Time Series Analysis of Stock Returns

2024-04-09

```
library(fpp2)
library(forecast)
library(fGarch)
library(astsa)
library(tseries)
```

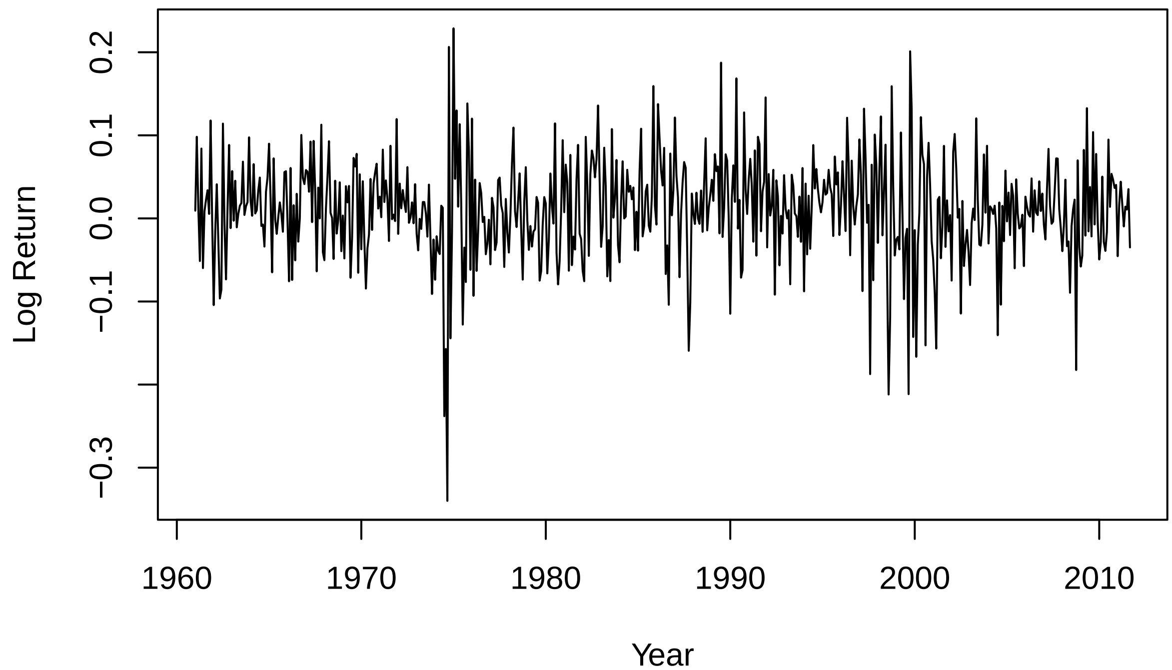
1)

The data set coca-cola.txt contains monthly stock returns of the Coca-Cola Company from January 1961 to September 2011. The simple returns are stored in the column called ko. In this analysis, we will transform the simple returns to log returns and analyze the time series for evidence of volatility clusters, serial correlation, and ARCH effects. We will also fit ARCH and GARCH models to the log returns and select the best model based on diagnostic checks and information criteria.

```
data <- read.table("coca-cola.txt", header = TRUE)
rt <- ts(log(data$ko+1), start = c(1961, 1), frequency = 12)
```

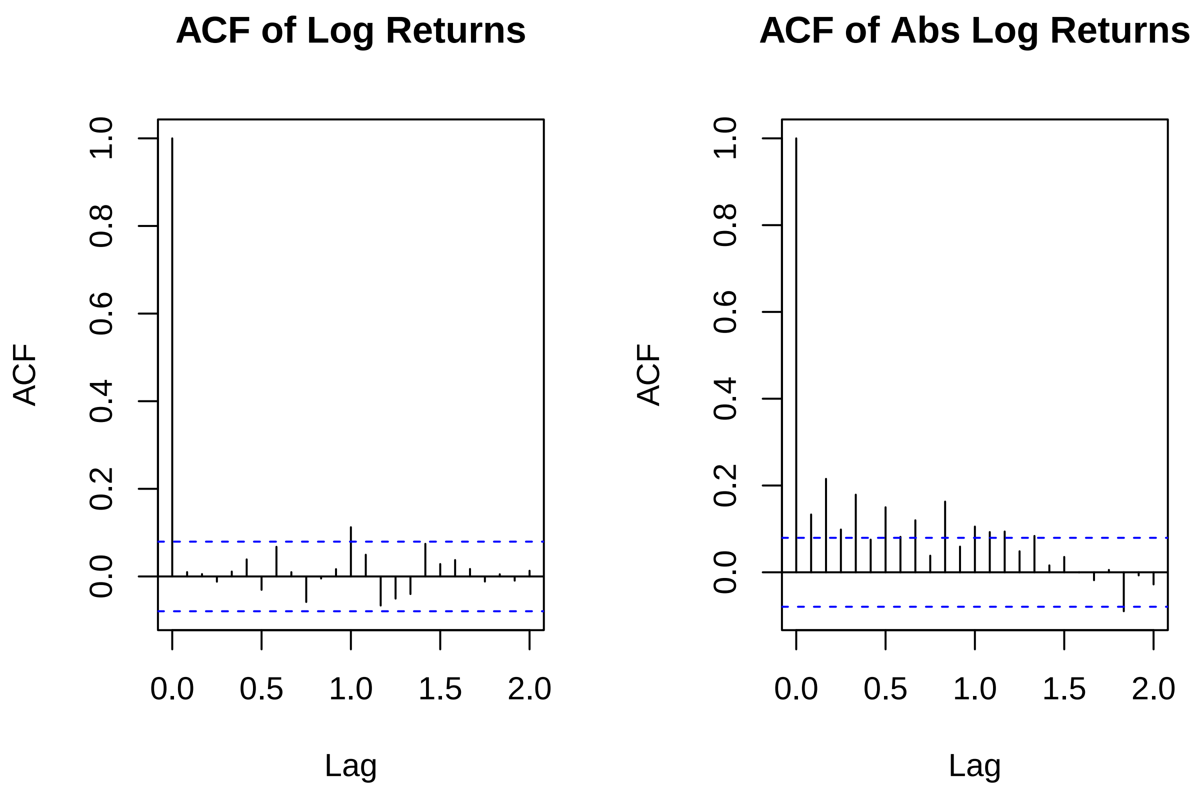
We transform the simple returns to log returns and convert them to a time series object for further analysis.

```
plot.ts(rt, ylab = "Log Return", xlab = "Year")
```



We see the presence of volatility clusters– the variance appears to be higher around 1975 and 2000. This suggests that the variance of the returns is not constant over time, which is typical for financial time series.

```r
par(mfcol=c(1,2))
acf(rt,lag=24, main="ACF of Log Returns")
acf(abs(rt),lag=24, main="ACF of Abs Log Returns")
```



The ACF plot of the log returns indicates an uncorrelated series, suggesting that the log returns are serially uncorrelated.

