MMF1921: Operations Research

Project 2 (Summer 2023)

Group Members: Joshua Kim, Ian Lee, Linlong Wu

Abstract: Financial portfolio optimization is a broadly studied topic in mathematics and statistics. This paper demonstrates the importance of incorporating various risk measures in portfolio optimization. This paper proposes Mean-Variance Optimization, CVaR optimization, Risk Parity optimization, and Max Sharpe ratio optimization. It also presents their performance with the given data and the developed models, and then compare the models with the assessment criteria. Finally, the paper suggests which optimization model is most effective and efficient to use under which circumstances.

Contents

1 Introduction

The use of mathematical optimization in constructing portfolios has been widely adopted in both academic and industry settings after Harry Markowitz published his paper, "Portfolio Selection", in 1952. In this project, we will make use of some well-established investment strategies to develop optimal portfolio models and investigate their performance under different market environments.

The main objective of this project is to investigate the impact of various optimization models on portfolio construction. Specifically, we will delve into three optimization models, namely conditional value-at-risk optimization (CVaR), risk parity portfolio optimization, and maximizing the Sharpe ratio. These models will be explored in greater detail to understand their effects and implications. The first model, CVaR, utilizes scenario-based optimization, whereby we will employ Monte Carlo simulations to generate these scenarios. On the other hand, risk parity and maximum Sharpe ratio are parametric optimization models that rely on asset expected returns and covariance matrix as inputs.

In this project, the period under examination spans from January 1st, 1992, to December 31st, 2016. The dataset comprises monthly-adjusted closing prices, wherein adjustments have been made to account for factors such as dividends, stock splits, and the risk-free rate. The report focuses on 20 assets, which are summarized in *Table 1* below.

Table 1: List of 20 assets by ticker

F	CAT	DIS	MCD	KO	PEP	WMT	C	WFC	JPM
AAPL	IBM	PFE	JNJ	XOM	MRO	ED	T	VZ	NEM

The provided factors are summarized in *Table 2*, including an extra column representing the risk-free rate. By employing the factor models and considering the monthly-adjusted returns, we calculate the covariance among the assets as well as the covariance among the factors. These covariance values, when combined, serve as input for our portfolio optimization process. The lookback period for calibration is fixed at 5 years.

Table 2: List of eight factors

Market ('Mkt_RF')	Size ('SMB')	Value ('HML')	Short-term reversal ('ST_Rev')
Profitability ('RMW')	Investment ('CMA')	Momentum ('Mom')	Long-term reversal ('LT_Rev')

In addition, this project utilize four different factor models to measure asset returns and covariance matrix:

- Ordinary least squares regression on all eight factors (OLS model)
- Least absolute shrinkage and selection operator model (LASSO model)
- Best subset selection model (BSS model)
- Principal component analysis (PCA model)

Subsequently, we will analyze the in-sample effects of our portfolios, and then we will assess the out-of-sample effects of our portfolios, with a specific focus on the following aspects:

- Comparing Sharpe ratio, CVaR with different confidence intervals and target returns for each portfolio.
- Examining the wealth evolution over time for each of the four portfolios.
- Selection of optimal Models and their test performance.

2 Factor Models

All of the following factor models are based on the following assumptions of an ideal multi-factor environment:

- $\text{cov}(f_i, \epsilon_j) = 0$ for all i, j
- $\text{cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$
- $\text{cov}(f_i, f_j) = 0$, for all $i \neq j$

where f_i is a factor and ϵ_i is the corresponding stochastic error term of f_i.

Note that the factor models used in this report do not respect the ideal environment, as a number of factors used are in fact correlated. For example, Size and Value, both being subsets of the market, are correlated.

2.1 OLS Model

Ordinary least squares (OLS) method is a common statistical technique for estimating the relationship between one or more independent quantitative variables and a dependent variable by the principle of least squares, i.e. minimizing the sum of the squares of the differences between the observed dependent variable and the linear output from the independent variables.

This analysis begins by examining a multi-factor OLS model, using all eight factors provided in *Table 2*. For the case of the OLS model, the return of asset i can be described by the following relationship:

$$r_i - r_f = \alpha_i + \textstyle\sum_{k=1}^{8} \beta_{ik} f_k + \epsilon_i$$

for $i \in \{1, \dots, 20\}$ where r_i is the return of asset i, r_f is the risk-free rate, f_k is the the return of factor k, α_i and β_{ik} are the intercept from regression and corresponding factor loading, respectively, and ϵ_i is the stochastic error term of the asset. We can then calibrate the OLS model by choosing a set of optimal regression coefficients that satisfy the minimization problem seen below.

$$\begin{aligned}
\min_{\mathbf{B}_i} \quad \|r_i - XB_i\|^2 &= \min_{\mathbf{B}_i} \quad (r_i - XB_i)^\top (r_i - XB_i) \\
&= \min_{\mathbf{B}_i} \quad r_i^\top ri - 2r_i^\top XB_i + B_i^t opXB_i + B_i^\top X^\top XB_i \\
&= \min_{\mathbf{B}_i} \quad f(B_i)
\end{aligned}$$

where $\mathbf{B}_i = [\alpha_i \quad \mathbf{V}_i^\top]^\top$ is the vector of regression coefficients, $\mathbf{V}_i = [\beta_{i1} \quad \beta_{i2} \quad \cdots \quad \beta_{i8}]^\top$ is the factor loadings, and $\mathbf{X} = [\mathbf{1} \quad \mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_8]$.

This is a convex quadratic unconstrained minimization problem, thus only the first order necessary conditions must be satisfied to ensure a global minimum. Therefore, the optimal regression coefficients will satisfy the following system of equations.

$$\begin{aligned}
\bigtriangledown f(B_i) &= -2X^\top r_i + 2X^\top X B_i = 0 \\
2X^\top X B_i &= 2X^\top r_i \\
B_i &= (X^\top X)^{-1} X^\top r_i
\end{aligned}$$

This implies a closed form solution to the OLS is possible with optimal regression coefficients B^* given by:

$$\therefore B^* = (X^\top X)^{-1} X^\top r$$

Finally, the residuals resulting from the optimal regression coefficients are:

$$\epsilon_i = \mathbf{r}_i - \mathbf{X}\mathbf{B}_i^*$$

which can be used to calculate the unbiased estimate of the residual variance:

$$\sigma_{\epsilon_i}^2 = \frac{1}{T-p-1} \|\epsilon_i\|_2^2$$

where we divide by the appropriate number of degrees of freedom $T - p - 1$. This is a result from having T observations but calculating $p + 1$ coefficients (one α and p number of β).

2.2 LASSO Model

One problem with including a large pool of factors is that we may run into risk of overfitting, thus making the models very sensitive to noise generated from estimating the parameters. Additionally, it reduces the interpretability of the model as accurate risk attribution is difficult when we have a very large number of factors in the model. One approach to avoid this is by introducing regularization to promote sparsity and reduce statistical overfitting.

Here, we implemented a penalized form of the LASSO (least absolute shrinkage and selection operator) model, one of the two sparse models in this report seen below:

$$\min_{\mathbf{B}_i} \quad \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 + \lambda \|\mathbf{B}_i\|_1$$

where $\mathbf{r}_i$ is the return of asset i, λ is the penalty parameter of our choice, and $\mathbf{B}_i$ is the vector of regression coefficients. Note that λ is selected by hyperparameter tuning and the optimal value of λ is determined to be 0.029, which is discussed in further detail in *Section* **??**. (Will figure out optimal value and section)

The ℓ_1 norm used by the LASSO model is continuous and convex but is not smooth everywhere. In order to address this issue that arises from the absolute value of $\mathbf{B}_i$, we make the following substitutions:

$$\begin{aligned}
\mathbf{B}_i &= \mathbf{B}_i^+ - \mathbf{B}_i^- \\
\|\mathbf{B}_i\|_1 &= \mathbf{B}_i^+ + \mathbf{B}_i^-
\end{aligned}$$

Using the substitutions, the penalized objective function of the LASSO model becomes solvable:

$$
\begin{aligned}
& \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 + \lambda\|\mathbf{B}_i\|_1 \\
= \quad & (r_i - \mathbf{X}\mathbf{B}_i)^\top (r_i - \mathbf{X}\mathbf{B}_i) + \lambda\|\mathbf{B}_i\|_1 \\
= \quad & r_i^\top r_i - 2r_i^\top \mathbf{X}\mathbf{B}_i + \mathbf{B}_i^\top \mathbf{X}^\top \mathbf{X}\mathbf{B}_i + \lambda\|\mathbf{B}_i\| \\
= \quad & r_i^\top r_i + \left[\begin{pmatrix} -2r_i^\top \mathbf{X} \\ 2r_i^\top \mathbf{X} \end{pmatrix} + \begin{pmatrix} \lambda \\ \vdots \\ \lambda \end{pmatrix} \right]^\top \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix} + \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & -\mathbf{X}^\top\mathbf{X} \\ -\mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix}
\end{aligned}
$$

Let $\mathbf{H} = \mathbf{2} \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & -\mathbf{X}^\top\mathbf{X} \\ -\mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{X} \end{bmatrix}, \quad \mathbf{c}^\top = \left[\begin{pmatrix} -2r_i^\top \mathbf{X} \\ 2r_i^\top \mathbf{X} \end{pmatrix} + \begin{pmatrix} \lambda \\ \vdots \\ \lambda \end{pmatrix} \right]^\top.$

The original problem is now in the standard form of MATLAB *QuadProg*:

$$
\begin{aligned}
\min_{\mathbf{B}_i^+, \mathbf{B}_i^-} \quad & \frac{1}{2} \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix}^\top \mathbf{H} \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix} + \mathbf{c}^\top \begin{bmatrix} \mathbf{B}_i^+ \\ \mathbf{B}_i^- \end{bmatrix} \\
\text{s.t.} \quad & \mathbf{B}_i^+, \mathbf{B}_i^- \geq \mathbf{0}.
\end{aligned}
$$

Note that $r_i^\top r_i$ is a constant term so it can be omitted in the objective function of the minimization problem. This standard form of *QuadProg* gives an equivalent solution to the original quadratic programming problem with linear constraints containing absolute values [1].

2.3 BSS Model

The best subset selection (BSS) model serves a similar purpose as the LASSO method where both models promote true sparsity. However, the BSS model introduces a cardinality constraint instead of penalizing $\mathbf{B}_i$ in ℓ_1 norm. In this project, we consider the constrained form of the BSS Model using all eight factors as inputs. The model is formulated as follows:

$$
\begin{aligned}
\min_{\mathbf{B}_i} \quad & \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\
\text{s.t.} \quad & \|\mathbf{B}_i\|_0 \leq K
\end{aligned}
$$

where $K = 4$ is the cardinality constraint parameter of our choice.

Given the discontinuity and non-convexity of the ℓ_0 norm, solving the model is computationally difficult. However, we can introduce auxiliary binary variables to address the ℓ_0 norm constraint, converting the problem into a mixed-integer quadratic program (MIQP). Employing the work of Bertsimas [2], the following is the derivation:

$$
\begin{aligned}
& \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\
= \quad & (r_i - \mathbf{X}\mathbf{B}_i)^\top (r_i - \mathbf{X}\mathbf{B}_i) \\
= \quad & r_i^\top r_i - 2r_i^\top \mathbf{X}\mathbf{B}_i + \mathbf{B}_i^\top \mathbf{X}^\top \mathbf{X}\mathbf{B}_i
\end{aligned}
$$

The original problem is now in the standard form of MATLAB *QuadProg*:

$$
\begin{aligned}
&\min_{\mathbf{B}_i} && \|\mathbf{r}_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\
&s.t. && \|\mathbf{B}_i\|_0 \le K \\
\implies &\min_{\mathbf{B}_i, y_i} && \frac{1}{2}\|r_i - \mathbf{X}\mathbf{B}_i\|_2^2 \\
&s.t. && lb \cdot y_i \le B_i \le ub \cdot y_i \\
& && y_i \in \{0,1\}, \\
& && \sum y_i \le K. \\
\implies &\min_{\mathbf{B}_i, y_i} && \frac{1}{2}\begin{bmatrix}\mathbf{B}_i \\ y_i\end{bmatrix}^\top \mathbf{H} \begin{bmatrix}\mathbf{B}_i \\ y_i\end{bmatrix} + \mathbf{c}^\top \begin{bmatrix}\mathbf{B}_i \\ y_i\end{bmatrix} \\
&\text{s.t.} && \mathbf{A}\begin{bmatrix}\mathbf{B}_i \\ y_i\end{bmatrix} \le \mathbf{b} \\
& && y_i \in \{0,1\}
\end{aligned}
$$

Where $\mathbf{H} = 2\begin{bmatrix}\mathbf{X}^\top\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}$, $c = \begin{bmatrix}-2r_i^\top\mathbf{X} \\ \mathbf{0}\end{bmatrix}$, $\mathbf{A} = \begin{bmatrix}-\mathbf{I}_d & lb \cdot \mathbf{I}_d \\ \mathbf{I}_d & -ub \cdot \mathbf{I}_d \\ \mathbf{0}_d & \mathbf{1}_d\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}0 \\ \vdots \\ 0 \\ K\end{bmatrix}$, $\mathbf{0}_d = \begin{bmatrix}0 \\ \vdots \\ 0\end{bmatrix}^\top$, $\mathbf{1}_d = \begin{bmatrix}1 \\ \vdots \\ 1\end{bmatrix}^\top$, $lb = -30$, $ub = 30$, and d is one greater than the number of parameters.

An upper bound and lower bound, ub and lb respectively, are selected to be sufficiently large but do not encapsulate the noise from integer programming in MATLAB. See [2] for further details.

2.4 Principal Component Analysis

In this investigation we implemented Principal Component Analysis (PCA) to construct our own factor model. We used PCA with the goal of creating a set factors intrinsic to our specific asset data set and reducing the dimensionality of the unknown feature space. More, specifically we used PCA to select the three most representative principal components as our factor returns and ignored the rest. Below is a detailed outline of our PCA implementation.

1. Compute the expected return of all assets $\boldsymbol{\alpha} = \mathbb{E}[\boldsymbol{R}]$, where $\boldsymbol{R}$ is the matrix of asset returns.
2. Compute the covariance matrix of the assets $\boldsymbol{\Sigma}$.
3. Eigendecompose the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$, where each column in $\boldsymbol{\Gamma}$ is an eigenvector γ_i normalized to 1, and $\boldsymbol{\Lambda}$ is a diagonal matrix where each diagonal element is an eigenvalue $\boldsymbol{\lambda}_i$.
4. Sort the eigenvectors by order of decreasing eigenvalue and collect them to create an ordered matrix $\boldsymbol{\Gamma}$.
5. Take the matrix of asset returns and subtract the corresponding expected return from each entry to obtain $\boldsymbol{R} - \boldsymbol{\alpha}$.
   - This ensures that each column has a mean of zero and our new matrix of asset returns is centered around the origin.

6. Compute the Principal Component matrix $\boldsymbol{p} = \boldsymbol{\Gamma}^\top(\boldsymbol{R} - \boldsymbol{\alpha})$.

   - The intuition behind this is to transform our origin-centered matrix to the eigenbasis. Since eigenvectors are orthogonal to each other, this ensures that the principal components are mutual independent.

   - We have achieved the goal of identifying the principal components of our asset return matrix. Furthermore, they are sorted in order of decreasing eigenvalue. We will select a subset of the principal components as our corresponding factor loadings.

7. Select the first p principal components from the Principal Component matrix corresponding to those with the highest eigenvalues.

   - In our case we selected $p = 3$ principal components to reduce the dimensionality of our feature space. The components with the highest corresponding eigenvalues were selected because they provide more information on the distribution of the data.

8. We can estimate the asset returns from the chosen factors as follows: $\hat{\boldsymbol{r}} = \boldsymbol{\alpha} + \boldsymbol{V}^\top \boldsymbol{f} + \boldsymbol{\epsilon}$, where $\boldsymbol{V}$ is the collection of first p eigenvectors, $\boldsymbol{f}$ is the chosen p factors, and $\boldsymbol{\epsilon}$ is the stochastic error term.

We implemented PCA analysis with the intention of creating a set of factors intrinsic to the data and reducing the dimensionality of the feature space. We have achieved this goal by transforming the feature space via PCA analysis onto the smaller space with the eigenbasis computed above. Our selected three factor returns are the most significant components of the data. Additionally the principal components are orthogonal to each other therefore, there is no correlation between the three factors, being mutually independent. the factors are also centered which means they have a mean of zero. This means that the model respects the ideal environment for a factor model.

