Operations Research Project 2

Automated Asset Management System

Junhao Wang, Fei Wei, Bona Zhang

May 18, 2021

Contents

1 Introduction

The aim of the project is to design an automated asset management system, i.e. an algorithmic trading system, with the concepts, methods and optimization models we have learned throughout the course. To narrow down the scope, we have set up the following investment competition rules:

- We are provided with three different training and competition datasets which consist of equities (both stocks and ETFs) and equity-based factors. The factor data consists of eight factors as shown in Table 1, with an additional column corresponding to the risk-free rate. The observation data is recorded monthly, and the unseen datasets may contain 15 to 40 assets each.

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

Table 1: List of Factors

- For all datasets, we reserved the first five years for calibration and estimation, and was not used for the out-of-sample analysis.

- Portfolios were re-balanced every six months.

2 Methodology

In the project, we firstly applied four factor models on the dataset to estimate μ and Q, which are the inputs for portfolio optimization models. After selecting the best-performance factor model based on some statistical measures, we then used five portfolio optimization strategies to build the best portfolio and trading algorithm.

2.1 Factor Models

Factor models are financial models that use *factors* — that can be technical, fundamental, macroeconomic or alternate to define a security's risk and returns. These variables are referred to as "factors" because of their financial relevance in explaining systematic return and risk. The factor models are linear regression models, as they define the securities returns to be a linear combination of factor returns weighted by the securities factor exposures. They have the general form of:

$$R_{security} = \beta_{factor\ 1}^{security} \times R_{factor\ 1} + \beta_{factor\ 2}^{security} \times R_{factor\ 2} + \cdots$$

Figure 1: General Form of Factor Model

The factor models that we used in this project include:

(1) Ordinary Least Squares (OLS) regression on all eight factors **(OLS model)**
(2) Fama-French three-factor model **(FF model)**
(3) Least absoluate shrinkage and selection operator model **(LASSO model)**
(4) Best subset selection model **(BSS model)**

2.1.1 OLS Model

The Ordinary Least Squares (OLS) regression is a statistical method of analysis estimating the relationship between the observed dependent variable and several explanatory variables. In our project, we used excess return as the dependent variable, and all the factors in Table 1 as the explanatory variables. The factor model for asset i can be expressed as:

$$r_i - r_f = \alpha_i + \sum_{k=1}^{8} \beta_{ik} f_k + \epsilon_i,\ i = 1, ..., 20$$

where r_i is the return of asset i, r_f is the risk-free rate, α_i is the intercept from regression, f_k is the return of factor k, β_{ik} is its corresponding factor loading, and ϵ_i is the stochastic error term of the asset.

For OLS model, we wish to minimize the sum of squared residuals, which is an unconstrained minimization problem:

$$\min_{\mathbf{B}_i} \ ||\mathbf{r}_i - \mathbf{X}\mathbf{B}_i||_2^2$$

where

- $\mathbf{r}_i$ is the observed returns of asset i over risk-free rate through time $t = 1, ..., T$, $r_i \in \mathbb{R}^T$
- $\mathbf{f} = [\mathbf{f}_1\ \mathbf{f}_2\ ...\ \mathbf{f}_p] \in \mathbb{R}^{T \times p}$ is the observed factor returns for p factors over time $t = 1, ..., T$
- $\mathbf{X} = [\mathbf{1}\ \mathbf{f}] \in \mathbb{R}^{T \times (p+1)}$ is the data matrix
- $\mathbf{V}_i = \begin{bmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{ip} \end{bmatrix} \in \mathbb{R}^p$ represents our factor loadings (or 'beta's)

- $\mathbf{B}_i = \begin{bmatrix} \alpha_i \\ \mathbf{V}_i \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{ip} \end{bmatrix} \in \mathbb{R}^{p+1}$ is the vector of regression coefficients for asset i

- $||\mathbf{r}_i - \mathbf{X}\mathbf{B}_i||_2^2 = (\mathbf{r}_i - \mathbf{X}\mathbf{B}_i)^T(\mathbf{r}_i - \mathbf{X}\mathbf{B}_i)$, which is the l_2 norm of $(\mathbf{r}_i - \mathbf{X}\mathbf{B}_i)$

To get the closed-form optimal solution for $\mathbf{B}_i$,

$$\begin{aligned} f(\mathbf{B}_i) &= ||\mathbf{r}_i - \mathbf{X}\mathbf{B}_i||_2^2 = (\mathbf{r}_i - \mathbf{X}\mathbf{B}_i)^T(\mathbf{r}_i - \mathbf{X}\mathbf{B}_i) \\ &= (\mathbf{r}_i)^T\mathbf{r}_i - 2(\mathbf{r}_i)^T\mathbf{X}\mathbf{B}_i + \mathbf{B}_i^T\mathbf{X}^T\mathbf{X}\mathbf{B}_i \\ \nabla f(\mathbf{B}_i) &= -2\mathbf{X}^T\mathbf{r}_i + 2\mathbf{X}^T\mathbf{X}\mathbf{B}_i \overset{set}{=} 0 \\ &\Longrightarrow \mathbf{X}^T\mathbf{X}\mathbf{B}_i = \mathbf{X}^T\mathbf{r}_i \\ &\Longrightarrow \mathbf{B}_i^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r}_i \end{aligned}$$

2.1.2 FF Model

The Fama-French three-factor model is a subset of our 'OLS model', where we only consider the three factors of 'Market', 'Size', and 'Value' from Table 1. The FF model for asset i can be expressed as:

$$r_i - r_f = \alpha_i + \beta_{im}(f_m - r_f) + \beta_{is}f_s + \beta_{iv}f_v + \epsilon_i,\ i = 1, ..., 20$$

where r_i is the return of asset i, r_f is the risk-free rate, α_i is the intercept from regresion, $(f_m - r_f)$ is the excess market return factor with β_{im} being its corresponding loading, f_s is the size factor with β_{is} being its corresponding factor loading, f_v is the value factor with β_{iv} being its corresponding factor loading, and ϵ_i is the stochastic error term of the asset. The only difference between FF model and OLS model is that under FF model, we only include the three factors ($p = 3$) as the explanatory variables, and use them for the derivation of $\mathbf{B}_i^*$. The closed form optimal solution for $\mathbf{B}_i$ is the same as the OLS model, which is:

$$\mathbf{B}_i^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r}_i$$

2.1.3 LASSO Model

The Least Absoluate Shrinkage and Selection Operator (LASSO) regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The LASSO procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of muticollinearity or when we want to automate certain parts of model selection, such as variable selection or parameter elimination. Lasso regression performs l_1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients. This type of regularization can result in sparse models with few coefficients, where some coefficients can become zero and eliminated from the model.

