MGMTMFE407-2 HW1

Cohort 2, Group 7 - Hyeuk Jung, Jiaqi Li, Xichen Luo, Huanyu Liu

January 13, 2019

Problem 1

Q1

Let $Y = \delta_t B_t$, then we can derive that:

$$\begin{aligned} P(Y \geq y) &= P(\delta_t B_t \geq y) \\ &= P(B_t = 1)P(\delta_t B_t \geq y | B_t = 1) + P(B_t = 0)P(\delta_t B_t \geq y | B_t = 0) \\ &= pP(\delta_t \geq y) \end{aligned}$$

Since $\delta_t \sim N(0, 1)$, $Y = \delta_t B_t \sim N(0, p)$.

We also have:

$$E(\delta_t^3) = E(\varepsilon_t^3) = 0$$
$$E(\delta_t^4) = E(\varepsilon_t^4) = 3$$
$$E((\delta_t B_t)^3) = 0, E((\delta_t B_t)^4) = 3p$$

Calculate mean, vairnace, skewness, and kurtosis:

$$
\begin{aligned}
r_t &= \mu + \sigma\varepsilon_t + B_t(\mu_J + \sigma_J\delta_t) \\
&= \mu + \sigma\varepsilon_t + B_t\mu_J + B_t\sigma_J\delta_t
\end{aligned}
$$

$$
\begin{aligned}
E(r_t) &= E(\mu + \sigma\varepsilon_t + B_t\mu_J + B_t\sigma_J\delta_t) \\
&= \mu + 0 + \mu_J p + 0 = \mu + \mu_J p
\end{aligned}
$$

$$
\begin{aligned}
Var(r_t) &= Var(\mu + \sigma\varepsilon_t + B_t\mu_J + B_t\sigma_J\delta_t) \\
&= \sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2 p^2
\end{aligned}
$$

$$
\begin{aligned}
S(r_t) &= S(\mu + \sigma\varepsilon_t + B_t\mu_J + B_t\sigma_J\delta_t) \\
&= 0 + 0 + \mu_J^3 \frac{1-2p}{p(1-p)} + 0 \text{ where } E((\delta_t B_t)^3) = 0 \\
&= \mu_J^3 \frac{1-2p}{p(1-p)}
\end{aligned}
$$

$$
\begin{aligned}
K(r_t) - 3 &= K(\mu + \sigma\varepsilon_t + B_t\mu_J + B_t\sigma_J\delta_t) - 3 \\
&= 3\sigma^2 + \mu_J^4 \frac{3p^2 - 3p + 1}{p(1-p)} + 3p\sigma_j^4 - 3 \text{ where } E((\delta_t B_t)^4) = 3p
\end{aligned}
$$

Q2

For daily log returns, data is skewed and has high kurtosis. Thus, a simple log normal model is not a good model to replicate the real market.

In the Bernoulli-normal model, a jump is introduced into the model, which helps improve the model.

