FA_Assignment8_Duo_Zhou

Duo Zhou

8/11/2020

Exercise 1 on page 239.

Answer questions using 5% significance level in tests and 10 lags of serial correlations for return series.

Consider daily returns of ETF SPDR S&P 500 from file d-spy-0111.txt. Transform the simple returns into log-returns

a). Is the expected log-return zero? Are there any serial correlations in the log returns? Is there ARCH effect in the log-returns?

Read the data for monthly log-returns and create returns.

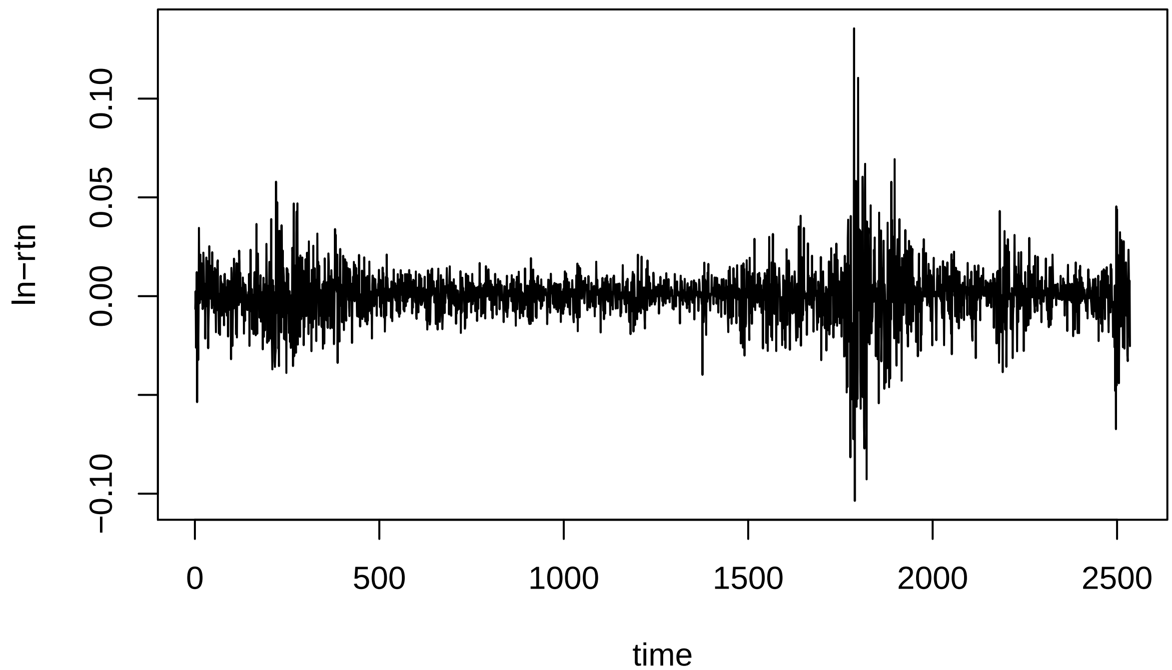
```
datapath <- "C:/Users/zd000/Desktop/MSCA/Financial Analytics/Assignments/week8/"
da=read.table(paste(datapath,"d-spy-0111.txt",sep="/"),header=T)
head(da)
```

```
##       date       rtn
## 1 20010904 -0.006395
## 2 20010905  0.002469
## 3 20010906 -0.025770
## 4 20010907 -0.018507
## 5 20010910  0.012233
## 6 20010917 -0.052249
```

```
log.rtn=log(da$rtn+1)
logrtn.ts=ts(log.rtn)
```

Plot returns

```
plot(logrtn.ts,type='l',xlab='time',ylab='ln-rtn') # time plot
```



Test H0: ="Mean Equals Zero"

```
t.test(log.rtn)  # testing the mean of returns
```

```
##
##  One Sample t-test
##
## data:  log.rtn
## t = 0.26515, df = 2534, p-value = 0.7909
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0004633792  0.0006082874
## sample estimates:
##   mean of x
## 7.24541e-05
```

The null hypothesis is rejected at 5% level, so $\mu_t = \mu + \epsilon_t$ and could be modeled by a stationary time series model.

Run Box-Ljung test with lag=10. H0: = "Serial Correlations are Zero".

```
Box.test(log.rtn,lag=10,type='Ljung')
```

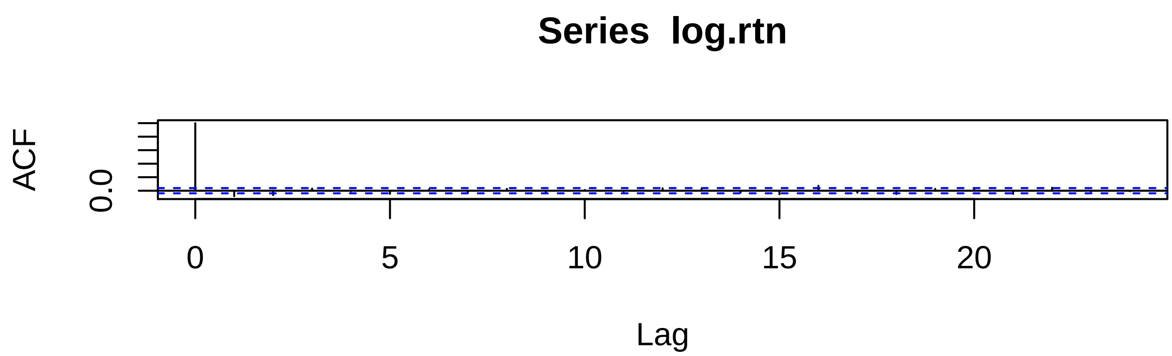
```
##
##  Box-Ljung test
```

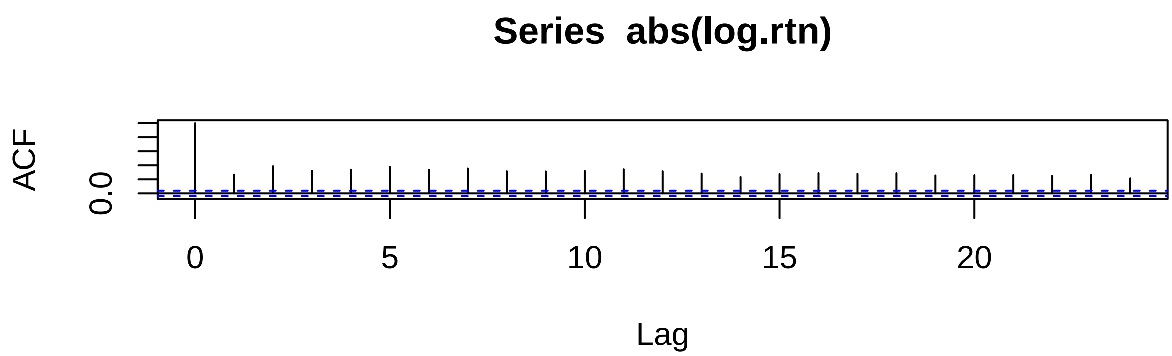
```
##
## data:  log.rtn
## X-squared = 40.38, df = 10, p-value = 1.452e-05
```

The null hypothesis can not be rejected at 5% level. There are serial correlations.

