Sin Cheuk Kit 54418030 cksin8-c@my.cityu.edu.hk

1. Data Description

Data	Description	Time Frame
R	Return of 49 stocks	From 197401 to 199712 (288 months)
RF	Return of risk-free asset	From 197401 to 201712 (528 months)
ZT	16 characteristics of 49 stocks	From **197312** to **201711** (528 months)
FT	8 Fama-French factors	From 197401 to 201712 (528 months)
XT	10 Goyel-Welch type predictors	From 197401 to 199712 (288 months)

In this report, time index t from 1 to 528. t = 1 refers to 197401 t = 528 refers to 201712

2. Objectives

This report is purposed to create a long-only tangency portfolio based on an evaluation using hold-out sample. For long-only portfolio, the weighting of each stock must be any positive number between 0 and 1, while the sum of the weights of stocks equals to 1.

The hold-out sample refers to the predicted return of 49 stocks from 199801 to 201712 (240 months). Therefore, the result should be a matrix with 240 months * 49 stock weighting = 11760 elements.

3. Modelling of Stock Returns

On top of the returns of 49 stocks from **197401(t = 1) to 199712 (t = 288)**, the prediction of excess returns of 49 stocks from 199801 **(t = 289) to 201712 (t = 528)** should be carried out before creating a long-only tangency portfolio.

All models should be conducted by R package "caret". To avoid **overfitting** problem, **10-folded cross validation** has been used to select model.

3.1. *Prediction of Excess Returns of 49 Stocks*

The **in-sample fitting** of excess returns of 49 stocks was carried out with 2 types of cross-sectional prediction.

- Cross Sectional Prediction with **Characteristics**
- Cross Sectional Prediction with **Factors**

3.1.1.Cross Sectional Prediction with Characteristics (in-sample)

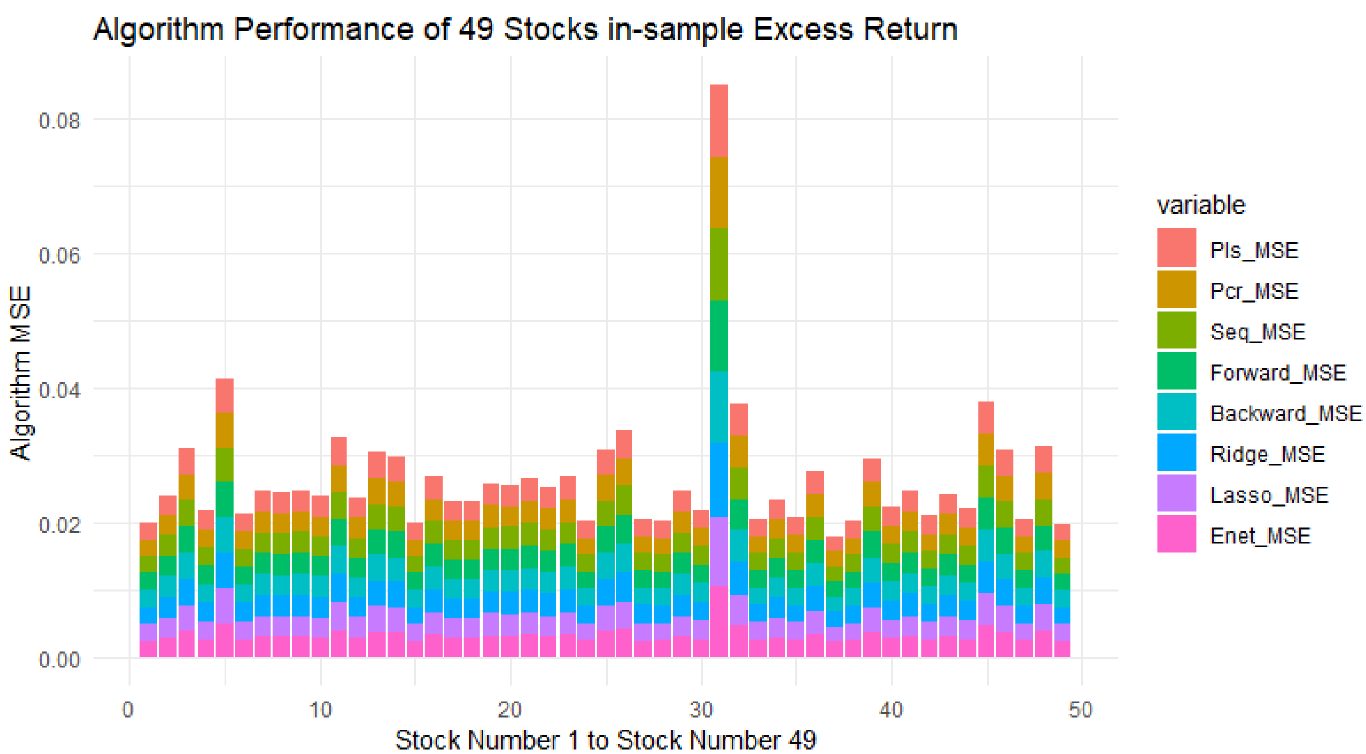
For each of the 49 stocks, 8 models were applied to fit and analyzed the **in-sample (197401 to 199712)** excess returns of stock at time **t,** using 16 characteristics at time **t-1** as predictors.

