HW 1
MFE 407: Empirical Methods in Finance
Professor Lochstoer
Group 6
Students: Xiahao Wang, Juan Manuel Ferreyra Maspero, Xinyue Zhu,
Yichu Li, Mu Lin

Problem 1: Modeling heavy-tails with jumps

1. Derive the mean, variance, skewness and excess kurtosis of this distribution.

$E(r_t) = E(\mu + \sigma * \epsilon + J_t)$

$= \mu + \sigma * E(\epsilon) + E(B_t * (\mu_j + \sigma_j * \delta_t))$

$= \mu + P * E(\mu_j + \sigma_j * \delta_t))$

$= \mu + P * \mu_j$, which is the mean

And

$Var(r_t) = Var(\mu + \sigma * \epsilon + J_t)$

$= Var(\sigma * \epsilon) + Var(J_t) + 2Cov(\sigma * \epsilon, J_t)$

$= \sigma^2 + Var(B_t * \mu_j) + Var(B_t * \sigma_j * \delta_t) + 2 * 0$

$= \sigma^2 + P * (1 - P) * \mu_j^2 + P * \sigma_j^2$, which is the variance

2. Explain carefully why this Bernoulli-normal mixture is potentially a better model of returns than the log-normal model.

Answer: Stock returns have fat tails that cannot be explained with a lognormal model. Unlike the lognormal model which assumes returns are normally distributed, this mixture model allows certain jumps that are normally distributed and happen with probability of p. This characteristics can be seen in the historical data where, in certain period, the volatility of return is more dramatic than others.

3. Suppose that log returns rt are simply i.i.d. distributed N (0.008, 0.0632). Plot a simulated series of 600 observations (50 years). Does it look like the data? What is missing?

```
library(xts)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(moments)
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:lubridate':
##
##      hour, isoweek, mday, minute, month, quarter, second, wday,
##      week, yday, year
```

```
## The following objects are masked from 'package:xts':
##
##      first, last
```

```r
library(sandwich)

# Problem 1
n <- 600
mu <- 0.00
sd <- 0.063
rt <- rnorm(600, mu, sd)

#Market log normal return
mkt_monthly_raw <- read.csv("CRSP_market_monthly.csv",header = TRUE, sep=",", skip=4)[-3,]
colnames(mkt_monthly_raw) <- c("caldt","a","b","sprtrn","indx")
mkt_monthly <- xts(x=mkt_monthly_raw$sprtrn, ymd(mkt_monthly_raw$caldt))
mkt_logret <- log(mkt_monthly + 1)
# Missing: No Clustering and Jump y_t and y_t+1 no related
plot(density(mkt_logret[200:800]), ylab="Log returns",
     xlab="Month",type="l", main = "Market Return VS Log normal distribution",col="blue")
lines(density(rt),col="green")
legend("topleft", c("Market Return", "Log normal distribution"),
       col=c("blue", "green"),cex=0.75, lwd=10)
```

