-Aigal Divya & Sawant Raghuraj

INDEX

-Aigal Divya & Sawant Raghuraj

Objective:

The goal of our project is to perform portfolio optimization for the stocks in the portfolio and understand the concepts from class to learn how a portfolio performs based on various scenarios.

Introduction:

In order to understand and apply the concepts learnt in class, we take into consideration a portfolio of 5 stocks from different sectors/domains. The industry domains were chosen from the existing domains which are well known and large-scale companies. The domain choice was arbitrary.

Data Collection:

For this project, we have taken into consideration data for 10 years ranging from 30th Nov 2006 to 13th Oct 2016. We have chosen this range span, because we can get data for the recession period of 2008, and we can understand how various factors affect the stock prices. Following is the table of the domain, the company and its symbol whose stocks we are analyzing for this project:

Sector/Domain	**Company (Symbol)**
Technology	GOOGL (Google)
Retail	WMT (Walmart)
Manufacturing	F (Ford)
Health care	PFE (Pfizer)
Oil & Gas	XOM (Exxon)

Apart from the stocks in our portfolio, we also require a risk-free asset which we have chosen as the 3-month treasury bill.

Data Sources:

http://www.nasdaq.com/symbol/xom/historical - Data source for stocks in our portfolio
https://fred.stlouisfed.org/series/TB3MS- Data source for the 3 month treasury bill

Initial Exploratory Data Analysis:

To begin with it is always important to understand the data at hand well. As an example, the data for Ford is shown by the **head()** command in R. The date, closing price, volume of stocks, and the opening, high and low prices are seen in the data. We have a similar data format for the other companies.

	date	close	volume	open	high	low
1	16:00	11.9	25,773,571	12.0	12.0	11.9
2	2016/10/13	11.9	39876170.0000	11.9	11.9	11.7
3	2016/10/12	12.0	20616640.0000	12.0	12.0	11.9
4	2016/10/11	12.0	39992310.0000	12.2	12.2	11.9
5	2016/10/10	12.1	24937110.0000	12.3	12.4	12.1
6	2016/10/07	12.3	22071310.0000	12.4	12.4	12.2

For the rest of our project, we have considered only the Closing prices for all stocks for all calculations. In order to create an initial exploratory plots for all the 5 stocks, we calculate the returns of the stocks and then plot this data. The returns are calculated as difference in the next and same day stock value for a particular company. The plot below shows the returns data plotted over the range of 2484 days which is equivalent to 10 years of data.

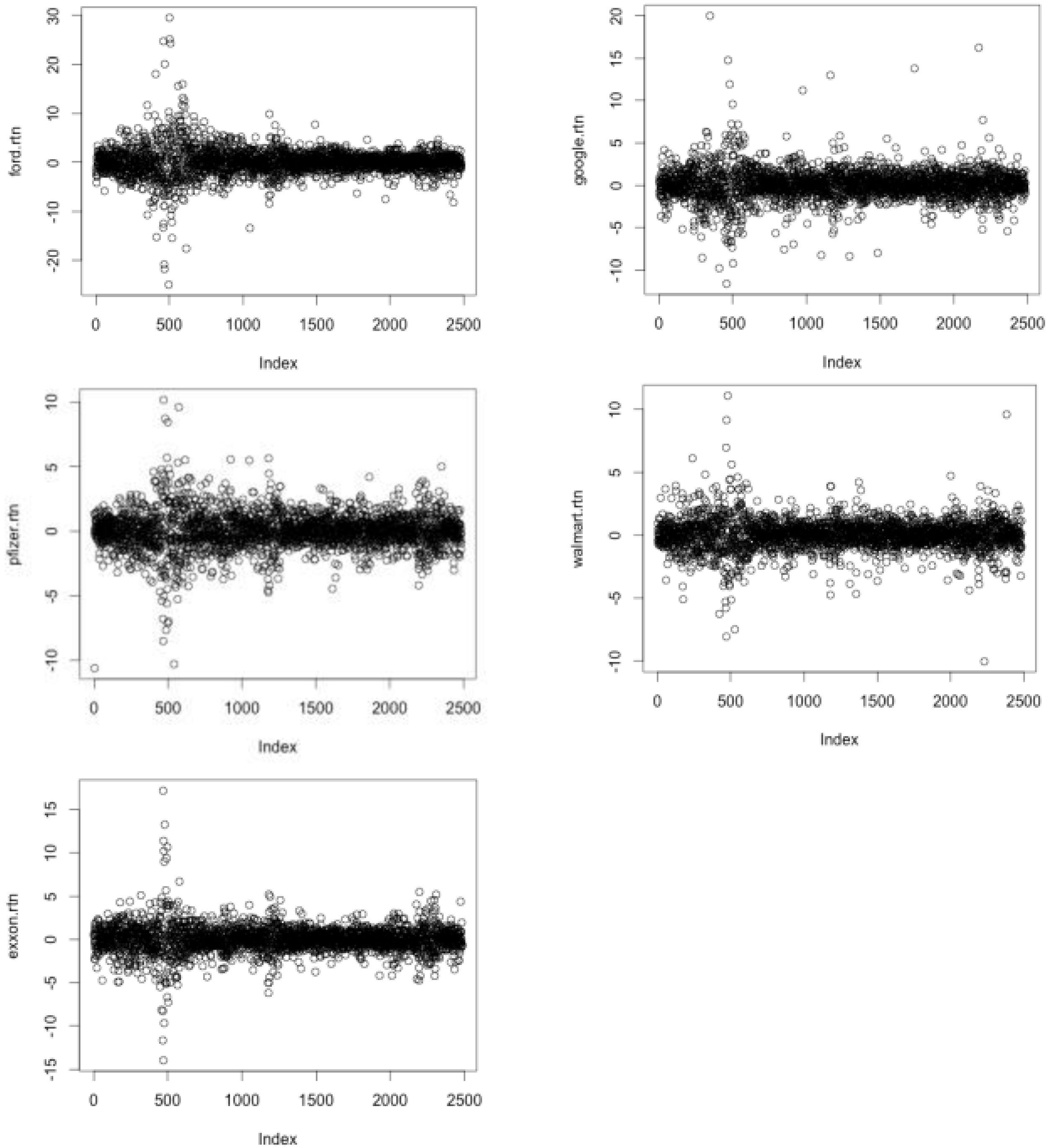


Figure 1: Plotting return for all stocks

Time-Series plots for the stock portfolio:

In order to understand the returns from the stock over a period of time, we use the time series analysis. A time-series plot here depicts the returns of the stock versus the time label for the return. Returns are plotted on the Y-axis and the X-axis depicts the time period. The plot shows a trend with spikes of low and high return values during a tenure.