In this project, we used the penalized form of LASSO on all eight factors, where the model is the following:

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

where $||\mathbf{B}_i||_1 = |\alpha_i| + \sum_{k=1}^{p} |\beta_{ik}|$ is the l_1 norm of $\mathbf{B}_i$. The L1-norm can be written as $-\mathbf{z} \leq \mathbf{B}_i \leq \mathbf{z}$ where $\mathbf{z} \in \mathbb{R}^{(p+1)\times 1}$. To solve this quadratic problem, we can apply function `quadprog` in Matlab. The default model is:

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}x^T H x + f^T x \\ \text{s.t.} \quad & Ax \leq b \\ & Aeq \cdot x = beq \\ & lb \leq x \leq ub \end{aligned}$$

The objective function can be transformed into:

$$||\mathbf{r}_i - \mathbf{X}\mathbf{B}_i||_2^2 + \lambda||\mathbf{B}_i||_1 = \mathbf{B}_i^T\mathbf{X}^T\mathbf{X}\mathbf{B}_i - 2\mathbf{X}^T\mathbf{r}_i\mathbf{B}_i + \mathbf{r}_i^T\mathbf{r}_i + \lambda\mathbf{1}^T\mathbf{z}$$

Since the $\mathbf{r}_i^T\mathbf{r}_i$ is a constant term, it does not impact the optimal solution. We write the matrix of our model as:

- $x = \begin{bmatrix} \mathbf{B}_i \\ \mathbf{z}_i \end{bmatrix} \in \mathbb{R}^{2(p+1)}$ is the objective variable
- $\mathbf{H} = 2\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{2(p+1)\times 2(p+1)}$ is the quadratic matrix in objective function
- $\mathbf{f} = \begin{bmatrix} -2\mathbf{X}^T\mathbf{r}_i \\ \lambda\mathbf{1}_{p+1,1} \end{bmatrix} \in \mathbb{R}^{2(p+1)}$ is the linear term in the objective function
- $\mathbf{A} = \begin{bmatrix} \mathbf{1}_{p+1} & -\mathbf{1}_{p+1} \\ -\mathbf{1}_{p+1} & -\mathbf{1}_{p+1} \end{bmatrix} \in \mathbb{R}^{2(p+1)\times 2(p+1)}$ is the inequality constraint matrix
- $\mathbf{b} = \mathbf{0} \in \mathbb{R}^{2(p+1)}$ is the inequality constraint vector
- $\text{lb} = -\infty \in \mathbb{R}^{2(p+1)}$ is the variables' lower bound
- $\text{ub} = \infty \in \mathbb{R}^{2(p+1)}$ is the variables' upper bound

As for the choice of λ, we implemented similar strategies in Project 1 to find the proper value, which should ideally result in a sparse factor model when there are only two to five factors are non-zero.

2.1.4 BSS Model

The Best Subset Selection (BSS) regression is a model selection approach that consists of testing all possible combination of the predictor variables, and then selecting the best model according to some statistical criteria. In the project, we used the constrained form of BSS model with all eight factors as inputs. The model is set up as the following:

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

where $||\mathbf{B}_i||_0 = \mathbb{1}_{\alpha_i \neq 0} + \sum_{k=1}^{p} \mathbb{1}_{\beta_{ik} \neq 0}$ is the l_0 norm of $\mathbf{B}_i$.

Since BSS involves l_0-norm constraint, we can present the problem into:

$$\begin{aligned}\min_{\widetilde{\mathbf{B}}_i} \quad & ||\widetilde{\mathbf{r}}_i - \widetilde{\mathbf{X}}\widetilde{\mathbf{B}}_i||_2^2 = \widetilde{\mathbf{B}}_i^T (\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})\widetilde{\mathbf{B}}_i - 2\widetilde{\mathbf{r}}_i^T \widetilde{\mathbf{B}}_i + \widetilde{\mathbf{r}}_i^T \widetilde{\mathbf{r}}_i \\ s.t. \quad & -\mathbf{B}_i - \mu \mathbf{z}_i \leq 0 \\ & \mathbf{B}_i - \mu \mathbf{z}_i \leq 0 \\ & \sum_{j=1}^{p+1} z_{ij} \leq K \\ & z_{ij} \in \{0, 1\}\end{aligned}$$

where

- $\widetilde{\mathbf{r}}_i = \begin{bmatrix} \mathbf{r}_i \\ 0 \end{bmatrix} \in \mathbb{R}^{2T}$ where $\mathbf{r}_i$ is the observed returns of asset i over risk-free rate through time $t = 1, \ldots, T, r_i \in \mathbb{R}^T$
- $\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2T \times 2(p+1)}$ is the data matrix
- $\widetilde{\mathbf{B}}_i = \begin{bmatrix} \mathbf{B}_i \\ \mathbf{z}_i \end{bmatrix}^T = \begin{bmatrix} \alpha_i & \beta_{i1} & \beta_{i2} & \cdots & \beta_{ip} & z_{i1} & \cdots & z_{i,p+1} \end{bmatrix}^T \in \mathbb{R}^{2(p+1)}$ is the objective variable

Since the $\widetilde{\mathbf{r}}_i^T\widetilde{\mathbf{r}}_i$ is a constant term, it does not impact the optimization solution. We use the function `gurobi` to solve the problem. In programming, we write the matrix of our model as:

$$\begin{aligned} \min_{x} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned}$$

where

- $\mathbf{Q} = \widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2(p+1)\times 2(p+1)}$ is the quadratic term in objective function
- $\mathbf{c} = -2\widetilde{\mathbf{X}}^T\widetilde{\mathbf{r}}_i = -2\mathbf{X}^T\mathbf{r}_i \in \mathbb{R}^{p+1}$ is the linear term in objective function
- $A = \begin{bmatrix} -I_{p+1} & -\mu I_{p+1} \\ -I_{p+1} & -\mu I_{p+1} \\ \mathbf{0}_{1\times(p+1)} & \mathbf{1}_{1\times(p+1)} \end{bmatrix} \in \mathbb{R}^{(2(p+1)+1)\times 2(p+1)}$ is the inequality constraint matrix
- $b = \begin{bmatrix} \mathbf{0}_{2p+1} \\ K \end{bmatrix} \in \mathbb{R}^{2(p+1)+1}$ is the inequality constraint vector

According to the conclusion from Project 1, larger values of K tend to return better results. Hence, we chose $K = 6$ here.

2.2 Estimation of Portfolio Optimization Strategies Input Parameters

After obtaining the optimal solution for $\mathbf{B}_i$, we can then estimate the input parameters for the portfolio optimization models, which are the expected asset return $\boldsymbol{\mu}$ and asset covariance matrix $\boldsymbol{Q}$. They have the vector notation in the form of:

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\alpha} + \boldsymbol{V}^T\bar{\boldsymbol{f}} \\ \boldsymbol{Q} &= \boldsymbol{V}^T\boldsymbol{F}\boldsymbol{V} + \boldsymbol{D} \end{aligned}$$

where

- $\bar{\boldsymbol{f}} \in \mathbb{R}^p$ is the vector of expected factor returns
- $\boldsymbol{F} \in \mathbb{R}^{p\times p}$ is the factor covariance matrix
- $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is the diagonal matrix of residual variances, i.e. the diagonal elements of $\boldsymbol{D}$ are the residual variances, $\sigma_{\epsilon_i}^2$ and all off-diagonal elements are zero. In particular, the residual variances are calculated by $\sigma_\epsilon^2 = \frac{1}{T-p-1}||\boldsymbol{\epsilon}||_2^2$, where we divide by the appropriate number of degree of freedom (DOF), $T-p-1$; and $\boldsymbol{\epsilon} = \boldsymbol{r} - \boldsymbol{X}\boldsymbol{B}$.