	For i = stock number 1 to stock number 49
Target variable:	Excess return of stock i at time **t**
Predictors:	16 characteristics of stock i at time **t-1**

The 8 models are shown as follows:

- Partial Linear Regression
- Principal Component Regression
- Best Subset Selection
- Forward Selection
- Backward Selection
- Lasso
- Ridge
- Elastic Net

The following graph shows MSE the in-sample forecasting excess returns of 49 stocks based on 8 models:



Detail measurement of **8 machine learning models'** performance will introduce in session 3.2

3.1.2. *Cross Sectional Prediction with Factors (in-sample)*

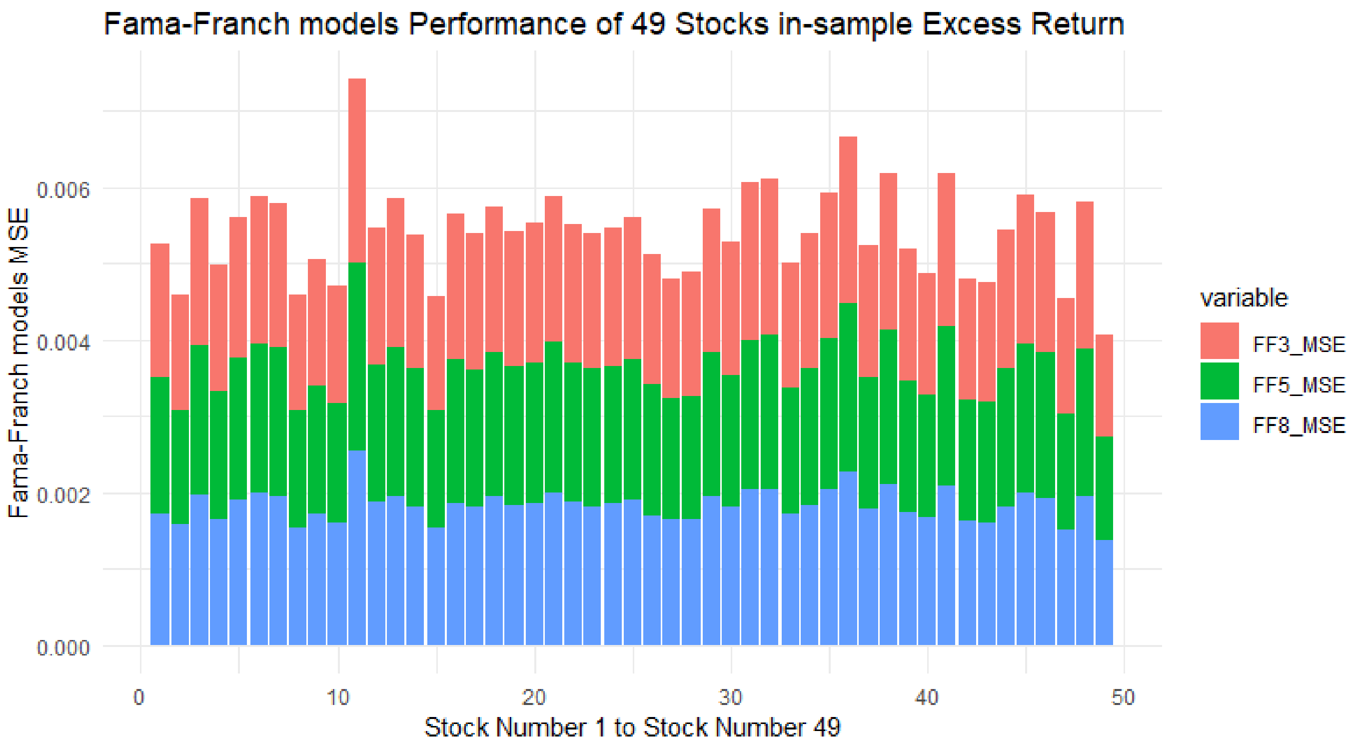
For each of the 49 stocks, **3 types of Fama-French model** were used to fit and analyze the **in-sample (197401 to 199712)** excess returns of stock based on an assessment of 8 factors.