Finally, here are the outputs of the PCA model that we obtained with the eigenbasis calculated above:

- Vector of asset expected returns, $\boldsymbol{\mu} \in \mathbb{R}^n$.
  - Note that $\boldsymbol{\mu} = \boldsymbol{\alpha} = \mathbb{E}[\boldsymbol{R}]$.
- Diagonal matrix of asset idiosyncratic variances, $\boldsymbol{D} \in \mathbb{R}^{n \times n}$.
  - For the scope of this project, we will assume that the asset residuals $\boldsymbol{\epsilon} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{0}, \boldsymbol{D})$ is normally distributed and uncorrelated. From $\boldsymbol{\epsilon_s} = \boldsymbol{R_s} - (\boldsymbol{\alpha} + \boldsymbol{V}^\top \boldsymbol{f}_s)$, we can calculate the unbiased estimate of the residual variance $\boldsymbol{\sigma}^2_{\boldsymbol{\epsilon}_s} = \frac{1}{T-p-1}\|\boldsymbol{\epsilon}_s\|_2^2$, where T is the number of observations, p is the number of factors, and $T - p - 1$ is the degree of freedom. Then the diagonal matrix of idiosyncratic variances can be constructed as follows:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{\sigma}^2_{\boldsymbol{\epsilon}_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\sigma}^2_{\boldsymbol{\epsilon}_2} & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \boldsymbol{\sigma}^2_{\boldsymbol{\epsilon}_n} \end{bmatrix}$$

- Matrix of factor loadings, $\boldsymbol{V} \in \mathbb{R}^{p \times n}$ for the first p principal components.
  - This is the first p components in $\boldsymbol{p}$.
- Factor covariance matrix, $\boldsymbol{F} \in \mathbb{R}^{p \times p}$.
  - By design, this matrix is diagonal because PCA factors are uncorrelated, thus $\boldsymbol{\sigma}_{i,j} = 0$ for $i \neq j$. Since the variance $\boldsymbol{\sigma_i^2} = \boldsymbol{\lambda}_i$, we can construct the factor covariance matrix as

follows:

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{\lambda}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\lambda}_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \boldsymbol{\lambda}_p \end{bmatrix}$$

- Asset covariance matrix, $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$
  - The asset covariance matrix can be calculated as follows: $\boldsymbol{Q} = \boldsymbol{V}^\top \boldsymbol{F} \boldsymbol{V} + \boldsymbol{D}$

3 Portfolio Optimization

3.1 Mean-Variance Optimization

Mean-variance optimization calculates the optimal portfolio by minimizing variance subject to our target expected return. Mean-variance optimization is defined as below:

$$\begin{aligned} \min_{x} \quad & x^\top \mathbf{Q} x \\ \text{s.t.} \quad & \mu^\top x \geq R \\ & \mathbf{1}^\top x = 1 \\ & x \geq \mathbf{0} \end{aligned}$$

where x is our portfolio weights, $\mathbf{Q}$ is our covariance matrix, μ is the expected returns of the assets, R is our target return, and $\mathbf{1}$ and $\mathbf{0}$ are the ones vector and zeroes vector respectively. Here, the target return is calculated as the geometric mean of the market factor for the calibration period.

This quadratic program was solved using MATLAB *QuadProg*. μ and $\mathbf{Q}$ are:

$$\begin{aligned} \mu &= \alpha + V^\top \overline{f} \\ Q &= V^\top F V + D \end{aligned}$$

where

$$V_i = \begin{bmatrix} \beta_{i1} \\ \beta_{i2} \\ \\ \beta_{pi} \end{bmatrix} \in \mathbb{R}^p, \quad D = \begin{bmatrix} \sigma^2_{\epsilon_1} & 0 & \cdots & \cdots & 0 \\ 0 & \sigma^2_{\epsilon_2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & \sigma^2_{\epsilon_{N-1}} & 0 \\ 0 & \cdots & \cdots & 0 & \sigma^2_{\epsilon_N} \end{bmatrix}$$

and $\sigma^2_{\epsilon_i}$ is the residual variance of asset i, $\sigma^2_{\epsilon_i} = \frac{1}{T-P-1}\|\epsilon_i\|_2^2$, $\epsilon_i = r_i - \mathbf{X}\mathbf{B}_i$.

MVO has been popularized for its intuitive appeal and theoretical property as the pareto-optimal in-sample allocation. Despite its wide application, it faces the following criticisms:

- Known as Markowitz's curse, it is numerically instable due to sensitivity to estimation error.
- It only considers the first two moments of the returns distribution from means and variances. It naively assumes the distribution is symmetric.
- It often produces portfolios that are over-concentrated.

While the MVO will serve as a baseline optimization model throughout the rest of the report, the following frameworks address some of the issues the MVO faces.

3.2 CVaR Optimization

3.2.1 Formulation

In previous optimization models we seek to minimize both up-side and down-side risk, however, intuitively speaking, we are not adverse to returns higher than our expected return. Therefore in this section we seek to implement an optimization model which minimizes solely down-side risk. We can define a metric value-at-risk (VaR_α) as follows:

$$VaR_\alpha(\boldsymbol{x}) = min\{\gamma \in \mathbb{R}^{n\times n} : \Psi(\boldsymbol{x},\gamma) \geq \alpha\}$$

where the cumulative distribution function, $\Psi(\boldsymbol{x},\gamma)$ is $\Psi(\boldsymbol{x},\gamma) = \int_{f(\boldsymbol{x},\boldsymbol{r})<\gamma} p(\boldsymbol{r})d\boldsymbol{r}$. Note: $f(\boldsymbol{x},\boldsymbol{r})$ is the loss in the event of our random asset return $\boldsymbol{r}$, and γ is the loss. Additionally, $p(\boldsymbol{r})$ is the probability density function of our vector of random asset returns.
VaR_α measures the minimum loss that we will match or exceed with probability $1-\alpha$. However, minimizing VaR_α is a non-convex activity, i.e. it is a difficult optimization problem. Hence we introduce a new variable conditional value-at-risk, defined below:

$$CVaR_\alpha(\boldsymbol{x}) = \frac{1}{1-\alpha}\int_{f(\boldsymbol{x},\boldsymbol{r})\geq VaR_\alpha(\boldsymbol{x})} f(\boldsymbol{x},\boldsymbol{r})p(\boldsymbol{r})d\boldsymbol{r}$$

This new variable measures the expected value of the losses that are greater than or equal to VaR_α. Some problems arise, however, as if we try to minimize $CVaR_\alpha(\boldsymbol{x})$, it still contains the $VaR_\alpha(\boldsymbol{x})$ term which is non-convex. In order to amend this, we introduce the auxiliary variable γ and function $F_\alpha(\boldsymbol{x},\gamma)$, defined below:

$$F_\alpha(\boldsymbol{x},\gamma) = \gamma + \frac{1}{1-\alpha}\int (f(\boldsymbol{x},\boldsymbol{r})-\gamma)^+ \, p(\boldsymbol{r}d\boldsymbol{r}$$

Where $(f(\boldsymbol{x},\boldsymbol{r})-\gamma)^+ = max\{0, f(\boldsymbol{x},\boldsymbol{r})-\gamma\}$, which can be replaced with the auxiliary variable z_s, defined below:

$$\begin{aligned} z_s &\geq 0, \quad s = 1,..,S \\ z_s &\geq f(\boldsymbol{x},\boldsymbol{r_s})-\gamma, \quad s = 1,..,S \end{aligned}$$

Finally, since $p(\boldsymbol{x})$ is difficult to solve for analytically, we will implement a scenario based approach using Monte Carlo simulations, defined in the following section. These scenarios, $\hat{\boldsymbol{r}}_{\boldsymbol{s}}$ for $s = 1,..,S$, are determined to be equally likely, therefore we can simplify the auxiliary function $F_\alpha(\boldsymbol{x},\gamma)$ as the estimation $\hat{F}_\alpha(\boldsymbol{x},\gamma)$, and with z_s introduced, it is defined below:

$$\hat{F}_\alpha(\boldsymbol{x},\gamma) = \gamma + \frac{1}{(1-\alpha)S}\sum_{s=1}^{S} z_s$$

Rendering our final $CVaR_\alpha(\boldsymbol{x})$ optimization model:

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z},\gamma} \quad & \gamma + \frac{1}{(1-\alpha)S}\sum_{s=1}^{S} z_s & \\
s.t. \quad & z_s \geq 0, & s = 1,..,S \\
& z_s \geq f(\boldsymbol{x}, \boldsymbol{r_s}) - \gamma, & s = 1,..,S \\
& \boldsymbol{x} \in \mathcal{X} &
\end{aligned}$$

where $\mathcal{X}$ are the budget and target return constraints. While CVaR optimization is similar to the MVO, it addresses the issue of the MVO that it naively assumes the returns distribution symmetry by using a tail-based risk measure. CVaR technique can especially come in handy when the client is interested in preventing worst-case scenarios.

3.2.2 Monte Carlo Simulations

An input unique to our $CVaR_\alpha(\boldsymbol{x})$ optimization model is $\hat{\boldsymbol{r}}_{\boldsymbol{s}}$, since we seek to form a discrete set of scenarios which mimic the probability distribution function $p(\boldsymbol{x})$. In order to get these scenarios, Monte Carlo simulations will be run under the assumption of Gaussian process.

The scenarios for the Gaussian process, $\boldsymbol{f}_s$, can be drawn from the multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{F})$, where $\boldsymbol{F}$ is the covariance matrix of the PCA factors. The Gaussian process is centered at $\boldsymbol{0}$ as the PCA factors have the mean $\boldsymbol{\mu} = 0$. The second random sampling comes from asset idiosyncratic noise, $\boldsymbol{\epsilon}_s \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$, where $\boldsymbol{D}$ is the $n \times n$ diagonal matrix of asset idiosyncratic variances, again from PCA.

This method for Monte Carlo simulation is then used to generate the asset returns scenarios, $\hat{\boldsymbol{r}}_s$, for $s = 1,..,S$ using the following equation:

$$\hat{\boldsymbol{r}}_s = \boldsymbol{\mu} + \boldsymbol{V}^T \boldsymbol{f}_s + \boldsymbol{\epsilon}_s$$

Note that the multivariate normal distributions are generated using MATLAB's built-in function *mvnrnd()*. 10,000 scenarios were run for each model and each asset.

3.3 Risk Parity Portfolio

The purpose of risk parity optimization, also known as equal risk contribution, is to generate a portfolio with equally distributed risk by allocating capital such that the contributions of risk from any one asset is equal to the contributions of another.

One of the key advantages of this optimization technique is that our portfolio is not subject to error in the estimations of expected returns, as the model minimizes solely based on the convariance matrix $\boldsymbol{Q}$. This optimization method looks to find a portfolio such that $\boldsymbol{R}_i = \boldsymbol{R}_j \ \forall \ i, j$, where $\boldsymbol{R}_i$ is the individual risk contribution of asset i, and $\boldsymbol{R}_i = \boldsymbol{x}^\top \boldsymbol{A}_i \boldsymbol{x}$. $\boldsymbol{A}_i \in \mathrm{I\!R}^{n \times n}$ is the individual risk contribution of asset i derived from $\boldsymbol{Q}$, defined below:

$$
\boldsymbol{A}_i = \begin{bmatrix} 0 & \cdots & 0 & \frac{1}{2}\sigma_{1i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 \\ \frac{1}{2}\sigma_{i1} & \cdots & \cdots & \frac{1}{2}\sigma_{ii} & \cdots & \cdots & \frac{1}{2}\sigma_{in} \\ 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{2}\sigma_{ni} & 0 & \cdots & 0 \end{bmatrix},
$$

An issue arises, however, as each $\boldsymbol{A}_i$ is indefinite, having its own positive eigenvalue and negative eigenvalue, with all other eigenvalues being zero, leading to non-convexity in minimization function, shown below, and many local minimum.

$$
\min_{\mathbf{x}} \quad \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^\top \boldsymbol{A}_i x_i) - (x_j^\top \boldsymbol{A}_j x_j))^2
$$

One workaround to this is to disallow short selling, resulting in the new optimization problem that is well-behaved within this smaller feasible region:

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^\top \boldsymbol{A}_i x_i) - (x_j^\top \boldsymbol{A}_j x_j))^2 \\
s.t. \quad & \mathbf{1}^\top \boldsymbol{x} = 1 \\
& \boldsymbol{x} \geq \mathbf{0}
\end{aligned}
$$

Some issues still remain, such as the 4^{th} degree polynomial being extremely non-linear. One solution to this, and what will be employed in this project, is rendering a convex model using some decision variable $\boldsymbol{y} \in \mathbb{R}^n$. Now define the function of this decision variable as:

$$
f(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} - c\sum_{i=1}^{n} \ln y_i
$$

where c is some arbitrary positive scalar. Since $\boldsymbol{Q}$ is PSD and natural logarithmic function is strictly concave, $f(\boldsymbol{y}$ is a strictly convex function, allowing us to take gradient, set it to zero and achieve a few desirable properties:

$$
\bigtriangledown f(\boldsymbol{y}) = \boldsymbol{Q}\boldsymbol{y} - c\boldsymbol{y}_i^{-1} = 0
$$

where $\boldsymbol{y}^{-1} = [\frac{1}{y_1}, ..., \frac{1}{y_n}]$. Therefore:

$$
\begin{aligned}
(\boldsymbol{Q}y)_i &= \frac{c}{y_i} \quad \forall i \\
(y_i \boldsymbol{Q}y)_i &= c \quad \forall i \\
(y_i \boldsymbol{Q}y)_i &= (y_j \boldsymbol{Q}y)_j \quad \forall i, j
\end{aligned}
$$

Achieving our desired outcome. Prohibiting short selling, and recovering the optimal portfolio weights after minimizing over our decision variable yields the risk parity optimization problem, implemented in MATLAB:

$$\begin{aligned}
\min_{\mathbf{y}} \quad & f(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} - c\sum_{i=1}^{n} \ln y_i \\
s.t. \quad & \boldsymbol{y} \geq \boldsymbol{0} \\
& x_i^* = \frac{y_i^*}{\sum_{i=1}^{n} y_i^*} \quad \forall i
\end{aligned}$$

Risk parity optimization method addresses the issue of the MVO that it often produces over-concentrated portfolios by intentionally distributing equal amount of risk contribution to each asset. Hence solutions tend to be very diversified. It also has an interesting property that the projected volatility of RP portfolio is between that of minimum-variance portfolio and equally-weighted portfolio, i.e. $\sigma_{MV} \leq \sigma_{RP} \leq \sigma_{EW}$.

3.4 Sharpe Ratio Optimization

Sharpe ratio can be thought of as the reward vs. risk of a given asset or portfolio, formally defined as:

$$SR_p = \frac{[r_i] - r_f}{\sqrt{\mathrm{var}(r_i - r_f)}} = \frac{\boldsymbol{\mu}^\top \boldsymbol{x} - r_f}{\sqrt{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}}$$

This function, however, is not linear, and a transformation must be made to amend this which is as follows:

$$\kappa = \frac{1}{(\boldsymbol{\mu} - \overline{\boldsymbol{r}}_f)^\top \boldsymbol{x}}, \quad \boldsymbol{y} = \kappa \boldsymbol{x}$$

Substituting this into the vector notation of SR_p yields:

$$\begin{aligned}
SR_p &= \frac{(\boldsymbol{\mu} - \overline{\boldsymbol{r}}_f)^\top \boldsymbol{x}}{\frac{1}{\kappa}\sqrt{\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y}}} \\
&= \frac{1}{\sqrt{\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y}}}
\end{aligned}$$

We can now maximize the Sharpe Ratio by minimizing $\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y}$ over $\boldsymbol{y}, \kappa$, and the constraints become:

$$(\boldsymbol{\mu} - \overline{\boldsymbol{r}}_f)^\top \boldsymbol{y} = 1, \boldsymbol{1}^\top \boldsymbol{y} = \kappa$$

Rending the following LP:

$$\begin{aligned}
\min_{\boldsymbol{y},\kappa} \quad & \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} \\
s.t. \quad & (\boldsymbol{\mu} - \overline{\boldsymbol{r}}_f)^\top \boldsymbol{y} = 1 \\
& \boldsymbol{1}^\top \boldsymbol{y} = \kappa \\
& \kappa \geq 0 \\
& \boldsymbol{y} \geq \boldsymbol{0}
\end{aligned}$$

Finally, the optimal portfolio weights may be recovered as:

$$x_i^* = \frac{y_i^*}{\kappa^*} = \frac{y_i^*}{\sum_{j=1}^{n} y_j^*}$$

Note that SR optimization model still yields a solution on the efficient frontier, but the one that has the greatest Sharpe ratio. The solution is sometimes referred to as the tangency portfolio as it is the tangent point between the capital allocation line (CAL) and the efficient frontier.

