Are put call ratio useful statistically?

Li Yuqiong

library(ggplot2)  
library(tidyr)

### import data  
setwd('/Users/yuqiongli/desktop/python/finance/sentiment0808')  
returns <- read.csv('returns.csv', header = TRUE, sep = ",") # note the address  
returns$Securities <- as.factor(returns$Securities)  
#typeof(returns$Date) -- Integer. Import from Excel, note starting date  
returns$Date <- as.Date(returns$Date, format = '%d/%m/%Y') # note the Uppercase Y

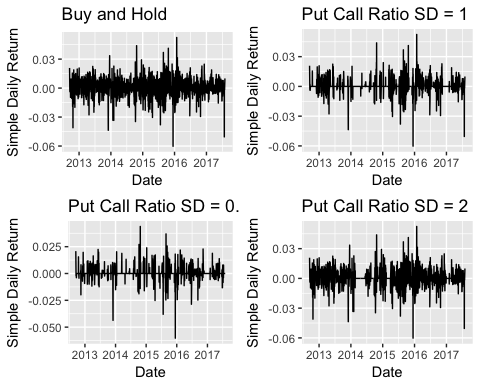
## Exploratory analysis

### Line Plots

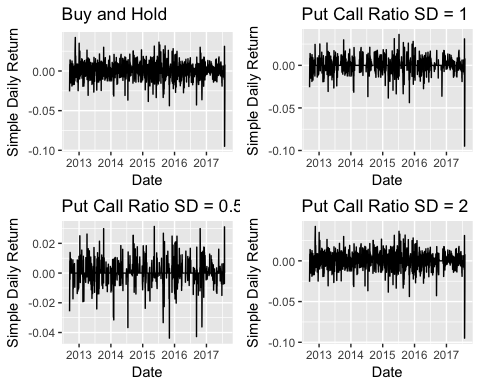
# Multiple plot function  
#  
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)  
# - cols: Number of columns in layout  
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.  
#  
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),  
# then plot 1 will go in the upper left, 2 will go in the upper right, and  
# 3 will go all the way across the bottom.  
# Source URL: http://www.cookbook-r.com/Graphs/Multiple\_graphs\_on\_one\_page\_(ggplot2)/   
#  
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {  
 library(grid)  
  
 # Make a list from the ... arguments and plotlist  
 plots <- c(list(...), plotlist)  
  
 numPlots = length(plots)  
  
 # If layout is NULL, then use 'cols' to determine layout  
 if (is.null(layout)) {  
 # Make the panel  
 # ncol: Number of columns of plots  
 # nrow: Number of rows needed, calculated from # of cols  
 layout <- matrix(seq(1, cols \* ceiling(numPlots/cols)),  
 ncol = cols, nrow = ceiling(numPlots/cols))  
 }  
  
 if (numPlots==1) {  
 print(plots[[1]])  
  
 } else {  
 # Set up the page  
 grid.newpage()  
 pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))  
  
 # Make each plot, in the correct location  
 for (i in 1:numPlots) {  
 # Get the i,j matrix positions of the regions that contain this subplot  
 matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))  
  
 print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,  
 layout.pos.col = matchidx$col))  
 }  
 }  
}

Plot the returns for three types of securities.

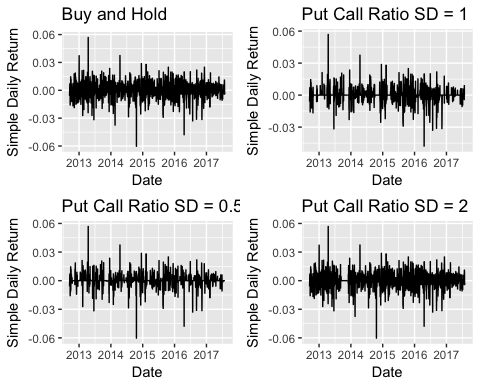
MMM <- returns[returns$Securities=='MMM', ]  
p1 <- ggplot(data = MMM, aes(x = Date, y = Buy\_Hold)) + geom\_line() +   
 ggtitle("Buy and Hold") + labs(y="Simple Daily Return")  
p2 <- ggplot(data = MMM, aes(x = Date, y = SD\_05)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 0.5") +labs(y="Simple Daily Return")  
p3 <- ggplot(data = MMM, aes(x = Date, y = SD\_1)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 1") + labs(y="Simple Daily Return")  
p4 <- ggplot(data = MMM, aes(x = Date, y = SD\_2)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 2") + labs(y="Simple Daily Return")  
multiplot(p1, p2, p3, p4, cols=2)



