Homework 5

SDGB 7844, Prof. Nagaraja

Yuwen Wu 3:30-5:30

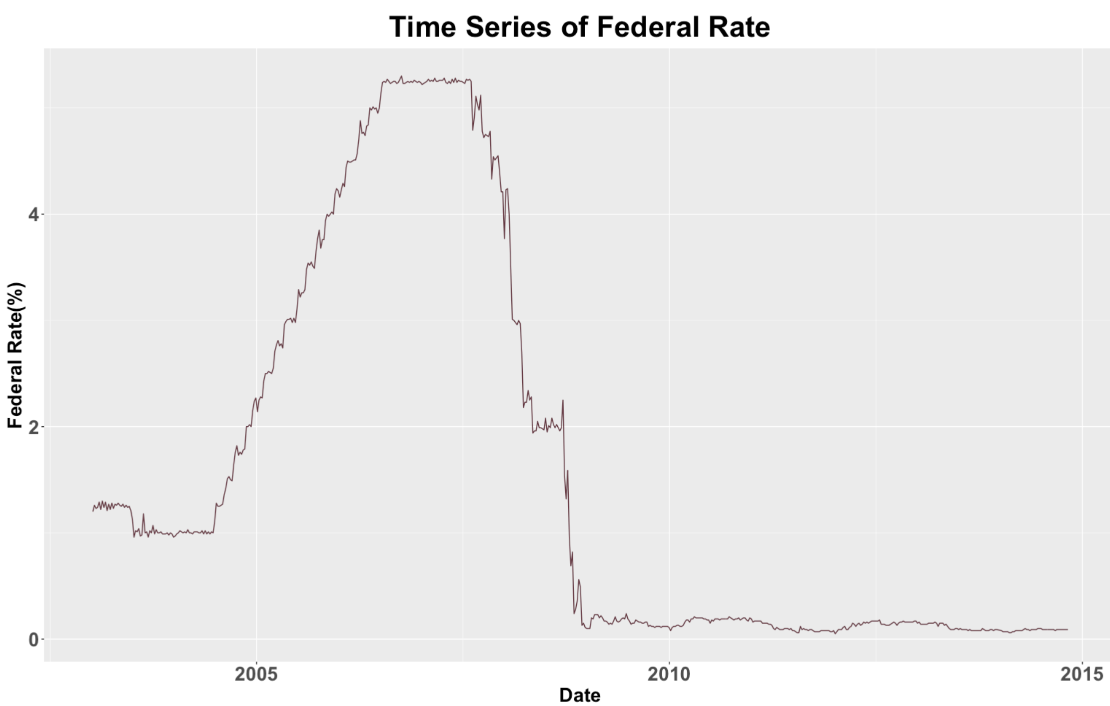
12/05/2018

Question 1

Solution:

```
# read data
data.x<-read_delim("asset_data.txt", delim=",", col_names=TRUE)
# check date
head(data.x)

## # A tibble: 6 x 4
##   date       close.spy close.tlt fed.rate
##   <date>         <dbl>     <dbl>    <dbl>
## 1 2003-01-02      91.1      86.3     NA
## 2 2003-01-03      91.4      86.5     NA
## 3 2003-01-06      93.0      86.2     NA
## 4 2003-01-07      92.7      86.6     NA
## 5 2003-01-08      91.4      87.0      1.2
## 6 2003-01-09      92.8      85.3     NA

# extract only the observations where the federal funds rate is available
data.clean<-data.x[complete.cases(data.x),]
# start date of data.clean
head(data.clean) ## the start date is 2003-01-08

## # A tibble: 6 x 4
##   date       close.spy close.tlt fed.rate
##   <date>         <dbl>     <dbl>    <dbl>
## 1 2003-01-08      91.4      87.0     1.2
## 2 2003-01-15      92.4      86.4     1.26
## 3 2003-01-22      88.2      87.9     1.23
## 4 2003-01-29      86.5      87.2     1.24
## 5 2003-02-05      84.8      87.5     1.29
## 6 2003-02-12      82.1      88.1     1.22

# end date of data.clean
tail(data.clean) ## the end date is 2014-10-29

## # A tibble: 6 x 4
##   date       close.spy close.tlt fed.rate
##   <date>         <dbl>     <dbl>    <dbl>
```

```
## 1 2014-09-24      200.      115.     0.09
## 2 2014-10-01      194.      118.     0.09
## 3 2014-10-08      197.      119.     0.09
## 4 2014-10-15      186.      123.     0.09
## 5 2014-10-22      193.      121.     0.09
## 6 2014-10-29      198.      119.     0.09

# Graph the federal funds interest rate as a time series.

theme.info <- theme(plot.title = element_text(size=30, face = "bold",
hjust=0.5),
                    axis.title = element_text(size=20, face = "bold"),
                    axis.text=element_text(size=20, face = "bold"),
                    legend.title=element_blank(),
                    legend.text = element_text(size = 20),
                    legend.position = "top")


p <- data.clean %>% ggplot(aes(x=date,y=fed.rate))+
  geom_line(color="#663A44")+
  ggtitle("Time Series of Federal Rate") +
  labs(y="Federal Rate(%)", x="Date") +
  theme.info

p
```



The start date of cleaned dataset is 2003-01-08 and the end date is 2014-10-29.

From the Federal Rate Time Series plot, it could be seen that the federal Rate increased from 2004 to 2008 and dramatically decreased from 2008 to 2009 and then stay in low federal rate approximately to zero from 2009 to 2014.

During the 2017 to 2018, the financial crisis happened. In order to increase the liquidity of the financial market, the FED decreased the fed rate to stable the market and install liquidity.

