Comparison of Factor Models for Parameter Estimation in Portfolio Optimisation

Pulkit Gaur
Hongyi (Henry) Wu
Jiaqi (Key) Feng

Date: 13 May 2020

Contents

1 Introduction

This report discusses the comparison of four different factor models - (Ordinary Least Squares**(OLS)** Regression, Fama–French**(FF)** three-factor model, Least absolute shrinkage and selection operator**(LASSO)** model, and Best subset selection**(BSS)** model), to estimate the parameters for portfolio optimisation.

Our portfolio will comprise of 20 stocks and we will use monthly **adjusted closing prices** from 31-Dec-2005 to 31-Dec-2016. We choose to use adjusted prices to incorporate the **corporate actions** like stock splits, dividends etc as adjusted close subtracts dividends paid from the close price and corrects the price as per the split ratio.

We further use the monthly factor data along with the asset excess returns, where we measure the return in excess of the risk-free rate, to calibrate our models to further estimate the asset expected returns and covariance matrix. We will use the estimates as parameters for our mean variance portfolio optimisation to select the optimal weights. And lastly, we will compare the in sample and out of sample performance of each model across different investment periods.

2 Factor Models

Asset returns are usually strongly related to the market financial factors like capital size and profitability. To define, **factor models** are financial models that incorporate factors (macroeconomic, fundamental and statistical) to determine the market equilibrium and calculate the required rate of return. Such models associate the return of a security to single or multiple risk factors in a linear model.

In this project, the returns of eight financial factors are utilized to estimate the expected returns and covariance matrix of twenty assets. This method can capture those assets' systematic return and risk while reducing the noises. The estimated returns and covariance will be used as input parameters for further portfolio optimization. In total, four factor models are applied and their performance would be discussed.

2.1 OLS Model

OLS is short for Ordinary Least Square. In statistics, it is a type of Ordinary Least Square method used to estimate the unknown parameters or coefficients of a linear regression model by minimizing the sum of the squares of the differences between the observed values and those predicted by the linear function.

In our problem, the dependent variables are assets returns and the independent variables are returns of the eight factors. This can be expressed as below

$$r_{it} = \alpha_i + \sum_{k=1}^{8} \beta_{ik} f_{kt} + \varepsilon_{it}$$

where $\boldsymbol{r_{it}}$ stands for the return for asset $\boldsymbol{i}$ at time $\boldsymbol{t}$, $\boldsymbol{f_{kt}}$ stands for the return for factor $\boldsymbol{k}(k = 1, 2, ..., 8)$ at time $\boldsymbol{t}$. $\boldsymbol{\alpha}_i$ and $\boldsymbol{\varepsilon_{it}}$ are intercept and residuals.

Meanings of factors $\boldsymbol{f}_k$ are displayed below

Table 1: Factor list of OLS model

$\boldsymbol{f}_1$	$\boldsymbol{f}_2$	$\boldsymbol{f}_3$	$\boldsymbol{f}_4$	$\boldsymbol{f}_5$	$\boldsymbol{f}_6$	$\boldsymbol{f}_7$	$\boldsymbol{f}_8$
Mkt RF	SMB	HML	ST Rev	RMW	CMA	Mom	LT Rev

In vector notation, let $\boldsymbol{r}_i$ be the return vector for asset $\boldsymbol{i}$, $\boldsymbol{X}$ be $[1, f_1, f_2, ..., f_8]$ and $\boldsymbol{B}_i$ be $[\alpha_i, \beta_{i1}, \beta_{i2}, ..., \beta_{i8}]$. For OLS method, we have the following unconstrained minimization problem,

$$\min_{\boldsymbol{B_i}} \quad \left\| \boldsymbol{r}_i - \boldsymbol{X B_i} \right\|_2^2$$

where $\left\| . \right\|_2$ is the ℓ_2 norm operator. If we expand the ℓ_2 norm and convert the constraints to the standard form, we get :

$$\min_{\boldsymbol{B_i}} \quad \boldsymbol{B_i}^{\boldsymbol{T}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X B_i} - 2\boldsymbol{r_i}^{\boldsymbol{T}} \boldsymbol{X B_i}$$

differentiate the function above and let it equal to 0, we get :

$$2\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X B_i} - 2\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{r}_{\boldsymbol{i}}^{\boldsymbol{T}} = \boldsymbol{0}$$

Then we get,

$$\boldsymbol{B}_i^* = (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{r_i}$$

The residuals for asset i from the regression are :

$$\varepsilon_i = \boldsymbol{r_i} - \boldsymbol{X B_i^*}$$

Thus, the corresponding residual variance is:

$$\sigma_{\varepsilon_i}^* = \frac{1}{T - p - 1} \left\| \boldsymbol{\varepsilon_i} \right\|_2^2$$

where T is the degree of freedom and p is the number of factors used in the regression. Next, let