MO <- returns[returns$Securities=='MO', ]  
p1 <- ggplot(data = MO, aes(x = Date, y = Buy\_Hold)) + geom\_line() +   
 ggtitle("Buy and Hold") + labs(y="Simple Daily Return")  
p2 <- ggplot(data = MO, aes(x = Date, y = SD\_05)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 0.5") +labs(y="Simple Daily Return")  
p3 <- ggplot(data = MO, aes(x = Date, y = SD\_1)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 1") + labs(y="Simple Daily Return")  
p4 <- ggplot(data = MO, aes(x = Date, y = SD\_2)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 2") + labs(y="Simple Daily Return")  
multiplot(p1, p2, p3, p4, cols=2)



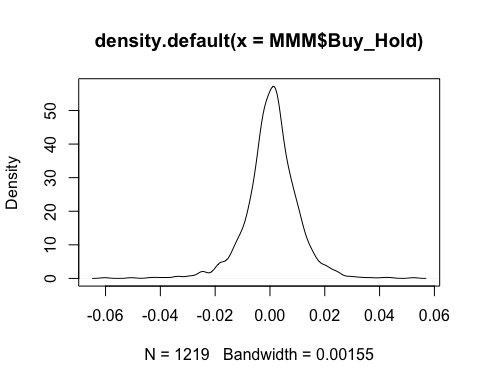
KO <- returns[returns$Securities=='KO', ]  
p1 <- ggplot(data = KO, aes(x = Date, y = Buy\_Hold)) + geom\_line() +   
 ggtitle("Buy and Hold") + labs(y="Simple Daily Return")  
p2 <- ggplot(data = KO, aes(x = Date, y = SD\_05)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 0.5") +labs(y="Simple Daily Return")  
p3 <- ggplot(data = KO, aes(x = Date, y = SD\_1)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 1") + labs(y="Simple Daily Return")  
p4 <- ggplot(data = KO, aes(x = Date, y = SD\_2)) + geom\_line() +   
 ggtitle("Put Call Ratio SD = 2") + labs(y="Simple Daily Return")  
multiplot(p1, p2, p3, p4, cols=2)



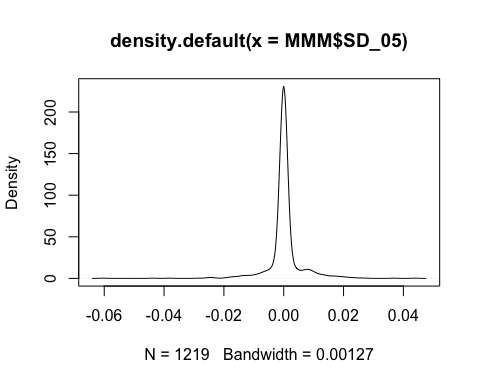
### Test for normality to see if can use ANOVA

#### First have a look at density

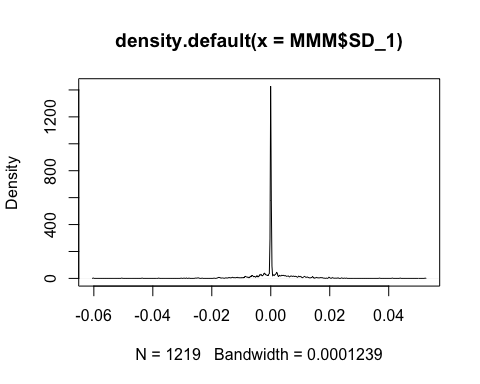
plot(density(MMM$Buy\_Hold))