Question 2

Solution:

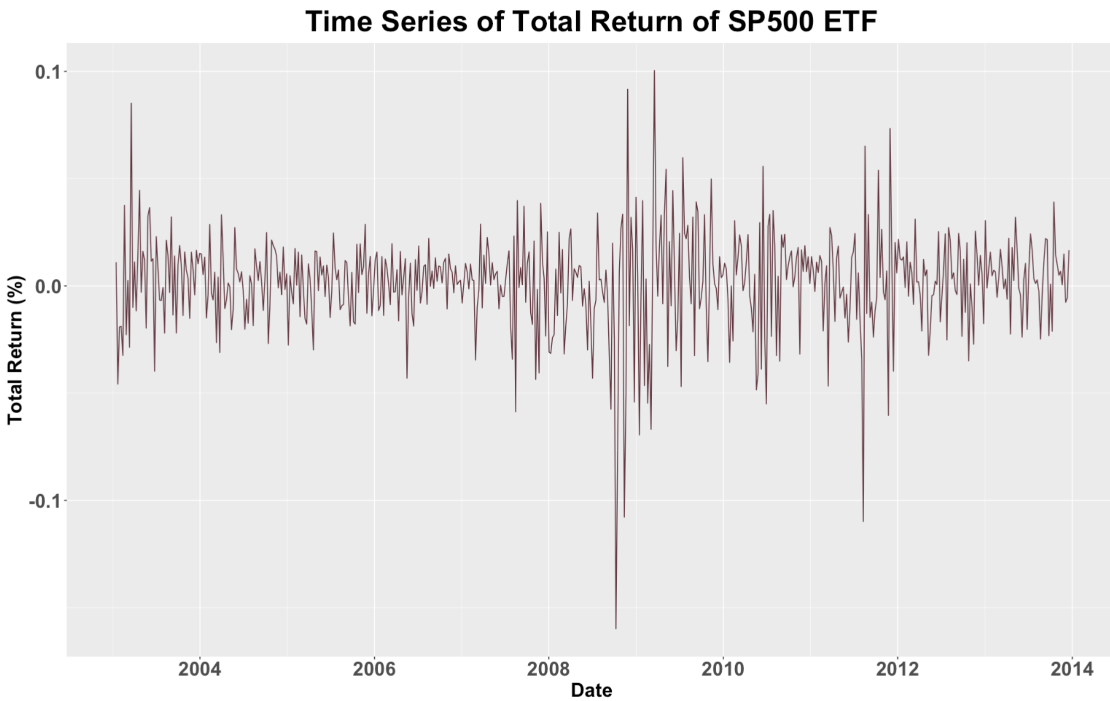
```
# split data into two parts by using specific date
# set the date
a=ymd("2014-01-01")
# set the training data
tr.data <- filter(data.clean,date < a)
# check data
head(tr.data)

## # A tibble: 6 x 4
##   date       close.spy close.tlt fed.rate
##   <date>         <dbl>     <dbl>    <dbl>
## 1 2003-01-08      91.4      87.0     1.2
## 2 2003-01-15      92.4      86.4     1.26
## 3 2003-01-22      88.2      87.9     1.23
## 4 2003-01-29      86.5      87.2     1.24
## 5 2003-02-05      84.8      87.5     1.29
## 6 2003-02-12      82.1      88.1     1.22

# set the test data
te.data <- filter(data.clean,date > a)
# check data
head(te.data)

## # A tibble: 6 x 4
##   date       close.spy close.tlt fed.rate
##   <date>         <dbl>     <dbl>    <dbl>
## 1 2014-01-08      184.      103.     0.08
## 2 2014-01-15      185.      104.     0.07
## 3 2014-01-22      184.      105.     0.07
## 4 2014-01-29      177.      108.     0.07
## 5 2014-02-05      175.      107.     0.07
## 6 2014-02-12      182.      106.     0.06

# check the observation number in training data
nrow(tr.data)

## [1] 570
```

```
# check the observation number in test data
nrow(te.data)

## [1] 43
```

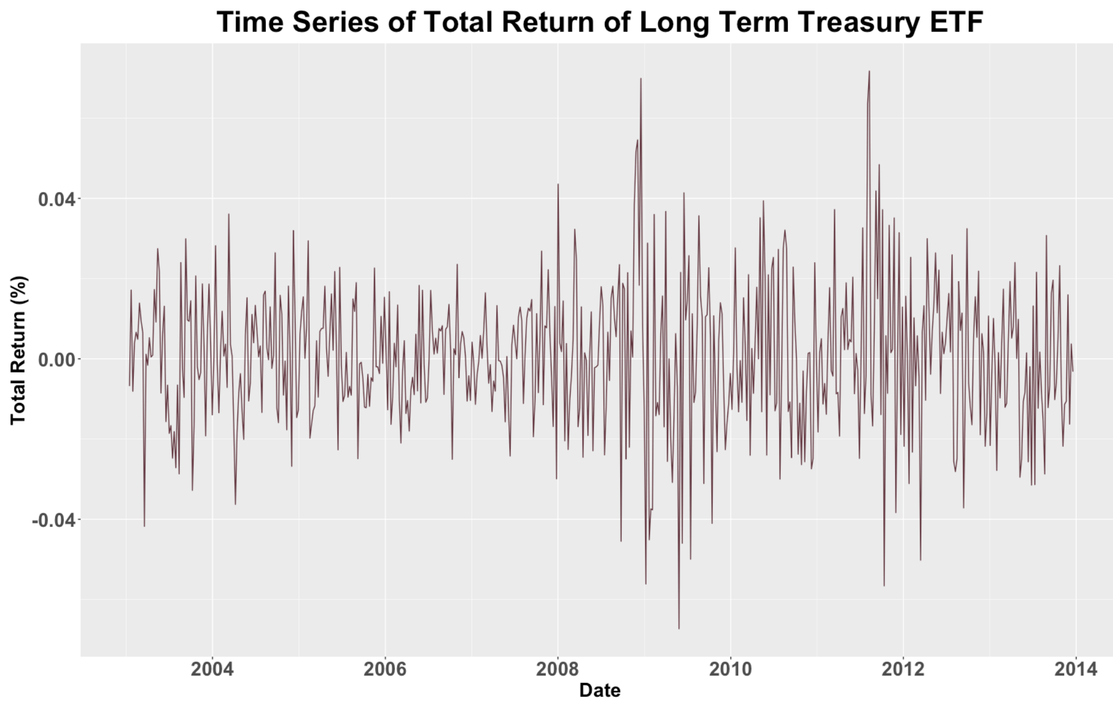
Question 3

```
# convert federal rate to decimal
tr.data$fed.rate<-tr.data$fed.rate/100
# compute the return of sp500 and ETF
n<-length(tr.data$close.spy)
spy<-NA
tlt<-NA
for(i in 2:n){
  r1 <- (tr.data$close.spy[i] - tr.data$close.spy[i-1])/tr.data$close.spy[i-
1]
  r2 <- (tr.data$close.tlt[i] - tr.data$close.tlt[i-1])/tr.data$close.tlt[i-
1]
  # calculate return by function
  spy <- c(spy,r1)
  tlt <- c(tlt,r2) # combine them together
}#end for Loop
tr.data <- tr.data %>% mutate(total_return_of_close_spy = spy,
                              total_return_of_close_tlt = tlt)


# check data

head(tr.data)

## # A tibble: 6 x 6
##   date       close.spy close.tlt fed.rate total_return_of…
total_return_of…
##   <date>         <dbl>     <dbl>    <dbl>            <dbl>
<dbl>
## 1 2003-01-08      91.4      87.0   0.012            NA               NA
## 2 2003-01-15      92.4      86.4   0.0126            0.0111          -
0.00678
## 3 2003-01-22      88.2      87.9   0.0123           -0.0458           0.0171
## 4 2003-01-29      86.5      87.2   0.0124           -0.0192          -
0.00808
## 5 2003-02-05      84.8      87.5   0.0129           -0.0188
0.00390
## 6 2003-02-12      82.1      88.1   0.0122           -0.0324
0.00663

p1 <- tr.data %>% ggplot(aes(x=date,y=total_return_of_close_spy))+
  geom_line(color="#663A44")+
  ggtitle("Time Series of Total Return of SP500 ETF") +
  labs(y="Total Return (%)", x="Date") +
  theme.info
```