Market Return VS Log normal distribution
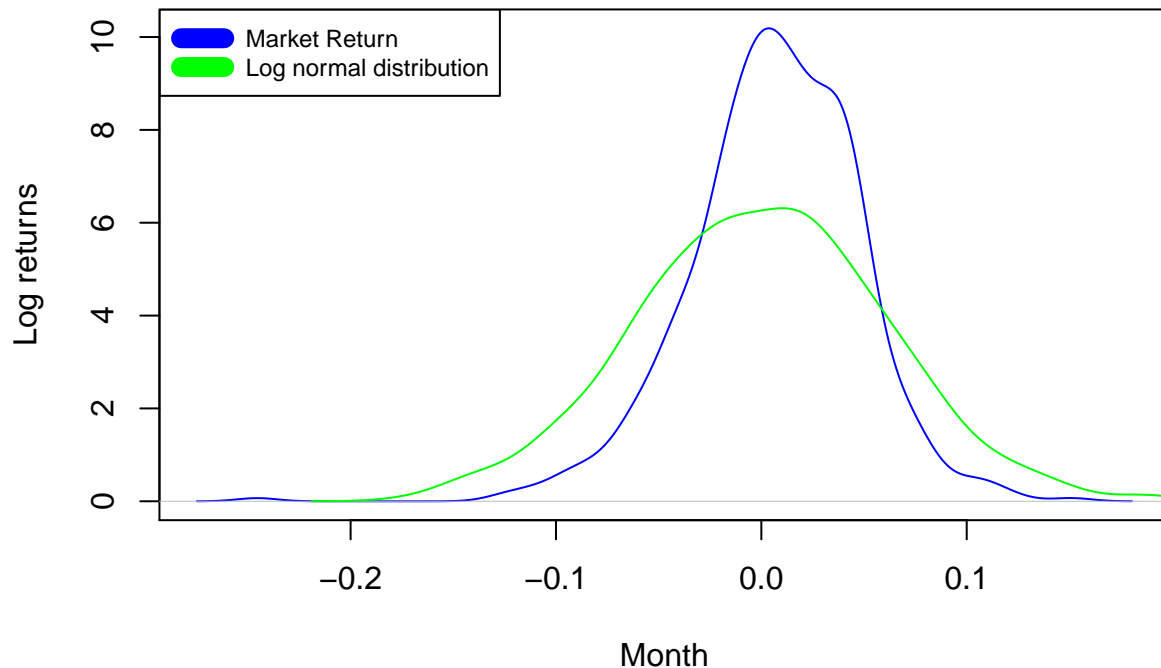
Answer: The simulation does not resemble the actual data. In fact, the actual distribution is 'taller' and 'fatter', which means there is excess kurtosis. This could be due to the fact that the log normal model doesnt consider jumps which happen in the real data.

Next, suppose that log returns rt are given by the above jump model with the following 5 parameters: (0.012, 0.05, 0.15, 0.03, 0.1) Using these parameter estimates, what are the unconditional mean, standard deviation, skewness and kurtosis of log stock returns? Again, plot a simulated series of 600 observations (50 years). Does it look like the data? What is missing

```
mean <- 0.012
sigma <- 0.05
prob <- 0.15
mu_j <- -0.03
sigma_j <- 0.1

set.seed(1234)
#simulate bernoulli
theta_t <- rnorm(n,0,1)
epsilon_t <- rnorm(n,0,1)
B_t <- rbinom(n, 1, prob)
J_t <- B_t * (mu_j + sigma_j * theta_t)
r_t <- mean + sigma * epsilon_t + J_t

simmean <- mean(r_t)
simsd <- sqrt(var(r_t))
simskw <- mean((r_t-simmean)^3/simsd^3)
simkurt <- mean((r_t-simmean)^4/simsd^4)

cat("The unconditional mean of the simulation is", simmean)
```

```
## The unconditional mean of the simulation is 0.005851911
```

```r
cat("The standard deviation of the simulation is", simsd)
```
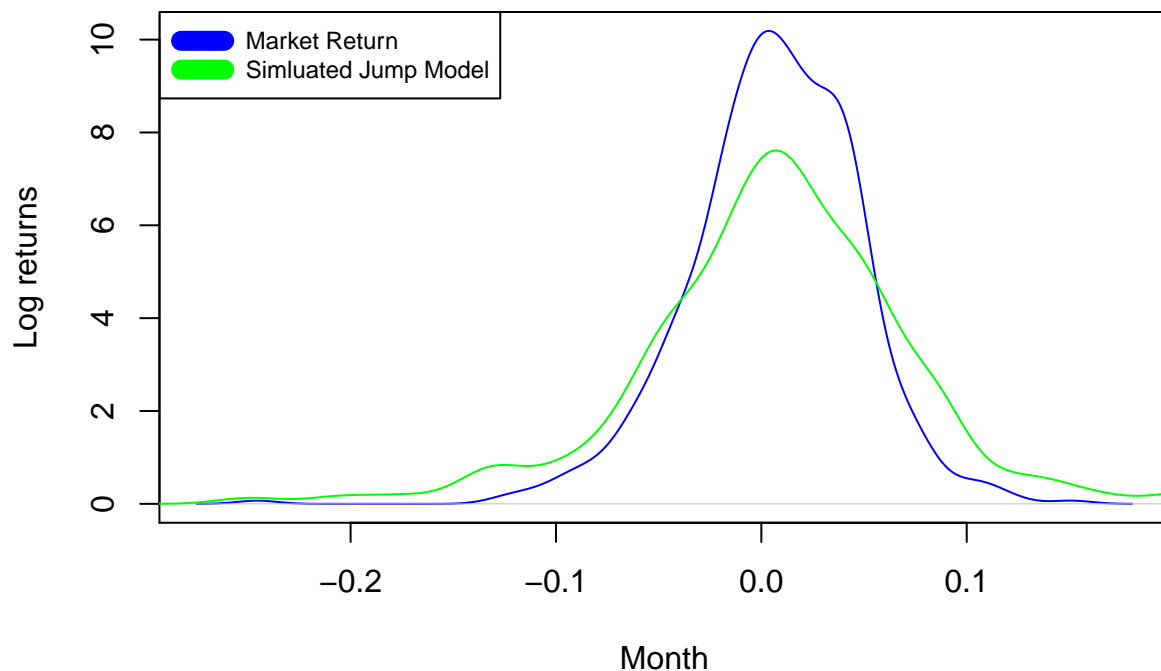
The standard deviation of the simulation is 0.07134357

```r
cat("The skewness of the simulation is", simskw)
```