Plot ACF of returns and ACF of absolute returns.

```
par(mfcol=c(2,1))
acf(log.rtn,lag=24) # ACF plots
acf(abs(log.rtn),lag=24)
```





It is hard to see the ACF of log returns to be outside the bonds. The ACF of absolute returns is definitely outside the bounds.

```
Box.test(abs(log.rtn),lag=10,type='Ljung')
```

```
##
##  Box-Ljung test
##
## data:  abs(log.rtn)
## X-squared = 2860.5, df = 10, p-value < 2.2e-16
```

The Box-Ljung test confirms rejection of zero serial correlation null hypothesis for absolute log returns at 5% level.

Testing for ARCH effect

Calculate squared residuals and apply Box-Ljung test.

```
## ARCH test
y=log.rtn-mean(log.rtn)
Box.test(y^2,lag=10,type='Ljung')
```

```
##
##  Box-Ljung test
##
## data:  y^2
## X-squared = 1895.3, df = 10, p-value < 2.2e-16
```

According to Box-Ljung test the zero correlations hypothesis is rejected at 5% level. This is an indication of ARCH effect.

The sourced script of the text book, archTest, applies Engle's test:

```
archTest <- function(rtn,m=10){
  # Perform Lagrange Multiplier Test for ARCH effect of a time series
  # rtn: time series
  # m: selected AR order
  #
  y=(rtn-mean(rtn))^2
  T=length(rtn)
  atsq=y[(m+1):T]
  x=matrix(0,(T-m),m)
  for (i in 1:m){
    x[,i]=y[(m+1-i):(T-i)]
  }
  md=lm(atsq~x)
  summary(md)
}
archTest(log.rtn,10)
```

```
##
## Call:
## lm(formula = atsq ~ x)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0041924 -0.0000982 -0.0000494  0.0000156  0.0150562
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.773e-05  1.262e-05   2.990  0.00282 **
## x1          -2.762e-03  1.977e-02  -0.140  0.88892
## x2           3.325e-01  1.963e-02  16.940  < 2e-16 ***
## x3          -1.520e-02  2.068e-02  -0.735  0.46259
## x4          -1.425e-02  2.066e-02  -0.690  0.49047
## x5           1.533e-01  2.047e-02   7.488 9.60e-14 ***
## x6           9.943e-02  2.049e-02   4.854 1.29e-06 ***
```

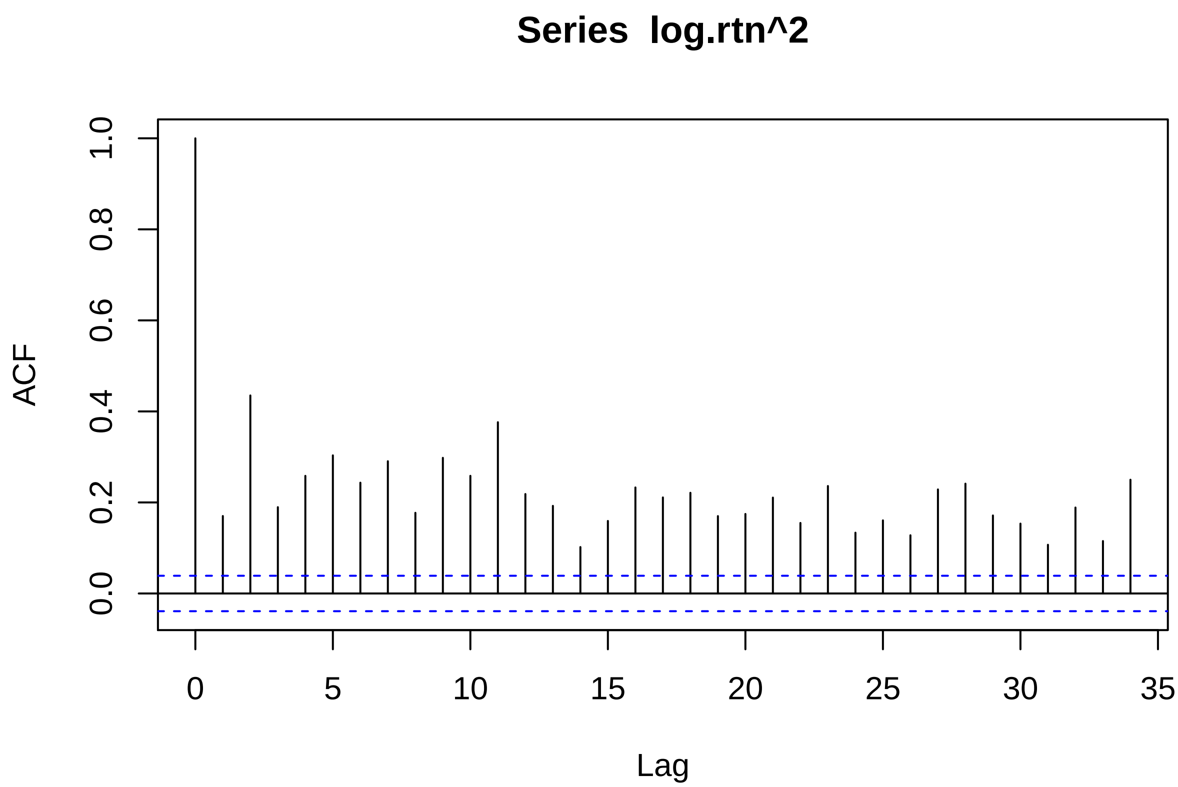
```
## x7           4.658e-02  2.058e-02   2.263  0.02370 *
## x8          -5.128e-02  2.060e-02  -2.489  0.01287 *
## x9           1.180e-01  1.956e-02   6.033 1.85e-09 ***
## x10          1.300e-01  1.969e-02   6.604 4.88e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0005718 on 2514 degrees of freedom
## Multiple R-squared:  0.2837, Adjusted R-squared:  0.2808
## F-statistic: 99.55 on 10 and 2514 DF,  p-value: < 2.2e-16
```

The output confirms ARCH effect.

