HW\_5\_Template

Lina Cao , Thursdays 1:15PM

SDGB 7844; Prof. Nagaraja; Fall 2017

## Question 1

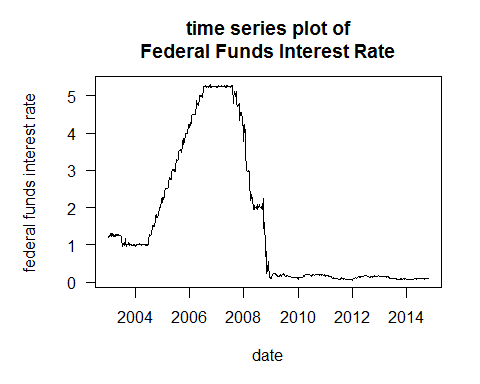
# read the data  
setwd("F:/0000Fordham New Life/4 SM&C.statistics methods and computation I/HW5")  
data.x <- read.csv("asset\_data.txt",stringsAsFactors = FALSE, header=TRUE)  
# turn first column from character string to a date  
data.x$date <- as.Date(data.x$date, format = "%Y-%m-%d")  
# extract weekly data (with federal funds rate available)  
data.clean <- data.x[complete.cases(data.x),]  
  
# the start date  
# locate where the start date appear  
s1 <- which(data.clean$date==min(data.clean$date))  
# extract the start date  
data.clean$date[s1]

## [1] "2003-01-08"

rm(s1)  
# the end date  
s2 <- which(data.clean$date==max(data.clean$date))  
data.clean$date[s2]

## [1] "2014-10-29"

rm(s2)  
  
# plot federal funds interest rate  
par(mar=c(5, 5, 4, 2)) # adjust plot margin  
plot(data.clean$date, data.clean$fed.rate,  
 main = "time series plot of\nFederal Funds Interest Rate", xlab = "date", ylab = "federal funds interest rate", las = TRUE,  
 type = "l")



The start date and end date of the dates that have federal fund interest rate is "2003-01-08" and "2014-10-29" respectively.

From the time series plot, (1) the federal interest rate was slightly over 1% from the start of 2003; (2) in the middle of 2003, it dropped to and remained stable around 1% until middle of 2004; (3) after that, the interest rate kept rising to over 5% until 2007, when crisis in the subprime mortgage market in the US occured; (4) interest rate declined dramatically from 2007's highest point(about 5.2%) to end of 2007(2%); (5) it went down continuously in 2008, since the crisis developed into a full-blown international banking crisis with the collapse of the investment bank Lehman Brothers on September 15, 2008; (6) from year 2009 then on, the federal funds interest rate remains in a low point of slightly over 0 (about 0.1%)

## Question 2

# data set (a) : training data  
aaa <- data.clean[data.clean$date<"2014-01-01",]  
# data set (b) : test data  
bbb <- subset(data.clean, data.clean$date>="2014-01-01" & data.clean$date<"2015-01-01")  
# count the observations  
nrow(aaa)

## [1] 570

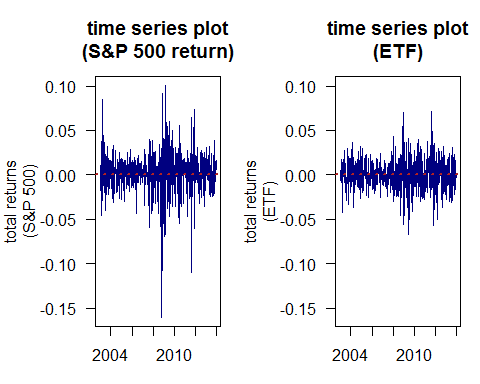
nrow(bbb)

## [1] 43

There are 570 observations in the first data set(date before 2014), while there are 43 observations in year 2014.