4 Factor Model Hyperparameter Tuning

Note: *Dataset 1* has been used for factor model hyperparameter tuning.

4.1 LASSO Model: Regularization Coefficient, λ

In *Section 2.2*, we discussed the penalized form of LASSO model with the ℓ_1 norm penalty parameter λ left undetermined. In this section we explore the most optimal value of λ through hyperparameter tuning.

When searching for the optimal λ, we seek to satisfy two criteria:

1. Ideally, the factor model selects two to five factors for each asset at any given period. Put differently, the number of non-zero factor loadings should be between two and five. The purpose of a sparse factor model is defeated when more than five factors are incorporated; at the other extreme, having less than two factors likely results in an underfitted model.

2. Given that the above condition is met, a higher correlation between the factors and observed returns is desirable.

Let us define a variable N_B: the number of assets which has two to five non-zero factor loadings under the LASSO factor model for a given λ value. A perfect model would yield $N_B = 20$, whereas a low N_B value indicates a poor sparsity. An initial search for $\lambda \in [0.01, 0.1]$ at an increment of 0.01 by computing average N_B values eliminates λ outside of the range $[0.03, 0.07]$, as summarized in *Table 3*, since we can reject average N_B values less than 16.

Subsequently, we perform a finer search within the range $\lambda = [0.03, 0.07]$ by calculating the average in-sample adjusted R^2 using the equations given below:

$$R^2_{adj} = 1 - (1 - R^2)\tfrac{n-1}{n-p-1}$$

$$\mathrm{R}^2 = 1 - \tfrac{SS_{res}}{SS_{tot}}, \quad SS_{tot} = \sum(y_i - \overline{y})^2, \quad SS_{res} = \sum(y_i - f_i)^2 = \sum \epsilon_i^2$$

where f_i is the predicted return of the asset at time i, $\overline{y}$ is the mean return for the asset over the investment period, and y_i is the actual return for the asset at time i. In general, R^2 either increases or remains the same when new predictors are introduced to the model, but high R^2 can indicate a problem in the model. This can cause misinterpretation of the model that is explained by variables. On the other hand, adjusted R^2 increases only when independent variable is significant and affects dependent variable. It penalizes for introducing extra predictor variables that don't improve the existing model.

Observe from *Figure 1* that as the value of λ increases, adjusted R^2 drops uniformly, manifesting an inverse relationship. Intuitively when we enlarge the regularization factor λ, the model is penalized harsher for non-zero factor loadings, yielding in a distorted objective function. This pushes the regularized least-squares optimization problem away from its optimal solution under no penalty terms, resulting in decreased explanatory power.

We choose the optimal λ value to be 0.03. This ensures that we have between 2 and 5 factors selected into the LASSO model. Out of the examined set of parameter values, $\lambda = 0.03$ stood out as the optimal value that yields the best adjusted R^2.

Table 3: Average N_B for varying λ

λ	Average N_B
0.01	6.05
0.02	14.05
0.03	16.6
0.04	17.45
0.05	17
0.06	17
0.07	16.75
0.08	15.95
0.09	14.6
0.1	13.85

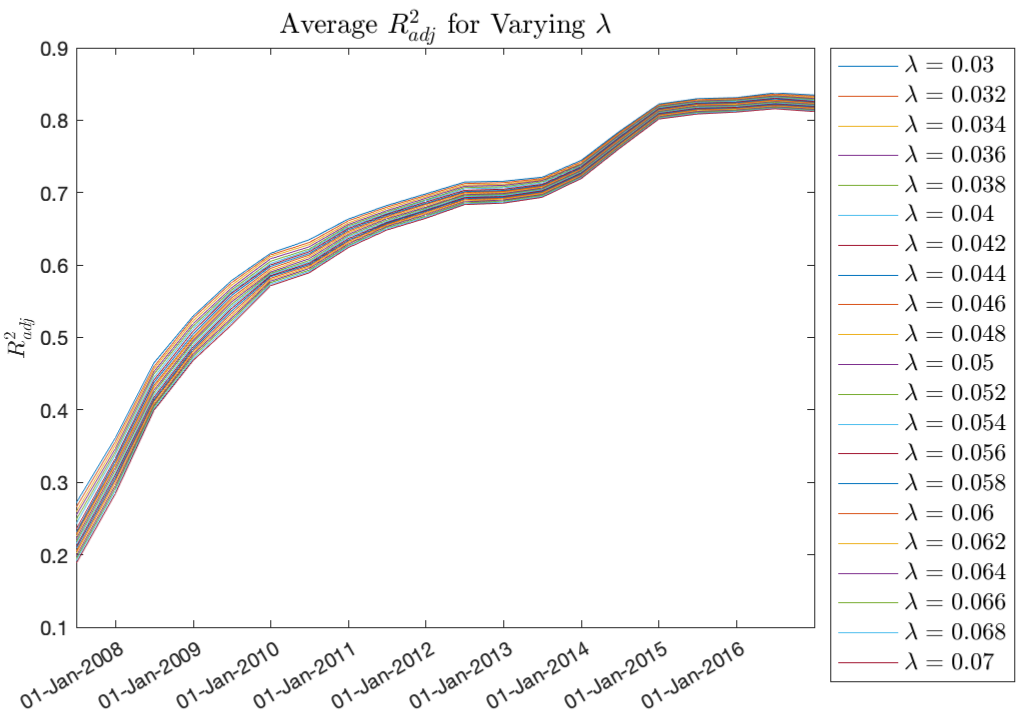


Figure 1: $\overline{R}^2_{adj}$ for $\lambda \in [0.03, 0.07]$ over 20 periods

4.2 BSS Model: Cardinality Constraint Parameter, K

Unlike the LASSO model, we cannot simply search for the optimal hyperparameter of the BSS model using in-sample correlations as BSS models impose strict cardinality constraints on the number of factors. For a linear regression model, adding a predictor variable can only increase the in-sample correlation coefficient; however, this is not necessarily desirable since it may lead to overfitting.

Instead, a heuristic out-of-sample analysis using the baseline MVO model can assist the search for the optimal K value. We repeatedly feed BSS factor models with varying $K \in \{2, 3, 4, 5\}$ into the baseline MVO, described in *Section 3.1*, and measure its ex-post Sharpe ratio over the entire investment horizon. We do not consider turnover as one of our metrics as MVO is known to generate over-concentrated portfolios. Hence even a reliable factor model may still produce high turnover rate. As shown in *Table 4*, the Sharpe ratio is maximized when $K = 5$, which will be our choice of the hyperparameter. This is evident in *Figure 11* as the plot of $K = 5$ reaches the greatest total wealth at the end of the investment horizon while resembling the volatility of the other graphs.

Table 4: Sharpe ratio of MVO portfolios generated using BSS factor model with varying $K \in \{2, 3, 4, 5\}$

K	Sharpe Ratio
2	0.17781
3	0.16591
4	0.17480
5	0.18990

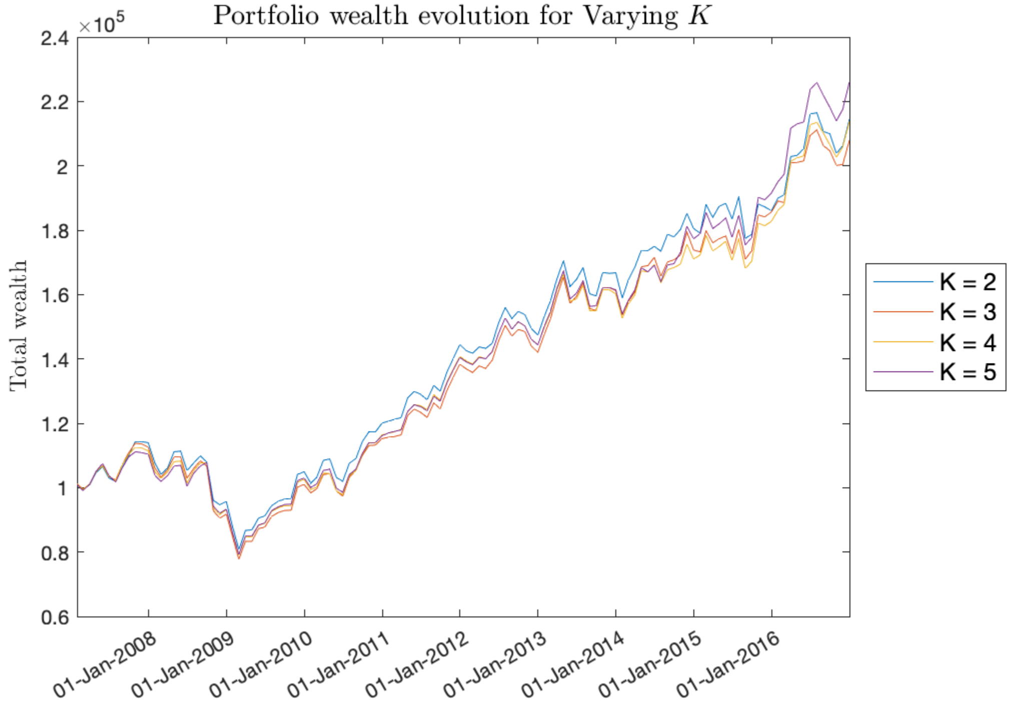


Figure 2: Portfolio wealth evolution for different $K \in \{2, 3, 4, 5\}$ over 20 periods

4.3 PCA Model: Number of Principal Components p

In this section we examine how well the PCA factor model fits the observed asset returns and attempt to discover the optimal number of principal components p. To quantify the goodness of fit, the proportion of variance explained, η^2 (analogous to R^2, the coefficient of determination, in multi-linear regression), can be employed. The formula for this can be seen below:

$$\eta_1^2 = \frac{\lambda_1}{\lambda_1 + \lambda_2 + ... + \lambda_P}$$
$$\eta_2^2 = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + ... + \lambda_P}$$
$$\vdots$$
$$\eta_P^2 = \frac{\lambda_1 + \lambda_2 + ... + \lambda_P}{\lambda_1 + \lambda_2 + ... + \lambda_P} = 1$$

where η_α^2 denotes $\sum_{i=1}^{\alpha} \eta_i^2$, and λ_j denotes the eigenvalue of the j^{th} factor, or covariance, of the resulting principal component in ascending order. The cumulative, as well as individual η^2 values are summarized below in scree plots for each investment period in *Figure 3*.

We choose $p = 3$ and the scree plots in *Figure 3* justify this decision via the elbow method, which looks to place the cutoff point for no further addition of factors where the scree plots have the largest change in slope, i.e. the elbow. By inspecting *Figure 3*, the elbows (cut-off points) are between factors 2 and 4, and the addition of factors 4 to 20 clearly having diminishing returns in terms of substantially increasing the cumulative η^2 can be seen above.

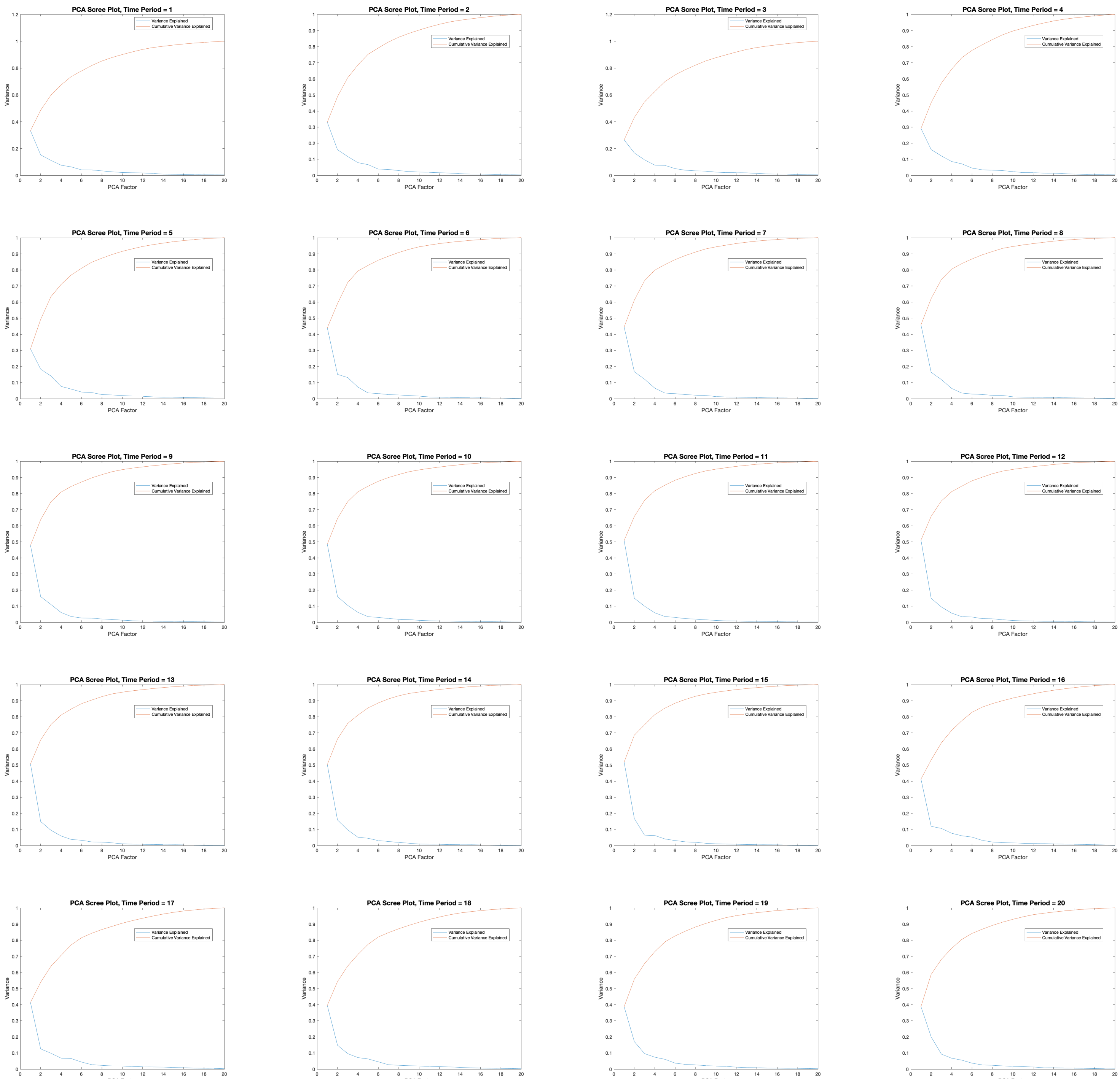


Figure 3: *PCA factors against variance at different time period*

5 Optimization Model Hyperparameter Tuning

Note: *Dataset 1* has been used for optimization model hyperparameter tuning.

5.1 CVaR Optimization: Confidence Interval

We begin the hyperparameter search of the CVaR model by finding the optimal confidence interval level α. By setting the target return of the model as 90% of the arithmetic mean of expected returns, $0.9 \cdot \overline{\mu}$, we iterate through all four factor models with their respective optimal hyperparameters found in *Section 4* and calculate Sharpe ratio and average turnover for different confidence level $\alpha \in \{0.7, 0.9, 0.95, 0.99\}$, summarized in *Table 5*.

Even though we are provided with two metrics – Sharpe ratio and average turnover – to optimize (former to maximize, latter to minimize), an optimization problem usually requires one and only one objective function. Thus we construct a heuristic objective $0.8 \cdot SR - 0.2 \cdot TO_{avg}$, where SR is the Sharpe ratio and TO_{avg} is the average turnover. This resembles the original problem (maximize Sharpe ratio and minimize the average turnover) while reflecting the competition grading scheme.

From *Table 6*, we observe that the score is the highest when $\alpha = 0.9$ for OLS and LASSO models and when $\alpha = 0.95$ for BSS and PCA models. This is expected since low confidence levels like $\alpha = 0.7$ carry much risk of large expected loss, resulting in high volatility and low Sharpe ratio; at the same time, very high confidience levels such as $\alpha = 0.99$ are too conservative and cannot take on high returns.

