Portfolio Optimization

## Part 1

In this project, we are going to form the optimal portfolio using the Sharpe ratio with historical and examine its performance over a test period. First of all, we load some packages and perform some data cleaning:

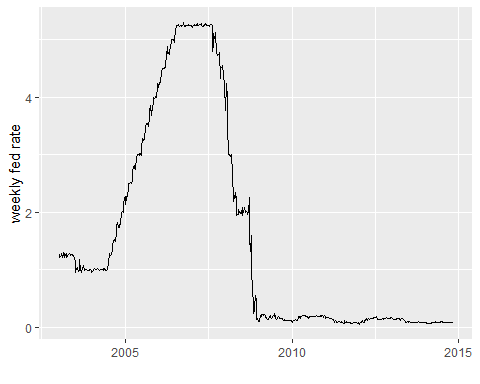
library(ggplot2)  
library(tidyverse)  
library(lubridate)  
  
# part 1  
# load the data  
data.x <- read.csv2('asset\_data.txt', sep = ',', stringsAsFactors=F)  
# extract only theobservations where the federal funds rate is available  
data.x <- na.omit(data.x)  
# convert datatypes  
data.x$date <- as.Date(data.x$date)  
data.x$close.spy <- as.numeric(data.x$close.spy)  
data.x$close.tlt <- as.numeric(data.x$close.tlt)  
data.x$fed.rate <- as.numeric(data.x$fed.rate)  
# check head and tail  
head(data.x)

## date close.spy close.tlt fed.rate  
## 5 2003-01-08 91.39 86.99 1.20  
## 10 2003-01-15 92.40 86.40 1.26  
## 14 2003-01-22 88.17 87.88 1.23  
## 19 2003-01-29 86.48 87.17 1.24  
## 24 2003-02-05 84.85 87.51 1.29  
## 29 2003-02-12 82.10 88.09 1.22

tail(data.x)

## date close.spy close.tlt fed.rate  
## 2951 2014-09-24 199.56 114.81 0.09  
## 2956 2014-10-01 194.35 118.23 0.09  
## 2961 2014-10-08 196.64 119.39 0.09  
## 2966 2014-10-15 186.43 122.53 0.09  
## 2971 2014-10-22 192.69 120.67 0.09  
## 2976 2014-10-29 198.11 119.46 0.09

# graph the federal funds interest rate as a time series  
ggplot(data.x, aes(date, fed.rate)) + geom\_line() +  
 xlab("") + ylab("weekly fed rate")



We find that:

1. the time series starts on 2003-01-08 and ends on 2014-10-29;
2. the federal funds interest rates were around 1% before 2005, increased to over 5% in 2007, and decreades sharply to zero after the most recent financial crisis.

## Part 2

We split the data into training set and test set based on whether they are before or in 2014:

# part 2  
# generate a column for year  
data.x <- as\_tibble(data.x)  
data.x$year <- year(data.x$date)  
# train set: all obs before 2014  
train\_set <- subset(data.x, year<2014)  
nrow(train\_set)

## [1] 570

# test set: all obs in 2014  
test\_set <- subset(data.x, year==2014)  
nrow(test\_set)

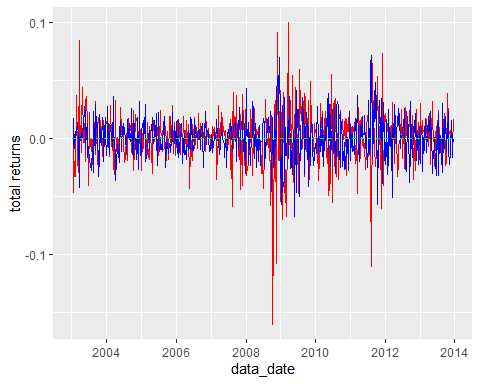
## [1] 43

We find that there are 570 observations in the training set and 43 observations in the test set.

