STAT0017 ICA 2 2018-19

Student number: xxx

2019-04-05

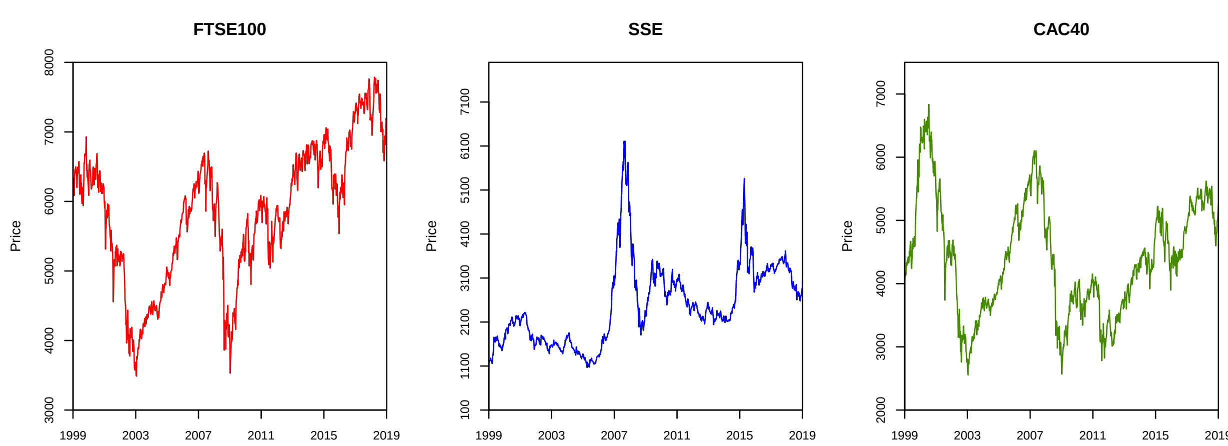
Contents

The first indices chosen are FTSE100, SSE and CAC40 and they are then plotted to observe the distribution of the data.

```
par(mfrow=c(1,3))
#FTSE100
plot(data$ftse100 ~ as.Date(data$date, "%d/%m/%y"), type = "l", xaxt = 'n', yaxt = 'n',
     xlab = "", ylab = "Price", col = "red", main = "FTSE100", xaxs = "i", yaxs = "i",
     ylim = c(3000,8000))
axis(side = 2, at = seq(3000,8000,1000), tick = T, cex.axis = 0.9)
axis.Date(side = 1, cex.axis = 0.9, at = seq(as.Date("1999/02/25"),
                                             as.Date("2019/02/28"), "4 years"))

#SSE
plot(data$sse ~ as.Date(data$date, "%d/%m/%y"), type = "l", xaxt = 'n', yaxt = 'n',
     xlab = "", ylab = "Price", col = "blue", main = "SSE", xaxs = "i",
     yaxs = "i", ylim = c(100,8000))
axis(side = 2, at = seq(100,8000,1000), tick = T, cex.axis = 0.9)
axis.Date(side = 1, cex.axis = 0.9,at = seq(as.Date("1999/02/25"),
                                            as.Date("2019/02/28"), "4 years"))

#CAC40
plot(data$cac40 ~ as.Date(data$date, "%d/%m/%y"), type = "l", xaxt = 'n', yaxt = 'n',
     xlab = "", ylab = "Price", col = "chartreuse4", main = "CAC40", xaxs = "i",
     yaxs = "i", ylim = c(2000,7500))
axis(2, at = seq(2000,7500,1000), tick = T, cex.axis = 0.9)
axis.Date(1, cex.axis = 0.9, at = seq(as.Date("1999/02/25"),
                                      as.Date("2019/02/28"), "4 years"))
```



From the plots, there seem to be linear trends in the data which may cause the data to be non-stationary.