For i = stock number 1 to stock number 49, the 3 types of Fama-French model are listed as follows:

Fama-French 3-factor model $$R_i - R_f = \alpha_i + \beta_{1,i}MKT + \beta_{2,i}SMB + \beta_{3,i}HML + \varepsilon_i$$
Fama-French 5-factor model $$R_i - R_f = \alpha_i + \beta_{1,i}MKT + \beta_{2,i}SMB + \beta_{3,i}HML + \beta_{4,i}RMW + \beta_{5,i}\text{CMA} + \varepsilon_i$$
Fama-French 8-factor model $$R_i - R_f = \alpha_i + \beta_{1,i}MKT + \beta_{2,i}SMB + \beta_{3,i}HML + \beta_{4,i}RMW + \beta_{5,i}\text{CMA} + +\beta_{6,i}MOM + \beta_{7,i}STR + \beta_{8,i}LTR + \varepsilon_i$$

Noted that **alpha** is set to be **0**, the forecasting excess returns of each stock will not be affected by draft term and can only be explained by Fama-French factor(s).

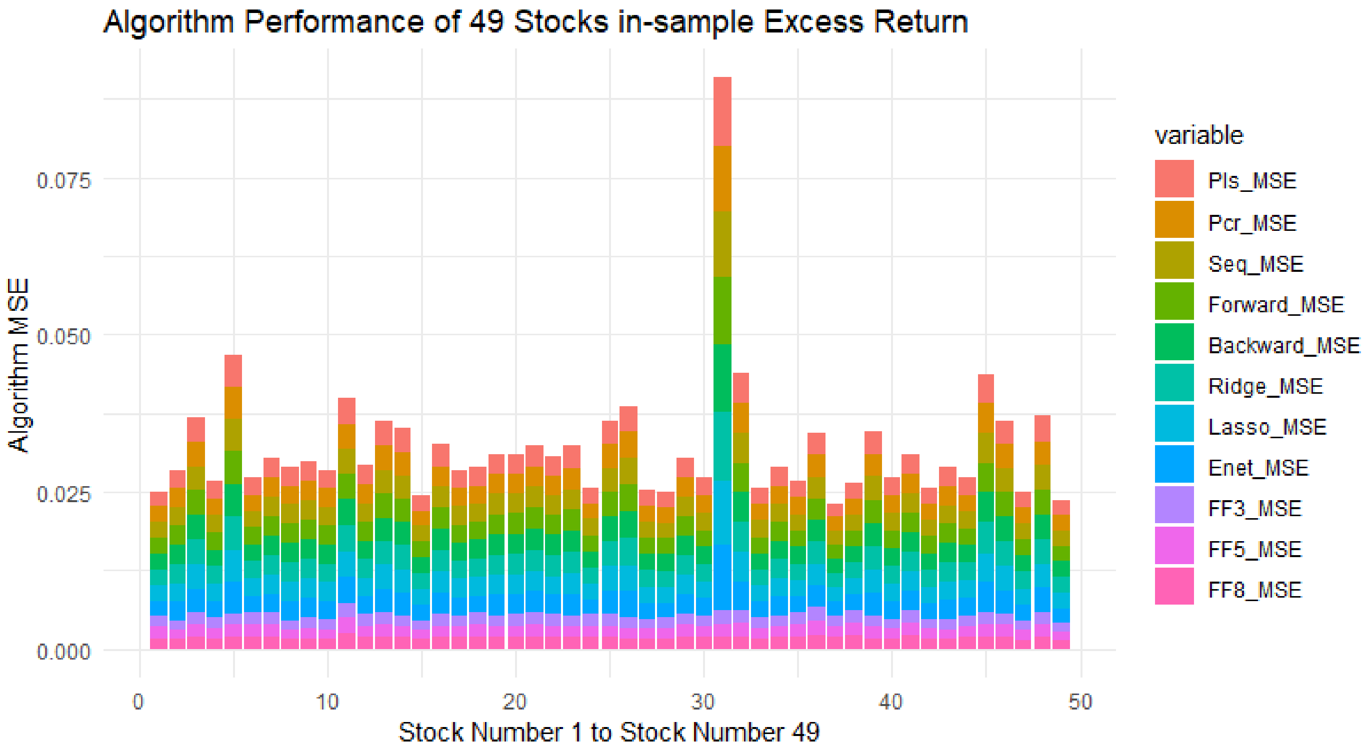
The following graph shows the in-sample forecasting excess returns of 49 stocks based on 3 types of Fama-French model MSE:



Detail measurement of **3 types of Fama-French model** performance will introduce in session 3.2

3.2. *Model Selection of Predicted Excess Returns of 49 Stocks (in-sample)*

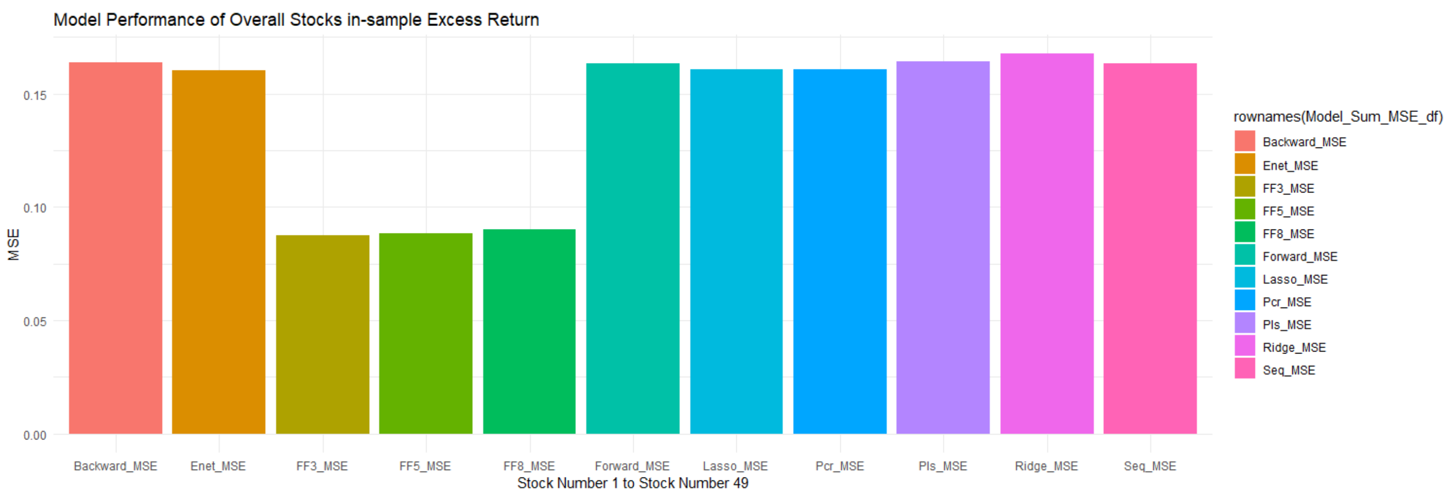
The following graph shows 8 machine learning method MSE and 3 types of Fama-French model MSE:



It was **not possible** to draw a conclusion from the results and figure out which model had the best performance. To illustrate, some of the stocks had the best performances using PLS model, while other stocks might have better performances with Fama-French 3-factor model.

Therefore, it was rational to evaluate the models using the **overall performance** of 49 stocks, instead of the individual performance of each stock.