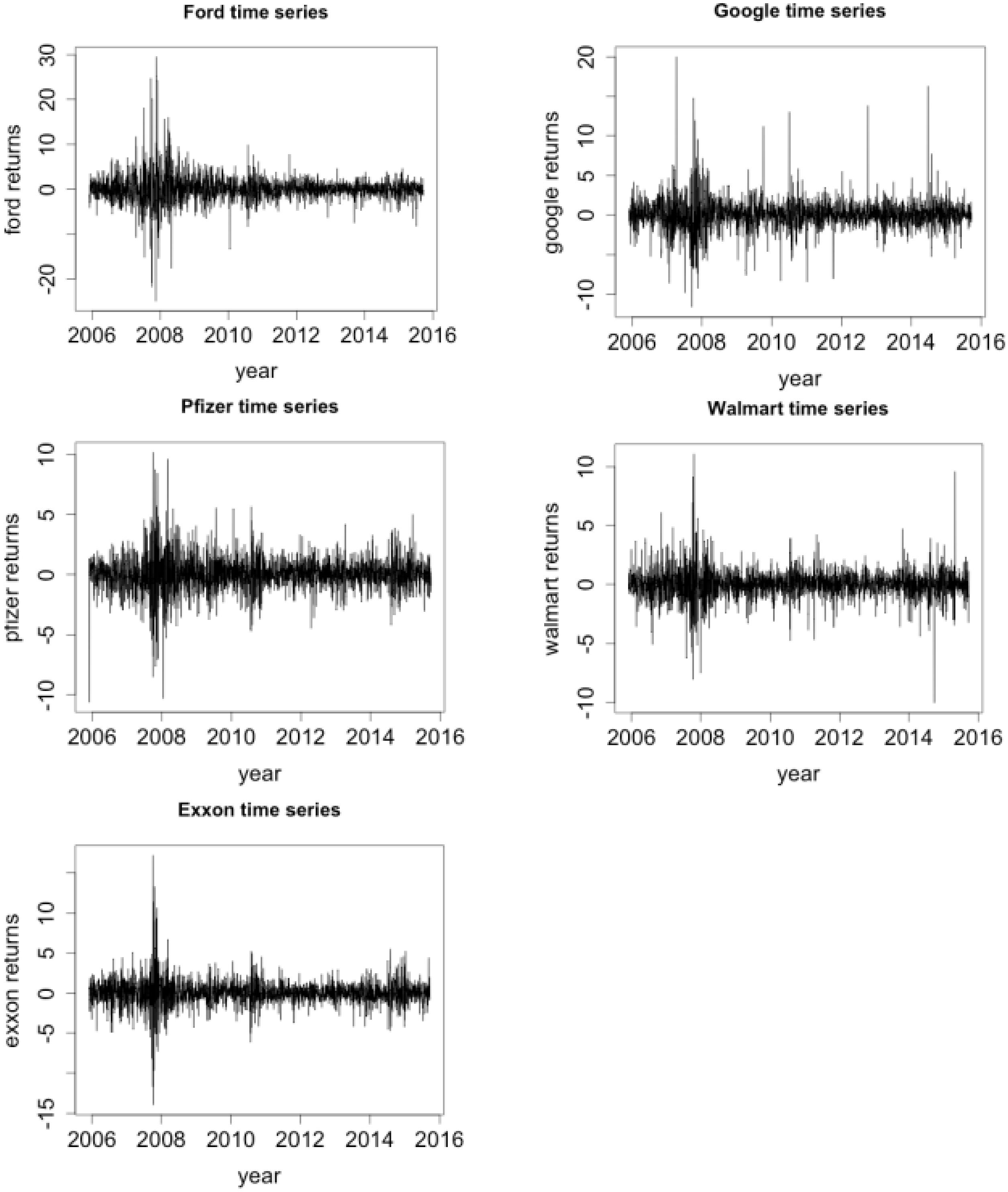


Figure 2: Time-series plot for all stocks

Analyzing the Time-series plot:

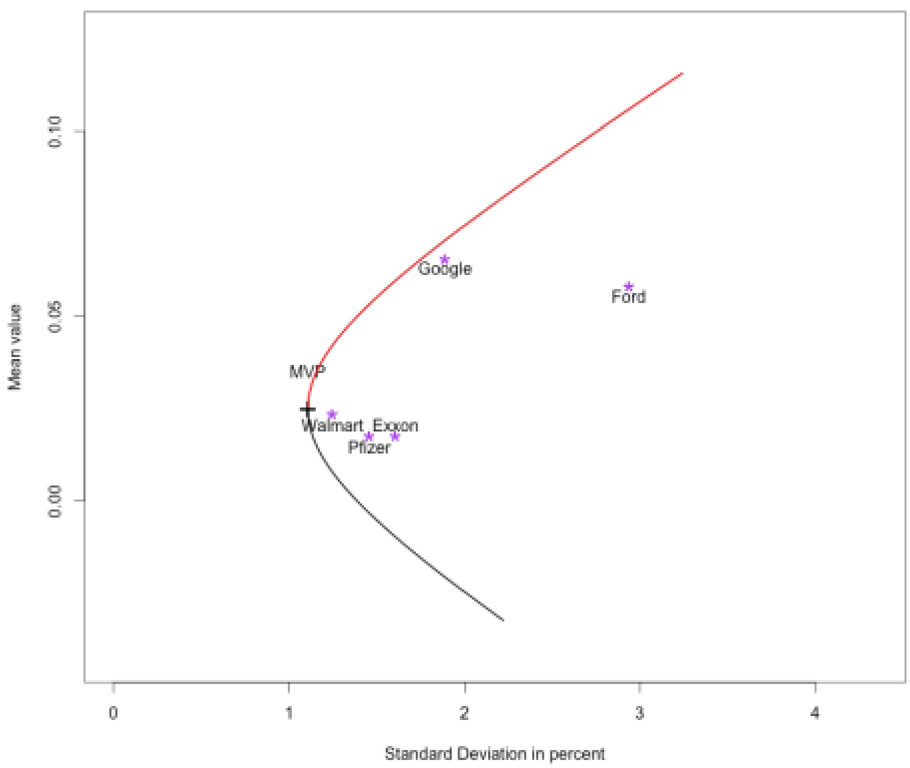
We can notice that during the period of 2008 in the time-series plot, all stocks have high and low spikes in the return value. This was the time of the 2007-2008 global recession. All stocks were affected during the recession period and hence the spikes of lows and then the recovery. Some major US banks defaulted and that had a cascading effect on the economy. Many other companies soon started defaulting and they were closed out leading to an economic crisis. After about more than a year the economy did stabilize again, and the same is seen in the time-series plot.

Efficient Frontier and various scenarios:

We now move forward to understanding the behavior of our portfolio in different scenarios. Starting from the basic scenario and considering the 5 risky assets from our portfolio we will move ahead to show the effect of a risk-free asset and correlated risky assets when short selling is allowed and other case when short selling is prohibited. These various scenarios are listed below.

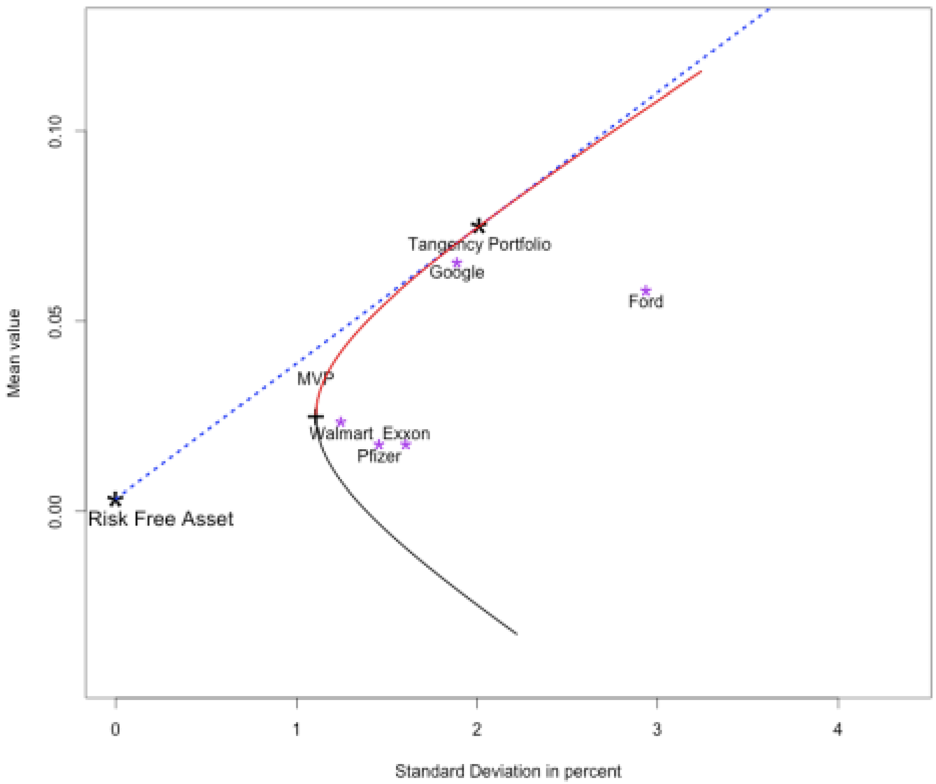
Scenario 1: Considering only risky assets:

In the first scenario that we try to explain, we consider our portfolio with the 5 stocks that we originally have. Here, the idea is to plot the efficient frontier for our 5 risky assets. For representational purpose, we plot the points of our stocks considering their mean values. We get the efficient frontier plot as follows. The plus sign (+) in our plot shows the point of minimum variance portfolio point. The X and Y axis represent the standard deviation in percentage and mean values respectively.



