Homework 2

Vladislav Zakatov

23 October 2015

This document has been created as part of the first homework assignment at CMF.

Initially, the working directory, system locale are set and the required packages are loaded.

```
##### Initialisation #####
library(quantmod)
library(ggplot2)
library(ghyp)
Sys.setlocale("LC_ALL","English")
```

The historical prices of S&P 500 index from 2013 up to present are loaded from Yahoo Finance using *getSymbols* function from *quantmod* package.

```
tickers = c("SPY") # specify the tickers to load

# environment where data will be stored
e <- new.env()

# suppress warning about future change in the function behaviour
options("getSymbols.warning4.0"=FALSE)

# load data from Yahoo Finance
getSymbols(tickers, src = "yahoo", from="2013-01-01", env=e)

# merge data into a single xts taking only adjusted prices
data = do.call(merge, eapply(e, Ad)[tickers])
names(data) = tickers
```

Daily returns are calculated and saved as *xts*.

```
##### Calculate returns #####
rets = apply(data, 2, function(x) diff(x)/x[-length(x)])
rets = xts(rets, order.by = index(data)[-1])
```

Empirical density is calculated and plotted together with normal distribution with appropriate parameters.

```
##### Density and Q-Q plot #####
dens <- density(rets, bw = "ucv", n = 1024)