$$\text{MSE of model m} = \sum_{i=1}^{49} MSE\ of\ stock\ i\ in\ model\ m$$

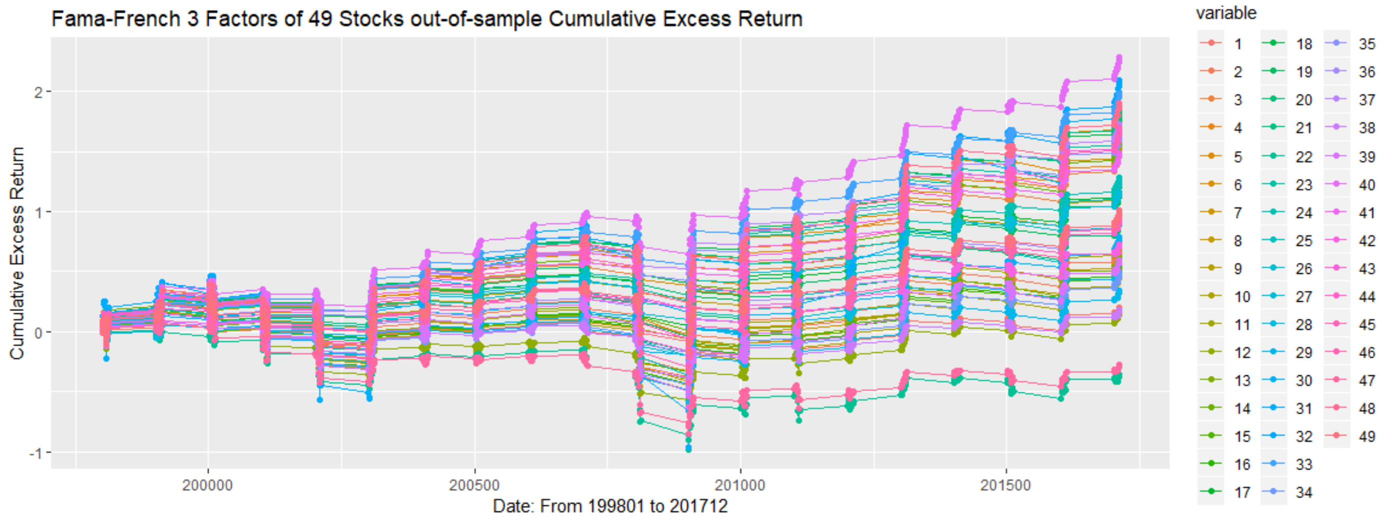


Based on the overall performance of 49 stocks derived from different models, **Fama-French 3-factor model** resulted in the lowest MSE (0.08768134). Hence, **Fama-French 3-factor model** had the best in-sample fitting performance on excess returns of stocks from 197401 to 199712.

3.3. Out-of-Sample Prediction of Excess Returns of 49 Stocks by Fama-French 3-Factor Model

An out-of-sample prediction of **excess returns** of 49 stocks from 199801 to 201712 (240 months) was conducted using Fama-French 3-factor model.

The following graph shows the **predicted excess returns** with Fama-French 3-factor model applied:

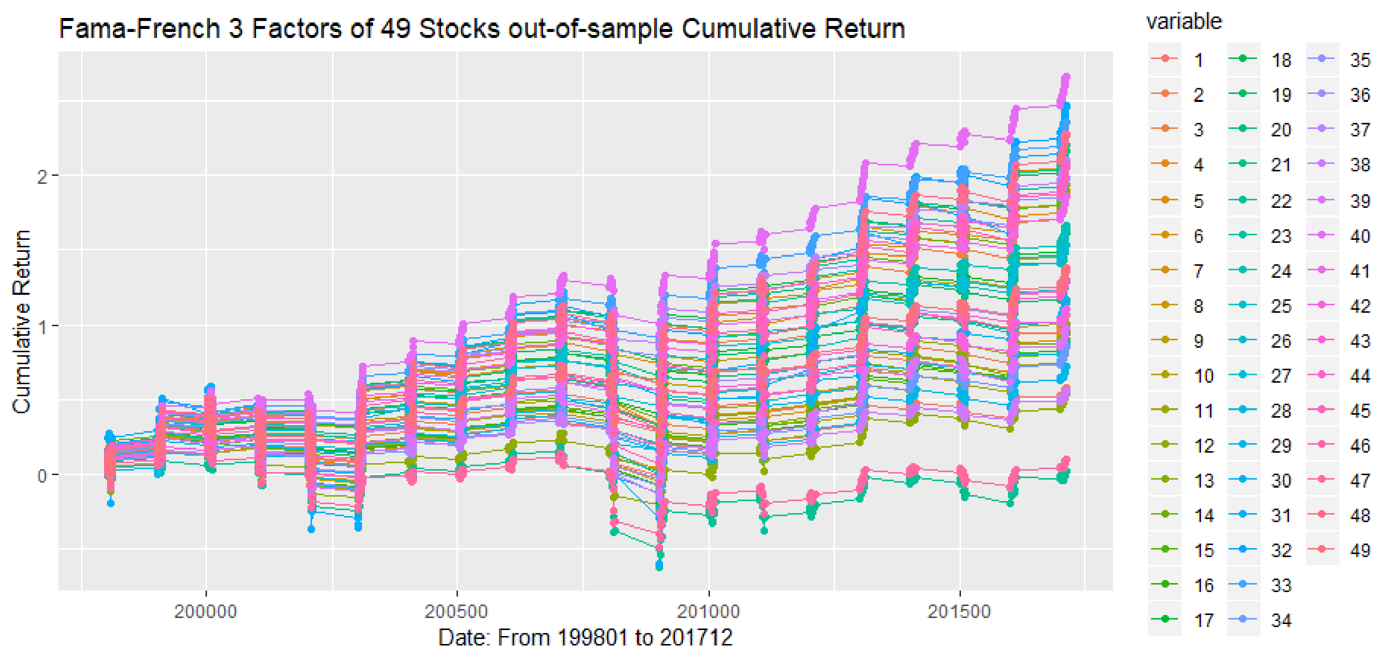


3.4. Out-of-Sample Prediction of 49 Stocks Return by Fama-French 3-Factor Model

In addition, the **predicted return** of 49 stocks from 199801 to 201712 (240 months) was computed using the following equation.

$$\hat{Return}_{i,t} = \hat{ExcessReturn}_{i,t} + Riskfree_t$$

The following graph shows the **predicted return** with Fama-French 3-factor model applied:



4. Tangency Portfolio

4.1. *Optimal Method*

Marto Carlo simulation method was applied to compute the tangency portfolio. In order to improve accuracy, given that the number of simulation is **1 million times, 1,000,000 scenario portfolios** with random stocks weighting have been conducted.