-Aigal Divya & Sawant Raghuraj

Scenario 2: Considering a risk-free asset and the risky assets: (Short selling is allowed)

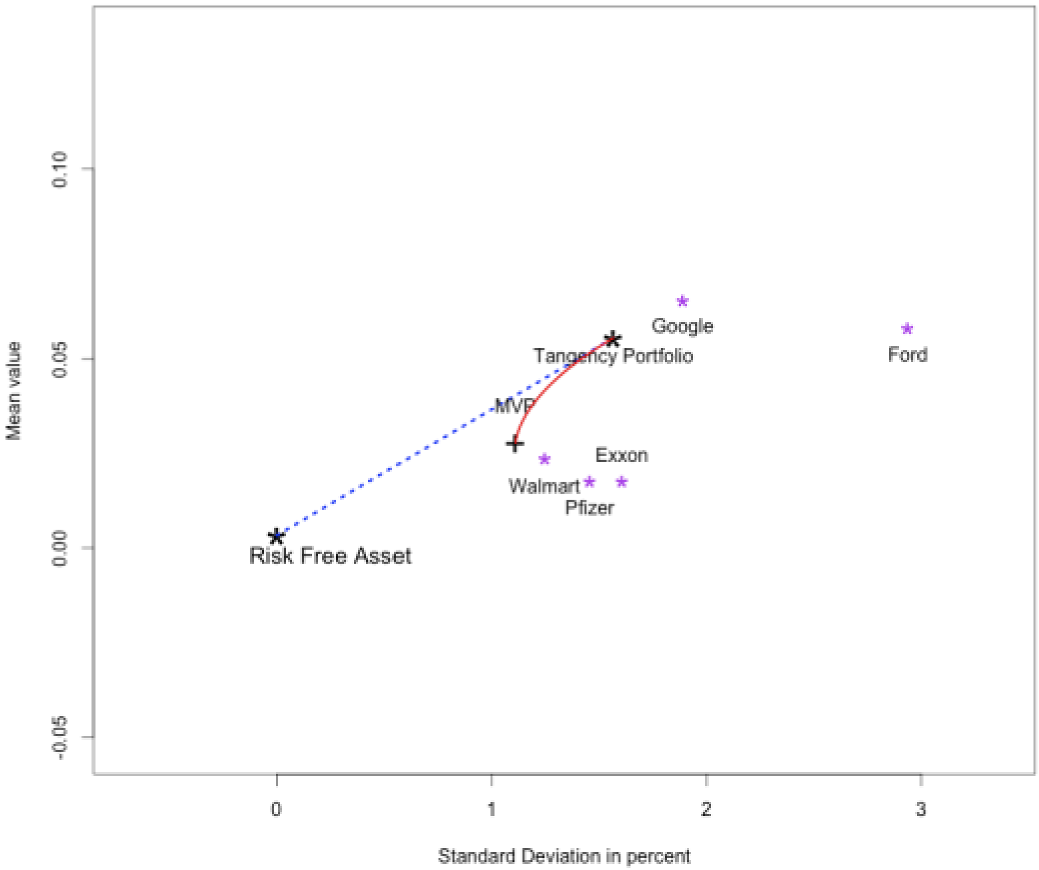


For the second scenario we combine a risk-free asset with our portfolio of risky-assets. We consider the 3 Month treasury bill as the risk free rate. The risk free rate for the 3 Month treasury bill is 0.033. The plot above shows the efficient frontier for the risky and risk-free assets. A line tangent from the risk-free asset point to the efficient frontier curve is drawn. The point of tangency is labeled on the plot. This tangency point is the one where we get the maximum returns and least risk only when we invest in our portfolio in proportion of the weights as determined by the tangency point. In other words, this is the point of maximum Sharpe ratio. In this scenario, we consider short selling, hence weights are negative. The weights for the tangency portfolio for the above plot are as follows:

Ford	**0.183**
Pfizer	**-0.248**
Walmart	**0.399**
Exxon	**-0.400**
Google	**1.066**

Scenario 3: Considering a risk-free asset and risky assets: (No Short Selling)

The other case we consider is when short selling is prohibited with risk-free and risky assets. For the no short selling case, we need to add a constraint of weights in the quadratic equation. In R, for solving the quadratic equation, we use **solve.QP** function from the **quadprog** package. Weights for no short selling scenario should be either zero or greater than 0, i.e. only those portfolios should be invested in which have positive weights. There are 4 things that differ in Scenario 3 from Scenario 2. They are the vector Amat, bvec and meq and the target expected returns value. No short sales constraint is enforced by placing an identity matrix in Amat and three-dimensional vector of zeros in bvec variable. The value for meq represents the number of constraints to be followed in our equation. Thus we use meq as 3 for solving our equation. The R program gives us the plot as below for the efficient frontier for our scenario.



From the weights we get for the tangency portfolio, we should invest only in Walmart and Google to obtain maximum returns and least risk, when we are not short selling. Also, note the values for the other stocks in the portfolio have turned to be zero. Thus we do not invest in the other 3 stocks, namely Ford, Pfizer and Exxon. The weights for the tangency portfolio for the above plot is as follows:

Ford	**0.0**
Pfizer	**0.0**
Walmart	**0.238**
Exxon	**0.0**
Google	**0.762**

Understanding relationships between stocks in the portfolio:

-Aigal Divya & Sawant Raghuraj

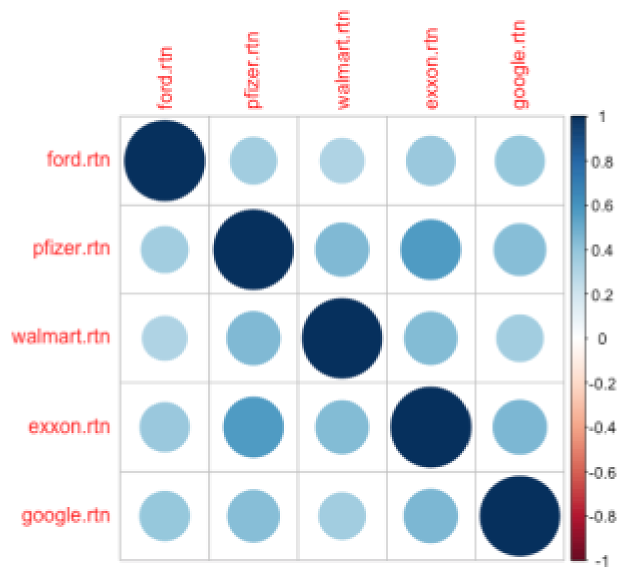


Figure 3: Correlation plot

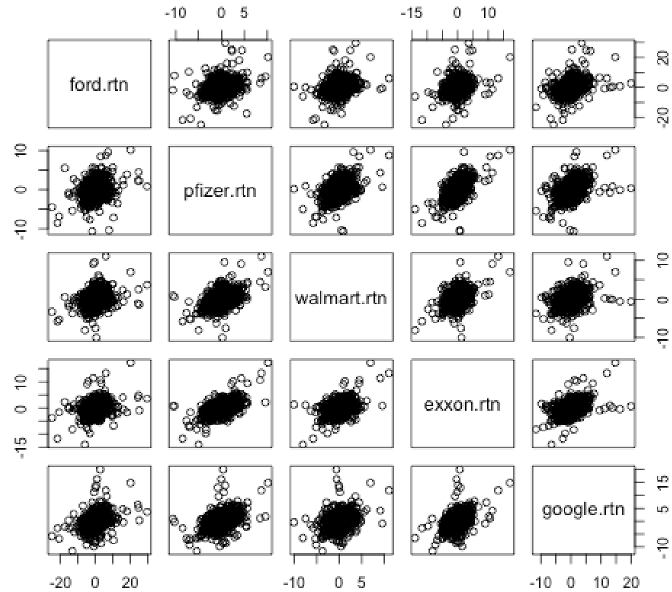


Figure 4: Pairs Scatterplot

Here, we try to understand the relationship between stocks in our portfolio. We created a correlation and pairs scatterplot for our portfolio. From the correlation plot, we can see that there is a correlation between Exxon and Pfizer. The correlation matrix shows a correlation of 0.56 for both these stocks. Also, the pairs plot helps to visualize that there is a positive correlation between both the stocks. The correlation between Pfizer and Exxon is unclear since both belong to different domains. While Pfizer is from the health care domain, Exxon is from the Oil and Gas sector. There does not seem to be much explanation between their correlation. The table below shows the values of the correlation coefficient for the different stocks and confirms that Pfizer and Exxon have 0.56 value of correlation co-efficient.

Correlation co-efficient values for the 5 stocks:

	ford.rtn	pfizer.rtn	walmart.rtn	exxon.rtn	google.rtn
ford.rtn	1	0.3408627	0.3095889	0.3747486	0.3845919
pfizer.rtn	0.3408627	1	0.4489563	0.5609928	0.4246371
walmart.rtn	0.3095889	0.4489563	1	0.4384863	0.3448595
exxon.rtn	0.3747486	0.5609928	0.4384863	1	0.4540902
google.rtn	0.3845919	0.4246371	0.3448595	0.4540902	1

Considering correlated assets in same domain:

As from the above analysis, correlation between Pfizer and Exxon is unclear, we move forward and add a stock from one of the domains already present in our portfolio. Hence, we take the stock Apple from the technology domain and couple it up with the 5 existing stocks. Now we have 6 stocks in our new portfolio. Below we will see the correlation and scatterplots for the new stock portfolio.

-Aigal Divya & Sawant Raghuraj

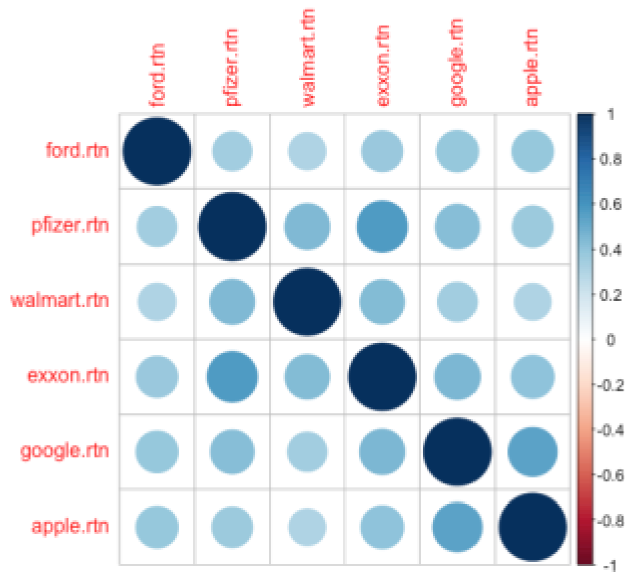


Figure 5: Correlation plot for new portfolio

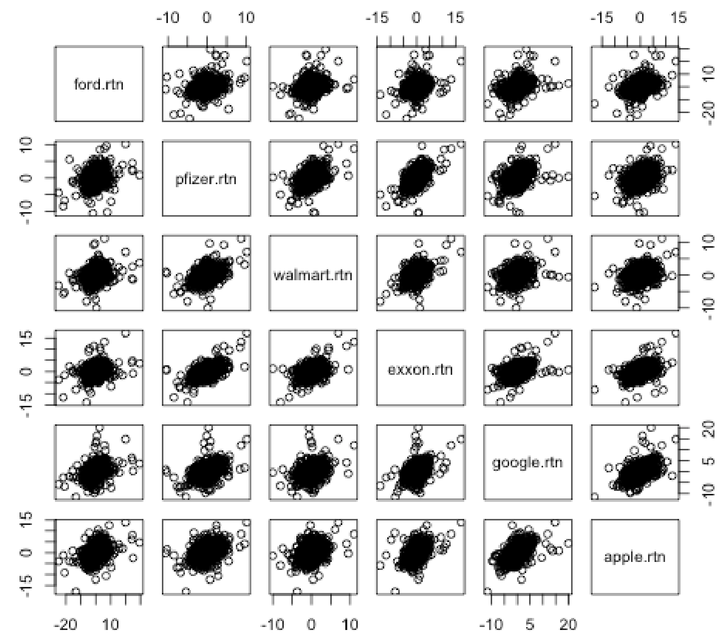


Figure 6: Scatterplot for new portfolio

The correlation plot states that along with the pair of correlated stocks; Pfizer and Exxon even Apple and Google are correlated to each other. When considering this scenario, we knew that such a correlation would be present and the various correlation plots from our analysis confirm these results. From the table below, the value of correlation between Apple and Google is 0.53. The pairs scatterplot also confirms the positive correlation between both pair of stocks. (Pfizer-Exxon and Apple-Google).

Correlation values for all the stocks:

	ford.rtn	**pfizer.rtn**	**walmart.rtn**	**exxon.rtn**	**google.rtn**	**apple.rtn**
ford.rtn	1	0.341	0.31	0.375	0.385	0.385
pfizer.rtn	0.341	1	0.449	0.561	0.425	0.368
walmart.rtn	0.31	0.449	1	0.438	0.345	0.302
exxon.rtn	0.375	0.561	0.438	1	0.454	0.408
google.rtn	0.385	0.425	0.345	0.454	1	0.53
apple.rtn	0.385	0.368	0.302	0.408	0.53	1

Scenario 4: Effect of correlated stocks in the portfolio

In this case, our motive is to find the effect of correlated stocks in our portfolio. When we analyze a portfolio and provide recommendation to a customer related to investing in a portfolio, we need to understand what effect do adding different types of stock does to a portfolio. Hence we consider this scenario and see the effect by plotting an efficient frontier for our new portfolio. The plot on the left is the efficient frontier of the original portfolio with 5 stocks. The plot on the right side is the one with 6 stocks including Apple. The plots are shown besides each other so as to make it easier in comparison.

