MMF1921 Operational Research

Project 1

Mean-Variance Portfolio Optimization under different Factor Models

by:

Manyi Luo 1003799419

Lujia Yang 1002955563

Xiaolu Xiong 1001126202

May 2022

Abstract

Factor models of security returns decompose the random return on each of a cross-section of assets into factor-related and asset-specific returns. It is widely used by the portfolio manager to forecast the future asset return and construct an investment strategy. To achieve a higher sharp ratio of the portfolio, we want to minimize the investment risk while achieving the target portfolio return. This project aims to compare the model performance and portfolio quality under four different factor models, including OLS (ordinary least squares), FF (Fama-French), and LASSO (Least Absolute Shrinkage and Selection Operator), and BSS (Best Subset Selection). The result suggests the LASSO model performs slightly better and can provide a well-diversified portfolio.

Contents

1 Introduction

In this project, we would like to investigate and compare four different factor models and implement these models to estimate the parameters required for portfolio optimization using mean-variance optimization (MVO) as our investment strategy in portfolio construction. These methods include OLS regression on all eight factors (OLS model), Fama–French three-factor model (FF model), Least absolute shrinkage and selection operator model (LASSO model), and Best subset selection model (BSS model).

Our investment universe in this project consists of 20 stocks ($n = 20$), with the company tickers shown below. We are given monthly adjusted closing prices corresponding to these 20 stocks from 31–Dec–2005 to 31–Dec–2016, and we can use the historical prices to compute our observed asset monthly returns. The reason why we chose adjusted closing prices instead of closing prices when using historical data for a backtest is that the adjusted price provides a more accurate measure: the adjusted closing price reflects a stock's value after accounting for any corporate actions, including dividends, stock splits, and new stock offerings, etc; while the closing price simply offers the cash value of the last transacted price before the market closes. Therefore, it's better to incorporate the adjusted closing price when making comparisons between stock prices that account for the true profitability of value-added stocks and dividends growth in asset allocation.

In addition to the historical prices, we are also given monthly factor returns for eight different factors corresponding to the period 31–Jan–2006 to 31–Dec–2016, which is shown below. This includes the monthly risk-free rate. As a result, we will use the factor models to estimate the asset expected returns and covariance matrix, using them as the inputs for portfolio optimization. Our investment strategy in this project will be mean-variance optimization (MVO). This will allow us to compare the out-of-sample performance of the portfolios built using the different factor models.

Table 1: List of assets by ticker

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

Table 2: List of factors

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

2 Methodology

The essence of factor models is to construct linear regression models with different factors, which are embedded with financial relevance in explaining systematic return and risk. In other words, we can attribute and measure an asset's return and risk based on its exposure to some relevant factors. For this project, we will focus on modelling the asset excess returns, where we measure the return in excess of the risk-free rate, by subtracting the monthly risk-free rate provided with the factor data from our monthly asset returns.

All the factors as shown in Table 2 can be stemmed out from synthetic portfolios of assets with shared properties. Thus, most factors will exhibit some degree of correlation, meaning that our factor models will not respect the ideal environment. Therefore, given the fact that covariances still convey valuable information about the asset risks, we must include these terms in our calculation of our asset covariance matrix.

2.1 OLS model

The ordinary least squares (OLS) model, it's intended to choose the coefficient of a multi-variable linear function of all eight factors introduced in Table 2, while minimizing the sum of the squares of the differences between the values of the stock's excess returns within the given data set. According to Table 1, the multi-factor model is constructed using all eight factors for each asset i, for $i = 1, \cdots, n; n = 20$. T is used to represent the total number of individual observations, i.e. $T = 48$. For asset i, it can be shown as the following.

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

Here, r_i is the return for asset i; r_f is the risk-free rate; α_i is the intercept from regression; f_k is the return of factor k and β_{ik} is the corresponding factor loading; ϵ_i is the stochastic error term of the asset (idiosyncratic risk). As a result, the monthly factor returns (see Table 2) for all eight factors and the monthly risk-free rate are provided in later sections for results and analysis.

In this case, the OLS model can be interpreted as an unconstrained minimization problem: the coefficients of eight factors (i.e. independent variables) are chosen such that the sum of squared residuals are minimized, which can be represented as the following.

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

To simplify the problem which captures the difference between the observed assets' returns and predicted returns by the OLS model, we denote the data set of intercept and observed factors in vector form using $\boldsymbol{X}$ ($\boldsymbol{X} \in \mathbb{R}^{T\times(p+1)}$). The regression coefficient matrix of intercept and factor loadings is represented by $\boldsymbol{B}_i$ ($\boldsymbol{B}_i \in \mathbb{R}^{(p+1)}$), while $||\cdot||$ calculates the squared Euclidean norm, which measures the distance of the vector coordinate from the defined starting point within the vector space.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{1} \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} \boldsymbol{1}\ \boldsymbol{f1} ... \boldsymbol{f8} \end{bmatrix}; \boldsymbol{B}_i = \begin{bmatrix} \alpha_i \\ \boldsymbol{V}_i \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \beta_{i1} \\ ... \\ \beta_{i8} \end{bmatrix}; \boldsymbol{V} = \begin{bmatrix} \beta_{1,1} & ... & \beta_{20,1} \\ \cdots & \cdots & \cdots \\ \beta_{1,8} & ... & \beta_{20,8} \end{bmatrix}_{8\times 20}$$

Through calculation and expansion, we can derive the following formulas and can further derive the First-order Necessary Condition (FONC) respectively.

$$||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 = (\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i)^T(\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i) = \boldsymbol{B}_i^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}\boldsymbol{B}_i + \boldsymbol{r}_i^T\boldsymbol{r}_i$$

$$\min_{\boldsymbol{B}_i} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 = \min_{\boldsymbol{B}_i} \boldsymbol{B}_i^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T \boldsymbol{X}\boldsymbol{B}_i$$

Set $L(\boldsymbol{B}_i) = \boldsymbol{B}_i \boldsymbol{B}_i^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T \boldsymbol{X}\boldsymbol{B}_i$, then:

$$\frac{\partial L(\boldsymbol{B}_i)}{\partial \boldsymbol{B}_i} = 2\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T \boldsymbol{X} = 0$$

Therefore, given 20 assets ($n = 20$), we can obtain the optimal solution, $\boldsymbol{B}_i^*$ and $\boldsymbol{B}^*$, whereas $\boldsymbol{R}$ indicates the matrix of asset returns ($\boldsymbol{R} \in \mathbb{R}^{T\times n}$).

$$\boldsymbol{B}_i^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{r}_i; \boldsymbol{B}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{R} = \begin{bmatrix} \alpha^{*T} \\ \boldsymbol{V}^* \end{bmatrix}_{9\times 20}$$

Note that the residual term can be written as $\epsilon_i = \boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i^*$, and we can use this vector of residuals to calculate the unbiased estimate of the residual variance $\sigma_{\epsilon_i^2} = \frac{1}{T-p-1}||\epsilon_i||_2^2$, whereas T is the total number of observations, p is the number of coefficients and $T - p - 1$ is the degree of freedom.