Non-stationarity will give rise to bias while forecasting and the model that will be considered assumes stationarity. Therefore, unit root tests are done to test if the time series data of the financial stock indices is stationary.

```
#FTSE100
c(paste("p-value for price:",round(unitrootTest(data$ftse100)@test$p.value[1],3)),
paste("p-value for log of price:",round(unitrootTest(log(data$ftse100))@test$p.value[1],3)))
```

```
## [1] "p-value for price: 0.671"        "p-value for log of price: 0.736"
```

```
#SSE
c(paste("p-value for price", round(unitrootTest(data$sse)@test$p.value[1],3)),
paste("p-value for log of price", round(unitrootTest(data$sse)@test$p.value[1],3)))
```

```
## [1] "p-value for price 0.605"        "p-value for log of price 0.605"
```

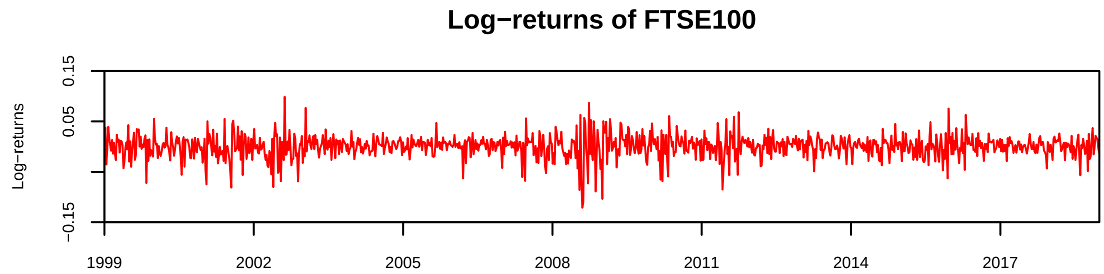
```
#CAC40
c(paste("p-value for price", round(unitrootTest(data$cac40)@test$p.value[1],3)),
paste("p-value for log of price", round(unitrootTest(log(data$cac40))@test$p.value[1],3)))
```

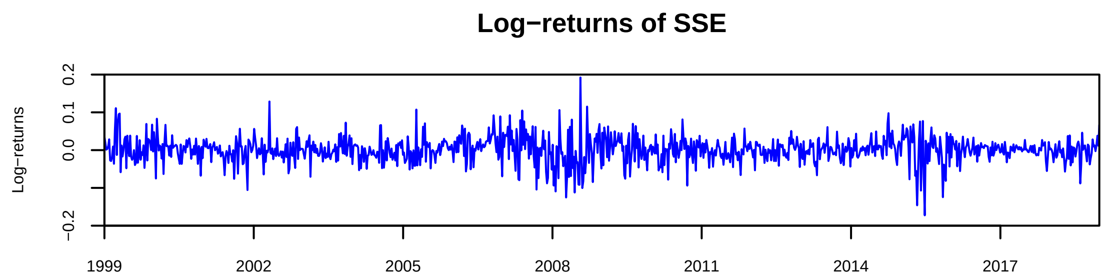
```
## [1] "p-value for price 0.666"        "p-value for log of price 0.758"
```

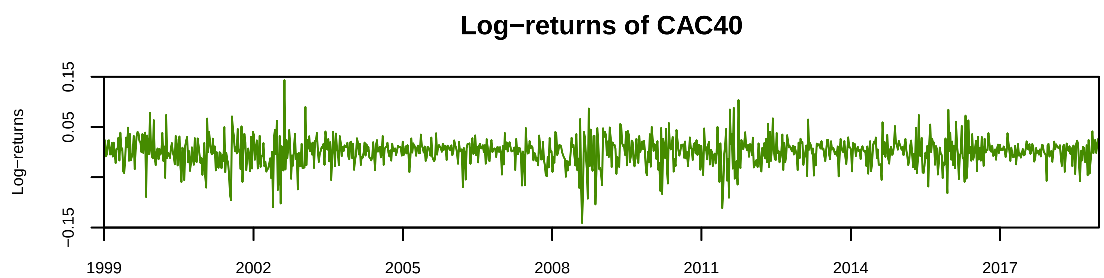
Hence, from all the p-values, it can be inferred that there are no sufficient evidence to suggest that the data are stationary. Taking the log is not enough to make the data stationary, so other transformations are needed to ensure stationarity of the data. The first order difference of the log of the prices, which is the log returns, is taken in an attempt to make the data stationary.