Table 5: Sharpe ratio and average turnover of CVaR optimization portfolios for varying confidence interval levels

	Sharpe Ratio				**Average Turnover**			
C.I.	**70%**	**90%**	**95%**	**99%**	**70%**	**90%**	**95%**	**99%**
OLS	0.17057	0.18923	0.18838	0.18167	0.38645	0.37283	0.3752	0.43917
LASSO	0.16149	0.17234	0.16653	0.18530	0.29555	0.26383	0.27524	0.31594
BSS	0.17591	0.17946	0.18107	0.18457	0.44719	0.41939	0.42578	0.4618
PCA	0.15735	0.16692	0.17508	0.16205	0.4264	0.38695	0.40126	0.46989

Table 6: Score of CVaR optimization portfolios for varying confidence interval levels

	Score for C.I.			
C.I.	**70%**	**90%**	**95%**	**99%**
OLS	0.059166	0.076818	0.075664	0.057502
LASSO	0.070082	0.085106	0.078176	0.085052
BSS	0.05129	0.05969	0.05970	0.055296
PCA	0.04060	0.056146	0.059812	0.035662

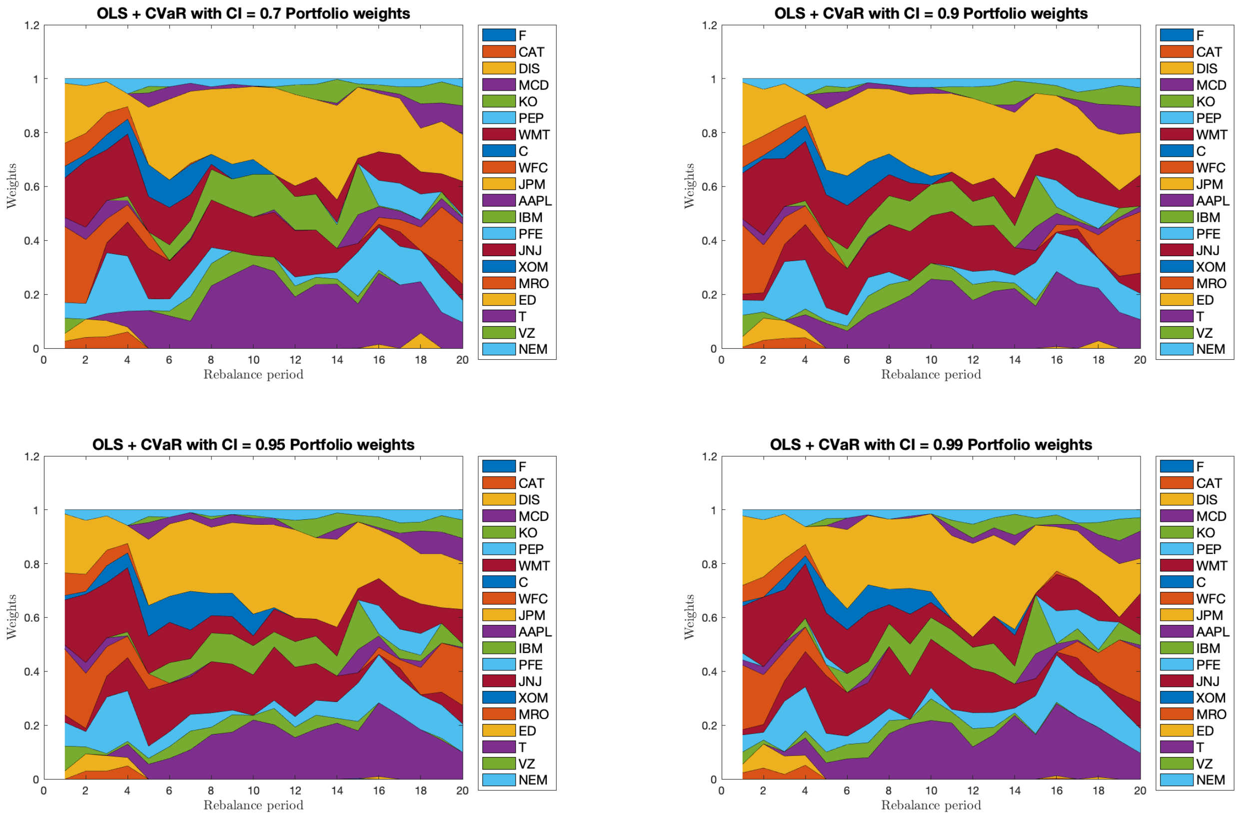


Figure 4: *OLS+CVaR Composition by different Confidence Intervals*

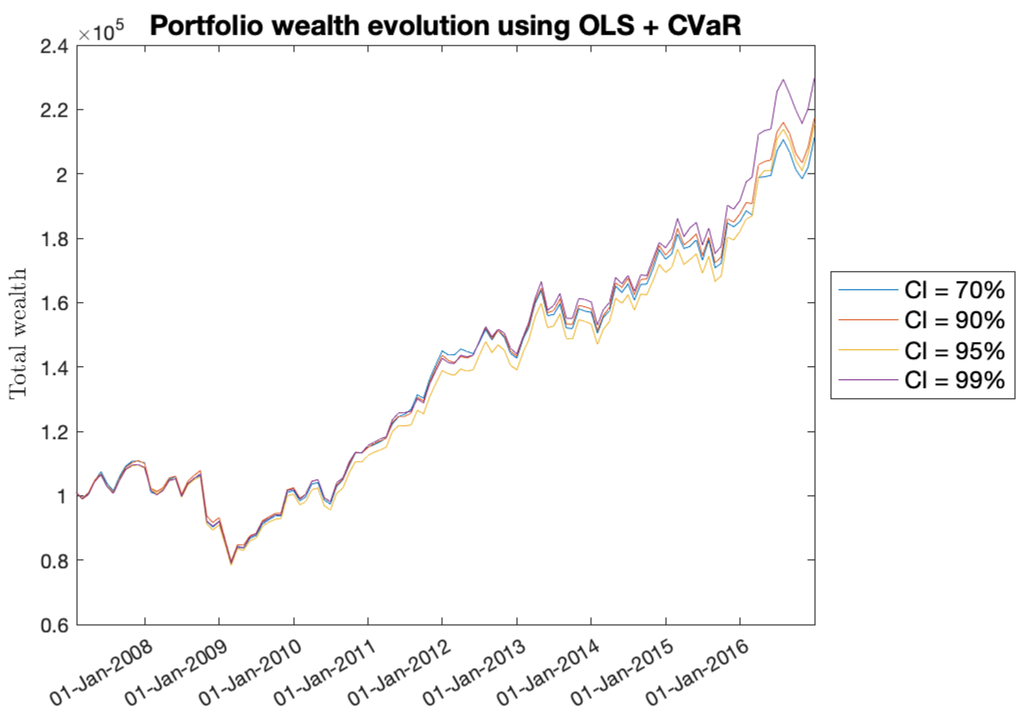


Figure 5: Portfolio Evolution of OLS+CVaR with different C.I.

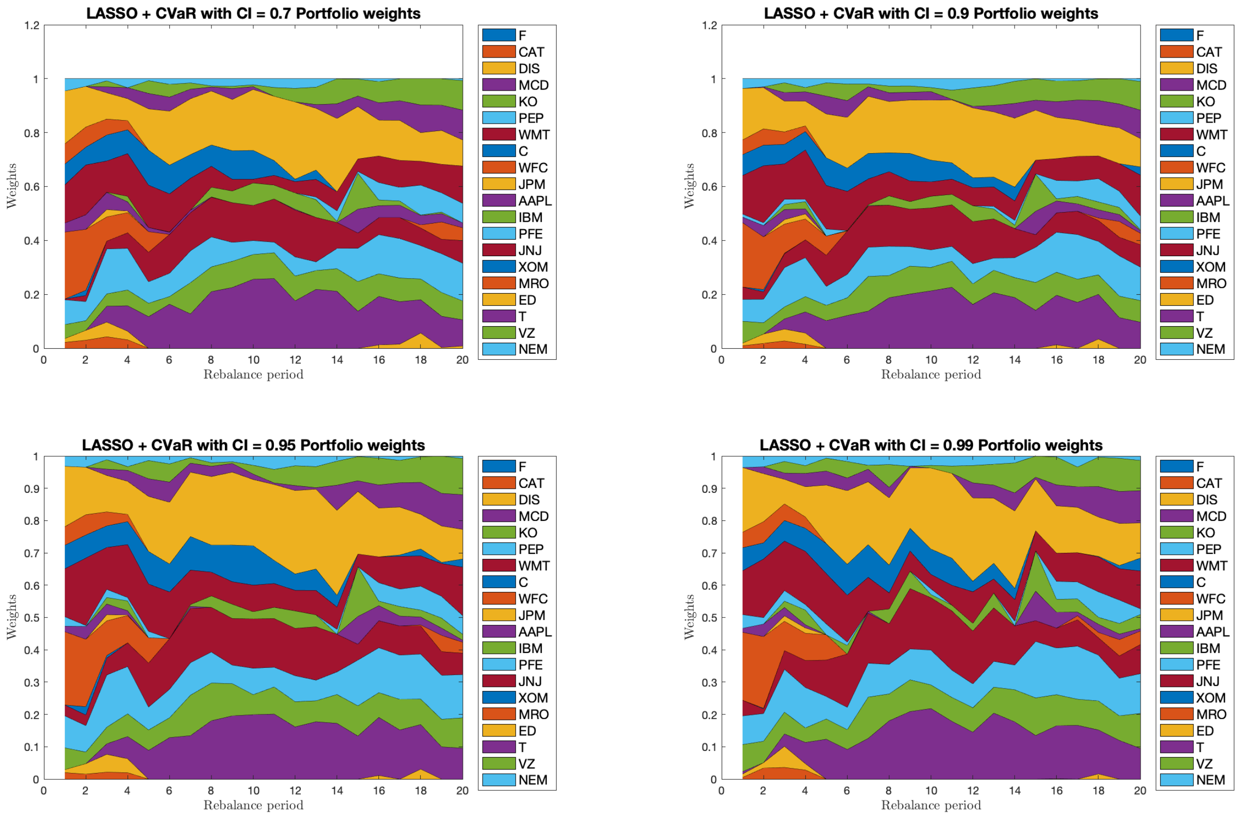


Figure 6: *LASSO+CVaR Composition by different Confidence Intervals*

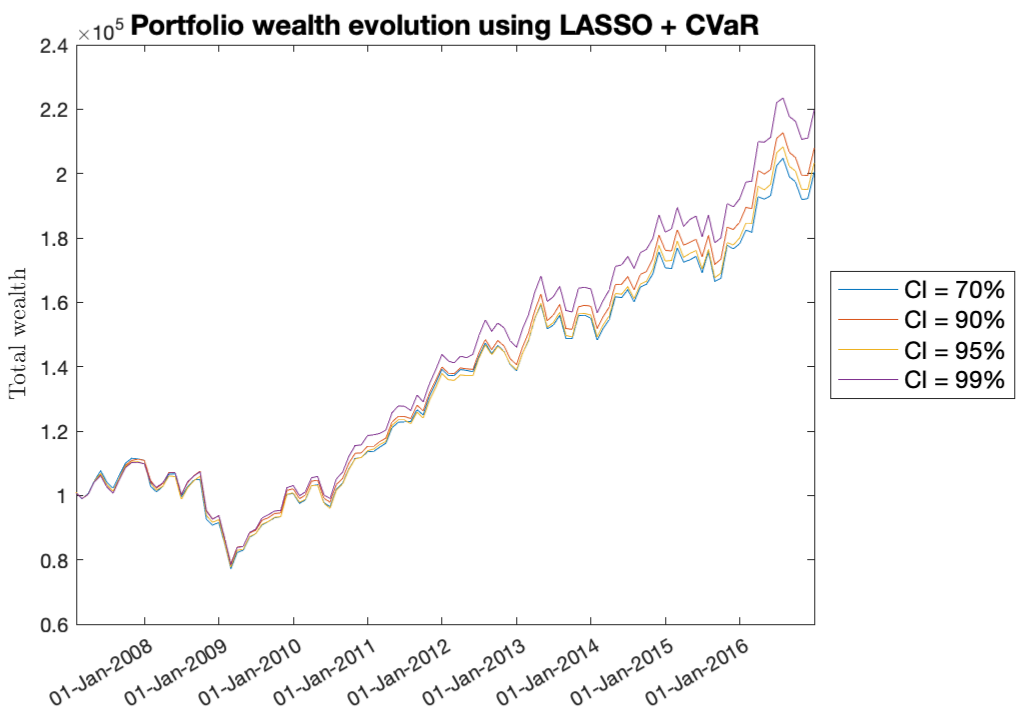


Figure 7: Portfolio Evolution of LASSO+CVaR with different C.I.

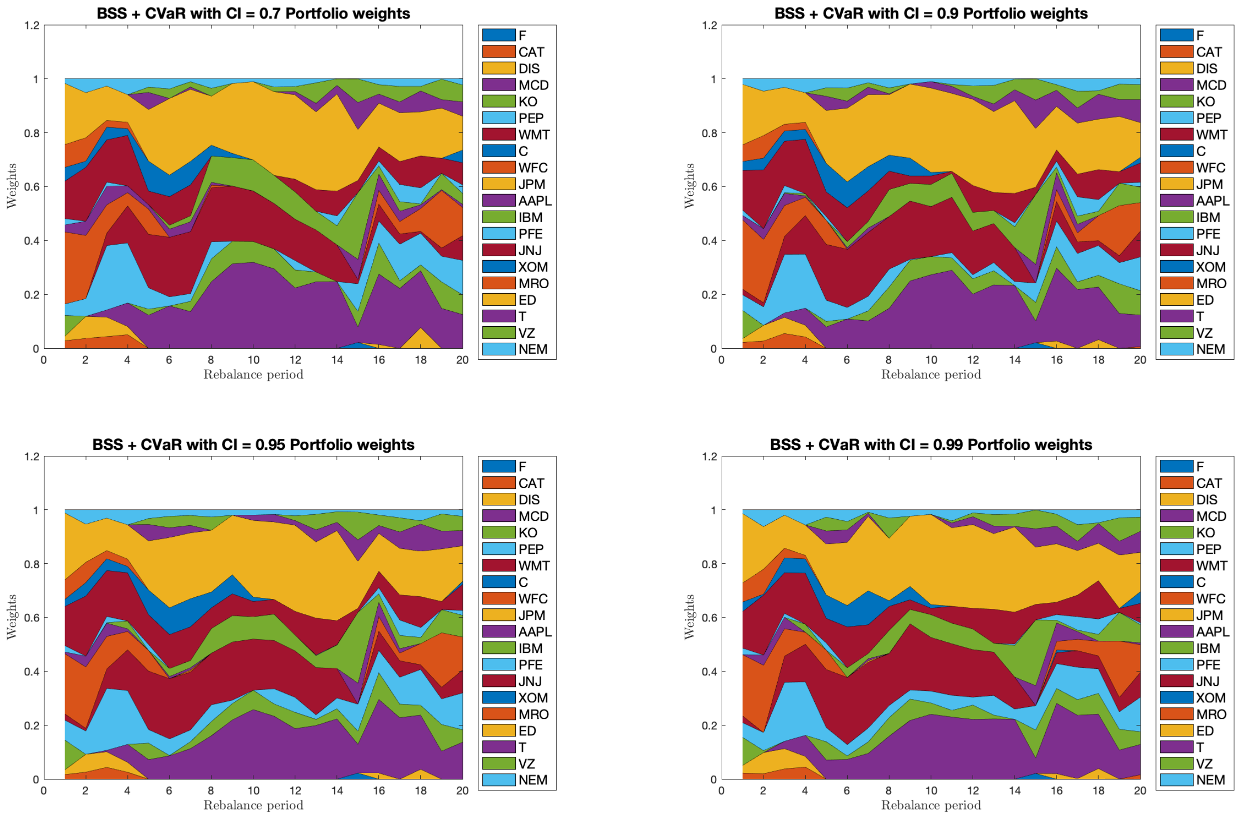


Figure 8: *BSS+CVaR Composition by different Confidence Intervals*

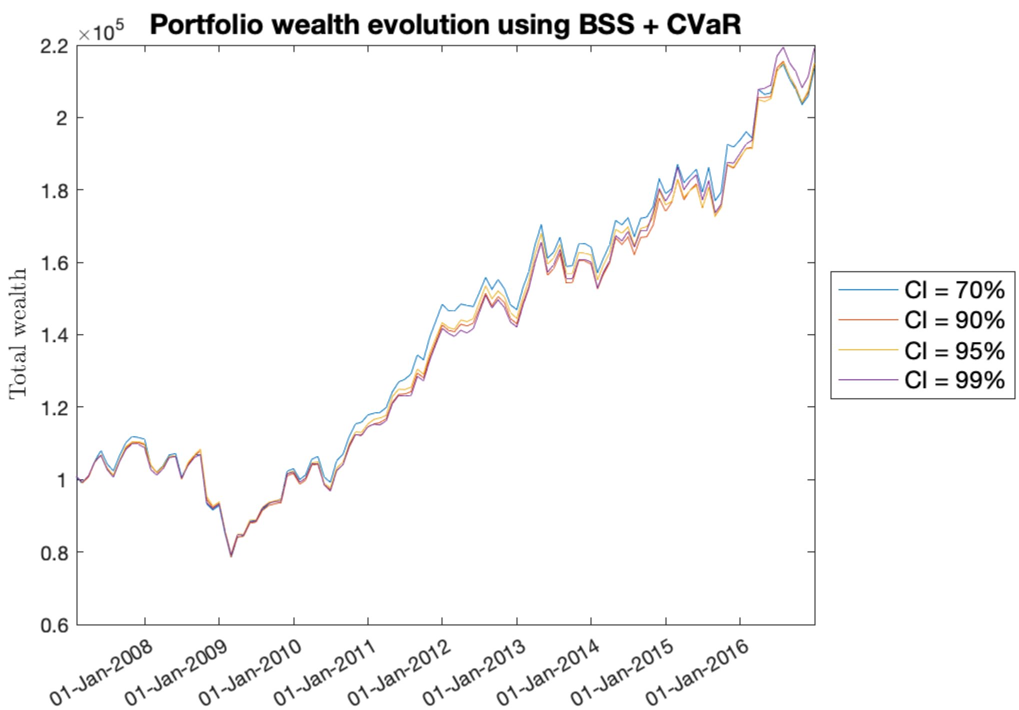


Figure 9: Portfolio Evolution of BSS+CVaR with different C.I.

