Equity Portfolio Optimization Methods in R

Author: Will Noone
DATS6103 Data Mining
The George Washington University

Abstract
The investment decision process can be viewed as a top-down process: capital allocation between a portfolio of risky assets and risk free assets and then asset allocation across broad asset classes or security selection of individual assets within each asset class. Optimal capital allocation is determined by maximizing the risk-return tradeoff given a set of constraints. Through an iterative introduction of assets into a portfolio, we seek to demonstrate the benefits of diversification in maximizing an efficient risk and return tradeoff We will briefly examine the process of efficient diversification from the ground up, starting with a menu of (3) risky assets, then incorporating assets in sets of (3) and examine the results. Effectively you can reduce your risk exposure while maintaining an expected overall portfolio return through governance of portfolio weights.

Introduction
The notion of diversification is often adage “Don’t put all your eggs in one basket”. Why? If one basket were to tip, some eggs would remain safely in another basket. To extend the metaphor, what is the best allocation of various types of eggs among a set of baskets of varying size/stability? Is there an algorithm to define the “best” allocation? The father of modern portfolio theory is economist Harry Markowitz, whose insights led to the development of his seminal theory of portfolio allocation under uncertainty, published in 1952 by the Journal of Finance. Markowitz applied mathematics to the analysis of the stock market as the topic for his PhD dissertation at the University of Chicago. While researching the then dated understanding of stock prices, Markowitz realized that the theory lacked an analysis of the impact of risk. Maximizing the risk-return tradeoff is the tenet of portfolio optimization.

Requisite Statistics and Background
<u>Portfolio Interactions</u>
Covariance is the expected product of the deviation of each return from its mean.

$$\mathrm{Cov}(ri, rj) = \frac{1}{T-1}\sum_{t=1}^{T}(ri,t\ -\ Eri)(rj,t\ -\ Erj)$$

Correlation is the covariance of the returns divided by the product of the standard deviation of each return.

$$\mathrm{Corr(ri, rj)} = \frac{\mathrm{Cov(ri, rj)}}{\mathrm{SD(ri)} * \mathrm{SD(rj)}}$$

<u>Aggregate Portfolio</u>
Expected return is the weighted-average of the expected returns

$$\mathrm{E(Rp)} = \mathrm{w1r1}\ +\ \mathrm{w2r2} + \cdots +\ \mathrm{wnrn} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{wiE(ri)}$$

Variance is the squared weighted variances incorporating covariance of all individual security pairs.

$$\mathrm{Var(Rp)} = \mathrm{w1^2Var(r1)} + \mathrm{w2^2Var(r2)} + \mathrm{2w1w2\ Cov(r1, r2)}$$

A Brief Statistical Argument for Diversification

Imagine you were offered (2) investment opportunities. Option 1 is (10) shares of the same stock with an expected return of 10%. Effectively you would have (10) perfectly correlated investments. All eggs are clearly in one basket. Option 2 is (10) shares in different stocks that are uncorrelated yet also promises an expected return of 10%. We assume that the standard deviation of all stocks to be equal between both scenarios. One can see that correlation among components of a portfolio inflates overall portfolio variance through the example below. Even though both options have equal potential return, option 2 achieves this by greatly reducing uncertainty.

$10 → Option 1: 10 shares all in stock A, 10% E(r)

$10 → Option 2: 10 shares 1 each in uncorrelated stocks {B,…,L}, 10% E(r)

where $\sigma_A = \sigma_B = \sigma_C = \sigma_D = \sigma_E = \sigma_F = \sigma_G = \sigma_H = \sigma_I = \sigma_J = \sigma_K = \sigma_L$

Because:

$Var(aX) = a^2 Var(X)$

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y)$

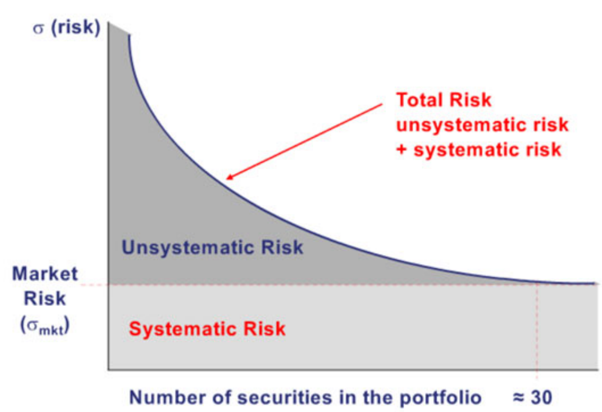
Variance Option 1 = $Var(10A) = 10^2\sigma^2 = 100\sigma^2$

Variance Option 2 = $Var(B+\ldots+L) = (1^2)Var(A) + \ldots + (1^2)Var(L) + 0 = 10\sigma^2$

The variance of Option 2 → variance Option 1 as correlated portfolio components are incorporated.