4.2. Portfolio Constraints

As only **positive stock weighting** should be involved in the **long-position** portfolio, stock weighting of each scenario portfolio was generated by any **normalized random number between 0 and 1**, with a **sum of stock weighting that equaled to 1** as a full-investment.

4.3. Useful formula

r_p Expected return of portfolio: $[\mathrm{W}]^T \times [R]$	[W] refers to a 49*1 stock weighting vector $[R]$ refers to 49*1 vector of the predicted excepted stocks return
σ_pStandard deviation of portfolio: $[\mathrm{W}]^T \times [Cov] \times [\mathrm{W}]$	$[Cov]$ refers to 49*49 variance-covariance matrix

4.4. Selection Criteria

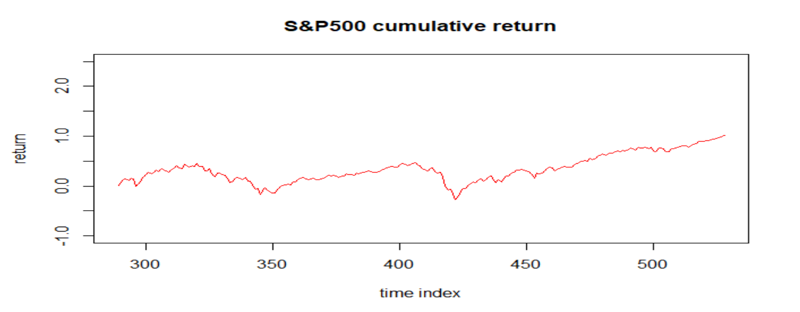
The scenario portfolios with the best performance in the following **3 criteria** was selected.

- **<u>Sharpe Ratio</u>**: $\frac{r_p}{\sigma_p}$

  r_p refers to the expected portfolio return, σ_p refers to the standard deviation of the portfolio

- **<u>Information Ratio over S&P 500</u>**: $\frac{r_p - r_{SP500}}{\sigma_p}$

  r_{SP500} refers to the expected return of S&P 500

Use R package "quantmod" to download daily data of S&P 500 from 199801 to 201712 (240 months) Then compute S&P 500 **monthly** return from 199801 to 201712 (240 months)	

- **<u>Information Ratio over equal weighting portfolio (EW)</u>**: $\frac{r_p - r_{EW}}{\sigma_p}$

  r_{EW} refers to the expected return of equal weighting portfolio

$$r_{EW} = \sum_{i=1}^{49} \frac{1}{49} r_i$$

For equal weighting portfolio, the weighting of each stock = $\frac{1}{49}$ = 0.02040816