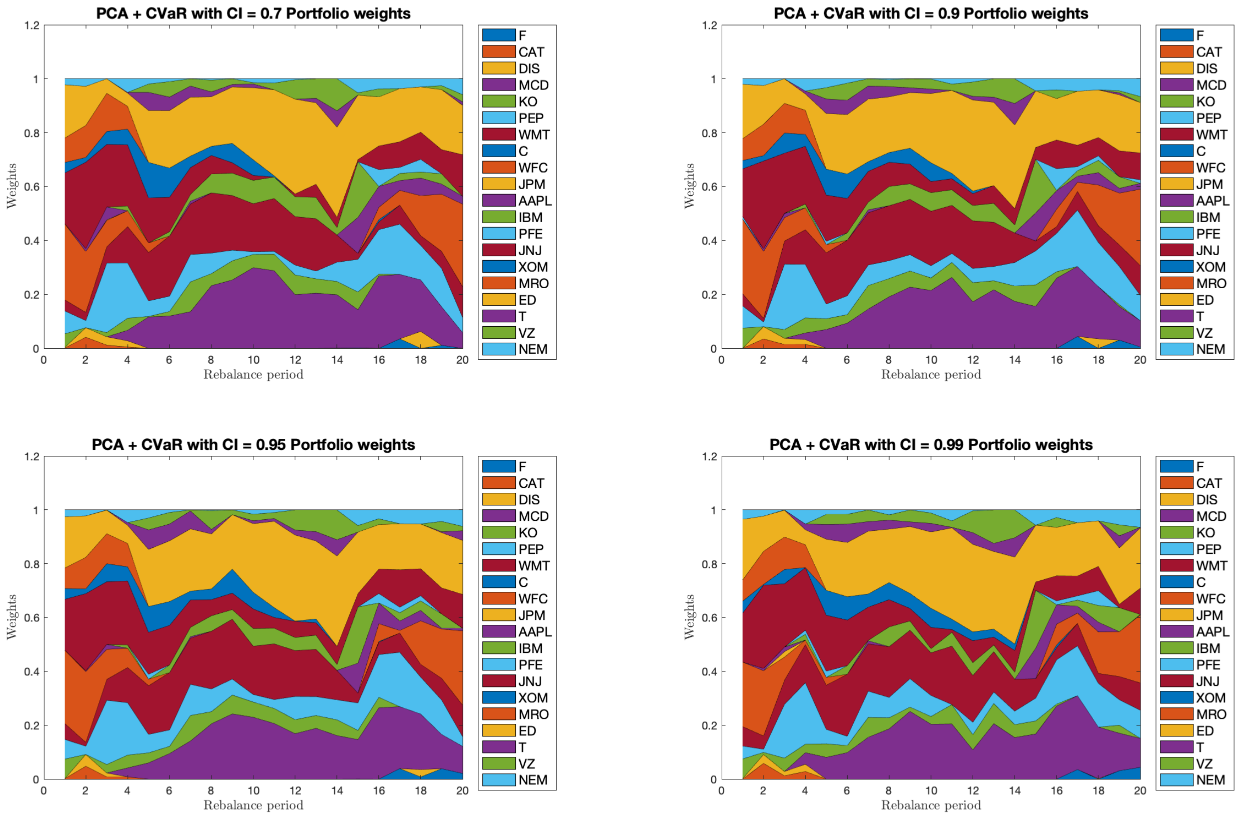


Figure 10: *PCA+CVaR Composition by different Confidence Intervals*

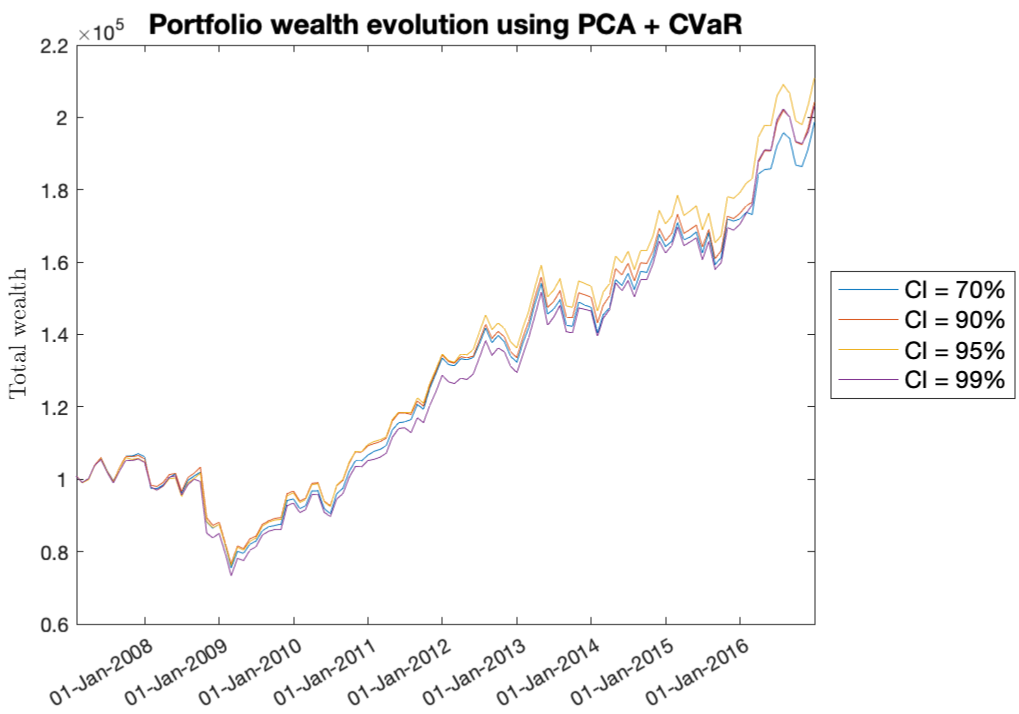


Figure 11: Portfolio Evolution of PCA+CVaR with different C.I.

5.2 CVaR Optimization: Target Return

In this section we search for the optimal value of the second hyperparameter of the CVaR optimization model: target return R_{target}. By setting the confidence level of the model α as 95%, we iterate through all four factor models with their respective optimal hyperparameters found in *Section 4* and calculate Sharpe ratio and average turnover for different confidence level $R_{target} \in \{0.7 \cdot \overline{\mu}, 0.9 \cdot \overline{\mu}, 1.0 \cdot \overline{\mu}\}$, summarized in *Table 7*.

Similar to *Section 5.1*, we employ the heuristic score function $0.8 \cdot SR - 0.2 \cdot TO_{avg}$, where SR is the Sharpe ratio and TO_{avg} is the average turnover, to consolidate the two metrics. From *Table 8*, we observe that the score is the highest when $R_{target} = 0.9 \cdot \overline{\mu}$ for OLS and when $R_{target} = 0.7 \cdot \overline{\mu}$ for LASSO, BSS and PCA models. While we did not anticipate the exact results a priori, it was expected that $R_{target} = \overline{\mu}$ would not be optimal since such constraint is too restrictive especially under bear market conditions where beating the historical market return is extremely difficult.

Table 7: Sharpe ratio and average turnover of CVaR optimization portfolios for varying target returns

	Sharpe Ratio			Average Turnover		
Mean Return	**70%**	**90%**	**100%**	**70%**	**90%**	**100%**
OLS	0.19233	0.19673	0.18658	0.36079	0.36796	0.39336
LASSO	0.16563	0.16789	0.17983	0.26436	0.28281	0.33363
BSS	0.17780	0.1809	0.18811	0.38705	0.43181	0.46311
PCA	0.17245	0.16854	0.1740	0.38296	0.40756	0.41069

Table 8: Score of CVaR optimization portfolios for varying target returns

	Score for Target Return		
Mean Return	**70%**	**90%**	**100%**
OLS	0.081706	0.083792	0.070592
LASSO	0.079632	0.07775	0.077138
BSS	0.06483	0.058358	0.057866
PCA	0.061368	0.05332	0.057062

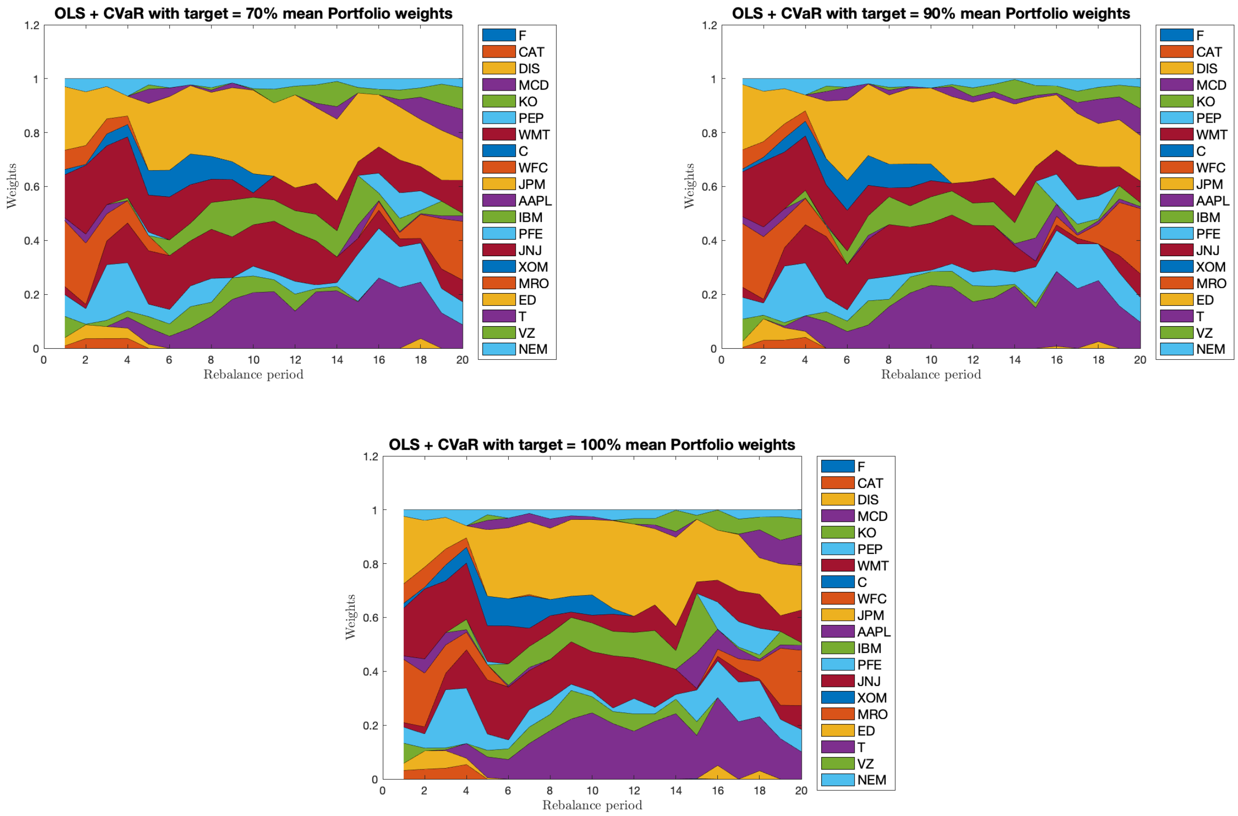


Figure 12: *OLS+CVaR Composition by different target returns*

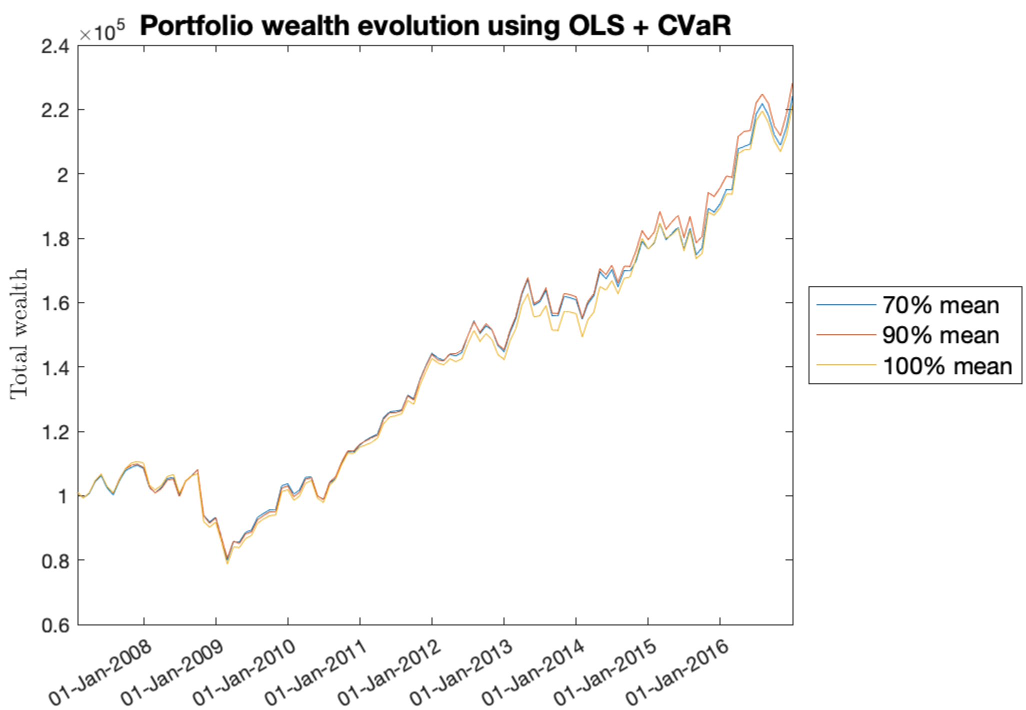
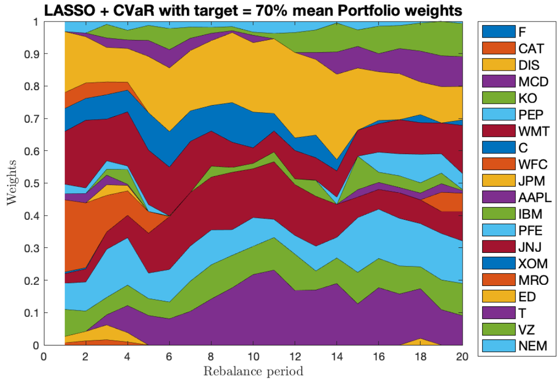
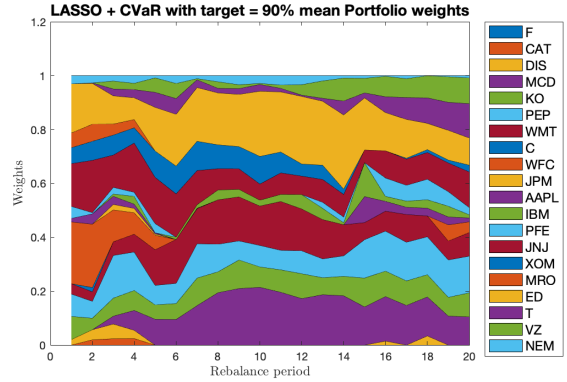


