**Investment Portfolio Analysis with R Language**

[Leave a reply](http://www.dataapple.net/?p=59#respond)

[Share on Facebook](https://www.facebook.com/sharer/sharer.php?u=http%3A%2F%2Fwww.dataapple.net%2F%3Fp%3D59&t=Investment+Portfolio+Analysis+with+R+Language)[Share on LinkedIn](http://www.linkedin.com/shareArticle?mini=true&url=http%3A%2F%2Fwww.dataapple.net%2F%3Fp%3D59&title=Investment+Portfolio+Analysis+with+R+Language)[Share on Twitter](https://twitter.com/share?url=http%3A%2F%2Fwww.dataapple.net%2F%3Fp%3D59&text=Investment+Portfolio+Analysis+with+R+Language)

*[Please note that this article is for studying R programming purpose only rather than for advices of investment]*

R has a wide application in finance analysis areas such as time series analysis, portfolio management, and risk management, with its basic functions and many professional packages in Finance. In this article, we will demonstrate how to help investors to optimize the investment portfolio with R language by a simple example.

**Scenario Introduction**

Peter is interested in investing the US stock markets and he has selected four stocks, 3M (MMM), Johnson & Johnson (JNJ), P&G (PG), and Wal-Mart (WMT) for consideration. To avoid “putting all eggs in one basket”, his investing strategy is to invest in all of the four stocks rather than to invest in only one stock.

In this case, what are the optimal asset allocation percentages among the four stocks, assuming no short? How can we use R language to conduct the portfolio management analysis for Peter?

**Basic Concepts**

Before we dive into the R programming, let’s review some of the basic concepts about the financial portfolio management.

**Expected Return**

Expected Return is a weighted-average return of a portfolio. The return of each asset in the portfolio can be the mean of return values of the asset in a past period of time; it can also be calculated by the capital asset pricing model (CAPM), which is used to determine a theoretically appropriate required rate of return of an asset1. It can also calculated by other factor models. We will use CAPM to calculate the expected return of each stock in this article.

**Risk**

Risk is the variation of the return of an asset. It can be calculated by the standard deviation of the historical return data; it can also be calculated by factor models. We will use one-factor model to calculate the variance of an asset and the covariance of two assets in a portfolio.

**Sharpe Ratio**

“The Sharpe ratio characterizes how well the return of an asset compensates the investor for the risk taken”2.  The higher Sharpe Ratio means the better investment option. In a set of risky assets, we can find the optimal portfolio asset allocations so that the Sharpe Ration is the largest.

**Capital Asset Pricing Model (CAPM) 1**

The CAPM is a model for pricing an individual security or portfolio. The formula of CAPM is in the following:-

E(Ri)  =  Rf + βi \* (E(Rm) – Rf)

Where:

E(Ri) is the expect return of an asset

Rf is the risk-free return, we usually use the interest rate of T-Bills as the Rf

E(Rm) is the expected return of the market, we usually use SP500 index return as the US stock market

βi is the sensitivity of the expected excess asset return to the expected excess market return. We can get the value of βi by do a regression on the historical data of the excess returns of an asset and the market.

**Explore the data**

The historical monthly return data of SP500, MMM, JNJ, PG, and WMT from April 2010 through April 2013, can be downloaded from [finance.yahoo.com](http://finance.yahoo.com). The one month T-Bills interest rate, which we use as the risky free return,  can be download from [here](http://http:/research.stlouisfed.org/fred2/series/TB4WK/downloaddata?cid=116). The SP500 returns are used as the market return. Here is the [csv file of the returns](http://www.dataapple.net/wp-content/uploads/2013/05/return.csv).