The skewness of the simulation is -0.5235854

```r
cat("The kurtosis of the simulation is", simkurt)
```

The kurtosis of the simulation is 7.056755

```r
plot(density(mkt_logret[200:800]), ylab="Log returns",
     xlab="Month",type="l", main = "Market Return VS Simluated Jump Model",col="blue")
lines(density(r_t),col="green")
legend("topleft", c("Market Return", "Simluated Jump Model"),
       col=c("blue", "green"),cex=0.75, lwd=10)
```

Market Return VS Simluated Jump Model

Answer: The modified simulation with 'jumps' fits the real data better especially in the tails. The center part, however, we are still somehow undershooting. The standard deviation of the jump model is higher that that in the actual market data. One way to show this is to set the sigma to a lower value, and the jump model will fit the market return graph better

```r
sigma <- 0.034
r_t_new <- mean + sigma * epsilon_t + J_t
plot(density(r_t_new),main="Market Return VS Simluated Jump Model with Variance",
     xlab="Return",ylab="Density",col="blue",ylim=c(0,10))
lines(density(mkt_logret[200:800]),col="green")
legend("topleft", c("Market Return", "Simluated Jump Model with Low Variance"),
       col=c("blue", "green"), lwd=10,cex = 0.6)
```

Market Return VS Simluated Jump Model with Variance
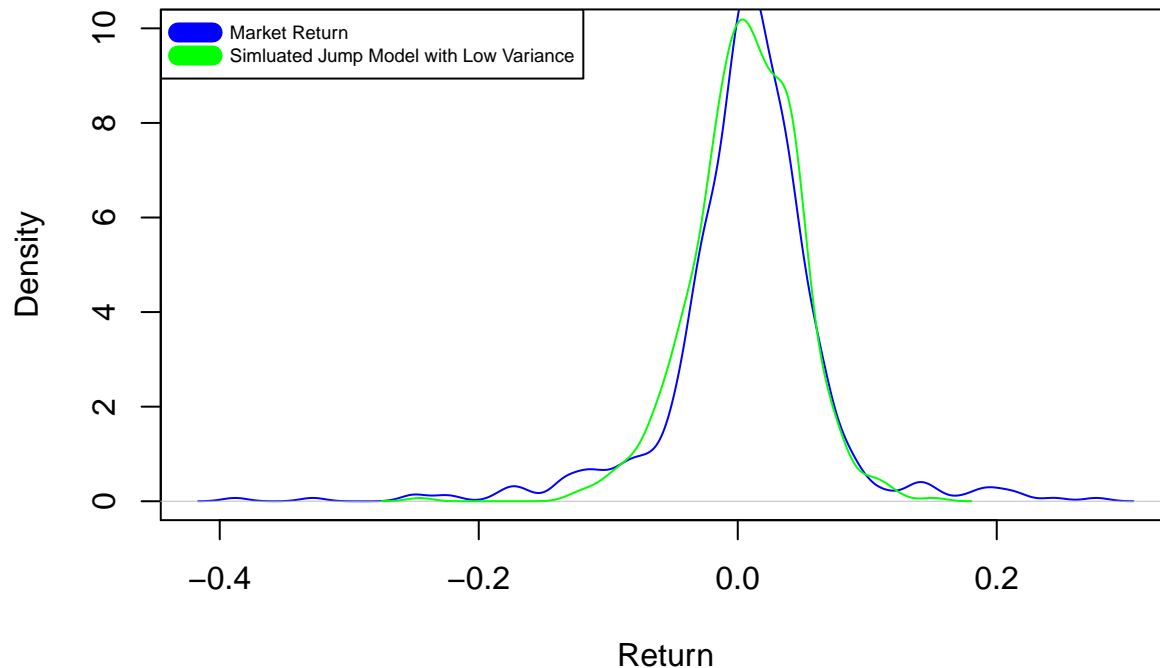
Problem 2: The Currency Carry Trade
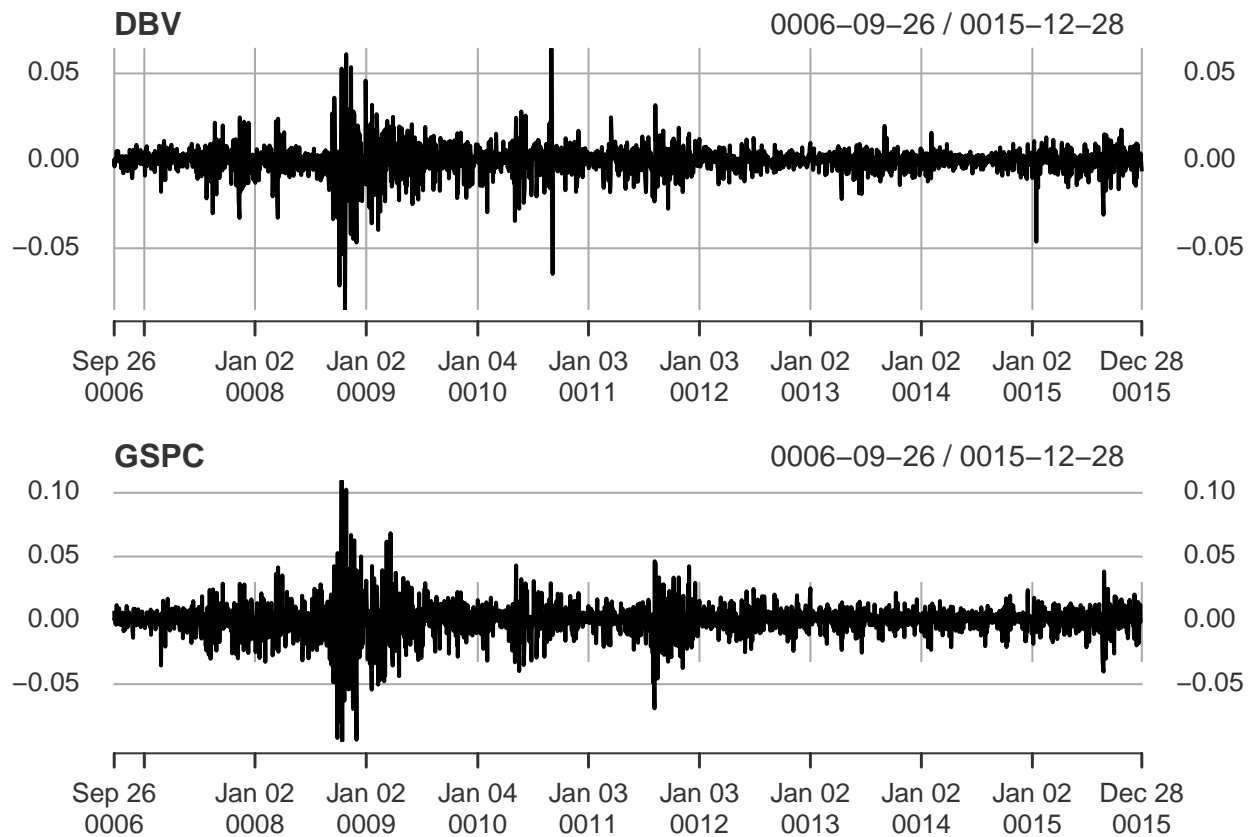
1. Visualizing the data.