df = data.frame(x = dens$x, y = dens$y, Type = "empirical")
df1 = data.frame(x = dens$x, y = dnorm(dens$x, mean = mean(rets), sd = sd(rets)),
                 Type = "theoretical")
df = rbind(df, df1) # merge datasets into a data frame for ggplot
rm(df1)

ggplot(df, aes(x,y, colour = Type)) + geom_line(size = 1) # densitities plot
```
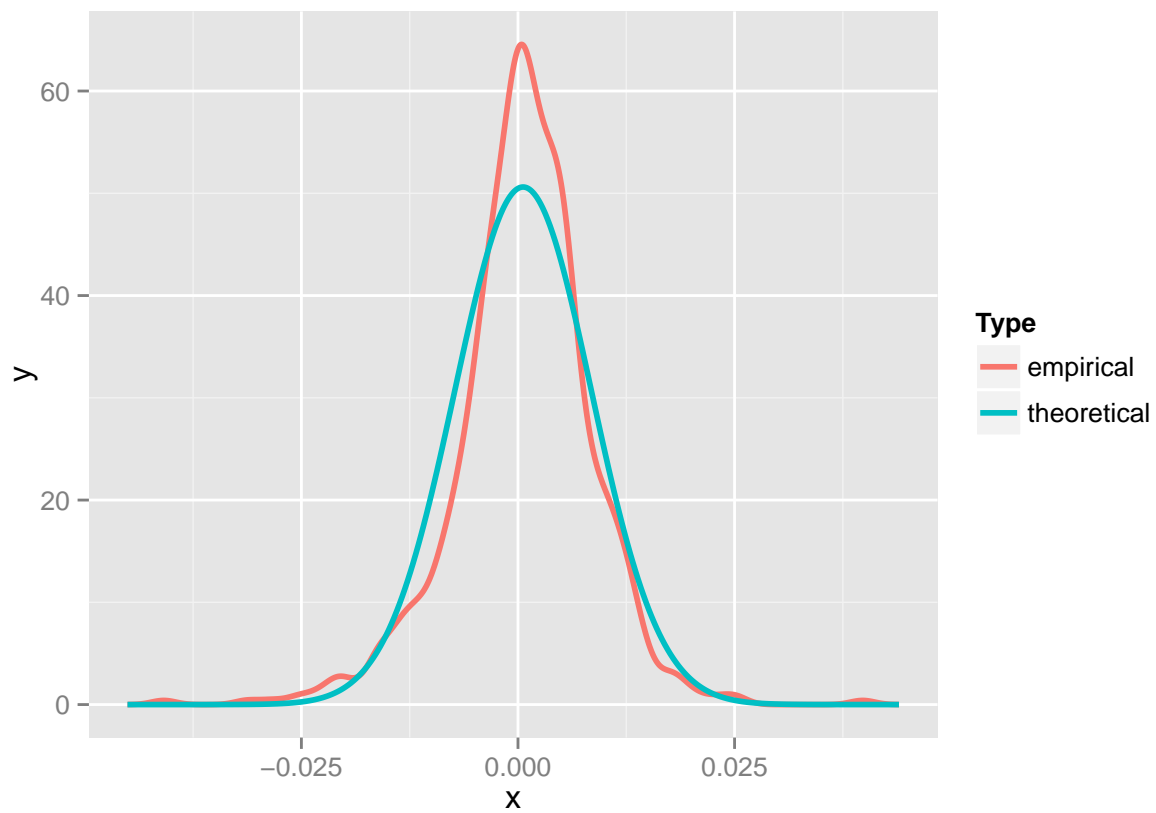
```
qqplot(rnorm(n=100^3,mean=mean(rets),sd=sd(rets)), as.numeric(rets),
       xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")
abline(0,1,lwd=2)
```
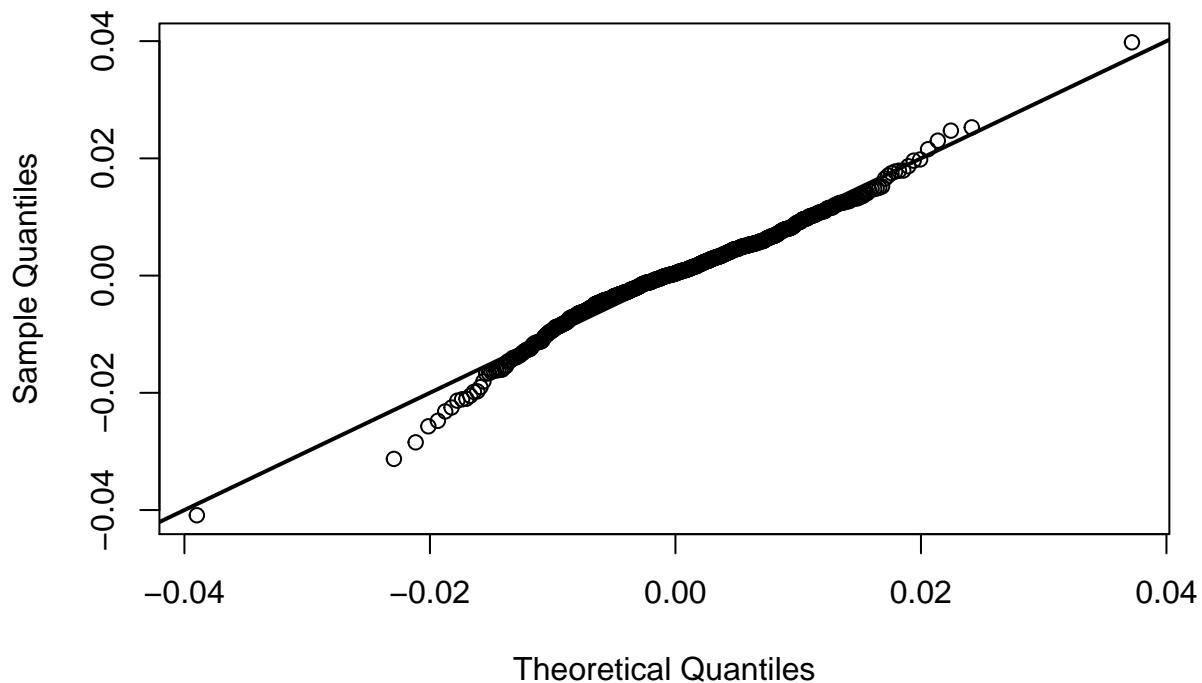
It can be seen from the graph and the Q-Q plot that the left tail of the S&P 500 Index returns distribution is much fatter than that of the normal distribution.

Tests for normality are then performed.

```
##### Normality tests #####
shapiro.test(as.numeric(rets))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(rets)
## W = 0.96862, p-value = 3.596e-11
```

```
ks.test(as.numeric(as.numeric(rets)), "pnorm", mean = mean(rets), sd = sd(rets))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(as.numeric(rets))
## D = 0.062606, p-value = 0.007835
## alternative hypothesis: two-sided
```

Both Shapiro-Wilk and Kolmogorov-Smirnov tests reject the null hypothesis that the returns are normally distributed with a large degree of confidence ($> 99\%$).

