between -2 and 2.
   - The sum of the weights assigned to each asset in the portfolio must equal 1 (i.e., ***we need to make sure all our money is actively working!*** So, let's diversify our investments while still following the Laws of Probability Theory).
   - Each individual weight can range from -2 to 2 (i.e., we can either go *all in* on one asset like *those wild investors* on WallStreetBets, or spread our investments more wisely).

Don't worry about all the fancy schmancy 'math'(matics); math is for nerds (yours truly, included). All math is, is it's another language. The more you practice it, the better you get.

Anyways, that's enough of my rambling and yapping. Let's explore the setup (in da next section *insert cool kid emoji*)!

3 Assumptions and (Analysis) Setup

So, if you made it this far, you deserve a cookie! 🍪

Nice job. 😎

Alrighty, enough shenanigans. Let's get (back) to work:

3.1 Setup

To make it easier, we will make the following assumptions for this (swag(gy)) project:

1. The portfolios will be *reallocated* (reoptimized) weekly from the beginning of **March 2007** to the end of **March 2024**.

2. Once the (fancy, math-y) models are made, let's think about three cases or situations for the input construction:

    - $\text{Long-Term Look-Back Period}: 120\,\text{Data Points}$ for estimation of a $\text{Sample Covariance \& Sample Mean}$; i.e., $\text{Scenario LT} \equiv S_{120}$.
    - $\text{Medium-Term Look-Back Period}: 90\,\text{Data Points}$ for estimation of a $\text{Sample Covariance \& Sample Mean}$; i.e., $\text{Scenario MT} \equiv S_{90}$.
    - $\text{Short-Term Look-Back Period}: 40\,\text{Data Points}$ for estimation of a $\text{Sample Covariance \& Sample Mean}$; i.e., $\text{Scenario ST} \equiv S_{40}$.
3. Consider two possible values for the **Target Beta** (again, ***not*** the colloquial slang term) $: 0 \ \& \ 1$.

4. Consider two possible values for the $\boldsymbol{\lambda}$ (the ***risk aversion parameter***; i.e., how much are you putting on black?) $: 0.10 \ \& \ 0.50$.

3.2 Analysis (Time) Periods

Alrighty, I am running out of time until the deadline at **11:59 PM EST** (😱), so (please) excuse me if I must ***lock in***, as the youngins colloquially say.

We will do the following:

- 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 post-COVID (**Period 5**).
- Run separate **backtests** for each sub-period when comparing strategies and assess the impact of the *term structure* (e.g., $S_{40}^{180} \text{ versus } S_{40}^{90}$).
- Run an entire period comparison from **March 1st, 2007** to **March 31st, 2024**.

3.3 (Back)Testing

I, unfortunately, do not have enough time to rigorously define *what* backtesting is, but, in simple terms, it's just using past or historical data (stock prices) and *see* how our math/code performs. If it does a good job, yay! If it doesn't, yikes. That's basically it (in a nutshell).

However, there are some logistical considerations that are important to note:

- Backtests are **not** forecasts; we can't ***double dip*** and use 'future' data for our optimization; that's cheating!

- Regarding 'rebalancing', we assume that we generate a new portfolio each week; that is, we have to run a new optimization every 5 days for a sequence of dates $t_i \in [t_1, \ldots, t_n], \ \forall\, i \in \{1, \ldots, n\}$. For the first date (t_1), we use the 60 previous days from historical data to estimate all inputs, run optimization, and store the weights. For the next date, we **roll the historical data window** by 5 days, re-estimate our inputs, and generate new weights. We (lather), rinse, and repeat until we reach our target data t_n.