$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_{20}]^T$ and $\boldsymbol{B} = [_{\mathbf{1}}, _{\mathbf{2}}, ..., _{\mathbf{20}}]$since we have 20 assets in total, and let $\boldsymbol{V}$ satisfies:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\alpha^T} \\ \boldsymbol{V} \end{bmatrix}_{9\times 20}.$$

As for factor returns, we let $\overline{f}$ be the average return for each factor, which is a 9×1 vector. Also, F is the covariance matrix of $[1, f_1, f_2, ..., f_8]$. Now we can output μ and Q as below :

$$\boldsymbol{\mu} = \boldsymbol{\alpha} + \boldsymbol{V}^T\overline{\boldsymbol{f}}$$
$$\boldsymbol{Q} = \boldsymbol{V}^T\boldsymbol{F}\boldsymbol{V} + \boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements being $\sigma^2_{\varepsilon_i}$:

$$\boldsymbol{D} = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{\varepsilon_{20}} \end{bmatrix}_{20\times 20}$$

In addition, we added the adjusted $\overline{R}^2$ as another output. For each asset i, we have:

$$\overline{r_i} = \frac{1}{T}\sum_{t=1}^{T} r_{it}$$

and then we can get total sum of square SS_{tot_i} and residual sum of squares SS_{res_i}

$$SS_{tot_i} = \sum_{t=1}^{T}(r_{it} - r_{it})$$
$$SS_{res_i} = \sum_{t=1}^{T}(r_{it} - \tilde{r_{it}})$$

where $\tilde{r_{it}}$ is element in $\boldsymbol{X}\boldsymbol{B_i^*}$. Then $\overline{R}^2$ can be interpreted as :

$$\overline{R}^2 = 1 - \frac{SS_{res_i}}{SS_{tot_i}}\frac{T-1}{T-1-p}$$

We calculated the variance directly and hence multiplied with $\frac{T-1}{T-1-p}$

2.2 FF Model

FF model is quite similar to OLS model. The only difference is that only three factors, which are Market, Size, and Value factors, are selected in the regression as shown below

Table 2: Factor list of FF model

$\boldsymbol{f}_1$	$\boldsymbol{f}_2$	$\boldsymbol{f}_3$
Mkt RF	SMB	HML

According to the Fama-French three-factor model, over the long-term, small companies outperform large companies, and value companies beat growth companies. In this model, $\boldsymbol{X}$ will be $[1, f_1, f_2, f_3]$ and $\boldsymbol{B}_i$ be will $[\alpha_i, \beta_{i1}, \beta_{i2}, \beta_{i3}]$. For FF method, we have the same unconstrained minimization problem as OLS,

$$\min_{\boldsymbol{B_i}} \quad \left\|\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B_i}\right\|_2^2$$

The process afterwards would be the same as with the OLS model.

2.3 LASSO Model

LASSO stands for **"Least Absolute Shrinkage and Selection Operator"**. The term is often used synonymous with 1-norm regularization. In its most basic form it refers to 1-norm regularized linear regression. The decisive property of LASSO regression is that the one-norm term enforces sparseness of the solution.We used the following formulation of LASSO to get our results.

$$\min_{\boldsymbol{B_i}, \boldsymbol{y_i}} \quad \left\|\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B_i}\right\|_2^2 + \lambda \left\|\boldsymbol{B_i}\right\|_1$$

As the ℓ_1 norm is not smooth everywhere, we use the penalised form of LASSO and introduce the auxiliary variables $\boldsymbol{y_i}$ to move the ℓ_1 norm from the objective to the constraints.

$$\begin{aligned} \min_{\boldsymbol{B_i}, \boldsymbol{y_i}} \quad & \left\|\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{B_i}\right\|_2^2 + \lambda \boldsymbol{1^T y} \\ \text{s.t.} \quad & \boldsymbol{y_i} \geq \boldsymbol{B_i} \\ & \boldsymbol{y_i} \geq -\boldsymbol{B_i} \end{aligned}$$

The structure of the problem ensures that, at optimality, the auxiliary variable $\boldsymbol{y_i}$ will be equal to the absolute value of the coefficients, i.e., $\boldsymbol{y_i} = |\boldsymbol{B_i}|$.

If we expand the ℓ_2 norm and convert the constraints to the standard form, we get :

$$
\begin{aligned}
\min_{\boldsymbol{B_i},\boldsymbol{y_i}} \quad & \boldsymbol{B_i}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{B_i} - 2\boldsymbol{r_i}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{B_i} + \lambda\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{y} \\
\text{s.t.} \quad & \boldsymbol{B_i} - \boldsymbol{y_i} \leq 0 \\
& -\boldsymbol{B_i} - \boldsymbol{y_i} \leq 0
\end{aligned}
$$

MATLAB's quadprog function takes the quadratic programming problems in the following format :

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\ \boldsymbol{x} + \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\
& \boldsymbol{A_{eq}}.\boldsymbol{x} = \boldsymbol{b_{eq}}
\end{aligned}
$$