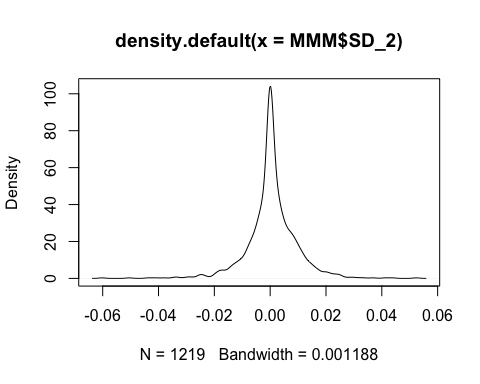
plot(density(MMM$SD\_05))



plot(density(MMM$SD\_1))



plot(density(MMM$SD\_2))

 Only Buy and Hold "looks" like a normal curve.

shapiro.test(MMM$Buy\_Hold)

##   
## Shapiro-Wilk normality test  
##   
## data: MMM$Buy\_Hold  
## W = 0.94898, p-value < 2.2e-16

shapiro.test(MMM$SD\_05)

##   
## Shapiro-Wilk normality test  
##   
## data: MMM$SD\_05  
## W = 0.61094, p-value < 2.2e-16

shapiro.test(MMM$SD\_1)

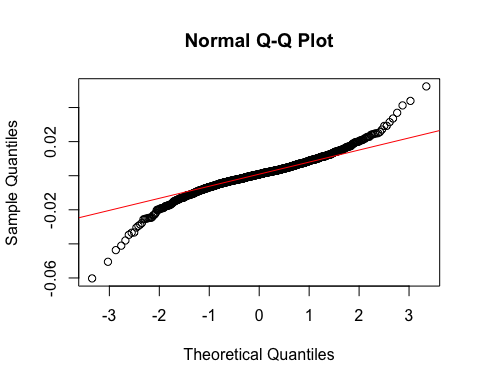
##   
## Shapiro-Wilk normality test  
##   
## data: MMM$SD\_1  
## W = 0.75122, p-value < 2.2e-16

shapiro.test(MMM$SD\_2)

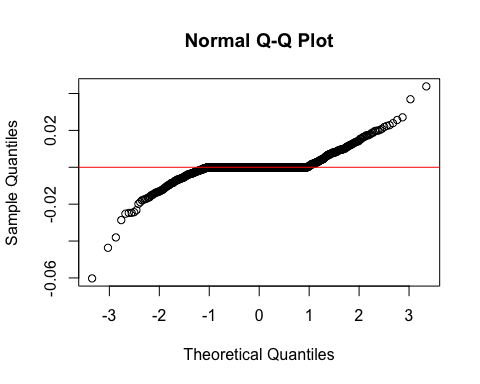
##   
## Shapiro-Wilk normality test  
##   
## data: MMM$SD\_2  
## W = 0.90921, p-value < 2.2e-16

For all four tests, p-value < 0.05 suggest that none of the four distributions are normal.

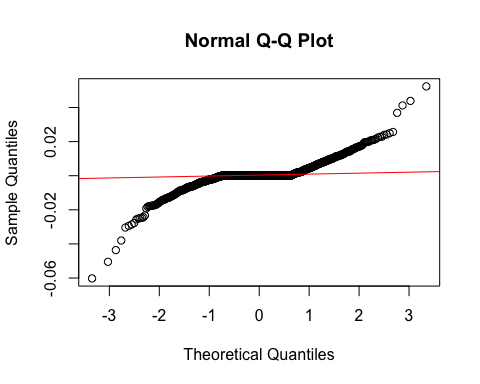
## Plot using a qqplot  
qqnorm(MMM$Buy\_Hold);qqline(MMM$Buy\_Hold, col = 2)