(a) Create time series plots of the daily log-returns for DBV and GSPC.

```r
dbv_raw <- read.csv("DBV.csv", header=TRUE, sep=",")
gspc_raw <- read.csv("GSPC.csv", header = TRUE, sep=",")

dbv_ts <- xts(x =dbv_raw$Adj.Close, as.POSIXct(dbv_raw$Date, format="%d/%m/%Y"))
gspc_ts <- xts(x=gspc_raw$Adj.Close, as.POSIXct(gspc_raw$Date, format="%d/%m/%Y"))

dbv_logret <- diff(log(dbv_ts), lag =1)[-1,]
gspc_logret <- diff(log(gspc_ts), lag=1)[-1,]

par(mfrow=c(2,1))
plot(dbv_logret, ylab = "Daily Log Return", xlab="Time",main ="DBV")
plot(gspc_logret, ylab = "Daily Log Return", xlab="Time",main ="GSPC")
```

DBV   0006–09–26 / 0015–12–28

0.05   0.05

0.00   0.00

−0.05   −0.05

Sep 26   Jan 02   Jan 02   Jan 04   Jan 03   Jan 03   Jan 02   Jan 02   Jan 02   Dec 28
0006   0008   0009   0010   0011   0012   0013   0014   0015   0015

GSPC   0006–09–26 / 0015–12–28

0.10   0.10

0.05   0.05

0.00   0.00

−0.05   −0.05

Sep 26   Jan 02   Jan 02   Jan 04   Jan 03   Jan 03   Jan 02   Jan 02   Jan 02   Dec 28
0006   0008   0009   0010   0011   0012   0013   0014   0015   0015