## Part 3

We convert the federal rate to decimal and compute the returns for both assets:

### from now on, work on train set###  
# part 3  
# fed rate to decimal  
train\_set$fed.rate <- train\_set$fed.rate / 100  
# SP500 returns  
train\_set$l\_sp500 <- c(NA, head(train\_set$close.spy, -1))  
train\_set$r\_sp500 <- train\_set$close.spy / train\_set$l\_sp500 - 1  
# bonds returns  
train\_set$l\_bonds <- c(NA, head(train\_set$close.tlt, -1))  
train\_set$r\_bonds <- train\_set$close.tlt / train\_set$l\_bonds - 1  
# y = 0  
train\_set$zeros <- 0  
# plot all out  
ggplot() +   
 geom\_line(data = train\_set, aes(x = date, y = r\_sp500), color = "red") +  
 geom\_line(data = train\_set, aes(x = date, y = r\_bonds), color = "blue") +  
 geom\_line(data = train\_set, aes(x = date, y = zeros), color = "green", linetype = "dashed") +  
 xlab('data\_date') + ylab('total returns')



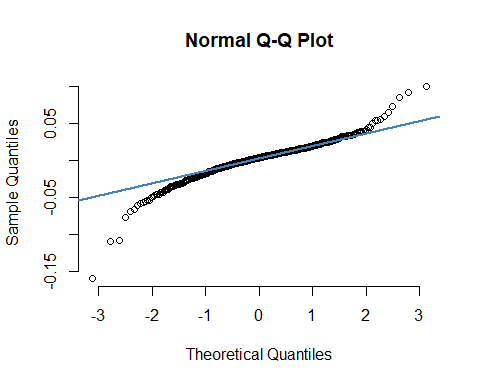
The red line is for SP500 and bule for tresury bonds. We can see that:

1. they are both fluctuate around zero line;
2. they seem to be negatively related: one gets peak while the other tends to be at the bottom.

## Part 4

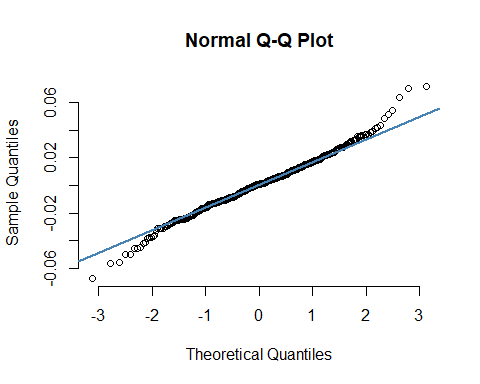
The QQ plots for SP500 is:

# part 4  
# normal quantile plots  
# for sp500  
qqnorm(train\_set$r\_sp500, pch = 1, frame = FALSE)  
qqline(train\_set$r\_sp500, col = "steelblue", lwd = 2)



For bonds:

# for bonds  
qqnorm(train\_set$r\_bonds, pch = 1, frame = FALSE)  
qqline(train\_set$r\_bonds, col = "steelblue", lwd = 2)



From these two figures we can see that the normality assumption seems to be satisified for both series.

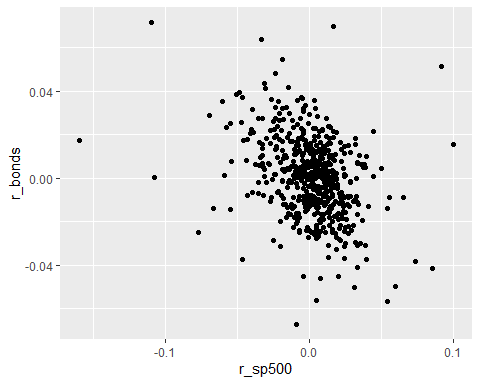
## Part 5

The correlation between two asset returns over the whole train set is -0.3439, and the scatter plot also indicate a negative relationship between these two:

# part 5  
# correlation all time  
cor(train\_set$r\_sp500, train\_set$r\_bonds, use = "complete.obs")

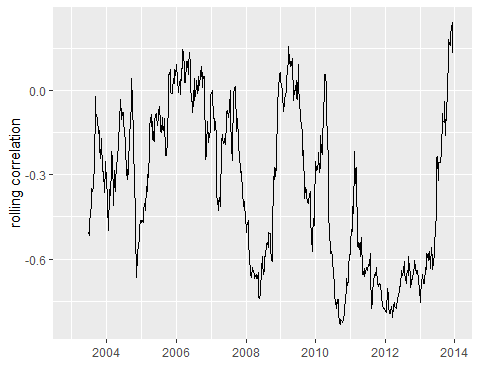
## [1] -0.3439013

# scatterplot  
ggplot() + geom\_point(data = train\_set, aes(x = r\_sp500, y = r\_bonds)) +  
 xlab('r\_sp500') + ylab('r\_bonds')



The rolling-window correlation is calculated as follows:

# correlation rolling window  
train\_set$rolling\_cor <- NA  
window\_length <- 24  
# using the last day of the window  
for (i in (1+window\_length):nrow(train\_set)){  
 sp500 <- train\_set$r\_sp500[(i-window\_length+1):i]  
 bonds <- train\_set$r\_bonds[(i-window\_length+1):i]  
 res <- cor(sp500, bonds)  
 train\_set$rolling\_cor[i] <- res  
}  
# plot it out  
ggplot(train\_set, aes(date, rolling\_cor)) + geom\_line() +  
 xlab("") + ylab("rolling correlation")



The rolling-window correlation is a better way to describe the relationship between two assets. In different economic conditions, the relationship between stock and bond can also vary: for xample, in the financial crisis in 2008, the stock price decreased and people tended to buy more tresury bonds, which resulted in a larger negative rolling correlation between these two assets. On the other hand, the overall correlation describes the average relationship between two assets across the sample period, and hence is less informative.

## Part 6

The Sharpe ratios for SP500 are calculated step-by-step:

# part 6  
# first, sp500  
# step 0  
# step 1  
train\_set$l\_frate <- c(NA, head(train\_set$fed.rate, -1))  
train\_set$e\_t <- train\_set$r\_sp500 - train\_set$l\_frate/52  
# step 2  
train\_set$g\_t <- NA  
train\_set$g\_t[1] <- 100  
for (i in 2:nrow(train\_set)){  
 train\_set$g\_t[i] <- train\_set$g\_t[i-1] \* (1 + train\_set$e\_t[i])  
}  
# step 3  
n\_years <- (nrow(train\_set)-1) / 52  
# step 4  
CAGR <- (train\_set$g\_t[nrow(train\_set)]/train\_set$g\_t[1])^(1/n\_years) - 1  
# step 5  
nu <- 52^0.5 \* sd(train\_set$e\_t, na.rm = TRUE)  
# step 6  
SR <- CAGR / nu  
SR

## [1] 0.2807176

We repeat it for bonds:

# second, bonds  
train\_set$e\_t <- train\_set$r\_bonds - train\_set$l\_frate/52  
# step 2  
train\_set$g\_t <- NA  
train\_set$g\_t[1] <- 100  
for (i in 2:nrow(train\_set)){  
 train\_set$g\_t[i] <- train\_set$g\_t[i-1] \* (1 + train\_set$e\_t[i])  
}  
# step 3  
n\_years <- (nrow(train\_set)-1) / 52  
# step 4  
CAGR <- (train\_set$g\_t[nrow(train\_set)]/train\_set$g\_t[1])^(1/n\_years) - 1  
# step 5  
nu <- 52^0.5 \* sd(train\_set$e\_t, na.rm = TRUE)  
# step 6  
SR <- CAGR / nu  
SR

## [1] -0.01095925

So the Sharpe ratio for SP500 is 0.2807, higher than -0.0110 for bonds. Hence SP500 ETF is a better investment: it achieves a better balance between expected returns and risks.