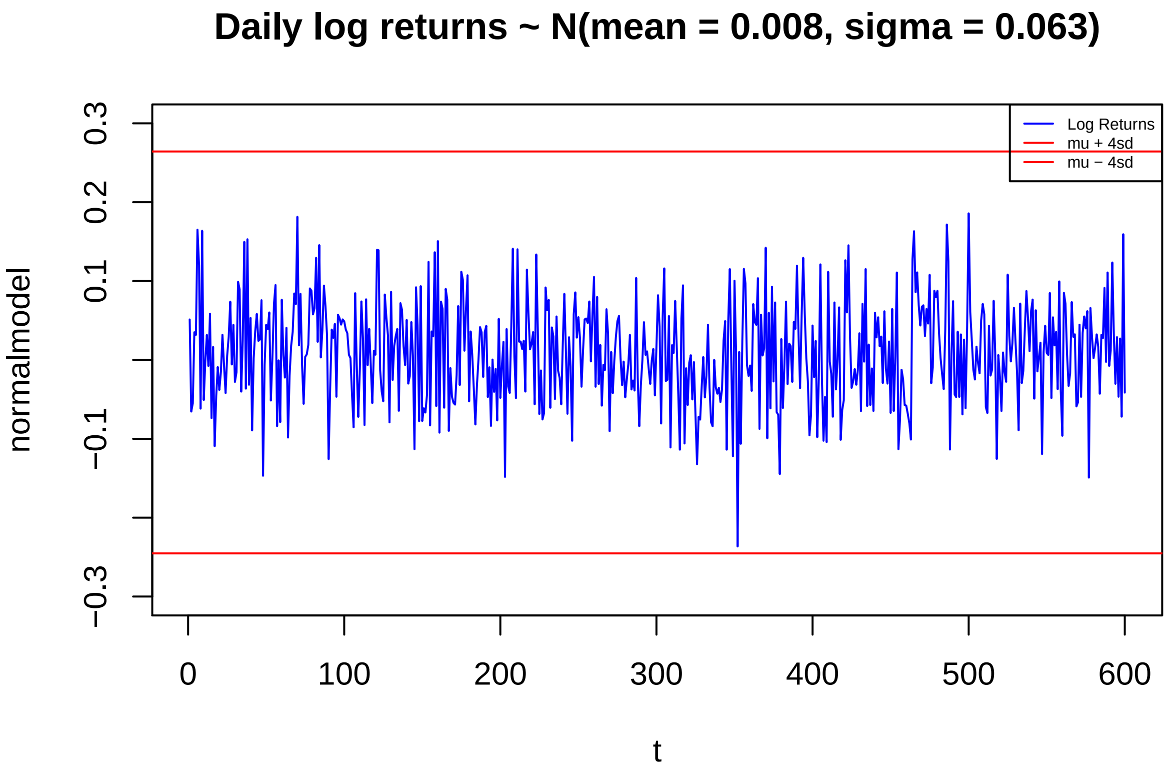
Q3

1) r ~ N(mean = 0.008, sd = 0.063), number of observations: 600 (50 years)

```
# 1) r ~ N(mean = 0.008, sd = 0.063), number of observations: 600 (50 years)
r = 0.008; sigma = 0.063; t = 1:600
sim <- function(r, sigma, t) {
  y = r + rnorm(length(t), sd = sigma)
}
normalmodel = sim(r, sigma, t)

#par(mfrow=c(1,2))
plot(t, normalmodel, pch=20, col="blue", ylim = c(-.3, .3), type = 'l')
abline(h = mean(normalmodel) + sd(normalmodel)*4, col = "red") # r + sigma*4
abline(h = mean(normalmodel) - sd(normalmodel)*4, col = "red") # r - sigma*4
```

```
title(paste("Daily log returns ~ N(mean = ", r,", sigma = ", sigma, ")", sep=""))
legend("topright", legend = c("Log Returns", "mu + 4sd", "mu - 4sd"), lty = 1, col = c("blue", "red", ":
```