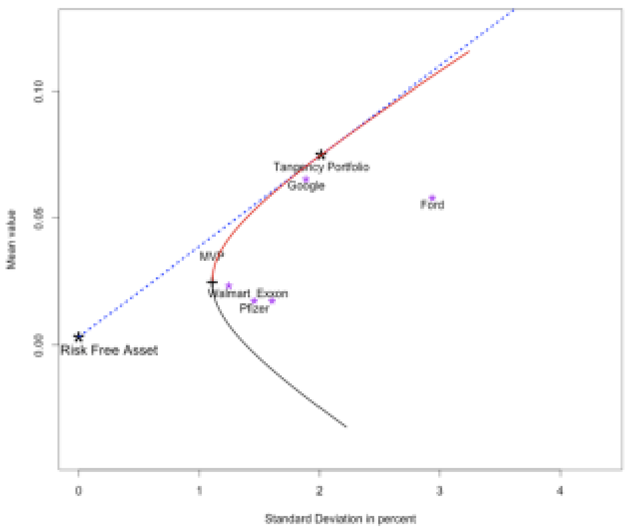


Figure 7: Original plot

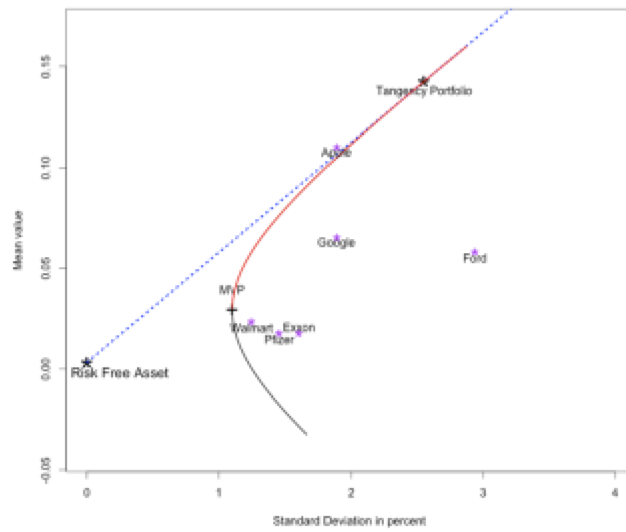


Figure 8: With 2 pair of Correlated stocks

From both the plots, we understand that after adding the stock of Apple in the portfolio, the tangency point shifted upward. The point of tangency for the original plot was mean = 0.07 and standard deviation in percent = 2.11. The point now shifted to mean = 0.143 and standard deviation in percent = 2.55. We manifest such a huge shift in the tangency point, as both Apple and Google are well performing stocks. Their mean return is the best for among the stocks in our portfolio. The mean return of Apple being 0.1101 and that of Google being 0.0655 which is higher than the other stocks in our portfolio. In this case, we have considered short selling as allowed, hence we get the following weights for the tangency portfolio:

Ford	**0.00891**
Pfizer	**-0.29124**
Walmart	**0.26221**
Exxon	**-0.48681**
Google	**0.36159**
Apple	**1.14535**

We can see from the weights that Apple gets more than proportion that is more than 100% of the investment, as there can be short selling in this case. The tangency point has shifter upwards, because both Apple and Google have positive correlation. With positive correlation, the two assets tend to move together which increases the volatility of the portfolio. Negative correlation is beneficial since it decreases risk. If the assets are negatively correlated, a negative return of one tends to occur with a positive return of the other so the volatility of the portfolio decreases. Also, we have understood that when correlation is small, then efficient portfolios have less risk for a given expected return compared to when correlation is large. As risk increases due to higher correlation, it is advisable to have diversity in the portfolio. Rather than more stocks of similar domain, we should keep correlated stocks to a minimum.

Value at Risk and Expected Shortfall of the portfolio:

Inspecting distribution of the data:

In order to work with our data, it is of utmost importance to understand your data well. For further computations, we require to check the data distribution First we check if our data is normally distributed. In order to do so, we use the Shapiro-Wilk test.
Below are the results for the Shapiro-Wilk test:

Stocks	W	p-value
Ford	0.83605	<2.2e-16
Pfizer	0.93319	<2.2e-16
Walmart	0.91056	<2.2e-16
Exxon	0.87977	<2.2e-16
Google	0.87978	<2.2e-16

From the Shapiro-Wilk test, we reject the null hypotheses that the datasets are normal, as the p-value is very small. The p-values are very small than 0.05 which is what we usually compare it with. Hence, the dataset is not normal. It is generally seen that financial data is not normally distributed.

Next, we check for t-distribution for our portfolio. To confirm the t-distribution, we use the **t.test**. Below are the results for the t.test. The table shows the t statistic, degrees of freedom and p-value for our portfolio.

Stocks	t	df	p-value
Ford	0.98771	2483	0.3234
Pfizer	1.7282	2483	0.08407
Walmart	0.60516	2483	0.5451
Exxon	0.94105	2483	0.3468
Google	0.54847	2483	0.5834

From the t.test, we fail to reject the null hypotheses that the datasets are in t-distribution, as the p-value is very large (larger than 0.05). Hence, the data for our portfolio is in t-distribution.

VaR and ES calculation:

Our investments cannot always lead into profits. When we invest in a portfolio, we do have to deal with loses at times. An investor would definitely like to make profit, but it is also good to know before-hand the maximum amount of money that you may lose in the entire transaction. The VaR (value at risk) is a bound such that the loss over a time-period is less than this bound with probability equal to the confidence coefficient ($1 - \alpha$). This means that if the confidence co-efficient is 95% and the time-period is 24 hours, then with a 95% confidence, we can predict that the maximum loss of our investment could be X amount.
Our portfolio analysis will not be complete until we calculate the VaR measure for our investment. Suppose we invest $10,000 in the portfolio (5 stocks and short selling is allowed), then value at risk for the next 24 hours when alpha = 0.05 comes out to be $254.9284. That is

with 95% confidence, the maximum loss for the portfolio is approximately $255 for a $10,000 investment. We have also calculated the VaR for the portfolio of 6 stocks (Correlated stocks Apple-Google) in our case. The VaR comes out to be $263. It is said that VaR discourages diversification. This is also seen from the calculated results of VaR for both the cases. The value increased from $255 for uncorrelated to $263 for correlated stocks.
New measures like the Expected Shortfall are also being used instead of VaR for calculating the maximum loss in an investment. For our case, the Expected shortfall value comes out to be $453.2902 for the 5 stock portfolio and $462 for 6 stock portfolio.
The following table shows the values of VaR and ES for an investment of $10,000 and at alpha = 0.05:

Portfolio composition	Value at Risk	Expected Shortfall
5 stocks, short selling allowed	$255	$453
6 stocks(Correlated), short selling allowed	$263	$462

Conclusions and Lessons learnt:

Throughout the entire project, we have learnt a lot of things. The following are the learnings and conclusions:

A) The first and foremost learning was that we need to understand the data that we are working with. Understanding the distributions for the data and then applying the appropriate formulations for the further computations is the most important learning.
B) Understanding and applying the concepts learnt in the class in our given scenarios. Some key concepts within Portfolio management like Efficient Frontier, Sharpe's Ratio, Minimum Variance Portfolio, Tangency Portfolio, how different constraints are added programmatically and their effect on the outcome.
C) By considering the various scenarios and breaking our project of portfolio management into several steps, we were able to monitor the changes and understand the results obtained in the gradual process. This iterative process showed that for every different scenario the weights on each asset changes given the constraint that we have maximum returns at minimum risk.
D) The best value for the tangency portfolio for this project was when short selling was allowed with a risk-free asset and the 5 risky asset portfolio. *Refer Scenario 2:* Considering a risk-free asset and the risky assets: (Short selling is allowed) *for details.*
E) After understanding the weights assigned for each asset, we wanted to understand that given the investment what would be the value at risk. To understand better, we checked this with two types of portfolios; namely correlated and uncorrelated and found that diversification of portfolio using the uncorrelated assets for investment would be comparatively safer than correlated assets. *Refer VaR and ES calculation: for details.*
F) From the results obtained we can say that it would be safe to have a diversified investment. Having all eggs in one basket would be unsafe, should the invested sector not perform well.
G) We also came across several investment options for common people with investments as low as $5 and monthly fees as low as $1. These applications advise an investor to diversify based on the preferences of the sectors. Examples being, Stash, Robinhood etc.