In order to incorporate mean-variance optimization, the expected return can be written as $\boldsymbol{\mu} = \boldsymbol{\alpha} + \boldsymbol{V^T}\bar{\boldsymbol{f}}$, whereas $\bar{\boldsymbol{f}}$ is used to represent the expected factor returns ($\bar{\boldsymbol{f}} \in \mathbb{R}^p$). Similarly, the covariance matrix $\boldsymbol{Q}$ can be expressed as $\boldsymbol{Q} = \boldsymbol{V^T}\boldsymbol{F}\boldsymbol{V} + \boldsymbol{D}$, whereas the factor covariance effect is $\boldsymbol{F} = cov(\boldsymbol{f}); \boldsymbol{F} \in p \times p$ and the diagonal matrix of residual variances is written as $\boldsymbol{D} \in n \times n$.

$$\boldsymbol{D} = \begin{bmatrix} \sigma_{\epsilon_1^2} & 0 & \cdots & 0 \\ 0 & \sigma_{\epsilon_2^2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_{\epsilon_{20}^2} \end{bmatrix}_{20\times 20}$$

.

2.2 FF model

The Fama–French (FF) three-factor model is a subset of the OLS model, where we use only the Market, Size, and Value factors from Table 2. The FF model is

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

Here, r_i is the return of asset i; r_f is the risk-free rate; α_i is the intercept from regression; $f_m - r_f$ is the excess market return factor and β_{im} is its corresponding factor loading; f_s is the size factor and β_{is} is its corresponding factor loading; f_v is the value factor and β_{iv} is its corresponding factor loading; ϵ_i is the stochastic error term of the asset (idiosyncratic risk). Note that we must only select the pertinent columns corresponding to the three Fama–French factors.

In general, factor model can combine a range of factors to model security returns in the form of linear combination. These factors could be either technical, fundamental, macroeconomic or alternate in viewing the security's performance measurement. Given the fact that FF model is similar to OLS model, we also have similar objective function for minimization, but the optimal solution has smaller matrix dimension due to three factors.

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1} \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} \mathbf{1}\ \boldsymbol{f}\mathbf{1} ... \boldsymbol{f}\mathbf{3} \end{bmatrix}; \boldsymbol{B}_i = \begin{bmatrix} \alpha_i \\ \boldsymbol{V}_i \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \beta_{i1} \\ ... \\ \beta_{i3} \end{bmatrix}; \boldsymbol{V} = \begin{bmatrix} \beta_{1,1} \ ... \ \beta_{20,1} \\ \cdots \\ \beta_{1,3} \ ... \ \beta_{20,3} \end{bmatrix}_{8 \times 20}$$

$$\boldsymbol{B}_i^* = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{r}_i; \boldsymbol{B}^* = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{R} = \begin{bmatrix} \alpha^{*T} \\ \boldsymbol{V}^* \end{bmatrix}_{4 \times 20}$$

The expected returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{Q}$ are obtained using the same approach as described in the previous section, OLS model.

2.3 LASSO model

Lasso stands for Least Absolute Shrinkage and Selection Operator, which is an alternative to the subset selection method for variable selection. It shrinks the regression coefficients toward zero by penalizing the regression model with a penalty term called L1-norm, which is the sum of the absolute coefficients.

The constrained form of LASSO can be expressed as the following.

$$\min_{\boldsymbol{B}_i, y} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2$$
$$s.t. ||\boldsymbol{B}_i||_1 \leq s$$

Here, $||\boldsymbol{B}_i||_1 \leq s$ and $s \in \mathbb{R}_+$, which is a positive parameter that limits the size of the sum of our regression coefficients in LASSO model. In other words, there always exists some value of $s \geq 0$ and some value of $\lambda \geq 0$ such that the optimal solution B_i^* is equivalent for both constrained form and its penalized form of LASSO.

The penalized form of LASSO is calibrated for this project, which can be described as the following.

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

Note that $|| \cdot ||_1$ is used to represent the l_1 norm, which is continuous and convex, but not smooth everywhere.

In order to translate the lasso model to a quadratic programming problem, we need to consider that the penalized form is equivalent with the constrained form. Introducing an auxiliary variable $\boldsymbol{y} \in \mathbb{R}^{p+1}$, we have below problem.

$$\begin{aligned} &\min_{\boldsymbol{B}_i, y} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 + \lambda \mathbf{1}^{\mathbf{T}}\mathbf{y} \\ &s.t. \quad \boldsymbol{y} \geq \boldsymbol{B}_i \\ &\qquad\ \ \boldsymbol{y} \geq -\boldsymbol{B}_i \end{aligned}$$

At optimality, $\boldsymbol{y}$ will be equal to $|\boldsymbol{B}_i|$. Therefore, for each asset class i, we can build a quadratic programming problem.

As

$$\begin{aligned} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 + \lambda||\boldsymbol{B}||_i &= \boldsymbol{r}_i^T\boldsymbol{r}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}(\boldsymbol{B}_i^+ - \boldsymbol{B}_i^-) + (\boldsymbol{B}_i^+ - \boldsymbol{B}_i^-)^T\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{B}_i^+\boldsymbol{B}_i^-) + \lambda\mathbf{1}^T(\boldsymbol{B}_i^+ + \boldsymbol{B}_i^-) \\ &= \boldsymbol{r}_i^T\boldsymbol{r}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}\boldsymbol{B}_i + \boldsymbol{B}_i^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_i + \lambda\mathbf{1}^T|\boldsymbol{B}_i| \end{aligned}$$

while setting $|\boldsymbol{B}_i| = \boldsymbol{B}_i^+ + \boldsymbol{B}_i^-$, $\boldsymbol{B}_i = \boldsymbol{B}_i^+ - \boldsymbol{B}_i^-$ and $\boldsymbol{B}_i^+, \boldsymbol{B}_i^- \geq 0$.

That is, we can write $\boldsymbol{B}_i$ as $\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{I}\end{bmatrix} \times \begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix}$, $|\boldsymbol{B}_i|$ as $\begin{bmatrix}\boldsymbol{I} & \boldsymbol{I}\end{bmatrix} \times \begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix}$, whereas $\boldsymbol{I}$ is an $p+1$ identity matrix, and p is the number of factors.

$$\begin{aligned} &= \boldsymbol{r}_i^T\boldsymbol{r}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}[\boldsymbol{I} - \boldsymbol{I}]\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} + ([\boldsymbol{I} - \boldsymbol{I}]\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix})^T\boldsymbol{X}^T\boldsymbol{X}[\boldsymbol{I} - \boldsymbol{I}]\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} + \lambda\mathbf{1}^T[\boldsymbol{I}\ \boldsymbol{I}]\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} \\ &= (\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix})^T[\boldsymbol{I} - \boldsymbol{I}]^T\boldsymbol{X}^T\boldsymbol{X}[\boldsymbol{I} - \boldsymbol{I}]\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} + (\lambda\mathbf{1}^T[\boldsymbol{I}\ \boldsymbol{I}] - 2\boldsymbol{r}_i^T\boldsymbol{X}[\boldsymbol{I} - \boldsymbol{I}])\begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} + \boldsymbol{r}_i^T\boldsymbol{r}_i \end{aligned}$$

Therefore , our quadratic programming is as follows.