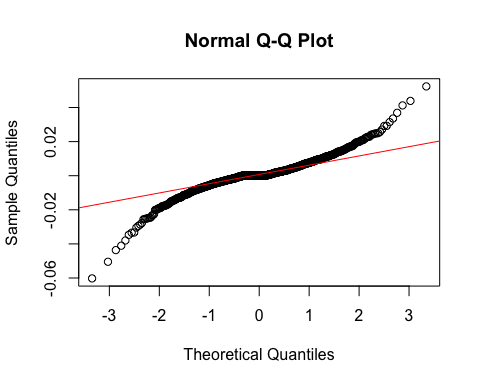
qqnorm(MMM$SD\_05);qqline(MMM$SD\_05, col = 2)



qqnorm(MMM$SD\_1);qqline(MMM$SD\_1, col = 2)

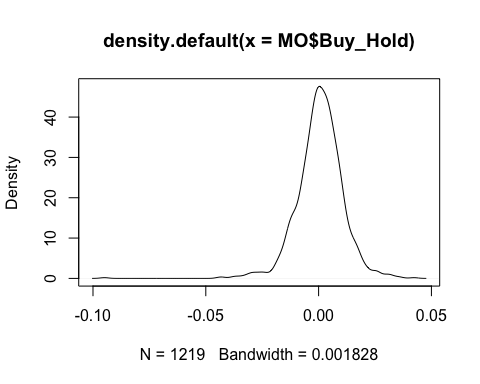


qqnorm(MMM$SD\_2);qqline(MMM$SD\_2, col = 2)

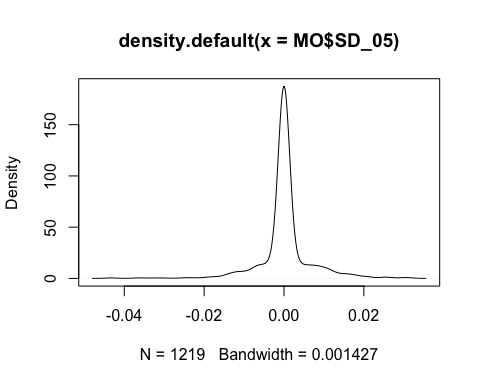
 Conclusion is that the normality assumptions of return sequences are violated. We have two options here. First, we can continue to use ANOVA with the possibility of a increased false positives. Also, it is possible to do a non-parametric test that does not require normality assumption, but the risk is the four distributions do not like the same from the above graph. Comparing these two I decide to go with ANOVA anyway.

Next, repeat the process for MO US Equity.

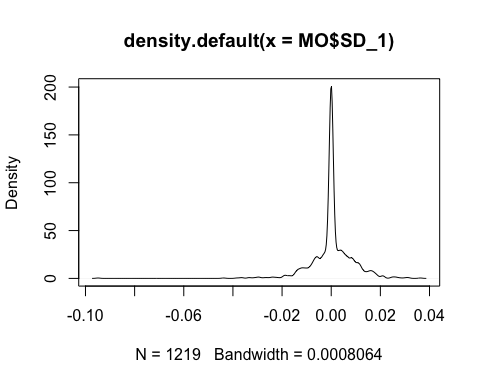
plot(density(MO$Buy\_Hold))



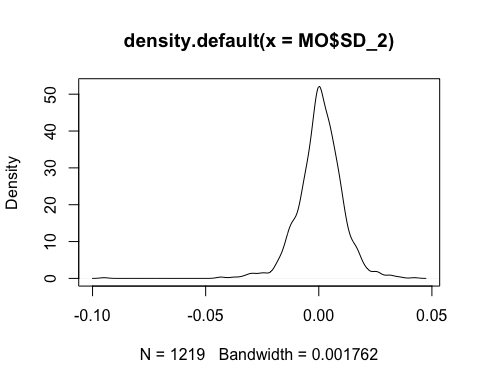
plot(density(MO$SD\_05))



plot(density(MO$SD\_1))



plot(density(MO$SD\_2))



shapiro.test(MO$Buy\_Hold)