4.5. *Fixed Model (Unconditional Mean-Variance Portfolio)*

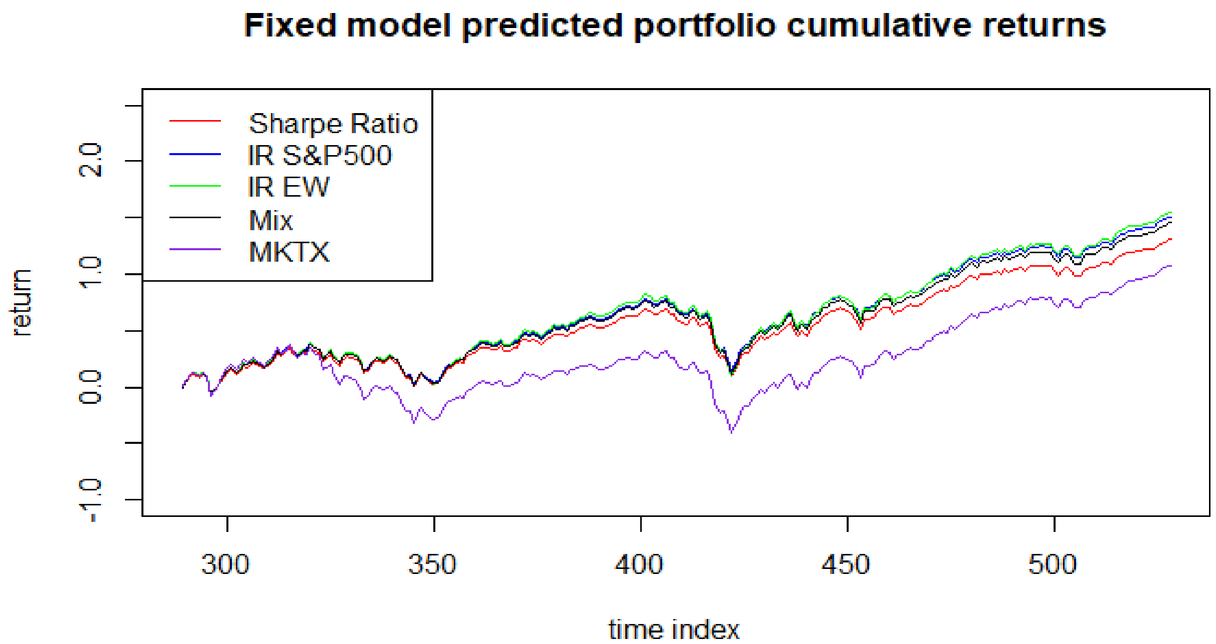
In fixed model, the tangency portfolio was computed using **unconditional mean vector** (length: 49*1) and **unconditional variance-covariance matrix** (length: 49*49) from 199801 to 201712 (240 months). As such, **weighting** of 49 stocks, **expected return** and **variance-covariance matrix** were fixed **(not time varying).**

The graph shows the cumulative portfolio returns with the highest

1. Sharpe Ratio
2. Information Ratio over S&P 500
3. Information Ratio over equal weighting portfolio

In addition, the performances of the portfolios have been compared using market excess return (MKTX).

From 199801 to 201712



To maximize Sharpe Ratio, Information Ratio over S&P 500 and Information Ratio over equal weighting portfolio were consider at same time, the "**Mix**" portfolio was created.

The weighting of the "**Mix**" portfolio has been obtained from the **weighted average** of the following 3 portfolios

1. Sharpe Ratio
2. Information Ratio over S&P 500
3. Information Ratio over equal weighting portfolio

To evaluate the overall performance of the portfolio generated from the fixed model, sharpe ratio of the "Mix" portfolio was computed.

Annual Sharpe Ratio of the "Mix" portfolio: $\frac{R_{mix}}{\sigma_{mix}} \times \sqrt{12} = 0.5887969$

4.6. *Dynamic Model (Conditional Mean-Variance Portfolio)*

Due to the fact that the excepted return and volatility of stocks may **change over time,** the tangency portfolio should be created with **dynamic model** instead of fixed model.

In dynamic model, the tangency portfolio was computed by **conditional mean vector** (length: 49*1) and **conditional variance-covariance matrix** (length: 49*49), in which the **weighting** of 49 stocks, **expected return** and **variance-covariance matrix** were dynamic **(time varying).**

Dynamic model was applied in **6 versions with different rolling period** to develop the tangency portfolio.

Rolling Period	**At Time t, Conditional Mean & Conditional Variance Based on:**
Quarterly (3 months)	time t-3 to time t
Semi-Annual (6 months)	time t-6 to time t
Annual (12 months)	time t-12 to time t
3 Years (36 months)	time t-36 to time t
5 Years (60 months)	time t-60 to time t
10 Years (120 months)	time t-120 to time t

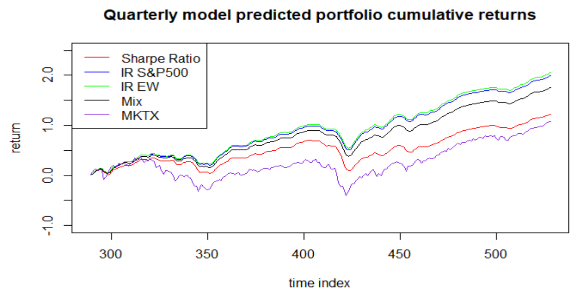
Take quarterly rolling period as example, **conditional mean** and **conditional variance at 199801 (t=289)** were calculated based on the period from 199710 (t=286) to 199801 (t=289).