So, if we compare our problem setup to the MATLAB's, we can devise that:

$$
\begin{aligned}
\boldsymbol{Q} \quad &= \boldsymbol{2X^TX} \\
\boldsymbol{c} \quad &= [-2\boldsymbol{r_i}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{B}_i; \lambda\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{y}] \\
\boldsymbol{A} &= \begin{bmatrix}
\begin{bmatrix}
1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1
\end{bmatrix}_{9\times18} \\
\begin{bmatrix}
-1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\
0 & -1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & -1 & 0 & 0 & \cdots & -1
\end{bmatrix}_{9\times18}
\end{bmatrix}_{18\times18} \\
\boldsymbol{b} &= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{18\times1} \quad .
\end{aligned}
$$

We have nine coefficients, $\boldsymbol{B_i}, (i = 1,\ ..., 9)$,to estimate for 8 factors and a intercept. And for each of these nine coefficients, we have 9 auxiliary variables $\boldsymbol{y_i}, (i = 1,\ ..., 9)$. Hence, in total we have to minimise for **18 variables**. For each factor, we have 2 constraints. Hence, in total we have **18 inequality constraints**.

The $\boldsymbol{Q}$ matrix is of the dimensions 18×18. The $\boldsymbol{c}$ vector is 18×1, where the first 9 elements correspond to factors and last 9 for auxiliary variables. After setting up the problem, we use the MATLAB function quadprog to get our results.

As part of this project, we had to select an appropriate value for $\boldsymbol{\lambda}$, penalisation factor, to be used.

Instead of choosing a common $\boldsymbol{\lambda}$ for all our models, we used a different $\boldsymbol{\lambda}$ for each stock in each model for each investment period. The process to choose the $\boldsymbol{\lambda}$ as follows-

1. Select a range of $\boldsymbol{\lambda}$s to choose from. We used $\boldsymbol{\lambda} \in [0.001, 0.002, 0.003, ..., 0.999, 1]$, with a step size of 0.001.

2. Then use each $\boldsymbol{\lambda_i}$ at each iteration and estimate the coefficients.

3. Check for the constraint, that 2 to 5 factors are non zero.

4. If the constraint satisfies, **stop** and choose these estimates. Else, continue with a new $\boldsymbol{\lambda_i}$.

Another way to choose $\boldsymbol{\lambda}$ would be to select all the models which satisfy the constraint of non zero factors and then to pick the $\boldsymbol{\lambda}$ which has the smallest mean square error among all the models. This approach, would lead to overfitting and the model won't be generalised. Hence, we used the process mentioned above. Also, we saw that each model has different number of non zero factors, because of the step size of 0.001. As we increase the step size, there is a chance that two or three factors become zero at once.

2.4 BSS Model

Best Subset Selection is the technique of choosing **K** variables out of available independent variables to minimize the RSS in a linear regression.

Let r_i denote i^{th} stock's return, F denote factor's return and $B_i = (b_i^j)_{9\times 1}$ to be the coefficient of the regression against ith stock. Similar to the **OLS**, our objective is to minimize the following:

$$\|\boldsymbol{r_i} - \boldsymbol{F}\boldsymbol{B_i}\|_2^2$$

which is equivalent to minimizing

$$-2\boldsymbol{r_i^T}\boldsymbol{F}\boldsymbol{B_i} + \boldsymbol{B_i^T}\boldsymbol{F^T}\boldsymbol{F}\boldsymbol{B_i}$$

under the constraint of

$$\|\boldsymbol{B_i}\|_0 = \mathbf{K}$$

To write this problem into the standard form, we introduce binary variables $\boldsymbol{Y_i} = (y_i^j)_{9\times 1}$, then we have **Mixed-Integer Quadratic Programming** problem :

$$
\begin{aligned}
&\min_{\boldsymbol{B_i}} \quad -2\boldsymbol{r_i^T F B_i} + \boldsymbol{B_i^T F^T F B_i} \\
&\text{s.t.} \quad -1000\boldsymbol{Y_i} \leq \boldsymbol{B_i} \leq 1000\boldsymbol{Y_i} \\
&-\boldsymbol{1^T Y_i = K} \\
&\boldsymbol{Y_i} \in \{0,1\}^9
\end{aligned}
$$

Note that the inequality constraint is to ensure that $b_i^j = 0$ when $y_i^j = 0$. And in this project, we choose K = 4, and run regression using factor returns (including constant term) against the returns of individual stocks. We used Gurobi's MATLAB API to run our BSS model.

3 Portfolio Optimization

Each of our four models return two set of parameters, $\boldsymbol{\mu}(20 \times 1)$, the expected asset returns and $\boldsymbol{Q}(20 \times 20)$, expected asset co-variance matrix, for each investment period.