## Part 7

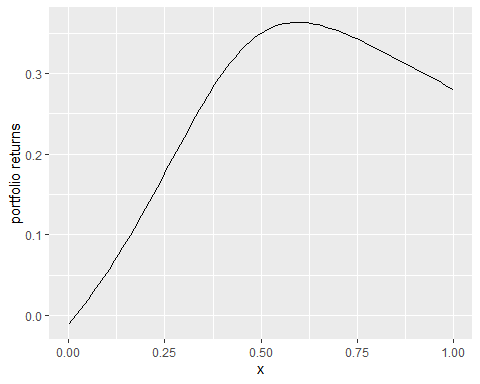
The above calculation procedure can be generalized and wrapped up in a function:

# part 7  
# the main function  
sr\_portfolio <- function(x, a1, a2, fr){  
 n\_obs <- length(a1)  
 y\_t\_1 <- c(NA, head(fr, -1))  
 srs <- rep(NA, length(x))  
 for (ix in 1:length(x)){  
 xx <- x[ix]  
 # step 0  
 r\_t <- xx\*a1 + (1-xx)\*a2  
 # step 1  
 e\_t <- r\_t - y\_t\_1/52  
 # step 2  
 g\_t <- rep(NA, n\_obs)  
 g\_t[1] <- 100  
 for (i in 2:n\_obs){  
 g\_t[i] <- g\_t[i-1] \* (1 + e\_t[i])  
 }  
 # step 3  
 n\_years <- (n\_obs-1) / 52  
 # step 4  
 CAGR <- (g\_t[n\_obs]/g\_t[1])^(1/n\_years) - 1  
 # step 5  
 nu <- 52^0.5 \* sd(e\_t, na.rm = TRUE)  
 # step 6  
 SR <- CAGR / nu  
 srs[ix] <- SR  
 }  
 return(srs)  
}  
# recompute SR for part 6  
x <- c(0.0, 1.0)  
a1 <- train\_set$r\_sp500  
a2 <- train\_set$r\_bonds  
fr <- train\_set$fed.rate  
sr\_portfolio(x, a1, a2, fr)

## [1] -0.01095925 0.28071757

We test its correctness with the portfolio in part 6 and find that they produce the same results. Next, we plot the function for different weights:

# plot it as a function of x  
base <- ggplot(data.frame(x = c(0, 1)), aes(x))  
base + stat\_function(fun = sr\_portfolio,   
 args = list(a1 = train\_set$r\_sp500, a2=train\_set$r\_bonds, fr=train\_set$fed.rate)) +  
 xlab("x") + ylab("portfolio returns")



It can be seen that the portfolio returns first increase as x increases from x=0 and then decrease until x=1 and there is a maxiumum around x=0.6.

## Part 8

We find the optimum weights as:

# part 8  
x\_opt <- optimize(sr\_portfolio, c(0, 1), tol=1e-5, maximum = TRUE,  
 a1 = train\_set$r\_sp500, a2=train\_set$r\_bonds, fr=train\_set$fed.rate)  
x\_opt

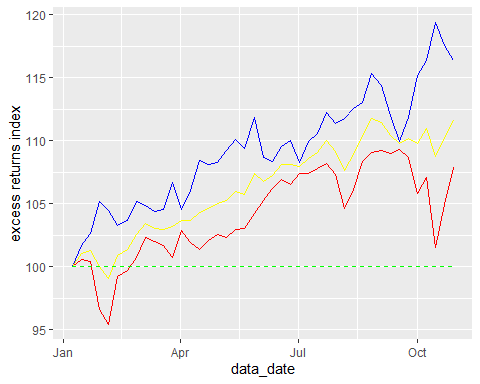
## $maximum  
## [1] 0.5958418  
##   
## $objective  
## [1] 0.3634139

So the optimal weight is x=0.5958 and the maximum Sharpe ratio is 0.3634. Recall that SP500 only gives us a Sharpe ratio of 0.2807 and bonds only gives us -0.0110, the combined portfolio is a better investment.