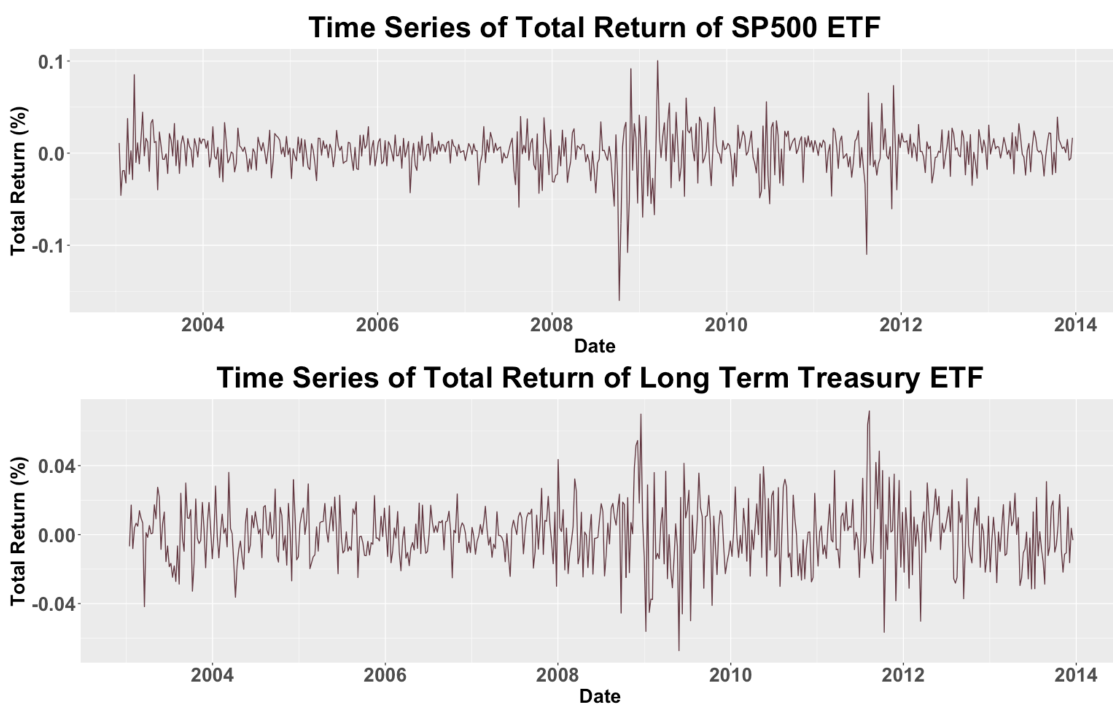
```
p1
```




```
p2 <- tr.data %>% ggplot(aes(x=date,y=total_return_of_close_tlt))+
  geom_line(color="#663A44")+
  ggtitle("Time Series of Total Return of Long Term Treasury ETF") +
  labs(y="Total Return (%)", x="Date") +
  theme.info

p2
```

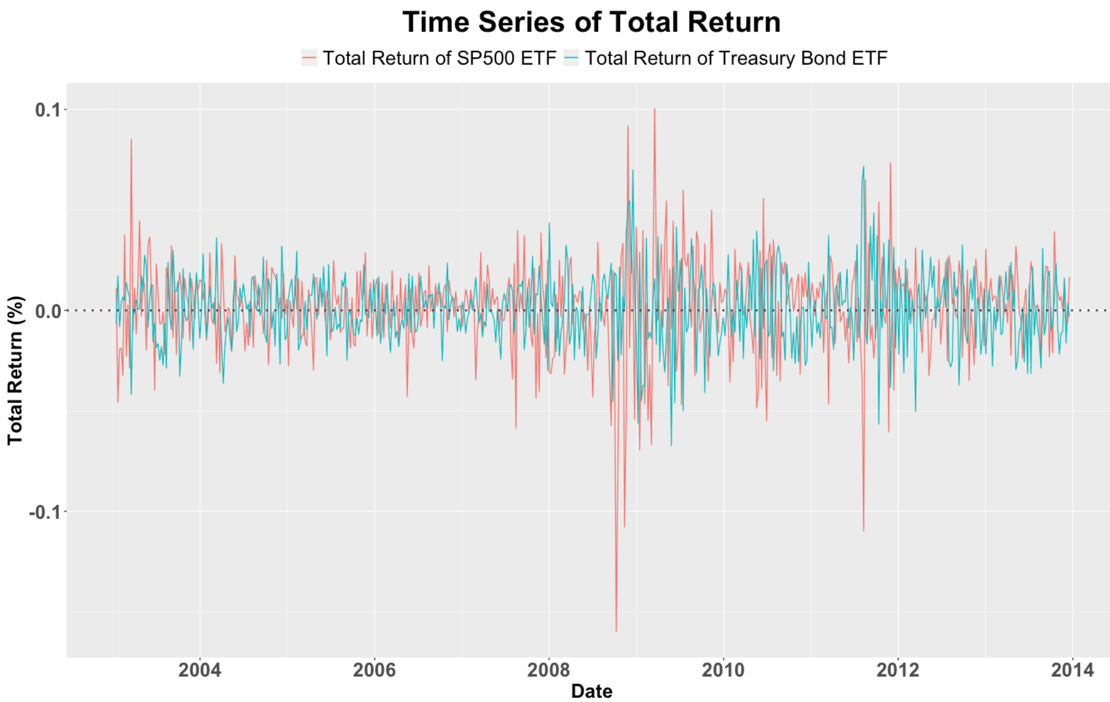



```
plot=grid.arrange(p1, p2, nrow = 2)
```




```
p3 <- ggplot()+
  geom_line(data = tr.data, aes(x = date, y = total_return_of_close_spy,
color = "Total Return of SP500 ETF"))+
  geom_line(data = tr.data, aes(x = date, y = total_return_of_close_tlt,
color = "Total Return of Treasury Bond ETF"))+
  geom_hline(yintercept=0, linetype=3, color="#663A44",size = 1) +
```

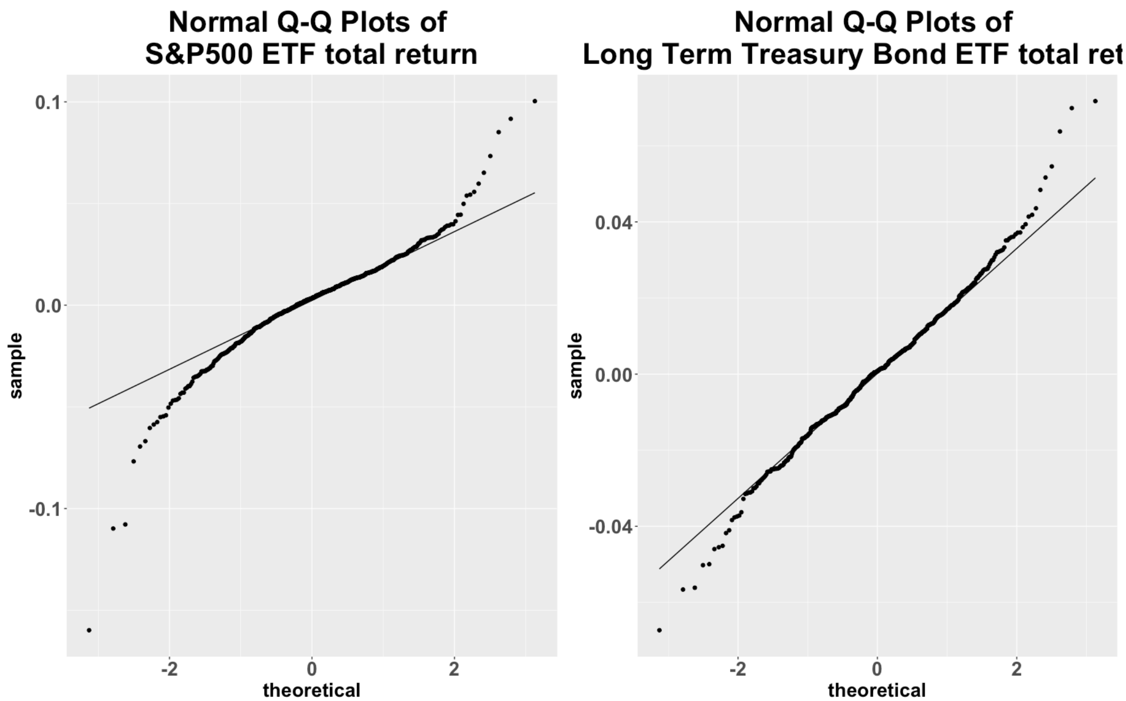
```
  ggtitle("Time Series of Total Return") +
  labs(x='Date',y='Total Return (%)')+
  theme.info

p3
```



From above time series plot , we can see that the S&P500 ETF was more fluctuated than long term treasury bond. In other words, The long term treasury bond has lower variance than S&P500. From 2008 to 2009, the total return decreased large from positive to negative because of the financial crisis in 2008 and in 2009.

Question 4

```
#qqplot for S&P500 ETF total return
qq.1 <- tr.data %>% ggplot(aes(sample = total_return_of_close_spy)) +
        stat_qq() +
        stat_qq_line() +
        ggtitle("Normal Q-Q Plots of \nS&P500 ETF total return") +
        theme.info
#qqplot for Long Term Treasury Bond ETF total return
qq.2 <- tr.data %>% ggplot(aes(sample = total_return_of_close_tlt)) +
        stat_qq() +
        stat_qq_line() +
        ggtitle("Normal Q-Q Plots of \nLong Term Treasury Bond ETF total
return") +
        theme.info

grid.arrange(qq.1, qq.2, nrow = 1)
```

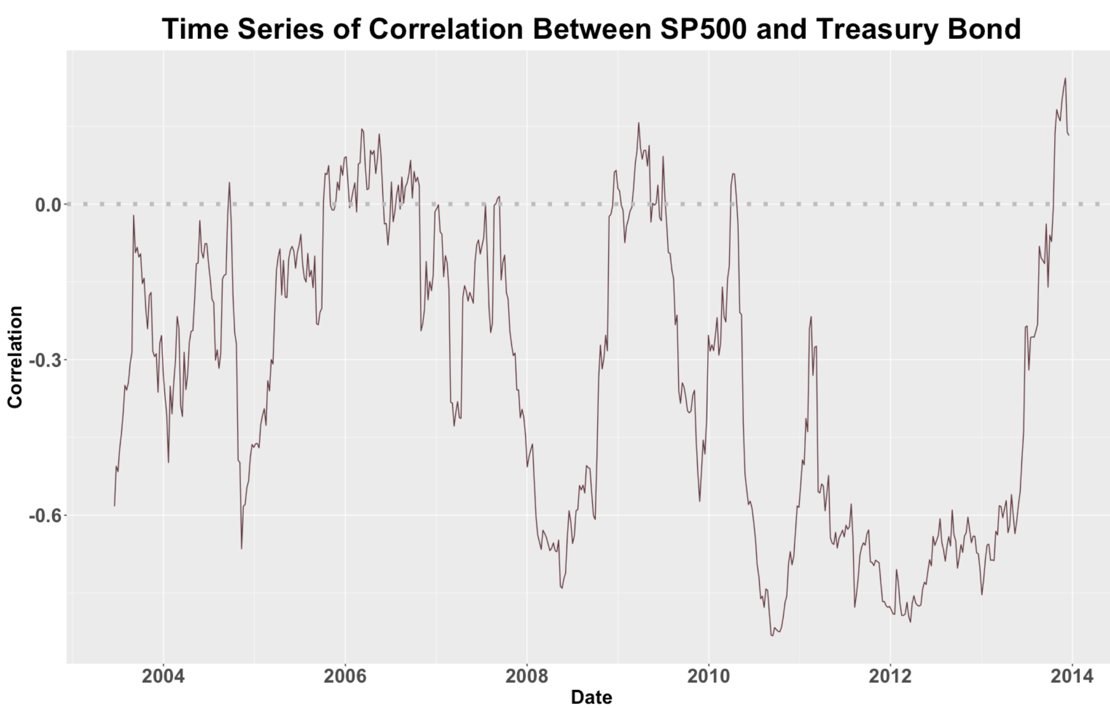


The Sharpe ratio calculation assumes that returns are normally distributed. From the above plots, we could see that these two datasets have fat(heavy) tailed and some outliers. The total return of SP500 ETF and total return of treasury bond ETF do not correspond to normal distribution. However, according to the Central Limit Theorem, when number of observations larger than 30, we can assume the data distribution is normally distributed.

Question 5

```
# Compute the correlation between S&P500 and Long Term Treasury Bond ETF
returns in the training set
cor(tr.data$total_return_of_close_spy,tr.data$total_return_of_close_tlt,
use="complete.obs")

## [1] -0.3439013

# Compute the total number we need to loop
n <- nrow(tr.data)-23

# Compute the rolling-window correlations
Corr<-NULL
for (i in 1:n){
    corr<-cor(tr.data[i:(i+23),5],tr.data[i:(i+23),6],
              use = "complete.obs")
    Corr<-c(Corr,corr)
}#end loop

head(Corr)

## [1] -0.5828004 -0.5054171 -0.5160272 -0.4710266 -0.4419905 -0.4023793
```

```
tr.data.p <- tr.data[24:570,] %>% mutate(correlation = Corr)
head(tr.data.p)

## # A tibble: 6 x 7
##   date       close.spy close.tlt fed.rate total_return_of…
total_return_of…
##   <date>         <dbl>     <dbl>    <dbl>            <dbl>
<dbl>
## 1 2003-06-18     102.       94.3  0.0125           0.0127           -0.0157
## 2 2003-06-25      97.5      93.7  0.0121          -0.0398           -
0.00658
## 3 2003-07-02      99.8      91.9  0.0113           0.0230           -0.0186
## 4 2003-07-09     101.       90.4  0.00960          0.00812          -0.0166
## 5 2003-07-16      99.9      88.2  0.0102          -0.00656          -0.0248
## 6 2003-07-23      99.2      86.6  0.0101          -0.00681          -0.0182
## # … with 1 more variable: correlation <dbl>

p4 <- tr.data.p %>% ggplot(aes(x=date,y=correlation))+
  geom_line(color="#663A44")+
  ggtitle("Time Series of Correlation Between SP500 and Treasury Bond") +
  labs(y="Correlation", x="Date") +
  geom_hline(yintercept=0, linetype=3, color="grey",size=2) +
  theme.info

p4
```



The correlation between this two portfolios is -0.3379974, which is negative, and the negative correlation presents the negative relationship between these two portfolios.

The rolling-window correlations would be the better way to describe the relationship between these two assets because in the different period the correlation would be change with the fluctuation of markets.

The one time correlation could not reflect the change of the correlation between these two assets. For example, in the financial crisis, the correlation decreased a lot since the Treasury Bonds' ETF is the subtitute of the SP500 ETF. When the return decreased in the SP500 ETF, more and more investors would buy Treasury Bonds as their risk free assets.

In this case, the rolling window correlations could be more representative.

Question 6

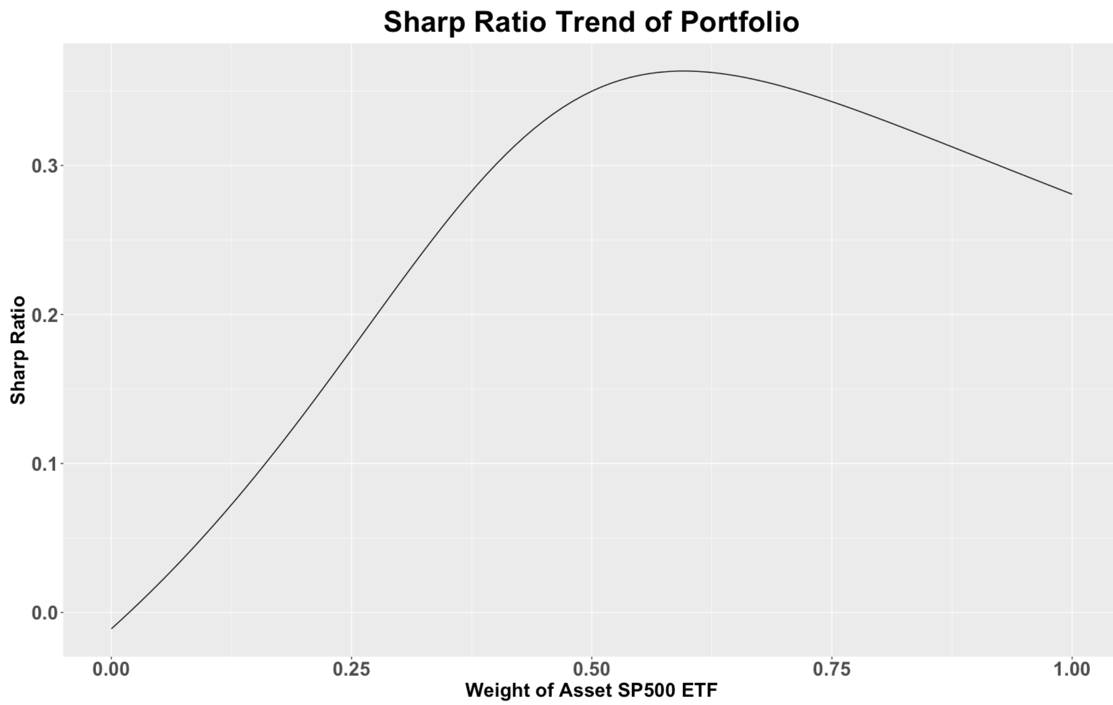
```
# step0:
rt_spy<- tr.data$total_return_of_close_spy
rt_tlt<- tr.data$total_return_of_close_tlt
yt <- tr.data$fed.rate
# step1:
# excess return
excess_return_spy <- NA
excess_return_tlt <- NA
n <- length(tr.data$total_return_of_close_spy)
for(i in 2:n){
  et1<-rt_spy[i]-(yt[i-1]/52)
  et2<-rt_tlt[i]-(yt[i-1]/52)
  excess_return_spy <- c(excess_return_spy,et1)
  excess_return_tlt <- c(excess_return_tlt,et2)
}#end for Loop

# step2:
# excess return index
excess_return_index_spy<-c(100)
excess_return_index_tlt<-c(100)
for(i in 2:n){
  gt1<-(excess_return_index_spy[i-1])*(1+(excess_return_spy[i]))
  gt2<-(excess_return_index_tlt[i-1])*(1+(excess_return_tlt[i]))
  excess_return_index_spy<-c(excess_return_index_spy,gt1)
  excess_return_index_tlt<-c(excess_return_index_tlt,gt2)
}#end for Loop

# step3:
# compute years
year_number<-(nrow(tr.data)-1)/52
year_number

## [1] 10.94231

# step4:
# Compute CAGR
cagr_spy<-
```

```
((excess_return_index_spy[n]/excess_return_index_spy[1])^(1/year_number))-1
cagr_spy

## [1] 0.04721991

cagr_tlt<-
((excess_return_index_tlt[n]/excess_return_index_tlt[1])^(1/year_number))-1
cagr_tlt

## [1] -0.001429458

# step5:
# annualized volatility
excess_return_spy <- tibble(excess_return_spy)
excess_return_tlt <- tibble(excess_return_tlt)
head(excess_return_spy)

## # A tibble: 6 x 1
##   excess_return_spy
##               <dbl>
## 1           NA
## 2            0.0108
## 3           -0.0460
## 4           -0.0194
## 5           -0.0191
## 6           -0.0327

sd(excess_return_spy$excess_return_spy,na.rm = TRUE)

## [1] 0.02332673

volatility_spy<-sqrt(52)*sd(excess_return_spy$excess_return_spy,na.rm = TRUE)

volatility_tlt<-sqrt(52)*sd(excess_return_tlt$excess_return_tlt,na.rm = TRUE)

# step6:
# Compute Sharp Ratio
sharp_spy<-cagr_spy/volatility_spy
sharp_tlt<-cagr_tlt/volatility_tlt

# check data
head(sharp_spy)

## [1] 0.2807176

head(sharp_tlt)

## [1] -0.01095925
```

The sharp ratio of SP500 ETF is 0.28072, and the sharp ratio of Long Term Treasury Bonds ETF is -0.0109. Considering the term of sharp ratio, which is measuring the excess return per unit the risk take, the SP500 ETF would be a better choice in this period.

Question 7

```
port<-function(x,
               asset1=tr.data$total_return_of_close_spy,
               asset2=tr.data$total_return_of_close_tlt,
               ffr=tr.data$fed.rate){
  if((x>=0)&(x<=1)){ # the weight of assets must be in [0,1]
    portfolio_return <-(x*asset1)+((1-x)*asset2)   # portifolio return
function
    excess_return<-NA        # excess return
    excess_index_return<-100 # excess index return
    for(i in 2:570){
      et<-portfolio_return[i]-(ffr[i-1]/52)
      excess_return<-c(excess_return,et)
    }
    for(j in 2:570){
      gt<-excess_index_return[j-1]*(1+excess_return[j])
      excess_index_return<-c(excess_index_return,gt)
    }

    year_number <-(570-1)/52
    cagr_portfolio<-
(excess_index_return[570]/excess_index_return[1])^(1/year_number)-1
    volatility_portfolio<-sqrt(52)*sd(excess_return,na.rm=TRUE)
    sharp_ratio_portfolio<-cagr_portfolio/volatility_portfolio
    return("Sharp Ratio"= sharp_ratio_portfolio) ## return list of sharpe
ratio of portfolio
  }else{
    return("ERROR: x should be between 0 and 1")
  }    # print error when x is not between o and 1
}

# Check the Sharp Ratio of Function
port(0.2)

## [1] 0.132891

#use "vectorized"  function to change the result to a vector in order for we
to make the plot
port.vectorized <- Vectorize(port, vectorize.args="x")
#make plot of the portfolio curve

plot <- ggplot(tibble(x = c(0, 1)), aes(x = x)) +
  stat_function(fun = port.vectorized)+
  ggtitle("Sharp Ratio Trend of Portfolio") +
  labs(y="Sharp Ratio", x="Weight of Asset SP500 ETF")+
  theme.info
plot
```




```
port(0.6)

## [1] 0.3633928
```

From the above plot, when the portfolio asset 1 weight nearly 0.6, the Sharp Ratio would approximate to the largest value in 0.3633928.

Question 8

Solution:

```
## optimum weight for each asset with the function optimize
optimize(port.vectorized,lower=0,upper=1,maximum = TRUE)

## $maximum
## [1] 0.5958502
##
## $objective
## [1] 0.3634139

## The weight for asset one
optimize(port.vectorized,lower=0,upper=1,maximum = TRUE)$maximum

## [1] 0.5958502

## 0.5958502

## The weight for asset two
1-optimize(port.vectorized,lower=0,upper=1,maximum = TRUE)$maximum
```

```
## [1] 0.4041498

## 0.4041498

## optimum sharp ratio
optimize(port.vectorized,lower=0,upper=1,maximum = TRUE)$objective

## [1] 0.3634139

## 0.3634139

## All invest in Long Term Treasury Bond ETF
port(0)

## [1] -0.01095925

## All invest in SP500 ETF
port(1)

## [1] 0.2807176
```

From above function, the optimum weight of SP500 ETF is 59.585% and the optimum weight of Long Term Treasury Bond ETF is 40.41%. In this case, the optimum sharp ratio should be 0.3634139.

According above analysis, I think we should invest in this optimum ratio becasuse in this ratio, we could earn highest excess return with the unit risk. If we invest all fund in Long Term Treasury Bond ETF, the sharp ratio would be -0.01095925, and if we invest all fund in SP500 ETF, the sharp ratio would be 0.2807176. These two sharp ratio all less than portfolio best allocation. Therefore, I would invest in portfolio to maximize the return and minimize the risk.

Question 9

Solution:

```
# change federal rate to decimal form in testing data
te.data$fed.rate<-te.data$fed.rate/100

# step1:
# total return for SP500 and Treasury Bonds ETF

n_test<-length(te.data$close.spy)
spy_test<-NA
tlt_test<-NA
for(i in 2:n_test){
  r1 <- (te.data$close.spy[i] - te.data$close.spy[i-1])/te.data$close.spy[i-
1]
  r2 <- (te.data$close.tlt[i] - te.data$close.tlt[i-1])/te.data$close.tlt[i-
1]
  # calculate return by function
```

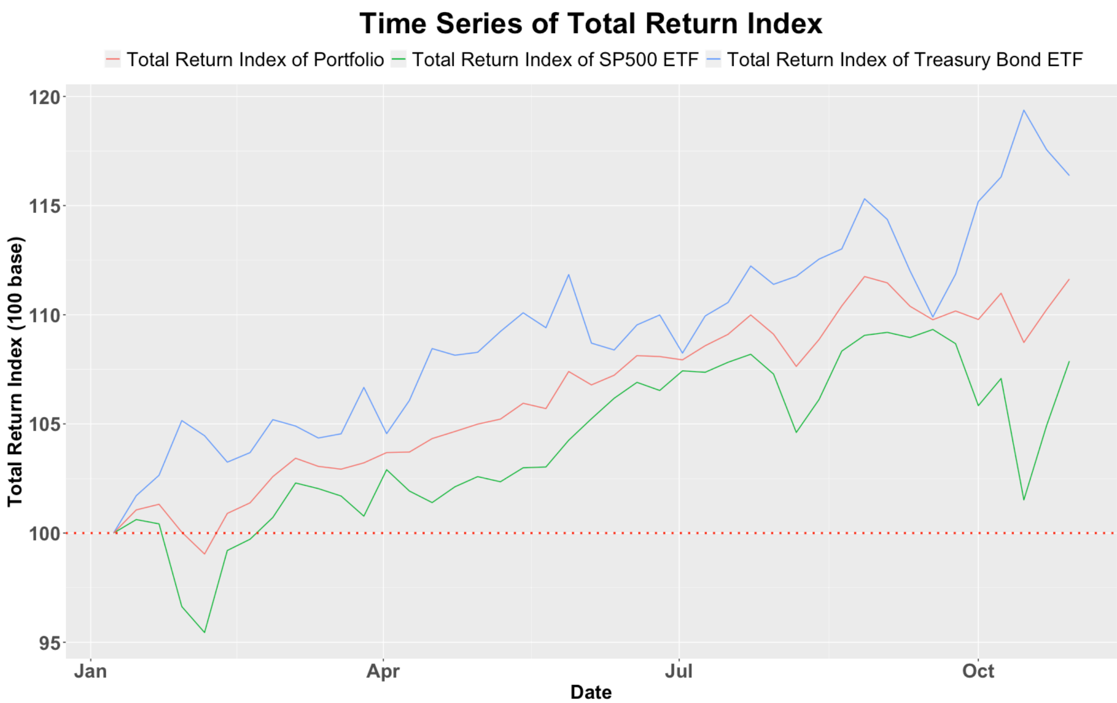
```r
  spy_test <- c(spy_test,r1)
  tlt_test <- c(tlt_test,r2) # combine them together
}#end for loop
te.data <- te.data %>% mutate(total_return_of_close_spy_test = spy_test,
                              total_return_of_close_tlt_test = tlt_test)

###############################
# step2:
# excess return for SP500 and Treasury Bonds ETF
rt_spy_test<- te.data$total_return_of_close_spy_test
rt_tlt_test<- te.data$total_return_of_close_tlt_test
yt_test <- te.data$fed.rate

excess_return_spy_test <- NA
excess_return_tlt_test <- NA
for(i in 2:n_test){
  et1<-rt_spy_test[i]-(yt_test[i-1]/52)
  et2<-rt_tlt_test[i]-(yt_test[i-1]/52)
  excess_return_spy_test <- c(excess_return_spy_test,et1)
  excess_return_tlt_test <- c(excess_return_tlt_test,et2)
}#end for loop

# step3:
# excess returns index for SP500 ETF and Long Term Treasury Bond ETF

excess_return_index_spy_test<-c(100)
excess_return_index_tlt_test<-c(100)
for(i in 2:n_test){
  gt1<-(excess_return_index_spy_test[i-1])*(1+(excess_return_spy_test[i]))
  gt2<-(excess_return_index_tlt_test[i-1])*(1+(excess_return_tlt_test[i]))
  excess_return_index_spy_test<-c(excess_return_index_spy_test,gt1)
  excess_return_index_tlt_test<-c(excess_return_index_tlt_test,gt2)
}#end for loop

te.data <- te.data %>% mutate(excess_return_spy_test =
excess_return_index_spy_test,
                              excess_return_tlt_test =
excess_return_index_tlt_test,
                              excess_return_index_spy =
excess_return_index_spy_test,
                              excess_return_index_tlt =
excess_return_index_tlt_test)


# step 4:
# testing data portfolio

port.return<-function(x,      #write function with four input
               asset1=te.data$total_return_of_close_spy_test,
```

```
                asset2=te.data$total_return_of_close_tlt_test,
                ffr=te.data$fed.rate){
  if((x>=0)&(x<=1)){  # the weight of assetmust be in the internal of [0,1]
    return_port<-(x*asset1)+((1-x)*asset2)   #portfolio rate
    excess_return<-rep(NA,43)
    index<-rep(NA,43)
    index[1]<-100
    for(i in 2:n_test){
      excess_return[i]<-return_port[i]-(ffr[i-1]/52)
    }#end for Loop
    for(j in 2:43){
      index[j]<-index[j-1]*(1+excess_return[j])
    }# end for Loop
    return(index)     #return excess index return
  }else{
    return("ERROR: x should be between 0 and 1")
  }
}
# use the optimum portfolio weight to calculate the test portfolio return
index_test<-port.return(optimize(port.vectorized,lower=0,upper=1,maximum =
TRUE)$maximum)
# add the test portfolio return to test data
te.data <- te.data %>% mutate(index_test = index_test)
# check data
head(te.data)
```

```
## # A tibble: 6 x 11
##   date       close.spy close.tlt fed.rate total_return_of…
total_return_of…
##   <date>         <dbl>     <dbl>    <dbl>            <dbl>
<dbl>
## 1 2014-01-08      184.      103. 0.0008           NA                NA
## 2 2014-01-15      185.      104. 0.0007            0.00621           0.0172
## 3 2014-01-22      184.      105. 0.0007           -0.00195
0.00920
## 4 2014-01-29      177.      108. 0.0007           -0.0377            0.0244
## 5 2014-02-05      175.      107. 0.0007           -0.0123           -
0.00658
## 6 2014-02-12      182.      106. 0.000600          0.0394           -0.0116
## # … with 5 more variables: excess_return_spy_test <dbl>,
## #   excess_return_tlt_test <dbl>, excess_return_index_spy <dbl>,
## #   excess_return_index_tlt <dbl>, index_test <dbl>
```

```
# plot the time series plot of these three assets
p5 <- te.data %>% ggplot()+
  geom_line(aes(x = te.data$date, y = te.data$excess_return_index_spy, color
= "Total Return Index of SP500 ETF"))+
  geom_line(aes(x = te.data$date, y = te.data$excess_return_index_tlt, color
= "Total Return Index of Treasury Bond ETF"))+
  geom_line(aes(x = te.data$date, y = te.data$index_test, color = "Total
```

```
Return Index of Portfolio"))+
  geom_hline(yintercept=100, linetype=3, color="red",size = 1) +
  ggtitle("Time Series of Total Return Index") +
  labs(x='Date',y='Total Return Index (100 base)')+
  theme.info

p5
```



From the plot above, we can see that the return of SP500 ETF and Long Term Treasury ETF are positive, which means we will receive more money than we invested at the end of investment period.

From the above picture, we also could see that the Long Term Treasury Bonds could bring the highest return index than our portfolio and SP500 ETF.

The portfolio return index is in the middle of other two ETF return index becasue the portfolio is combined of these two assets.

Question 10

Solution:

```
# Check the perfermance at the end of the test period for each asset in
addition to the risk-free interest rate

# SP500ETF
te.data$excess_return_index_spy[n_test]

## [1] 107.8763
```

```r
# Long Term Treasury ETF
te.data$excess_return_index_tlt[n_test]

## [1] 116.376

# Portfolio
te.data$index_test[n_test]

## [1] 111.6367
```

The 107.8763 we would be have if we invest all in SP500 ETF. The 116.376 we would be have if we invest all in Long Term Treasury Bonds ETF. The 111.6367 we would have if we invest with optimum weight in SP500 ETF and Treasury Bond ETF.

From above analysis, I think the optimum weight of portfolio in the trainining data is not good in test data and the portfolio is not good performed in the test period than Long Term Treasury Bonds ETF. Additionally, the sharp ratio is based on risk or variance, in other words, the optimum portfolio is the portfolio with the highest excess return based on the risk or variance. Futhermore, in this case, we cannot sell out product, which means we cannot build the portfolio which has better return than the better one of these two assets. Also, the best weight is calculated on training data, and this would not good for test data to use because the variance and return are changing everyday and the best weight also change.