We now fit the best model from Generalized Hyberbolic Distributions class based on AIC criterion.

3

```
##### Fit GHD #####
aic.uv <- stepAIC.ghyp(rets,dist=c("ghyp", "hyp", "NIG", "VG", "t", "gauss"),
                       symmetric=NULL,silent=TRUE)
```

```
summary(aic.uv$best.model)
```

```
## Asymmetric Normal Inverse Gaussian Distribution:
##
## Parameters:
##     alpha.bar           mu          sigma          gamma
##   1.104524543  0.001960849  0.007776621 -0.001371890
##
## Call:
## stepAIC.ghyp(data = rets, dist = c("ghyp", "hyp", "NIG", "VG",      "t", "gauss"), symmetric = NULL, s
##
## Optimization information:
## log-Likelihood:              2456.206
## AIC:                         -4904.412
## Fitted parameters:           alpha.bar, mu, sigma, gamma;  (Number: 4)
## Number of iterations:        189
## Converged:                   TRUE
```
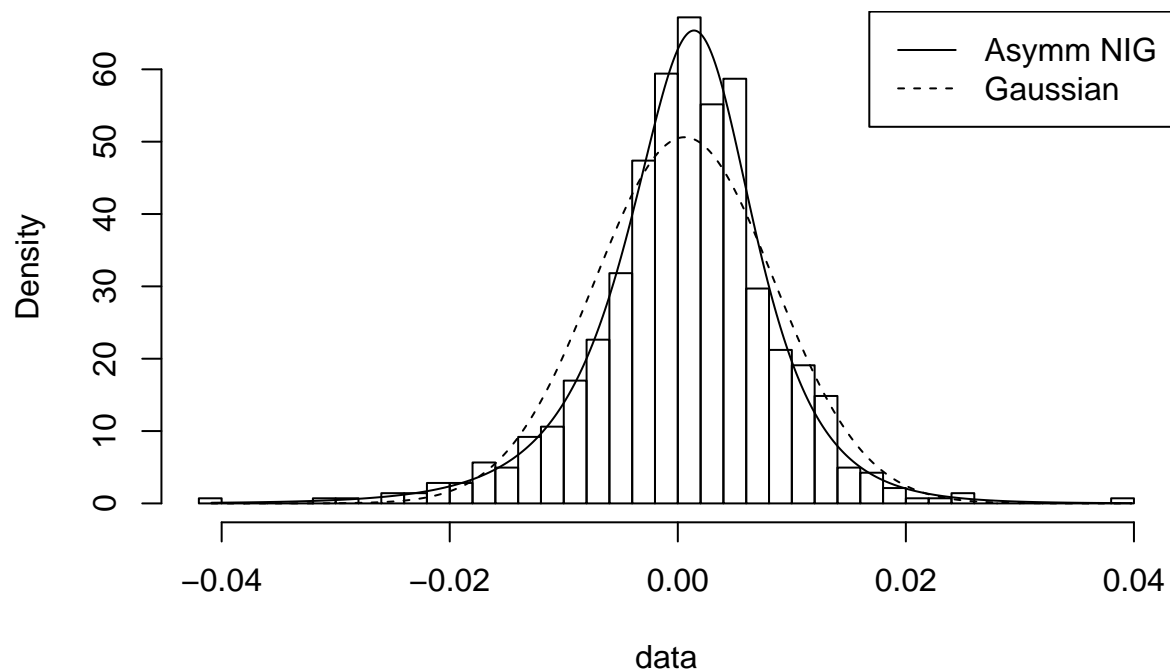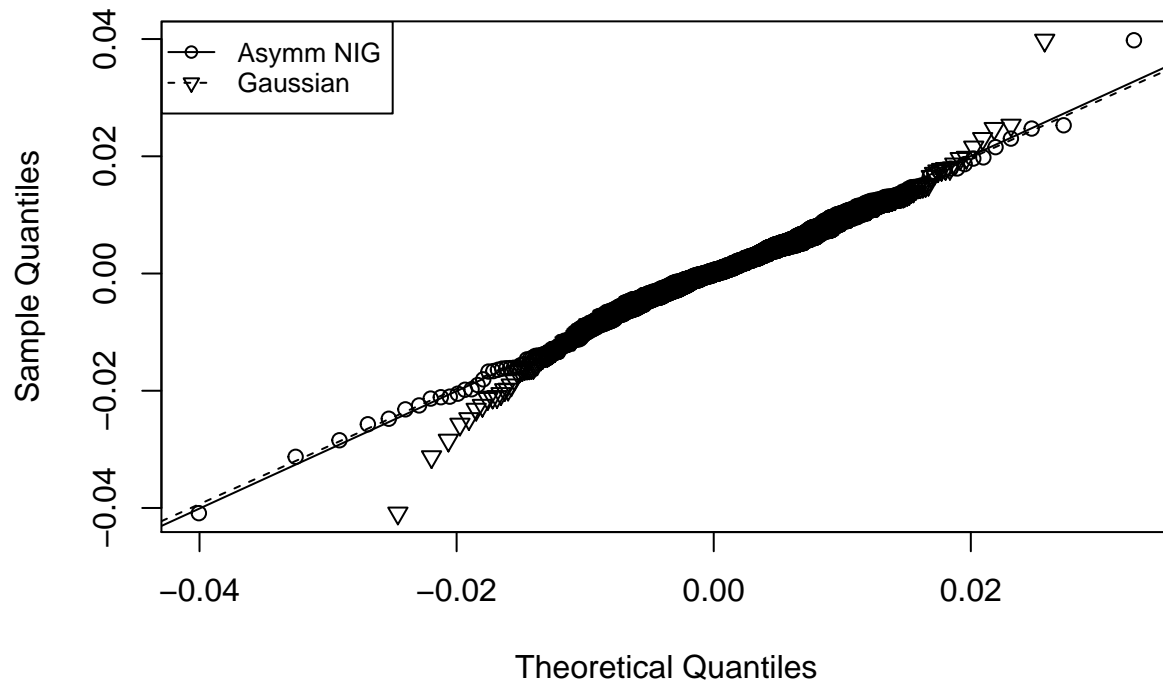
```
hist(aic.uv$best.model)
```