(b) Create histograms of the daily log-returns for DBV and GSPC.

```
colnames(dbv_logret) <- paste0("Log_Return")
colnames(gspc_logret) <- paste0("Log_Return")
ggplot(dbv_logret, aes(x=Log_Return)) + geom_histogram(colour="black",
fill="white",bins=100) + ggtitle("DBV Log Return")
```
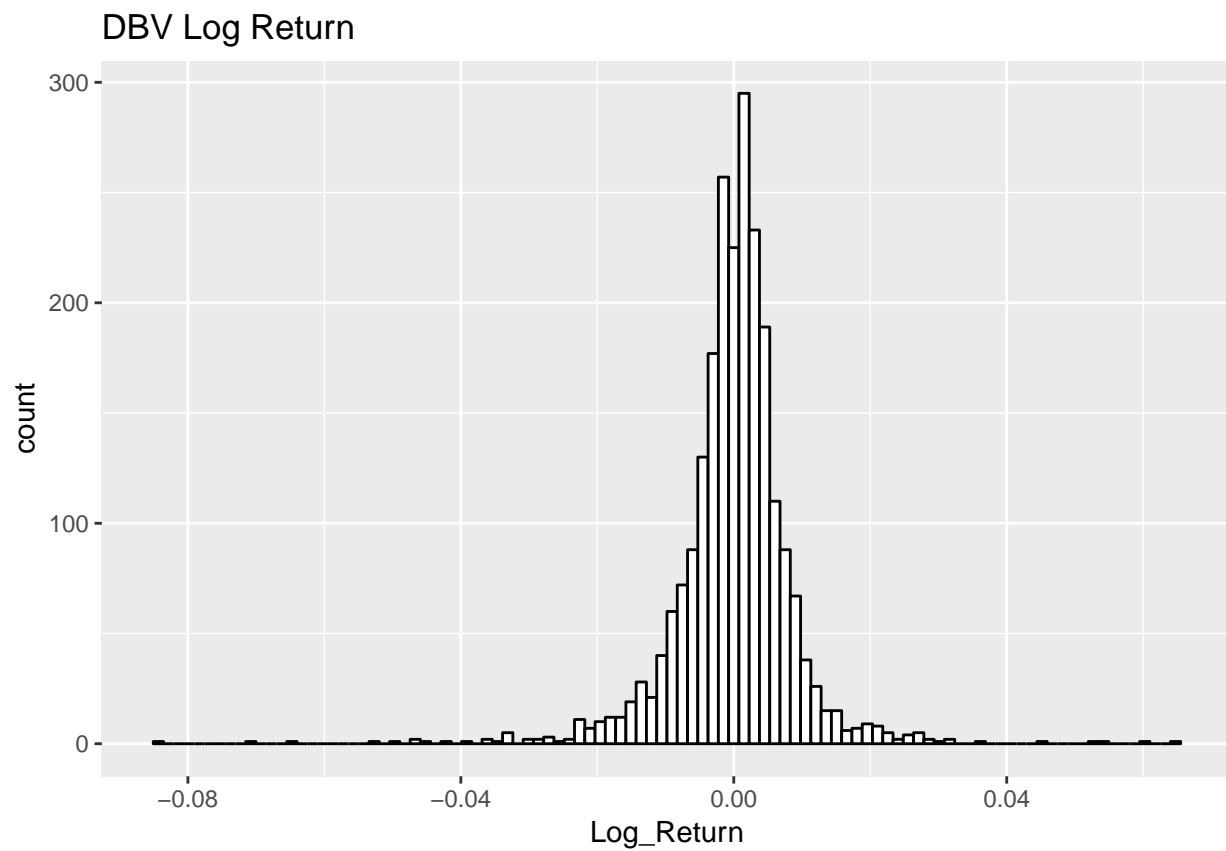
DBV Log Return


```r
ggplot(gspc_logret, aes(x=Log_Return)) + geom_histogram(colour= "black",
fill="white",bins=100)+ ggtitle("GSPC Log Return")
```
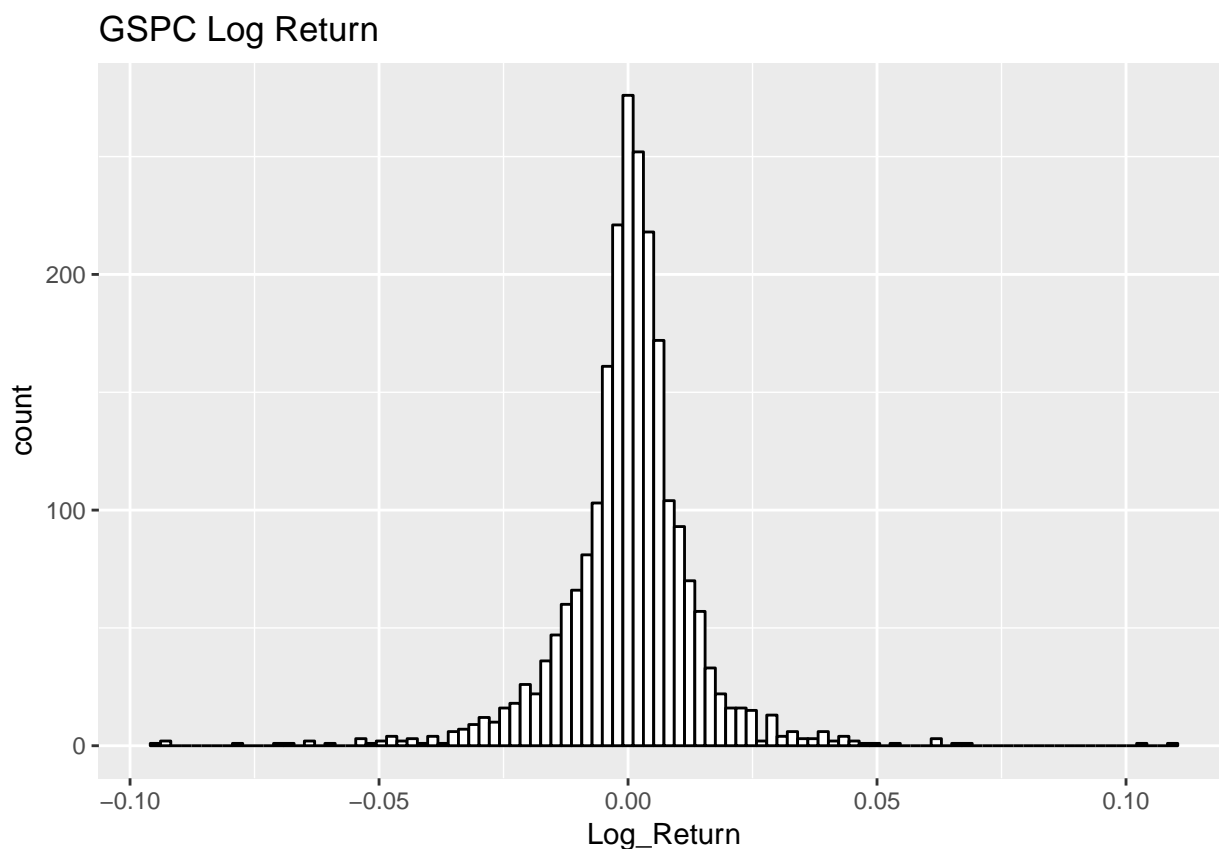
GSPC Log Return