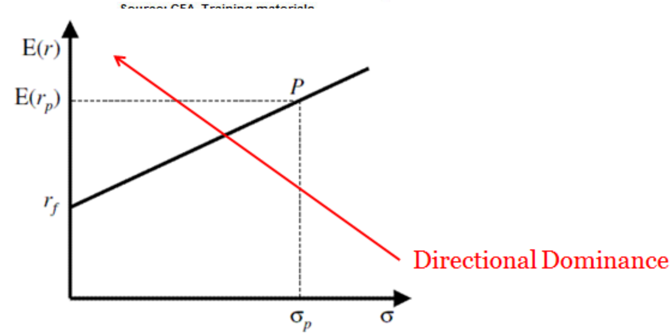
Theoretical Framework

When considering a stock, risk can be broken out into (2) parts. First are macro conditions from the general economy (interest rates, inflation, business cycle). While these conditions can be forecasted, there is considerable uncertainty to their prediction and all affect the expected return of the portfolio. This is known as systematic risk and would be common to all investment options in our universe. Second are the micro conditions (firm specific) supporting the value of the stock such as R&D, products, debt structure, etc. This is known as nonsystematic or diversifiable risk. Empirical research has shown that nonsystematic risk can be eliminated as n = 30 stocks are incorporated into a portfolio.



Risk-Return Tradeoff

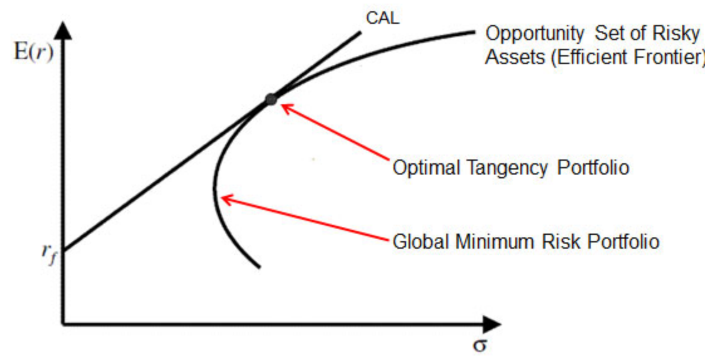
In return-variance space, Portfolios oriented left/up will dominate with respect to another located to the right/low. The Sharpe Ratio is the slope of a line through a given portfolio located in return-variance space. To earn the highest possible expected return for any level of volatility, you must find the



portfolio that generates the steepest possible line when combined with the risk-free investment.

$$\text{Sharpe Ratio(slope of CML)} = \frac{\text{Portfolio Excess Return}}{\text{Portfolio Volatility}} = \frac{E(R_p) - r_f}{SD(R_p)}$$

The line is known as the Capital Allocation Line. Where it touches the efficient frontier corresponds to the optimal risky portfolio. This portfolio is also called the mean–variance tangency portfolio. The region outside of the feasible set is unachievable by holding risky assets alone. No portfolios can be constructed corresponding to the points in this region. Points below the frontier are inefficient combinations of weighted portfolio assets. A rational investor will hold a portfolio only on the frontier.



<u>Optimal Tangency Solution</u>
To specify the parameters for the tangency portfolio, we derive a covariance matrix of the component assets. Then by constructing the following system with co-variances as coefficients, the solution yields our optimal weights.

Covariance Matrix	A	B	C
A	σ^2_A	$\sigma_{A,B}$	$\sigma_{A,C}$
B	$\sigma_{B,A}$	σ^2_B	$\sigma_{B,C}$
C	$\sigma_{C,A}$	$\sigma_{C,B}$	σ^2_C

$$(\sigma^2_A)z_1 + (\sigma_{A,B})z_2 + (\sigma_{A,C})z_3 = E(r_1) - r_f$$
$$(\sigma_{B,A})z_1 + (\sigma^2_B)z_2 + (\sigma_{B,C})z_3 = E(r_2) - r_f$$
$$(\sigma_{C,A})z_1 + (\sigma_{C,B})z_2 + (\sigma^2_C)z_3 = E(r_3) - r_f$$

$$w_1 = z_1/(z_1 + z_2 + z_3)$$
$$w_2 = z_2/(z_1 + z_2 + z_3)$$
$$w_3 = z_3/(z_1 + z_2 + z_3)$$

Data Management
We assume the following analysis to be constrained to a short term horizon for which the skewness that characterizes long term horizons is absent. The assumption of normality is sufficient to describe the returns in our data sets. We will be concerned with portfolio means, standard deviation, and correlation among portfolio components. We also assume that the weights of the portfolio are constrained by zero as a lower bound, meaning that short selling (betting a stock will go down) is not introduced; only long positions. Implementing a solution in R requires gathering price data, setting the data as a time series object, converting to periodic returns, and merging TS vectors in a portfolio. Packages like {quantmod} provide exceling tools to collect data. We created a download function {yahooDownload} which draws data from a .csv file from URL also composed by the function. Once return vectors are assembled we can perform portfolio optimization.

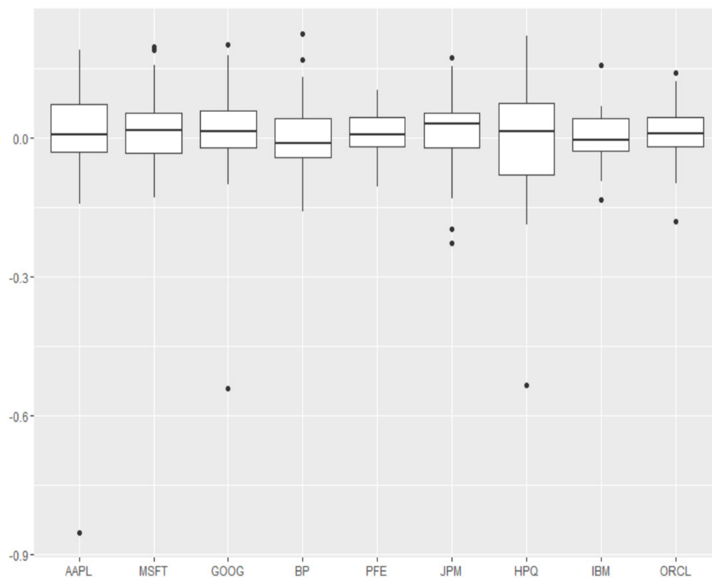
Portfolio Management
Our analysis could be divided into three parts: 1. Data exploratory analysis is conducted to nine stocks in portfolio A by checking means, variances and correlations of their returns. 2. Still considering only portfolio A, four methods – Feasible Portfolio, Minimum Risk Efficient Portfolio, Global Minimum Risk Portfolio and Tangency Portfolio, are applied to optimize it. We want to see differences in performance between those methods. 3. Diversification effect is discussed. Efficient frontiers of portfolio A, B and C are compared.

Data Exploratory Analysis

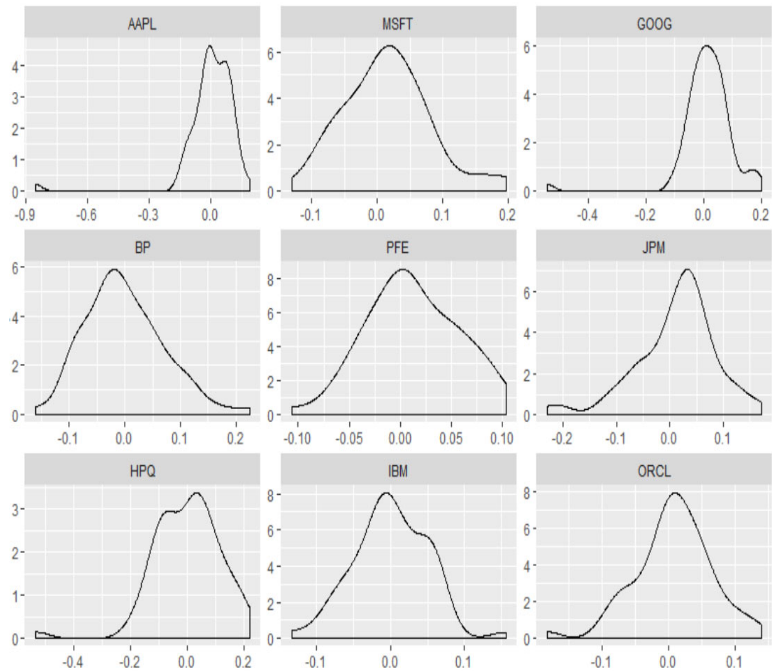
In this part, mean and variance of each stock in portfolio A is calculated and visualized by boxplot and density plot.

```
> colMeans
   AAPL    MSFT    GOOG      BP     PFE     JPM     HPQ     IBM    ORCL
-0.0013  0.0151  0.0121  0.0003  0.0123  0.0120 -0.0018  0.0001  0.0081
> colVars
  AAPL   MSFT   GOOG     BP    PFE    JPM    HPQ    IBM   ORCL
0.0183 0.0045 0.0094 0.0053 0.0021 0.0058 0.0147 0.0025 0.0034
```

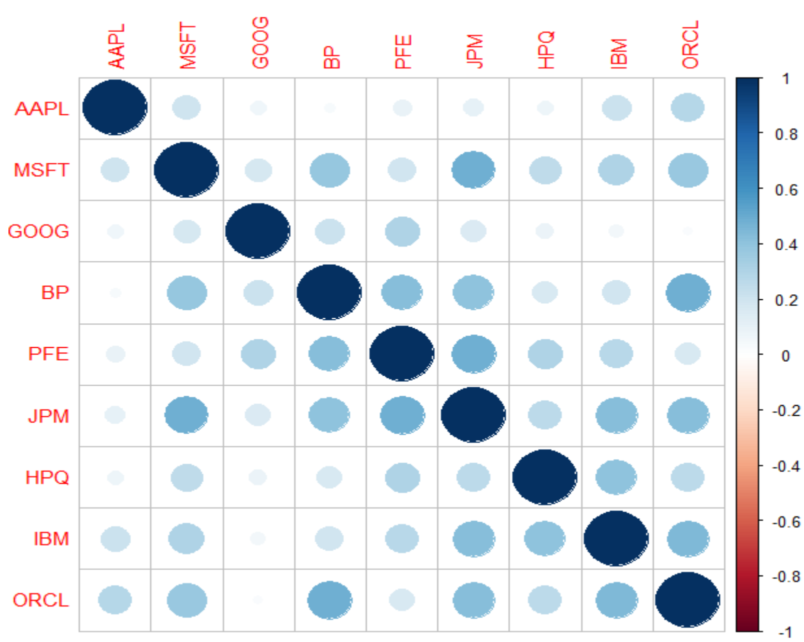
From the above numbers, we can tell that MSFT, GOOG, PFE and JPM are the top 4 most profitable stocks.



Boxplot of The Nine Stocks' Returns



From the above plots, AAPL and HPQ have variances that are significantly larger than others'. Next, correlation matrix of those stocks is shown by a correlation plot:



MSFT, as the most profitable stock, has relatively high correlation with BP, PFE, IBM and ORCL. Another profitable one as JPM, has high correlation with MSFT, BP, PFE, IBM and PFE, most parts of the portfolio.
One more stock that should be paid attention to is BP, highly correlated with 4 of the others.

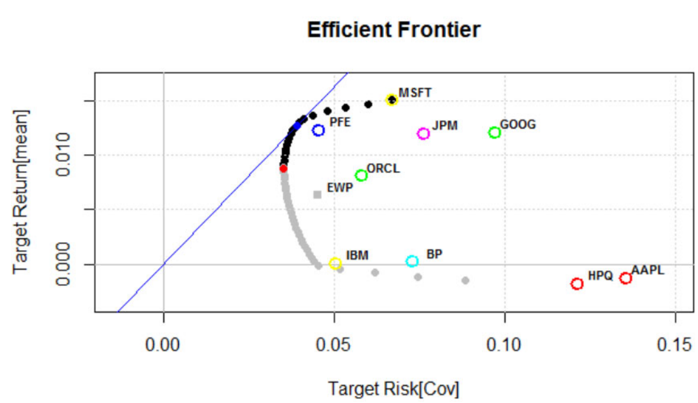
Portfolio Optimization for Portfolio A

Four methods – Feasible Portfolio, Minimum Risk Efficient Portfolio, Global Minimum Risk Portfolio and Tangency Portfolio, are considered for optimization, that is, to lower portfolio risk.
Feasible Portfolio - any combination of assets into a particular portfolio. For example, the equal weights portfolio is such a portfolio.
Minimum Risk Efficient Portfolio - portfolio with the lowest risk for a given target return.
Global Minimum Variance Portfolio – the efficient portfolio with the lowest possible risk.
Tangency Portfolio - calculated by minimizing the Sharpe Ratio for a given risk-free rate.