##   
## Shapiro-Wilk normality test  
##   
## data: MO$Buy\_Hold  
## W = 0.94911, p-value < 2.2e-16

shapiro.test(MO$SD\_05)

##   
## Shapiro-Wilk normality test  
##   
## data: MO$SD\_05  
## W = 0.72552, p-value < 2.2e-16

shapiro.test(MO$SD\_1)

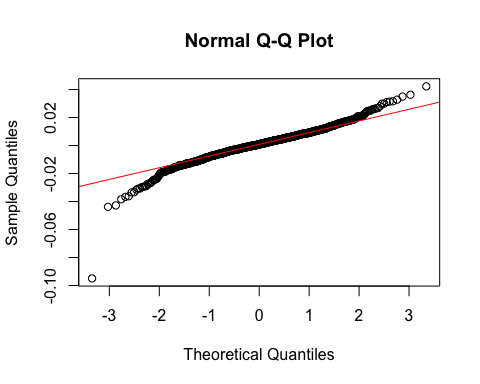
##   
## Shapiro-Wilk normality test  
##   
## data: MO$SD\_1  
## W = 0.85317, p-value < 2.2e-16

shapiro.test(MO$SD\_2)

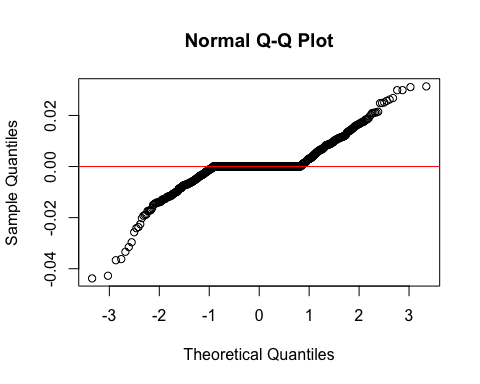
##   
## Shapiro-Wilk normality test  
##   
## data: MO$SD\_2  
## W = 0.94598, p-value < 2.2e-16

For all four tests, p-value < 0.05 suggest that none of the four distributions are normal.

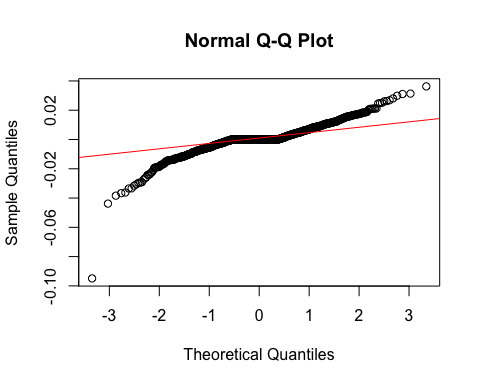
## Plot using a qqplot  
qqnorm(MO$Buy\_Hold);qqline(MO$Buy\_Hold, col = 2)



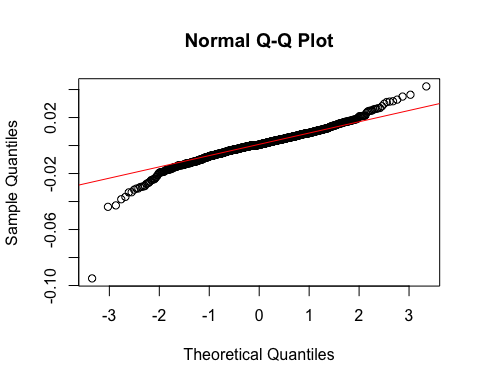
qqnorm(MO$SD\_05);qqline(MO$SD\_05, col = 2)



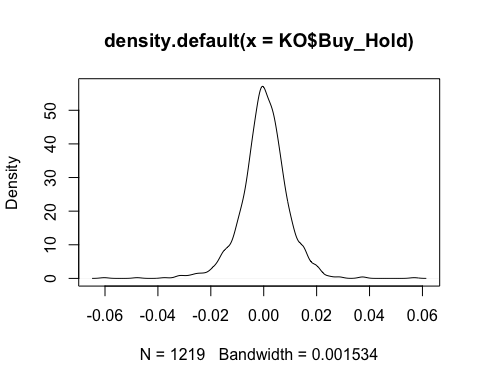
qqnorm(MO$SD\_1);qqline(MO$SD\_1, col = 2)