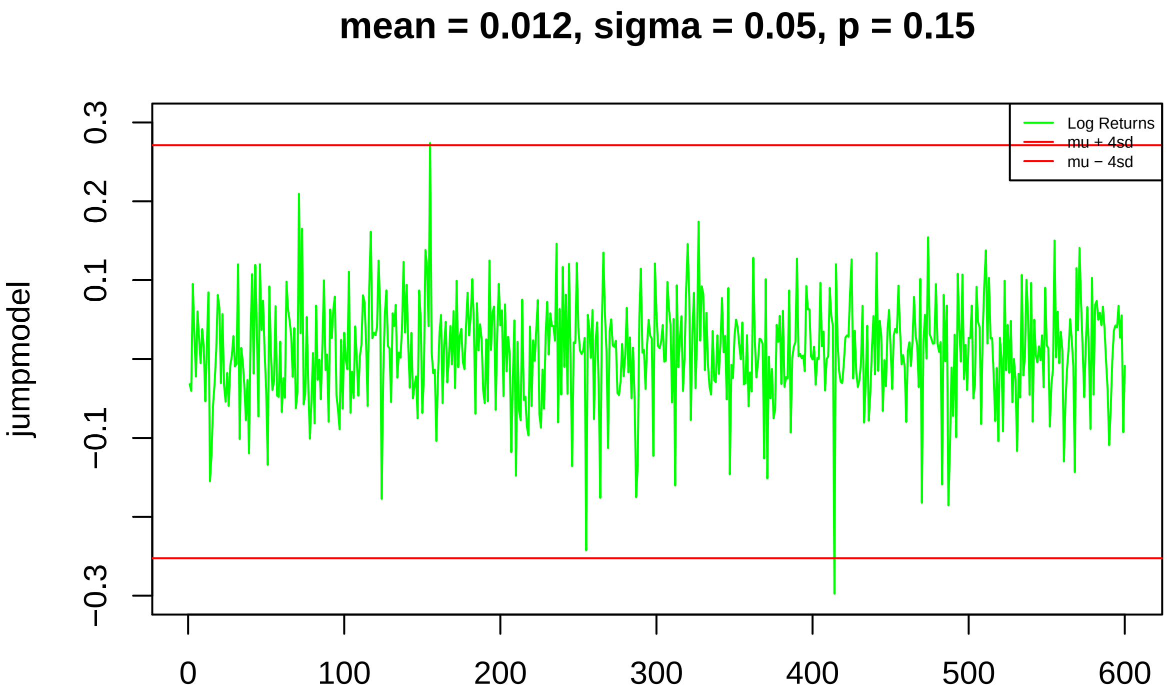


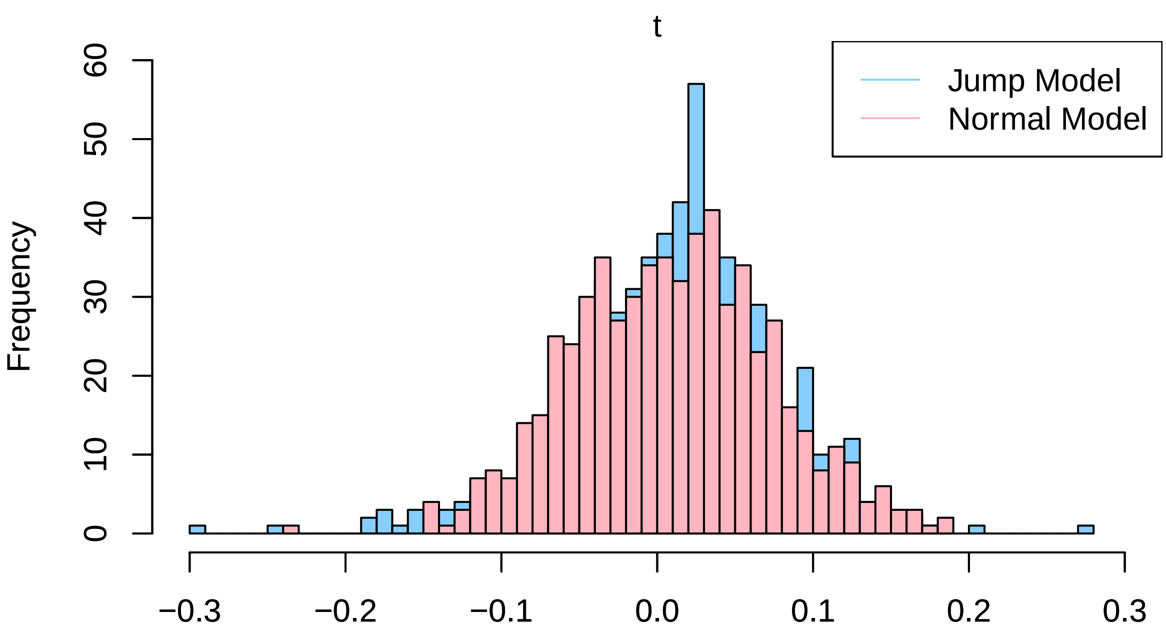
Compared to the data given in the lecture 1, all observations are within the +/- 4σ range. That being said, the simulation has no extreme case, which exists in the real market data.