2.3 Portfolio Optimization Strategies

After estimating μ and Q from Section 2.2, we can use them as the inputs for the investment strategies to optimize our portfolio. Portfolio optimization consists of determining a set of assets, and their respective portfolio participation weights, which satisfy the investors' objective. The objective typically maximizes factors such as expected return, or minimizes costs like financial risk. We have implemented the following five different portfolio optimization strategies to help us find the best portfolio thus trading algorithm:

(1) Mean-Variance Optimization **(MVO)**
(2) **Robust** Mean-Variance Optimization **(MVO)**
(3) Conditional Value-at-Risk **(CVaR)** Optimization
(4) Risk-Parity **(RP)** Optimization
(5) **Sharpe Ratio Maximize** Optimization

The details of the strategies as well as the selection of their corresponding parameters are illustrated as follows.

2.3.1 MVO

The first strategy we used is MVO, where we focused on two population characteristics - the mean and the covariance matrix of asset excess returns to optimize our portfolio. We used the version of MVO that seeks to minimize variance subject to a target expected return, with the setup as follows:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{\mu}^T \boldsymbol{x} \geq R_{\text{target}} \\ & \mathbf{1}^T \boldsymbol{x} = 1 \\ & \boldsymbol{x} \geq 0 \end{aligned}$$

where R_{target} is our target return, which is the average expected excess return of the market for the corresponding calibration period (i.e. we take the geometric mean of the market factor using the historical data pertinent to the calibration period). The target return will change every time we rebalance our portfolios. Besides, we set $\boldsymbol{x} \geq 0$ because short sales are not allowed. The Matlab function `quadprog` is used to solve for the optimal asset allocation.

2.3.2 Robust MVO

Under the nominal MVO assumptions, we make our investment decisions based solely on the estimated parameters, ignoring estimation error as well as their impact during optimization. However, the estimation errors may lead us to construct portfolios that do not align well with future realizations. In order to incorporate the uncertainty into the optimization model due to noisy expected returns, we introduced the robust MVO strategy. It constructed an ellipsoidal uncertainty set around the expected returns as the penalty on our targets returns with the value of $\epsilon_2||\boldsymbol{\Theta^{1/2}x}||_2$, regarding an appropriate confidence level. In this project, we used the alternative robust MVO to get an optimal trade-off between the risk and penalized expected returns. The optimization have the following formulation:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \lambda \boldsymbol{x^T Q x} - \boldsymbol{\mu}^T \boldsymbol{x} + \epsilon_2||\boldsymbol{\Theta^{1/2}x}||_2 \\ \text{s.t.} \quad & \boldsymbol{1x^T} = 1 \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

where

- ϵ_2 is the radius that bounds the standardized distance between μ and μ^{true}
- $\Theta = \frac{1}{T} diag(diag(Q))$, which is the measure of the uncertainty to standardize the estimated expected excess returns

2.3.3 Selection of λ for Robust MVO

To select the best value for λ (taking dataset 1 as an example), we firstly plotted the portfolio values for λ = 0.01, 0.02, 0.2, 1, and 2 as out-of-sample analysis, as shown in Figure 2. According to the graph, the portfolio values for different λ nearly coincide with each other for small values of λ. Larger values of λ tend to return lower portfolio values in the end. We may not be able to distinguish the differences when $\lambda = 0.01, 0.02$ and 0.2 in terms of the portfolio values.

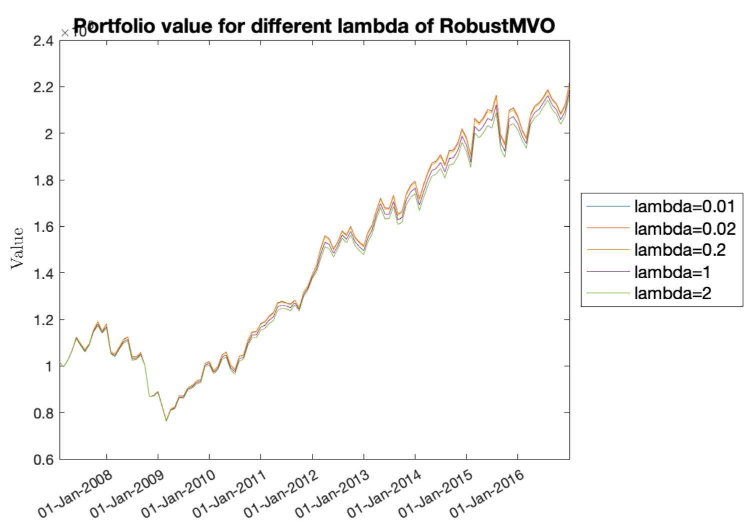


Figure 2: Portfolio Value for Different λ

Apart from the portfolio values, we also used several measurements to evaluate the portfolio performances, including:

- Geometric mean of excess returns
- Standard deviation (std) of excess returns, which measures the volatility of excess returns
- Sharpe ratio (more in Section 2.3.7), is the average return earned in excess of the risk-free rate per unit of volatility or total risk. A high Sharpe ratio is good when compared to similar portfolios or funds with lower returns.
- Average turnover rates, where turnover rate is the percentage of a mutual fund or other portfolio's holdings that have been replaced in a given year. In general, a low turnover ratio is desirable over a high turnover ratio. The rationale is that there are transaction costs involved with making trades (buying and selling securities).

The summary of portfolio performance measurements for various values of λ are displayed in Figure 3, where we see that when $\lambda = 0.02$, the portfolio has the highest geometric mean of excess returns as well as the second highest Sharpe ratio. Therefore, we chose $\lambda = 0.02$ as the optimal value.