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{f}^T\boldsymbol{x} \\ &s.t. \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{H} &= \begin{bmatrix}\boldsymbol{I} & -\boldsymbol{I}\end{bmatrix}^T \boldsymbol{X}^T\boldsymbol{X}\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{I}\end{bmatrix}^T \\ \boldsymbol{f} &= -[\boldsymbol{I} - \boldsymbol{I}]^T\boldsymbol{X}^T\boldsymbol{r}_i + \frac{\boldsymbol{\lambda}}{2}[\boldsymbol{I}\ \boldsymbol{I}]^T\mathbf{1} \\ \boldsymbol{x} &= \begin{bmatrix}\boldsymbol{B}_i^+ \\ \boldsymbol{B}_i^-\end{bmatrix} \end{aligned}$$

2.4 BSS model

The constrained form of Best Subset Selection (BSS) model incorporates all eight factors as inputs. The model is constructed by solving the following optimization problem:

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

The l_0 norm can impose a cardinality limit on the coefficients, which is to limiting the number of factors used in the model. Consider adding the constraints for the BSS model, it can be expressed as

$$\begin{aligned}&\min_{\boldsymbol{B}_i,\boldsymbol{y}} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 \\ &s.t. \quad L\boldsymbol{y} \leq \boldsymbol{B}_i \leq U\boldsymbol{y} \\ &\qquad \boldsymbol{1}^T\boldsymbol{y} \leq \boldsymbol{K}\end{aligned}$$

where $y_j \in \{0, 1\}$ for $j = 1, \cdots, p+1$, representing auxiliary binary variables. L and U are lower and upper bounds. K is the cardinality constraints ($K \in \mathbb{Z}_+$), which is a positive integer.
since $\boldsymbol{r}_i^T\boldsymbol{r}_i$ is a constant term and can be ignored

$$||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 = \boldsymbol{B}_i^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}\boldsymbol{B}_i + \boldsymbol{r}_i^T\boldsymbol{r}_i$$

$$\min_{\boldsymbol{B}_i,\boldsymbol{y}} ||\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i||_2^2 \Longleftrightarrow \min_{\boldsymbol{B}_i,\boldsymbol{y}} \boldsymbol{B}_i^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_i - 2\boldsymbol{r}_i^T\boldsymbol{X}\boldsymbol{B}_i$$

The above model can be converted to a quadratic programming problem which can be solved with quadprog().

$$\begin{aligned}&\text{minimize} \quad \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ &s.t. \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\end{aligned}$$

Let $\boldsymbol{H} = \boldsymbol{X}^T\boldsymbol{X} \in \mathbb{R}^{(2p+2)\times(2p+2)}$, $\boldsymbol{c}^T = -2\boldsymbol{r}_i^T\boldsymbol{X} \in \mathbb{R}^{(2p+2)\times 1}$.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{B}_i \\ \boldsymbol{y}_i \end{bmatrix} \in \mathbb{R}^{(2p+2)\times 1}, \quad \boldsymbol{A} = \begin{bmatrix} -\boldsymbol{I} & L\boldsymbol{I} \\ -\boldsymbol{I} & U\boldsymbol{I} \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ ... \\ 0 \end{bmatrix}_{(2p+2)\times 1}$$

And

$$\boldsymbol{Aeq} = \begin{bmatrix} \boldsymbol{0}_{1\times(p+1)} & \boldsymbol{1}_{1\times(p+1)} \end{bmatrix}, \quad \boldsymbol{beq} = \begin{bmatrix} K \\ ... \\ K \end{bmatrix}_{(p+1)\times 1}$$

Thus the above mixed-integer quadratic program (MIQP) can be solved in MATLAB with Gurobi functions. After obtaining $\boldsymbol{x}$, specifically $\boldsymbol{x}_i$ for each asset, we can further derive $\boldsymbol{B}_i = [\boldsymbol{1}_{(p+1)\times 1} \ \boldsymbol{0}_{(p+1)\times 1}]$ as well as $\boldsymbol{B} = [\boldsymbol{B}_1 \ \boldsymbol{B}_2 \ \cdots \ \boldsymbol{B}_{20}]$, and calculate $\boldsymbol{\mu} = \alpha^* + \boldsymbol{V}^{*T}\bar{\boldsymbol{f}}$ ($\bar{\boldsymbol{f}}$ represents the vector of expected factor returns) as well as $\boldsymbol{Q} = \boldsymbol{V}^{*T}\boldsymbol{F}\boldsymbol{V}^* + \boldsymbol{D}$, similar as what we derived in previous sections. Note that the residual term can be written as $\epsilon_i = \boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B}_i^*$, and we can use this vector of residuals to calculate the unbiased estimate of the residual variance $\sigma_{\epsilon_i^2} = \frac{1}{39}||\epsilon_i||_2^2$.

3 R^2 for In-Sample Analysis

The coefficient of determination, R^2, is commonly used in the context of predictive statistical models. It is based on the proportion of total variation of the outcomes explained by the model and can measure how well the model is fitted. Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model: the statistics take over-fitting into account, since adding more independent variables always tends to increase the R-squared value.

$$SS_{\text{res}} = \sum_i (y_i - \hat{y}_i)^2 = \sum_i (r_i - \boldsymbol{X}\boldsymbol{B}_i)^2 = \sum_i e_i^2$$

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2 = \sum_i (r_i - \bar{r})^2$$

$$Adjusted\ R^2 = 1 - \frac{SS_{\text{res}} \cdot (t - p - 1)}{SS_{\text{tot}} \cdot (t - 1)}$$

where n is the number of asset classes, p is the number of factors, and t is the number of observations. We will use the adjusted R^2 as a metric in the in-sample analysis. We will create a portfolio and calculate the expected return for every model of each period. By observing the adjusted R^2 of each model of the five periods, we will have an idea of how the models are performing.

4 Sharp Ratio for Out-of-Sample Analysis

In order to evaluate our model's out-of-sample performance, we will calculate the portfolio's average return, variance and sharp ratio for each model of each test period. In addition, we will calculate the risk metrics Sharpe Ratio. Sharpe ratio is widely used for getting risk-adjusted return, and it can help us to understand the model's risk adjusted performance: a greater the Sharpe Ratio indicates better risk-adjusted performance, and it can be expressed as the below function.

$$\text{Sharpe Ratio} = \mathbb{E}[R_i - R_f]/\sigma_a$$

where R_i is the return of portfolio. R_f is the risk-free rate and σ_p is the standard deviation.

5 Portfolio Optimization

The mean-variance analysis is a tool that investors use to help spread risk in their portfolios. Using this approach, the investors can measure an assets' risk, expressed as the "variance", then compares that with the assets' likely return. The goal of a mean-variance optimization is to maximize an investments' reward based on its risk, as we intend to minimize variance subject to a target expected return to optimize our portfolios. Therefore, we can use the four models to estimate the expected returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{Q}$ for each factor model, and then use it as the input for Mean-Variance Optimization (MVO) problem.

$$
\begin{aligned}
\text{minimize} \quad & \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \\
s.t. \quad & \boldsymbol{\mu}^T \boldsymbol{x} \geq R \\
& \boldsymbol{1}^T \boldsymbol{x} = 1 \\
& \boldsymbol{x} \geq 0
\end{aligned}
$$

Here, $\boldsymbol{x} = [x_1 \ x_2 \ \cdots x_2 0]$, where $x_i \geq 0$ for $i = 1, 2, \cdots, 20$, which represents the weight of individual asset in the portfolio without allowing short-selling. Investors need to allocate asset appropriately to achieve the target rate of return. $\boldsymbol{x}$ is the weight allocate to each asset, the sum of which is 1.

Therefore, we will use the MVO results to further evaluate the impact of the four different factor models in the out-of-sample performance section. To provide a highly summarized overview, we simulate an investment horizon within the range of five years (i.e. beginning of 2012 until the end of 2016). For each of the five investment periods, we re-calibrate and re-calculate our portfolio's expected returns $\boldsymbol{\mu}$ and the covariance $\boldsymbol{Q}$ by using preceding historical returns at the end of each year, and eventually calculate the asset allocation $\boldsymbol{x}$ for the optimal portfolios.

6 Simulation Results

6.1 LASSO Model: The impact of Lambda

One obvious advantage of LASSO regression is that it produces more interpretive models that incorporate only a reduced set of predictors. It might perform better when there exist large coefficients for certain factors. To ensure a decent performance of LASSO, selecting a good value of λ is critical. Thus conducting in-sample and out-of-sample analyses with a different set of lambdas to evaluate the model performance is needed.

6.1.1 In-Sample Analysis

To analyze the performance of the model for in-sample testing, adjusted R^2 would be a good metric. As it is a good indicator of model accuracy. We computed the adjusted R^2 for Lasso model with a set of $\lambda = 0, 0.01, 0.02...0.06$, we also computed the 5-year average for the adjusted R^2. We can observe from the table that as λ increases, the adjusted R^2 tends to decrease. That is, as λ increases, the model accuracy drops. As mentioned in the previous section of the methodology, it shrinks the regression coefficients toward zero by penalizing the regression model with a penalty term λ. As λ increases to infinite, the impact of the shrinkage penalty will increase, and the ridge regression coefficients will get close to zero. As we know that an ideal model should only contain two to five factors, the value of the λ should be just right, not too big that all the coefficients are zero, and not too small that allows the model to contain too many unnecessary factors. Therefore we would also like to consider the out-of-sample analysis to choose a good model. As we can see from the below table, the 5-year average portfolio return is also provided and will be further discussed in the next section.

Year	λ**=0**	λ**=0.01**	λ**=0.02**	λ**=0.03**	λ**=0.04**	λ**=0.06**
2012	0.480	0.438	0.0393	0.359	0.334	0.296
2013	0.477	0.405	0.0338	0.296	0.268	0.220
2014	0.436	0.324	0.0244	0.198	0.164	0.101
2015	0.397	0.285	0.0166	0.107	0.071	0.012
2016	0.439	0.343	0.0244	0.167	0.071	0.017
5yr Avg R^2	0.446	0.359	0.277	0.225	0.181	0.129
5yr Avg PortReturn	0.68%	0.69%	0.68%	0.69%	0.68%	0.69%

Table 1: Average adjusted R-Squared for LASSO under different λ

6.1.2 Out-of-Sample Analysis

The below plot 1 combines the 5-year average of adjusted R^2 and the portfolio's excess return. We can observe that the adjusted R^2 is decreasing and the average portfolio excess return is oscillating with a slightly increasing trend. It appears that the sample portfolio return is slightly increasing while λ increases.

In order to see the effect of λ more clear, we plotted the out-of-sample portfolio value with different λ in 2. And we can see that indeed as λ increases, the sample portfolio value increases, and the increase of return is quite small. we believe a relatively small λ could better balance the model accuracy versus potential portfolio return. Therefore, $\lambda = 0.01$ is selected to be used in portfolio optimization.

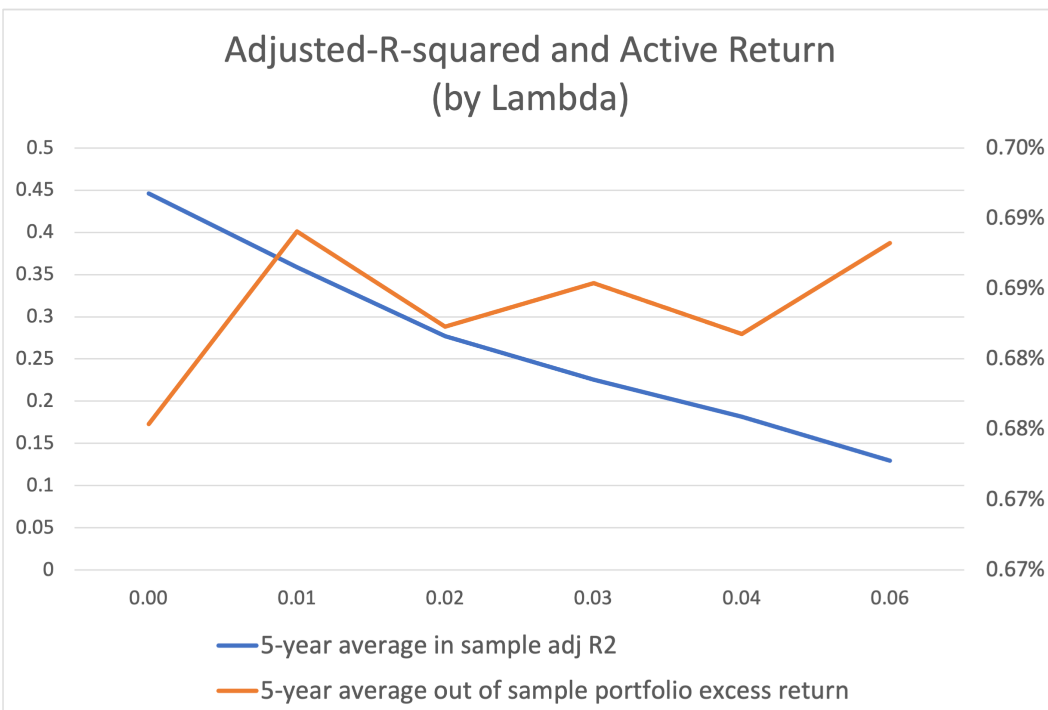

Figure 1: Comparison of Adjusted R-Square

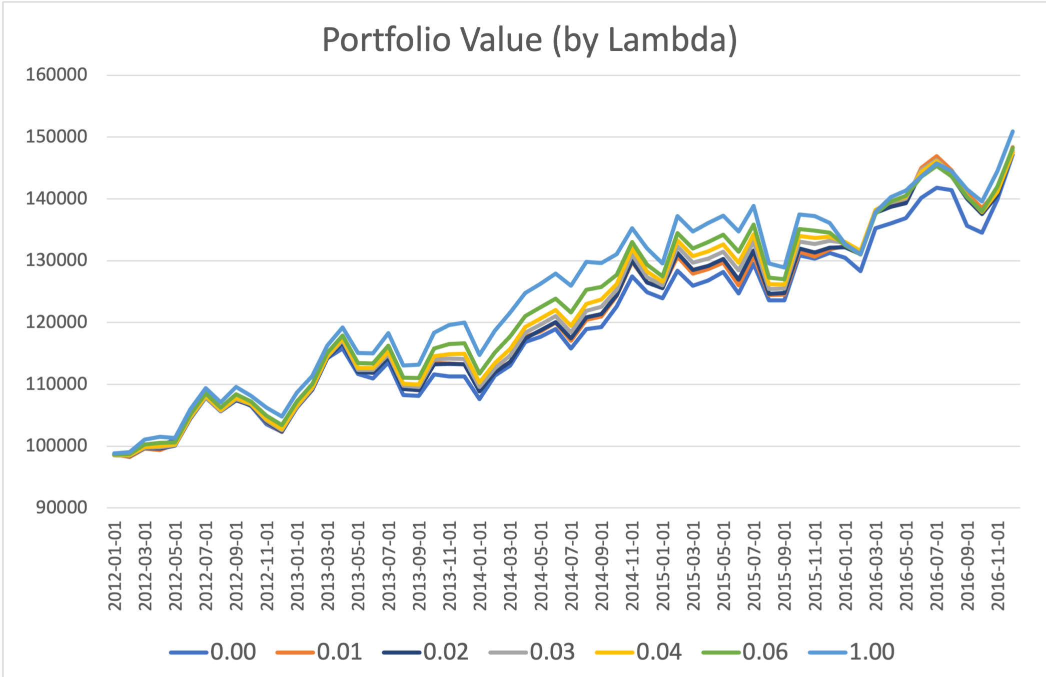

Figure 2: Comparison of Portfolio Values

6.2 BSS Model: The impact of K

By considering all possible combinations of independent variables, the BSS model is intended to find the subset of variables that best predict the outcome when building a regression model. It removes irrelevant variables by choosing the model given different factors, through both perspectives of determining the best model of each size and the best overall model.

Therefore, the advantage of performing the BSS model can be summarized as the following. First, it improves the accuracy of the regression model when eliminating unnecessary predictors and forming a generalization identity. Second, it provides a reproducible approach to reduce the number of predictors, which eventually makes the model easier to interpret the data by having less tendency to overfit.

6.2.1 In-Sample Analysis

Similar to the rationale provided in the LASSO model, the adjusted R^2 is a suitable metric as it poses penalty on bigger models when there are additional variables that provide least improvement and eventually lead to a low accuracy.

We computed the adjusted R^2 for BSS model with a set of $= 4, 5, ...8$, using four factor models as the basis. Therefore, ranging between five investment periods to investigate the impact of cardinality, we can observe that as K increases, the adjusted R^2 faces a drastic growth between $K = 4$ and $K = 5$, while the trend smooths out from adding additional K. This is also shown in the table below, as $K = 5$ is optimal solution. In this case, K acts as a penalty term and the impact of the shrinkage penalty will increase as K grows beyond the optimality. In the ideal model, the value of K should be just right, not too big that all it allows the model to contain too many unnecessary factors, or too small with low adjusted R^2, indicating the model lacks ability to interpret. Therefore, we would like to further conduct out-of-sample analysis in later section for model selection.

Year	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**
2008	0.3557	0.3568	0.3569	0.3569	0.3569
2009	0.3234	0.3265	0.3269	0.3270	0.3270
2010	0.3253	0.3261	0.3261	0.3261	0.3261
2011	0.3033	0.3064	0.3070	0.3070	0.3070
2012	0.3257	0.3290	0.3295	0.3295	0.3295
5yr Avg R^2	0.3267	0.3290	0.3293	0.3293	0.3293
5yr Avg PortReturn	0.67%	0.67%	0.65%	0.65%	0.65%

Table 2: Average adjusted R-Squared for BSS under different K

6.2.2 Out-of-Sample Analysis

The below plot 3 combines the 5-year average of adjusted R^2 and the portfolio excess return. We can observe that the value of adjusted R^2 examines a general increasing trend with a decreasing slope, whereas the average portfolio excess return decreases drastically between $K = 5$ and $K = 6$, remaining a slightly flat trend in other K values.

To see the effect of K more clear, we plotted the out-of-sample portfolio value with different values of K in 4, and we can see that $K = 5$ outperforms among the five curves, while the rest of the parameter selections are generally consistent with each other, as the portfolio value grows through time. As a result, we can conclude that a relatively intermediate selection of the number of factors, specifically $K = 5$, can create a better balance of the model accuracy versus potential portfolio return. Therefore, K is selected to be used in portfolio optimization.

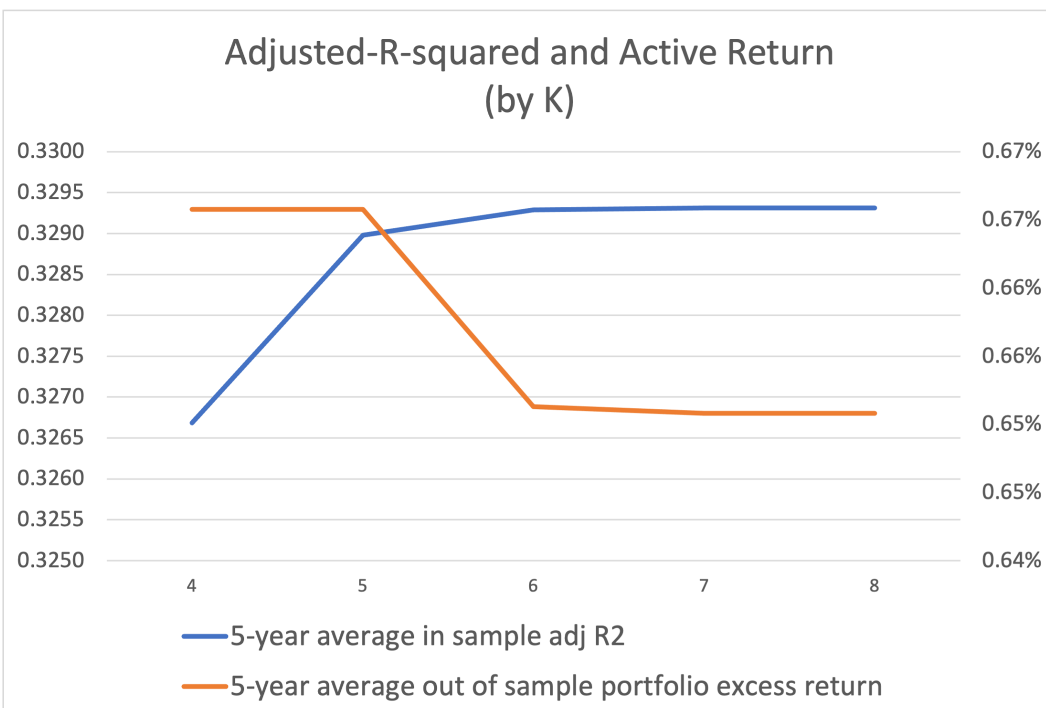

Figure 3: Comparison of Adjusted R-Square

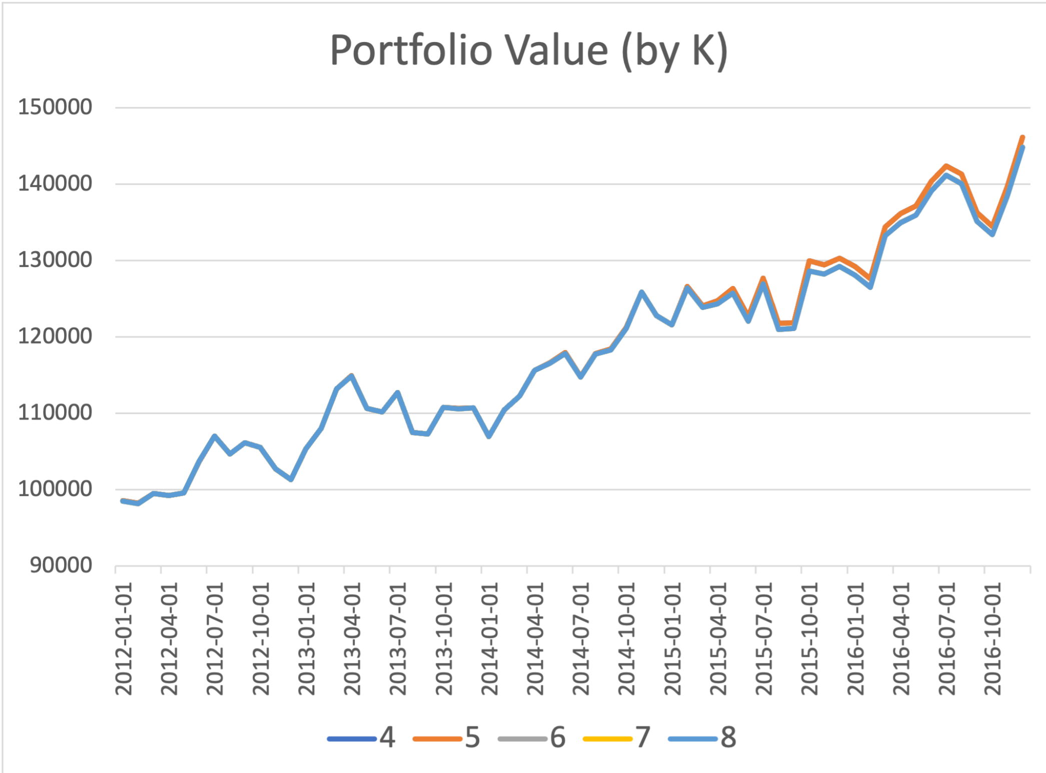

Figure 4: Comparison of Portfolio Values

6.3 Portfolio Simulation and Analysis

Based on the previous parameter selection process for the LASSO and BSS model. We select a reasonable $\lambda = 0.01$ for LASSO, and K = 5 for BSS, to generate final result. Additionally, for the LASSO method, we want to set an appropriate upper and lower bond to ensure the feasible region of finding the optimal value is not too large or too small, which can save the computational resources while avoiding a limited feasible region. The upper and lower bound of the LASSO model are set to be -4 and 4 respectively. The following analysis for both in-sample and out-of-sample performance is based on the final selected model parameters.

6.3.1 In-Sample Model Performance

In this project, we use adjusted-R-square (R^2_{adj}) as the measure of fit of the regression models to assess their in-sample quality. We prefer to use an adjusted 0squared rather than an R-squared because it provides a more precise view of the trade-off between model complexity and prediction accuracy. It considered how many independent variables are added to the model when estimating the asset expected excess return. The higher the R^2_{adj}, the better the model quality.

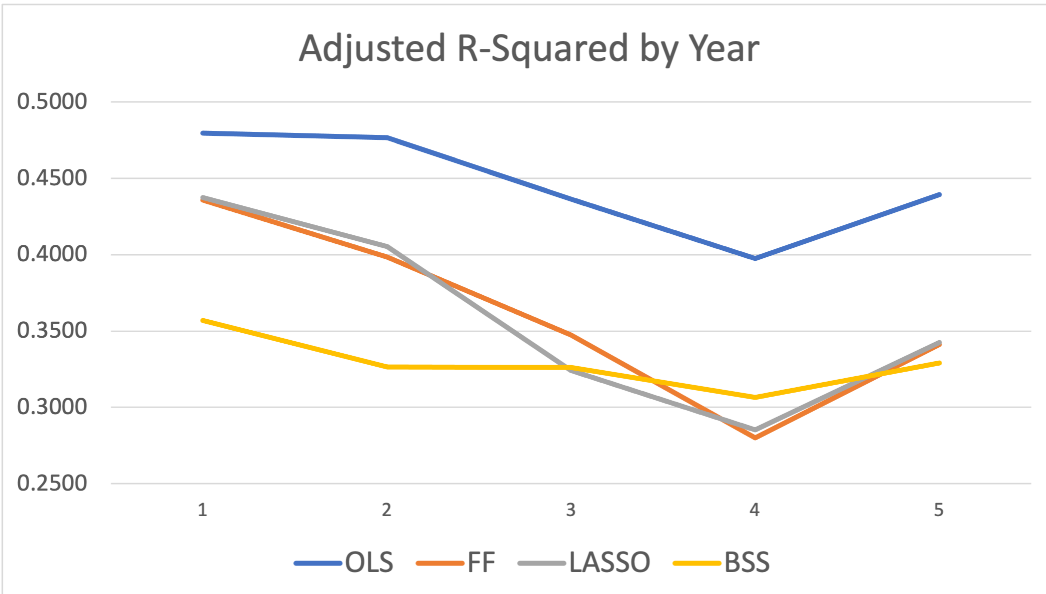

Figure 5: Comparison of portfolio values

Year	**OLS (8 factor)**	**FF (3 factor)**	**LASSO (λ = 0.01)**	**BSS (K = 5)**
2012	0.4797	0.4358	0.4376	0.3568
2013	0.4767	0.3984	0.4053	0.3265
2014	0.4364	0.3473	0.3241	0.3261
2015	0.3974	0.2800	0.2854	0.3064
2016	0.4393	0.3411	0.3426	0.3291

Table 3: Annual adjusted R-Squared for different factor models

Based on the figure and table, we find the OLS model has the best quality with the highest R^2_{adj} over years, where the mean value is 0.4459. On the other hand, the R^2_{adj} for the BSS model is relatively low, with the mean value for a different testing year equal to 0.329.

It is reasonable because there are only 8 factors in the database, which is not too large, and the OLS model tends to include all 8 variables in estimation. However, by set K = 5, the BSS model only includes 5 factors, which reduces its accuracy compared to the OLS model. FF and LASSO models have similar R^2_{adj} over different years. Based on the estimated α, β for each asset using the LASSO model, we observed the coefficient is large for certain factors and close to 0 for the others, which is significantly different from the OLS model (where the coefficient is more balanced for each of the 8 factors). So intuitively, LASSO only heavily weighs 3-5 factor and the FF model only take 3 factors. Moreover, there may be some factor correlation between FF selected ($Mkt - RF$,SMB, HML) and the other factors (RMW, CMA, Mom, $STRev$, $LTRev$), which means there can be alternative to each other. Therefore, the performance of FF and LASSO may be similar.

6.3.2 Portfolio Performance

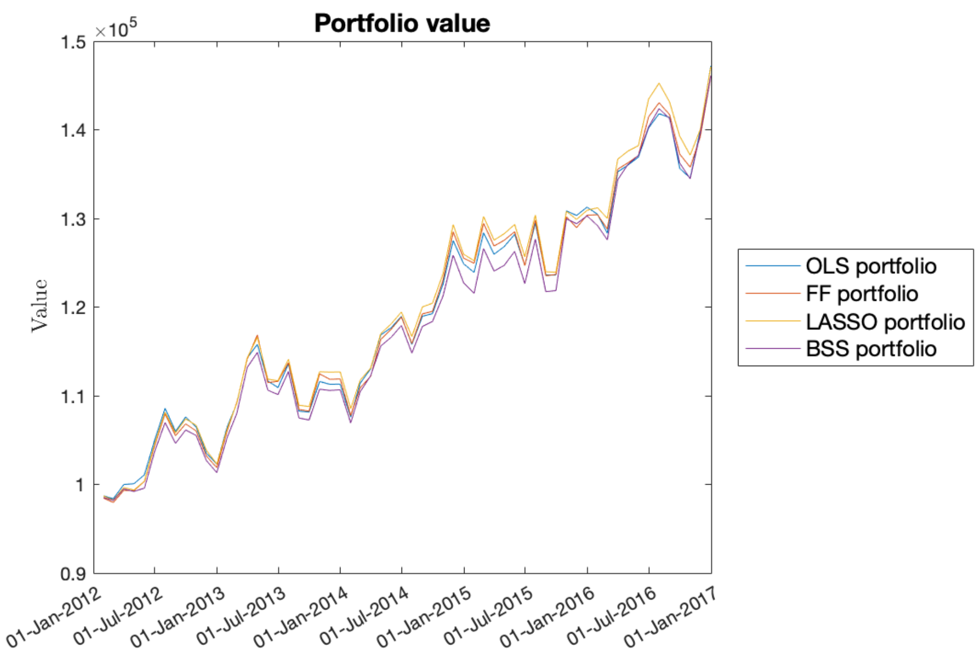

Figure 6: Comparison of portfolio values

Factor Model	OLS	FF	LASSO	BSS
Average Excess Return	0.68%	0.67%	0.67%	0.67%
Volatility	2.61%	2.62%	2.60%	2.62%
Sharpe Ratio	25.87%	25.39%	25.92%	25.41%

Table 4: Performance Metrics for Portfolios

In the simulated portfolio, we found the portfolio values over years are close for these four-factor models. The table and plot demonstrate better performance of the LASSO portfolio and the worst performance of the BSS portfolio. The gap in portfolio values among these models is during the year 2015, which can be explained by the difference in portfolio asset allocation in this rebalance year based on the plot in the next section. The average excess return for FF, LASSO, and BSS is 0.67% and the average excess return for OLS is 0.01% higher than the other three. The volatility for OLS and LASSO is 2.61% and 2.6% respectively, which are better than FF and BSS. It is because FF and BSS models include fewer factors in the prediction, leading to lower prediction accuracy. OLS includes 8 factors and has the best in-sample quality. However, it failed to provide the best portfolio in the out-of-sample portfolio simulation. There could be a model over-fitting issue if we use the OLS method. Finally, we observed the LASSO portfolio is the best because it has the highest Sharpe ratio (25.92%), meaning that the LASSO method can achieve the lowest risk while meeting the target return. In other words, by taking one unit of extra risk, investors can benefit more from the LASSO portfolio compared to the other three portfolios. One advantage of LASSO is that it favors subsets of features that have less collinearity, but BSS may fail to exclude highly correlated variables, resulting in higher asset correlation risks.

6.3.3 Portfolio Asset Allocation

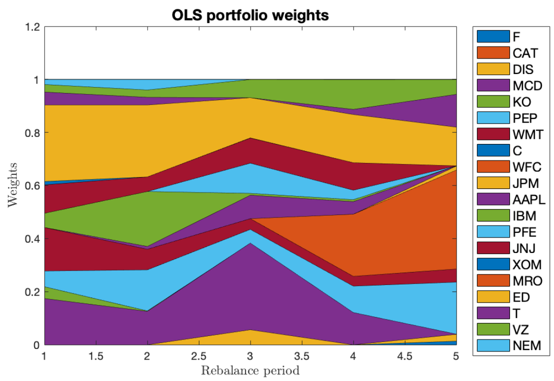

Figure 7: OLS Portfolio

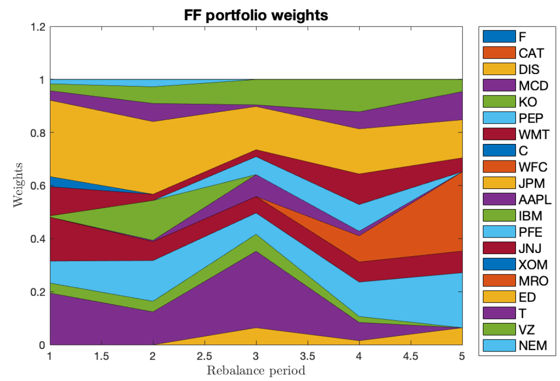

Figure 8: FF Portfolio

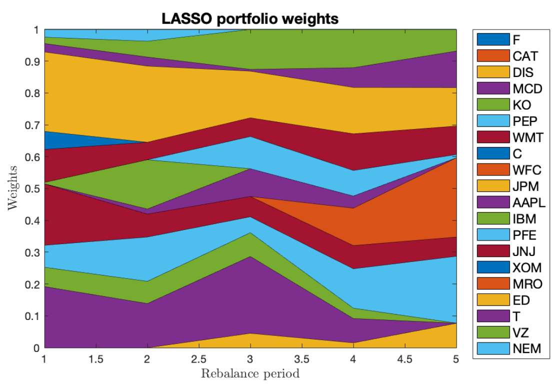

Figure 9: LASSO Portfolio

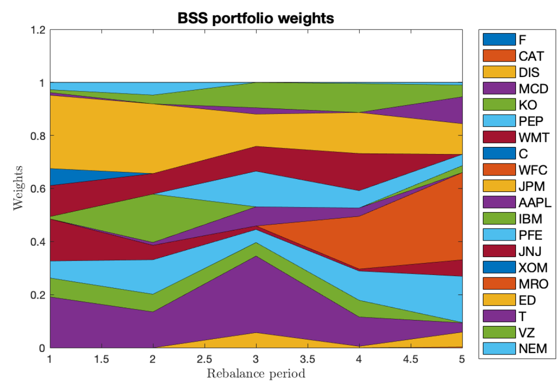

Figure 10: BSS Portfolio

Through presenting the four area plots, we can see the investment asset are well-diversified under each of these four portfolios, as the optimal weights aren't concentrated after portfolio re-balancing. The plot of portfolio weights shows the change in asset allocation at each rebalancing point. For example, all these portfolios tend to hold a large portion of AAPL in the first 4 years and sell about half of its share in the year 2016. It makes sense because we can see the stock return for AAPL drop in the year 2015 (see Figure 11), thus the model forecasting may result in a lower return for AAPL in the next year and reduce the weights on AAPL in 2016, though the stock price rebound. Meanwhile, these four portfolios hold ticker C for the first 2 years, and then they removed this asset. It can also be explained by the fact that the price of ticker C is in a slightly decreasing trend. However, the weight for

IBM (the green area at the bottom of each plot) is different, as the stock return for IBM fluctuates over the 10 years. It is harder for the linear factor model to forecast asset return with a vague trend. OLS portfolio only holds IBM in the first 2 years. FF and BSS portfolios hold a small portion of IBM without obvious change in weights during the 5 years. It may be because these two models include fewer factors, and the predicted return is more stable over different years. Additionally, LASSO holds about 5% of IBM in the first 4 years and then reduce its holding in the year 2016 to achieve a better asset diversification, resulting in a better Sharpe ratio.

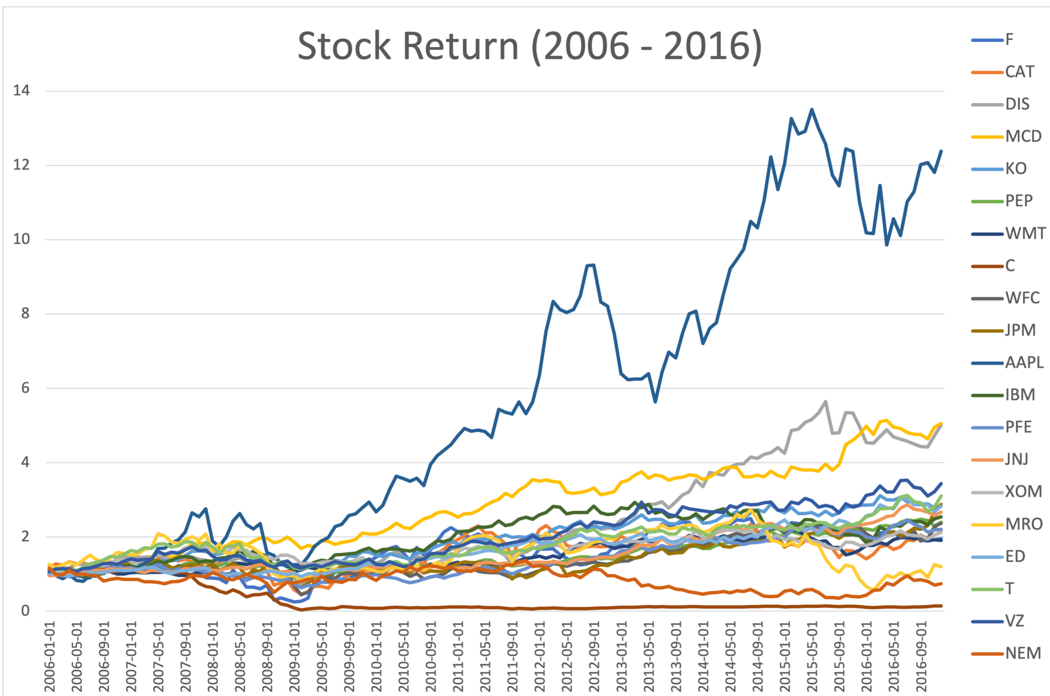

Figure 11: 10 Year Stock Return

7 Conclusion and Discussion

Mean-variance optimization (MVO) is a strategy that investors use to diversify portfolio risk. Investors measure an asset's risk and then compare it to the asset's expected return. The goal of the mean-variance optimization is to maximize the return on investment in terms of risk. The input of MVO is the expected asset return and the volatility, which can be estimated using factor models. The OLS model considered all factors and is more complex than the Farma-French three-factor model. OLS model has the highest in-sample prediction accuracy but does not necessarily achieve the best portfolio sharp ratio in the out-of-sample validation. A reasonable selection of λ and K is important for LASSO and BSS models. We see there is a trade-off between the model accuracy and the simulated portfolio value. In the portfolio simulation with ($\lambda = 0.01$ and K = 5), we found LASSO provides the best Sharpe ratio. The asset allocation result indicates all these four-factor methods can achieve decent performance. To react to the market timely, investors need to select an appropriate factor model to forecast the asset return when applying the mean-variance optimization strategy.

The reason why we conduct in-sample and out-of-sample analyses are to evaluate the financial performance of the factor models and portfolios. To further discuss our model results, we propose that providing adjustments to the current BSS model can yield potential improvements to the model accuracy, for example: further defining the variable lower bounds might help to generate a more persuasive adjusted-R-square ratio compared to the OLS model, as the in-sample data could become a better fit in penalizing models when using additional factors. However, having the model to better explain the variability of the expected excess return couldn't guarantee the out-of-sample performance for other model portfolios.