Team work and collaboration:

A) **Project Selection**: Divya contributed with two topics in sports and finance domain and Raghuraj suggested one topic in Real Estate and finance. After discussing with Prof. K.C. Chang we decided to consider portfolio management.

B) **Project Objectives**: Divya did research on the sources and some candidate stocks for our project and Raghuraj suggested the idea of breaking the project into different scenarios. Divya also suggested to consider finding the risk of the investment.

C) **Data Exploration**: Divya and Raghuraj individually started their efforts to explore the data, transform the data from stocks, visualize the data.

D) **Programming**: While major part of coding was done by Divya along with some key research on constraints, Raghuraj looked at reverse engineering the portfolio management code provided by Prof. and understanding what each component in the code did along with research from websites. [4][5]

E) **Results**: Divya and Raghuraj worked together to understand the output of each scenario and interpreted the results.

F) **Presentation and Paper**: The detailed aspects of the project report were provided by Divya, while Raghuraj looked at the presentation slides and researched about the applications in line with the project. [3]

Computation

```
#Loading all packages
library(corrplot)
library(MASS)
library(quadprog)
library(fBasics)
library(quantmod)

# read the data
# Ford data
ford_data <- read.csv("/Users/divya/OneDrive/Data
Analytics/Fall 2016/OR
538/Project/Data/F_HistoricalQuotes.csv")
# Google data
google_data <-
read.csv("/Users/divya/OneDrive/Data Analytics/Fall
2016/OR
538/Project/Data/GOOGL_HistoricalQuotes.csv")
# Pfizer data
pfizer_data <-
read.csv("/Users/divya/OneDrive/Data Analytics/Fall
2016/OR
538/Project/Data/PFE_HistoricalQuotes.csv")
# Walmart data
walmart_data <-
read.csv("/Users/divya/OneDrive/Data Analytics/Fall
2016/OR
538/Project/Data/WMT_HistoricalQuotes.csv")
# Exxon data
exxon_data <-
read.csv("/Users/divya/OneDrive/Data Analytics/Fall
2016/OR
538/Project/Data/XOM_HistoricalQuotes.csv")
# Apple data
apple_data <- read.csv("/Users/divya/OneDrive/Data
Analytics/Fall 2016/OR
538/Project/Data/AAPL_HistoricalQuotes.csv")

#calculate the returns

n = length(ford_data$close[-1])
# converting to percentages
ford.rtn =  100 * (ford_data$close[-1][2:n ] /
ford_data$close[-1][1:(n-1) ] - 1)
google.rtn = 100 * (google_data$close[-1][2:n ] /
google_data$close[-1][1:(n-1) ] - 1)
pfizer.rtn = 100 * (pfizer_data$close[-1][2:n ] /
pfizer_data$close[-1][1:(n-1) ] - 1)
walmart.rtn = 100 * (walmart_data$close[-1][2:n ] /
walmart_data$close[-1][1:(n-1) ] - 1)
exxon.rtn = 100 * (exxon_data$close[-1][2:n ] /
exxon_data$close[-1][1:(n-1) ] - 1)
apple.rtn = 100 * (apple_data$close[-1][2:n ] /
apple_data$close[-1][1:(n-1) ] - 1)
```

```
#plot the returns
plot(ford.rtn)
plot(google.rtn)
plot(pfizer.rtn)
plot(walmart.rtn)
plot(exxon.rtn)
plot(apple.rtn)

#Time Series plots
year = seq(2005+231/253,2015+184/253,1/253)
length(year) #2484

# Plots for all the 5 stocks
plot(year,ford.rtn,ylab="ford returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Ford time series")
plot(year,google.rtn,ylab="google returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Google time series")
plot(year,pfizer.rtn,ylab="pfizer returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Pfizer time series")
plot(year,walmart.rtn,ylab="walmart returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Walmart time series")
plot(year,exxon.rtn,ylab="exxon returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Exxon time series")
plot(year,apple.rtn,ylab="apple returns",type="l",xlab="year",cex.lab=1.5,
    cex.axis=1.5,cex.main=1.3, main= "Apple time series")

#Binding all 5 stock return data
stocksData = cbind(ford.rtn, pfizer.rtn, walmart.rtn, exxon.rtn, google.rtn)

#Binding stock return data also considering Apple
stocksData_6 = cbind(ford.rtn, pfizer.rtn, walmart.rtn, exxon.rtn,  google.rtn, apple.rtn)

cor(stocksData)
cov(stocksData)
corrplot(cor(stocksData))
pairs(stocksData)

#Plotting the efficient frontier for the risky assets only i.e.the stocks

mean_vect = apply(stocksData,2,mean)
cov_mat = cov(stocksData)
sd_vect = sqrt(diag(cov_mat))
M = length(mean_vect)
Amat = cbind(rep(1,M),mean_vect)  # set the constraints matrix

muP = seq(min(mean_vect)-0.05,max(mean_vect)+0.05,length=300)  # set of 300 possible target values
# for the expect portfolio return1
sdP = muP # set up storage for standard deviations of portfolio returns
weights = matrix(0,nrow=300,ncol=M) # storage for portfolio weights

# for each expected return (constraint), find the minimum risk/variance portfolio using
# quadratic programming
for (i in 1:length(muP))  # find the optimal portfolios for each target expected return
{
  bvec = c(1,muP[i])  # constraint vector
  result =
   solve.QP(Dmat=2*cov_mat, dvec=rep(0,M), Amat=Amat, bvec=bvec, meq=2)
  sdP[i] = sqrt(result$value)
  weights[i,] = result$solution
}

plot(sdP,muP,type="l",xlim=c(min(sdP)-1.1,max(sdP)+1.1),ylim=c(min(muP)-0.01,max(muP)+0.01),
    xlab = "Standard Deviation in percent", ylab = "Mean value", lwd= 2,lty=1)  #  plot
# the efficient frontier (and inefficient frontier)
sharpe = muP/sdP # compute Sharpe ratios
ind = (sharpe == max(sharpe)) # Find maximum Sharpe ratio
options(digits=3)
weights[ind,] # Find tangency portfolio# show line of optimal portfolios
# show line of optimal portfolios
ind2 = (sdP == min(sdP)) # find the minimum variance portfolio
points(sdP[ind2],muP[ind2],cex=2,pch="+") # show minimum variance portfolio
text(sdP[ind2], muP[ind2]+0.010,'MVP', cex=1)  # plot the minimum variance portfolio
ind3 = (muP > muP[ind2])
lines(sdP[ind3],muP[ind3],type="l",xlim=c(min(sdP)-1.1,max(sdP)+1.1),
```

```
    ylim=c(min(muP)-0.01,max(muP)+0.01),col=
"red",lwd=2)  #  plot the efficient frontier
text(sd_vect[1],mean_vect[1]-0.003,"Ford",cex=1)
points(sd_vect[1],mean_vect[1],cex=2,col =