Lambda	Geo Mean of Excess Returns	Std of Excess Returns	Sharpe Ratio	Avg Turnovers
0.01	0.0060	0.0360	0.1673	0.3822
0.02	0.0060	0.0360	0.1673	0.3820
0.2	0.0060	0.0359	0.1672	0.3788
1	0.0059	0.0350	0.1669	0.3678
2	0.0058	0.0350	0.1665	0.3592

Figure 3: Summary of Portfolio Performance Measurements for Different λ

2.3.4 RP Optimization

Risk parity, also known as 'Equal Risk Contribution', is a portfolio allocation strategy that uses risk to determine allocations across various components of an investment portfolio. The risk parity approach to portfolio construction seeks to allocate investment capital on a risk-weighted basis to optimally diversify investments, viewing the risk and return of the entire portfolio as one. By applying the strategy, we aim to equalize the risk distribution for each asset by risk diversification while also takes the risk covariances into consideration. In this project, we considered the convex reformulation of the original optimization, which has the formulation as follows for some variable $\boldsymbol{y} \in \mathbb{R}^n$ and $K > 0$:

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & \frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{y} - K\sum_{i=1}^{n} \ln(y_i) \\ \text{s.t.} \quad & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

Let $f(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{y} - K\sum_{i=1}^{n}\ln(y_i)$. It is a strictly convex function since the covariance matrix $\boldsymbol{Q}$ is PSD, and $K\sum_{i=1}^{n}\ln(y_i)$ is a strictly concave function for some $K > 0$. Hence, there exists a unique global minimum to the problem. To find the optimal solution, we derived its gradient and set it to zero:

$$\nabla f(\boldsymbol{y}) = \boldsymbol{Q}\boldsymbol{y} - K\boldsymbol{y}^{-1} \overset{set}{=} 0,$$
$$\text{where } \boldsymbol{y}^{-1} = [1/y_1, 1/y_2, ..., 1/y_n]^T$$

After finding the optimal solution of $\boldsymbol{y}^*$, we can get the optimal asset weights by:

$$x_i^* = \frac{y_i^*}{\sum_{i=1}^{n} y_i^*}$$

where $\boldsymbol{x}^*$ is unique.

2.3.5 Selection of K for RP Optimization

To find the most optimal value of K, we also plotted the portfolio values and calculated portfolio performance measures for different values of K, as shown in Figure 4 and 5. Since the portfolio values as well as the measurements coincide with each other, we found that the portfolio values are independent of our initial choice of K. Here, we chose $K = 5$.

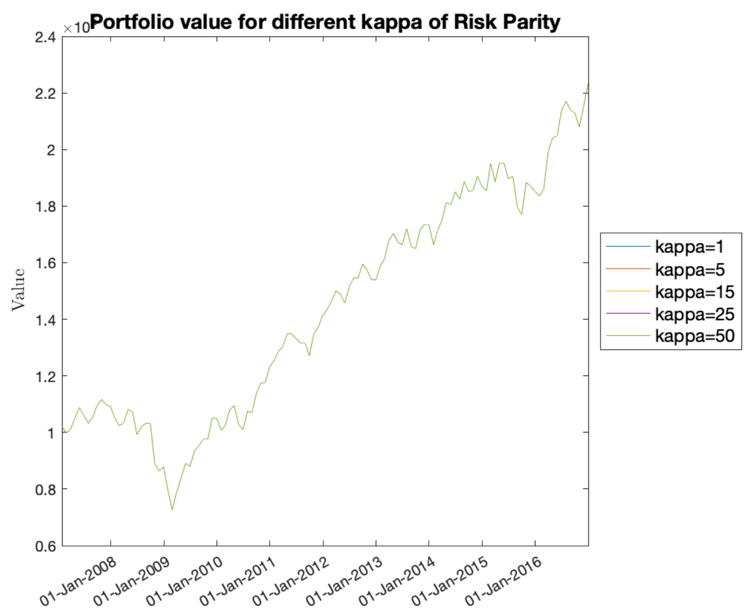


Figure 4: Portfolio Value for Different K

Kappa	Geo Mean of Excess Returns	Std of Excess Returns	Sharpe Ratio	Avg Turnovers
1	0.0061	0.0357	0.1699	0.1567
5	0.0061	0.0357	0.1699	0.1567
15	0.0061	0.0357	0.1699	0.1567
25	0.0061	0.0357	0.1699	0.1567
50	0.0061	0.0357	0.1699	0.1567

Figure 5: Summary of Portfolio Performance Measurements for Different K

2.3.6 CVaR Optimization

Another approach that we used is the CVaR optimization strategy. For continuous distributions, CVaR is defined as the conditional expected loss under the condition that it exceeds VaR. This risk measure is also known as Mean Excess Loss, Mean Shortfall, or Tail Value-at-Risk. Compared to VaR method, CVaR not only considered the magnititude of losses exceeding VaR, but it is also sub-additive and convex. $CVaR_\alpha$ has the mathematical form of:

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

Recall the definition of VaR_α,

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

VaR_α can be regarded as the minimum γ that satisfies the constraint $\Psi(\boldsymbol{x},\gamma) \geq \alpha$. Hence, let $\gamma \in \mathbb{R}$ serve as a placeholder for VaR_α during optimization, we have the definition of $CVaR_\alpha$ with respect to γ as:

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

where $a^+ = \max\{a, 0\}$.

Since $p(\boldsymbol{r})$ is sometimes difficult to solve analytically, we used a scenario-based representation, $\hat{\boldsymbol{r_s}}$ for $x = 1, ..., S$, where $\hat{\boldsymbol{r_s}} \in \mathbb{R}^n$ is the realization of scenario s. Assume all scenarios are equally likely, then the approximation of $F_\alpha(\boldsymbol{x},\gamma)$ can be written as:

$$\tilde{F}_\alpha(\boldsymbol{x},\gamma) = \gamma + \frac{1}{(1-\alpha)S}\sum_{s=1}^{S}\left(f(\boldsymbol{x},\hat{\boldsymbol{r_s}}) - \gamma\right)^+$$

where we can deal with the non-smoothness of $\tilde{F}_\alpha(\boldsymbol{x},\gamma)$ by introducing an auxiliary variable for each $s = 1, ..., S$,

$$\left(f(\boldsymbol{x},\hat{\boldsymbol{r_s}}) - \gamma\right)^+ = \begin{cases} Z_s \geq 0, \\ Z_s \geq f(\boldsymbol{x},\hat{\boldsymbol{r_s}}) - \gamma \end{cases}$$

Hence, the $CVaR_\alpha$ optimization problem can be formulated as the following:

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

where

- The linear loss function $\boldsymbol{f(x, \hat{r_s})} = -\boldsymbol{\hat{r_s}^T x}$
- $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{1^T x} = 1; ...; \boldsymbol{\mu^T x} \geq R\}$
- α is the confidence interval of our portfolio $\boldsymbol{x} \in \mathbb{R}^n$

2.3.7 Sharpe Ratio Maximize Optimization

Since Sharpe ratio is an important measure to the performance of our optimized portfolio, we introduced the Sharpe Ratio strategy to maximize this measure. Sharpe ratio is a measure of excess return received for the extra volatility endured to hold a riskier asset. The higher the Sharpe ratio, the better the portfolio. Mathematically, the Sharpe ratio has the form:

$$Sharpe\ Ratio = \frac{\bar{r} - r_f}{\sigma}, \text{ where } \sigma \text{ is the standard deviation of the portfolio.}$$

