Title
R functions development for stockPortfolio package

Author
Luo, Rui

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

R functions development for *stockPortfolio* package

A dissertation submitted in partial satisfaction of the

requirements for the degree Master of Science

in Statistics

by

Rui Luo

2013

ABSTRACT OF THE DISSERTATION

R functions development for *stockPortfolio* package

by

Rui Luo

Master of Science in Statistics

University of California, Los Angeles, 2013

Professor Qing Zhou, Chair

Modern portfolio theory is a statistical framework to allocate investment assets properly, with the aim of reducing risk by diversification. In the past decades, a variety of index and group models (with different covariance assumption) have been proposed to optimize the portfolio, including Single Index Model, Constant Correlation Model, Multi-Group Model, and Multi-Index Model. An R package "*stockPortfolio*" is developed by Drs. Christou and Diez, and fully implemented Single Index Model, Constant Correlation Model. Besides, this package also includes functions to download historical data from Yahoo Finance, build models, estimate optimal portfolios, and test portfolios. However, *stockPortfolio* package does not include the optimization functions for Multi-Group & Multi-Index Models with & without short selling. Besides, corresponding covariance matrix calculation functions are also not completed for these two models. My thesis will implement all these missing elements and aims to make a complete *stockPorfolio* package. The performance of these functions will be evaluated with multiple different historical data sets.

The dissertation of Rui Luo is approved.

Nicolas Christou

Alan Yuille

Qing Zhou, Committee Chair

University of California, Los Angeles

2013

I dedicate my thesis to to my family

Kangjin Luo, Jianqing Lu, and Chaochao Cai

for their unconditional love

and support of my study abroad

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to acknowledge the members of my thesis committee, Drs. Qing Zhou, Nicolas Christou, and Alan Yuille, for their suggestions. I especially would like to thank Professor Nicolas Christou for providing me with the invaluable guidance and support to write this thesis

1 Introduction

Introduced by Harry Markowitz in the 1950s, modern portfolio theory (MPT) is a theory which reduces the risk of investment by diversification. It attempts to construct optimal portfolio by carefully calculating the proportions of various assets with math and statistical method. Under this theory, risk-averse investors can construct an "efficient frontier" of optimal portfolios and maximize portfolio expected return for a given level of market risk, or equivalently minimize portfolio risk based on given portfolio expected return. This theory emphasizes that risk is an inherent part of higher reward, but one can reduces risk without reducing expected return by diversification.

After the proposal of MPT, a series of extensive studies were carried out to optimize the portfolio with index and group models. The most frequently used models include Single Index Model, Constant Correlation Model, Multi-Group Model, and Multi-Index Model. All these models require certain assumption about covariance matrix and can simplify the input for portfolio optimization problem. Each model will produce a unique result, and lead to slightly different allocation of stocks. An R package "*stockPortfolio*" is developed by Drs. Christou and Diez, in which fully implemented Single Index Model, Constant Correlation Model. Besides, this package also includes functions to download historical data from Yahoo Finance, build models, estimate optimal portfolios, and test portfolios.

However, *stockPortfolio* package mainly focuses on Single Index Model, Constant Correlation Model, and does not include the optimization functions for Multi-Group & Multi-Index Models with & without short selling. Besides, corresponding covariance matrix calculation functions are also not completed for these two models. My thesis will implement all these missing elements and aims to make a complete *stockPorfolio* package. The performance of these functions will be evaluated with multiple historical data sets downloaded from Yahoo Finance.

2 Review of modern portfolio theory

2.1 Concepts and Key terms

MPT is a statistical framework to allocate investment assets properly, with the aim of reducing investment risk by diversification. By combining different assets with weak correlation, or negative correlation, it is possible for risk-averse investors to construct optimal portfolios which maximizes portfolio expected return for a given level of market risk, or equivalently minimizes portfolio risk based on given portfolio expected return [1, 2]. MPT defines an asset's return as a random variable (i.e. normal distributed), and risk as standard deviation of return. The portfolio is modeled as weighted combination of assets, so that the return of a portfolio is easily the weighted linear combination of all assets' returns [3, 4]. MPT also assumes that investors are rational and markets are efficient [5].

Return: The definition of return varies across different investment vehicles, for example, the return of debt instruments may include price appreciation, interest or principle payments. For stocks, the return contains both capital appreciation (the price of the stock rises) and dividends [6-8]. The expected return for single asset is calculated based on historical performance, and portfolio return is the weighted combination of the constituent assets' returns (Equation 1).

$$E\left(R_p\right) = \sum_i w_i E\left(R_i\right), \quad \text{(Equation 1)}$$

where R_p is the return of the portfolio, R_i is the return for ith asset. w_i is the weight for corresponding asset, and the sum of w_i is 1.

Risk: In the MPT, risk is the measure of variability for the expected return. And the risk of single asset is simply the standard deviation of its historical return, which is valid if asset returns are jointly normally distributed or elliptically distributed [9, 10]. Thus, the risk for portfolio is easily the standard deviation of portfolio historical return (Equation 2-3).

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}, \text{ (Equation 2)}$$

where σ_p^2 is the variance of the portfolio, σ_i is the risk for asset i. w_i is the weight for corresponding asset. ρ_{ij} is the correlation coefficient between the returns of assets i and j.

So, the risk of the portfolio is:

$$\sigma_p = \sqrt{\sigma_p^2}, \text{ (Equation 3)}$$

Diversification: By investing assets with weak or negative correlated assets, the risk of the portfolio can be diversified and reduced. Diversification may allow portfolio to reduce the risk while maintain the same expected return. In order to reach best diversified portfolio, it is really important to understand the underlying covariance matrix across all assets, and several index models have been proposed which will be introduced in later section. Besides, the correlation matrix may vary with time, and one needs to update covariance matrix based on real market [8, 11-13].

Efficient frontiers: Every possible combination of the risky assets (without risk free assets) can be plotted in risk-expected return space. And the left boundary of all plots is a hyperbola.

Optimal portfolios lying on the upper edge offers the highest expected return for a defined level of risk or the lowest risk for a given level of expected return. Thus it is called efficient frontier[1, 14].

Risk free: Risk free assets are investment with no risk of financial loss. For example, treasuries are considered to be risk free because they are backed by the U.S. government. We can consider incorporating risk free assets inside the portfolio by borrowing or lending at risk free rate (*R_{rf}*). Since the return is constant, the risk (standard deviation) of the return on the risk free asset should be zero. Besides, the correlation between returns from risk free asset and any others should be zero. The return of combined portfolio can be expressed as Equation 4.

$$R_c = (1-x)R_f + xR_p, \text{ (Equation 4)}$$

where R_c is the expected return for combined portfolio, and R_p is the expected return for portfolio without risk free assets. x is the fraction of portfolio without risk free assets, while (1-x) is the fraction of risk free assets.

The risk of combined portfolio can be expressed as Equation 5.

$$\sigma_c = \left[(1-x)^2 \sigma_{rf}^2 + x^2\sigma_p^2 + 2x(1-\mathrm{x})\sigma_p\sigma_{rf}\rho\right]^{1/2}, \text{ (Equation 5)}$$

Since the correlation between risk free asset and previous portfolio is zero, the final risk can be simplified as Equation 6.

$$\sigma_c = x\sigma_p, \text{ (Equation 6)}$$

Combined Equation 4 and 6, we get Equation 7.

$$R_c = R_{rf} + \left(\frac{R_p - R_{rf}}{\sigma_p} \right) \sigma_c \text{, (Equation 7)}$$

It is obvious that, all combined portfolios lie on a straight line, with intercept as R_{rf} and slope as $\left(\frac{R_p - R_{rf}}{\sigma_p} \right)$. Besides, this line also passes through (σ_p , R_p).

2.2 Portfolio optimization models

In the past decades, a variety of index and group models have been proposed to optimize the portfolio extensively, including Single Index Model, Constant Correlation Model, Multi-Group Model, Multi-Index Model[15-18]. Each of these models makes assumption about the underlying covariance structure for stocks, and leads to a unique stock allocation solution.

Single Index Model

The Single Index Model is a simple asset-pricing model for portfolio construction. Because of its simplification, it is the oldest and widely used model. The model is expressed as Equation 8.

$$R_i = \alpha_i + \beta_i R_m + \varepsilon_i \text{, (Equation 8)}$$

where R_i is the return of stock *i*, α_i is the interpret, β_i is the slope of stock *i* to the market return, R_m is the return of market, and ε_i is the residual which follows normal distribution.

Under this model, we can estimate the mean return of *i*th stock as $\overline{R_i} = \alpha_i + \beta_i R_m$. And the risk of *i*th stock is $\alpha_i = \sqrt{\beta_i^2 \sigma_m^2 + \sigma_\varepsilon^2}$, and the covariance between stock *i* and stock *j* is $\beta_i \beta_j \sigma_m^2$.

Constant Correlation Model

Constant Correlation Model is another model that is used to define the covariance structure in the portfolio. It assumes that the correlation between all pairs of stocks is the same. The most important definition in Constant Correlation Model is the estimate of correlation $\bar{\rho}$ in the portfolio, and it is usually defined as Equation 9.

$$\bar{\rho} = \frac{\sum_{i=1}^{N}\sum_{j=1;i\neq j}^{N}\rho_{ij}}{\frac{N(N-1)}{2}} \text{ , (Equation 9)}$$

ρ_{ij} is the correlation between stock *i* and *j*, and N is the number of stocks.

Multi-Group Model

The assumption of constant correlation model is generally too strong in reality. To improve that, Multi-Group Model is introduced, which assumes that the correlation between stocks from the same group is identical. The correlation of any stocks between two groups is also constant, while varies from different pair of groups. With this definition, the correlation matrix is partitioned into several submatrices. The correlations are the same within each submatrix, while their values might be different among the submatrices. A concrete example of this is that, we suppose stocks 1, 2, 3 belong to industry 1 (bank stocks), and stocks 4, 5, 6 belong to industry 2 (biomedical stocks). Then the correlations are as follows:

Group 1: $\rho_{12} = \rho_{13} = \rho_{23} = \rho_{11}$; Group 2: $\rho_{45} = \rho_{46} = \rho_{56} = \rho_{22}$; Between Group1 and Group 2: $\rho_{14} = \rho_{15} = \rho_{16} = \rho_{24} = \rho_{25} = \rho_{26} = \rho_{34} = \rho_{35} = \rho_{36} = \rho_{12}$. Where ρ_{11} is the correlation from Group 1, and ρ_{22} is the correlation in Group 2, while ρ_{12} is the constant correlation between Group 1 and Group2. And the correlation matrix is like below:

$$\rho = \left(\begin{array}{ccc|ccc} 1 & \rho_{11} & \rho_{11} & \rho_{12} & \rho_{12} & \rho_{12} \\ \rho_{11} & 1 & \rho_{11} & \rho_{12} & \rho_{12} & \rho_{12} \\ \rho_{11} & \rho_{11} & 1 & \rho_{12} & \rho_{12} & \rho_{12} \\ \hline \rho_{21} & \rho_{21} & \rho_{21} & 1 & \rho_{22} & \rho_{22} \\ \rho_{21} & \rho_{21} & \rho_{21} & \rho_{22} & 1 & \rho_{22} \\ \rho_{21} & \rho_{21} & \rho_{21} & \rho_{22} & \rho_{22} & 1 \end{array}\right)$$

Multi-Index Model

Multi-Index Model is an improvement of Single-Index Model. It is found that, other than the general market index, there are other factors will affect the stock price. Multi-Index Model uses economic factors and structures to explain the stock movements above and beyond what market indexes already explain, so here we use Multi-Index Model to predict correlation of stocks by grouping them by industry. In this model, we assume that stocks are linearly related to the group index (industry) and the industry is linearly related to the market index. The model is expressed as the following equations (Equation 10-11).

$$R_i = \alpha_i + \beta_i I_j + \varepsilon_i \text{, (Equation 10)}$$

$$I_j = r_j + b_j R_m + c_j \text{ , (Equation 11)}$$

where R_i is the return of stock *i*, I_j is the return of index *j*, R_m is the return for market, ∂_i & β_i are the parameter for stock *i*, and r_j & b_j are the parameter with industry index *j*, ε_i is the error term for stock *i*, and c_j is the error term for industry *j*, both error terms follow normal distribution.

Based on this equation, we have the risk of stock *i* as Equation 12.

$$\sigma_i = \sqrt{\beta_i^2\left(b_j^2\sigma_m^2 + \sigma_{cj}^2\right) + \sigma_{\varepsilon_i}^2}$$, (Equation 12)

The covariance between stock *i* and stock *k* from the same industry can be expressed as $\beta_i\beta_j\left(b_j^2\sigma_m^2 + \sigma_{c_j}^2\right)$. And the covariance between stock *i* and stock *k* from different industries can be expressed as $\beta_i\beta_k b_j b_i \sigma_m^2$.

2.3 *stockPortfolio* R package

stockPorfolio R package is originally developed by Dr. David Diez and Dr. Nicolas Christou. This package implements different quantitative approaches for portfolio optimization, including Single Index Model, Constant Correlation Model, Multi-Group Model, Multi-Index Model. Besides, the package includes functions to download historical data from Yahoo Finance, build models, estimate optimal portfolios, and test portfolios. A large range of graphical features has also been included for visual understanding.

This package provides a good framework for portfolio construction, and a common starting point is *getReturns* function, which can be used to obtain stock data using an Internet connection. Using an object of class "stockReturns" from the *getReturns* function, one can build a stock model using the *stockModel* function. There are four model options in *stockModel*: no model, Single Index Model, Constant Correlation Model, and Multi-Group Model. After a stock model has been built, the user can obtain an estimate of the optimal portfolio allocation among those stocks using *optimalPort* function. Additionally, one can test out models and portfolios on data sets that are either supplied by the user or are output from *getReturns* function. Besides, this package also offers two specialty plotting functions: *portPossCurve* function plots the portfolio possibilities curve based on a model, and *portCloud* function plots a cloud of possible portfolios based on a model.

However, this package mainly focuses on portfolio optimization with Single Index Model and Constant Correlation Model, and does not include the optimization functions for Multi-Group & Multi-Index Models with or without short selling. Besides, corresponding covariance matrix calculation functions are also not completed for these two models. My thesis will implement all these missing elements and aims to make a complete *stockPorfolio* package.

3 Data and portfolio inputs

In this study, three data sets (stock94, stock99, and stock04) embedded in in *stockPortfolio* library will be used. All these three data sets contain monthly price from 25 exactly same variables (24 stocks and S&P500 index), but covers different time windows. Among these 24 stocks, 6 industries are covered: 5 are from Biotechnology (AMGN, GILD, CELG, GENZ, BIIB), 5 are from Electrical Utilities (SO, DUK, D, HE, EIX), 4 are from Fuel Refining (IMO, MRO, HES, YPF), 3 are from Machinery (CAT, DE, HIT), 3 are from Major Airlines (LUV, CAL, AMR) and 4 are from Money Center Banks (C, KEY, WFC, JPM). As shown in their name, stock94 data contains stock prices from October 1994 to September 1999, stock99 data contains stock prices from October 1999 to September 2004, and stock 04 data contains stock prices from October 2004 to September 2009. The reason to utilize these data sets is that they contain the index price as well as stocks from different industries, which serve as a good cohort to test different portfolio models.

In order to acquire these data, the following R code will be used:

```
library(stockPortfolio)
data(stock94)
data(stock99)
data(stock04)
```

Besides, corresponding annotation files for these 25 variables can be found in

data(stock94Info).

Below is the example of the data matrix:

```
                      C         KEY         WFC         JPM
1999-09-01 -0.030756087 -0.04751502 -0.03908356 -0.07470246
1999-08-02 -0.002981261 -0.06342711  0.02557015  0.08608361
1999-07-01 -0.058917836 -0.01955868 -0.08764187 -0.10487445
1999-06-01  0.075431034 -0.07513915  0.06873315  0.19505737
```

After acquiring the data, the Standard and Poor's 500 Index and 24 monthly stock prices are converted to monthly return based on Equation 13.

$$R_i = \frac{p_i - p_{i-1}}{p_{i-1}}$$, (Equation 13)

where p_i is price from *i*th month, and R_i is the return for *i*th month,

4 Functions development for Multi-Group and Multi-Index Models

4.1 No shot selling and corresponding efficient frontier

Traditionally, people invest a stock by buying corresponding shares, and make profit when it appreciates in value. Shorting is the opposite: an investor makes money only when underlying stocks falls in value. More specifically, a short sale is the sale of a security that borrowed from seller, which is promised return in future. If the price of stock drops, you can buy back the stock at the lower price than sale prices, and make a profit on the difference. If the price of the stock rises, you have to buy it back at the higher price, and you lose money.

Short selling is widely used by hedge fund or other investors during the portfolio optimization. This means they are protecting other long positions with offsetting short positions. All the assets with negative proportion during the portfolio construction should be short. However, some market (i.e. merging market) prohibits short selling. No short selling would prevent investors from hedging their long positions, hence may not reach the best combination of assets. In Figure 1 a, the blue dots show all the potential combinations in return and risk space when short selling allowed. In Figure 1 b, the red dots correspond to assets combination when short selling is prohibited. Obviously, red dots are just part of blue dots.

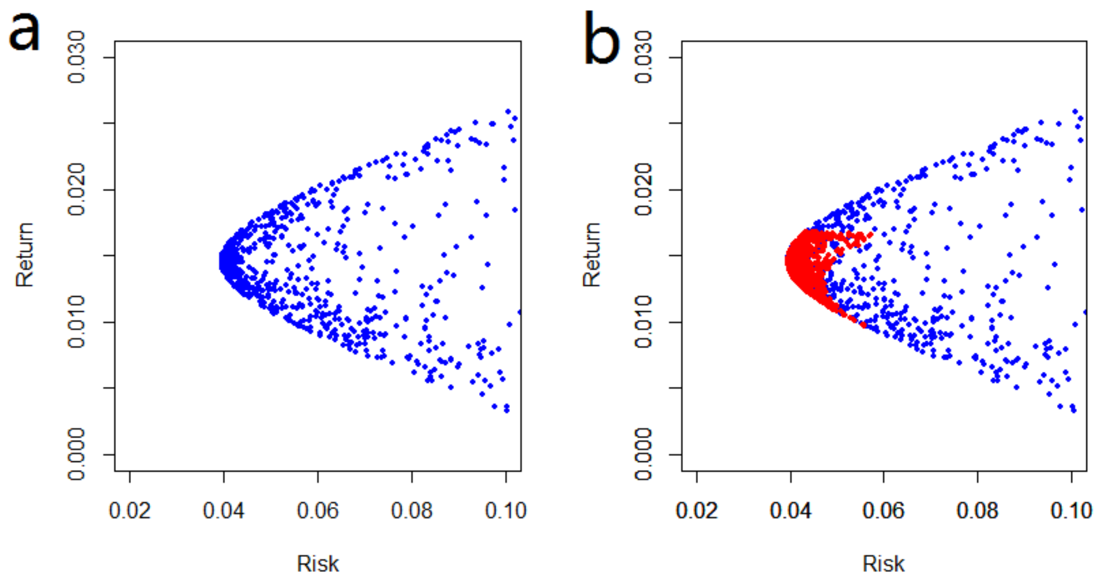

Figure 1: Change of portfolio combination space when short selling is not allowed. (a) The blue dots represent different combinations of stocks in return and risk space when short selling is not allowed. (b)The red dots represent combinations of stocks when no short selling.

4.2 Kuhn-Tucker Conditions

The Kuhn-Tucker Conditions (also known as KKT conditions) are the first order necessary conditions for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. With nonlinear programming, KKT generalizes the method of Lagrange multipliers, which allows only equality constraints. Usually, equations from KKT conditions cannot be solved directly, except in the few cases where a closed form solution can be derived analytically. In general, many optimization algorithms can be interpreted as methods for numerically solving the KKT system of equations[19].

When short sales are allowed, what we need to do is to simply maximize the slope to find the tangent to the efficient frontier with Equation 14.

$$\max \theta = \frac{\bar{R}_p + R_{rf}}{\sigma_p} \text{ , (Equation 14)}$$

where $\overline{R_p}$ is the expected return of portfolio without risk free assets, and σ_p is the corresponding standard deviation. R_{rf} is the return of risk free assets, and the sum of proportions $\sum x_i = 1$.

To find the maximum value of θ, we take the derivatives of θ conditioned on each x_i, and solve the following equations:

$$\frac{\mathrm{d}\theta}{\mathrm{d}x_i} = z_i^2\sigma_i^2 + \sum_{j \neq i}^{N} z_j \partial_{ij} = 0 \text{ , (Equation 15)}$$

$$\overline{R_i} - R_{rf} = z_i^2\sigma_i^2 + \sum_{j \neq j}^{N} z_i \sigma_{ij} \text{ , (Equation 16)}$$

For short sales not allowed, there are more constraints on the range of x_i. It has to be non-negative. We need to utilize Kuhn-Tucker Conditions to solve the problem.

We take the same derivative to each x_i as in short sales allowed. When the maximum of θ occurs at $x_i < 0$, then $\frac{\mathrm{d}\theta}{\mathrm{d}x_i} < 0$, otherwise $\frac{\mathrm{d}\theta}{\mathrm{d}x_i} = 0$. This reaches the first Kuhn-Tucker Condition (Equation 17).

$$\frac{\mathrm{d}\theta}{\mathrm{d}x_i} + u_i = 0 \text{ , (Equation 17)}$$

Regarding the u_i, if the maximum occurs at $x_i > 0$, then $u_i = 0$. If maximum occurs at $x_i = 0$, then $\frac{\mathrm{d}\theta}{\mathrm{d}x_i} < 0$ and $u_i > 0$. In summary we have $x_i u_i = 0, \mathrm{x_i} \geq 0, \mathrm{u_i} \geq 0$, which is second Kuhn-Tucker Condition.

Now, the optimization question we need to solve is changed into Equation 18.

$$\overline{R_i} - R_{rf} = z_i^2 \sigma_i^2 + \sum_{j \neq j}^{N} z_i \sigma_{ij} - u_i \text{ , (Equation 18)}$$

where $x_i u_i = 0, \mathrm{x_i} \geq 0, \mathrm{u_i} \geq 0$ for all stocks.

This optimization process highly depends on the covariance format. It is very complicated and taking a lot of iterations in Multi-Group and Multi-Index models, considering the complicated covariance structure. Many optimization packages, like *optimx* package, can be used to solve this problem with non-linear optimization and quadratic programming. We will apply these methods and develop corresponding functions in the next section.

4.3 Optimization functions

In this section, we mainly focus on the development of missing function in *stockPorfolio* R package, including the covariance matrix estimation function for Multiple Index Model, optimization function for both Multi-Group and Multi-Index Model with or without short selling. The performance of these functions will be tested in the following section.

(1) Updating of *stockModel* function (covariance matrix estimation). Only section for Multiple Index Model is updated and shown as below, and this module will work when the user set module as MIM. The theory of estimation can be found in section 2.2. For complete function, please refer to *stockPorfolio* library in CRAN.

```
if(tM$model == 'MIM'){
    tM$theIndex <- tM$ticker[index]
    tM$ticker   <- tM$ticker[-index]
    M.returns   <- tM$returns[,index]
    VM          <- var(M.returns)
    tM$VM       <- VM
    tM$industry <- as.character(tM$industry)
    tM$industry <- tM$industry[-index]
    tM$returns  <- tM$returns[,-index]
    I.returns   <- apply(tM$returns,1,tapply,tM$industry,mean)

    # get the mean return of each industry
    I.returns   <- t(I.returns)
    Icov        <- cov(I.returns)
    IcovM       <- cov(I.returns,M.returns)
    Ib          <- IcovM/VM
    n           <- dim(Icov)[1]+1
    IMSE        <- (n-1)*(diag(Icov)-Ib^2*VM)/(n-2)

    Iindex <- matrix(0,ncol=5,nrow=length(tM$ticker))
    colnames(index) <- c("alpha","beta","MSE","b","IMSE")
    for(i in 1:(ncol(I.returns))){
           loc.col <- which(tM$industry %in% colnames(I.returns)[i])
           Return.use <- data.frame(tM$returns[,loc.col],Industry=I.returns[,i])
           R.use <- apply(Return.use,2,mean)
           Cov.use <- cov(Return.use)
```

```
            Grc.use<- getRegCoef(R.use,Cov.use,ncol(Return.use),dim(Cov.use)[1])
            Iindex[loc.col,1] <- Grc.use$alpha
            Iindex[loc.col,2] <- Grc.use$beta
            Iindex[loc.col,3] <- Grc.use$MSE
            Iindex[loc.col,4] <- Ib[i]
            Iindex[loc.col,5] <- IMSE[i]
        }

        covMatrix <- matrix(0,ncol=ncol(tM$returns),nrow=ncol(tM$returns))
        Iindex     <- data.frame(Iindex)
        beta2     <- Iindex$beta %*% t(Iindex$beta)
        b2        <- Iindex$b %*% t(Iindex$b)
        covMatrix <- beta2 %*% b2 * VM

        for(i in 1:ncol(covMatrix)){
        loc.row<- is.element(tM$industry,tM$industry[i])
        covMatrix[loc.row,i] <- beta2[loc.row,i] * Iindex$IMSE[i]
    }

    diag(covMatrix) <- diag(covMatrix)+Iindex$MSE
    tM$COV <- covMatrix
    tM$rho <- getCorr(tM$COV, tM$industry)
    tM$index.matrix <-Iindex
}
```

(2) Updating of *optimalPort* function (proportion of assets estimation). The optimization models for Multi-Group and Multi-Index models are not available, and we aim to add these two modules for both short selling and no short selling conditions. This optimization process highly depends on the covariance format. It is very complicated and taking a lot of iterations in Multi-Group and Multi-Index models, considering the complicated covariance structure. Here, we will use *optimx* package to solve this problem. The function shown here is not limited to these two modules, it can be generalized to other modules with correct covariance input.

```
optimalPort<-function(model, Rf=NULL, shortSell=T, eps=10^(-4)){
       #===> setup the model <===#
       op <- list()
```

```
        class(op) <- "optimalPortfolio"
        op$model  <- model
        op$X      <- NA
        op$R      <- NA
        op$risk   <- NA

        COV=model$COV
        R=model$returns
        Rf=model$Rf
        CountN=length(model$ticker)
        fr <- function(x){
        -(sum(x*R)-Rf)/sqrt(sum(t(x*COV)*x))
        }
        if(shortSell==NULL||shortSell==T){
        ans1<-optimx(rep(1, CountN),
       fr,method = "L-BFGS-B",lower=rep(1,CountN),upper=rep(1,CountN),itnmax=1000)
        }
        else {
        ans1<-optimx(rep(1, CountN),
       fr,method = "L-BFGS- B",lower=rep(0,CountN),upper=rep(1,CountN),itnmax=1000)
       }

       op$X=(ans1$par[[1]])/sum((ans1$par[[1]]))*100
       op$R= as.matrix(op$X)%*%as.matrix(R)
       op$risk=sqrt(as.matrix(op$X)%*%as.matrix(COV)%*%t(as.matrix(op$X)))

       return(op)
}
```

4.4 Validation of functions and results

In order to validate the accuracy of our updated function, here we carry out comprehensive evaluation with three data sets (stock94, stock99, and stock04) embedded in in *stockPortfolio* library. All these three data sets contain monthly price from 25 exactly same variables (24 stocks and S&P500 index), but covers different time windows. As shown in their name, stock94 data contains stock prices from October 1994 to September 1999, stock99 data contains stock prices from October 1999 to September 2004, and stock 04 data contains stock prices from October

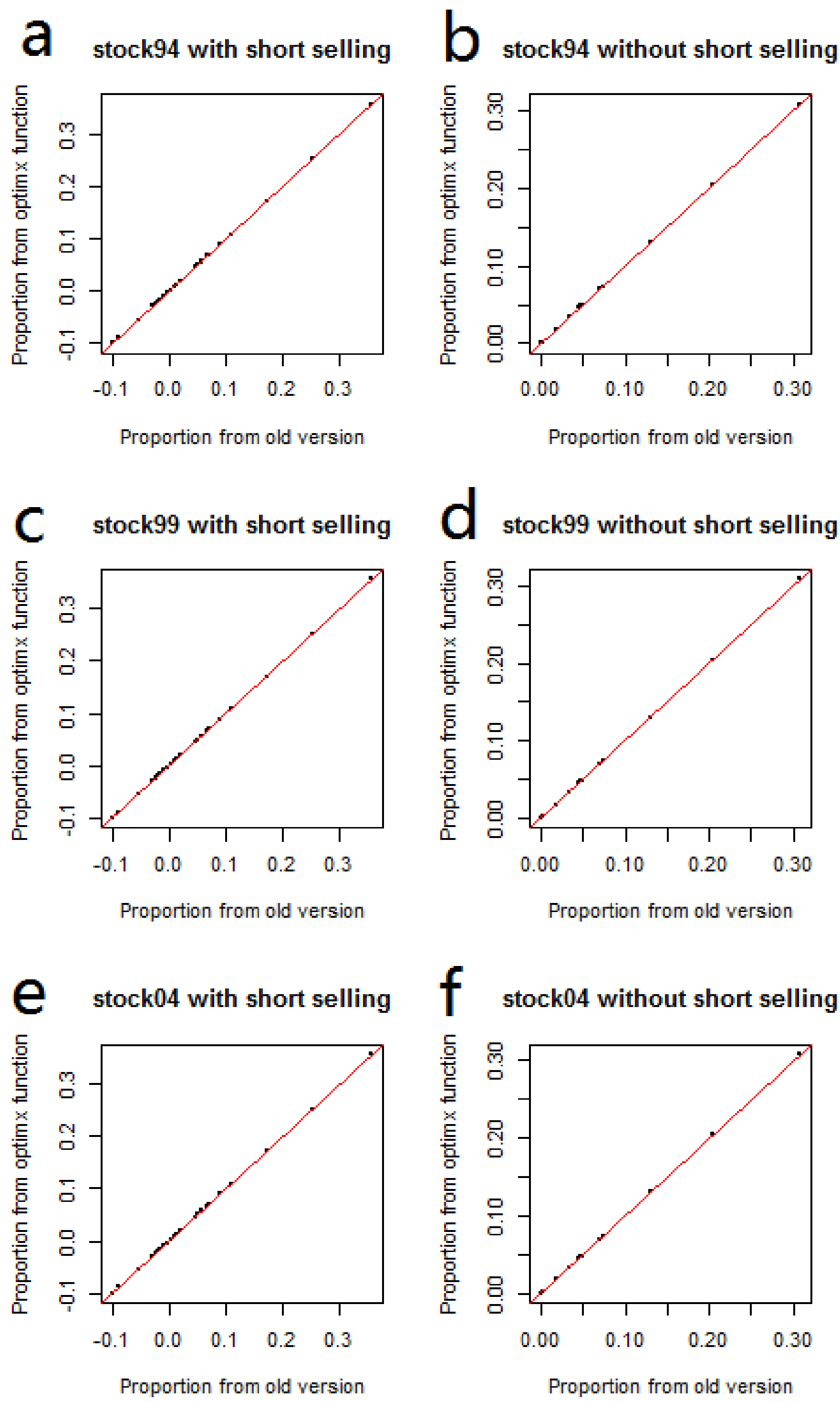

Figure 2: Comparison of results from old function and *optimx* function. In this figure, three rows correspond to three different data sets, stock 94, stock 99, and stock 04. Two columns correspond to short selling and no short selling. Each dot stands for one stock, and the red line is the bench line (y=x).

2004 to September 2009. The new function apply *optimx* package, and work in different way as previous function. So, we try to compare the results generated from previous function and new function under Constant Correlation Model. As shown in Figure 2, no matter with or without short selling, new function with *optimx* generates exactly same result as result from old function. This conclusion is further validated with other two models.