2) r = mean + sd*e_t + J_t

```
# e_t ~ N(0, 1)
# J_t; mean = mean(J)*p, var = mean(J)^2*p*(1-p) + var(J)*p
mean = 0.012; sigma = 0.05; p = 0.15; mean_j = -0.03; sigma_j = 0.1
sim2 <- function(mean, sigma, p, mean_j, sigma_j, t) {
  y = mean + sigma*rnorm(length(t), sd = 1) + rbernoulli(length(t), p)*(mean_j + sigma_j*rnorm(length(t
}
jumpmodel = sim2(mean, sigma, p, mean_j, sigma_j, t)

#par(mfrow=c(1,2))
plot(t, jumpmodel, pch=20, col="green", ylim = c(-.3, .3), type = 'l')
abline(h = mean(jumpmodel) + sd(jumpmodel)*4, col = "red") # mean + mean_j*p + sigma*4
abline(h = mean(jumpmodel) - sd(jumpmodel)*4, col = "red")
title(paste("mean = ", mean,", sigma = ", sigma, ", p = ", p, sep=""))
legend("topright", legend = c("Log Returns", "mu + 4sd", "mu - 4sd"), lty = 1,
       col = c("green", "red", "red"), cex = .5)
```





Unconditional mean, var, skewness, and kurtosis of the jump model:

```
## [1] "mean =  0.00931816388993834"
```

```
## [1] "variance =  0.00428419028921374"
```

```
## [1] "skewness =  -0.354995031641882"
```

```
## [1] "kurtosis =  4.45364824684534"
```

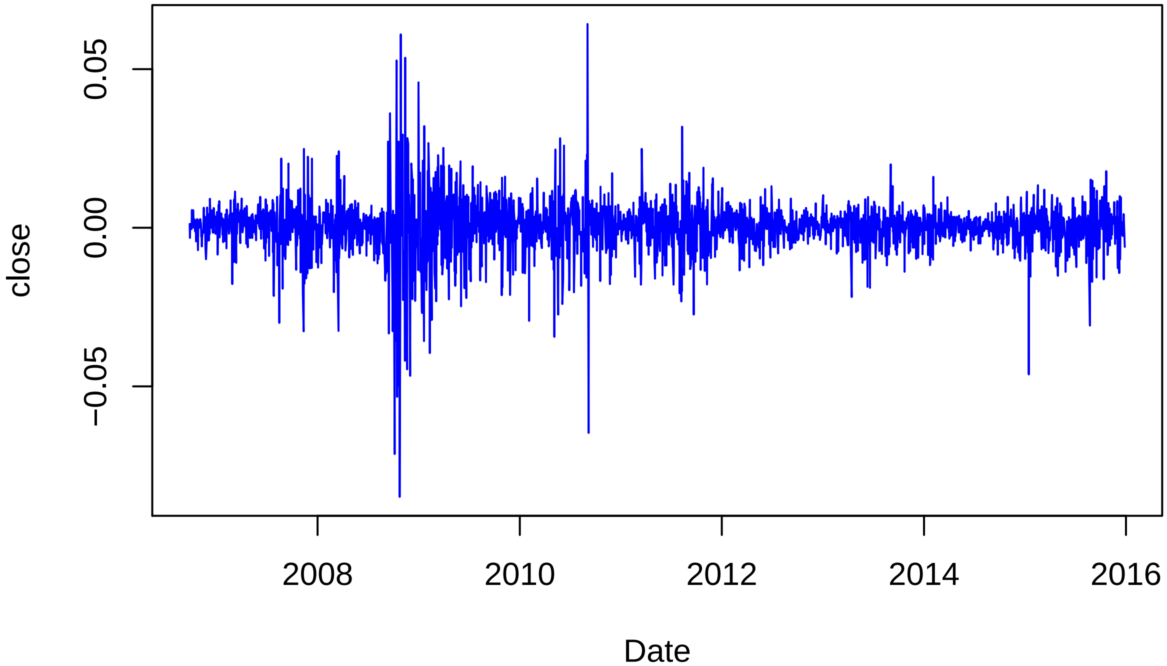
Compared to the simulation above, the simulated results are more silimar to the real market data. We can observe extreme cases reaching +/- 4σ range, which implies that the distribution of the result would have fat tails.