Taking tangency portfolio as an example, R outputs a result table like this:

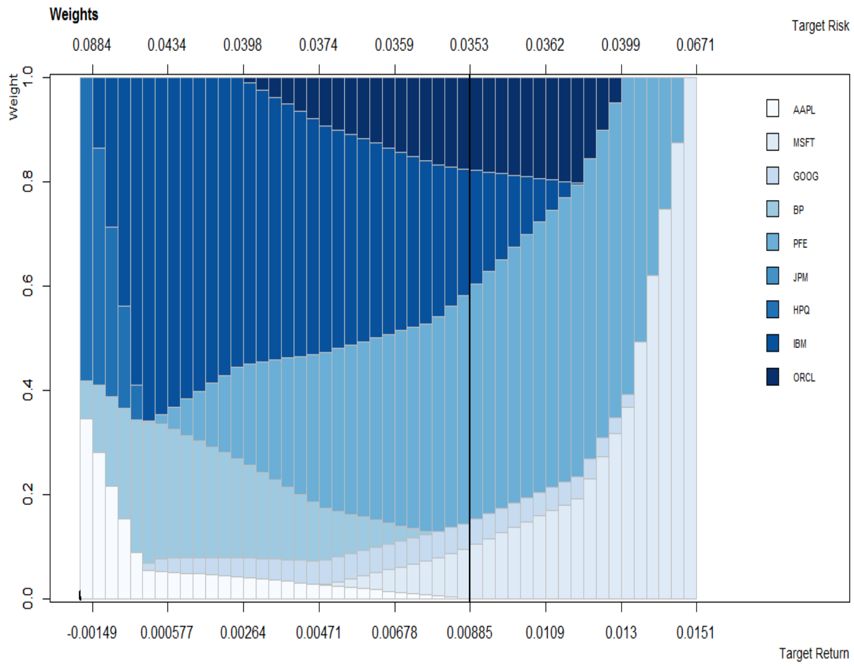
```
 MV Tangency Portfolio
 Estimator:         covEstimator
 Solver:            solveRquadprog
 Optimize:          minRisk
 Constraints:       LongOnly

Portfolio Weights:
  AAPL   MSFT   GOOG     BP    PFE    JPM    HPQ    IBM   ORCL
0.0000 0.2852 0.0334 0.0000 0.5939 0.0000 0.0000 0.0000 0.0875

Covariance Risk Budgets:
  AAPL   MSFT   GOOG     BP    PFE    JPM    HPQ    IBM   ORCL
0.0000 0.3372 0.0318 0.0000 0.5752 0.0000 0.0000 0.0000 0.0559

Target Returns and Risks:
  mean    Cov   CVaR    VaR
0.0127 0.0390 0.0701 0.0585
```

Two sets of values inside red boxes are most concerned – Portfolio Weights and Mean and Variance for Target Returns. In this portfolio, MSFT, GOOG, PFE and ORCL are assigned with a nonzero weight. PFE as a stock with the lowest risk, is given the largest weight. PFE is followed by ORCL, which has the second lowest risk. MSFT and GOOG as profitable ones are added into the optimized portfolio. The stocks highly correlated with those 4 selected stocks are removed, especially JPM even with a large return. The mean and variance of target returns are compared with performance of other methods.