## Question 3

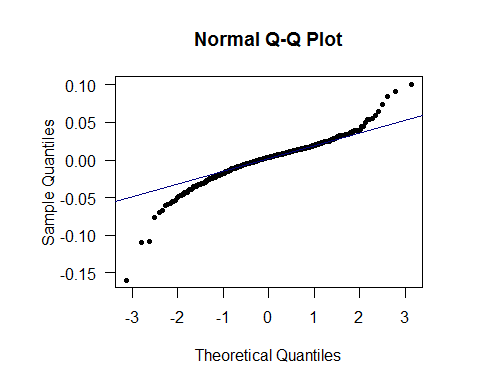
# convert federal fund rate into decimal  
y <- aaa$fed.rate/100  
  
# deal with S&P500 , ETF return  
ccc <- aaa$close.spy  
ddd <- aaa$close.tlt  
rrr1 <- rep(NA, times = length(ccc))  
rrr2 <- rep(NA, times = length(ddd))  
  
for (i in 2:length(ccc)){  
   
 rrr1[i] <- (ccc[i]-ccc[i-1])/ccc[i-1]  
   
} # end for loop1  
  
for (j in 2:length(ddd)){  
   
 rrr2[j] <- (ddd[j]-ddd[j-1])/ddd[j-1]  
   
} # end for loop2  
rm(i,j,ccc,ddd)  
# add both sets to training set (aaa)  
maaa <- data.frame(aaa,rrr1,rrr2)  
  
par(mfrow=c(1,2)) # set plot frame  
par(mar=c(3, 5, 4, 1)) # adjust plot margin  
# training data plot  
# check the min\max of training plot to set identical y=axis  
# min(maaa$rrr1,na.rm = TRUE) , min=-0.1598  
# max(maaa$rrr1,na.rm = TRUE) , max=0.1004  
plot(maaa$date, maaa$rrr1, las=TRUE,  
 type = "l", col="navy",  
 ylim = c(-0.16,0.10),  
 xlab = "year", ylab = "total returns\n(S&P 500)",  
 main = "time series plot\n(S&P 500 return)")  
# add line at y=0  
abline(h=0, col="firebrick", lty=3, lwd=2.5)  
# test data plot  
plot(maaa$date, rrr2, las=TRUE,  
 type = "l", col="navy",  
 ylim = c(-0.16,0.10),  
 xlab = "year", ylab = "total returns\n(ETF)",  
 main = "time series plot\n(ETF)")  
# add line at y=0  
abline(h=0, col="firebrick", lty=3, lwd=2.5)



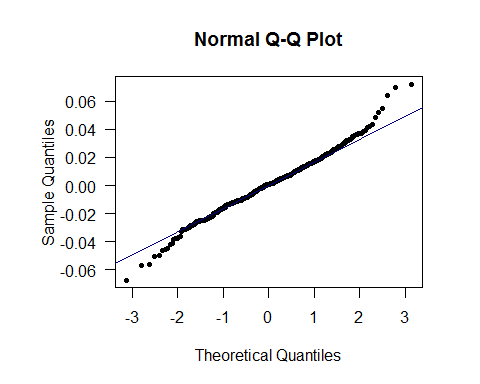
1. Two returns all fluctuate around 0% in the long term. However, the fluctuation of ETF assets is much more stable than S&P 500.
2. The returns tend to have same trend. In middle of 2008 and end of 2011, there are two times when huge volatility occured. And in both plots, the returns showed a similar shape at those two time periods.

## Question 4

# quantile plot for S&P500  
par(mar=c(5,6,4,3))  
qqnorm(maaa$rrr1,pch=20,las=TRUE)  
qqline(maaa$rrr1,col="navy",lwd=1.5)



# quantile plot for ETF asset  
par(mar=c(5,6,4,3))  
qqnorm(maaa$rrr2,pch=20,las=TRUE)  
qqline(maaa$rrr2,col="navy",lwd=1.5)



I don't think they are normally distributed according to the shape. Both quantile plots show a slightly "s-shape", which indicates a thicker tail than normal distribution. There are many outliers in both distribution if we regard them as normal distribution. However, with a large number of data size(570), the final result of the calculated value of cdf or pdf can be quite close to normal distribution on certain x value.