Plotting the log returns, the variance seems to be much smaller than before the transformation and the trend in the data seems to have been removed as well.

```
par(mfrow=c(3,1))
#FTSE100
log_returns_ftse100 <- diff(log(data$ftse100), lag = 1, na = remove)
plot(log_returns_ftse100~as.Date(data$date[2:length(data$date)], "%d/%m/%y"), type = "l",
     yaxt = 'n', xaxt = 'n', xlab = "", ylab = "Log-returns", ylim = c(-0.15,0.15),
     xaxs = "i", yaxs = "i", col = "red", cex.lab = 0.8, main = "Log-returns of FTSE100")
axis(2, at = seq(-0.15,0.15,0.1), tick = T, cex.axis = 0.7)
axis.Date(1, cex.axis=0.7, at=seq(as.Date("1999/03/04"), as.Date("2019/02/28"), "3 years"))
#SSE
log_returns_sse <- diff(log(data$sse), lag = 1, na = remove)
plot(log_returns_sse~as.Date(data$date[2:length(data$date)], "%d/%m/%y"), type = "l",
     yaxt = 'n', xaxt = 'n', xlab = "", ylab = "Log-returns", ylim = c(-0.2,0.2),
     xaxs = "i", yaxs = "i", col = "blue", cex.lab = 0.8, main = "Log-returns of SSE")
axis(2, at = seq(-0.2,0.2,0.1), tick = T, cex.axis = 0.7)
axis.Date(1, cex.axis=0.7, at=seq(as.Date("1999/03/04"), as.Date("2019/02/28"), "3 years"))
#CAC40
log_returns_cac40 <- diff(log(data$cac40), lag = 1)
plot(log_returns_cac40~as.Date(data$date[2:length(data$date)], "%d/%m/%y"), type = "l",
     yaxt = 'n', xaxt = 'n', xlab = "", ylab = "Log-returns", ylim = c(-0.15,0.15),
     xaxs = "i", yaxs = "i", col="chartreuse4", cex.lab = 0.8, main="Log-returns of CAC40")
axis(2, at = seq(-0.15,0.15,0.1), tick = T, cex.axis = 0.7)
axis.Date(1, cex.axis=0.7, at=seq(as.Date("1999/03/04"), as.Date("2019/02/28"), "3 years"))
```







The means of the log-returns, which are now constant, are all approximately 0. The variance of the log-returns do seem to be relatively constant, so unit root tests are carried out again to verify them.

```r
#FTSE100
paste("p-value for the log-return:", unitrootTest( log_returns_ftse100 )@test$p.value[1])
```

```
## [1] "p-value for the log-return: 6.45559299192264e-41"
```

```r
#SSE
paste("p-value for the log-return:", unitrootTest( log_returns_sse )@test$p.value[1])
```

```
## [1] "p-value for the log-return: 1.14050630493526e-40"
```

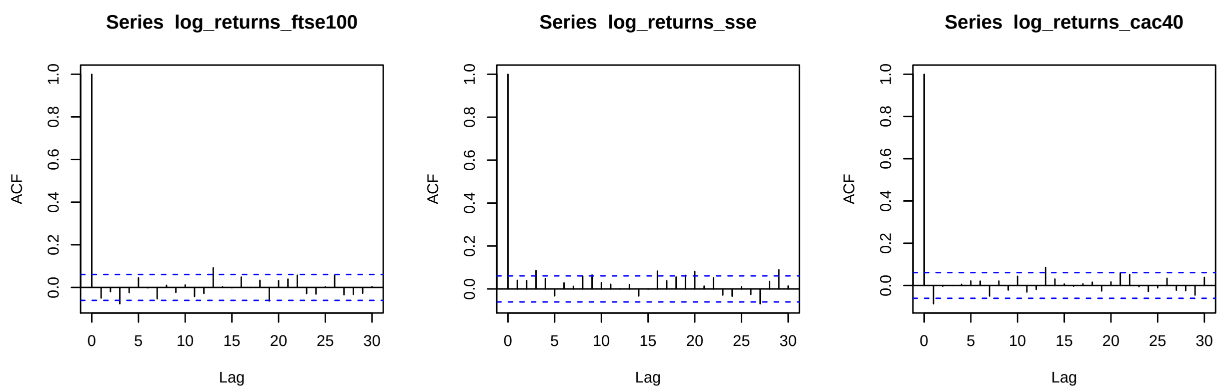
```r
#CAC40
paste("p-value for the log-return:", unitrootTest( log_returns_cac40 )@test$p.value[1])
```

```
## [1] "p-value for the log-return: 7.07499436539168e-41"
```

From the tests, there are insufficient evidence to conclude that the log-return data are non-stationary. However, there seem to be volatility clustering in the data which motivates the use of GARCH model to fit the data.