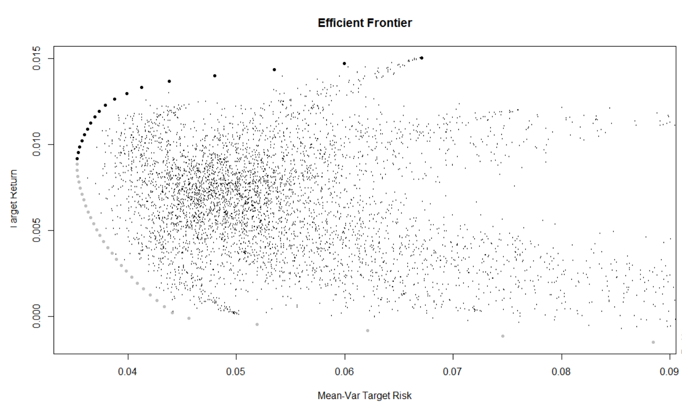
Efficient Frontier of Tangency Portfolio A

In the above plot, the black dots shape the efficient frontier curve. The blue line is the tangency line, which intersects the curve at tangency point that represents tangency portfolio. The colorful points with stock labels show the location of each stock in the return*risk plot. The ones closer to the efficient frontier are more likely to be selected.



Performance Plot of Portfolio A

The above performance plot is a more numerical version of the return-risk curve. It shows how weights' change affects portfolio risk and return. The black solid line is the global minimum risk, and the efficient frontier is right to the black line.



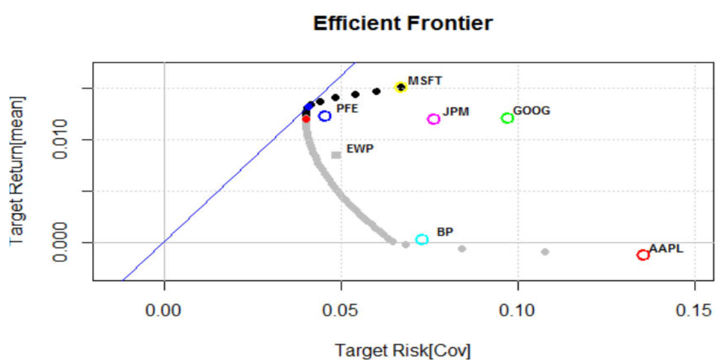
Monte Carlo Simulation of Portfolio A

Each dot in the above plot is a sample portfolio of A. There are totally 1000 thousand simulations. All points lie on the right side of the efficient frontier.

Diversification Effect on Portfolio Performance

The below is tangency result of portfolio B (6 stocks):

```
Title:
 MV Tangency Portfolio
 Estimator:         covEstimator
 Solver:            solveRquadprog
 Optimize:          minRisk
 Constraints:       LongOnly

Portfolio Weights:
  AAPL   MSFT   GOOG     BP    PFE    JPM
0.0000 0.3300 0.0314 0.0000 0.6386 0.0000

Covariance Risk Budgets:
  AAPL   MSFT   GOOG     BP    PFE    JPM
0.0000 0.3757 0.0287 0.0000 0.5956 0.0000

Target Returns and Risks:
  mean    Cov   CVaR    VaR
0.0132 0.0408 0.0724 0.0591
```

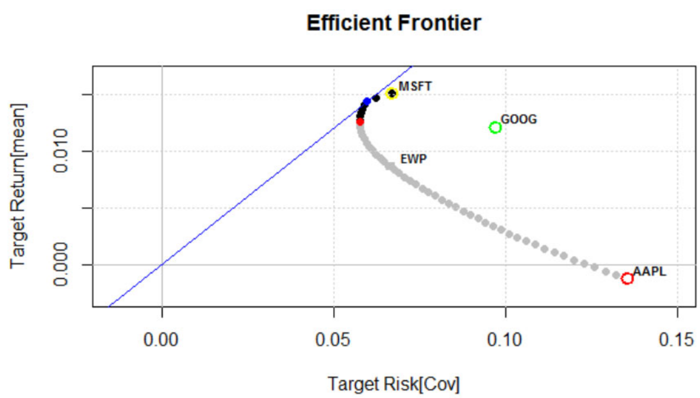


Optimization Result and Efficient Frontier of Portfolio B

Compared to portfolio A (9 stocks), the target return increases from 0.0127 to 0.0132 by 0.0005, and the risk increases from 0.0390 to 0.0408 by 0.0018, more significantly than the return.