The Mean Variance Optimisation problem, which takes those parameters as inputs, is shown below

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \boldsymbol{x}^T \boldsymbol{Q}\ \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{\mu}^T \boldsymbol{x} \geq R \\
& \boldsymbol{1}^T \boldsymbol{x} = 1 \\
& (x_i \geq 0, \quad i = 1,\ ...,\ n)
\end{aligned}
$$

Our objective is to minimise the variance of the portfolio by choosing the optimal set of weights(x) for each asset, subject to the constraint of **atleast** achieving the target return R, which is the geometric mean of market excess return during the corresponding calibration period. And the sum of the weights is 1. In our formulation, we are not allowing short selling, hence, the weight of each asset should be non negative.

In this particular project, we have data from $2006-01-31$ to $2016-12-31$. However, we only used data starting from $2008-01-31$. We set up our time window as 5 year and did a rolling calibration. For instance, for the first time window, we utilized data from $2008-01-01$ to $2011-12-31$ as calibration set and $2012-01-01$ to $2012-12-31$ as test set. In the calibration part, we would get $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ as input parameters for further MVO optimization. We would also got in-sample metrics, which will be discussed in the next section. Then, the $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ would be fed into the MVO and output the optimized weights for each assets in the portfolio for the year of 2012. Portfolio value and relevant out-of-sample metrics would be calculated in this time.

In total, we would have 5 rolling windows. In other words, portfolio values from 2012 to 2016 for the 4 models and equal weighted portfolio would be displayed eventually. The yearly adjusted assets weights and portfolio metrics would also be shown in the next sections.

4 In-sample Analysis

In this section, we will show the mean and standard deviation of adjusted $R2$ from multiple optimizations. After calibration, we had $20 \times 5 \times 4$ adjusted $R2$ values for the 20 assets over the 5 investment periods for the 4 factor models. So, we calculate mean and standard deviation for each year for each model having 20 observations.

From Table 3, we can see that OLS models always has the highest and the most stable adjusted R^2 among all methods. This result is expected because OLS included 8 factors, and all of them are valid in the industry. For the other three models, there are no significant differences in either mean or standard deviation data. For FF and LASSO, the adjusted R^2 in the first two years are higher than the last three years, but for BSS, there is no evident difference across time periods.

But overall, we can observe that the models are stable over time.

Table 3: R^2 of four models in each calibration

		OLS	FF	LASSO	BSS
2011	Mean	0.480	0.436	0.395	0.356
	Std.	0.211	0.190	0.222	0.226
2012	Mean	0.477	0.398	0.375	0.323
	Std.	0.230	0.230	0.252	0.251
2013	Mean	0.436	0.347	0.317	0.325
	Std.	0.213	0.220	0.245	0.238
2014	Mean	0.397	0.280	0.288	0.303
	Std.	0.207	0.232	0.250	0.243
2015	Mean	0.439	0.341	0.331	0.326
	Std.	0.182	0.205	0.190	0.194

5 Out-of-sample Analysis

We use various performance and risk measures to evaluate our portfolios in out of sample data. We calculate **portfolio returns, volatility, Sharpe ratio, maximum drawdown, skewness** and **value at risk** .

1. **Mean Portfolio Return :** Average monthly return for 60 months.
2. **Monthly Volatility:** Standard deviation of monthly returns.
3. **Sharpe Ratio :** $\frac{MonthlyRiskFreeRate - MeanPortfolioReturn}{MonthlyVolatility}$
4. **Maximum Drawdown :** It is the maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained. We used Matlab's **Financial Toolbox** to calculate it.
5. **Skewness :**It is the average cube deviation from the mean, divided by the cube of the standard deviation. We want to push the distribution to the right, into the positive skewed territory to improve the odds of seeing higher, positive return.
6. **VaR (Value at Risk) :** It is a general measure of risk defined as the predicted worst-case loss with a specific confidence level (for example, 95%) over a period of time. We calculated 90% and 95% VaR, using the **Financial Toolbox** of MATLAB.
7. **CAGR :** It is the rate of return that would be required for an investment to grow from its beginning balance to its ending balance, assuming the profits were reinvested at the end of each year of the investment's lifespan.
8. **Percentage Return :** $(V^{2016}_{portfolio} - V^{2010}_{portfolio})/1000$

Table 4 displays all the metrics for the four portfolios corresponding to the four models and the equally weighted portfolio:

Table 4: Portfolio Metrics

	OLS	**FF**	**LASSO**	**BSS**	**EW**
CAGR	8.13%	7.95%	**8.41%**	7.67%	10.60%
Percentage Change	48.76%	47.52%	**50.78%**	45.60%	66.89%
Mean Return	0.70%	0.68%	**0.72%**	0.66%	0.90%
Monthly Volatility	2.58%	2.60%	**2.57%**	2.65%	3.02%
Sharpe Ratio	0.237	0.230	**0.246**	0.217	0.270
Max Drawdown	6.85%	7.50%	**6.56%**	7.06%	11.40%
Skewness	-0.067	-0.151	-0.113	**-0.001**	-0.098
VaR at 90%	-2.61%	-2.65%	**-2.58%**	-2.74%	-2.97%
VaR at 95%	-3.55%	-3.60%	**-3.51%**	-3.70%	-4.07%

We also created an **equally weighted portfolio** as a naive benchmark which just allocates equal weight to all the stocks at the start of any investment period.

Figure 1 shows the evolution of wealth for each of the five portfolios. We can clearly observe that the four models move in a very correlated manner and have similar movements. **Table 4** supports our observation and we can also infer that the LASSO portfolio has the best metrics among the four portfolios whereas BSS portfolio is the worst.

But these sophisticated techniques are overperformed by the naive equally weighted portfolio which has better return metrics than others but is compromising with the risk metrics. Still, the Sharpe Ratio which is risk adjusted return is better for the EW portfolio with higher VaR and maximum drawdown.

Figure 2 shows the changes per period in the composition of our portfolios. **Figure 3** shows the percentage of wealth allocated to each stock by each model for each investment period. We can clearly see our portfolios are fairly concentrated in stocks with lower variance and higher expected returns.

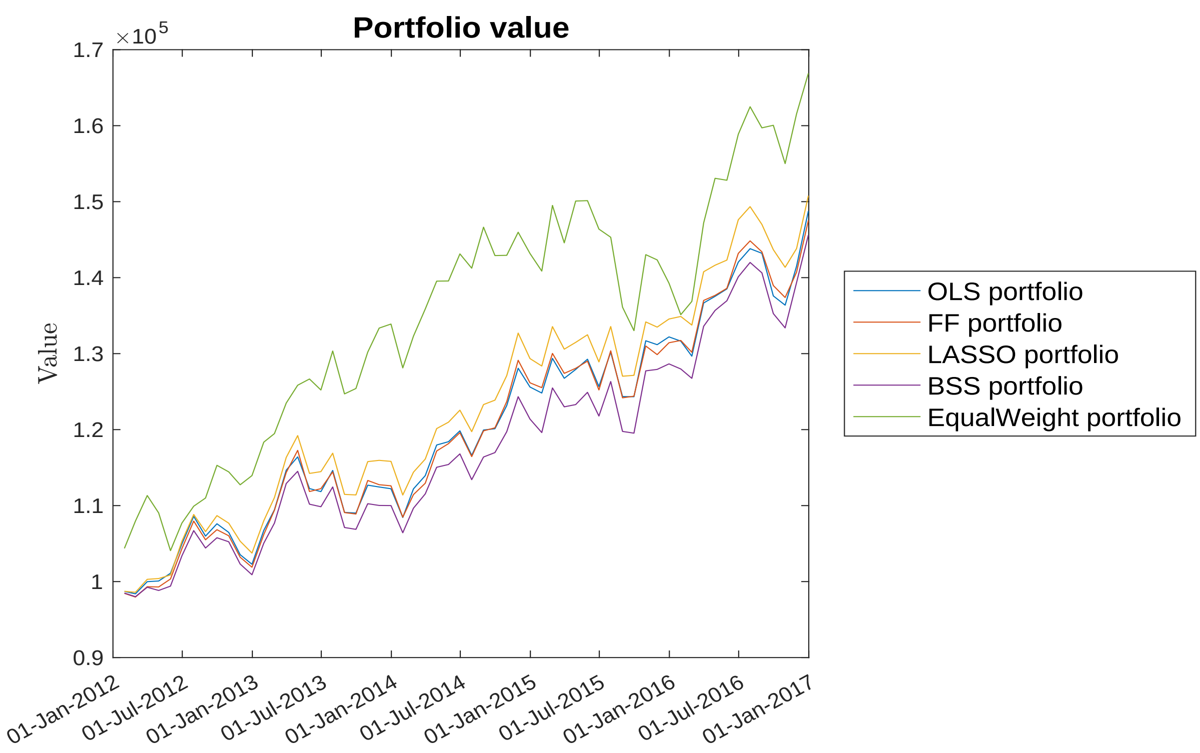

Figure 1: Portfolios Comparison

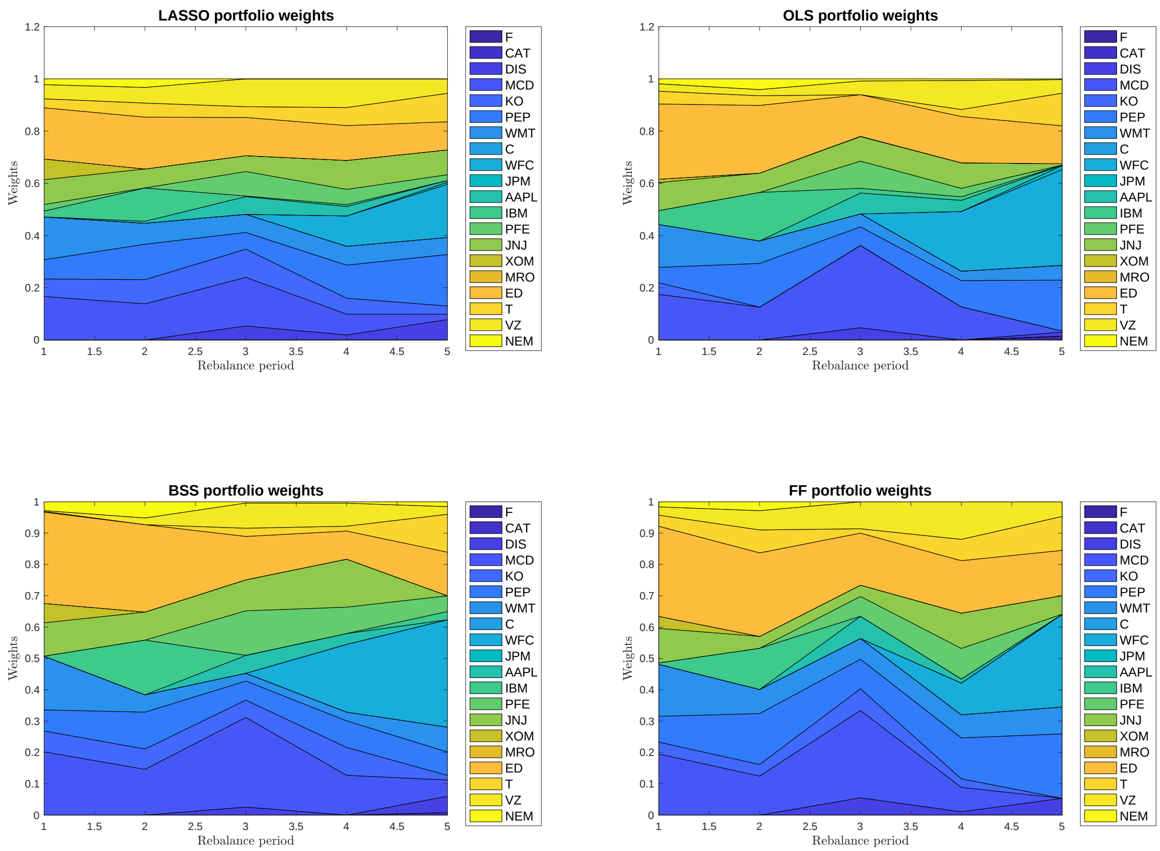

Figure 2: Portfolio Weights across Investment Periods

Ticker	Period 1				Period 2				Period 3				Period 4				Period 5			
	OLS	FF	LASSO	BSS	OLS	FF	LASSO	BSS	OLS	FF	LASSO	BSS	OLS	FF	LASSO	BSS	OLS	FF	LASSO	BSS
F	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.47%	0.00%	0.00%	0.00%
CAT	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.79%
DIS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.67%	5.53%	5.36%	2.57%	0.00%	1.06%	1.92%	0.00%	1.54%	5.35%	7.84%	5.21%
MCD	17.46%	19.50%	16.66%	20.19%	12.56%	12.47%	13.84%	14.62%	31.49%	27.88%	18.61%	28.62%	12.68%	7.80%	7.96%	12.69%	0.42%	0.00%	2.04%	5.22%
KO	4.44%	3.78%	6.73%	6.66%	0.03%	3.69%	9.26%	6.48%	0.04%	6.91%	10.79%	5.51%	0.00%	2.73%	6.11%	8.90%	0.00%	0.00%	3.12%	1.45%
PEP	5.90%	8.25%	7.34%	6.74%	16.74%	16.27%	13.58%	11.77%	7.12%	9.45%	6.40%	6.09%	10.09%	13.05%	12.66%	8.51%	19.49%	20.59%	19.75%	7.45%
WMT	16.43%	16.58%	16.40%	17.14%	8.61%	7.69%	8.04%	5.53%	4.97%	6.61%	6.97%	2.49%	3.57%	7.36%	7.23%	2.76%	5.67%	8.57%	6.51%	7.94%
C	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
WFC	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	22.87%	10.14%	11.63%	21.63%	36.67%	29.56%	20.36%	34.29%
JPM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.51%	0.00%	0.80%	0.00%
AAPL	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.74%	0.00%	8.00%	7.09%	6.78%	5.68%	4.26%	1.31%	3.61%	3.47%	0.21%	0.00%	0.68%	0.00%
IBM	5.33%	0.47%	2.25%	0.03%	18.58%	13.14%	12.77%	17.47%	1.81%	0.00%	0.22%	0.06%	1.36%	0.00%	0.71%	0.00%	0.00%	0.00%	0.00%	2.63%
PFE	0.00%	0.00%	2.52%	0.00%	0.00%	0.00%	0.02%	0.00%	10.42%	6.32%	9.37%	14.19%	3.26%	9.75%	5.84%	8.42%	0.00%	0.06%	2.20%	5.01%
JNJ	10.73%	11.01%	9.52%	10.62%	7.39%	3.77%	7.24%	8.94%	9.41%	3.60%	6.06%	9.89%	9.76%	11.26%	11.09%	15.31%	0.50%	5.92%	9.49%	0.01%
XOM	1.26%	3.82%	7.90%	6.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%
MRO	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ED	28.83%	28.77%	19.61%	29.13%	25.94%	26.68%	19.87%	27.90%	15.95%	16.60%	14.64%	13.85%	17.73%	16.74%	13.36%	8.97%	14.53%	14.44%	10.78%	13.91%
T	4.87%	3.55%	3.45%	0.27%	3.71%	7.32%	5.38%	0.00%	0.00%	1.41%	4.14%	2.62%	2.69%	6.80%	6.90%	1.56%	12.51%	10.83%	10.90%	12.10%
VZ	2.85%	2.64%	5.43%	0.25%	2.36%	6.15%	5.98%	2.13%	5.26%	8.60%	10.65%	8.03%	11.11%	12.00%	10.99%	7.34%	5.16%	4.67%	5.53%	2.44%
NEM	1.89%	1.62%	2.20%	2.79%	4.10%	2.83%	3.29%	5.16%	0.80%	0.00%	0.00%	0.40%	0.62%	0.00%	0.00%	0.44%	0.31%	0.00%	0.00%	1.55%

Figure 3: Portfolio Weights

6 A deeper dive into Best Subset Selection model results

In this section, we show the out-of-sample results with different values of **K** in the BSS model. Figure 4 shows the trend of portfolios with different **K**. We observe a significant gap between **K**=3 and others, and the strategy has the best performance in the sense of total return when **K**=5 or 6.

Table 5 showed the relevant stats. One pattern is that for **K**=3, 4 or 5, the larger **K** is, the higher the return and the Sharpe Ratio are. When **K**=5 or 6, the risk profiles of the strategy are very similar.

Figure 4: Trend of BSS model for different K

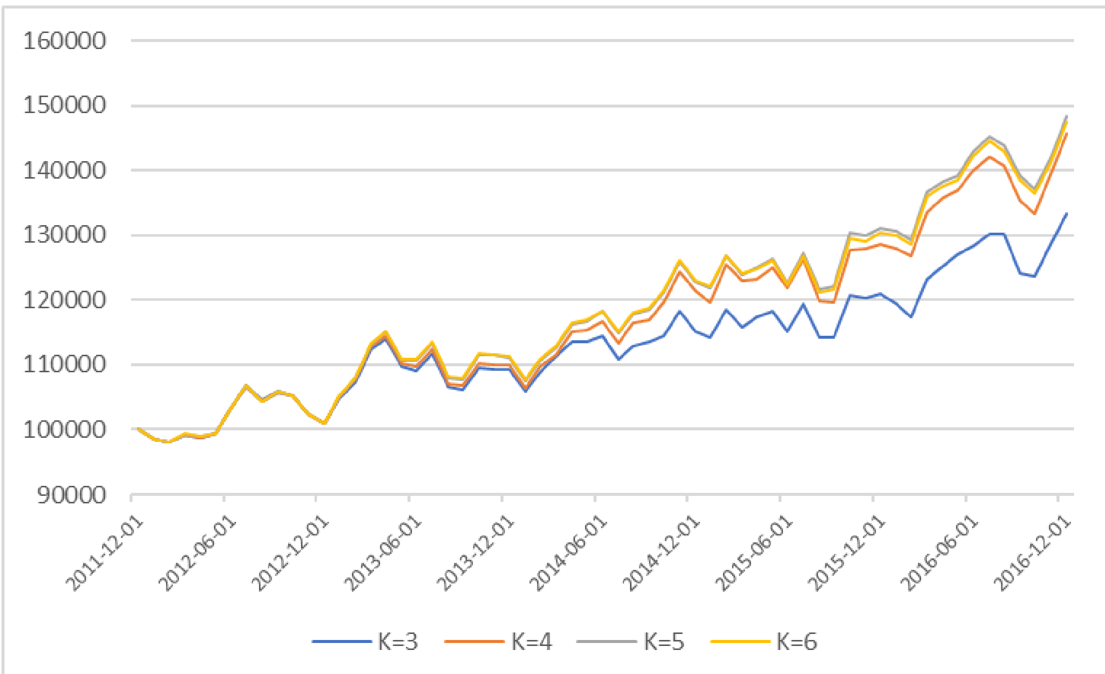

Table 5: Return-risk profile of BSS strategies for different K

	Return	Std.	Sharpe	Skewness	Max Drawdown
K=3	6.12%	8.68%	0.705	-0.109	-6.63%
K=4	7.96%	9.27%	0.858	-0.001	-6.76%
K=5	8.31%	9.07%	0.917	0.012	-6.27%
K=6	8.20%	9.02%	0.909	-0.023	-6.26%

Table 6 shows the mean and standard deviation for adjusted R^2. Similarly, when **K** = 3,4, or 5, the R^2 is increasing, but the data will not change if we further increase **K** to 6.

Table 6: R^2 of BSS model for different K

		2012	2013	2014	2015	2016
K=3	Mean	0.347	0.313	0.318	0.291	0.314
	Std.	0.228	0.253	0.239	0.247	0.198
K=4	Mean	0.356	0.323	0.325	0.303	0.326
	Std.	0.226	0.251	0.238	0.243	0.194
K=5	Mean	0.357	0.327	0.326	0.306	0.329
	Std.	0.226	0.250	0.238	0.242	0.194
K=6	Mean	0.357	0.327	0.326	0.307	0.329
	Std.	0.226	0.250	0.238	0.242	0.194

7 Conclusion and Discussion

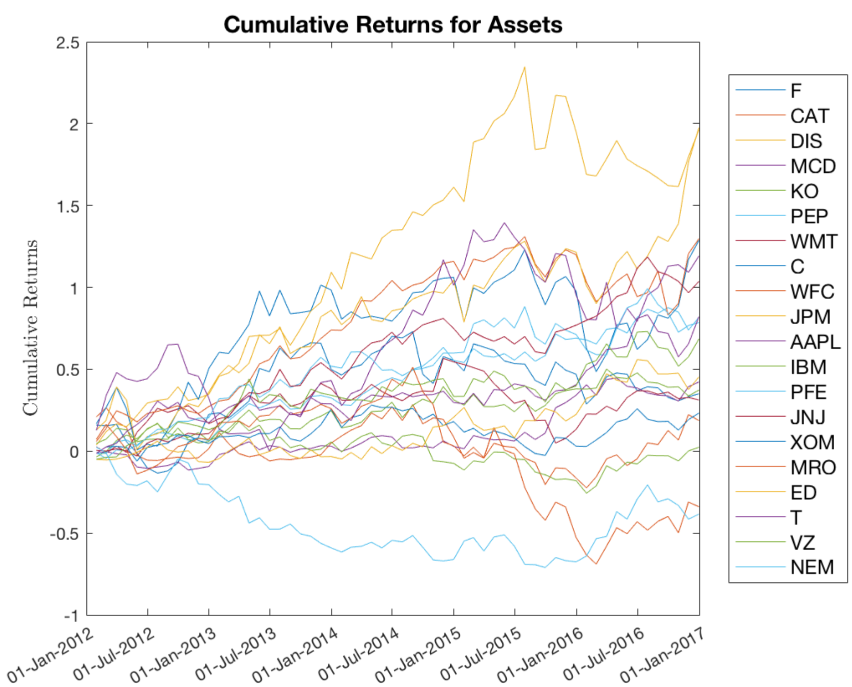

Figure 5: Cumulative Returns for Assets

To our surprise, the **equal-weighted portfolio** outperformed all other factor models in this project, in terms of return metrics. After analyzing the data, we can conclude that this might be caused by the variance-minimizing objective function in the MVO method. It is clear that from 2012 to 2017, most of the time there was a **bull market**. Thus, it should be better to hold some assets with higher variance, which also means higher expected returns, during the period. However, as our purpose was to minimize the risk, we lost the chance to hold some multibagger assets like DIS, which has the highest cumulative returns from 2012 to 2017.

That being said, the factor models did provide us a portfolio with less variance. For example, MCD and ED both had a high weights in the factor portfolios. In addition, we can see that all the factor models did not give a high weight to or did not even select NEM or MRO, which performed worst during the period. Therefore, **factor models did effectively reduce the risk** and can help to avoid some bad(risky) investments.

Among the four factor models, We can see that the LASSO model performed the best in out of sample. We can guess the reason is that our LASSO algorithm is quite dynamic and flexible. Since we tried various λ, for each asset in each investment period, until 2 to 5 factors were selected, we might be able to choose the best factors and the best *number of* factors for each assets and period. Because the returns of different assets might be explained better by different factor combination, the dynamic LASSO could better capture this point. OLS and FF had almost equivalent performance, but we can see that sometimes OLS performed better and vice versa. Again, this can support our view that during different period, different combination of factors can better explain the returns. But what confused us is the performance of BSS. While BSS is most close to LASSO, which also embraceswe a dynamic advantage, it performed worst. One plausible reason is that we fixed $K = 4$, which made BSS not as flexible as LASSO.

Table 7: Number of stocks with weight ≥ 0.1 %

	OLS	**FF**	**LASSO**	**BSS**
Period 1	11	11	**12**	10
Period 2	9	10	**11**	9
Period 3	11	11	**12**	12
Period 4	12	12	**13**	12
Period 5	13	8	**13**	13

There are some drawbacks though, MVO creates problem of over allocation and is highly dependent on our prior estimates of expected return(overestimation risk). If we have over diversification or invest a very small fraction of amount in many stock, we can incur heavy transaction costs, market impact and slippages while re-balancing our portfolios. There has to be a consideration of re-balancing costs and size of the portfolio while going for diversification and optimisation.