#read the historical return data from .csv file

> return <- read.csv(file.choose(),header=T)

#list each data column names and the first six lines of the dataset

> head(return)

Month T.Bills    SP500       MMM       JNJ         PG         WMT

1 2010/4/1  0.0015 -0.08197592 -0.100012182 -0.084992210 -0.017177262 -0.05195068

2 2010/5/1  0.0015 -0.05388238 -0.004060639  0.013053348 -0.018198198 -0.04925373

3 2010/6/1  0.0008  0.06877783  0.083038869 -0.016433240  0.027711507  0.06503700

4 2010/7/1  0.0016 -0.04744917 -0.076044673 -0.009303209 -0.024464286 -0.01495052

5 2010/8/1  0.0015  0.08755110  0.103897868  0.086814872  0.005125389  0.06755024

6 2010/9/1  0.0012  0.03685594 -0.028666339  0.028566390  0.068111455  0.01201442

We also calculated the excess return data set for our further usage in this [excess return csv file](http://www.dataapple.net/wp-content/uploads/2013/05/ExcessRt.csv).The excess return is equal to monthly return minus monthly T-Bills interest rate.

#read the excess return from .csv file

>exReturn <- read.csv(file.choose(),header=T)

#list each data column name and the first five lines of the dataset

>head(exReturn)

Month   SP500\_ExcessR  MMM\_ExcessR JNJ\_ExcessR  PG\_ExcessR WMT\_ExcessR

1 2010/4/1   -0.08347592 -0.101512182 -0.08649221 -0.018677262 -0.05345068

2 2010/5/1   -0.05538238 -0.005560639  0.01155335 -0.019698198 -0.05075373

3 2010/6/1    0.06797783  0.082238869 -0.01723324  0.026911507  0.06423700

4 2010/7/1   -0.04904916 -0.077644673 -0.01090321 -0.026064286 -0.01655052

5 2010/8/1    0.08605110  0.102397868  0.08531487  0.003625389  0.06605023

6 2010/9/1    0.03565594 -0.029866339  0.02736639  0.066911455  0.01081442

Let’s further explore the means and standard deviations of the excess returns.

# remove the date column

exReturn <- exReturn[,-1]

# mean and standard deviation

Mean=sapply(exReturn,mean)

SD=sapply(exReturn,sd)

cbind(Mean,SD)

Mean         SD

**SP500\_ExcessR   0.008837961    0.04298597**

MMM\_ExcessR   0.008758935    0.05381220

JNJ\_ExcessR      0.010740620    0.03862065

PG\_ExcessR      0.008919860    0.03794766

WMT\_ExcessR    0.012765898    0.04456828

We will use the mean and standard deviation of SP500 excess returns as the market excess return and risk (standard deviation) in the following parts.

**CAPM Analysis**

According to the CAPM formula, we will first get the beta of each stock by regressions; then further calculate the expected return of each stock and the covariance matrix of the four stocks; finally we can calculate the optimal asset allocations (weights) of the portfolio consisting of the four stocks.

**Beta**

We can conduct the regression on the excess returns of each stock and the excess return of market (SP500). Take the MMM stock for example as follows.

# conduct the regression to get the beta of MMM

>lm.MMM<- lm(MMM\_ExcessR ~ SP500\_ExcessR,df)

>summary(lm.MMM)

Call:

lm(formula = MMM\_ExcessR ~ SP500\_ExcessR, data = df)

Residuals:

Min        1Q    Median        3Q       Max

-0.066906 -0.015134  0.006473  0.019116  0.053404

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept)   **-0.0005611**  0.0049388  -0.114     0.91

SP500\_ExcessR  **1.0545439**  0.1140273   9.248 6.29e-11 \*\*\*

—

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: **0.02941** on 35 degrees of freedom

Multiple R-squared: 0.7096,  Adjusted R-squared: 0.7013

F-statistic: 85.53 on 1 and 35 DF,  p-value: 6.291e-11

The beta of MMM is 1.0545439 as shown in the above. The CAPM formula of MMM is:-

E(RMMM) – Rf  =βMMM \* (E(Rm) – Rf), that is,

Expected excess return of MMM =βMMM \* Expected excess return of SP500,

Where:

The expected excess return of SP500 is the mean value, which is 0.008837961, of the excess SP500 return in the past three years.