Figure 13: Portfolio Evolution of OLS+CVaR with different target returns





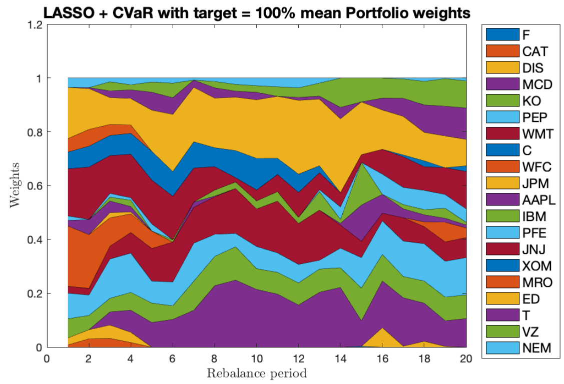


Figure 14: *LASSO+CVaR Composition by different target returns*

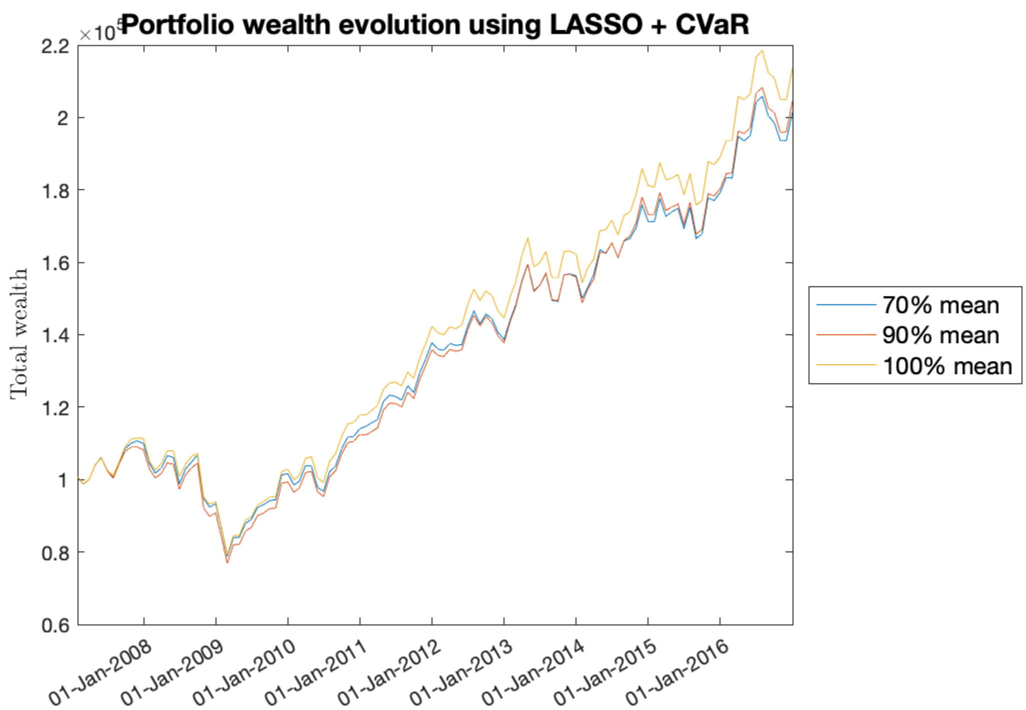


Figure 15: Portfolio Evolution of LASSO+CVaR with different target returns

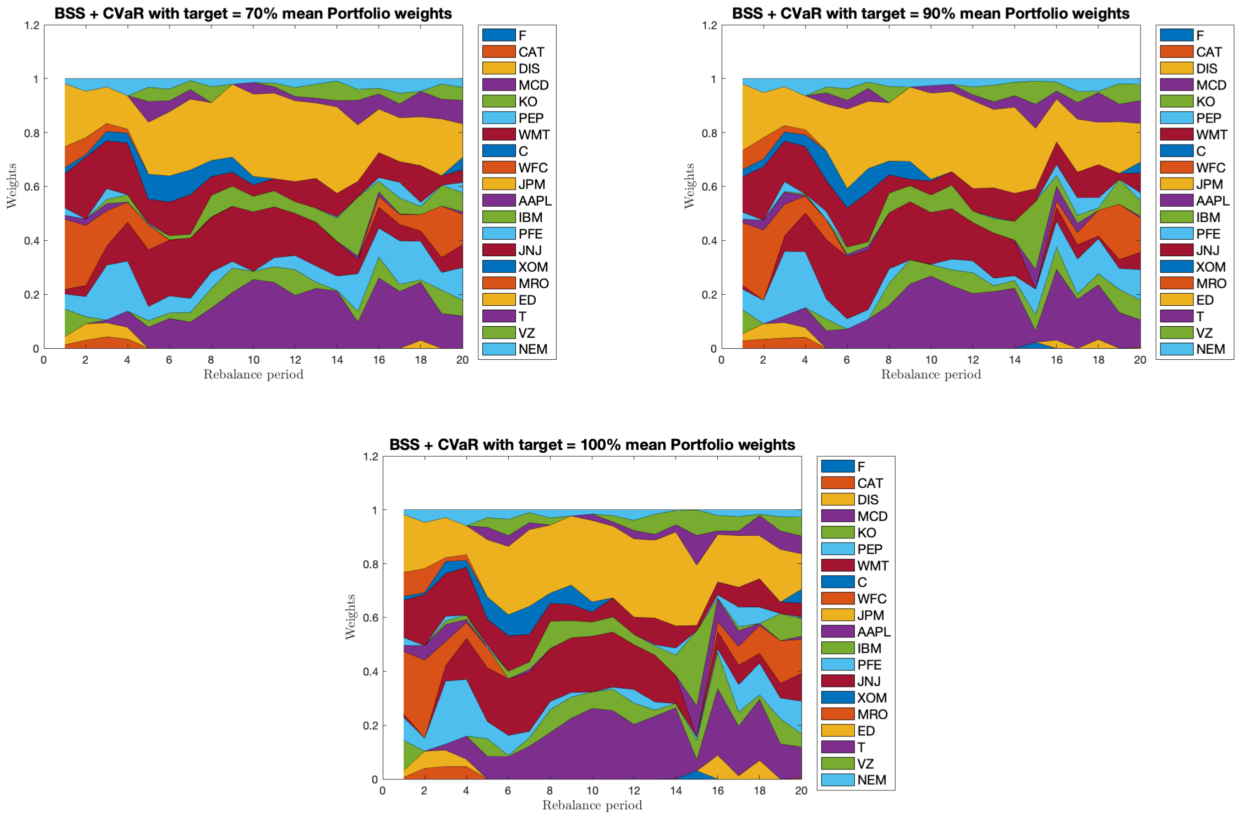


Figure 16: *BSS+CVaR Composition by different target returns*

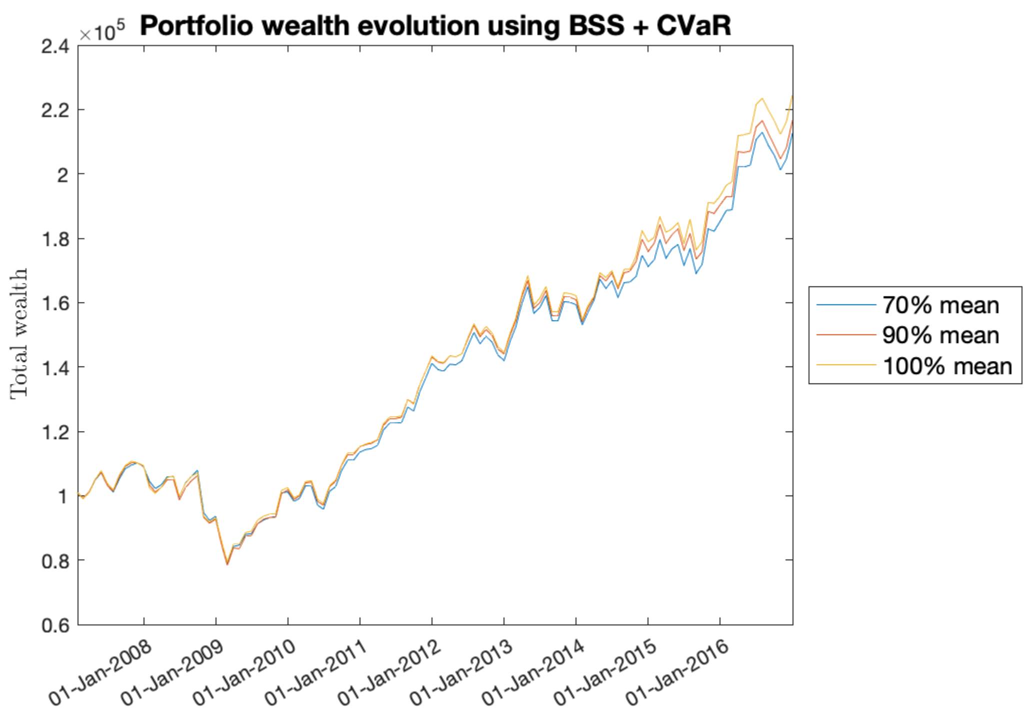


Figure 17: Portfolio Evolution of BSS+CVaR with different target returns

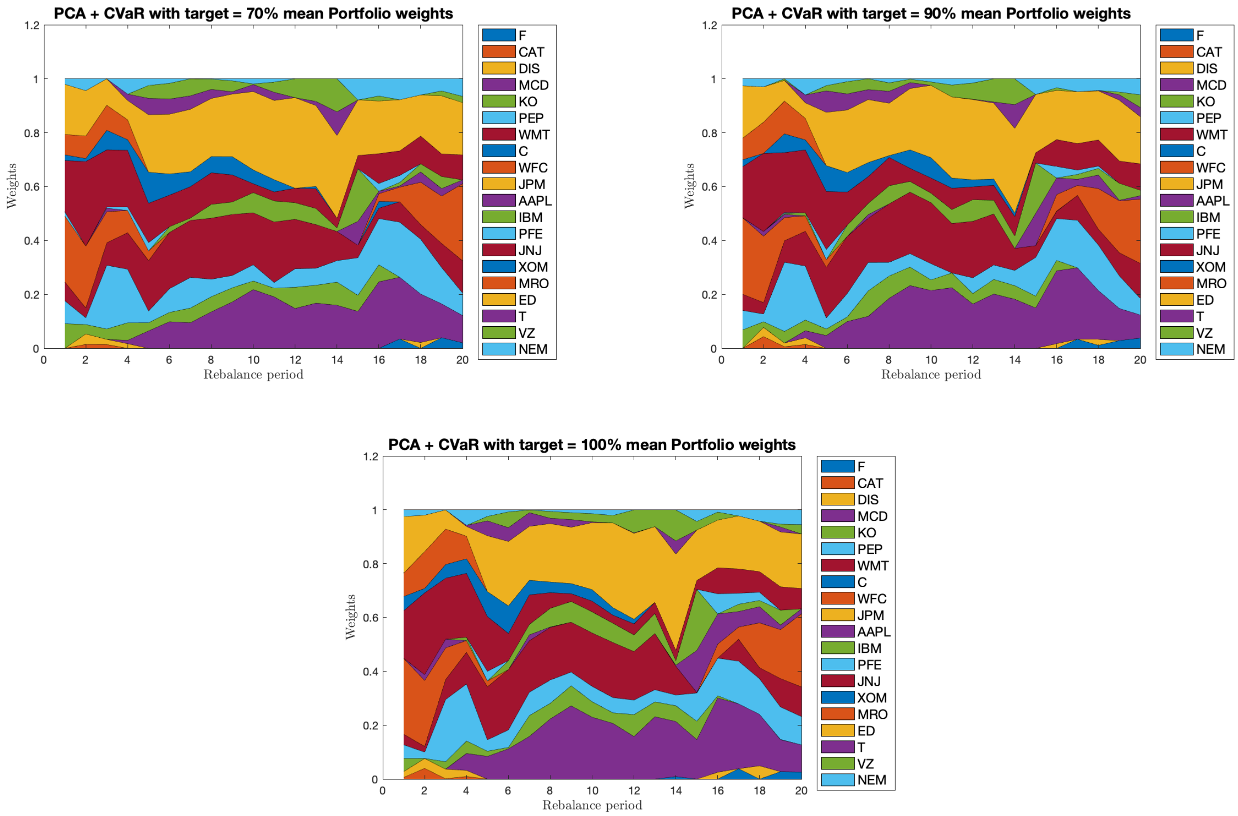


Figure 18: *PCA+CVaR Composition by different target returns*

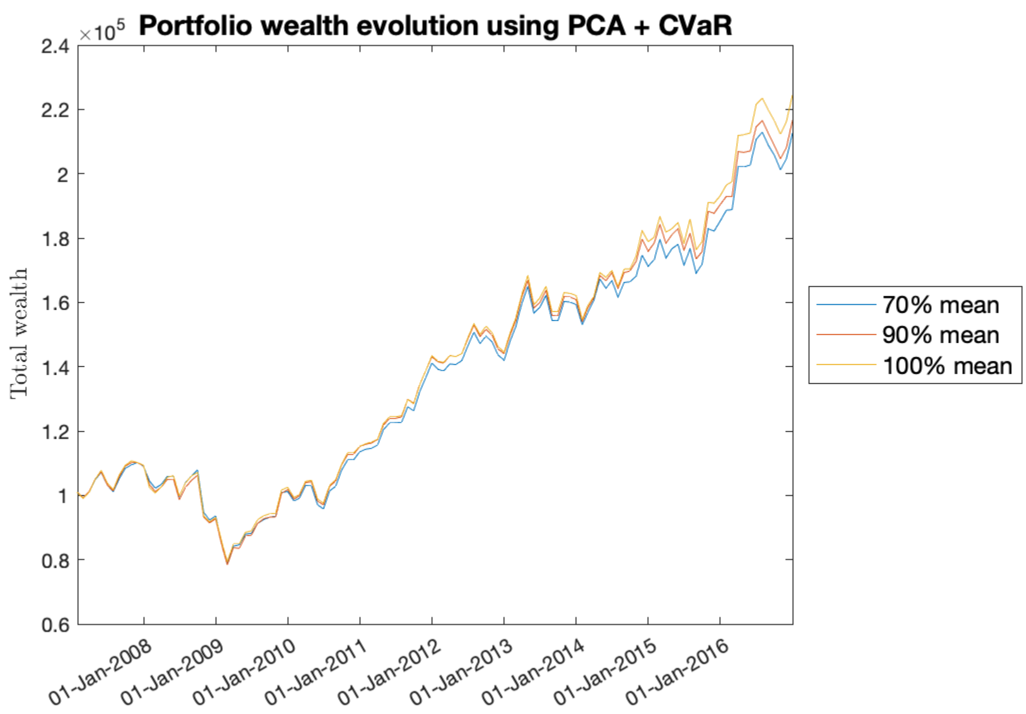


Figure 19: Portfolio Evolution of PCA+CVaR with different target returns

6 Model Selection

Note: *Dataset 3* has been used for regime detection and model selection.

6.1 Regime Detection

Financial markets often exhibit rapid changes in behavior; the switches in market regimes can be hard to observe yet they may pose challenges to the statistical assumptions of our model. In this section we attempt to identify two market regimes – a volatile regime and a stable regime – using a simple model based on asset volatilities. We exploit a naive but also a widely accepted assumption that markets tend to volatile when it is bearish. The expectation is that each regime has different optimal combination of factor model and portfolio optimization model.