Discussions

In this thesis, we implement optimization function for both Multi-Group & Multi-Index Models with or without short selling. The performance of updated function is validated with three different data sets. Besides, corresponding covariance matrix estimation function is also integrated. The application of new optimization function is not limited to Multi-Group & Multi-Index Models, and can be applied to any model with known covariance matrix and return. At the same time, this new method needs to take a lot of iterations, and is not computationally efficient.

Multiple covariance models have been proposed to better capture the underlying covariance structure, but it remains challenge to accurate estimate the covariance matrix. One of the important reason is the covariance matrix varies with time, and covariance matrix generated by historical data cannot well represent future data. Study the change trend of covariance matrix with time remains interesting.

Portfolio optimization requires lots of assumption, while many of them are not validated in the real financial market. Another problem is that MPT simplify the investment with standard variance and mean return, while these two cannot capture all essential information for the market.

In practice, investors must reach predictions based on historical measurements of asset return and volatility for these values in the equations.

References

1. Markowitz H: Portfolio selection; efficient diversification of investments. New York,: Wiley; 1959.

2. Elton EJ: Modern portfolio theory and investment analysis, 7th edn. Hoboken, N.J.: John Wiley & Sons; 2007.

3. Sherman EJ, International Gold Corporation.: The Application of modern portfolio theory to gold, 1974-1983. New York: International Gold Corporation; 1984.

4. Francis JC, Kim D: Modern portfolio theory : foundations, analysis, and new developments + website. Hoboken, N.J.: Wiley; 2013.

5. RC M: An Analytic Derivation of the Efficient Portfolio Frontier. *Journal of Financial and Quantitative Analysis* 1972, 7:1851-1872.

6. Brock WA, LeBaron BD, National Bureau of Economic Research: A dynamic structural model for stock return volatility and trading volume. In: *NBER working paper series,*. Cambridge, MA: National Bureau of Economic Research,; 1995: 46 p. ill. 22 cm.

7. Andersen TG, National Bureau of Economic Research: The distribution of stock return volatility. In: *NBER working paper series no 7933.* Cambridge, MA.: National Bureau of Economic Research,; 2000: 22, 17 p. ill. 22 cm.

8. Bekaert G, Hodrick RJ, Zhang X, National Bureau of Economic Research: International stock return comovements. In: *NBER working paper series,*. Cambridge, Mass.: National Bureau of Economic Research,; 2005: 56 p. ill. 22 cm.

9. **Jacques K, Nigro P, United States. Office of the Comptroller of the Currency.: Risk-based capital, portfolio risk, and bank capital a simultaneous equations approach. In: *Economic & policy analysis working paper 94-6.* Washington, D.C.: Office of the Comptroller of the Currency,; 1994.**

10. **Connor G, Goldberg LR, Korajczyk RA: Portfolio risk analysis. Princeton, N.J.: Princeton University Press; 2010.**

11. **Goetzmann WN, Kumar A, National Bureau of Economic Research: Equity portfolio diversification. In: *NBER working paper series,.* Cambridge, MA.: National Bureau of Economic Research,; 2001: 37 p. ill, 22 cm.**

12. **Goetzmann WN, Massa M, Simonov A, National Bureau of Economic Research: Portfolio diversification and city agglomeration. In: *NBER working paper series,.* Cambridge, Mass.: National Bureau of Economic Research,; 2004: 44, 42 p. 22 cm.**

13. **Neaime S, American University of Beirut. Institute of Financial Economics.: Portfolio diversification and financial integration of MENA stock markets. Beirut?: Institute of Financial Economics, American University of Beirut; 2004.**

14. **Michaud RO: Efficient asset management : a practical guide to stock portfolio optimization and asset allocation. Boston, Mass.: Harvard Business School Press; 1998.**

15. **Park RC: Portfolio optimization through statistical modeling and forecasting. *M s.* UCLA; 2006.**

16. **Prigent J-L: Portfolio optimization and performance analysis. In: *Chapman & Hall/CRC financial mathematics series.* Boca Raton: Chapman & Hall/CRC,; 2007: xvi, 434 p. ill. 425 cm.**

17. **Wah E: Portfolio optimization through data conditioning and aggregation. *M s.* UCLA; 2010.**

18. **Pfaff B: Financial risk modelling and portfolio optimization with R. Chichester, West Sussex: Wiley; 2013.**

19. **Rentmeesters MJ: A theory of multi-objective optimization : comprehensive Kuhn Tucker conditions for lexicographic and Pareto optima.** ***Ph D , Information and Computer Science.*** **University of California, Irvine; 1998.**