As an alternative to the standard MVO as mentioned, we consider a different portfolio optimization task. We denote risk-free interest rate as r_f and formulated the problem as following:

$$\max_{\boldsymbol{x}} \frac{\boldsymbol{\mu^T x} - r_f}{\sqrt{\boldsymbol{x^T Q x}}} \tag{1}$$

$$\text{s.t.} \sum_j x_j = 1 \tag{2}$$

$$\boldsymbol{x} \geq \boldsymbol{0} \tag{3}$$

The objective function is naturally difficult to solve. However, it is feasible to reduce it to a standard convex quadratic program under a reasonable assumption - there exists a vector x which satisfies the following inequality:

$$\mu^T - r_f \geq 0$$

Given an asset vector we define:

$$f(x) = \frac{\mu^T x - r_f}{\sqrt{x^T Q x}}$$

Since we know $\sum_j x_j = 1$,

$$f(x) = \frac{\mu^T x - r_f}{\sqrt{x^T Q x}} = \frac{\mu^T x - r_f \sum_j x_j}{\sqrt{x^T Q x}} = \frac{\hat{\mu}^T x}{\sqrt{x^T Q x}}$$

we define $\hat{\mu_j} = \mu_j - r_f$ for each j. Based on this fact, we have a observation: For any vector $\boldsymbol{x}$ satisfies $\sum_j x_j = 1$ and any scalar $\lambda \geq 0, \quad f(\lambda x) = f(x)$ We can verify the observation by letting $y = \lambda x$, then $\sqrt{y^T Q y} = \lambda \sqrt{x^T Q x}$ and similarly we can check $\hat{\mu}^T y = \lambda \hat{\mu}^T x$.

Now we can rewrite our optimization problem as the following:

$$\max_{\boldsymbol{x}} \frac{1}{\sqrt{\boldsymbol{y^T Q y}}} \tag{4}$$

$$\text{s.t. } \boldsymbol{\hat{\mu}^T y} = 1 \tag{5}$$

$$\boldsymbol{y} \geq \boldsymbol{0} \tag{6}$$

To verify the problem (1) and (4) are equivalent, we suppose that $\bar{y}$ is an optimal solution to (4). To be noticed that because of (5), $\bar{y}$ cannot be the zero vector. In addition by (6), we can get $\sum_j \bar{y}_i \geq 0$. We define a new vector $\bar{x}$ which satisfies the following equation:

$$\bar{x} = \frac{\bar{y}}{\sum_j \bar{y}_i}$$

By construction,

$$\sum_j \bar{x_j} = 1$$

Therefore, $\bar{x}$ is feasible for quadratic program (4). Moreover, as observed before, $f(\bar{x}) = f(\bar{y}) = \frac{1}{\sqrt{y^T Q y}}$ because of $\hat{\mu}^T \bar{y} = 1$

In conclusion, the value of problem (1) is at least as large as the value of problem (4). The converse can be proved by the similar way. Therefore, problem (1) and (4) are equivalent.

We rewrite the problem (4) in the following standard quadratic program:

$$\min_{\boldsymbol{x}} \sqrt{\boldsymbol{y^T Q y}}$$

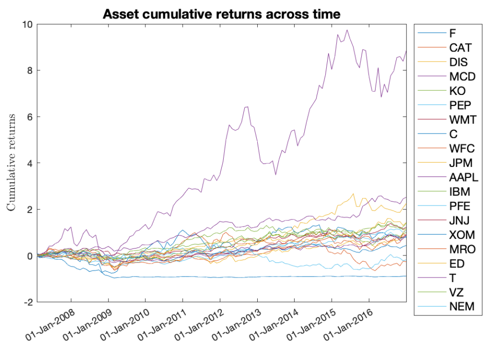
$$\text{s.t. } \boldsymbol{\hat{\mu}^T y = 1}$$

$$\boldsymbol{y} \geq \boldsymbol{0}$$

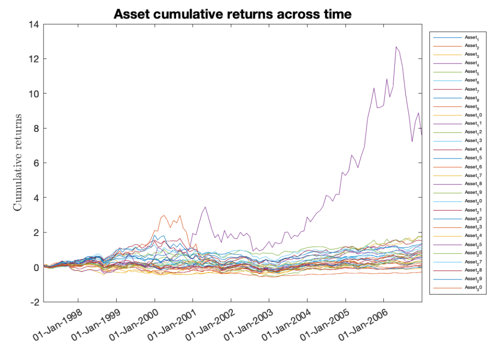
3 Model Selection Process and Testing

3.1 Overview of Datasets

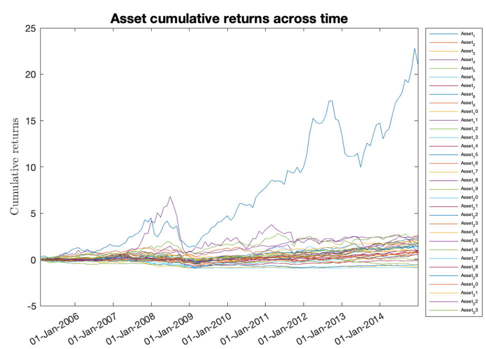
In this project, we are given three datasets, representing different investment periods and market conditions. Figure 6 shows the plots of cumulative returns for the three datasets to provide an overview.



(a) Dataset 1 Asset Cumulative Returns



(b) Dataset 2 Asset Cumulative Returns



(c) Dataset 3 Asset Cumulative Returns

Figure 6: Asset Cumulative Returns

3.2 Selection of Factor Models

To compare the measure of fit between the four factor models, we use the coefficient of determination R^2. In statistics, R^2 represents the proportion of the variance in the dependent variables that is predictable from the independent variables. The closer R^2 is to 1, the better fitness the linear regression model. Mathematically, R^2 is defined as:

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

where SS_{tot} denotes the total sum of squares of the data and SS_{reg} denotes the regression sum of squares. For our project, they are in the following forms for asset i:

$$\begin{aligned} SS_{tot} &= \sum_{t=1}^{T} r_t^{(i)} - \bar{r}^{(i)} \\ SS_{reg} &= \sum_{t=1}^{T} r_t^{(i)} \mathbf{X}\mathbf{B}_i - \bar{r}^{(i)} \end{aligned}$$

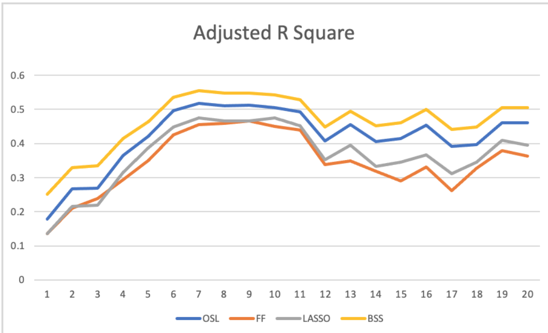
where $\bar{r}^{(i)}$ is the mean return of asset i.

On the other hand, to account for the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model, we calculated adjusted R^2 value ($\bar{R}^2$) instead to compare the measures of fit. $\bar{R}^2$ adjusts the number of explanatory variables excluding constant(p) by the degree of freedom relative to the magnitude of the data (T) and has the form of:

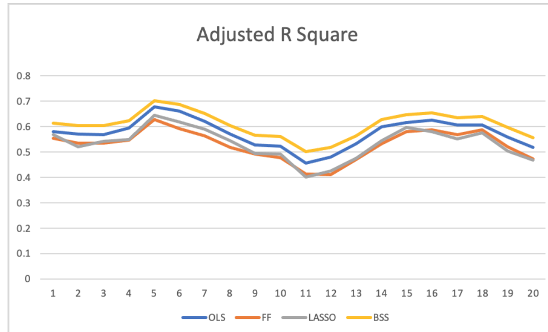
$$\bar{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-p-1}$$

Note that by the setup of the project, the total number of explanatory variables (p) for the 4 factor models are not the same.