The below is tangency result of portfolio B (6 stocks):

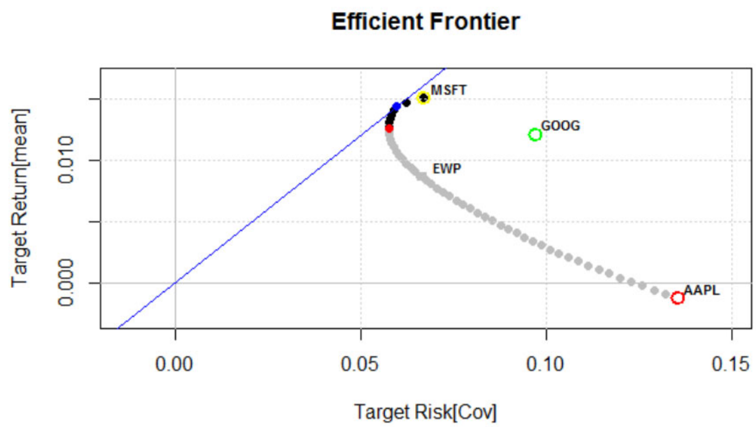
```
Title:
 MV Tangency Portfolio
 Estimator:         covEstimator
 Solver:            solveRquadprog
 Optimize:          minRisk
 Constraints:       LongOnly

Portfolio Weights:
  AAPL   MSFT   GOOG
0.0000 0.7748 0.2252

Covariance Risk Budgets:
  AAPL   MSFT   GOOG
0.0000 0.8105 0.1895

Target Returns and Risks:
  mean    Cov   CVaR    VaR
0.0144 0.0598 0.0888 0.0774
```



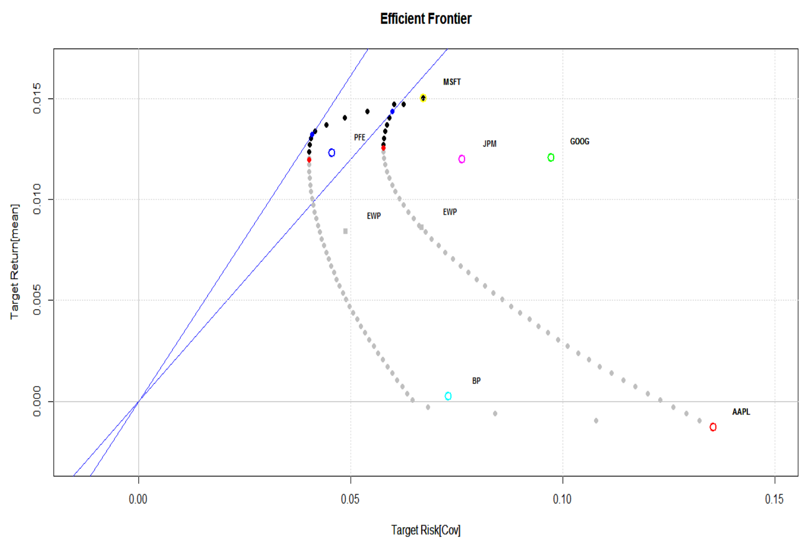
Optimization Result and Efficient Frontier of Portfolio C

Compared to portfolio B (6 stocks), the target return increases from 0.0132to 0.0144 by 0.0012, and the risk increases from 0.0408 to 0.0598 by 0.0190, more significantly than the return.
Compared to portfolio A (9 stocks), the target return increases from 0.0127 to 0.0144 by 0.0014, and the risk increases from 0.0390 to 0.0598 by 0.0208, more significantly than the return, almost 15 times.

So far, the numerical results already show that, diversification improves portfolio's performance, by lowering portfolio risk in a large amount, with sacrifice of little return.
If we plot efficient frontiers of those three portfolio together, a clearer view is shown:

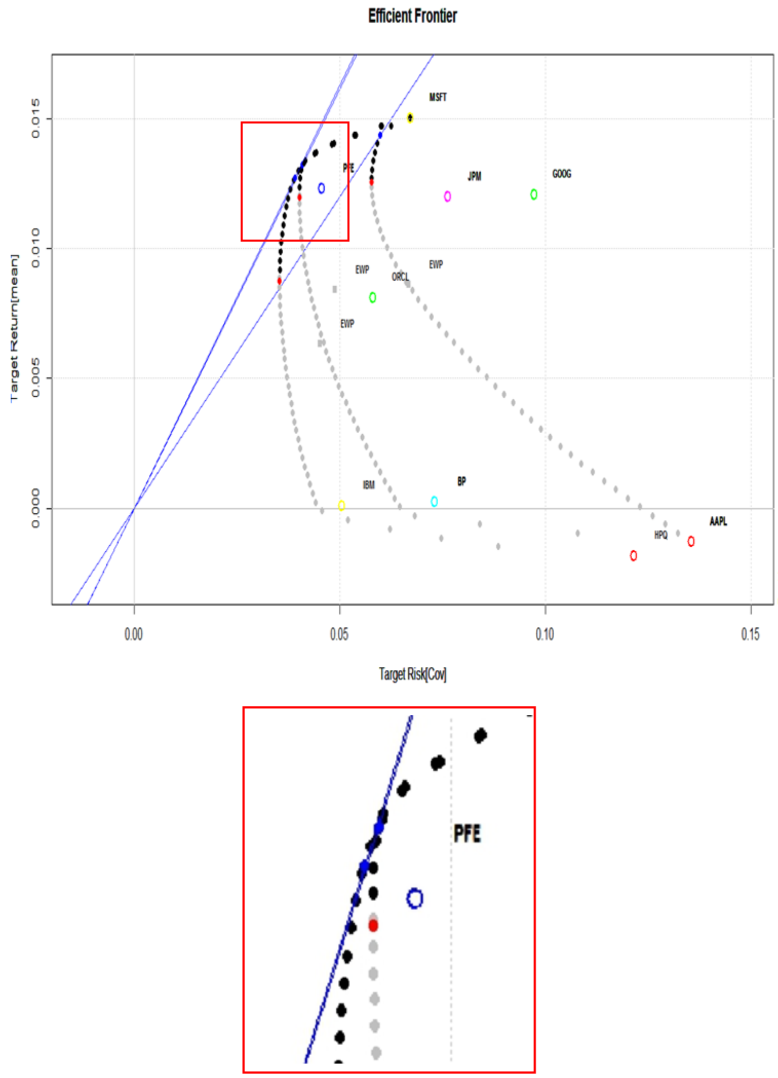


Efficient Frontier of Portfolio C



Diversification Effect with Portfolio B and C

Compared to the first plot, the efficient frontier is largely pushed to the left, resulting in lower risks.



Diversification Effect with Portfolio B, C and A

Extension of portfolio B to A, leads to a slight decrease in risk. But there appears another question: why this time the decrease in risk is much smaller?

Looking back to the new 3 stocks added to portfolio B to form A, they are IBM, ORCL and HPQ, and they are all in IT industry. They are highly correlated with each other and with the stocks already in the previous portfolio. So, even increasing the number of stocks considered, the improvement is very small. That also indicates that, diversification not only means more stocks considered to construct a portfolio, but more importantly, less correlated stocks.

	Total Portfolilo	Tangency Portfolio			
	stocks(n)	Markowitz Selection	E(r)	Cov	Sharpe
Portfolio A	9	MSFT,GOOG,PFE,ORCL	0.0127	0.0390	0.3256
Portfolio B	6	MSFT,GOOG,PFE	0.0132	0.0408	0.3235
Portfolio C	3	MSFT,GOOG	0.0144	0.0598	0.2408

Summary Table of Portfolio A, B and C

The above summary shows that, as more stocks introduced into portfolio, the mean of return is decreased a bit, but risk is lowered at a large amount and Sharpe ratio is increased significantly.

Conclusions

Of the four methods – Feasible Portfolio, Minimum Risk Efficient Portfolio, Global Minimum Risk Portfolio and Tangency Portfolio, Tangency Portfolio, by minimizing the Sharpe Ratio for a given risk-free rate, is the optimal result for portfolio optimization, which considers an equilibrium between return and risk.
Extension of portfolio show a better performance. Diversification improves portfolio's performance, by lowering portfolio risk in a large amount, with sacrifice of little return. And diversification here not only means more stocks considered to construct a portfolio, but more importantly, less correlated stocks.

Future work: 1. The numbers of stocks considered will be extended to hundreds and thousands to see the improvement at different number scale. 2. Analysis of how to diversify portfolio so that portfolio return will be improved.

References

Bodie, Kane, Marcus, Mohanty, “Investments”. 2008
Markowitz, H.M. (March 1952). "Portfolio Selection". The Journal of Finance. 7 (1): 77–91
Ross, Westerfield, Jaffe, “Corporate Finance”. 2010
Wurtz, Lam, Ellis, Chalabi, “Basic R for Finance”. 2010
Wurtz, Lam, Ellis, Chen, “Portfolio Optimization with R”. 2009