## Part 9

Now, we test the performance of over the observations in 2014:

### from now on, work on test set###  
# part 9  
# redo part 3  
# fed rate to decimal  
test\_set$fed.rate <- test\_set$fed.rate / 100  
# SP500 returns  
test\_set$l\_sp500 <- c(NA, head(test\_set$close.spy, -1))  
test\_set$r\_sp500 <- test\_set$close.spy / test\_set$l\_sp500 - 1  
# bonds returns  
test\_set$l\_bonds <- c(NA, head(test\_set$close.tlt, -1))  
test\_set$r\_bonds <- test\_set$close.tlt / test\_set$l\_bonds - 1  
# optimal portfolio  
test\_set$r\_portf <- x\_opt$maximum[1] \* test\_set$r\_sp500 + (1 - x\_opt$maximum[1]) \* test\_set$r\_bonds  
# redo part 6  
# first, sp500  
# step 0  
# step 1  
test\_set$l\_frate <- c(NA, head(test\_set$fed.rate, -1))  
test\_set$e\_t <- test\_set$r\_sp500 - test\_set$l\_frate/52  
# step 2  
test\_set$g\_t <- NA  
test\_set$g\_t[1] <- 100  
for (i in 2:nrow(test\_set)){  
 test\_set$g\_t[i] <- test\_set$g\_t[i-1] \* (1 + test\_set$e\_t[i])  
}  
  
# second, bonds  
test\_set$e\_t1 <- test\_set$r\_bonds - test\_set$l\_frate/52  
# step 2  
test\_set$g\_t1 <- NA  
test\_set$g\_t1[1] <- 100  
for (i in 2:nrow(test\_set)){  
 test\_set$g\_t1[i] <- test\_set$g\_t1[i-1] \* (1 + test\_set$e\_t1[i])  
}  
  
# third, optimal portfolio  
# second, bonds  
test\_set$e\_t2 <- test\_set$r\_portf - test\_set$l\_frate/52  
# step 2  
test\_set$g\_t2 <- NA  
test\_set$g\_t2[1] <- 100  
for (i in 2:nrow(test\_set)){  
 test\_set$g\_t2[i] <- test\_set$g\_t2[i-1] \* (1 + test\_set$e\_t2[i])  
}  
  
# plot all out  
test\_set$const <- 100  
ggplot() +   
 geom\_line(data = test\_set, aes(x = date, y = g\_t), color = "red") +  
 geom\_line(data = test\_set, aes(x = date, y = g\_t1), color = "blue") +  
 geom\_line(data = test\_set, aes(x = date, y = g\_t2), color = "yellow") +  
 geom\_line(data = test\_set, aes(x = date, y = const), color = "green", linetype = "dashed") +  
 xlab('data\_date') + ylab('excess returns index')



Still, we use red for SP500, blue for bonds, while the yellow line is for the combined portfolio with optimal weight found in part 8. We find that:

1. for most periods, the three time series are above 100, i.e., they mostly yield positive excess returns;
2. the bonds perform the best, following by the combined portfolio while the SP500 performs the worst.

## Part 10

The money one would earn in addition to the risk free rate at the end of 2014 with $100 at the beginning of 2014 is:

# part 10  
cat('sp500 only', test\_set$g\_t[nrow(test\_set)], '\n')

## sp500 only 107.8763

cat('bonds only', test\_set$g\_t1[nrow(test\_set)], '\n')

## bonds only 116.376

cat(' portfolio', test\_set$g\_t2[nrow(test\_set)], '\n')

## portfolio 111.6367

That is, we get 108 with SP500, 116 with bonds, and 112 with combined portfolio. The portfolio seems to underperform in the test set. This might because the optimal weight we get from the test set is for the entire period over different economic conditions, while the test set is just for one year, where economic condition matters more. To get a better performance, we should take the economic situation into account and design different portfolios for different situations.