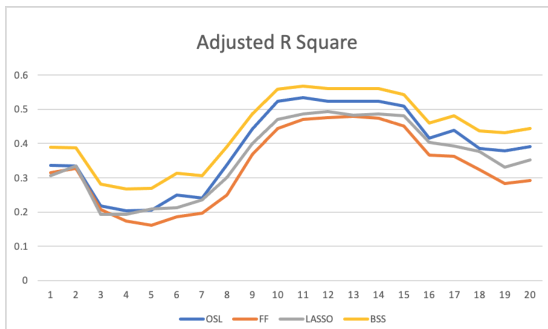
We recorded the R^2 mean values for all three datasets and plotted the line graphs as shown in Figure 7. (The R^2 mean values can be found in tables in Appendix A.) According to the the graph, the BSS model with $K = 4$ has the largest value among the 4 factor models throughout the investment period. Thus, it outperforms the rest of three factor models in explaining the asset returns as well as having highest credibility on estimation of expected returns μ and covariance matrix Q served as optimization models input parameters. Besides, the OSL model has second best performance whose adjusted R^2 is noticeably less than the best BSS model in the charts across the years. Moreover, FF and Lasso models perform the worst in explaining the asset returns. Therefore, we decided to use the BSS model results for further analysis.



(a) Dataset 1 Adjusted R^2



(b) Dataset 2 Adjusted R^2



(c) Dataset 3 Adjusted R^2

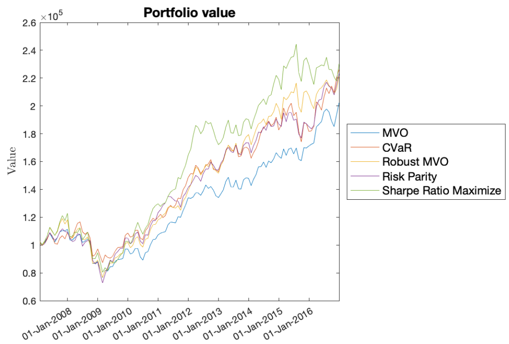
Figure 7: Adjusted R^2

3.3 Selection of Portfolio Optimization Strategies

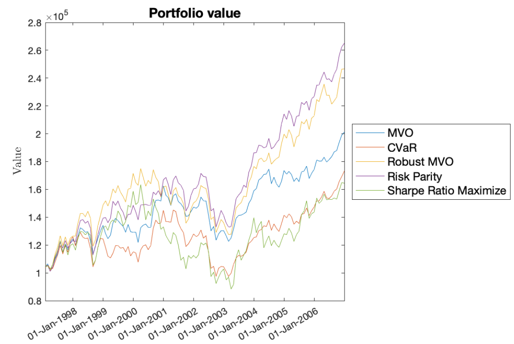
In this section, the out-of-sample analysis will be conducted for the five strategies to evaluate their financial performance of the corresponding portfolios, which will help us to select the best strategy.

3.3.1 Portfolio Values

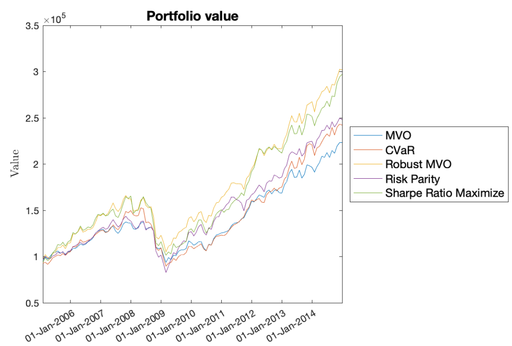
We have plotted the evolution of portfolio values within the investment period for all three datasets, as shown in Figure 8. In general, the differences between each strategy in the first two to three years are small, where the lines almost coincide with each other. While the gaps become increasingly large as time goes by. According to dataset 1 and 3, Sharpe ratio maximize and Robust MVO ($\lambda = 0.02$) optimizations have the best performances in returning the highest portfolio values in the end. While for dataset 2, RP method has the highest portfolio values towards the end, and Robust MVO ($\lambda = 0.02$) portfolio ranks the second. Based on the portfolio value evolution, it is not clear which strategy is the best, so we need to take on more portfolio performance measures.



(a) Dataset 1 Portfolio Values



(b) Dataset 2 Portfolio Values



(c) Dataset 3 Portfolio Values

Figure 8: Portfolio Values

3.3.2 Measurements of Portfolio Performance

The portfolio performance measures for all three datasets are recorded in Figure 9, where the better strategies are colored in darker green. To see the performance of each strategy clearer, we set up a "Score" column, which equals to $0.8 \times Sharpe\ ratio + 0.2 \times (-1) \times Avg\ Turnovers$. Since we aim to maximize Sharpe ratio and minimize the average turnovers, the higher the score, the better the strategy's performance.

For dataset 1, there is no big difference in the geometric means of excess returns. RP strategy tends to return the best result with the highest score since it has the second highest Sharpe ratio, and the lowest average turnovers which is much lower compared to others. This means that the RP portfolio has high excess return received for the extra risk endured with the lowest percentage of assets being replaced in the given period. Its standard deviation of excess returns is also reasonably low, meaning the risk/volatility of the portfolio is low. Robust MVO with $\lambda = 0.02$ has the second best performance with slightly lower Sharpe ratio and mean of excess returns as well as a bit higher risks and turnovers compared to the RP method. Similarly to dataset 1, RP methods also returns the best portfolio in general with the highest score for dataset 2, where it has the highest geometric mean of excess returns, second highest Sharpe ratio and the lowest average turnovers. Robust MVO with $\lambda = 0.02$ also ranks the second in performance here, with a slightly worse performance compared to the RP optimization. On the other hand, for dataset 3, even though RP method has not only low geometric mean of excess returns and Sharpe ratio, but also high volatility, it still returns the lowest average turnovers, which drives its score to the second best. Robust MVO with $\lambda = 0.02$ has the best performance in terms of Sharpe ratio and Average turnovers, it also has the second highest geometric mean of excess returns. Its standard deviation of excess returns as well as the average turnovers are also reasonably low. Even though it ranks the second for dataset 3, the difference between RP and robust MVO is pretty small. To get the best overall portfolio optimization performance in terms of Sharpe ratio and turnovers, we decided to choose the RP optimization strategy as our final method since it has the highest score in general, bsaed on the given datasets. Note that the most optimal method may vary depending on different investment periods as well as bull or bear market conditions.

Strategy	Geo Mean of Excess Returns	Std of Excess Returns	Sharpe Ratio	Avg Turnovers	Score
MVO	0.0052	0.0331	0.1574	0.5181	0.0223
Robust MVO (Lambda = 0.02)	0.0060	0.0360	0.1673	0.3820	0.0575
RP	0.0061	0.0357	0.1699	0.1567	0.1045
CVaR	0.0061	0.0345	0.1777	0.8314	-0.0242
Sharpe Ratio Maximization	0.0063	0.0389	0.1618	0.8024	-0.0311

(a) Dataset 1 Portfolio Performance Measures

Strategy	Geo Mean of Excess Returns	Std of Excess Returns	Sharpe Ratio	Avg Turnovers	Score
MVO	0.0026	0.0346	0.0744	0.6434	-0.0692
Robust MVO (Lambda = 0.02)	0.0042	0.0413	0.1023	0.3209	0.0177
RP	0.0049	0.0391	0.1256	0.1293	0.0746
CVaR	0.0014	0.0398	0.0343	0.8880	-0.1502
Sharpe Ratio Maximization	0.0009	0.0566	0.0157	1.1037	-0.2082

(b) Dataset 2 Portfolio Performance Measures

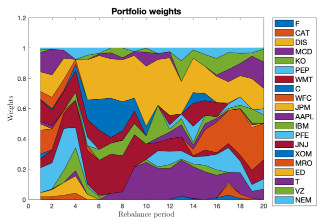
Strategy	Geo Mean of Excess Returns	Std of Excess Returns	Sharpe Ratio	Avg Turnovers	Score
MVO	0.0060	0.0315	0.1904	0.5680	0.0387
Robust MVO (Lambda = 0.02)	0.0085	0.0367	0.2309	0.3670	0.1113
RP	0.0068	0.0381	0.1787	0.1943	0.1041
CVaR	0.0069	0.0345	0.2007	0.9124	-0.0219
Sharpe Ratio Maximization	0.0086	0.0399	0.2157	0.7591	0.0207

(c) Dataset 3 Portfolio Performance Measures

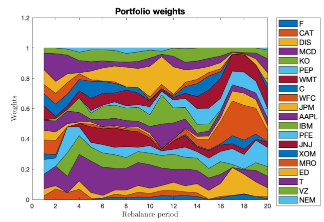
Figure 9: Summary of Portfolio Performance Measures

3.3.3 Portfolio Weights

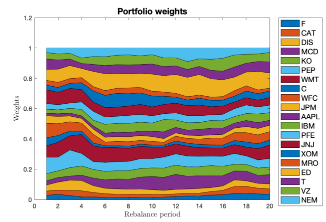
To see how the portfolio composition evolves over the investment period, we plotted the evolution of portfolio weights. Since results are similar for each dataset, here we took dataset 1 as an example, and the plot is shown in Figure 10. In particular, it shows how our optimal weights changes every time we rebalanced a portfolio for the five strategies. (Portfolio weights for other datasets are attached in Appendix B.) As demonstrated in the graphs, Robust MVO and RP portfolios tend to have the most diversified portfolios, and the RP portfolio has the most steady composition of stocks. On the other hand, the portfolios of CVaR and Sharpe Ratio Maximize Optimizations seem to concentrate more on few stocks over time, where we can see there are periods that only 3 or 4 stocks took up the most proportions of the portfolios.



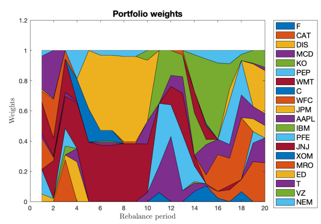
(a) MVO Portfolio Weights



(b) Robust MVO Portfolio Weights



(c) RP Portfolio Weights



(d) CVaR Portfolio Weights

(e) Sharpe Ratio Maximize Portfolio Weights

Figure 10: Dataset 1 Portfolio Weights

3.3.4 Running Time

Apart from the former measures, we also recorded running time for the five strategies, taking dataset 1 as an example, where quicker running time is colored in lighter green in Figure 11. Compared with Figure 9 (a), we found that the better the performance of the strategy, the larger the running time tends to be. Therefore, the running time can be generally regarded as a sacrifice to get a better performance in portfolio optimization. On the other hand, this sacrifice is not large since there are only seconds differences in running time.

Strategy	Running Time (s)
MVO	5.4844
Robust MVO (Lambda = 0.02)	7.1226
RP	8.3699
CVaR	5.6937
Sharpe Ratio Maximization	5.0944

Figure 11: Dataset 1 Running Time

4 Conclusion

In this project, we firstly applied four factor models (OLS model, FF model, LASSO model, and BSS model) to the three datasets to find the best one in predicting the input parameters μ and Q for portfolio optimization strategies. By analyzing the line plots of adjusted R^2 mean values, we see that the BSS model with $K = 6$ has the closest values to 1 among the 4 factor models throughout the years. Hence, it has the best performance in explaining the asset returns as well as the most reliable estimation of expected returns μ and covariance matrix Q for optimization strategy inputs. Therefore, the BSS model with $K = 6$ result is chosen to proceed on further analysis.

We have implemented five different portfolio optimization strategies, including MVO, robust MVO, RP, CVaR and Sharpe ratio maximize. To find the best approach, we firstly plotted the evolution of portfolio values throughout the investment period for the 3 datasets, but found that it is not clear to distinguish between the methods. Hence, we moved on to see a set of portfolio performance measures, such as geometric mean of excess returns, standard deviation of excess returns, Sharpe ratio, and average turnovers. In order to clarify the performances according to our competition rule, we built up the "Score" column to capture their differences in Sharpe ratio and turnovers. By comparing these measures, we found that the RP method tends to return the best results and highest score overall. Besides, its portfolio also has the most diversified and steadiest composition of stocks over time. Hence, by the conclusion under these three datasets, the RP method is selected as our final optimization strategy. On the other hand, the most optimal method may vary depending on different investment periods as well as bull or bear market conditions.

In addition, running times for the five strategies were also recorded, where we found that the better the performance of the strategy, the larger the running time tends to be. Hence, we can generally regard running time as a sacrifice to get a better performance portfolio. While even though the running time differs, the gaps between each method are only in seconds, which are pretty small.