## Question 5

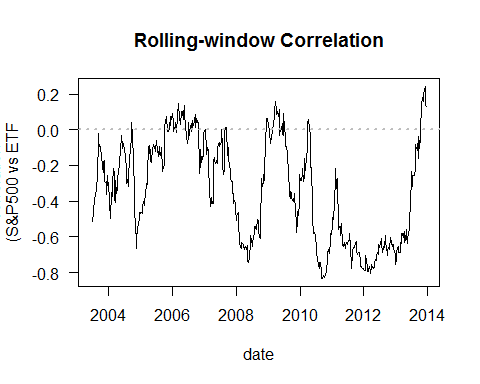
The correlation of S&P500 and ETF is -0.3439.

I think the rolling-window correlation is a better way to describe the relationship. Because the correlation can only show the overall time correlationship between the assets. However, the rolling-window method can give a dynamic impression of the change in correlation. From the plot, the correlation between two assets changes along with the time. Before 2007, the correlation fluctuated around 0, while the correlation tends to be under 0 after 2007. S&P500 and ETF assets are negatively correlated for most of the time after 2007. In late 2013 the correlation tends to have an upper trend and the relationship becomes positive related.

# correlation between S&P500 and ETF  
cor(rrr1[-1],rrr2[-1])

## [1] -0.3439013

# do calculation in a rolling-window  
corr <- rep(NA, times=546)  
for (i in 1:546){  
 corr[i] <- cor(rrr1[(i+1):(i+24)],rrr2[(i+1):(i+24)])  
} # end for loop  
rm(i)  
  
# plot rolling-window correlation  
# use only the date except for the first 24 days to make the point plotted on the last day of the window  
plot(maaa$date[-24:-1],corr,  
 xlab = "date", ylab = "correlation\n(S&P500 vs ETF",  
 main = "Rolling-window Correlation",  
 las = TRUE, pch = 20, cex = 0.8, type = "l")  
abline(h=0, col="gray", lty=3, lwd=2)



## Question 6

# S0:denote r and y for each week  
# rrr1 : return on S&P500   
# rrr2 : return on ETF  
# y : federal funds interest rate  
y <- maaa$fed.rate/100  
  
# S1:excess returns  
e1 <- rep(NA, times=570)  
e2 <- rep(NA, times=570)  
 for (i in 2:570){  
 e1[i] <- rrr1[i]-y[i-1]/52  
 e2[i] <- rrr2[i]-y[i-1]/52  
 }#end for loop  
rm(i)  
  
# S2:convert into index  
g1 <- 100  
gsp <- rep(NA, times = 570)  
getf <- rep(NA, times = 570)  
 for (i in 1:570){  
 if(i==1){  
 gsp[i] <- g1  
 getf[i] <- g1  
 }else{  
 gsp[i] <- gsp[i-1]\*(1+e1[i])  
 getf[i] <- getf[i-1]\*(1+e2[i])  
 }# end if  
 }#end for loop  
rm(i)  
  
# S3:compute number of years  
n <- (nrow(maaa)-1)/52  
  
# S4:CAGR  
cagr.sp <- (gsp[570]/g1)^(1/n)-1  
cagr.etf <- (getf[570]/g1)^(1/n)-1  
  
# S5:annualized volatility  
vsp <- sqrt(52)\*sd(e1,na.rm = TRUE)  
vetf <- sqrt(52)\*sd(e2,na.rm = TRUE)  
  
# S6:Sharpe Ratio  
sr.sp <- cagr.sp/vsp  
sr.etf <- cagr.etf/vetf  
sr.sp

## [1] 0.2807176

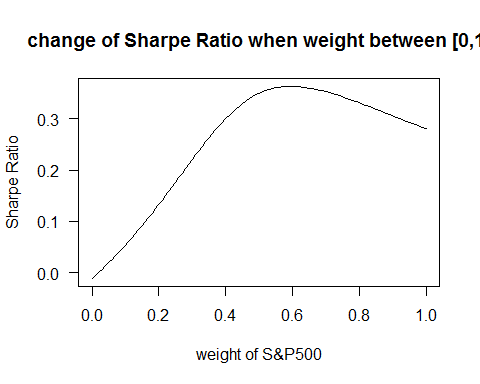
sr.etf

## [1] -0.01095925

I think S&P500 is better. The Sharpe ratio of S&P500 (0.2807) is greater than zero, which means taking risk on every unit, investor can earn 0.28 unit. However, the Sharpe ratio of ETF assets is less than zero. Take more risk means lose more on this asset, though loss on every unit risk is slightly far from zero.