We measure the overall volatility of our investment universe by first computing the variance of each traded asset from the six most recent months then averaging them as a proxy for the market volatility. By visually inspecting *Figure 20*, we can draw the threshold at 0.01% monthly volatility to divide our observation periods into the specified regimes. Evident from *Figure 20*, volatility of our investment universe (composed of 20 invested assets) spiked above the threshold of 0.01 during the Great Financial Crisis between 2008 and 2010, while the rest of the investment horizon sat below our threshold.

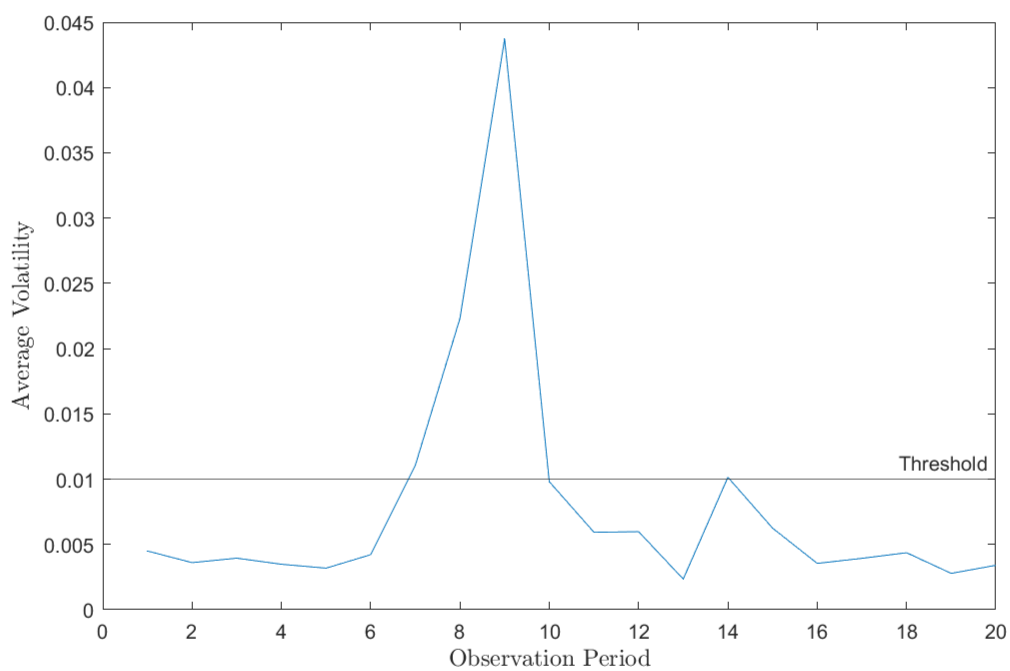


Figure 20: Average asset volatility per observation period and the threshold of 0.01

6.2 Selecting Optimal Model Pairs

By identifying an optimal pair of factor models and portfolio models for each regime, our hope is to establish a robust investment framework unaffected by market conditions. To do so, we run grid search of models on both regime separately from a previously unseen dataset. As seen in *Section 6.1*, we classified the period between July 2008 and December 2010 as volatile market and the rest as stable.

Table 9: Sharpe ratio and average turnover of portfolios under stable market condition

	Sharpe Ratio				Average Turnover			
Optimization	**MVO**	**CVaR**	**RP**	**Max S.R.**	**MVO**	**CVaR**	**RP**	**Max S.R.**
OLS	0.27633	0.25961	0.30615	0.30671	0.33343	0.31463	0.12107	0.50123
LASSO	0.28598	0.28835	0.29872	0.32021	0.28486	0.24998	0.12711	0.44611
BSS	0.2785	0.27481	0.3058	0.2944	0.34433	0.33337	0.12586	0.53518
PCA	0.2721	0.26758	0.30954	0.29917	0.32628	0.31408	0.1232	0.4898

Table 10: Scores of portfolios under stable market condition

	Score for Optimization			
Optimization	**MVO**	**CVaR**	**Risk Parity**	**Max Sharpe Ratio**
OLS	0.1544	0.1448	0.2207	0.1451
LASSO	0.1718	0.1807	0.2136	0.1669
BSS	0.1539	0.1532	0.2195	0.1285
PCA	0.1524	0.1513	0.2230	0.1414

Table 11: Sharpe ratio and average turnover of portfolios under volatile market condition

	Sharpe Ratio				Average Turnover			
Optimization	**MVO**	**CVaR**	**RP**	**Max S.R.**	**MVO**	**CVaR**	**RP**	**Max S.R.**
OLS	-0.14519	-0.14728	-0.034897	-0.096316	0.63193	0.63906	0.22158	0.69747
LASSO	-0.12807	-0.14682	-0.038917	-0.083225	0.52731	0.51095	0.24075	0.61269
BSS	-0.13893	-0.15267	-0.036176	-0.10692	0.70094	0.68606	0.22741	0.74573
PCA	-0.14451	-0.14761	-0.036272	-0.1292	.68048	0.67592	0.25407	0.86421

Table 12: Scores of portfolios under volatile market condition

	Score for Sharpe Ratio			
Optimization	**MVO**	**CVaR**	**RP**	**Max S.R.**
OLS	-0.2425	-0.2456	-0.0722	-0.2165
LASSO	-0.2079	-0.2196	-0.0793	-0.1891
BSS	-0.2513	-0.2593	-0.0744	-0.2347
PCA	-0.2517	-0.2533	-0.0798	-0.2762

We iterate through all four factor models and four optimization models with their respective optimal hyperparameter choices found in *Section 4* and *Section 5*. Similar to out-of-sample parameter selection process, we use the heuristic function $0.8 \cdot SR - 0.2 \cdot TO_{avg}$, where SR is the Sharpe ratio and TO_{avg} is the average turnover, to consolidate the two metrics.

Under the stable market scenarios, we observe that the combination of the PCA factor model and the risk parity model yields the highest score. On the other hand, the OLS factor model and the risk parity model pair outputs the best score under the volatile market condition. It is not surprising that the risk parity model, also known as the "all-weather fund", has been selected for both regimes because it diversifies across all assets such that each of them share equal risk contribution to the portfolio. In effect, the portfolio weights are very stable with minimal changes over time, resulting in low turnover. *Figure 22*, *Figure 23*, *Figure 24*, and *Figure 25* all illustrate how portfolio weights of the risk parity model remain reasonably consistent under any factor models.

One hypothesis for why the OLS model is preferred over the PCA model during volatile periods is because the observed asset returns are too noisy and do not reflect the current market situation. Hence parameter estimation, i.e. the covariance matrix Q for the risk parity model, is conducted more reliably using the factors. On the other hand, the latent principal components extracted from the returns data using the PCA model can be useful during stable periods when there exist less noise in the market.

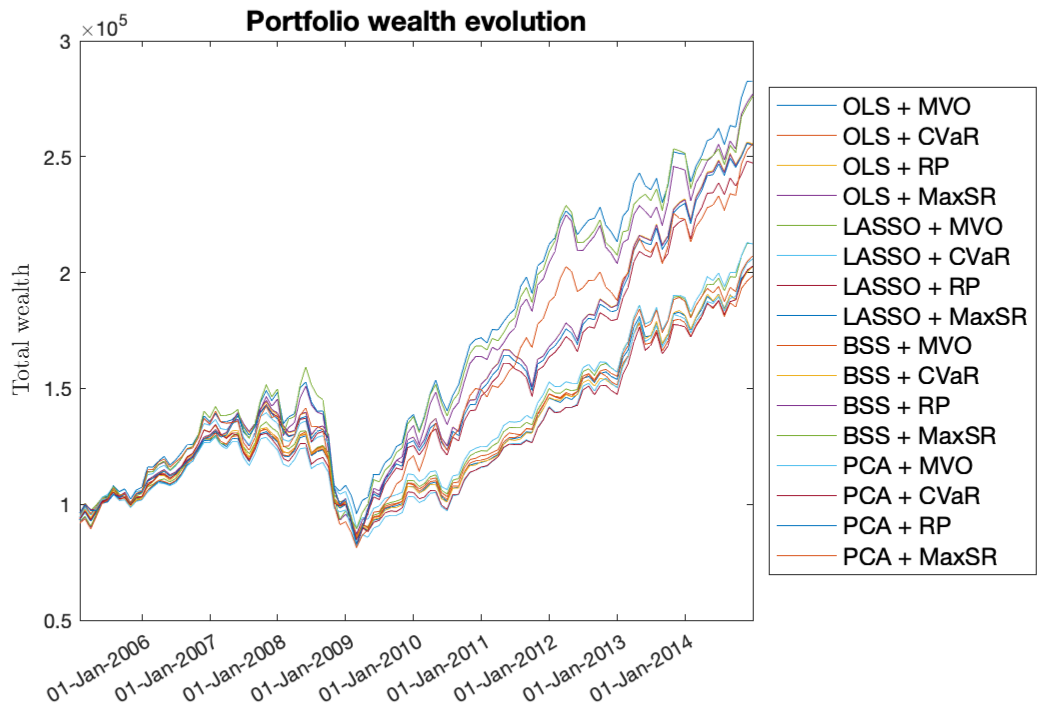


Figure 21: Portfolio evolution on a test dataset

The last bells-and-whistles of this investment framework is to take the linear combination of the portfolios generated from the two model pairs. Formally speaking, say $\mathbf{x}_{t,s}$ is the portfolio produced by the stable regime optimal model pair, i.e. PCA and risk parity model, and $\mathbf{x}_{t,v}$ is the portfolio produced by the volatile regime optimal model pair, i.e. OLS and risk parity model, at a given period t. If the average asset volatility from the past six months at time t is $\sigma^2_{t,avg}$, the final portfolio generated by the framework would be:

$$x_t = c_t \cdot x_{t,s} + (1 - c_t) \cdot x_{t,v}$$

where $c_t = \max(0, 1 - \frac{\sigma^2_{t,avg}}{0.02})$. To better see how this formula behaves, the final portfolio will be an even mix between $x_{t,s}$ and $x_{t,v}$ when the average volatility is 0.01, the threshold dividing the two regimes. Moreover when $\sigma^2_{t,avg} > 0.02$, the final portfolio is identical to $x_{t,v}$ as the market is very volatile.

We adopt this technique to minimize turnover by preventing large changes in portfolio composition. If the framework is designed as a binary decision problem, i.e. $x_{t,s}$ when in stable regime and $x_{t,v}$ when in volatile regime, it may result in high turnover if the two atomic portfolios contrast hugely. Having a mixture of the two introduces a smooth transition between regimes.