Problem 2

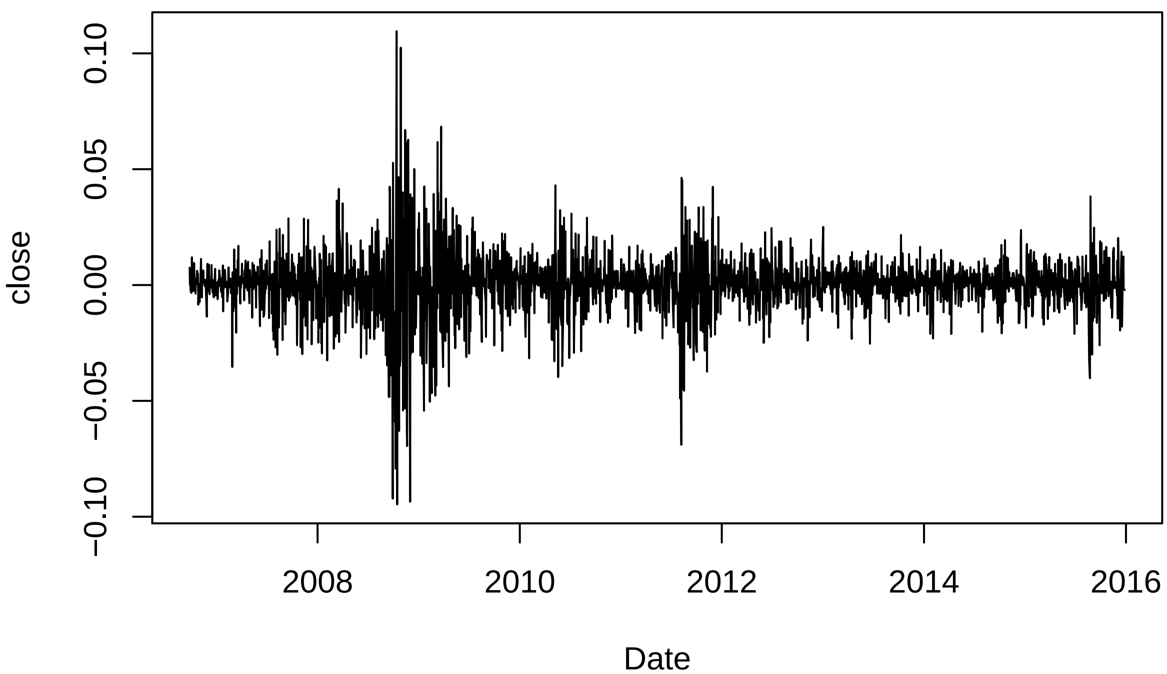
Q1. Visualizing the data

(a) Time series plots of the daily log-returns for DBV and GSPC

```
with(dbv, plot(Date, close, type = 'l', col = "blue"))
```

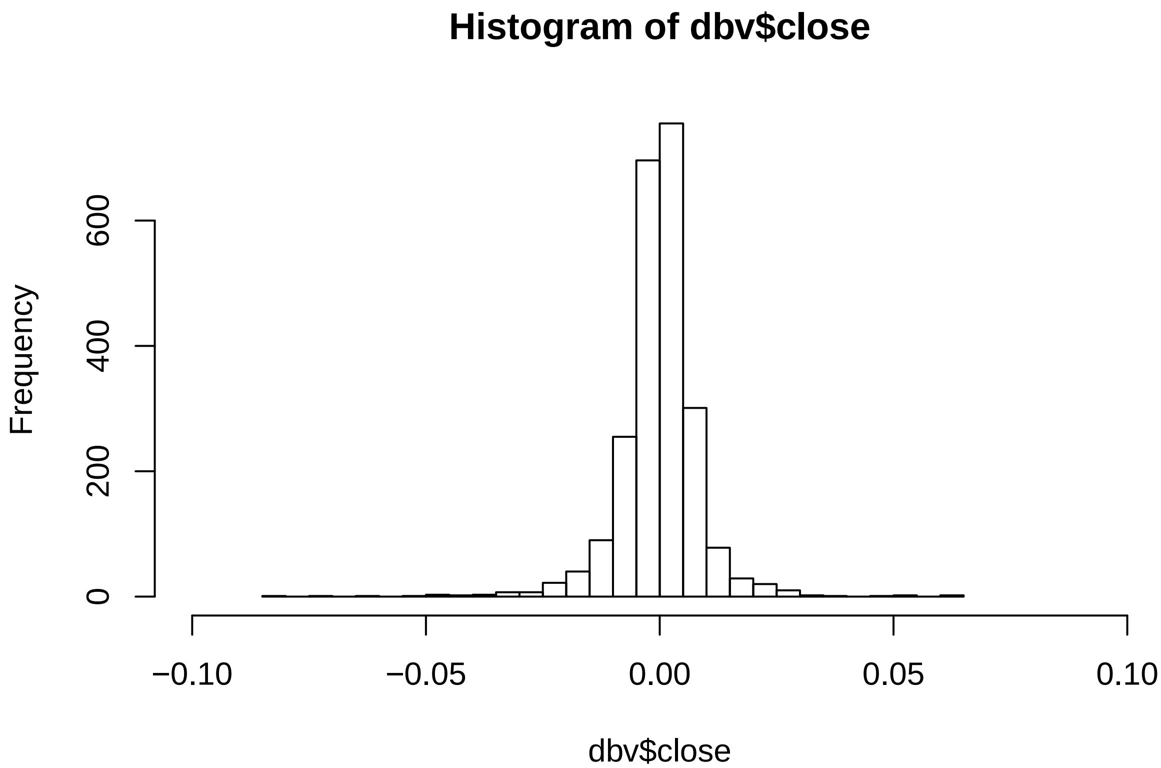



```
with(gspc, plot(Date, close, type = 'l'))
```