The GARCH process models the volatilty of the log returns and the ARMA process models the time series itself. The ACF plot of the data are used to determine the parameters of the AR model.

```
par(mfrow=c(1,3))
acf(log_returns_ftse100)
acf(log_returns_sse)
acf(log_returns_cac40)
```



For the log-returns of FTSE100, there is a spike at lag 3 and then no spikes after, hence an AR(3) model is chosen;

```
model_ftse100 <- garchFit(formula = ~ arma(3,0)+garch(1,1), data = log_returns_ftse100,
                          trace = F, cond.dist = "sstd")
```

For the log-returns of SSE, similarly there is a spike at lag 3 and then no spikes after, therefore an AR(3) model is chosen;

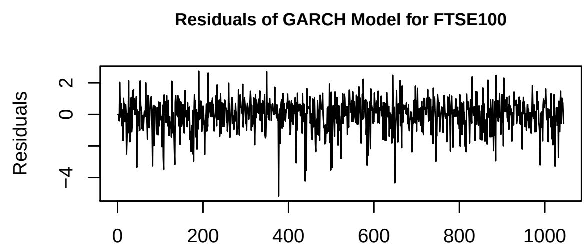
```
model_sse <- garchFit(formula = ~ arma(3,0)+garch(1,1), data = log_returns_sse,
                      trace = F, cond.dist = "sstd")
```

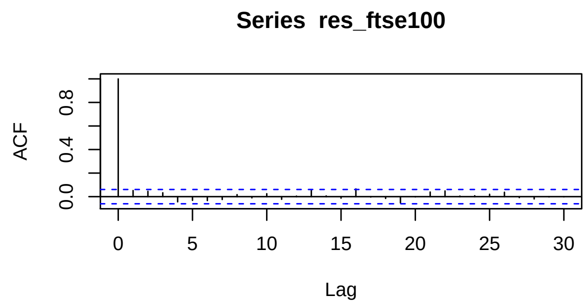
For the log-returns of CAC40, there is a spike at lag 1 and then no spikes after, so an AR(1) model is chosen.

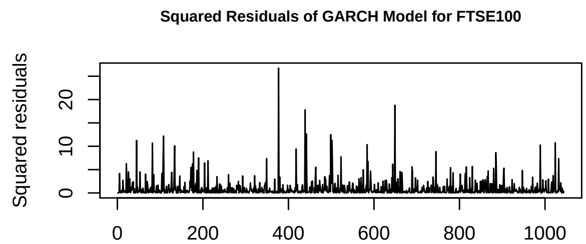
```
model_cac40 <- garchFit(formula = ~ arma(1,0)+garch(1,1), data = log_returns_cac40,
                        trace = F, cond.dist = "sstd")
```

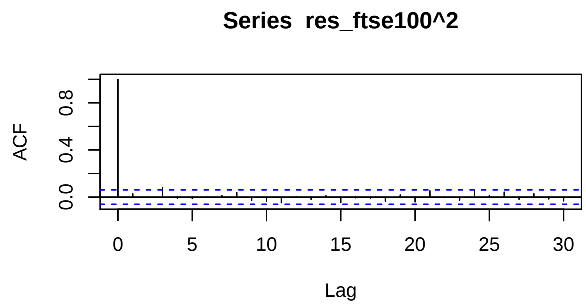
Now plotting the residuals and square of residuals as well as their respective ACF, the plots all seem to look like realisations of discrete white noise processes, which indicates good fit of the ARMA-GARCH models.

```
par(mfrow=c(2,2))
#FTSE100
res_ftse100 <- residuals(model_ftse100, standardize = T)
plot(res_ftse100,type='l',ylab="Residuals",xlab="",main="Residuals of GARCH Model for FTSE100",
     cex.main=0.9)
acf(res_ftse100)
plot(res_ftse100^2,type='l',ylab="Squared residuals",xlab="",
     main="Squared Residuals of GARCH Model for FTSE100",cex.main=0.8)
acf(res_ftse100^2)
```