The ACF plot of the absolute value of the log returns indicates a dependent series as it has significant ACFs, which is consistent with the presence of volatility clustering.

```
t.test(rt)
```

```
##
##  One Sample t-test
##
## data:  rt
## t = 4.2198, df = 608, p-value = 2.819e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.005655242 0.015501584
## sample estimates:
##  mean of x
## 0.01057841
```

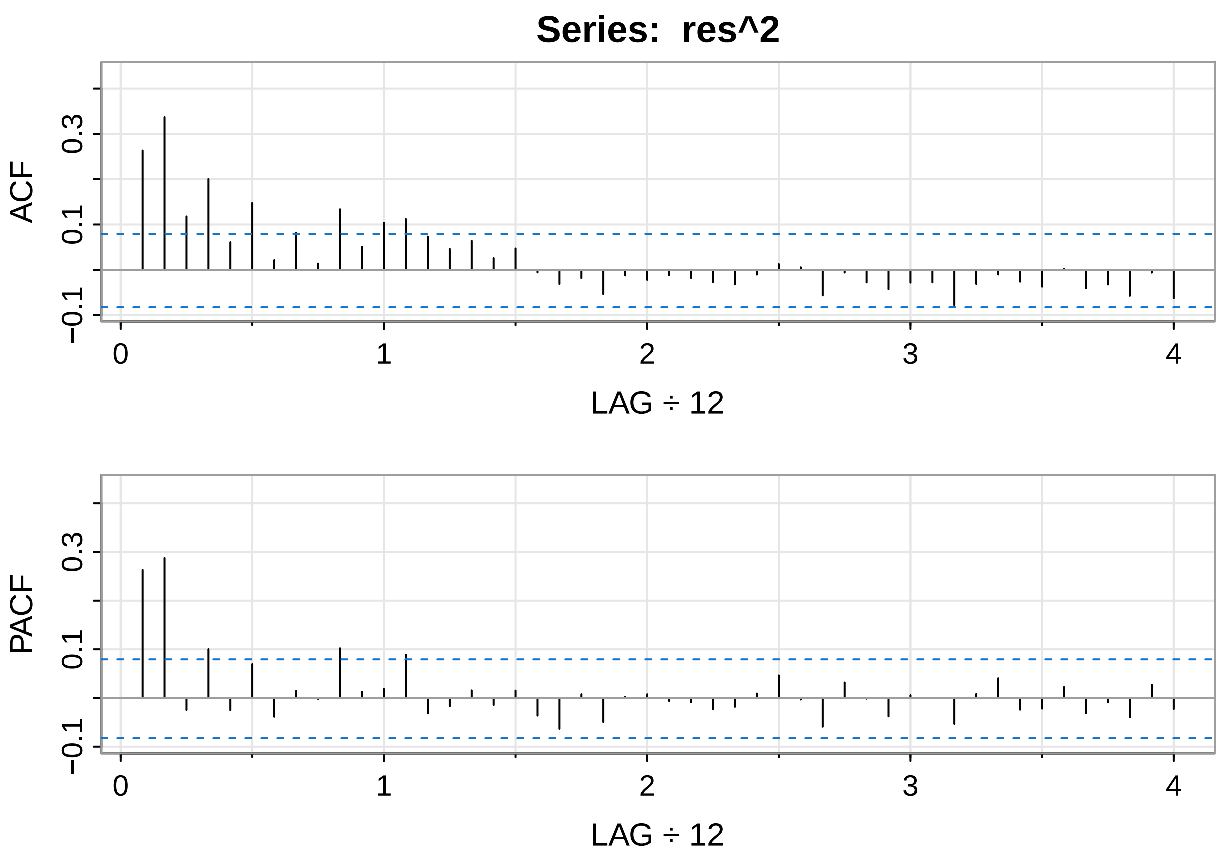
The mean of the log returns is significantly different from 0, indicating that there is a non-zero expected return for Coca-Cola stock during the period analyzed.

```
res<-rt-mean(rt)
Box.test(res^2,lag=12,type='Ljung')
```

```
##
##  Box-Ljung test
##
## data:  res^2
## X-squared = 184.93, df = 12, p-value < 2.2e-16
```

The Ljung test is significant, indicating the presence of ARCH effects in the log returns. This suggests that modeling the conditional variance with an ARCH model is appropriate.

```r
astsa::acf2(res^2)
```




```
##       [,1] [,2]  [,3] [,4]  [,5] [,6]  [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.26 0.34  0.12  0.2  0.06 0.15  0.02 0.08 0.01  0.13  0.05  0.10  0.11
## PACF 0.26 0.29 -0.02  0.1 -0.03 0.07 -0.04 0.01 0.00  0.10  0.01  0.02  0.09
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.07  0.05  0.06  0.03  0.05 -0.01 -0.03 -0.02 -0.05 -0.01 -0.02 -0.01
## PACF -0.03 -0.02  0.02 -0.01  0.02 -0.04 -0.06  0.01 -0.05  0.00  0.01 -0.01
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF  -0.02 -0.03 -0.03 -0.01  0.01  0.01 -0.06 -0.01 -0.03 -0.04 -0.03 -0.03
## PACF -0.01 -0.02 -0.02  0.01  0.05  0.00 -0.06  0.03  0.00 -0.04  0.01  0.00
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48]
## ACF  -0.08 -0.03 -0.01 -0.03 -0.04  0.00 -0.04 -0.03 -0.06 -0.01 -0.06
## PACF -0.05  0.01  0.04 -0.02 -0.02  0.02 -0.03 -0.01 -0.04  0.03 -0.02
```

When examining the PACF plot, we see that the first 2 lags are significant, as well as lag 4, 10, and 13.