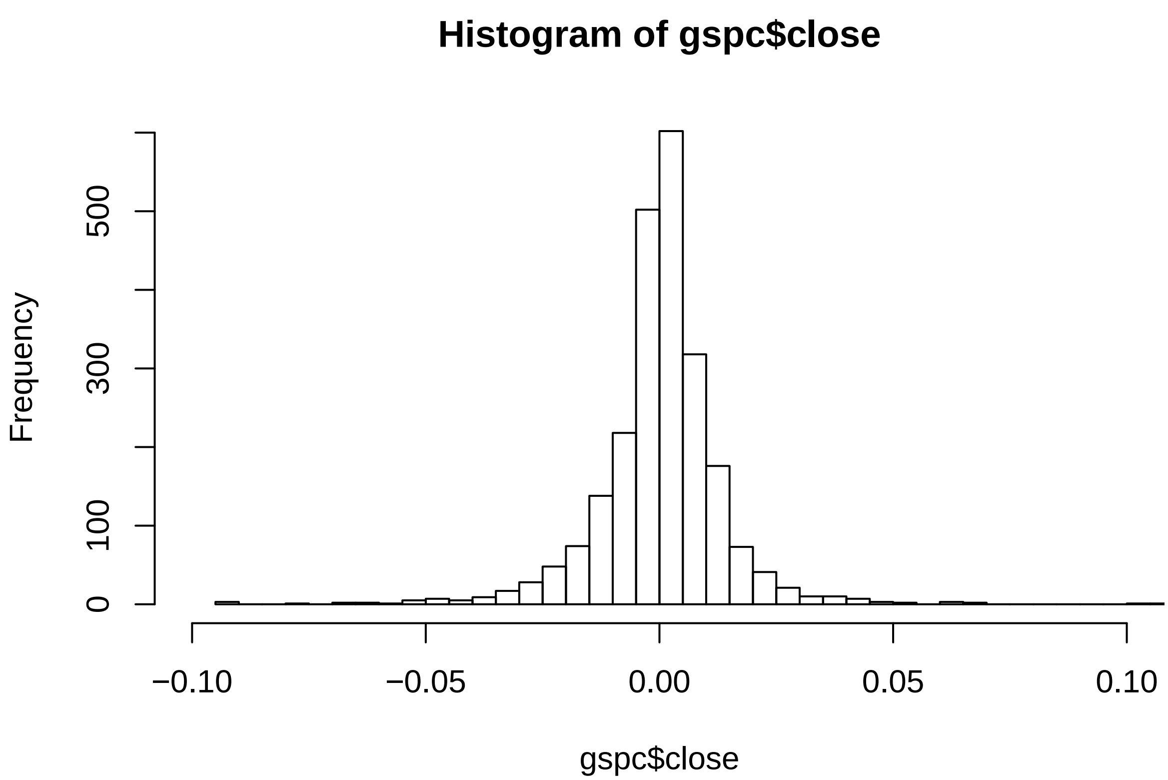


(b) Histograms of the daily log-returns for DBV and GSPC

```r
hist(dbv$close, breaks = 50, xlim = c(-.1, .1))
```




```r
hist(gspc$close, breaks = 50, xlim = c(-.1, .1))
```