"purple",pch="*")
text(sd_vect[2],mean_vect[2]-0.003,"Pfizer",cex=1)
points(sd_vect[2],mean_vect[2],cex=2,col =
"purple",pch="*")
text(sd_vect[3],mean_vect[3]-
0.003,"Walmart",cex=1)
points(sd_vect[3],mean_vect[3],cex=2,col =
"purple",pch="*")
text(sd_vect[4],mean_vect[4]+0.003,"Exxon",cex=1)
points(sd_vect[4],mean_vect[4],cex=2,col =
"purple",pch="*")
text(sd_vect[5],mean_vect[5]-0.003,"Google",cex=1)
points(sd_vect[5],mean_vect[5],cex=2,col =
"purple",pch="*")

####################################
#Plotting the efficient frontier for short selling
#Short selling is allowed
mean_vect = apply(stocksData,2,mean)
cov_mat = cov(stocksData)
sd_vect = sqrt(diag(cov_mat))
M = length(mean_vect)
Amat = cbind(rep(1,M),mean_vect)  # set the
constraints matrix

muP = seq(min(mean_vect)-
0.05,max(mean_vect)+0.05,length=300)  # set of 300
possible target values
# for the expect portfolio return
sdP = muP # set up storage for standard deviations
of portfolio returns
weights = matrix(0,nrow=300,ncol=M) # storage for
portfolio weights

# for each expected return (constraint), find the
minimum risk/variance portfolio using
# quadratic programming
for (i in 1:length(muP))  # find the optimal portfolios
for each target expected return
{
  bvec = c(1,muP[i])  # constraint vector
  result =
    solve.QP(Dmat=2*cov_mat, dvec=rep(0,M),
Amat=Amat, bvec=bvec, meq=2)
  sdP[i] = sqrt(result$value)
  weights[i,] = result$solution
}

plot(sdP,muP,type="l",xlim=c(min(sdP)-
1.1,max(sdP)+1.1),ylim=c(min(muP)-
0.01,max(muP)+0.01),
    xlab = "Standard Deviation in percent", ylab =
"Mean value", lwd= 2,lty=1)  #  plot
# the efficient frontier (and inefficient frontier)
mufree = 0.0033 # input value of risk-free interest
rate = 3 Month treasury bill
points(0,mufree,cex=3,pch="*")  # show risk-free
asset
text(0.25, mufree-0.005,'Risk Free Asset', cex=1.2)  #
plot the Risk Free Asset
sharpe =( muP-mufree)/sdP # compute Sharpe ratios
ind = (sharpe == max(sharpe)) # Find maximum
Sharpe ratio
options(digits=3)
weights[ind,] # Find tangency portfolio# show line of
optimal portfolios
lines(c(0,max(sdP)+2),mufree+c(0,max(sdP)+2)*(mu
P[ind]-mufree)/sdP[ind],col="blue",lwd=3,lty=3)
# show line of optimal portfolios
points(sdP[ind],muP[ind],cex=3,pch="*") # show
tangency portfolio
text(sdP[ind],muP[ind]-0.005,'Tangency Portfolio',
cex=1)  # plot the tangency portfolio
ind2 = (sdP == min(sdP)) # find the minimum
variance portfolio
points(sdP[ind2],muP[ind2],cex=2,pch="+") # show
minimum variance portfolio
text(sdP[ind2], muP[ind2]+0.010,'MVP', cex=1)  #
plot the minimum variance portfolio
ind3 = (muP > muP[ind2])
lines(sdP[ind3],muP[ind3],type="l",xlim=c(min(sdP)-
1.1,max(sdP)+1.1),
     ylim=c(min(muP)-0.01,max(muP)+0.01),col=
"red",lwd=2)  #  plot the efficient frontier
text(sd_vect[1],mean_vect[1]-0.003,"Ford",cex=1)
points(sd_vect[1],mean_vect[1],cex=2,col =
"purple",pch="*")
text(sd_vect[2],mean_vect[2]-0.003,"Pfizer",cex=1)
points(sd_vect[2],mean_vect[2],cex=2,col =
"purple",pch="*")
text(sd_vect[3],mean_vect[3]-
0.003,"Walmart",cex=1)
points(sd_vect[3],mean_vect[3],cex=2,col =
"purple",pch="*")
text(sd_vect[4],mean_vect[4]+0.003,"Exxon",cex=1)
points(sd_vect[4],mean_vect[4],cex=2,col =
"purple",pch="*")
text(sd_vect[5],mean_vect[5]-0.003,"Google",cex=1)
points(sd_vect[5],mean_vect[5],cex=2,col =
"purple",pch="*")

####################################
```

```
#Plotting the efficient frontier for NO short selling,
w>=0
mean_vect_no = apply(stocksData,2,mean)
cov_mat_no = cov(stocksData)
sd_vect_no = sqrt(diag(cov_mat_no))
M = length(mean_vect_no)
Amat_no = cbind(rep(1,M),mean_vect_no,
diag(1,nrow=M))  # set the constraints matrix
muP_no =
seq(min(mean_vect_no)+0.01,max(mean_vect_no)-
0.01 ,length=300)  # set of 300 possible target values
# for the expect portfolio return1
sdP_no = muP_no # set up storage for standard
deviations of portfolio returns
weights = matrix(0,nrow=300,ncol=M) # storage for
portfolio weights

# for each expected return (constraint), find the
minimum risk/variance portfolio using
# quadratic programming
for (i in 1:length(muP_no))  # find the optimal
portfolios for each target expected return
{
  bvec_no = c(1, muP_no[i], rep(0,M)) # constraint
thresholds: no short selling, min weights are zero
  result =
    solve.QP(Dmat=2*cov_mat_no, dvec=rep(0,M),
Amat=Amat_no, bvec=bvec_no, meq=3)
  sdP_no[i] = sqrt(result$value)
  weights[i,] = result$solution
}

plot(sdP_no,muP_no,type="l",xlim=c(min(sdP_no)-
1.8,max(sdP_no)+1.8),ylim=c(min(muP_no)-
0.08,max(muP_no)+0.08),
     xlab = "Standard Deviation in percent", ylab =
"Mean value", lwd= 2,lty=1)  #  plot
# the efficient frontier (and inefficient frontier)
mufree = 0.0033  # input value of risk-free interest
rate = 3 Month treasury bill
points(0,mufree,cex=3,pch="*")  # show risk-free
asset
text(0.25, mufree-0.005,'Risk Free Asset', cex=1.2)  #
plot the Risk Free Asset
sharpe = (muP_no-mufree)/sdP_no # compute
Sharpe ratios
ind = (sharpe == max(sharpe)) # Find maximum
Sharpe ratio
options(digits=3)
weights[ind,] # Find tangency portfolio# show line of
optimal portfolios
lines(c(0,sdP_no[ind]),c(mufree,muP_no[ind]),col="b
lue",lwd=3,lty=3)
# show line of optimal portfolios
points(sdP_no[ind],muP_no[ind],cex=3,pch="*") #
show tangency portfolio
text(sdP_no[ind],muP_no[ind]-0.005,'Tangency
Portfolio', cex=1)  # plot the tangency portfolio
ind2 = (sdP_no == min(sdP_no)) # find the minimum
variance portfolio
points(sdP_no[ind2],muP_no[ind2],cex=2,pch="+") #
show minimum variance portfolio
text(sdP_no[ind2], muP_no[ind2]+0.010,'MVP',
cex=1)  # plot the minimum variance portfolio
ind3 = (muP_no > muP_no[ind2])
lines(sdP_no[ind3],muP_no[ind3],type="l",xlim=c(mi
n(sdP_no)-1.8,max(sdP_no)+1.8),
      ylim=c(min(muP_no)-
0.02,max(muP_no)+0.02),col= "red",lwd=2)  #  plot
the efficient frontier
text(sd_vect_no[1],mean_vect_no[1]-
0.007,"Ford",cex=1)
points(sd_vect_no[1],mean_vect_no[1],cex=2,col =
"purple",pch="*")
text(sd_vect_no[2],mean_vect_no[2]-