Histogram of data

```
qqghyp(aic.uv$best.model)
```

Generalized Hyperbolic Q–Q Plot

The best model was estimated to be the Asymmetric Normal Inverse Gaussian Distribution. It can be easily seen from the distribution plot and Q-Q plot that this distribution fits the data much better, with almost all the points lying on the straight lying passing with zero intercept and slope of 1.

```
rets.ghyp = fit.ghypuv(rets, symmetric = FALSE, silent = TRUE)

lik.ratio.test(rets.ghyp,aic.uv$best.model,conf.level=0.95)
```

```
## $statistic
##         L
## 0.9973635
##
## $p.value
## [1] 0.9420736
##
## $df
## [1] 1
##
## $H0
## [1] TRUE
```

The likelihood ratio test fails to reject the null hypothesis that the more generalized model (GHD) has the same explanatory power as the best fit model based on AIC criterion, hence we can proceed with modelling S&P 500 index returns with the Asymmetric Normal Inverse Gaussian Distribution.

We now calculate value at risk for the whole period (2013 – present)

```r
##### Calculate VaR #####
alpha <- 0.1
N <- 10^6

rets.sim <- rghyp(n=N,object=aic.uv$best.model)
rets.sim <- sort(rets.sim)
VaR <- rets.sim[alpha*N]
ES <- mean(rets.sim[1:(alpha*N-1)])

print(VaR)
```

```
## [1] -0.008645906
```

```r
print(ES)
```

```
## [1] -0.01456473
```

The esimated value at risk is about -0.86% and expected shortfall is equal to -1.46%.

We now calculate estimate the VaR curve with window of length 130 trading days.

```r
##### Calculate VaR curve #####
VaR_curve <- numeric()
h <- 0.5 * 260 # window length

for (i in (h+1):length(rets))
{
  h.rets <- rets[(i-h):(i-1)]
  rets.fit <- stepAIC.ghyp(h.rets,dist=c("ghyp", "hyp", "NIG", "VG", "t", "gauss"),
                           symmetric=NULL,silent=TRUE)
  VaR_curve[i-h] <- qghyp(alpha,object=rets.fit$best.model)
}
```

```r
fact <- rets[(h+1):length(rets)]
plot(as.numeric(fact),type="l", ylab = "Return")
lines(VaR_curve,col="red")
```
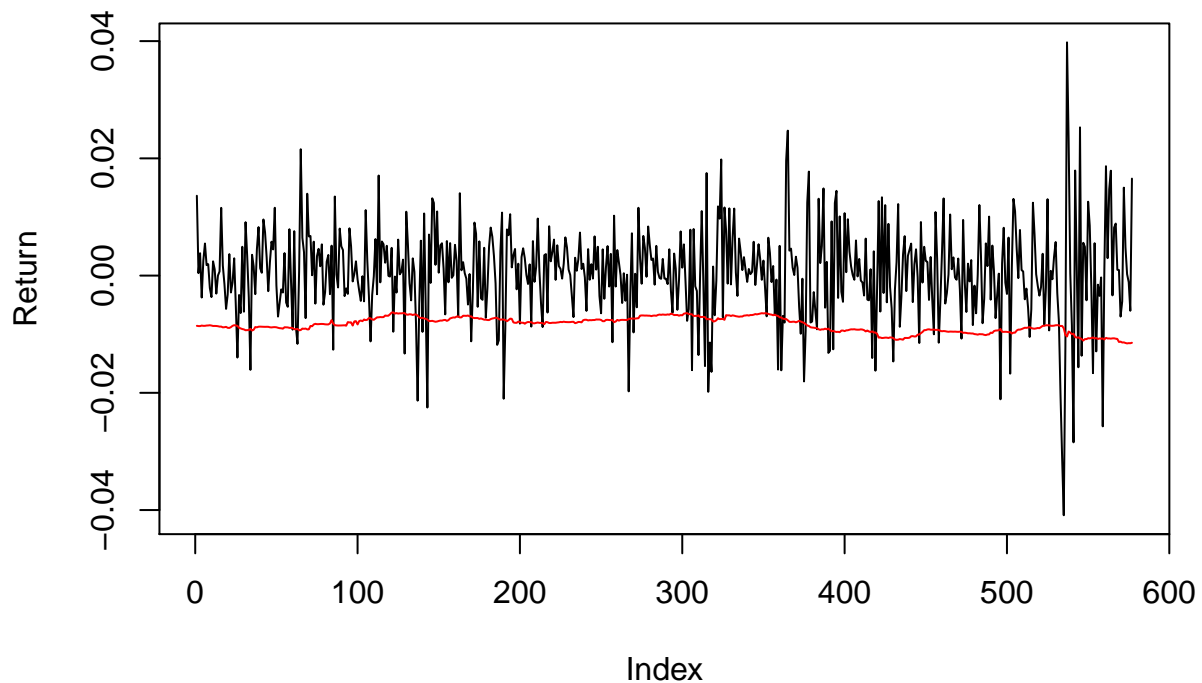
The graph shows that the resulting VaR curve is almost straight, which provides evidence that our calculations of VaR were correctly modelled and reliable

To prove the relevance of the VaR curve we use the Kupiec test to check whether the number of actual returns lower than VaR is not greater than it is to be expected.

```
##### Kupiec test #####
K <- sum(fact<VaR_curve); alpha0 <- K/length(rets)
S <- -2*log((1-alpha)^(length(rets)-K)*alpha^K)+
  2*log((1-alpha0)^(length(rets)-K)*alpha0^K)
p.value <- 1-pchisq(S,df=1)

##### Loss functions #####
L.Lo <- sum((fact-VaR)^2*(fact<VaR))/K
L.BI <- sum((fact-VaR)/VaR*(fact<VaR))/K

print(alpha)
```

```
## [1] 0.1
```

```
print(alpha0)
```

```
## [1] 0.09052334
```

```r
print(p.value)
```

```
## [1] 0.394084
```

```r
print(L.Lo)
```

```
## [1] 6.70238e-05
```

```r
print(L.BI)
```

```
## [1] 0.6420005
```

Given the inital confidence level of 10%, we get 9.05% of actual returns lower than VaR which seems to be a good approximation. The Kupiec test also fails to reject the null hypothesis with a very high p-value. We can conclude that the estimated VaR is reliable and statistically significant. We also include the results for Lopez and Blanco-Ihle loss functions, however these results are irrelevant since only one model was chosen to be the best and there is nothing to compare.