## Question 7

# function  
sr <- function(x=0.5,r1=rrr1,r2=rrr2,rf=y){  
   
 if((x>=0)&(x<=1)){  
   
 # S0:obtain portfolio return  
 rp <- x\*r1 + (1-x)\*r2  
 # denote return number as m  
 m <- length(rp)  
   
 # S1:excess returns  
 ep <- rep(NA, times=m)  
   
 for (i in 2:m){  
  
 ep[i] <- rp[i]-(y[i-1]/52)  
  
 }#end for loop 1  
 rm(i)  
  
 # S2:convert into index  
 g1 <- 100  
 grp <- rep(NA, times = m)  
   
 for (j in 1:570){  
   
 if(j==1){  
 grp[j] <- g1  
 }else{  
 grp[j] <- grp[j-1]\*(1+ep[j])  
 }# end inner if  
   
 }#end for loop 2  
 rm(j)  
  
 # S3:compute number of years  
 n <- (m-1)/52  
  
 # S4:CAGR  
 cagr.rp <- ((grp[m]/g1)^(1/n))-1  
  
 # S5:annualized volatility  
 vrp <- sqrt(52)\*sd(ep,na.rm = TRUE)  
  
 # S6:Sharpe Ratio  
 sr.rp <- cagr.rp/vrp  
  
 # return(list("Sharpe Ratio"=sr.rp,"portfolio weight"=x))  
 return(sr.rp)  
   
 }else(  
   
 return("Error!Weights should be between 0 and 1.")  
   
 ) # end if/else  
   
}#end of function  
  
# when weight input is a vector  
# set :SIMPLIFY as "TRUE", the output will converge to a vector rather than print separately  
sr.vec <- Vectorize(sr, vectorize.args = "x", SIMPLIFY = TRUE)  
# plot the Sharpe Ratio's change  
# training set data is named maaa  
# since the default value is from maaa when we write the function before vectorizing, parameters are not specified in curve  
curve(sr.vec, from = 0, to = 1,  
 xlab = "weight of S&P500", ylab = "Sharpe Ratio", las = TRUE,  
 main = "change of Sharpe Ratio when weight between [0,1]")



The curve shows an upside down "U" shape, and Sharpe Ratio is highest when the proportion of S&P500 is around 0.6.

## Question 8

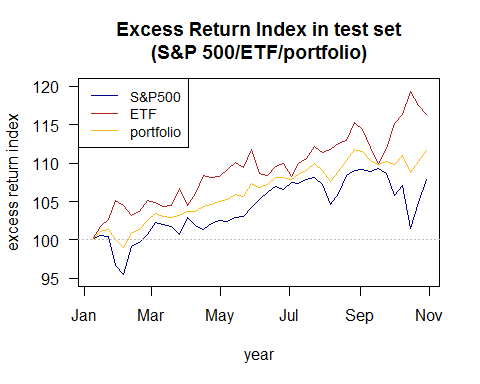
optimize(sr.vec, lower = 0, upper = 1, maximum = TRUE)

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

From the optimization, the proportion of S&P500 in the portfolio that derives largest Sharpe Ratio is 0.5959, the corresponding Sharpe Ratio is 0.3634. Therefore, 59.59% of the fund should be used to invest in S&P500, and 40.01& should be invest in ETF bond, in order to get the largest Sharpe Ratio. When we invest in S&P500 only (in which case the weight is 1), the Sharpe Ratio is 0.2807. When we invest in ETF only (in which case the weight is 0), the Sharpe Ratio is -0.011. Sharpe ratio is a metrics to evaluate the performance on an asset when take one unit risk. A larger number would be better. Therefore I would recommend to invest a portfolio with 59.59% in S&P500 and the rest be ETF.