Q2. Shape of Return Distribution

(a) H0: skewness = 0, a = 0.05

```r
# DBV (Deutsche Bank)
mean_dbv = mean(dbv$close)
sd_dbv = sd(dbv$close)
n_dbv = length(dbv$close)
skew_dbv <- sum((dbv$close-mean_dbv)^3/(sd_dbv^3))/(n_dbv)
skew_dbv_t <- skew_dbv / sqrt(6/n_dbv)
if(abs(skew_dbv_t) > 1.96) {
  cat(paste("|t| =", round(abs(skew_dbv_t), 4), "> 1.96: reject the null hypothesis\n (Skewness of DBV i
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## |t| = 16.6811 > 1.96: reject the null hypothesis
##  (Skewness of DBV is significantly different from zero.)
```

```r
# GSPC (S&P 500)
mean_gspc = mean(gspc$close)
sd_gspc = sd(gspc$close)
n_gspc = length(gspc$close)
skew_gspc <- sum((gspc$close-mean_gspc)^3/(sd_gspc^3))/(n_gspc)
skew_gspc_t <- skew_gspc / sqrt(6/n_gspc)
if(abs(skew_gspc_t) > 1.96) {
  cat(paste("|t| =", round(abs(skew_gspc_t), 4), "> 1.96: reject null hypothesis\n (Skewness of GSPC is
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## |t| = 6.3815 > 1.96: reject null hypothesis
##  (Skewness of GSPC is significantly different than zero.)
```

(b) H0: excess kurtosis = 0, a = 0.05

```r
# DBV (Deutsche Bank)
kurt_dbv <- sum((dbv$close-mean_dbv)^4/(sd_dbv^4))/(n_dbv) - 3 # Excess kurtosis
kurt_dbv_t <- kurt_dbv / sqrt(24/n_dbv)
if(abs(kurt_dbv_t) > 1.96) {
  cat(paste("|t| =", round(abs(kurt_dbv_t), 4), "> 1.96: reject null hypothesis\n (Kurtosis of DBV is si
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## |t| = 133.0419 > 1.96: reject null hypothesis
##  (Kurtosis of DBV is significantly different than zero.)
```

```r
# GSPC (S&P 500)
kurt_gspc <- sum((gspc$close-mean_gspc)^4/(sd_gspc^4))/(n_gspc) - 3
kurt_gspc_t <- kurt_gspc / sqrt(24/n_gspc)
if(abs(kurt_gspc_t) > 1.96) {
  cat(paste("|t| =", round(abs(kurt_gspc_t), 4), "> 1.96: reject null hypothesis\n (Kurtosis of GPSC is
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## |t| = 96.4317 > 1.96: reject null hypothesis
##  (Kurtosis of GPSC is significantly different than zero.)
```

(c) Jarque-Bera test (H0: skewwness & excess kurtosis = 0 (log return ~ Normal), a = 0.05)

```r
#jb = ( skew_dbv^2 / (6/length(dbv$close)) ) + (kurt_dbv^2 / (24/length(dbv$close)))
# DBV (Deutsche Bank)
jb_dbv <- skew_dbv_t^2 + kurt_dbv_t^2
jb_dbv_chi <- qchisq(0.95, df = 2)
if(jb_dbv > jb_dbv_chi) {
  paste(round(jb_dbv, 4), ">", round(jb_dbv_chi, 4), ": reject the null (DBV is not normally distribute
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## [1] "17978.4135 > 5.9915 : reject the null (DBV is not normally distributed.)"
```

```r
# GSPC (S&P 500)
jb_gspc <- skew_gspc_t^2 + kurt_gspc_t^2
jb_gspc_chi <- qchisq(0.95, df = 2)
if(jb_gspc > jb_gspc_chi) {
  paste(round(jb_gspc, 4), ">", round(jb_gspc_chi, 4), ": reject the null (GSPC is not normally distrib
} else {
  paste("Fail to reject the null hypothesis.")
}
```

```
## [1] "9339.7935 > 5.9915 : reject the null (GSPC is not normally distributed.)"
```

Q3.