0.007,"Pfizer",cex=1)
points(sd_vect_no[2],mean_vect_no[2],cex=2,col =
"purple",pch="*")
text(sd_vect_no[3],mean_vect_no[3]-
0.007,"Walmart",cex=1)
points(sd_vect_no[3],mean_vect_no[3],cex=2,col =
"purple",pch="*")
text(sd_vect_no[4],mean_vect_no[4]+0.007,"Exxon"
,cex=1)
points(sd_vect_no[4],mean_vect_no[4],cex=2,col =
"purple",pch="*")
text(sd_vect_no[5],mean_vect_no[5]-
0.007,"Google",cex=1)
points(sd_vect_no[5],mean_vect_no[5],cex=2,col =
"purple",pch="*")

####################################
#When we have correlated stocks:
#Exxon and Pfizer and Google and Apple

cor(stocksData_6)
cov(stocksData_6)
corrplot(cor(stocksData_6))
pairs(stocksData_6)

#Short selling is allowed
mean_vect = apply(stocksData_6,2,mean)
cov_mat = cov(stocksData_6)
sd_vect = sqrt(diag(cov_mat))
M = length(mean_vect)
Amat = cbind(rep(1,M),mean_vect)  # set the
```

```
constraints matrix

muP = seq(min(mean_vect)-
0.05,max(mean_vect)+0.05,length=300)  # set of 300
possible target values
# for the expect portfolio return1
sdP = muP # set up storage for standard deviations
of portfolio returns
weights = matrix(0,nrow=300,ncol=M) # storage for
portfolio weights

# for each expected return (constraint), find the
minimum risk/variance portfolio using
# quadratic programming
for (i in 1:length(muP))  # find the optimal portfolios
for each target expected return
{
  bvec = c(1,muP[i])  # constraint vector
  result =
    solve.QP(Dmat=2*cov_mat, dvec=rep(0,M),
Amat=Amat, bvec=bvec, meq=2)
  sdP[i] = sqrt(result$value)
  weights[i,] = result$solution
}

plot(sdP,muP,type="l",xlim=c(min(sdP)-
1.1,max(sdP)+1.1),ylim=c(min(muP)-
0.01,max(muP)+0.01),
     xlab = "Standard Deviation in percent", ylab =
"Mean value", lwd= 2,lty=1)  #  plot
# the efficient frontier (and inefficient frontier)
mufree = 0.0033  # input value of risk-free interest
rate = 3 Month treasury bill
points(0,mufree,cex=3,pch="*")  # show risk-free
asset
text(0.25, mufree-0.005,'Risk Free Asset', cex=1.2)  #
plot the Risk Free Asset
sharpe =( muP-mufree)/sdP # compute Sharpe ratios
ind = (sharpe == max(sharpe)) # Find maximum
Sharpe ratio
options(digits=3)
weights[ind,] # Find tangency portfolio# show line of
optimal portfolios
lines(c(0,max(sdP)+2),mufree+c(0,max(sdP)+2)*(mu
P[ind]-mufree)/sdP[ind],col="blue",lwd=3,lty=3)
# show line of optimal portfolios
points(sdP[ind],muP[ind],cex=3,pch="*") # show
tangency portfolio
text(sdP[ind],muP[ind]-0.005,'Tangency Portfolio',
cex=1)  # plot the tangency portfolio
ind2 = (sdP == min(sdP)) # find the minimum
variance portfolio
points(sdP[ind2],muP[ind2],cex=2,pch="+") # show
minimum variance portfolio
text(sdP[ind2], muP[ind2]+0.010,'MVP', cex=1)  #
plot the minimum variance portfolio
ind3 = (muP > muP[ind2])
lines(sdP[ind3],muP[ind3],type="l",xlim=c(min(sdP)-
1.1,max(sdP)+1.1),
     ylim=c(min(muP)-0.01,max(muP)+0.01),col=
"red",lwd=2)  #  plot the efficient frontier
text(sd_vect[1],mean_vect[1]-0.003,"Ford",cex=1)
points(sd_vect[1],mean_vect[1],cex=2,col =
"purple",pch="*")
text(sd_vect[2],mean_vect[2]-0.003,"Pfizer",cex=1)
points(sd_vect[2],mean_vect[2],cex=2,col =
"purple",pch="*")
text(sd_vect[3],mean_vect[3]-
0.003,"Walmart",cex=1)
points(sd_vect[3],mean_vect[3],cex=2,col =
"purple",pch="*")
text(sd_vect[4],mean_vect[4]+0.003,"Exxon",cex=1)
points(sd_vect[4],mean_vect[4],cex=2,col =
"purple",pch="*")
text(sd_vect[5],mean_vect[5]-0.003,"Google",cex=1)
points(sd_vect[5],mean_vect[5],cex=2,col =
"purple",pch="*")
text(sd_vect[5],mean_vect[6]-0.003,"Apple",cex=1)
points(sd_vect[5],mean_vect[6],cex=2,col =
"purple",pch="*")

####################################
# Check if the data is normal by using the shapiro
wilk test
shapiro.test(ford.rtn) #W = 0.83605, p-value < 2.2e-
16
shapiro.test(google.rtn) #W = 0.87978, p-value <
2.2e-16
shapiro.test(pfizer.rtn) #W = 0.93319, p-value < 2.2e-
16
shapiro.test(walmart.rtn) #W = 0.91056, p-value <
2.2e-16
shapiro.test(exxon.rtn) #W = 0.87977, p-value <
2.2e-16
shapiro.test(apple.rtn) #W = 0.94347, p-value < 2.2e-
16
# Data is not normal

# Now we proceed to check for t-distribution
t.test(ford.rtn) # t = 0.98771, df = 2483, p-value =
0.3234
t.test(google.rtn) # t = 1.7282, df = 2483, p-value =
0.08407
t.test(pfizer.rtn) # t = 0.60516, df = 2483, p-value =
0.5451
t.test(walmart.rtn) # t = 0.94105, df = 2483, p-value
```

```
= 0.3468
t.test(exxon.rtn) # t = 0.54847, df = 2483, p-value =
0.5834
t.test(apple.rtn) # t = 2.6256, df = 2483, p-value =
0.008702

####################################
#Calculating Value at Risk and Expected Shortfall

fitt = fitdistr(stocksData/100,"t") # Dividing by 100
because earlier we had multiplied the returns by 100
param = as.numeric(fitt$estimate)
alpha = 0.05
mean = param[1]
df = param[3]
sd = param[2] * sqrt((df)/(df-2))
lambda = param[2]  #  scale parameter (STD)
qalphat = qt(alpha,df=df)
VaR_part = -10000*(mean + lambda*qalphat)
IEVaR = (stocksData < qalphat)
# Expected shortfall
es1 = dt(qalphat,df=df)/(alpha)
es2 = (df + qalphat^2) / (df - 1)
es3 = -mean+lambda*es1*es2
ES_part = 10000*es3
VaR_part # 254.9284
ES_part # 453.2902

fitt = fitdistr(stocksData_6/100,"t") # Dividing by 100
because earlier we had multiplied the returns by 100
param = as.numeric(fitt$estimate)
alpha = 0.05
mean = param[1]
df = param[3]
sd = param[2] * sqrt((df)/(df-2))
lambda = param[2]  #  scale parameter (STD)
qalphat = qt(alpha,df=df)
VaR_part = -10000*(mean + lambda*qalphat)
IEVaR = (stocksData_6 < qalphat)
# Expected shortfall
es1 = dt(qalphat,df=df)/(alpha)
es2 = (df + qalphat^2) / (df - 1)
es3 = -mean+lambda*es1*es2
ES_part = 10000*es3
VaR_part # 263
ES_part # 462
```

References:

1. Ruppert, D. (2011). Statistics and data analysis for financial engineering. New York: Springer.

2. Professor K.C Chang all notes and charts

3. https://www.stashinvest.com/

4. http://www.wdiam.com/2012/06/10/mean-variance-portfolio-optimization-with-r-and-quadratic-programming/

5. http://economistatlarge.com/portfolio-theory/r-optimized-portfolio