2. Shape of Return Distribution:

(a) Test the null that the skewness of daily log returns is zero at the 5% significance level.

```r
testSkewness <- function(skew_val,size){
  return(skew_val/sqrt(6/size))
}


testKurtosis <- function(kurt_val, size){
  return((kurt_val-3)/sqrt(24/size))
}


Jarque_Bera_test <- function(skew_val, kurt_val, size){

  return((skew_val^2)/(6/size) + (kurt_val -3)^2/(24/size))
}



tval_dbv_skew <- testSkewness(skewness(dbv_logret), length(dbv_logret))
tval_gspc_skew <- testSkewness(skewness(gspc_logret), length(gspc_logret))

# apply t-test on knewness, two_tail 5% sig level
if(abs(tval_dbv_skew)>1.96){
  cat("We reject the null hypothesis that the skewness of DBV is 0. ")
} else{
  cat("We fail to reject the nul hypothesis that the skewness of DBV is 0. ")
}
```

```
## We reject the null hypothesis that the skewness of DBV is 0.

if(abs(tval_gspc_skew)>1.96){
  cat("We reject the null hypothesis that the skewness of GSPC is 0. ")
} else{
  cat("We fail to reject the nul hypothesis that the skewness of GSPC is 0. ")
}

## We reject the null hypothesis that the skewness of GSPC is 0.
```

(b) Test the null that the excess kurtosis of daily log returns is zero at the 5% significance level.

```
tval_dbv_kurt <- testKurtosis(kurtosis(dbv_logret), length(dbv_logret))
tval_gspc_kurt <- testKurtosis(kurtosis(gspc_logret), length(gspc_logret))

if(abs(tval_dbv_kurt)>1.96){
  cat("We reject the null hypothesis that the excess kurtosis of DBV is 0. ")
} else{
  cat("We fail to reject the nul hypothesis that the excess kurtosis of DBV is 0. ")
}

## We reject the null hypothesis that the excess kurtosis of DBV is 0.

if(abs(tval_gspc_kurt)>1.96){
  cat("We reject the null hypothesis that the excess kurtosis of GSPC is 0. ")
} else{
  cat("We fail to reject the nul hypothesis that the excess kurtosis of GSPC is 0. ")
}

## We reject the null hypothesis that the excess kurtosis of GSPC is 0.
```

(c) Test the null that the daily log returns are normally distributed at the 5% significance level using the Jarque-Bera test.

```
dbv_jb_result <- Jarque_Bera_test(skew_val = skewness(dbv_logret),
kurt_val = kurtosis(dbv_logret), size = length(dbv_logret))

gspc_jb_result <- Jarque_Bera_test(skew_val = skewness(gspc_logret),
kurt_val = kurtosis(gspc_logret), size = length(gspc_logret))


if(dbv_jb_result>qchisq(0.95,2)){
  cat("We reject the null hypothesis that the log-return of DBV is normal. ")
}else{
  cat("We fail to reject the null hypothesis that the log-return of DBV is normal. ")
}

## We reject the null hypothesis that the log-return of DBV is normal.

if(gspc_jb_result>qchisq(0.95,2)){
  cat("We reject the null hypothesis that the log-return of GSPC is normal. ")
}else{
  cat("We fail to reject the null hypothesis that the log-return of GSPC is normal. ")
}

## We reject the null hypothesis that the log-return of GSPC is normal.
```