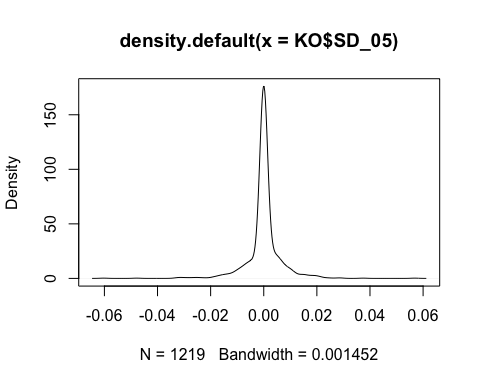
qqnorm(MO$SD\_2);qqline(MO$SD\_2, col = 2)

 Conclude that distribution is not normal and thus ANOVA might have higher false positive results.

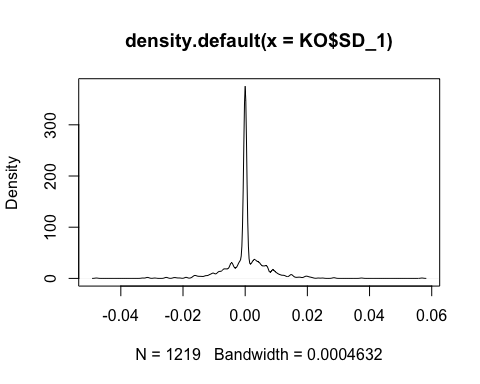
plot(density(KO$Buy\_Hold))



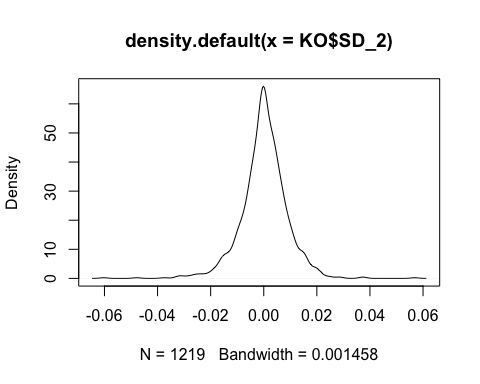
plot(density(KO$SD\_05))



plot(density(KO$SD\_1))



plot(density(KO$SD\_2))



shapiro.test(KO$Buy\_Hold)

##   
## Shapiro-Wilk normality test  
##   
## data: KO$Buy\_Hold  
## W = 0.95661, p-value < 2.2e-16

shapiro.test(KO$SD\_05)

##   
## Shapiro-Wilk normality test  
##   
## data: KO$SD\_05  
## W = 0.715, p-value < 2.2e-16

shapiro.test(KO$SD\_1)

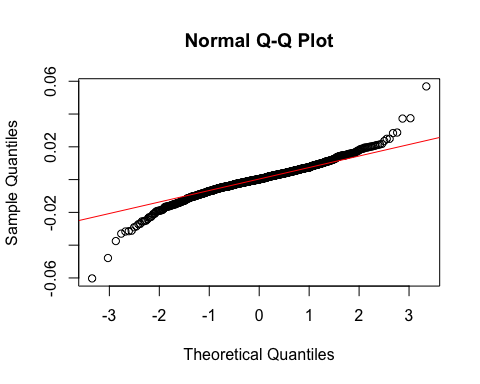
##   
## Shapiro-Wilk normality test  
##   
## data: KO$SD\_1  
## W = 0.83306, p-value < 2.2e-16

shapiro.test(KO$SD\_2)