	DBV (Deutsche Bank)	GSPC (S&P 500)
Skewness	-0.8464924	-0.323834
t-test	-16.6811274	-6.381530
Excess Kurtosis	13.5025619	9.786951
t-test	133.0419239	96.431683
JB-test	17978.4135231	9339.793486
p-value JB	0.0000000	0.000000

From the result, it is hard to say that log returns of both the DBV and S&P500 indices are following the normal distribution. For the DBV index, its negative skewness and high kurtosis shows that the index has been exposed to more extreme events compared to the S&P500.

Q4.

```r
# 1. scale the data to hold annual return = 0.2, annual sd = 0.4 (sharpe ratio = E(r)/sd = 0.5)
daily_target_r <- 0.2/252
daily_target_sd <- 0.4/sqrt(252)
```

```r
# DBV
mean_dbv_daily <- mean(dbv$close)
sd_dbv_daily <- sd(dbv$close)
dbv$ret <- ((dbv$close - mean_dbv_daily)/sd_dbv_daily)* daily_target_sd + daily_target_r

## Annual returns
dbv_annual <- apply.yearly(dbv[, .(Date, close)], function(x) prod(1 + x, na.rm = T) - 1)
#mean(dbv_annual) # -0.003420513
dbv_lever <- apply.yearly(dbv[, .(Date, ret)], function(x) prod(1 + x, na.rm = T) - 1)
#mean(dbv_lever) # 0.2006492

# GSPC
mean_gspc_daily <- mean(gspc$close)
sd_gspc_daily <- sd(gspc$close)
gspc$ret <- ((gspc$close - mean_gspc_daily)/sd_gspc_daily)* daily_target_sd + daily_target_r
## Annual returns
gspc_annual <- apply.yearly(gspc[, .(Date, close)], function(x) prod(1 + x, na.rm = T) - 1)
#mean(gspc_annual) # 0.04344666
gspc_lever <- apply.yearly(gspc[, .(Date, ret)], function(x) prod(1 + x, na.rm = T) - 1)
#mean(gspc_lever) # 0.1985398
```

	Original DBV	Leveraged DBV	Original GSPC	Leveraged GSPC
Daily Return	-0.0000147	0.0007937	0.0001882	0.0007937
Annual Return	-0.0034205	0.2006492	0.0434467	0.1985398
Standard Deviation	0.0087652	0.0251976	0.0134675	0.0251976
SD (annual)	0.1331622	0.4111677	0.1927480	0.3753094
Skewness	-0.8464924	-0.8464924	-0.3238340	-0.3238340
Kurtosis	16.5025619	16.5025619	12.7869509	12.7869509

While we can manipulate the return and standard deviation, it does not affect the skewness and the kurtosis of the distribution. Therefore, under the same sharpe ratios for equity and currency market, investing in equity market has less risk.

Q5.

```r
# Classic OLS (under normality)
out <- lm(dbv$close ~ gspc$close)
ols <- summary(out)$coef[1:2, 1:2] # coefficient and standard errors

# Regression allowing non-normality and HAC
library(sandwich)
hac <- sqrt(diag(vcovHC(out, type = "HC")))
```

	Standard OLS	Non-normality and HAC
Slope	0.0101102	0.0181383
Intercept	0.0001361	0.0001367

While we are assuming constant volatility for all residuals under homogeneity, heteroskedastic standard errors does not hold this assumption and accepts wider/narrower intervals of the residuals.

In general, residuals increases as the value of independent variable increases. Thus, the robust standard errors are larger than non-robust (homogeneity) standard errors, evaluating coefficients conservatively (as it decreases the t-stat of the coefficients).

In the result above, the robust standard error of the slope and the intercept holds larger values.