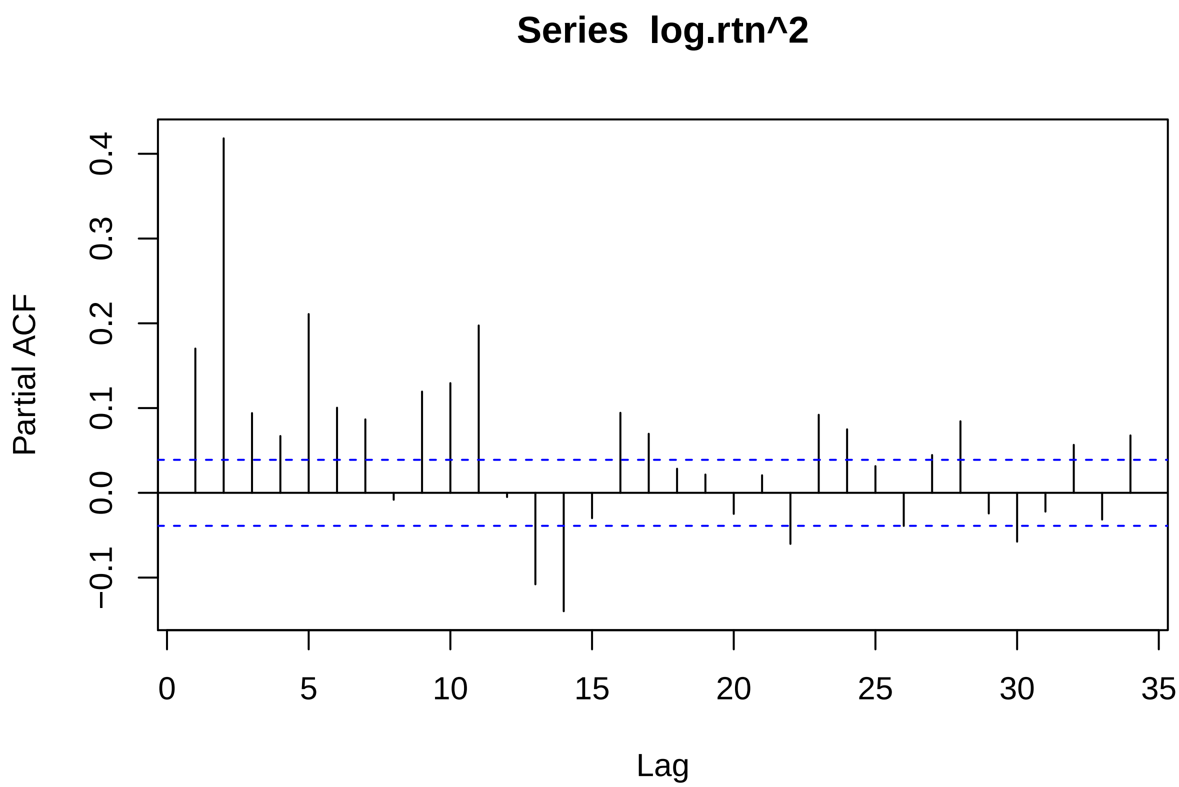
b). Fit Gaussian ARMA-GARCH model for the log-return series. Perform model checking. Obtain the QQ-plot of the standardized residuals. Write down the fitted model. [Hint: use GARCH(2,1)]

Look at ACF and PACF of the squared log returns.

```
acf(log.rtn^2)
```




```
pacf(log.rtn^2)
```



The PACF shows the 2nd lag is most significant. We can try ARCH(2). ACF shows many lags of significance. It is conventional to just try GARCH(1) for the sigma squared effect.
Fit GARCH(2,1) with Gaussian innovations.

```r
m1=garchFit(~1+garch(2,1),data=log.rtn,trace=F)
summary(m1)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~1 + garch(2, 1), data = log.rtn, trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(2, 1)
## <environment: 0x00000000185202b0>
##  [data = log.rtn]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu       omega      alpha1      alpha2       beta1
## 5.7243e-04  2.3226e-06  1.9896e-03  1.1165e-01  8.7049e-01
##
```

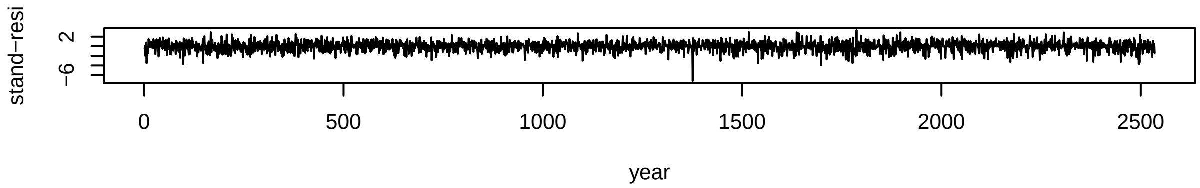
```
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     5.724e-04   1.707e-04    3.353   0.0008 ***
## omega  2.323e-06   4.938e-07    4.704 2.56e-06 ***
## alpha1 1.990e-03   1.031e-02    0.193   0.8469
## alpha2 1.117e-01   1.691e-02    6.604 4.01e-11 ***
## beta1  8.705e-01   1.457e-02   59.730  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  7976.237    normalized:  3.146445
##
## Description:
##  Tue Aug 11 03:23:38 2020 by user: zd000
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  341.8337  0
##  Shapiro-Wilk Test  R    W      0.9847918 7.758679e-16
##  Ljung-Box Test     R    Q(10)  16.68652  0.08159467
##  Ljung-Box Test     R    Q(15)  23.20357  0.07991225
##  Ljung-Box Test     R    Q(20)  27.01178  0.1349328
##  Ljung-Box Test     R^2  Q(10)  7.011437  0.7243648
##  Ljung-Box Test     R^2  Q(15)  8.756466  0.8899028
##  Ljung-Box Test     R^2  Q(20)  9.758849  0.9723368
##  LM Arch Test       R    TR^2   7.741879  0.8049624
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -6.288944 -6.277430 -6.288952 -6.284767
```

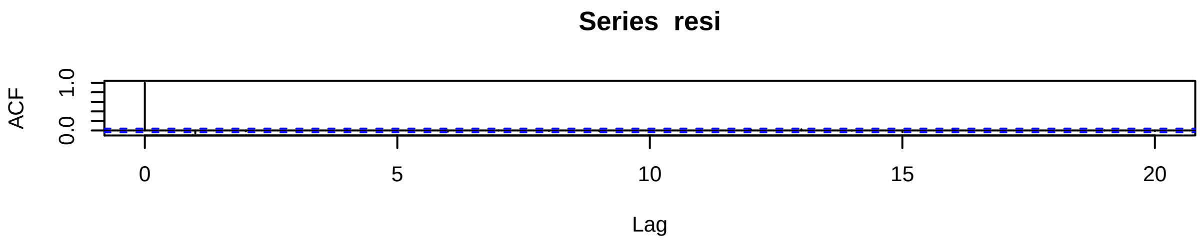
The resulting model is: $r_t = \mu + a_t$ $a_t = \sigma_t \epsilon_t$ $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \beta_1 \sigma_{t-1}^2$ $\mu =$ 5.7243e-04 $\alpha_0 =$ 2.3226e-06 $\alpha_1 =$ 1.9896e-03 α_2= 1.1165e-01 and β_1= 8.7049e-01 Note that alpha1 is not significant.

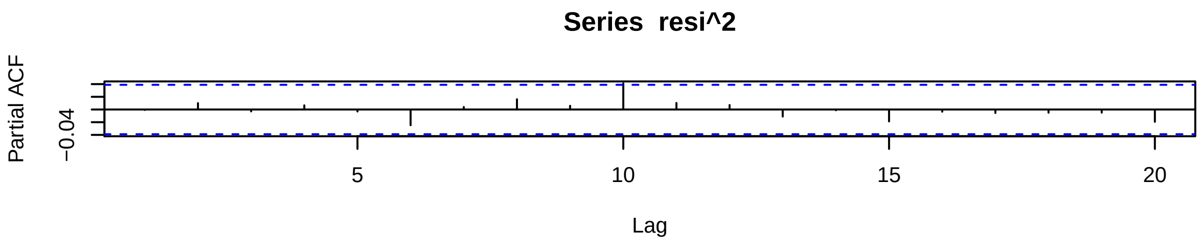
Based on the Ljung-Box Test, at 5% level and lag=10, 15 and 20, we can not reject the H0: there is zero auto correlation for reresiduals and residual squares.