For the sake of starting with simpler models, we begin with ARCH(2)

```r
result2<-fGarch::garchFit(~1+garch(2,0), data=rt, trace=F)
summary(result2)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(2, 0), data = rt, trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(2, 0)
## <environment: 0x0000017a57c065d8>
##  [data = rt]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##        mu      omega     alpha1     alpha2
## 0.0121485  0.0027599  0.0552768  0.1865166
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     0.0121485   0.0023095    5.260 1.44e-07 ***
## omega  0.0027599   0.0002121   13.015  < 2e-16 ***
## alpha1 0.0552768   0.0290296    1.904 0.056890 .
## alpha2 0.1865166   0.0544875    3.423 0.000619 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  858.0511    normalized:  1.408951
##
## Description:
##  Mon Aug 12 18:54:14 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                 Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  93.8581873 0.000000e+00
##  Shapiro-Wilk Test  R    W       0.9804732 2.920460e-07
##  Ljung-Box Test     R    Q(10)   8.2390534 6.054991e-01
##  Ljung-Box Test     R    Q(15)  19.1088965 2.088416e-01
##  Ljung-Box Test     R    Q(20)  22.8550477 2.959680e-01
##  Ljung-Box Test     R^2  Q(10)  18.5330834 4.661003e-02
##  Ljung-Box Test     R^2  Q(15)  29.7820455 1.273214e-02
##  Ljung-Box Test     R^2  Q(20)  33.4991235 2.971943e-02
##  LM Arch Test       R    TR^2   18.5069704 1.011418e-01
```

```
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.804766 -2.775788 -2.804851 -2.793493
```

If the residuals are independent, the Ljung-Box tests for the residuals and squared residuals for all lags should be insignificant. This isn't the case for this model and therefore we deem it inadequate.

I chose to fit an ARCH(4) as the fourth lag was also significant, and it is the next simplest model we can try:

```r
result4<-fGarch::garchFit(~1+garch(4,0), data=rt, trace=F)
summary(result4)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(4, 0), data = rt, trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(4, 0)
## <environment: 0x0000017a4cc47090>
##  [data = rt]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##        mu      omega     alpha1     alpha2     alpha3     alpha4
## 0.0118011  0.0019974  0.0566803  0.1585269  0.0897013  0.1655745
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     0.0118011   0.0022052    5.351 8.72e-08 ***
## omega  0.0019974   0.0002587    7.721 1.15e-14 ***
## alpha1 0.0566803   0.0322304    1.759  0.07865 .
## alpha2 0.1585269   0.0536603    2.954  0.00313 **
## alpha3 0.0897013   0.0465835    1.926  0.05415 .
## alpha4 0.1655745   0.0582145    2.844  0.00445 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  866.6294    normalized:  1.423037
##
## Description:
##  Mon Aug 12 18:54:14 2024 by user: prana
##
##
```

```
## Standardised Residuals Tests:
##                                 Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  84.3726423 0.000000e+00
##  Shapiro-Wilk Test  R    W       0.9820426 8.205934e-07
##  Ljung-Box Test     R    Q(10)   7.8941291 6.391775e-01
##  Ljung-Box Test     R    Q(15)  17.1087898 3.124027e-01
##  Ljung-Box Test     R    Q(20)  20.9865198 3.979264e-01
##  Ljung-Box Test     R^2  Q(10)   9.0681283 5.256504e-01
##  Ljung-Box Test     R^2  Q(15)  14.7422574 4.701373e-01
##  Ljung-Box Test     R^2  Q(20)  18.2762342 5.692169e-01
##  LM Arch Test       R    TR^2   11.1129229 5.192668e-01
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.826369 -2.782903 -2.826561 -2.809460
```

The ARCH(4) model has satisfactory diagnostics as the Ljung-Box tests for the residuals and squared residuals for all lags are insignificant. However, some terms in the model are insignificant, which raises concerns about the model's efficiency.

Next, we fit a GARCH(1,1) model with normal white noise to see if it provides a better fit:

```r
result11<-fGarch::garchFit(~1+garch(1,1),data=rt,trace=F)
summary(result11)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(1, 1), data = rt, trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(1, 1)
## <environment: 0x0000017a4fc955d0>
##  [data = rt]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##        mu      omega     alpha1      beta1
## 0.0123677  0.0002592  0.0987809  0.8297573
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     1.237e-02   2.267e-03    5.455 4.90e-08 ***
## omega  2.592e-04   8.641e-05    3.000   0.0027 **
## alpha1 9.878e-02   2.261e-02    4.368 1.25e-05 ***
## beta1  8.298e-01   3.393e-02   24.458  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  869.3329    normalized:  1.427476
##
## Description:
##  Mon Aug 12 18:54:14 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                   Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  83.094389 0.0000000000
##  Shapiro-Wilk Test  R    W       0.982898 0.0000014707
##  Ljung-Box Test     R    Q(10)   9.877630 0.4512941449
##  Ljung-Box Test     R    Q(15)  18.695474 0.2278666489
##  Ljung-Box Test     R    Q(20)  21.543451 0.3657888818
##  Ljung-Box Test     R^2  Q(10)  12.543353 0.2503352632
##  Ljung-Box Test     R^2  Q(15)  13.048732 0.5985337527
##  Ljung-Box Test     R^2  Q(20)  14.330256 0.8133641860
##  LM Arch Test       R    TR^2   10.797458 0.5463518008
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.841816 -2.812838 -2.841901 -2.830543
```

The model diagnostics for the GARCH(1,1) model are all satisfactory– the Ljung-Box tests for the residuals and squared residuals for all lags are insignificant, indicating a good fit.

To explore alternative distributions, we fit GARCH(1,1) models with Student t and skewed Student t distributions for the error term:

```
result11std<-fGarch::garchFit(~1+garch(1,1),data=rt,trace=F, cond.dist = "std")
summary(result11std)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(1, 1), data = rt, cond.dist = "std",
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(1, 1)
## <environment: 0x0000017a50c49a38>
##  [data = rt]
##
## Conditional Distribution:
##  std
##
## Coefficient(s):
##         mu       omega      alpha1       beta1       shape
## 0.01270784  0.00022276  0.10421057  0.83594037  7.09676187
##
```

```
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     0.0127078   0.0021240    5.983 2.19e-09 ***
## omega  0.0002228   0.0000909    2.451 0.014257 *
## alpha1 0.1042106   0.0283507    3.676 0.000237 ***
## beta1  0.8359404   0.0380905   21.946  < 2e-16 ***
## shape  7.0967619   1.8562544    3.823 0.000132 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  881.8586    normalized:  1.448044
##
## Description:
##  Mon Aug 12 18:54:14 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                 Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  84.744570 0.000000e+00
##  Shapiro-Wilk Test  R    W       0.982921 1.494278e-06
##  Ljung-Box Test     R    Q(10)  10.285960 4.157735e-01
##  Ljung-Box Test     R    Q(15)  18.947594 2.161185e-01
##  Ljung-Box Test     R    Q(20)  21.619696 3.614983e-01
##  Ljung-Box Test     R^2  Q(10)  11.672337 3.075841e-01
##  Ljung-Box Test     R^2  Q(15)  11.947614 6.829899e-01
##  Ljung-Box Test     R^2  Q(20)  13.256380 8.661148e-01
##  LM Arch Test       R    TR^2    9.814007 6.322738e-01
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.879667 -2.843445 -2.879800 -2.865576
```

The Ljung-Box tests for the residuals and squared residuals for all lags are insignificant. Therefore, this model has satisfactory diagnostics.

```r
result11sstd<-fGarch::garchFit(~1+garch(1,1),data=rt,trace=F, cond.dist = "sstd")
summary(result11sstd)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(1, 1), data = rt, cond.dist = "sstd",
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ 1 + garch(1, 1)
## <environment: 0x0000017a4c2fcaf8>
```

```
##  [data = rt]
##
## Conditional Distribution:
##  sstd
##
## Coefficient(s):
##         mu       omega      alpha1       beta1        skew       shape
## 0.01205847  0.00022557  0.10244468  0.83580831  0.94768525  7.41293838
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     1.206e-02   2.244e-03    5.373 7.74e-08 ***
## omega  2.256e-04   9.111e-05    2.476 0.013293 *
## alpha1 1.024e-01   2.777e-02    3.689 0.000225 ***
## beta1  8.358e-01   3.810e-02   21.939  < 2e-16 ***
## skew   9.477e-01   5.651e-02   16.771  < 2e-16 ***
## shape  7.413e+00   2.047e+00    3.622 0.000292 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  882.2713    normalized:  1.448721
##
## Description:
##  Mon Aug 12 18:54:15 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                  Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  85.9768377 0.000000e+00
##  Shapiro-Wilk Test  R    W       0.9827566 1.334121e-06
##  Ljung-Box Test     R    Q(10)  10.2930366 4.151713e-01
##  Ljung-Box Test     R    Q(15)  18.9901797 2.141792e-01
##  Ljung-Box Test     R    Q(20)  21.7154877 3.561465e-01
##  Ljung-Box Test     R^2  Q(10)  11.8412771 2.958197e-01
##  Ljung-Box Test     R^2  Q(15)  12.1191114 6.699918e-01
##  Ljung-Box Test     R^2  Q(20)  13.4619514 8.566961e-01
##  LM Arch Test       R    TR^2   10.0044188 6.155730e-01
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.877738 -2.834272 -2.877930 -2.860829
```

The Ljung-Box tests for the residuals and squared residuals for all lags are insignificant. Therefore, this model has satisfactory diagnostics.

k)

ARCH(4): AIC -2.826369, BIC -2.782903, SIC -2.826561, HQIC -2.809460

GARCH(1,1)[Norm]: AIC -2.841816, BIC -2.812838, SIC -2.841901, HQIC -2.830543

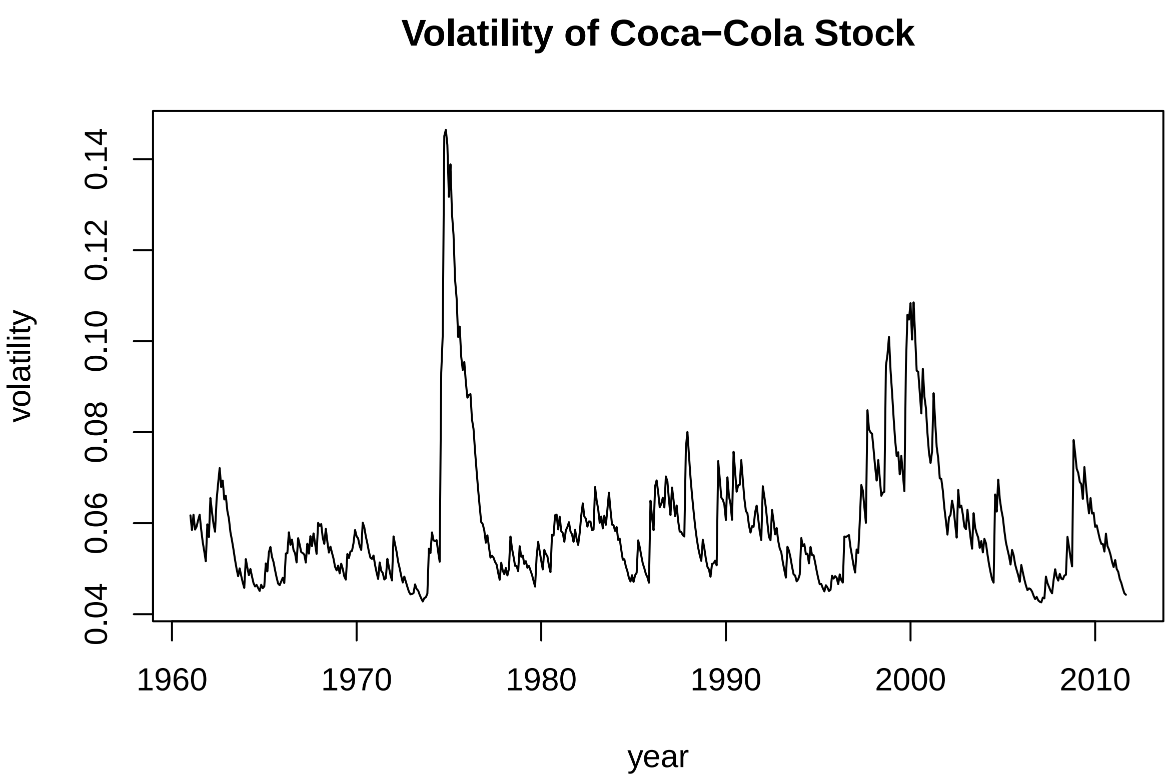
GARCH(1,1)[std]: AIC -2.879667 BIC -2.843445 SIC -2.879800 HQIC -2.865576

GARCH(1,1)[sstd]: AIC -2.877738 BIC -2.834272 SIC -2.877930 HQIC -2.860829

Although the ARCH(4) model has the lowest AIC and BIC, it had 2 insignificant terms. Therefore, I am choosing to go forward with the GARCH(1,1) model with the Student t distribution as it has the next lowest AIC/BIC and satisfactory diagnostics.

For the selected model, we plot the volatility over time to observe periods of higher volatility:

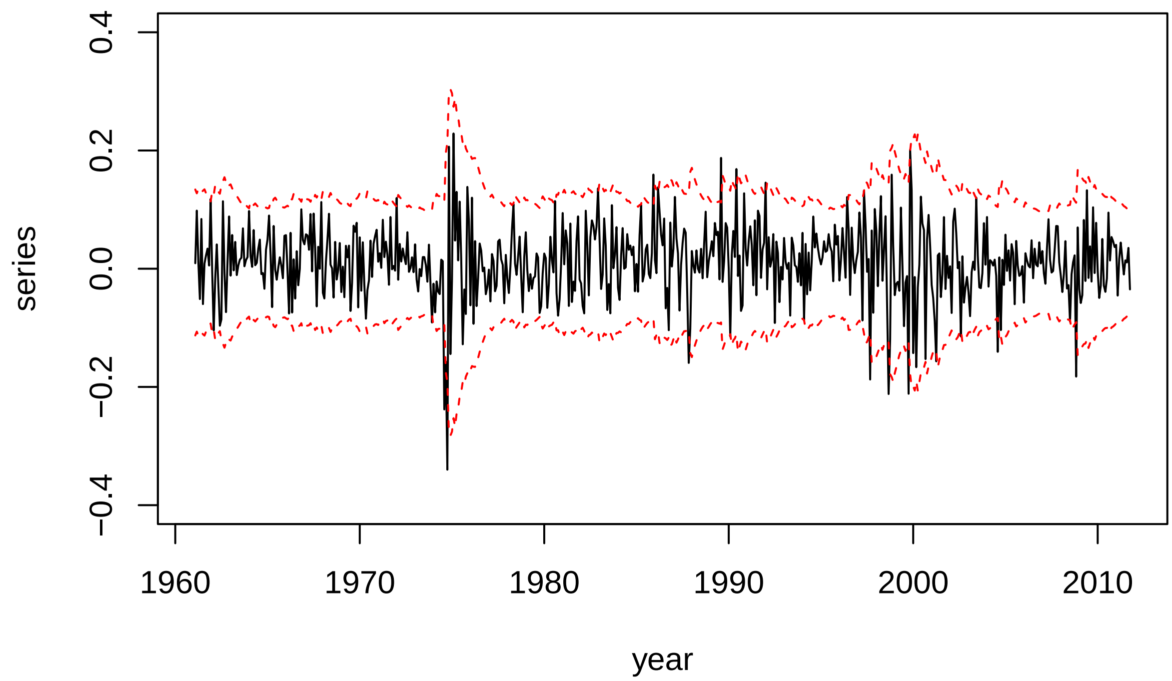
```
par(mfcol=c(1,1))

vol.result<-fGarch::volatility(result11)
vol.result<-ts(vol.result,frequency=12,start=c(1961,1))
plot(vol.result,xlab='year',ylab='volatility',type='l',
     main="Volatility of Coca-Cola Stock")
```



~1975 and around the 2000s seem to be more volatile, which aligns with historical economic events during those periods.

Next, we create a plot of the log returns with prediction intervals by adding and subtracting 2 times the estimated volatility to the mean.

```
upper<-0.01057841+2*vol.result ##0.01057841 is the mean that we found earlier
lower<-0.01057841-2*vol.result
tdx<-c(1:609)/12+1961
plot(tdx,rt,xlab='year',ylab='series',type='l',ylim=c(-0.4,0.4))
lines(tdx,upper,lty=2,col='red')
lines(tdx,lower,lty=2,col='red')
```



The prediction intervals do capture the log returns for the most part except the large negative spikes in log returns.

We forecast the volatility for the next 12 months using the selected GARCH(1,1) model.

```r
(predict(result11, n.ahead=12)[,3])^2
```

```
##  [1] 0.002106877 0.002215519 0.002316398 0.002410067 0.002497043 0.002577804
##  [7] 0.002652793 0.002722423 0.002787077 0.002847111 0.002902855 0.002954616
```

Finally, we calculate the long-term forecast of volatility as l to infinity and compare it with the 200-month ahead forecast.

```r
summary(result11)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~1 + garch(1, 1), data = rt, trace = F)
##
## Mean and Variance Equation:
```

```
##  data ~ 1 + garch(1, 1)
## <environment: 0x0000017a4fc955d0>
##  [data = rt]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu      omega     alpha1      beta1
## 0.0123677  0.0002592  0.0987809  0.8297573
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     1.237e-02   2.267e-03    5.455 4.90e-08 ***
## omega  2.592e-04   8.641e-05    3.000   0.0027 **
## alpha1 9.878e-02   2.261e-02    4.368 1.25e-05 ***
## beta1  8.298e-01   3.393e-02   24.458  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  869.3329    normalized:  1.427476
##
## Description:
##  Mon Aug 12 18:54:14 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                 Statistic      p-Value
##  Jarque-Bera Test   R    Chi^2  83.094389 0.0000000000
##  Shapiro-Wilk Test  R    W       0.982898 0.0000014707
##  Ljung-Box Test     R    Q(10)   9.877630 0.4512941449
##  Ljung-Box Test     R    Q(15)  18.695474 0.2278666489
##  Ljung-Box Test     R    Q(20)  21.543451 0.3657888818
##  Ljung-Box Test     R^2  Q(10)  12.543353 0.2503352632
##  Ljung-Box Test     R^2  Q(15)  13.048732 0.5985337527
##  Ljung-Box Test     R^2  Q(20)  14.330256 0.8133641860
##  LM Arch Test       R    TR^2   10.797458 0.5463518008
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.841816 -2.812838 -2.841901 -2.830543
```

```r
omega <-0.0002592
alpha1 <-0.0987809
beta1 <-0.8297573
omega/(1-alpha1-beta1)
```

```
## [1] 0.003627113
```

```r
(predict(result11, n.ahead=200)[,3][200])^2
```

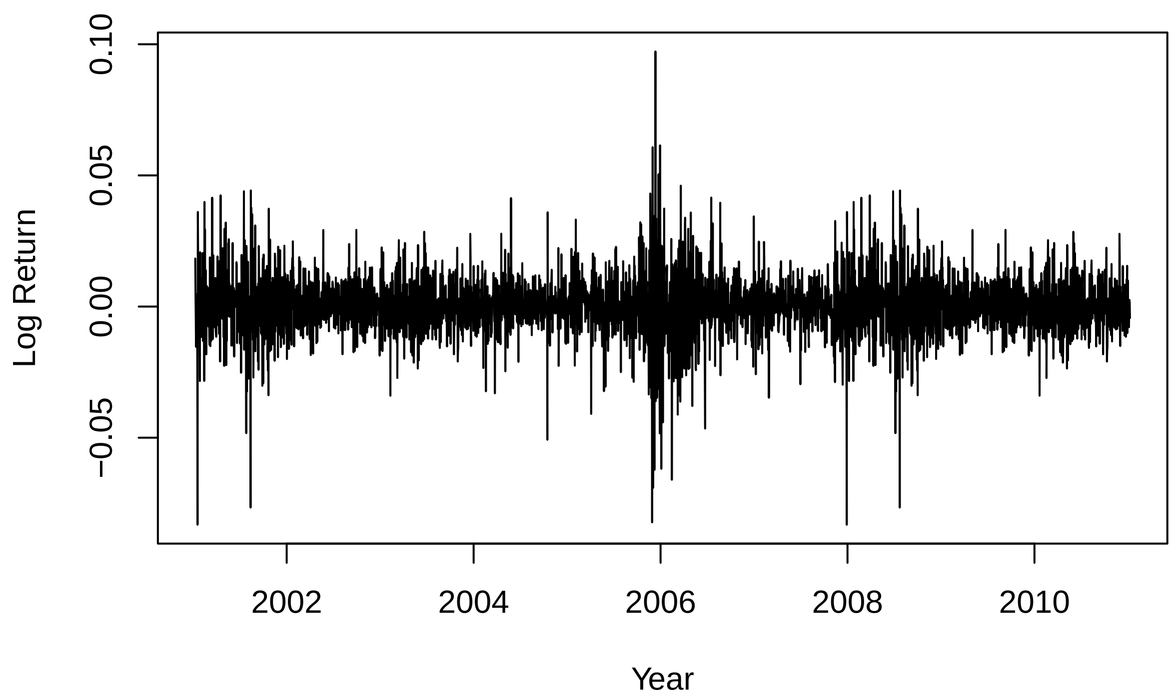
```
## [1] 0.003627162
```

The values are close to each other, confirming the model's consistency for long-term forecasts.

Now, we move on to the analysis of daily returns for Procter & Gamble (P&G) stock from September 1, 2001, to September 30, 2011. We transform the simple returns to log returns for analysis.

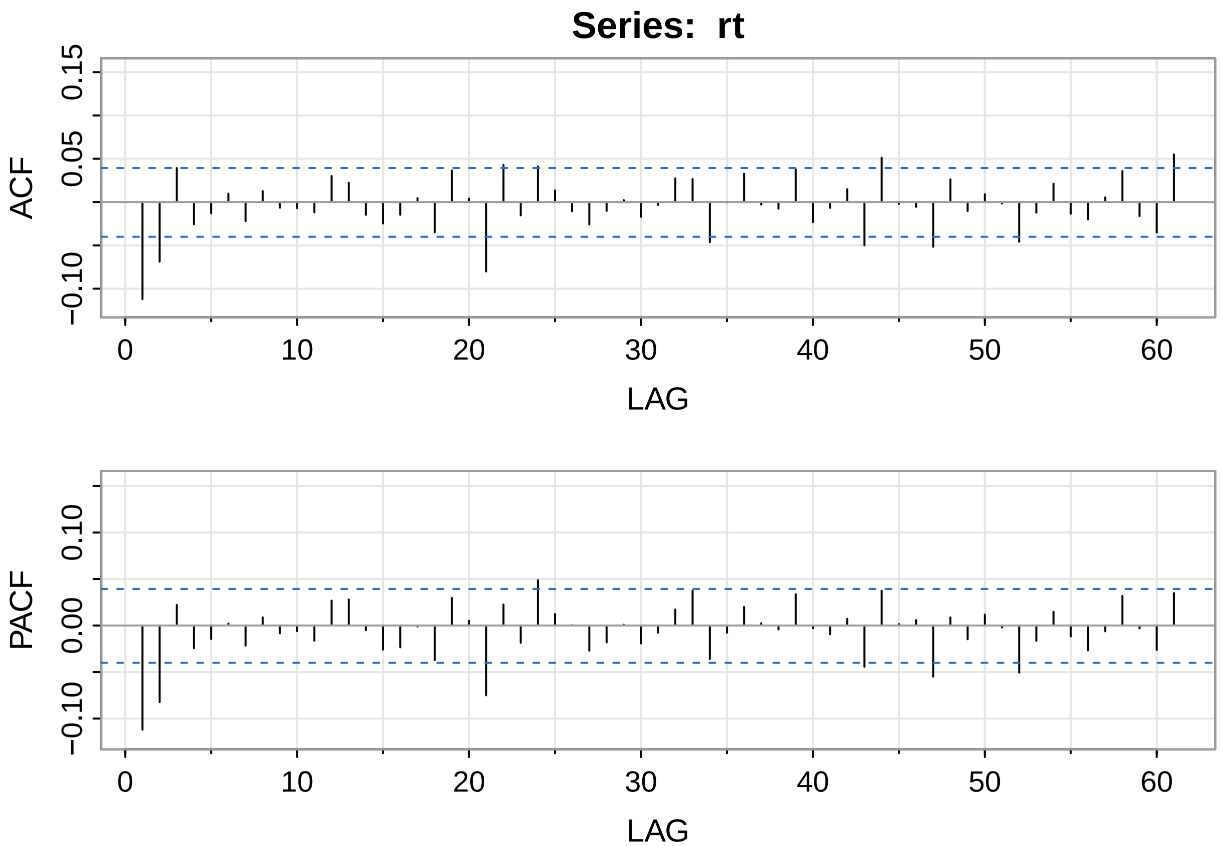
```r
data <- read.table("proctor-gamble.txt", header = TRUE)

rt<-log(data$rtn+1)
```

```r
rtts <- ts(log(data$rtn+1), start = c(2001, 9), end = c(2011, 9), frequency = 365)
plot.ts(rtts, ylab = "Log Return", xlab = "Year")
```



There is evidence of volatility clusters. We see an increased variance most notably around 2006 but there are also other periods (like late 2001 and 2008) where we see irregularities in variance.

```r
astsa::acf2(rt)
```




```
##        [,1]  [,2] [,3]  [,4]  [,5] [,6]  [,7] [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  -0.11 -0.07 0.04 -0.03 -0.01 0.01 -0.02 0.01 -0.01 -0.01 -0.01  0.03  0.02
## PACF -0.11 -0.08 0.02 -0.02 -0.01 0.00 -0.02 0.01 -0.01 -0.01 -0.02  0.03  0.03
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF  -0.01 -0.02 -0.01     0 -0.03  0.04  0.00 -0.08  0.04 -0.02  0.04  0.01
## PACF  0.00 -0.03 -0.02     0 -0.04  0.03  0.01 -0.08  0.02 -0.02  0.05  0.01
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF  -0.01 -0.03 -0.01     0 -0.02  0.00  0.03  0.03 -0.05  0.00  0.03     0
## PACF  0.00 -0.03 -0.02     0 -0.02 -0.01  0.02  0.04 -0.04 -0.01  0.02     0
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49]
## ACF  -0.01  0.04 -0.02 -0.01  0.01 -0.05  0.05     0 -0.01 -0.05  0.03 -0.01
## PACF  0.00  0.03  0.00 -0.01  0.01 -0.04  0.04     0  0.01 -0.05  0.01 -0.01
##      [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57] [,58] [,59] [,60] [,61]
## ACF   0.01     0 -0.05 -0.01  0.02 -0.01 -0.02  0.01  0.04 -0.02 -0.04  0.06
## PACF  0.01     0 -0.05 -0.02  0.01 -0.01 -0.03 -0.01  0.03  0.00 -0.03  0.03
```

As we can see from the ACF plot, the log returns are serially correlated. Significant lags suggest that the returns are not independent.

Given the significant lags at 1 and 2 in the PACF, an AR(2) model might be appropriate for the mean equation of the log returns.

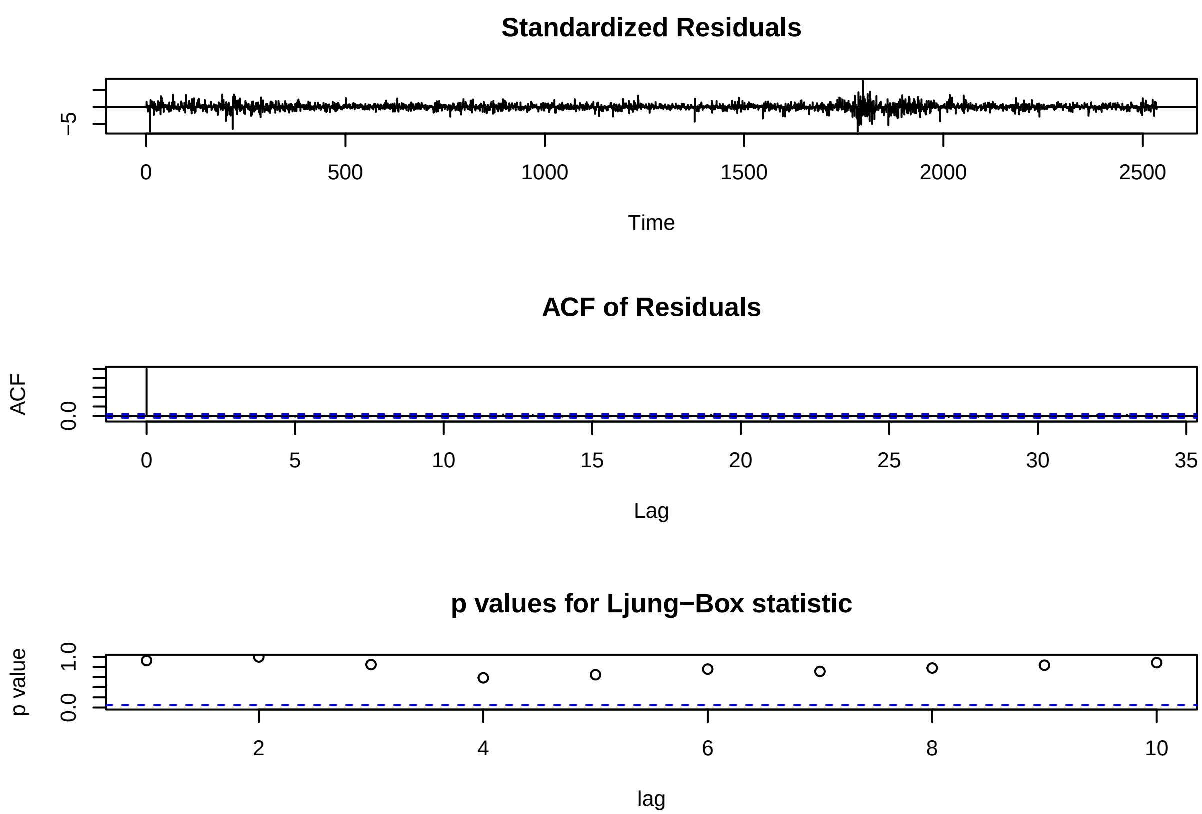
```
m.eqn<-arima(rt, order=c(2,0,0))
m.eqn
```

```
##
```

```
## Call:
## arima(x = rt, order = c(2, 0, 0))
##
## Coefficients:
##           ar1      ar2  intercept
##       -0.1214  -0.0824      3e-04
## s.e.   0.0198   0.0198      2e-04
##
## sigma^2 estimated as 0.0001414:  log likelihood = 7638.05,  aic = -15268.1
```

We fit an AR(2) model using the arima() function.

```
tsdiag(m.eqn)
```

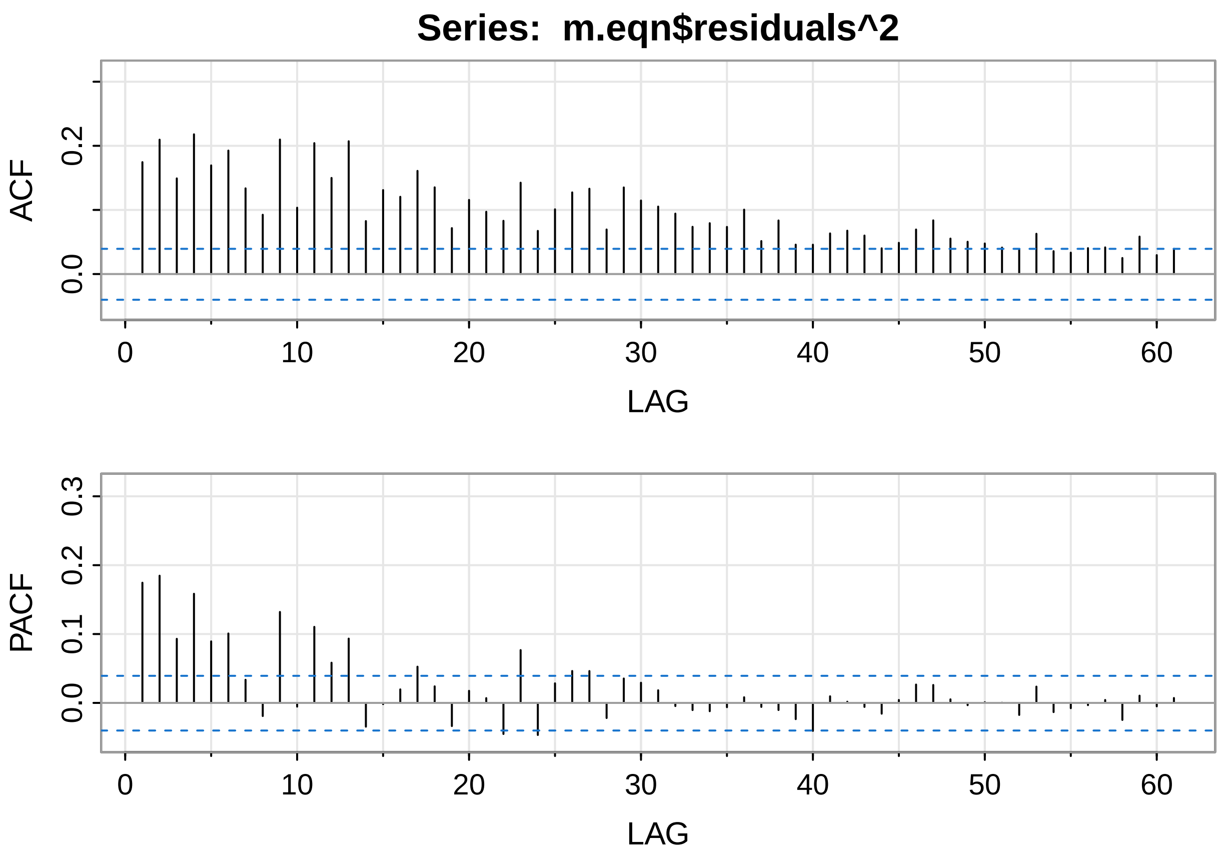


ACF of Residuals are all white noise and Ljung-Box p values are all insignificant, therefore we can conclude that AR(2) was appropriate.

```
Box.test(m.eqn$residuals^2, lag=10, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  m.eqn$residuals^2
## X-squared = 740.01, df = 10, p-value < 2.2e-16
```

```r
astsa::acf2(m.eqn$residuals^2)
```




```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]  [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.17 0.21 0.15 0.22 0.17 0.19 0.13  0.09 0.21  0.10  0.20  0.15  0.21
## PACF 0.17 0.18 0.09 0.16 0.09 0.10 0.03 -0.02 0.13 -0.01  0.11  0.06  0.09
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.08  0.13  0.12  0.16  0.14  0.07  0.12  0.10  0.08  0.14  0.07  0.10
## PACF -0.03  0.00  0.02  0.05  0.02 -0.03  0.02  0.01 -0.05  0.08 -0.05  0.03
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF   0.13  0.13  0.07  0.14  0.11  0.11  0.09  0.07  0.08  0.07  0.10  0.05
## PACF  0.05  0.05 -0.02  0.04  0.03  0.02  0.00 -0.01 -0.01 -0.01  0.01 -0.01
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49]
## ACF   0.08  0.05  0.05  0.06  0.07  0.06  0.04  0.05  0.07  0.08  0.06  0.05
## PACF -0.01 -0.02 -0.04  0.01  0.00 -0.01 -0.02  0.00  0.03  0.03  0.01  0.00
##      [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57] [,58] [,59] [,60] [,61]
## ACF   0.05  0.04  0.04  0.06  0.04  0.03  0.04  0.04  0.03  0.06  0.03  0.04
## PACF  0.00  0.00 -0.02  0.02 -0.01 -0.01  0.00  0.00 -0.02  0.01  0.00  0.01
```

Since the Ljung-Box test is significant and the squared residuals are correlated, we have evidence of ARCH effects.

```r
result<-fGarch::garchFit(~arma(2,0)+garch(1,1), data=rt, trace=F)
summary(result)
```

```
##
```

```
## Title:
##  GARCH Modelling
##
## Call:
##  fGarch::garchFit(formula = ~arma(2, 0) + garch(1, 1), data = rt,
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(2, 0) + garch(1, 1)
## <environment: 0x0000017a555dc370>
##  [data = rt]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##          mu          ar1          ar2        omega       alpha1        beta1
##  4.7559e-04  -8.2946e-02  -4.6892e-02   2.1291e-06   4.7260e-02   9.3491e-01
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##          Estimate  Std. Error  t value Pr(>|t|)
## mu      4.756e-04   1.946e-04    2.444   0.0145 *
## ar1    -8.295e-02   2.123e-02   -3.908 9.31e-05 ***
## ar2    -4.689e-02   2.111e-02   -2.222   0.0263 *
## omega   2.129e-06   6.575e-07    3.238   0.0012 **
## alpha1  4.726e-02   9.700e-03    4.872 1.10e-06 ***
## beta1   9.349e-01   1.392e-02   67.142  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  7934.971    normalized:  3.130166
##
## Description:
##  Mon Aug 12 18:54:17 2024 by user: prana
##
##
## Standardised Residuals Tests:
##                                   Statistic   p-Value
##  Jarque-Bera Test   R    Chi^2  1830.9659711 0.0000000
##  Shapiro-Wilk Test  R    W         0.9661574 0.0000000
##  Ljung-Box Test     R    Q(10)     1.1425598 0.9996839
##  Ljung-Box Test     R    Q(15)     5.7785264 0.9832326
##  Ljung-Box Test     R    Q(20)    11.6879150 0.9263990
##  Ljung-Box Test     R^2  Q(10)     8.3033415 0.5992334
##  Ljung-Box Test     R^2  Q(15)    10.2217879 0.8055617
##  Ljung-Box Test     R^2  Q(20)    14.9596703 0.7787106
##  LM Arch Test       R    TR^2      7.1847525 0.8451666
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
```

```
## -6.255599 -6.241781 -6.255610 -6.250586
```

The model has all satisfactory diagnostics as the Ljung-Box tests for the residuals and squared residuals for all lags are insignificant.

$$r_t = .0004756 - .08295a_{t-1} - .04689a_{t-2} + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = .000002129 + .047260a_{t-1}^2 + .93491\sigma_{t-1}^2$$

This is the final estimated ARMA(2,0)-GARCH(1,1) model for the P&G log returns.