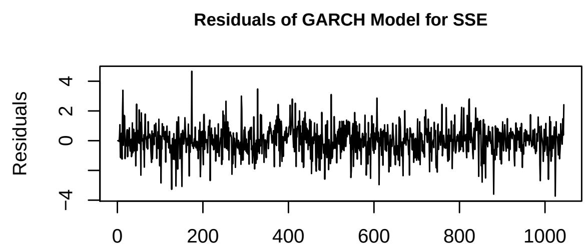


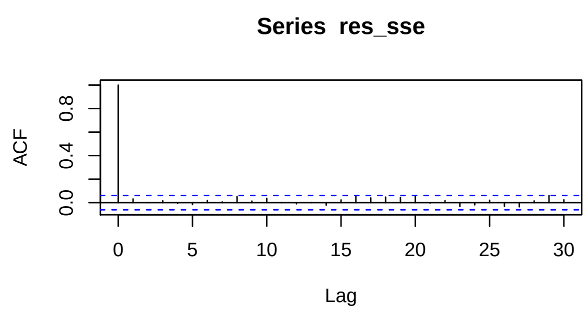


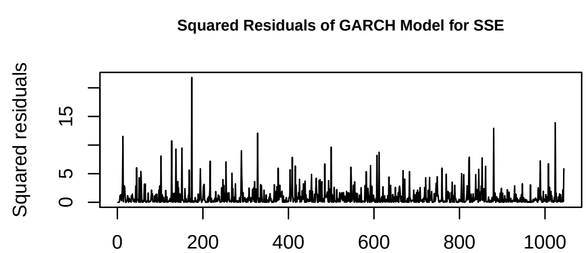


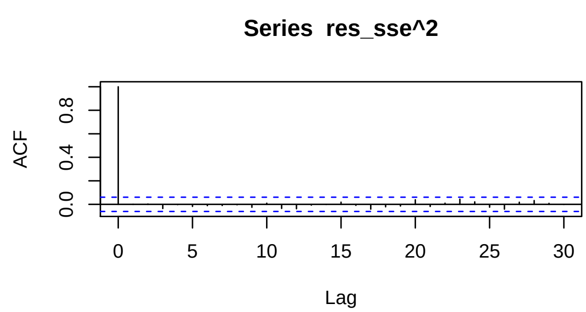



```r
#SSE
res_sse <- residuals(model_sse, standardize = T)
plot(res_sse,type="l",ylab="Residuals",xlab="",main="Residuals of GARCH Model for SSE",
     cex.main=0.9)
acf(res_sse)
plot(res_sse^2,type="l",ylab="Squared residuals",xlab="",
     main="Squared Residuals of GARCH Model for SSE",cex.main=0.8)
acf(res_sse^2)
```

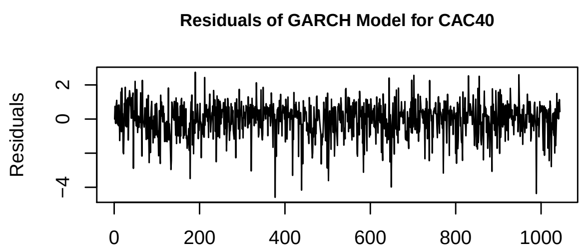


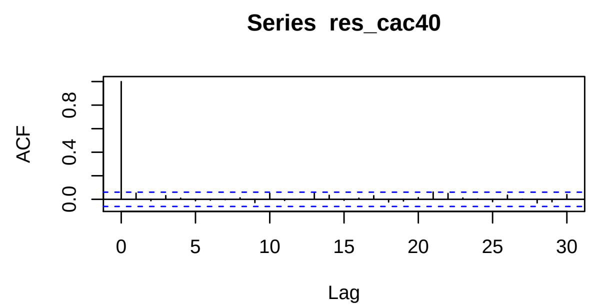


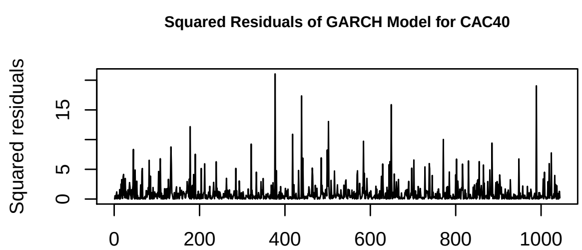


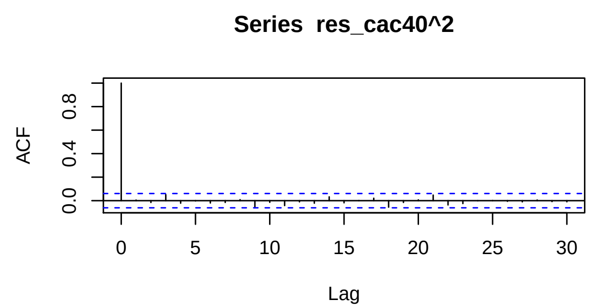


```r
#CAC40
res_cac40 <- residuals(model_cac40, standardize = T)
plot(res_cac40,type="l",ylab="Residuals",xlab="",main="Residuals of GARCH Model for CAC40",
     cex.main=0.9)
acf(res_cac40)
plot(res_cac40^2,type="l",ylab="Squared residuals",xlab="",
     main="Squared Residuals of GARCH Model for CAC40",cex.main=0.8)
acf(res_cac40^2)
```









Then to test for the GARCH effect, i.e the conditional heteroscedasticity, the Ljung-Box tests are done on all the square of the residuals. The Ljung-Box test is also carried out on the residuals to test for the autocorrelation effect.

```r
#FTSE100
c(paste("p-value for the residuals:", round(Box.test(res_ftse100, lag = 20,
                                        type = c("Ljung-Box"), fitdf = 0)$p.value,3)),
paste("p-value for the squared residuals:", round(Box.test(res_ftse100^2, lag = 20,
                                            type = c("Ljung-Box"), fitdf = 0)$p.value,3)))
```

```
## [1] "p-value for the residuals: 0.255"
## [2] "p-value for the squared residuals: 0.44"
```

```r
#SSE
c(paste("p-value for the residuals:", round(Box.test(res_sse, lag = 20,
                                        type = c("Ljung-Box"), fitdf = 0)$p.value,3)),
paste("p-value for the squared residuals:", round(Box.test(res_sse^2, lag = 20,
                                            type = c("Ljung-Box"), fitdf = 0)$p.value,3)))
```

```
## [1] "p-value for the residuals: 0.395"
## [2] "p-value for the squared residuals: 0.968"
```

```r
#CAC40
c(paste("p-value for the residuals:", round(Box.test(res_cac40, lag = 20,
                                        type = c("Ljung-Box"), fitdf = 0)$p.value,3)),
```

```r
paste("p-value for the squared residuals:", round(Box.test(res_cac40^2, lag = 20,
                                              type = c("Ljung-Box"), fitdf = 0)$p.value,3)))
```

```
## [1] "p-value for the residuals: 0.807"
## [2] "p-value for the squared residuals: 0.707"
```

For the FTSE100 model, the p-value in the first test indicates that there is insufficient evidence to conclude that the residuals up to lag 20 are dependent. With that, it is implied that the model chosen is good enough to capture the autocorrelation in the series. In the second test, the p-value indicates that there is no evidence to support that the squared residuals up to lag 20 are dependent. This means that the ARMA-GARCH model that was chosen adequately describes the time-varying volatility process.

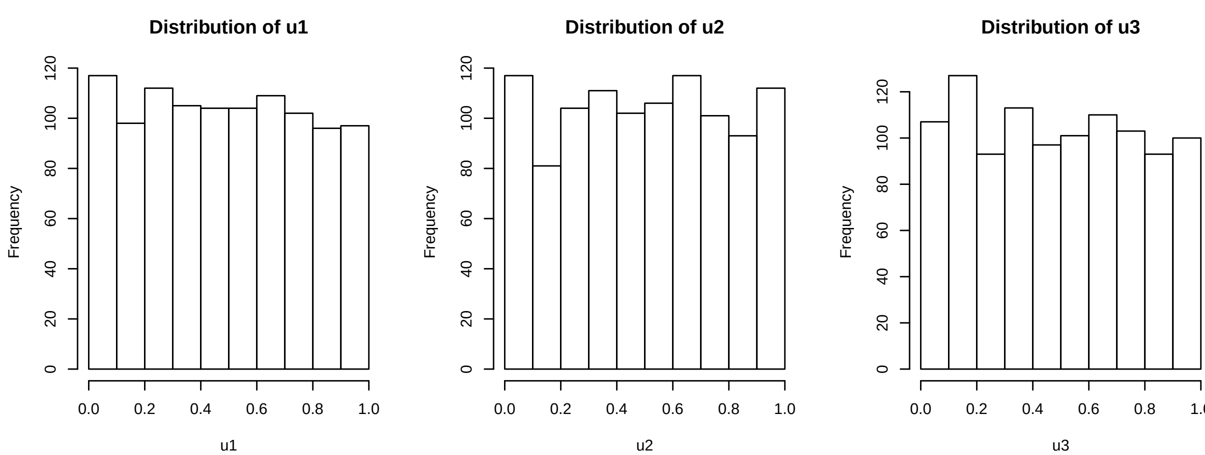
Similarly for both the SSE and CAC40 model, the p-values in the first tests indicate that there are insufficient evidence to conclude that the residuals up to lag 20 are dependent. In the second tests, the p-values indicate that there are no evidence to support that the squared residuals up to lag 20 are dependent.

To build the copula, marginal probabiltiy distributions of the log-returns have to be uniform. Due to the distributions of all three of the residuals being heavy-tailed, the skewed Student's t-distribution is used to model the marginal probability of the log-returns of all the chosen stock indices. The standardized residuals are then converted to U(0,1) samples by using the probability integral transformation method.

```r
par(mfrow = c(1, 3))
shape1 <- coef(model_ftse100)[9]
skew1 <- coef(model_ftse100)[8]
u1 <- psstd(res_ftse100, mean = 0, sd = 1, nu = shape1, xi = skew1)
hist(u1,main="Distribution of u1")

shape2 <- coef(model_sse)[9]
skew2 <- coef(model_sse)[8]
u2 <- psstd(res_sse, mean = 0, sd = 1, nu = shape2, xi = skew2)
hist(u2,main="Distribution of u2")

shape3 <- coef(model_cac40)[7]
skew3 <- coef(model_cac40)[6]
u3 <- psstd(res_cac40, mean = 0, sd = 1, nu = shape3, xi = skew3)
hist(u3,main="Distribution of u3")
```



From the histograms plotted, it seems like the distribution of u1, u2 and u3 are approximately uniform. Several tests are carried out to determine if u1, u2 and u3 are distributed uniformly. Firstly, the Lilliefors-corrected Kolmogorov-Smirnov Goodness-of-Fit test is implemented as the population parameters are unknown and the sample statistics are used to estimate them. The null hypothesis of the test is that the sample, either u1, u2 or u3, is drawn from the uniform distribution.

Observing all three of the p-values, it can be concluded that there are no significant evidence to say that u1, u2 and u3 are not uniformly distributed. Secondly, the Anderson-Darling tests of goodness-of-fit to a uniform distribution are performed. Similarly, in these tests, the null hypothesis is that the CDF of u1, u2 or u3 has a uniform distribution.

```r
ADtest1 <- ad.test(u1, null = "punif")
ADtest1$p.value
```

```
## [1] 0.413301
```

```r
ADtest2 <- ad.test(u2, null = "punif")
ADtest2$p.value
```

```
## [1] 0.8637571
```

```r
ADtest3 <- ad.test(u3, null = "punif")
ADtest3$p.value
```

```
## [1] 0.1819566
```

With all three of the p-values being significant, it cannot be concluded that u1, u2 and u3 are drawn from a distribution that is not uniform. Therefore, it is inferred that u1, u2 and u3 indeed come from distribution of U[0,1].