## Question 9

######  
# S1 : convert federal interest rate into decimal form  
######  
rf <- bbb$fed.rate/100  
  
######  
# S2 : compute returns for S&P500 and ETF  
######  
# convert rate into decimal  
spy <- bbb$close.spy/100  
tlt <- bbb$close.tlt/100  
rrr3 <- rep(NA, times = length(spy))  
rrr4 <- rep(NA, times = length(tlt))  
  
# return for S&P500 in test set  
for (i in 2:length(spy)){  
 rrr3[i] <- (spy[i]-spy[i-1])/spy[i-1]  
} # end for loop  
  
# return for ETF in test set  
for (j in 2:length(tlt)){  
 rrr4[j] <- (tlt[j]-tlt[j-1])/tlt[j-1]  
} # end for loop  
rm(i,j)  
  
# add return series to test set (bbb)  
mbbb <- cbind(bbb,spy,tlt,rf)  
  
# compute the return of portfolio  
w <- optimize(sr.vec, lower = 0, upper = 1, maximum = TRUE)$maximum  
port <- w\*rrr3 + (1-w)\*rrr4  
# add portfolio return to test set (bbb)  
mbbb <- cbind(mbbb, port)  
  
######  
# S3 : compute excess returns index  
######  
# S3-1:excess returns  
q <- nrow(mbbb)  
e3 <- rep(NA, times=q)  
e4 <- rep(NA, times=q)  
e5 <- rep(NA, times=q)  
 for (i in 2:q){  
 e3[i] <- rrr3[i]-rf[i-1]/52  
 e4[i] <- rrr4[i]-rf[i-1]/52  
 e5[i] <- port[i]-rf[i-1]/52  
 }#end for loop  
rm(i)  
  
# S3-2:convert into index  
g1 <- 100  
g3 <- rep(NA, times = q)  
g4 <- rep(NA, times = q)  
g5 <- rep(NA, times = q)  
 for (i in 1:q){  
 if(i==1){  
 g3[i] <- g1  
 g4[i] <- g1  
 g5[i] <- g1  
 }else{  
 g3[i] <- g3[i-1]\*(1+e3[i])  
 g4[i] <- g4[i-1]\*(1+e4[i])  
 g5[i] <- g5[i-1]\*(1+e5[i])  
 }# end if  
 }#end for loop  
rm(i)  
# add index into same test set  
mbbb <- cbind(mbbb, g3,g4,g5)  
  
######  
# S4 : plot three excess return index on one plot  
######  
plot(mbbb$date, mbbb$g3, las=TRUE,  
 type = "l", col="navy", lwd=1.8,  
 ylim = c(95,120),  
 xlab = "year", ylab = "excess return index",  
 main = "Excess Return Index in test set\n(S&P 500/ETF/portfolio)")  
lines(mbbb$date, mbbb$g4, lwd=1.8, col="firebrick")  
lines(mbbb$date, mbbb$g5, lwd=1.8, col="darkgoldenrod1")  
# add legend to identify different lines by color  
legend("topleft", legend=c("S&P500","ETF","portfolio"), box.lwd = 0,  
 col=c("navy","firebrick","darkgoldenrod1"), lty=1, lwd=1, cex=0.9)  
# add abline  
abline(h=100, col="gray",lty=3)



Overall, the three line has a similar trend, starting from 100, the indexes all went up with fluctuation. And end up with each high point at the end of 2014. ETF performed best among the three assets. It is the only one that had index above benchmark, 100, all the time. Portfolio with highest Sharpe Ratio also did well, however there is one week when its excess return index is below 100. Generally, the fluctuation of it is always smaller than S&P500 when there is an obvious movement for all assets' excess return. S&P500 had several weeks when the excess return ratio is less than 100, but later in 2014, the return kept increasing in a slight way.

## Question 10

#combine excess return index  
index.test <- cbind(g3,g4,g5)  
colnames(index.test) <- c("S&P500", "ETF","portfolio")  
  
# calculate the excess at the end of 2014  
index.test[nrow(index.test),]

## S&P500 ETF portfolio   
## 107.8763 116.3760 111.6367

After extracting the excess return at the last week of 2014, we can see ETF produced a better job. We can earn 116.38 dollar in addition to the risk-free interest rate if we invest 100 dollar at the start of 2014. However that of the same investment in portfolio and S&P500 is 111.64 and 107.88 dollar respectively. In conclusion, it seems the portfolio does not perform well in the test set. We can also tell from the plot above, where ETF's line is always at the top of other two lines.