Let us analyze the residuals

```
resi=residuals(m1,standardize=T)
tdx=c(1:length(resi))
par(mfcol=c(3,1))
plot(tdx,resi,xlab='year',ylab='stand-resi',type='l')
acf(resi,lag=20)
pacf(resi^2,lag=20)
```

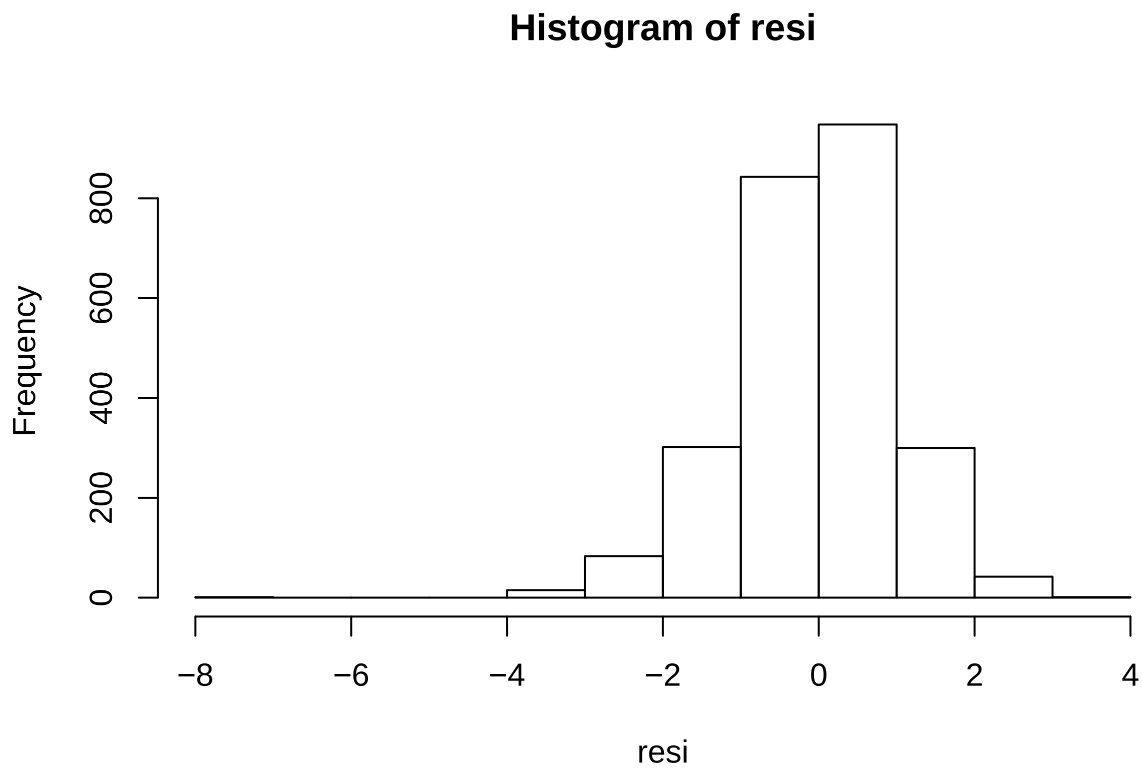




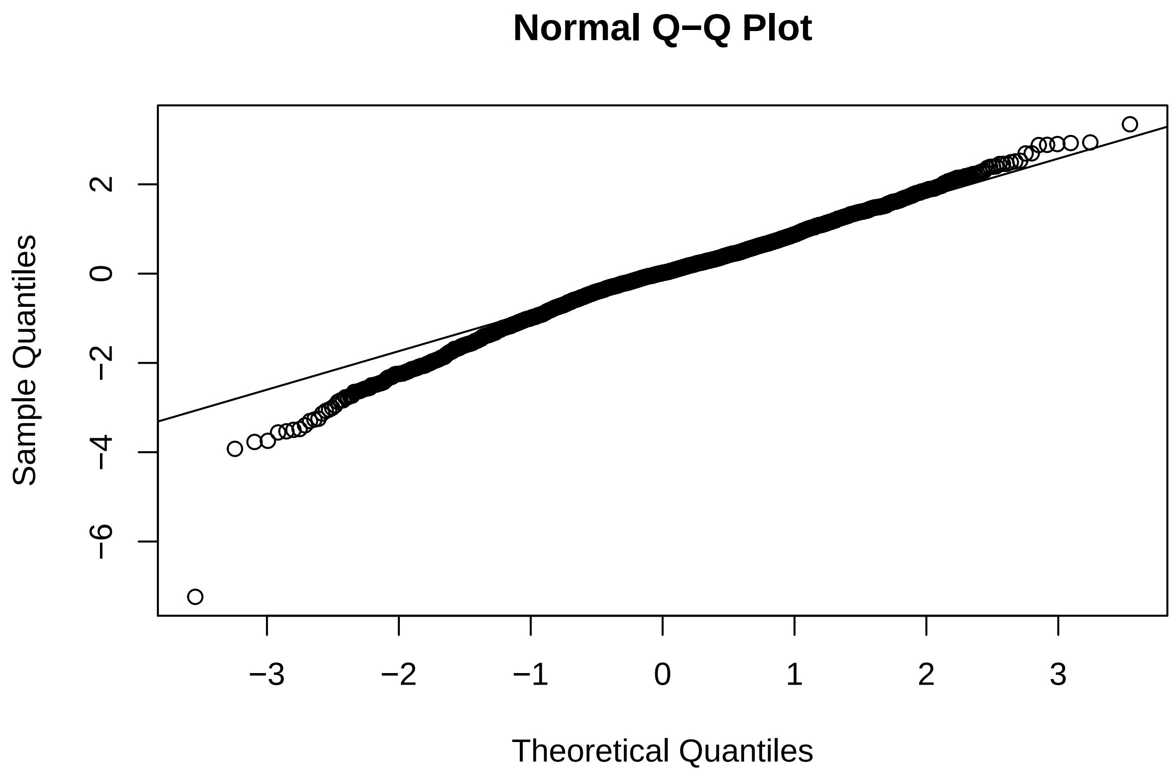


The ACF and PACF confirms that standardized residuals do not have autocorrelation in the first 20 lags. Plot the QQ-plot and histogram of the standardized residuals.

```r
hist(resi)
```