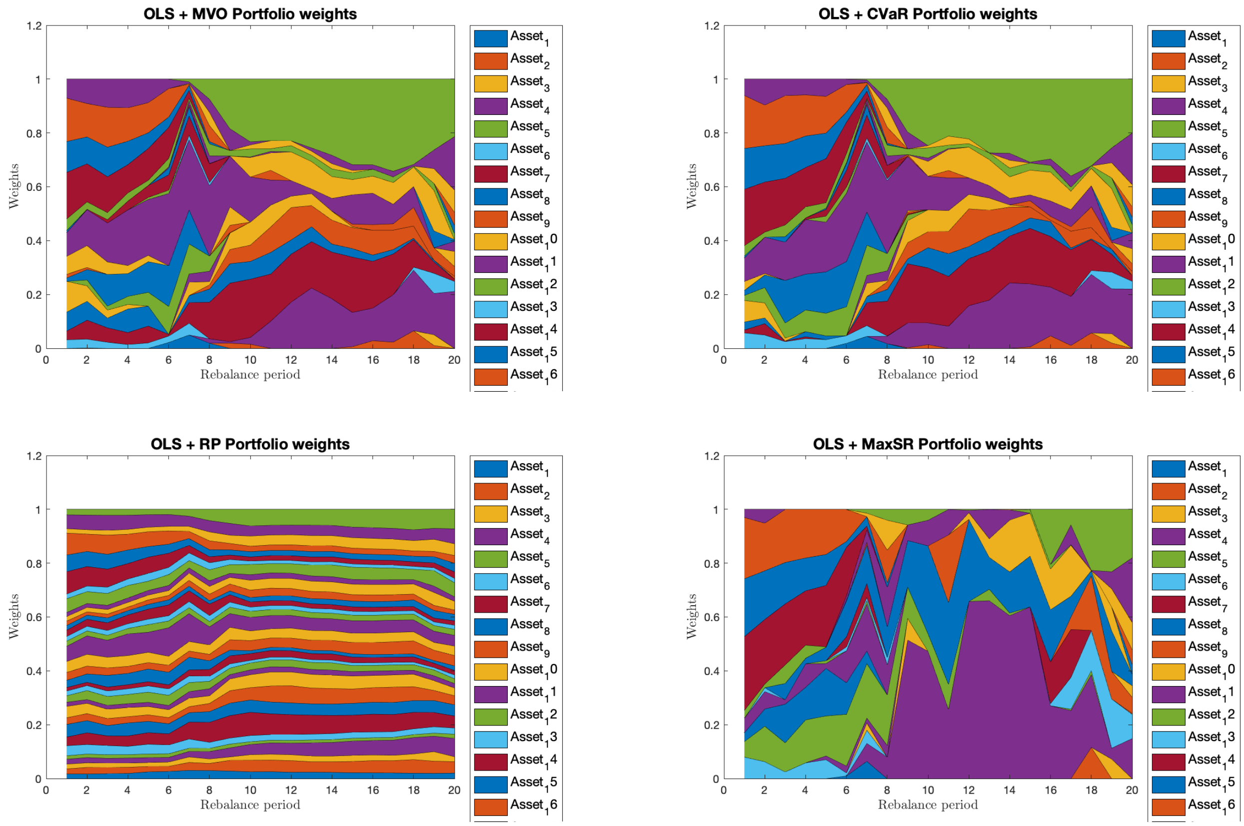


Figure 22: *Portfolio Composition using OLS*

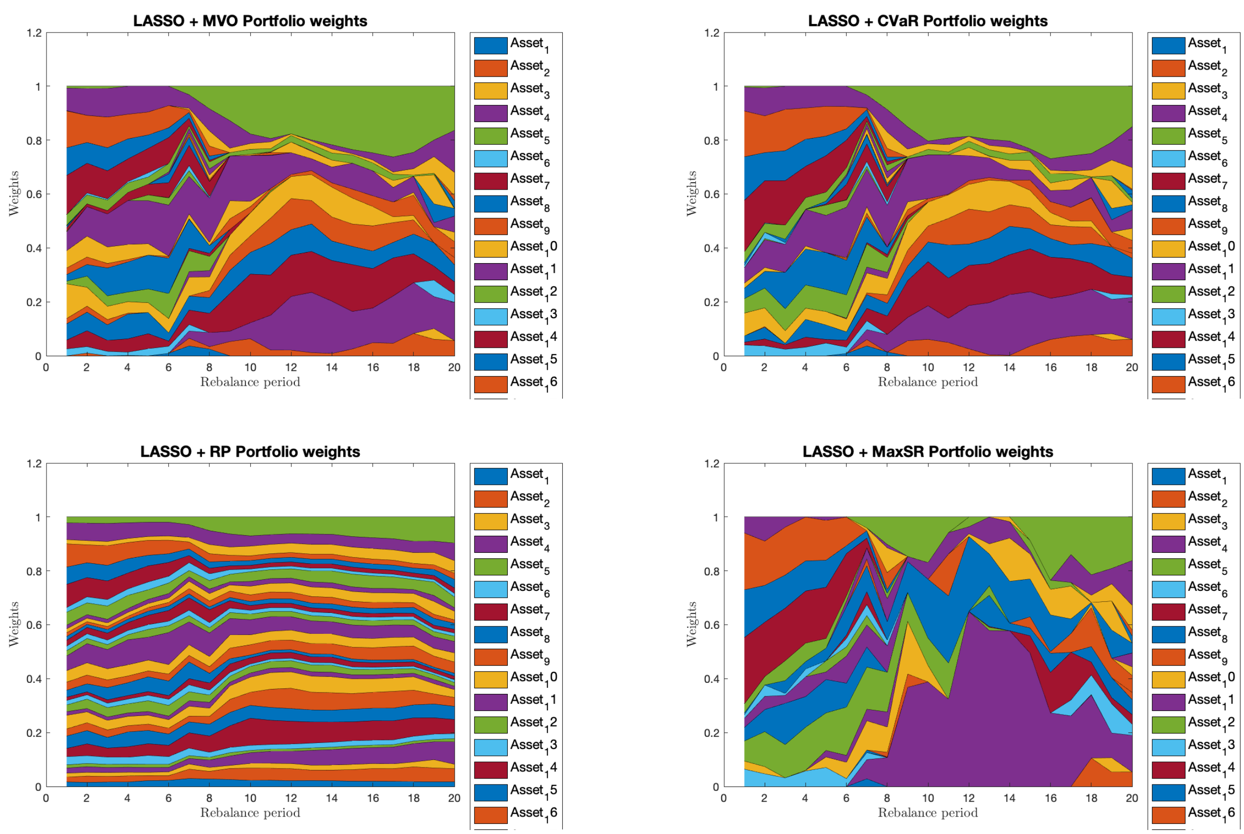


Figure 23: *Portfolio Composition using LASSO*

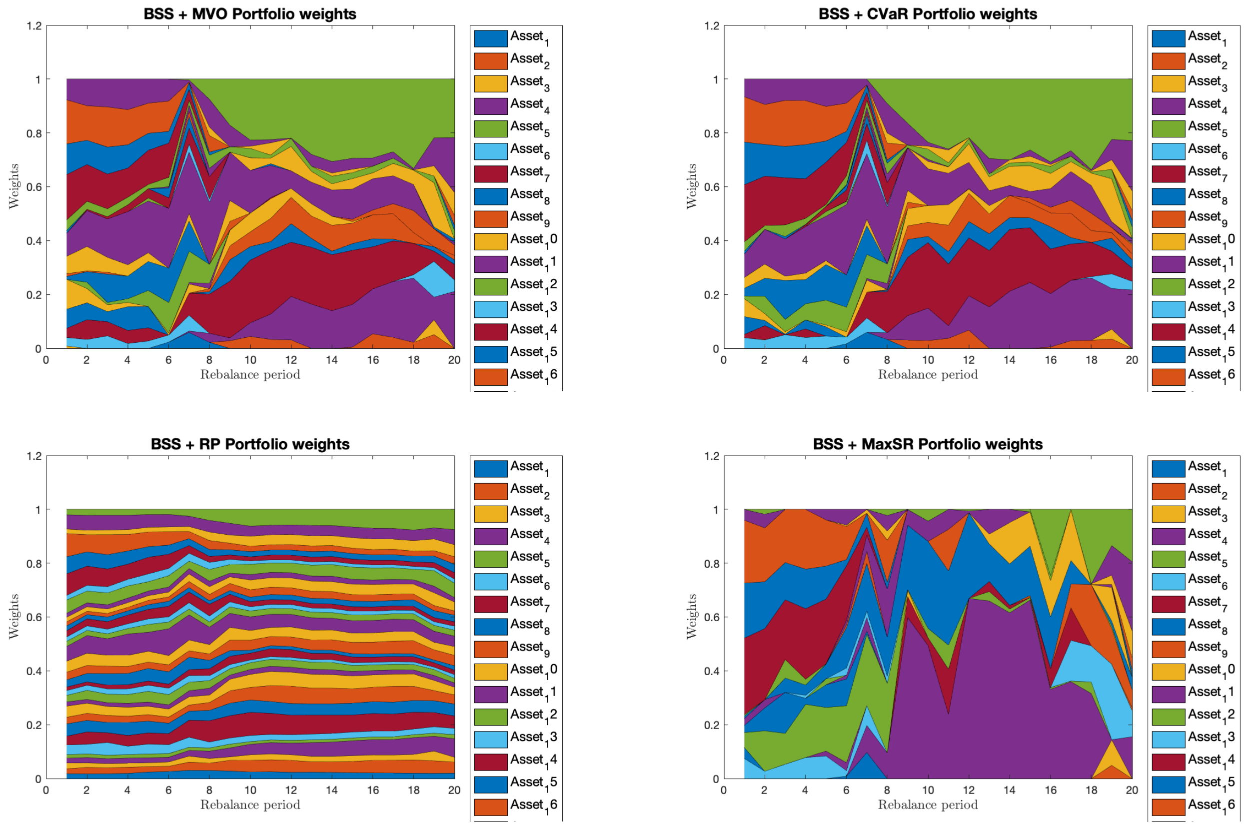


Figure 24: *Portfolio Composition using BSS*

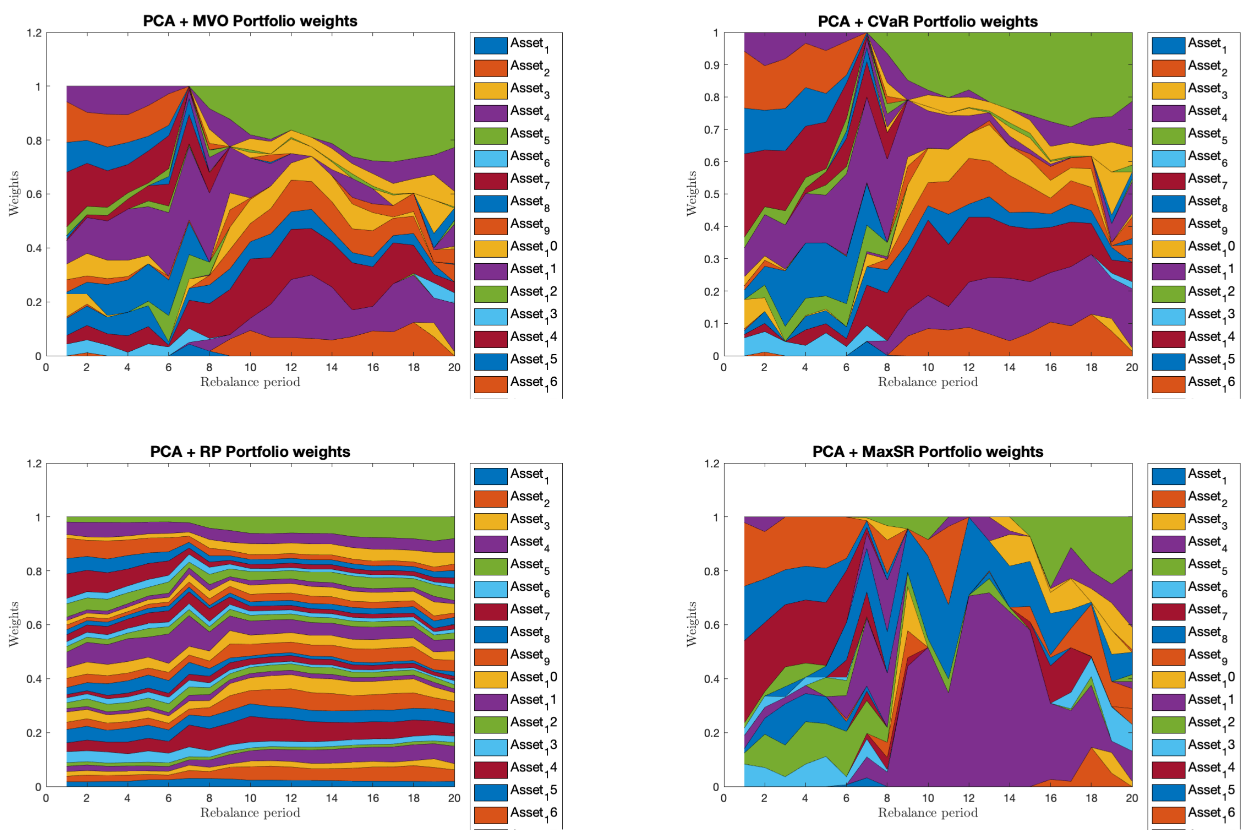


Figure 25: *Portfolio Composition using PCA*

7 Test Performance

Note: *Dataset 2* has been used for measuring the test performance.

After undergoing hyperparameter tuning from *Dataset 1* and model selection process from *Dataset 3*, we finally use *Dataset 2* to measure the final performance as it is the only previously unseen data. We compare the final framework with the baseline model and three other portfolios based on four metrics: average return μ, monthly standard deviation σ, Sharpe ratio SR, and average turnover TO_{avg}.

To begin with, it is promising to see that the final selected framework outperforms the baseline in Sharpe ratio and average turnover by a large margin as these two are the grading criteria of the competition. Likewise, a randomly selected model pair (LASSO and Sharpe ratio optimization model) does not surpass our final model in the two metrics. It is interesting but also anticipated that the final result would fall in between the stable regime portfolio and the volatile one as it is a mixture of the two. The turnover is greater than both of them as one would be required to switch between the portfolios more frequently, yielding a higher turnover.

Figure 26 indicates a strong portfolio performance for two reasons:

- Not only is the wealth evolution plot rising upward to the right, it demonstrates a relatively small drawdown during the dot-com bubble recession between 2002 and 2003. This signifies robustness of the portfolio in the face of market volatility and economic downturns. The minimal drawdown during such a significant financial crisis indicates the portfolio's ability to withstand adverse market conditions and preserve its value.
- Throughout the investment horizon, the portfolio weights in the composition plot show minimal fluctuations, resulting in a low turnover. This stability in portfolio weights reflects the disciplined approach to investment management and indicates a long-term investment strategy focused on maintaining consistent asset allocation.

Table 13: Performance metrics of various portfolio frameworks

	Baseline	Random	Stable	Volatile	Final
Metric	OLS + MVO	LASSO + MaxSR	PCA + RP	OLS + RP	Stable + Volatile
μ	0.0064	0.0098	0.0086	0.0087	0.0087
σ	0.0353	0.0508	0.0394	0.0396	0.0395
SR	0.080945	0.10937	0.12481	0.12535	0.12507
TO_{avg}	0.39652	0.63413	0.10406	0.10211	0.1042

8 Conclusion

The report investigates modern quantitative asset management techniques from a set of four factor models – OLS, LASSO, BSS, and PCA – and four portfolio construction methods – MVO, CVaR, RP, and MaxSR. The aim of the study is to devise an optimal investment strategy using a combination of the models in a systematic manner.

To optimize the hyperparameters of the LASSO model, among the various parameter values explored, $\lambda = 0.03$ emerged as the optimal choice, providing the highest adjusted R^2 from the set of sparsity-promoting values of λ. Moreover we have also determined that the Sharpe ratio of an MVO portfolio under the BSS model is maximized when $K = 5$, making it our preferred choice for the hyperparameter selection. Its plot corresponding to $K = 5$ exhibits the highest total wealth at the

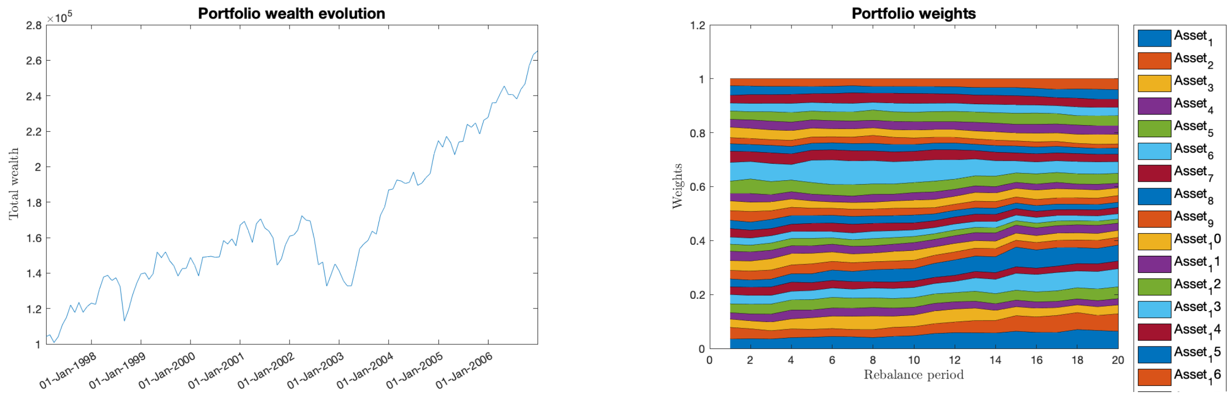


Figure 26: *Portfolio wealth evolution (left) and portfolio composition (right) on the test dataset*

end of the investment horizon, while still maintaining a comparable level of volatility to the other plotted curves. After careful consideration of the number of principal components, we have selected $p = 3$ as our choice. This decision is supported by the scree plots depicted where the elbow method is employed to determine the cutoff point beyond which no further addition of factors is necessary. Upon examining we observe that the elbows occur between factors 2 and 4, and it is evident that the inclusion of factors 4 to 20 results in diminishing returns, as the substantial increase in cumulative η^2 becomes less pronounced.

When it comes to the hyperparameter search for the CVaR model, we first establish a heuristic score funtion $0.8 \cdot SR - 0.2 \cdot TO_{avg}$, then discover that the highest score is achieved when $\alpha = 0.9$ for the OLS and LASSO models and $\alpha = 0.95$ for the BSS and PCA models respectively. This outcome is in line with expectations since extremely low confidence levels like $\alpha = 0.7$ carry a substantial risk of significant expected loss, leading to increased volatility and a lower Sharpe ratio. Conversely, very high confidence levels such as $\alpha = 0.99$ are excessively conservative and may hinder the ability to attain high returns. Likewise, for the optimization of target return R_{target}, we note that the score reaches its peak when R is set to $0.9 \cdot \overline{\mu}$ for the OLS model. In contrast, for the LASSO, BSS, and PCA models, the highest score is achieved when R is equal to $0.7 \cdot \overline{\mu}$.

During periods of rapid market changes, the switch in market regime can be a challenging task. In our analysis, we calculate the average variance of all traded asset based on the data from the six most recent months to identify distinct market regimes. By setting a threshold at 0.01% monthly volatility, we are able to separate our observation periods into the specified market conditions. In stable market scenario, we observe that the combination of the PCA factor model and the risk parity model produces the highest score. This is primarily due to the diversification across all assets as each asset contributes an equal share of risk to the portfolio. In contrast, the OLS model is preferred over other models because the observed asset returns exhibit significant noise and fail to accurately reflect the current market situation. The final investment framework takes the linear combination of the two portfolios based on the market volatility to ensure smooth transition between the regimes and minimize turnover.

8.1 Next Steps

- Additional features which can undergo hyperparameter tuning:
  - Number of years of lookback period for calibration – while it has been fixed at 5 years throughout the report, one could try shorter, longer, or variable length periods.
  - Target return of the MVO model – since we treated the MVO as a baseline model, we

simply fixed the target return as the mean of the asset returns during the lookback period.

  - Number of simulations – while we fixed the number of simulations for the CVaR optimization model as 10,000 iterations, we could try less for faster computation or more for robust scenario generationo.

- Including turnover as part of the optimization problem:
  - None of the current implementation of optimization models incorporates the change in portfolio composition into their objective function or constraints.
  - For example, one can add it to the objective of the MVO problem by formulating it as $\min \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} + \gamma\|\mathbf{x} - \mathbf{x}_0\|_1$, where γ is a turnover penalty coefficient.

- Higher moments of distribution to generate Monte Carlo simulations:
  - Skewness is the 3^{rd} order standardized moment of a probability distribution. It may be thought of as how symmetrical a distribution is about its mean, where negative skewness represents a mean below the mode, and vice-versa for positive skewness. When skewness is equal to zero, the mean, median and mode are all the same.
  - The adjusted Fisher-Pearson coefficient of skewness (used by MATLAB) is calculated as $S = \frac{\sqrt{N(N-1)}}{N-2} \cdot \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^3}{N\sigma^3}$.
  - Kurtosis is the 4^{th} order standardized moment of a probability distribution. Kurtosis may be thought of as how fat the tails of a distribution are, where large values of kurtosis indicate a greater number of outliers than small values.
  - Kurtosis is defined as $K = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^4}{N\sigma^4}$.
  - For the stochastic process with higher moments, skewness (3^{rd}-order moment) and kurtosis (4^{th}-order moment) will be incorporated into the generation of the factor scenarios, $\boldsymbol{f}_s$, and ϵ_s generated as before. Higher moment distributions can be generated using MATLAB's built-in function *pearsrnd()*.
  - The additional moments can be helpful if the underlying factor subject to simulations does not follow a Gaussian distribution.

References

[1] J. Busa. “Solving Quadratic Programming Problem with Linear Constraints Containing Absolute Values”. In: *Acta Electrotechnica et Informatica* 12.3 (2012), pp. 11–18. DOI: 10.2478/v10198-012-0024-4.

[2] D. Bertsimas, A. King, and R. Mazumder. “Best subset selection via a modern optimization lens”. In: *The annals of statistics* (2016), pp. 813–852.