3. Compare all of these numbers in (a) and (b) to the same numbers for daily log returns on the S&P 500 measured over the same sample in one single table.

```r
DT <- data.table(
  Portfolio = c("Mean", "Standard Deviation", "Skewness", "t-test","Kurtosis", "t-test", "JB-test"),
  DBV = c(mean(dbv_logret), sd(dbv_logret), skewness(dbv_logret), tval_dbv_skew,
          kurtosis(dbv_logret), tval_dbv_kurt, dbv_jb_result),
  GSPC = c(mean(gspc_logret), sd(gspc_logret), skewness(gspc_logret), tval_gspc_skew,
           kurtosis(gspc_logret), tval_gspc_kurt, gspc_jb_result)
)
DT
```

```
##              Portfolio           DBV          GSPC
## 1:                Mean -1.469803e-05  1.882230e-04
## 2: Standard Deviation  8.765226e-03  1.346752e-02
## 3:            Skewness -8.470376e-01 -3.240426e-01
## 4:              t-test -1.669187e+01 -6.385641e+00
## 5:            Kurtosis  1.651674e+01  1.279793e+01
## 6:              t-test  1.331816e+02  9.653990e+01
## 7:             JB-test  1.801595e+04  9.360729e+03
```

4. Suppose you are a fund manager with a target return in mind (say 20 % per annum), and suppose that the ratio of expected returns to standard deviation is 0.50 for both investments. Using these numbers, discuss the di?erent nature of the risks you would face if you invested in equity or currency markets to achieve that target return (with the appropriate leverage).

Answer: Both DBV and GSPC are not normally distributed, therefore we need to consider higher order moments. Sharp ratio only uses mean and variance, the first two moments. - Higher skewness suggest higher downside risk. - Higher kurtosis suggest higher tail risk

DBV has higher skewness and kurtosis, which leads to higher downside risk. Because of the short position in the currency trading, fund manager is exposed to even higher downside risk.

5. Regress the log DBV returns on the log GSPC returns. Report the standard errors of the slope and intercept coe? cients using both standard OLS assumptions and allowing for non-normalities and heteroskedastic errors (White standard errors). Explain why the two are di?erent and give the intuition for why White standard errors are smaller/larger than the classic OLS standard errors in this case.

```r
reg <- lm(as.numeric(dbv_logret) ~ as.numeric(gspc_logret))
summary(reg)
```

```
##
## Call:
## lm(formula = as.numeric(dbv_logret) ~ as.numeric(gspc_logret))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.069616 -0.003081  0.000239  0.003296  0.058668
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -9.579e-05  1.361e-04  -0.704    0.482
## as.numeric(gspc_logret)  4.309e-01  1.011e-02  42.616   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006571 on 2328 degrees of freedom
## Multiple R-squared:  0.4382, Adjusted R-squared:  0.438
## F-statistic:  1816 on 1 and 2328 DF,  p-value: < 2.2e-16
```

```
white_stderr <- sqrt(diag(vcovHC(x=reg, type = "HC")))
intercept_se <- as.numeric(white_stderr[1])
beta_se <- as.numeric(white_stderr[2])
cat("Intercept standard error is",paste0(intercept_se))
```

Intercept standard error is 0.000136734249580345

```
cat("Beta (GSPC Log return) standard error is",paste0(beta_se))
```

Beta (GSPC Log return) standard error is 0.0181383312209999

Answer: White Standard Error are larger than the classic OLS standard in this case. OLS regression gives equal weight to all observations, but when heteroscedasticity is present, the cases with larger error have more "pull" than other observations. In this case the variance of the error terms increase with the disstance between gspc log return (independent variable) and its mean