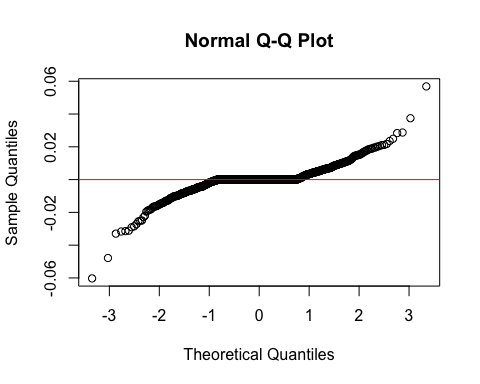
##   
## Shapiro-Wilk normality test  
##   
## data: KO$SD\_2  
## W = 0.94974, p-value < 2.2e-16

For all four tests, p-value < 0.05 suggest that none of the four distributions are normal.

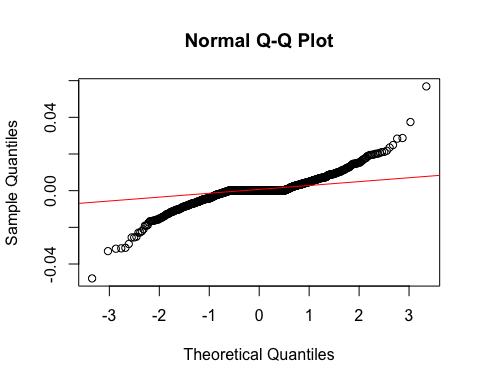
## Plot using a qqplot  
qqnorm(KO$Buy\_Hold);qqline(KO$Buy\_Hold, col = 2)



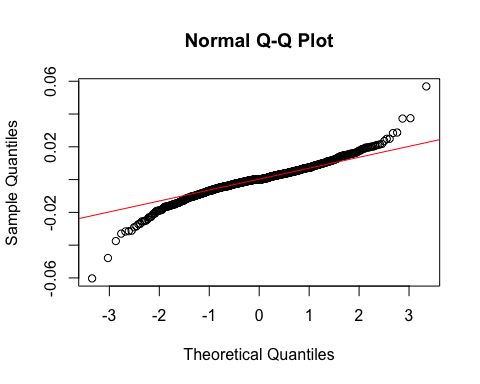
qqnorm(KO$SD\_05);qqline(KO$SD\_05, col = 2)



qqnorm(KO$SD\_1);qqline(KO$SD\_1, col = 2)



qqnorm(KO$SD\_2);qqline(KO$SD\_2, col = 2)

 Conclude that distribution is not normal and thus ANOVA might have higher false positive results.

## ANOVA test

First need to regroup data. Create a strategy variable and conversion from wide format to long format.

returns\_long <- gather(returns, strategy, daily\_return, Buy\_Hold:SD\_2, factor\_key=TRUE)

MMM\_long <- returns\_long[returns\_long$Securities=='MMM', ]  
fit <- aov(daily\_return ~ strategy, data = MMM\_long)  
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)  
## strategy 3 0.0001 4.003e-05 0.613 0.606  
## Residuals 4872 0.3180 6.528e-05

Even with an possibly larger probability of false positive, the null is not rejected because p-value not significant at 0.05 level. Accept null that the four groups have equal means, Conclusion is that for MMM US Equity, Buy and Hold and three other strategies have equal means. That is, put/call ratio strategy does not make statistiacally differences.

MO\_long <- returns\_long[returns\_long$Securities=='MO', ]  
fit <- aov(daily\_return ~ strategy, data = MO\_long)  
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)  
## strategy 3 0.0001 2.229e-05 0.275 0.844  
## Residuals 4872 0.3951 8.110e-05

Conclude that for MO US Equity put/call ratio strategy does not have significance effects on group means. Rationale similar with above.

KO\_long <- returns\_long[returns\_long$Securities=='KO', ]  
fit <- aov(daily\_return ~ strategy, data = KO\_long)  
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)  
## strategy 3 0.00002 6.310e-06 0.101 0.959  
## Residuals 4872 0.30413 6.242e-05

Conclude that for KO US Equity put/call ratio strategy does not have significance effects on group means. Rationale similar with above.