5 Appendix A - R^2 Values

Period	1	2	3	4	5	6	7	8	9	10
OLS	0.179095	0.2670425	0.2688398	0.3644828	0.4209619	0.4957159	0.5167953	0.5098597	0.5120448	0.5054406
FF	0.1357094	0.2103131	0.2379262	0.2940464	0.350997	0.425796	0.4546044	0.4597905	0.4663316	0.449868
LASSO	0.1349295	0.2151	0.218273	0.3159374	0.3874296	0.4480458	0.4741556	0.4656424	0.4659312	0.4746603
BSS	0.2506579	0.3301774	0.3348664	0.4140723	0.4640108	0.5356654	0.5549884	0.547429	0.5486509	0.5427007

Period	11	12	13	14	15	16	17	18	19	20
OLS	0.4931353	0.4080874	0.456139	0.4048887	0.4150311	0.4535266	0.3912543	0.3969341	0.4607654	0.4612096
FF	0.4400536	0.3387579	0.348019	0.3183504	0.2907096	0.3306338	0.2618906	0.3280584	0.3785149	0.362502
LASSO	0.452449	0.3522997	0.3943745	0.3325338	0.3455452	0.3665668	0.3118692	0.3454099	0.4092888	0.3957679
BSS	0.5278552	0.449024	0.4946544	0.4519474	0.4610631	0.5002461	0.4418855	0.4482605	0.5049546	0.5056943

Figure 12: Dataset 1 Adjusted R^2

Period	1	2	3	4	5	6	7	8	9	10
OLS	0.58072354	0.57025841	0.56789396	0.59313659	0.6781903	0.6599445	0.61982427	0.56924399	0.52814364	0.52251754
FF	0.55423218	0.53516321	0.53577503	0.54680651	0.62727485	0.59264033	0.56317169	0.51767343	0.49072193	0.47721318
LASSO	0.56842665	0.52023831	0.54138642	0.54987686	0.64396169	0.61728834	0.59036887	0.54451445	0.49455517	0.49259344
BSS	0.61436166	0.60381405	0.60259281	0.62366039	0.70177938	0.68805324	0.65019077	0.60452255	0.56606283	0.56111043

Period	11	12	13	14	15	16	17	18	19	20
OLS	0.45682504	0.47902935	0.53187383	0.59801246	0.61473688	0.62433117	0.60616285	0.60620363	0.55759574	0.51738577
FF	0.41392291	0.4104281	0.47088591	0.53302347	0.58047475	0.58806489	0.56765559	0.58660901	0.52107747	0.47162885
LASSO	0.40068709	0.42415922	0.47420352	0.54517112	0.59647083	0.57996307	0.55167409	0.57416203	0.50274255	0.46761531
BSS	0.50041351	0.51809182	0.56333702	0.62676492	0.64575052	0.65437351	0.63457346	0.63832042	0.59552486	0.55674766

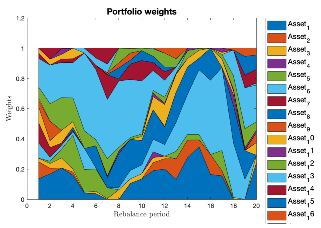
Figure 13: Dataset 2 Adjusted R^2

Period	1	2	3	4	5	6	7	8	9	10
OLS	0.33669918	0.33442082	0.21820099	0.20344506	0.20558434	0.25003589	0.24142978	0.33822514	0.44212644	0.52307749
FF	0.31515703	0.32758351	0.20678267	0.1730911	0.16137565	0.18697687	0.19627607	0.24954475	0.36863046	0.44485634
LASSO	0.30561183	0.33407489	0.19237712	0.1924881	0.20901302	0.21234395	0.23499827	0.30093369	0.40063714	0.47030972
BSS	0.38976033	0.38795326	0.28191921	0.26752622	0.26962216	0.31395405	0.3067801	0.3904148	0.48670925	0.55957285

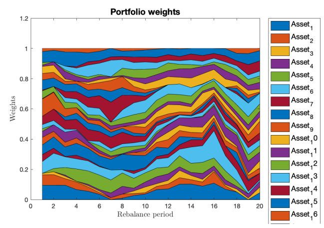
Period	11	12	13	14	15	16	17	18	19	20
OLS	0.53345567	0.52321386	0.52394154	0.52435172	0.50852058	0.41612867	0.43846957	0.38565492	0.37929135	0.39181592
FF	0.47030964	0.47601094	0.47977652	0.47336927	0.45139273	0.36592445	0.36333031	0.32449546	0.28301268	0.29189162
LASSO	0.48710693	0.49417952	0.4833502	0.48674448	0.48165401	0.40291434	0.39359709	0.37621963	0.33067515	0.351413
BSS	0.56841122	0.56044288	0.5602795	0.5608714	0.54373791	0.46010207	0.48143218	0.43721277	0.43101874	0.44336177

Figure 14: Dataset 3 Adjusted R^2

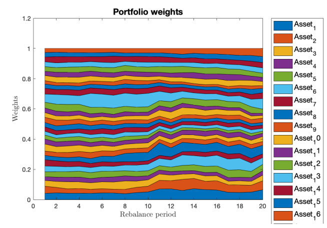
6 Appendix B - Portfolio Weights for Datasets 2 and 3



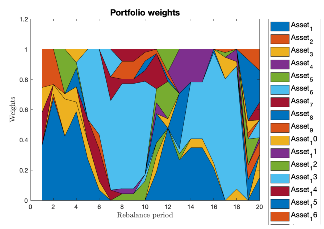
(a) MVO Portfolio Weights



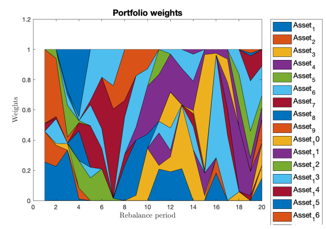
(b) Robust MVO Portfolio Weights



(c) RP Portfolio Weights

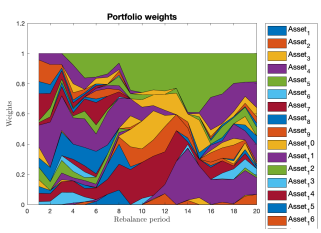


(d) CVaR Portfolio Weights

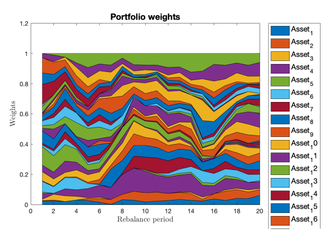


(e) Sharpe Ratio Portfolio Weights

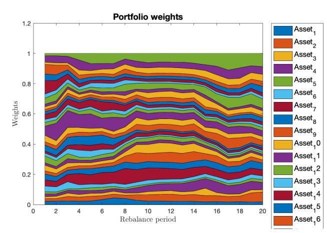
Figure 15: Dataset 2 Portfolio Weights



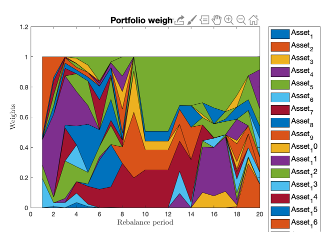
(a) MVO Portfolio Weights



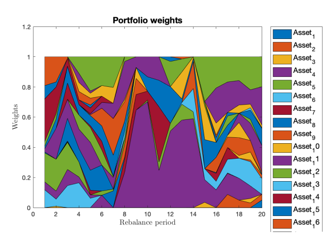
(b) Robust MVO Portfolio Weights



(c) RP Portfolio Weights



(d) CVaR Portfolio Weights



(e) Sharpe Ratio Portfolio Weights

Figure 16: Dataset 3 Portfolio Weights

7 Reference

Maximizing the Sharpe Ratio. (n.d.). Retrived May 18, 2021 from
https://people.stat.sc.edu/sshen/events/backtesting/reference/maximizing