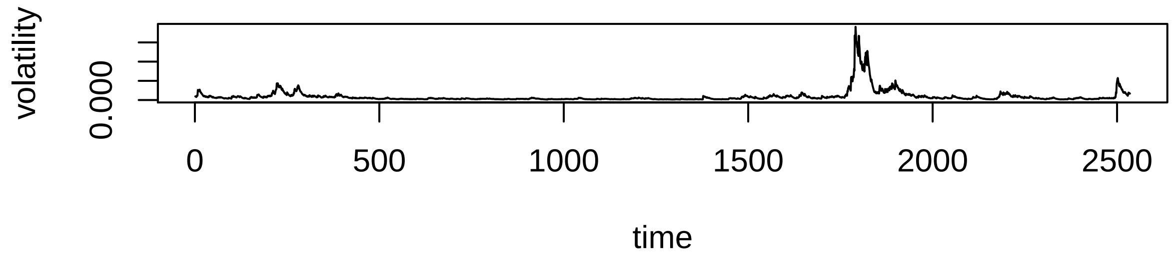

```
qqnorm(resi)
qqline(resi)
```

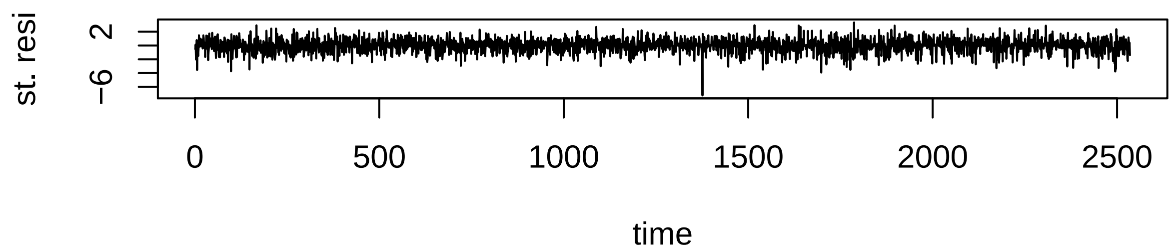


The qq plot and histogram of the standardized residuals show a left skewness and some kurtosis. This is also confirmed by the rejection of H0 from both Jarque-Bera Test and Shapiro-Wilk Test.

Plot standardized residuals and conditional variances.

```
v1<-m1@h.t
vol=ts(v1)
res=ts(resi)
par(mfcol=c(2,1))  # Show volatility and residuals
plot(vol,xlab='time',ylab='volatility',type='l')
plot(res,xlab='time',ylab='st. resi',type='l')
```

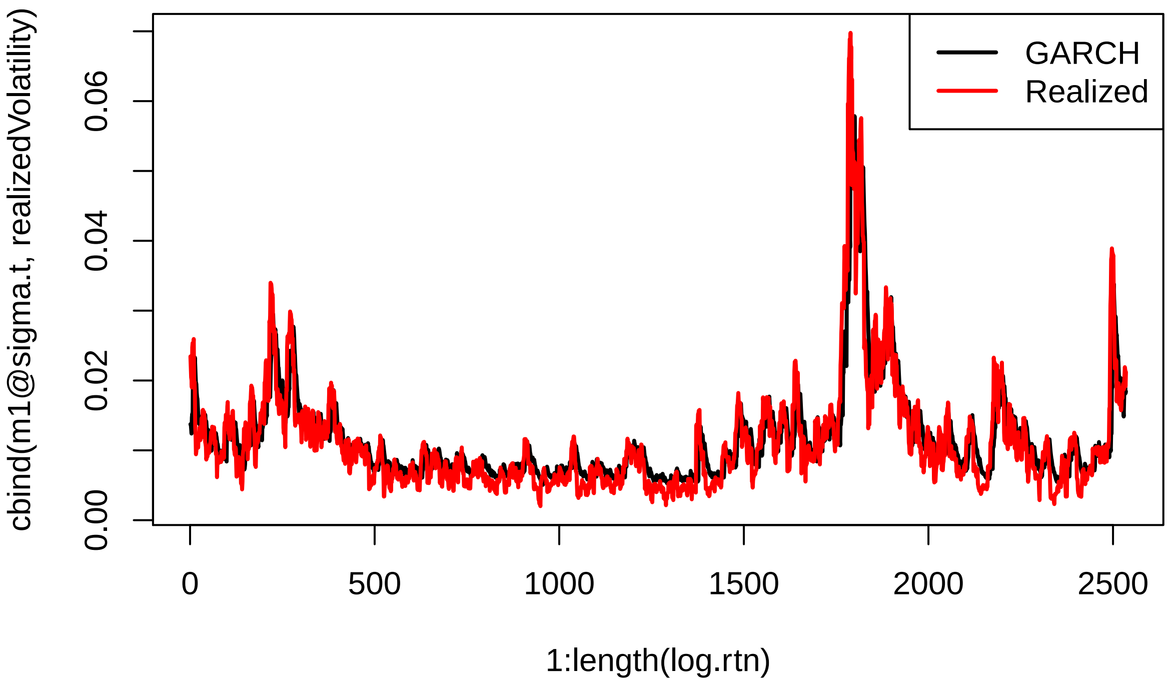




Standardized residuals have only a few outliers, look pretty good. Conditional variances show characteristic jumps and clustering. Volatility was high during the 2008 ressesion.

Compare volatility estimated by GARCH with historical realized volatility.

```
realizedVolatility<-rollapply(log.rtn,width=10,partial=T,by=1, sd)
matplot(1:length(log.rtn),cbind(m1@sigma.t,realizedVolatility),type="l",lty=1,lwd=2)
legend("topright",legend=c("GARCH","Realized"),
       lty=1,col=c("black","red"),lwd=2)
```

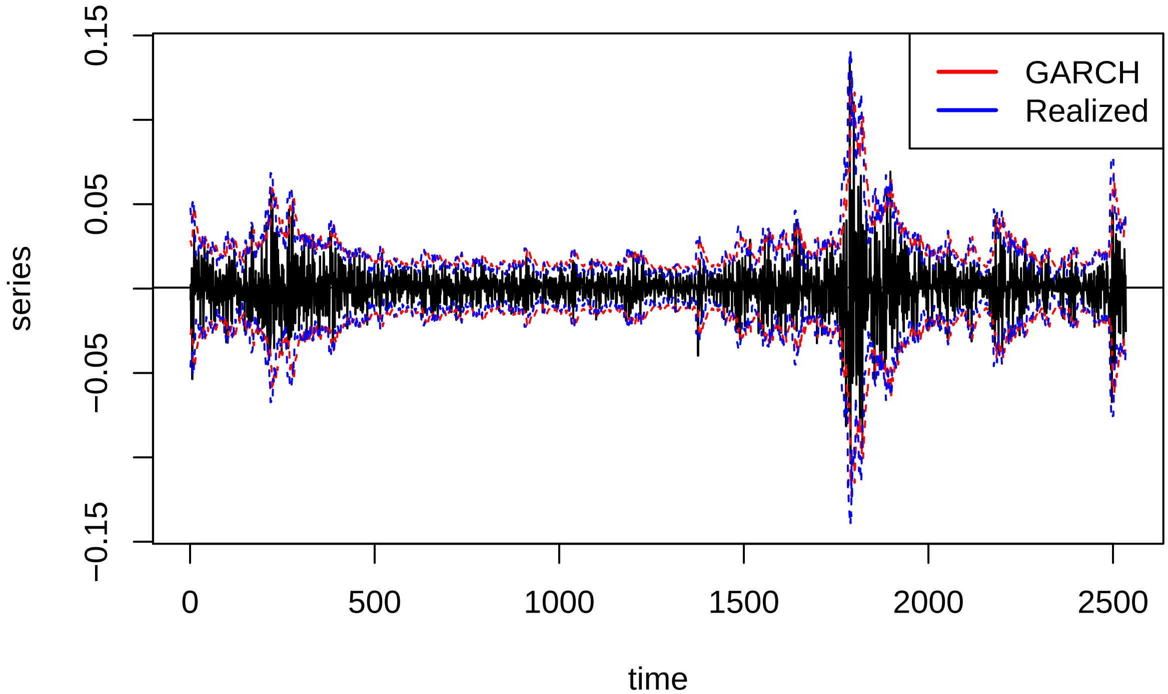



```
cor(cbind(sigma=m1@sigma.t,RV=realizedVolatility))
```

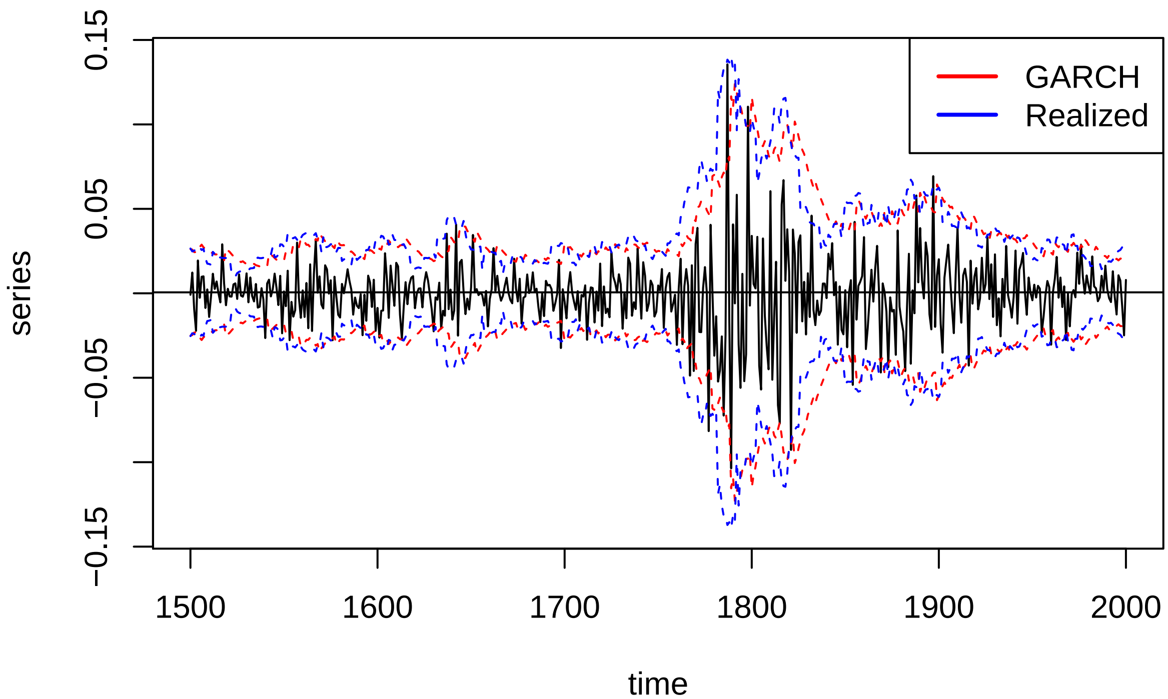
```
##          sigma       RV
## sigma 1.000000 0.887338
## RV    0.887338 1.000000
```

The Predicted Velotility is very similar to realized velotility, the correlation is very high 0.887338.

```
# Obtain plot of predictive intervals
par(mfcol=c(1,1))
uppGarch=m1@fitted+2*sqrt(v1)
lowGarch=m1@fitted-2*sqrt(v1)
uppReal=m1@fitted+2*realizedVolatility
lowReal=m1@fitted-2*realizedVolatility
tdx=c(1:length(resi))
plot(tdx,log.rtn,xlab='time',ylab='series',type='l',ylim=c(-0.14,0.14))
lines(tdx,uppGarch,lty=2,col='red')
lines(tdx,lowGarch,lty=2,col='red')
lines(tdx,uppReal,lty=2,col='blue')
lines(tdx,lowReal,lty=2,col='blue')
abline(h=m1@fitted[1])
legend("topright",legend=c("GARCH","Realized"),
       lty=1,col=c("red","blue"),lwd=2)
```




```
plot(tdx[1500:2000],log.rtn[1500:2000],
     xlab='time',ylab='series',type='l',ylim=c(-0.14,0.14))
lines(tdx[1500:2000],uppGarch[1500:2000],lty=2,col='red')
lines(tdx[1500:2000],lowGarch[1500:2000],lty=2,col='red')
lines(tdx[1500:2000],uppReal[1500:2000],lty=2,col='blue')
lines(tdx[1500:2000],lowReal[1500:2000],lty=2,col='blue')
abline(h=m1@fitted[1])
legend("topright",legend=c("GARCH","Realized"),
       lty=1,col=c("red","blue"),lwd=2)
```



Realized volatility shows stronger and faster reaction to shocks as well as stronger and faster decline than predicted volatility. The Predicted Volatility captures most of the variance magnitude very well even during the shock period. The GARCH(2,1) model is adequate and can be used for prediction.

c). Build an ARMA-GARCH model with Student t innovations for the log-return series. Perform model checking and write down the fitted model.

Obtain the same model with Student innovations.

```
# Student-t innovations
m2=garchFit(~1+garch(2,1),data=log.rtn,trace=F,cond.dist="std")
summary(m2)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~1 + garch(2, 1), data = log.rtn, cond.dist = "std",
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(2, 1)
## <environment: 0x0000000018505040>
##  [data = log.rtn]
```

```
##
## Conditional Distribution:
##  std
##
## Coefficient(s):
##         mu       omega      alpha1      alpha2       beta1       shape
## 7.2454e-04  1.5793e-06  6.1117e-03  1.1435e-01  8.7418e-01  7.5068e+00
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     7.245e-04   1.613e-04    4.493 7.01e-06 ***
## omega  1.579e-06   4.988e-07    3.166  0.00155 **
## alpha1 6.112e-03   1.272e-02    0.480  0.63094
## alpha2 1.143e-01   2.086e-02    5.482 4.21e-08 ***
## beta1  8.742e-01   1.629e-02   53.657  < 2e-16 ***
## shape  7.507e+00   1.137e+00    6.604 4.00e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  8015.731    normalized:  3.162024
##
## Description:
##  Tue Aug 11 03:23:39 2020 by user: zd000
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  484.6936  0
##  Shapiro-Wilk Test  R    W      0.982789  0
##  Ljung-Box Test     R    Q(10)  16.4353   0.08783114
##  Ljung-Box Test     R    Q(15)  22.44416  0.09667569
##  Ljung-Box Test     R    Q(20)  26.27182  0.1570209
##  Ljung-Box Test     R^2  Q(10)  5.461128  0.8583262
##  Ljung-Box Test     R^2  Q(15)  7.906807  0.9274645
##  Ljung-Box Test     R^2  Q(20)  9.649639  0.9740892
##  LM Arch Test       R    TR^2   6.005807  0.9157891
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -6.319314 -6.305497 -6.319326 -6.314302
```

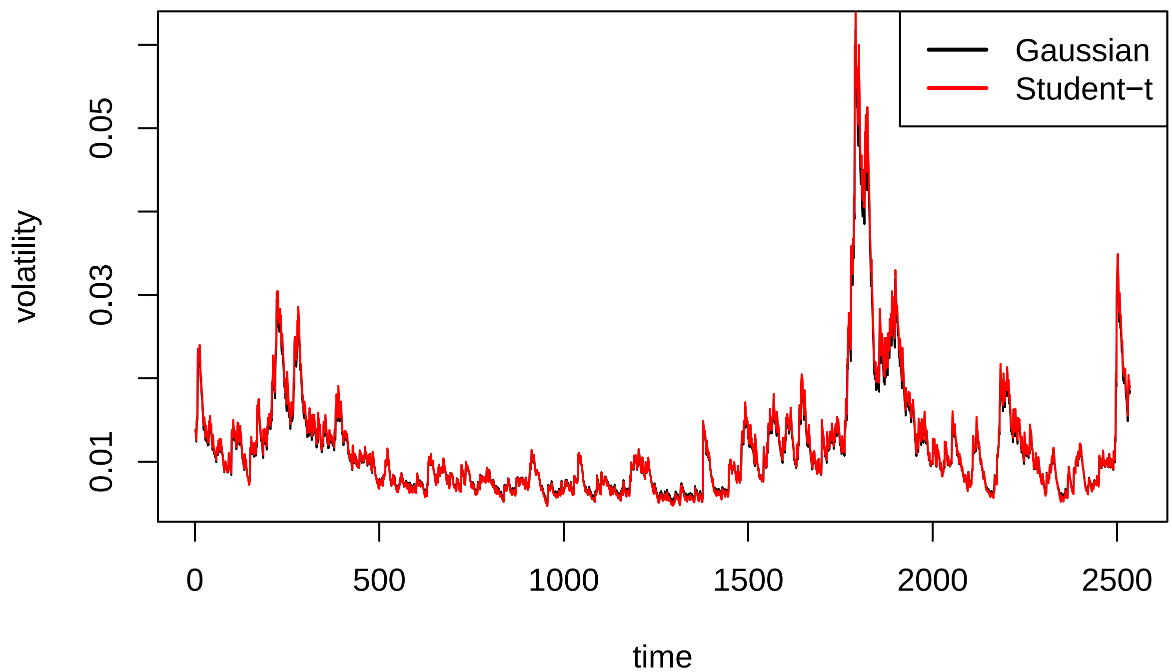
The resulting model is: $r_t = \mu + a_t$ $a_t = \sigma_t \epsilon_t$ $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \beta_1 \sigma_{t-1}^2$ μ = 7.2454e-04 α_0 = 1.5793e-06 α_1 = 6.1117e-03 α_2= 1.1435e-01 and β_1= 7.5068e+00
Note that alpha1 is still not significant but improved.

Based on the Ljung-Box Test, at 5% level and lag=10, 15 and 20, we can not reject the H0: there is zero auto correlation for reresiduals and residual squares.

The same model with the student innovaion produce smaller AIC and BIC. Since we confirmed kurtosis from the previous standarized residual analysis, the smaller AIC and BIC from the new model are expected.

Plot and compare the volatilities estimated by the 2 models.

```r
v2<-m2@h.t
plot(tdx,sqrt(v1),xlab='time',ylab='volatility',type='l')
lines(tdx,sqrt(v2),xlab='time',ylab='volatility',type='l',col="red")
legend("topright",legend=c("Gaussian","Student-t"), lty=1,col=c("black","red"),lwd=2)
```



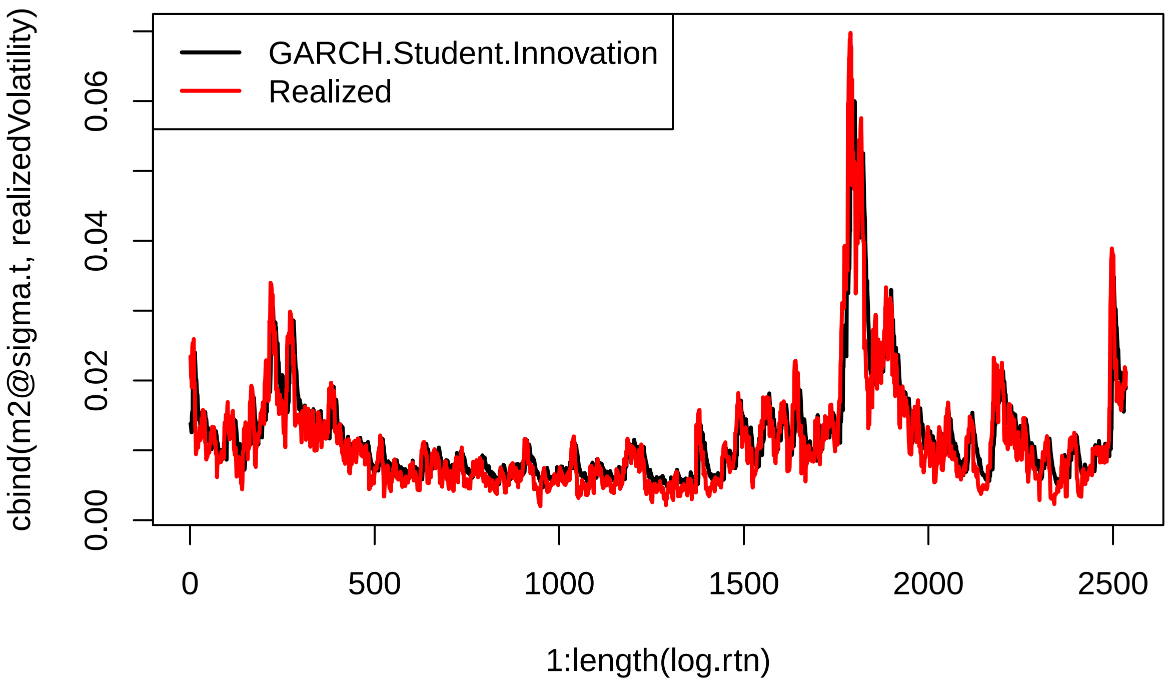

```r
cor(cbind(v1,v2))
```

```
##           v1        v2
## v1 1.0000000 0.9999218
## v2 0.9999218 1.0000000
```

Both model volatility prediction are practically the same. Cross-correlation matrix shows that they are almost perfectly correlated.

Volatility estimated by GARCH(2,1) with student innocations vs historical realized volatility.

```r
realizedVolatility<-rollapply(log.rtn,width=10,partial=T,by=1, sd)
matplot(1:length(log.rtn),cbind(m2@sigma.t,realizedVolatility),type="l",lty=1,lwd=2)
legend("topleft",legend=c("GARCH.Student.Innovation","Realized"),
       lty=1,col=c("black","red"),lwd=2)
```

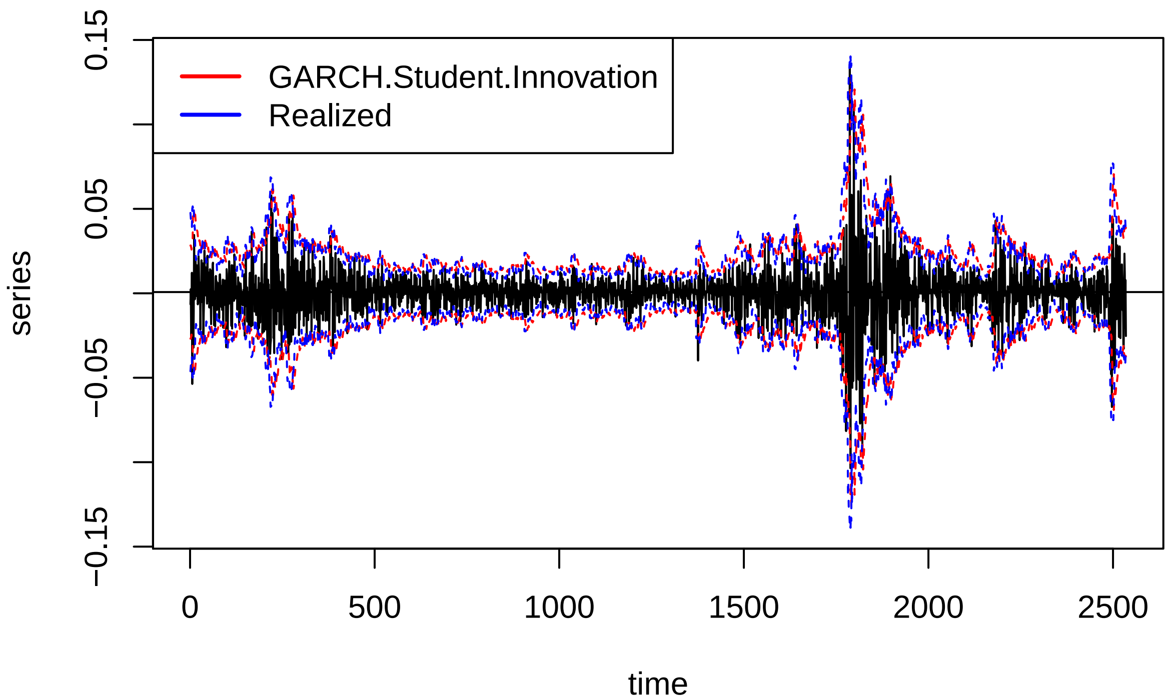



```
cor(cbind(sigma=m2@sigma.t,RV=realizedVolatility))
```

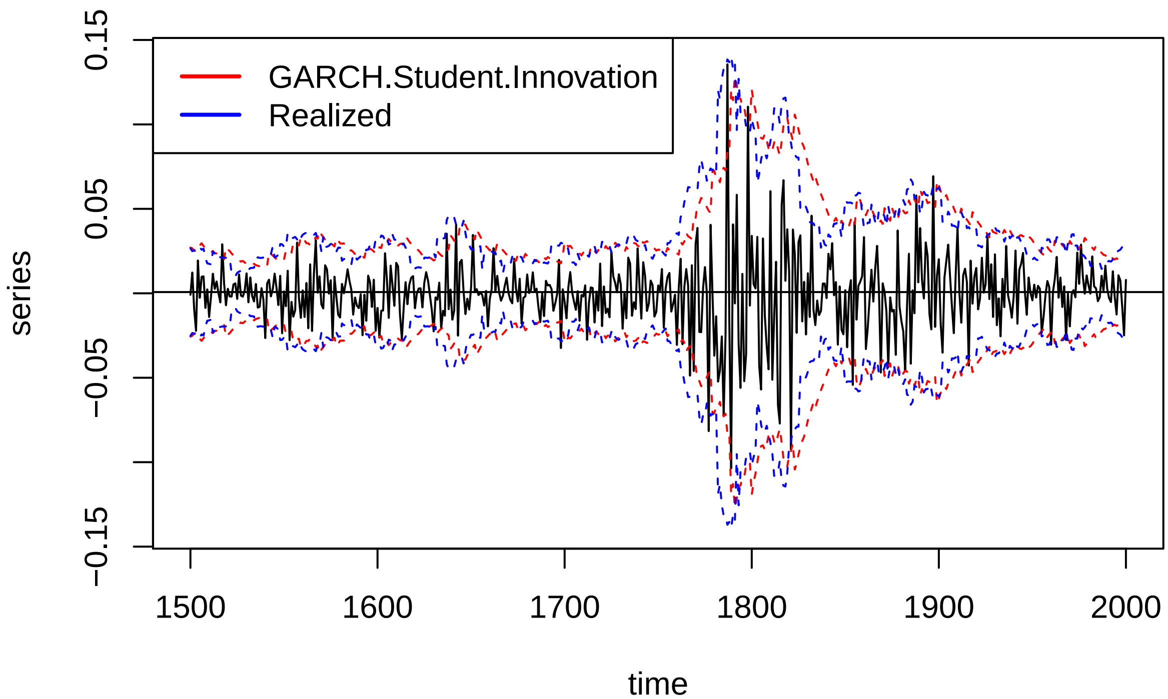
```
##           sigma        RV
## sigma 1.0000000 0.8880311
## RV    0.8880311 1.0000000
```

The Predicted Velotility of GARCH(2,1) Student Innovation correlates even better with the realized velotility.

```
# Obtain plot of predictive intervals
par(mfcol=c(1,1))
uppGarch.SI=m2@fitted+2*sqrt(v2)
lowGarch.SI=m2@fitted-2*sqrt(v2)
uppReal1=m2@fitted+2*realizedVolatility
lowReal1=m2@fitted-2*realizedVolatility
tdx=c(1:length(resi))
plot(tdx,log.rtn,xlab='time',ylab='series',type='l',ylim=c(-0.14,0.14))
lines(tdx,uppGarch.SI,lty=2,col='red')
lines(tdx,lowGarch.SI,lty=2,col='red')
lines(tdx,uppReal1,lty=2,col='blue')
lines(tdx,lowReal1,lty=2,col='blue')
abline(h=m2@fitted[1])
legend("topleft",legend=c("GARCH.Student.Innovation","Realized"),
       lty=1,col=c("red","blue"),lwd=2)
```




```r
plot(tdx[1500:2000],log.rtn[1500:2000],
     xlab='time',ylab='series',type='l',ylim=c(-0.14,0.14))
lines(tdx[1500:2000],uppGarch.SI[1500:2000],lty=2,col='red')
lines(tdx[1500:2000],lowGarch.SI[1500:2000],lty=2,col='red')
lines(tdx[1500:2000],uppReal1[1500:2000],lty=2,col='blue')
lines(tdx[1500:2000],lowReal1[1500:2000],lty=2,col='blue')
abline(h=m2@fitted[1])
legend("topleft",legend=c("GARCH.Student.Innovation","Realized"),
       lty=1,col=c("red","blue"),lwd=2)
```



Like the original GARCH(2,1) model, Realized volatility shows stronger and faster reaction to shocks as well as stronger and faster decline than predicted volatility from the new model. The Predicted Volatility captures most of the variance magnitude very well even during the shock period. The GARCH(2,1) Student Innovations model is also great for predicting this data set.