So, the expected excess return of MMM = 1.0545439 \* 0.008837961 = 0.009320017.

According to Single-Factor Model, the variance of MMM is:-

σMMM2= βMMM2σM2+ σ2(e)

Where,

σM2 is the variance of SP500, which is 0.0429859752

σ2(e) is the variance of residual standard error, which is 0.029412, as shown in the above output.

So, σMMM2 = 1.05454392 \* 0.0429859752 + 0.029412 = 0.0540352772

We can follow the same procedures to get the betas, variances, and expected excess returns of the stocks of JNJ, PG, and WMT as follows.

MMM          JNJ           PG           WMT

Variances        0.0540352772 0.039036492   0.0384049982 0.0451064012

beta             1.0545439      0.438671     0.347552             0.39777

Expected Excess Rt         0.009320017   0.003876957   0.003071651   0.003515476

Based on the above results, we can further calculate the covariance in pairs of the four stocks by the formula of σij= βi βjσM2 .  So, for the covariance between MMM and JNJ,

σMMM,JNJ = βMMM βJNJσM2 = 1.0545439 \* 0.438671 \* 0.0429859752 = 0.000854786.

We can calculate the other covariances by the same way and build the covariance matrix as follows.

MMM          JNJ           PG           WMT

MMM      0.002919811   0.000854786   0.000677233   0.000775087

JNJ       0.000854786  0.001523848   0.000281716   0.000322422

PG        0.000677233  0.000281716   0.001474944   0.00025545

WMT        0.000775087   0.000322422   0.00025545     0.002034587

**Mean-Variance Portfolio Optimization**

Now with the expected excess returns and the covariance matrix ready, we can conduct the Mean-Variance Portfolio Optimization to get the optimal allocation (weight) of each stock so that the Sharpe Ratio of the portfolio is maximized.

Here we will use Guy Yollin’s “effFrontier” and “maxSharpe” functions3, which use the core function of “portfolio.optim” in the “tseries” R package, for the calculations of efficient frontier and maximum Sharpe Ratio.

#load tseries package

library(tseries)

#build the expected excess return matrix and covariance matrix for the parameters of “maxSharpe” #function. The numbers are from the calculations in the previous parts.

> averet <- matrix(c(0.009320017,0.003876957,0.003071651,0.003515476), nrow=1)

>rcov<-matrix(c(0.002919811,0.000854786,0.000677233,0.000775087,0.000854786,0.001523848,0.000281716,0.000322422,0.000677233,0.000281716,0.001474944,0.00025545,0.000775087,0.000322422,0.00025545,0.002034587),nrow=4)

# get the optimal weights

> weights <- maxSharpe(averet,rcov,shorts=F)

> weights

[1] 0.5779719 0.1794985 0.1332271 0.1093025

According to the above output, Peter can allocate about **58% of his capital in MMM, 18% in JNJ, 13% in PG, and 11% in WMT**, to get a maximum Sharpe Ratio roughly, which means a maximum value of expected excess return per risk unit.

**Summary**

In this article, we went through some basic related finance concepts and demonstrated how to use R to conduct the portfolio management.

**References**

1. <http://en.wikipedia.org/wiki/Capital_asset_pricing_model>

2. <http://en.wikipedia.org/wiki/Sharpe_ratio>

3. <http://www.rinfinance.com/RinFinance2009/presentations/yollin_slides.pdf>

4. HKUST Prof. Fei Ding’s lecture slides.

**Data Sources**

Stock:  <http://finance.yahoo.com/>

T-Bills:  <http://research.stlouisfed.org/fred2/series/TB4WK/downloaddata?cid=116>

*Author: Jack Han. All rights reserved. 转载须以超链接形式标明文章原始出处和作者信息*