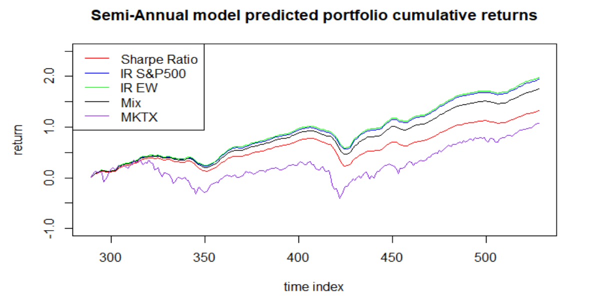
The following graphs show the cumulative portfolio returns, comparing among 6 different rolling periods, with the highest

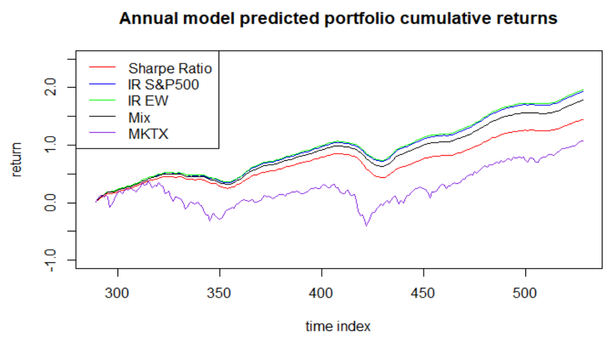
1. Sharpe Ratio
2. Information Ratio over S&P 500
3. Information Ratio over equal weighting portfolio

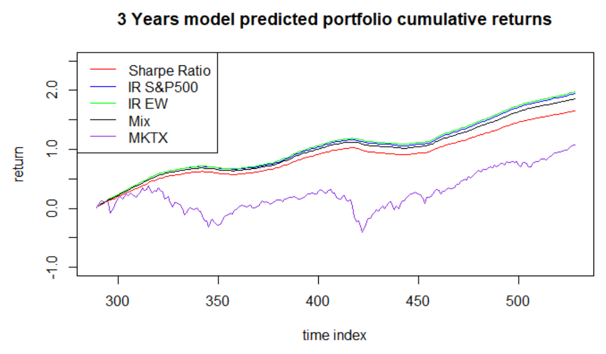
Noted that the weighting of the "**Mix**" portfolio is obtained from the weighted average of the above 3 portfolios, portfolio returns can be compared using the data of market excess returns (MKTX).

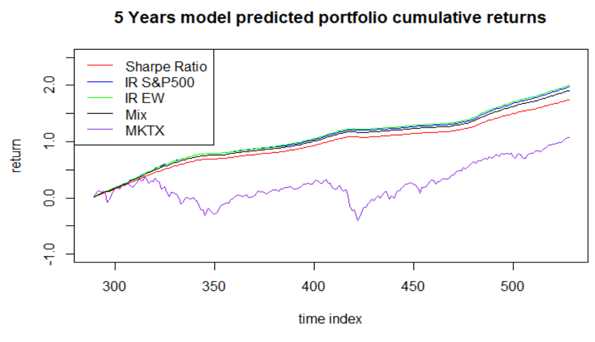
Cumulative Return of Portfolio from 199801 to 201712

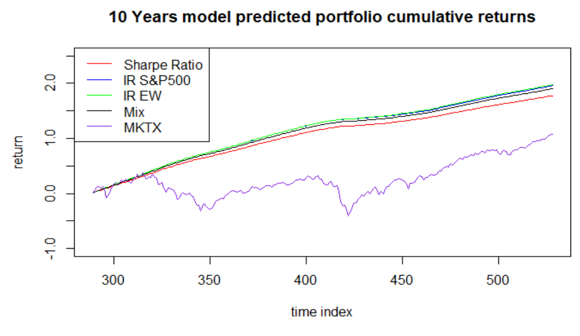












Analyzed from the above graphs, **the wider the rolling period**, the smoother the **curve** of the cumulative portfolio returns.

Rolling Period	Annual Sharpe Ratio of "mix" portfolio: $\frac{R_{mix}}{\sigma_{mix}} \times \sqrt{12}$
Quarterly (3 months)	1.75992
Semi-Annual (6 months)	2.26082
Annual (12 months)	3.11145
3 Years (36 months)	5.78136
5 Years (60 months)	8.69345
10 Years (120 months)	**11.92745 (the highest)**

It can also be observed that the model accuracy increased with the range of historical stock return data, suggesting that the **wider the rolling period**, the larger the annual Sharpe Ratio of the "mix" portfolio.

5. Conclusion

Fama-French 3-factor model had the best performance in in-sample fitting of excess returns of all 49 stocks. This model was also applied to conduct the out-of-sample prediction of excess returns of 49 stocks. The out-of-sample predicted returns of stocks were then calculated by adding excess returns of stocks and risk-free rate together. Further, it is found that the tangency portfolio, **developed using dynamic model with a rolling period of 10 years**, had the highest annual Sharpe ratio.

The cumulative returns of the **selected tangency portfolio** are shown below: