Time Series 413, Assignment 6

Volatility Models (TS6)

Reed Ballesteros

Northwestern University SPS, Fall 2022
MSDS-413-DL
Instructor: Dr. Jamie D. Riggs, Ph.D
2022-10-31

The following list defines the data sets and their respective variables. The daily returns of Microsoft stock ($msft$) stock for the period given in the starter R script. The data are available from yahoo! finance (https://finance.yahoo.com/quote/MSFT/history?p=MSFT) and in the file **MSFT.csv**. "Date" "Open" "High" "Low" "Close" "Adj.Close" "Volume"

- Date: year month day
- Open: daily opening price
- High: daily high
- Low: daily low
- Close: daily closing price
- Adj.Close: adjusted daily closing price
- Volume: daily volume

The monthly returns of Boeing (BA) stock for the period given in the starter R script. The data are available from yahoo! finance (https://finance.yahoo.com/quote/BA/history?p=BA) and in the file **BA.csv**.

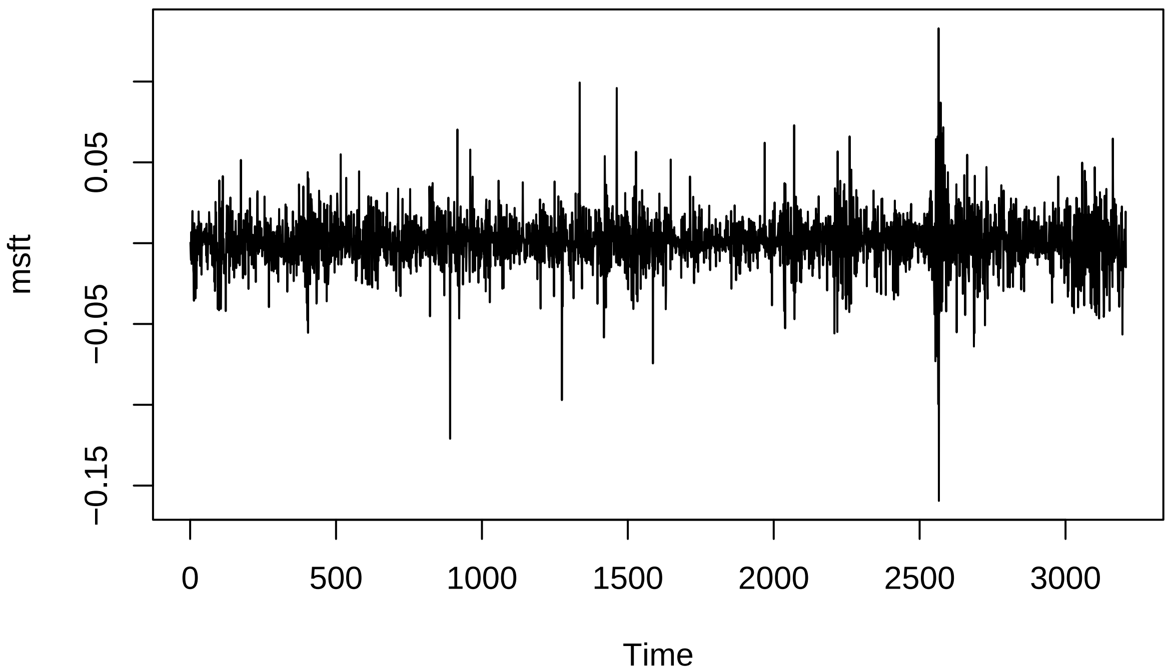
- Date: year month day
- Open: monthly opening price
- High: monthly high
- Low: monthly low
- Close: daily closing price
- Adj.Close: adjusted monthly closing price
- Volume: monthly volume

1. MSFT (30 points)

Consider the daily returns (msft) of Microsoft stock from January 4, 2010 to April 17, 2020. Construct time series models of the Microsoft daily adjusted closing price as follows.

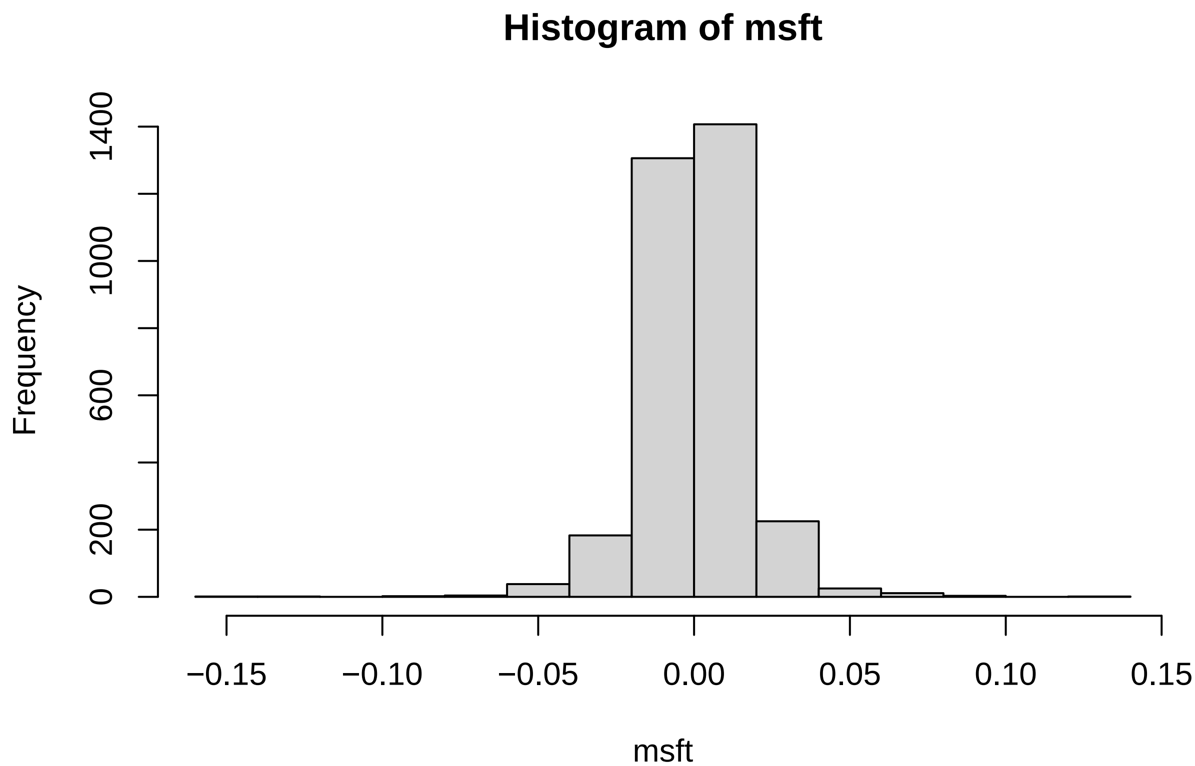
1.1. Use EDA on the msft daily log returns. Is the expected log return zero? Why? Are there any serial correlations in the log returns? Why?

Let us plot the msft daily log returns:



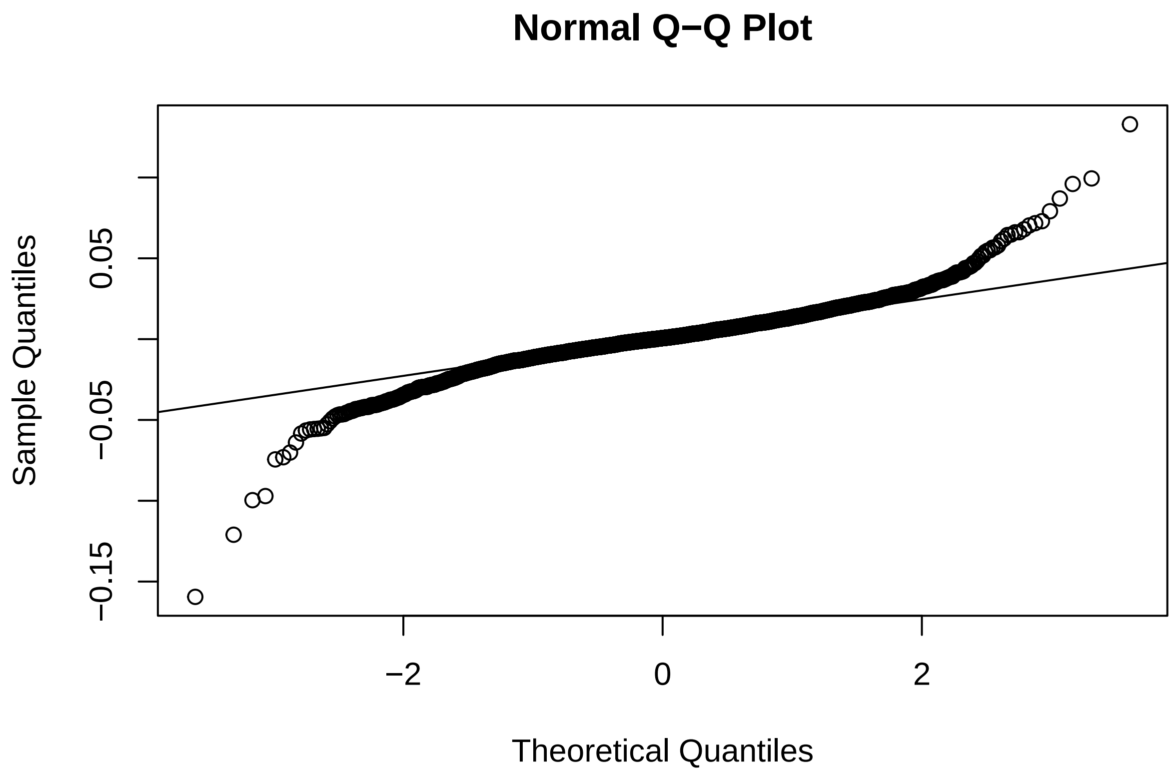
The plot of log(msft) may show mean 0, and the outliers in the plot might display non-contant variance.

Histogram - log(msft):



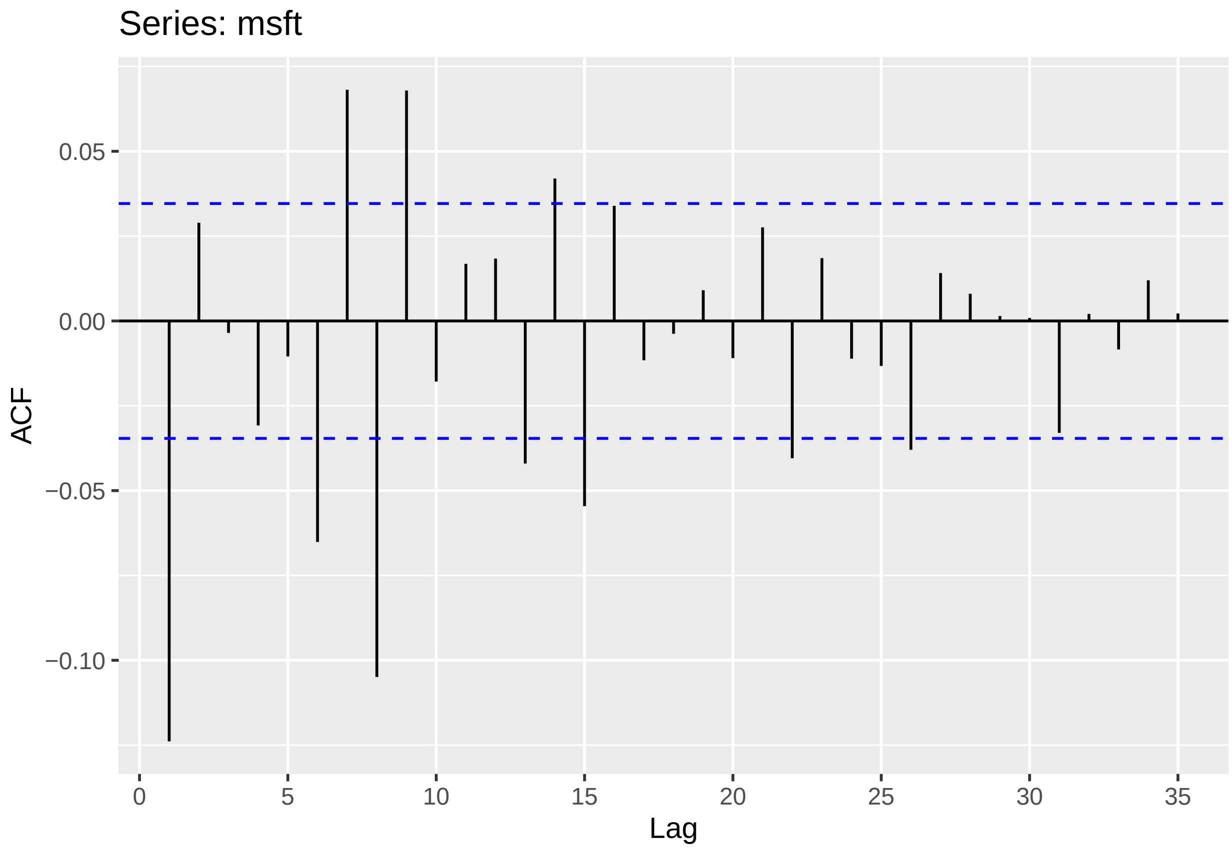
We can observe a left-skewed, tall distribution, thus showing non-normalcy in respect to a Gaussian PDF.

Q-Q Plot - log(msft):



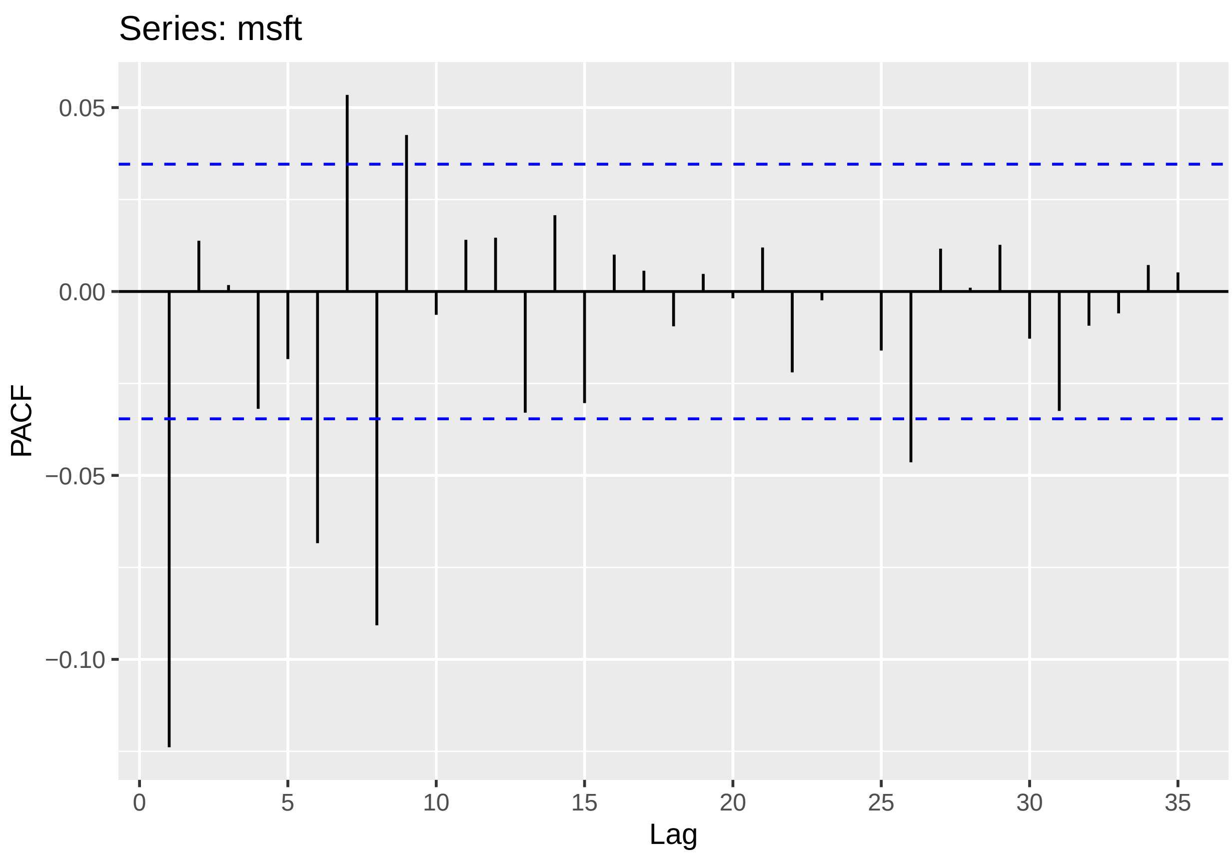
We can observe very thick tails, indicating very tall Kurtosis, thus showing non-normalcy in respect to a Gaussian PDF.

ACF Plot - log(msft):



We observe a fairly stationary ACF plot, and from it we determine to use an MA(1) component for an ARMA model.

PACF Plot - log(msft):



We observe a fairly stationary ACF plot, and from it we determine to use an AR(1) component for an ARMA model.

T-Test Mean of Zero - log(msft):

```
##
##  One Sample t-test
##
## data:  data
## t = 2.5181, df = 3206, p-value = 0.01185
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.0001589344 0.0012770451
## sample estimates:
##    mean of x
## 0.0007179897
##
## T-Test: mean *NOT* statistically zero, linear trend present ->
## reject H0
```

```
## [1] FALSE
```

While the 95% confidence internval (CI) does not include 0, the lower and upper bounds are quite close to zero, as well as the mean of x. In this case, we would call the expected mean of log(msft) to be very close to zero, in some ways accepted to be zero, with alsmot no signs of a linear trend.

Normalcy: Skewness - log(msft):

```
##       skew     lwr.ci     upr.ci
## -0.2421876 -0.2499237 -0.2036944
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The log(msft) data has a distribution with left skewness, thus showing non-normalcy in respect to a Gaussian PDF.

Normalcy: (excess) Kurtosis - log(msft):

```
##     kurt   lwr.ci   upr.ci
## 8.822661 8.873920 9.201317
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The log(msft) data has a distribution with tall (excess) Kurtosis, thus showing non-normalcy in respect to a Gaussian PDF.

Constant Variance: Breusch-Pagan Test log(msft)

```
##
##  studentized Breusch-Pagan test
##
## data:  lm(data ~ seq(1, length(data)))
## BP = 19.048, df = 1, p-value = 1.275e-05
##
## Breusch-Pagan: *NON*-constant variance, possible clustering,
## heteroscedastic -> reject H0
```

```
##    BP
## FALSE
```

While we can see signs of non-constant variance in the time series plot, the Breusch-Pagen test confirms it.

Correlation: Box-Ljung test - log(msft)

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 178.4, df = 30, p-value < 2.2e-16
##
## Box-Ljung: implies dependency present over 30 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

With a Box-Ljung test p-value < 0.05 for the log(msft) data, we observe lag dependency and thus serial autorrelation.

1.2. Write a mean model (arima to forecast expected values) to be fitted based on your EDA. Construct a mean model for the log returns. Is there an ARCH effect in the log return series? Why?

We create the following mean model based on the EDA above using AR(1) and MA(1):

```
p <- 1; d <- 0; q <- 1
m <- Arima(msft,order=c(p,d,q),include.mean=TRUE)
```

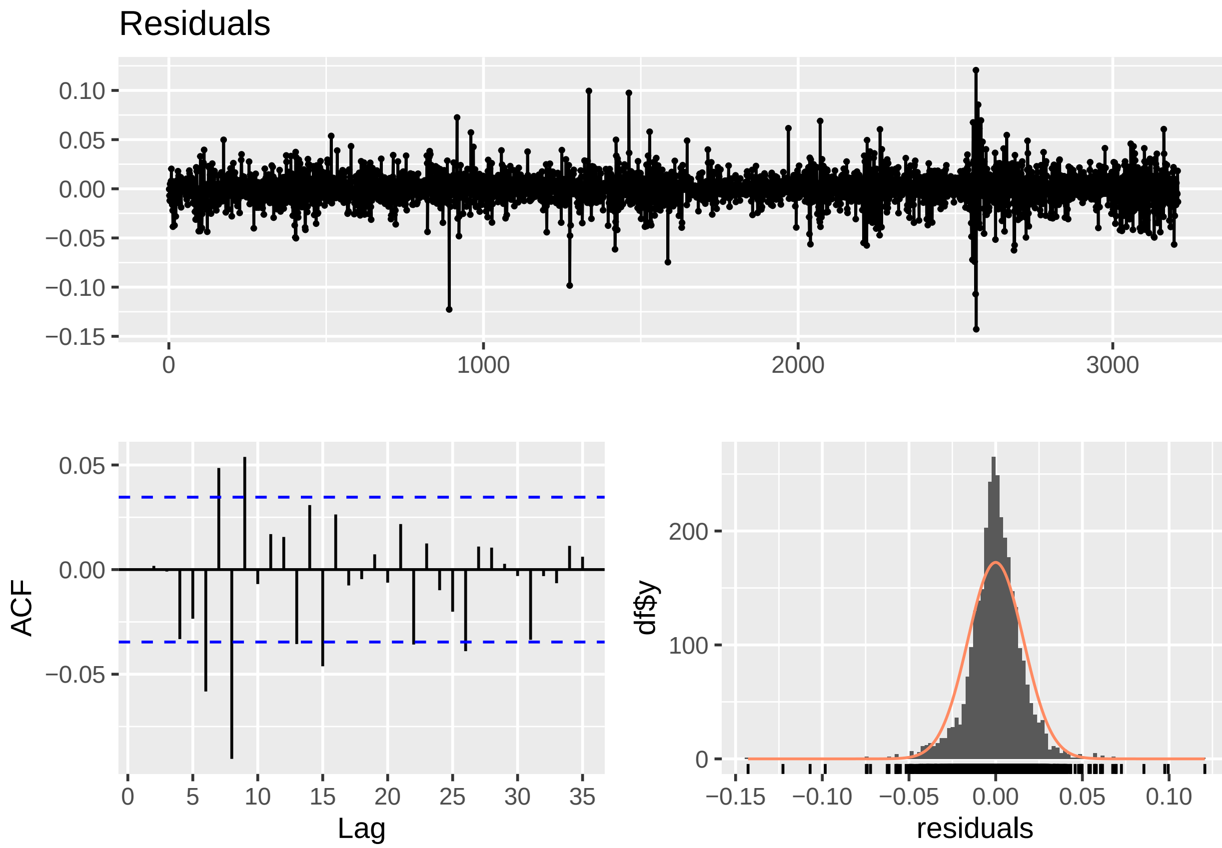
```
## Series: msft
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##           ar1     ma1   mean
##       -0.2260  0.1037  7e-04
## s.e.   0.1256  0.1281  3e-04
##
## sigma^2 = 0.0002568:  log likelihood = 8707.2
## AIC=-17406.39   AICc=-17406.38   BIC=-17382.1
##
## Training set error measures:
##                         ME       RMSE        MAE MPE MAPE      MASE
## Training set -2.663597e-06 0.01601867 0.01110696 NaN  Inf 0.6679811
##                      ACF1
## Training set 3.339609e-05
```

Based on the summary above, we create the following equation:

$$y_t = \phi_1 y_{t-1} + \theta_1 z_{t-1} + \mu$$

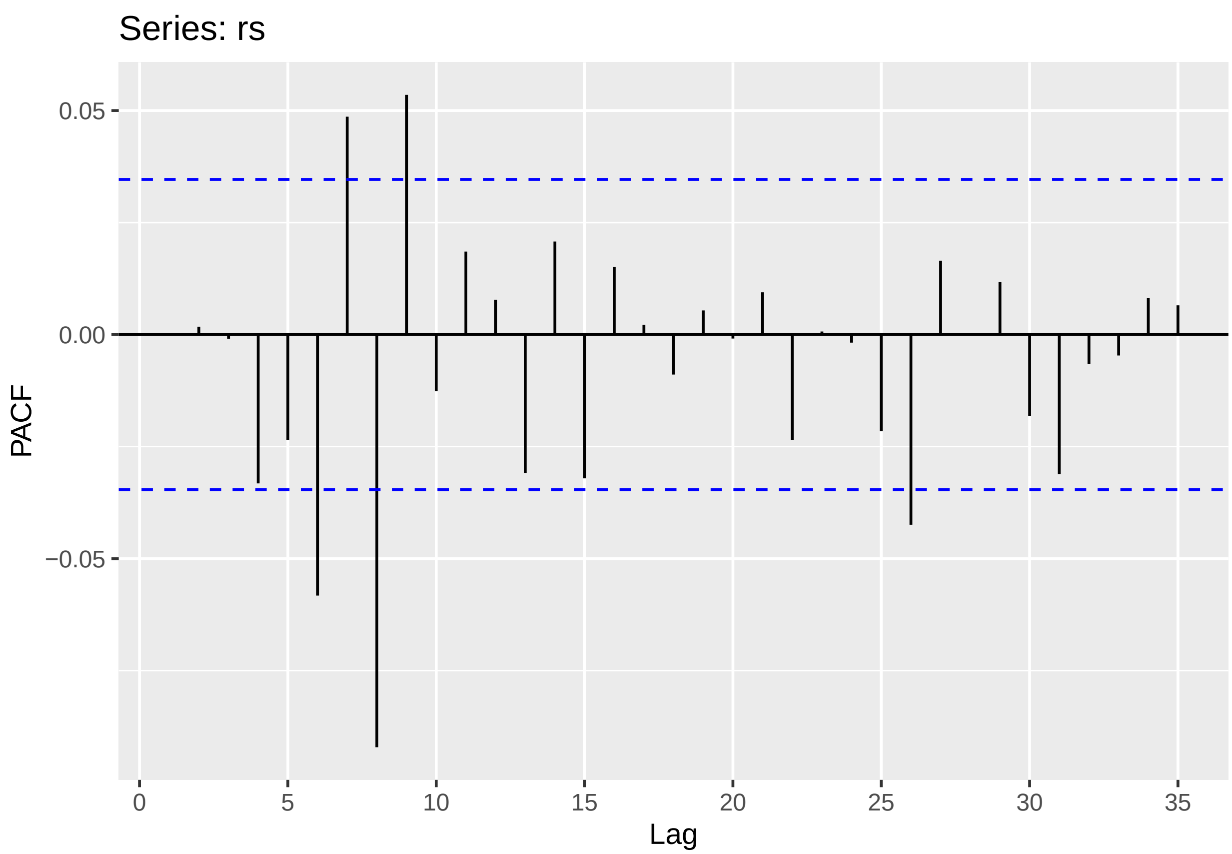
Let us perform model diagnostics.

Check residuals:



From the plots above, we observe a fairly plot showing constant variance and mean 0 (we will test to confirm below). The ACF plot looks fairly stationary, while the distribution looks to be slightly left skewed with very tall Kurtosis around mean 0.

Stationarity: PACF Plot



The PACF plot further supports stationarity for the model.

Mean Zero: T-Test

```
##
##  One Sample t-test
##
## data:  data
## t = -0.0094151, df = 3206, p-value = 0.9925
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0005573617  0.0005520345
## sample estimates:
##     mean of x
## -2.663597e-06
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains zero, therefore the mean is statistically 0 and the linear trend is removed.

Normalcy: Skewness

```
##       skew     lwr.ci     upr.ci
```

```
## -0.2859103 -0.3310652 -0.2901452
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF

## [1] FALSE
```

The model has exhibits left skewness, showing non-normalcy in respect to a Gaussian PDF.

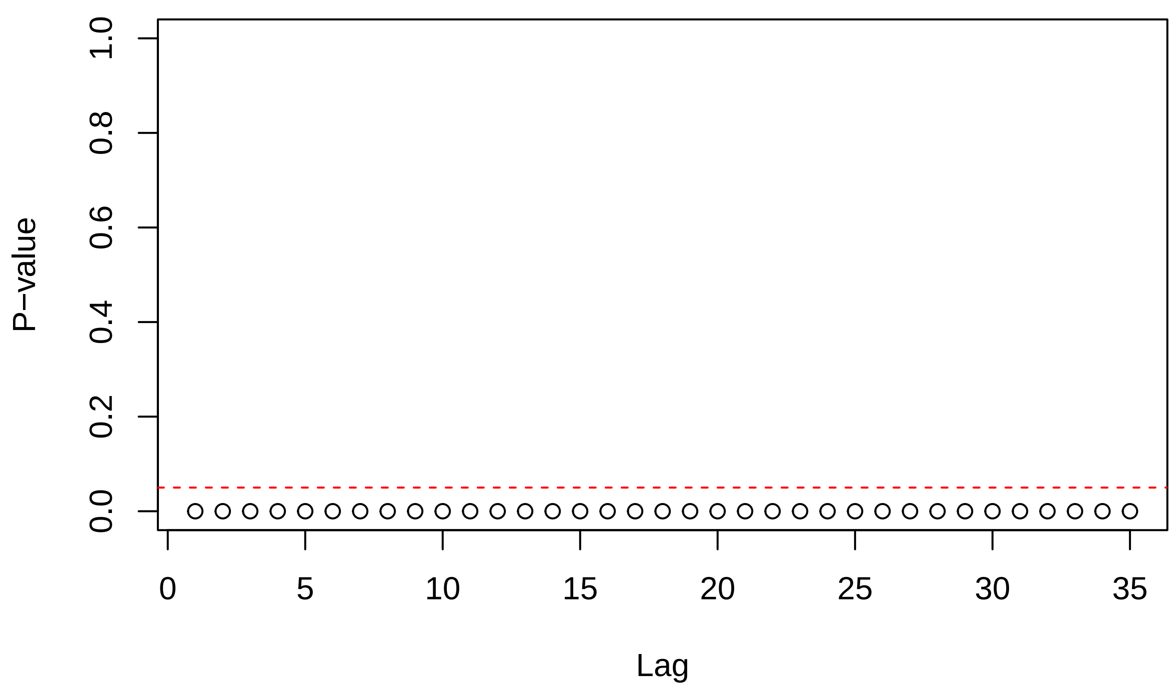
Normalcy: (excess) Kurtosis

```
##     kurt   lwr.ci   upr.ci
## 7.548113 7.631222 7.856361
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF

## [1] FALSE
```

The model has exhibits tall (excess) Kurtosis, showing non-normalcy in respect to a Gaussian PDF.

Constant Variance: McLeod-Li Test




```
## McLeod-Li: *NON*-constant variance, heteroscedastic -> reject H0
## McLeod-Li: Lags < 0.05:
## 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35

## [1] FALSE
```

The model exhibits non-constant variance with many lags with a p-value of under 0.05.

Lag independence: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 91.657, df = 30, p-value = 3.675e-08
##
## Box-Ljung: implies dependency present over 30 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

Box-Ljung test shows model has lag dependence and therefore serial autocorrelation.

Stationarity: ADF test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -11.1468
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:03 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test shows the model is stationary.

Stationarity: KPSS

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.1015
##
```

```
## Critical value for a significance level of:
##                  10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test shows the model is stationary.

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 551.04, df = 2, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 2 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung ARCH(m) test results in a p-value $<$ 0,05 for the model, implying non-constant variance, possible variance clustering, which results with ARCH effects present.

1.3. Fit a Gaussian ARMA-GARCH volatility model to the log return series. Obtain the normal QQ-plot of the standardized residuals, and write the model to be fitted. Is the model adequate? Persistent? Why?

We create the following mean model based on the EDA above using AR(1) and MA(1):

```
m <- garchFit(~arma(1,1)+garch(1,1),data=msft,trace=F)  # arma from EDA
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(1, 1) + garch(1, 1), data = msft, trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(1, 1) + garch(1, 1)
## <environment: 0x0000018ba7ff2960>
##  [data = msft]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##          mu          ar1          ma1        omega       alpha1        beta1
##  2.4540e-04   7.6995e-01  -8.0791e-01   2.4041e-05   1.4679e-01   7.6165e-01
##
```

```
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##          Estimate  Std. Error  t value Pr(>|t|)
## mu      2.454e-04   1.272e-04    1.930   0.0537 .
## ar1     7.699e-01   1.123e-01    6.854 7.17e-12 ***
## ma1    -8.079e-01   1.031e-01   -7.834 4.66e-15 ***
## omega   2.404e-05   4.200e-06    5.724 1.04e-08 ***
## alpha1  1.468e-01   2.122e-02    6.917 4.62e-12 ***
## beta1   7.616e-01   3.162e-02   24.086  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  8977.943    normalized:  2.799483
##
## Description:
##  Tue Mar 28 19:56:03 2023 by user: Reed
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  5009.72   0
##  Shapiro-Wilk Test  R    W      0.9536068 0
##  Ljung-Box Test     R    Q(10)  12.67103  0.242651
##  Ljung-Box Test     R    Q(15)  20.80202  0.1432751
##  Ljung-Box Test     R    Q(20)  22.23971  0.3276465
##  Ljung-Box Test     R^2  Q(10)  5.138948  0.8817048
##  Ljung-Box Test     R^2  Q(15)  6.534645  0.9693019
##  Ljung-Box Test     R^2  Q(20)  7.223212  0.9958862
##  LM Arch Test       R    TR^2   5.463446  0.9406846
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -5.595225 -5.583863 -5.595232 -5.591152
```

Based on the summary above, we create the following equation:

$$r_t = \omega + \phi_1 y_{t-1} + \theta_1 z_{t-1} + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1} + \mu$$

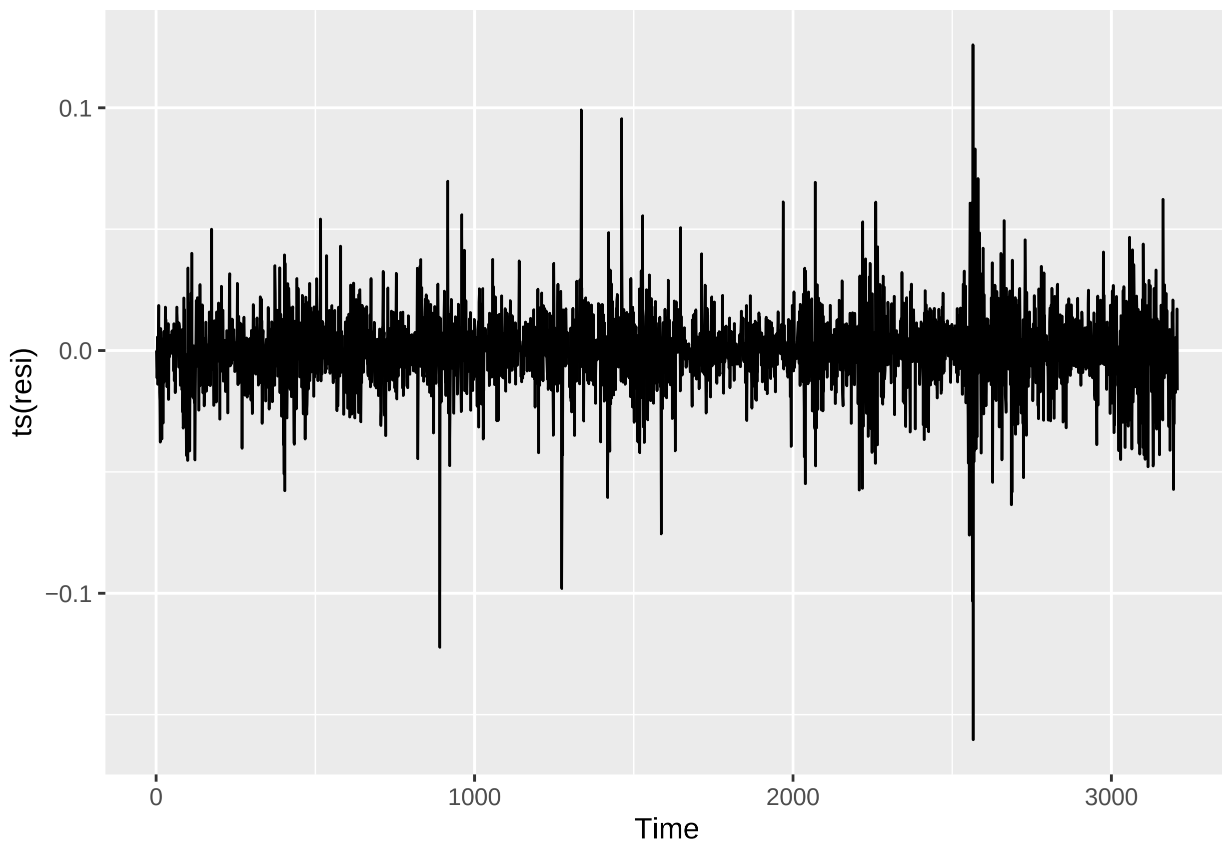
Adding the α and β coefficients together:

```r
sum(coef(m)[(length(coef(m))-1):(length(coef(m)))]) # 0.9084379 - slow to revert back to mean
```

```
## [1] 0.9084374
```

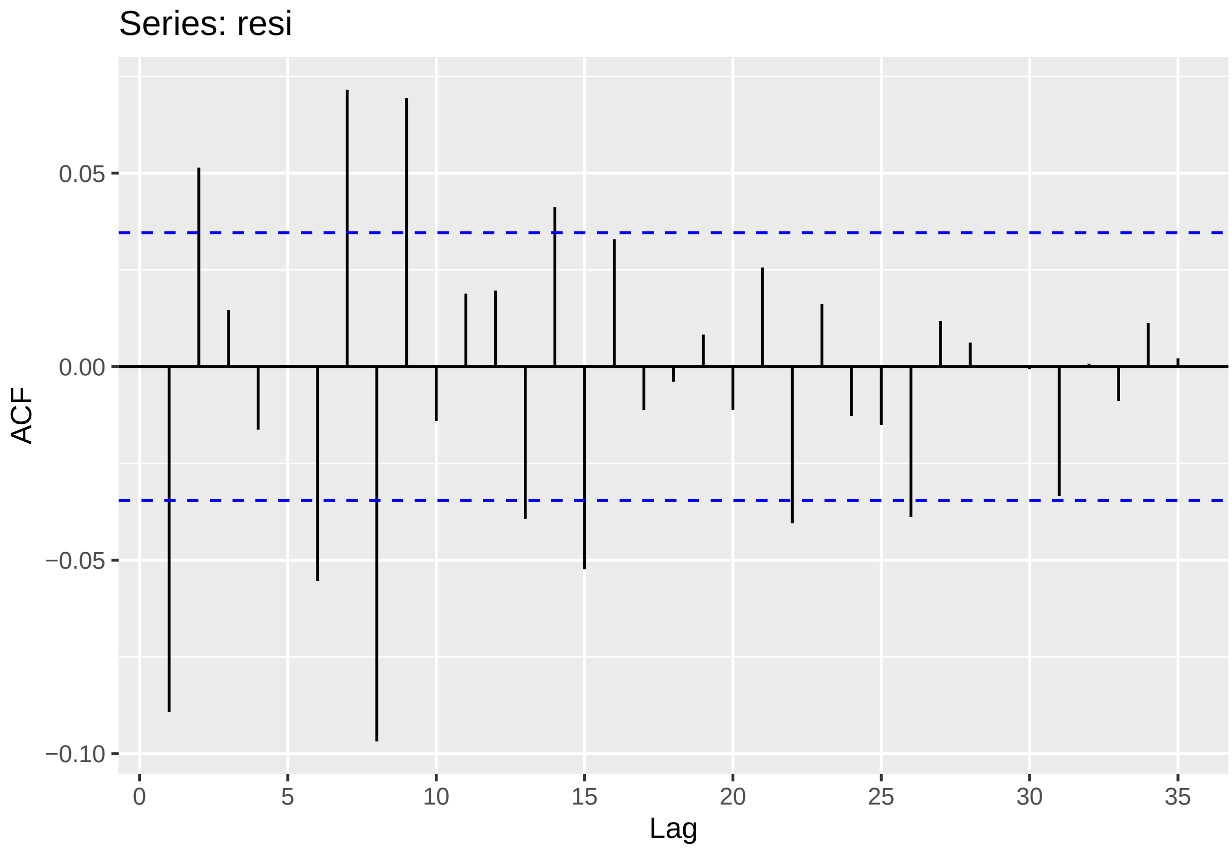
The combined value is 0.9084, which is near 1, which makes the model's volatility slow to revert back to mean.

Let us perform model diagnostics.



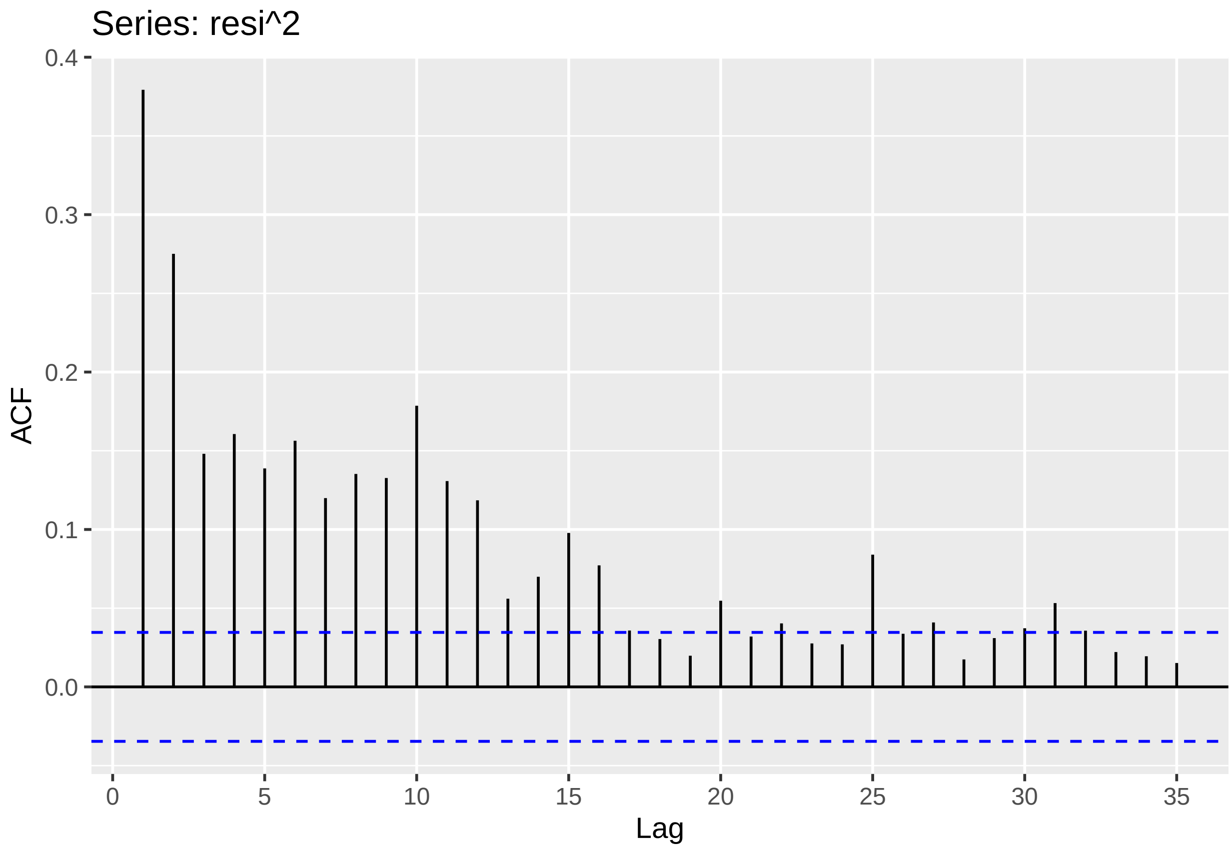
The plot may exhibit mean 0 and possible constant variance.

Stationarity: ACF plot for residuals



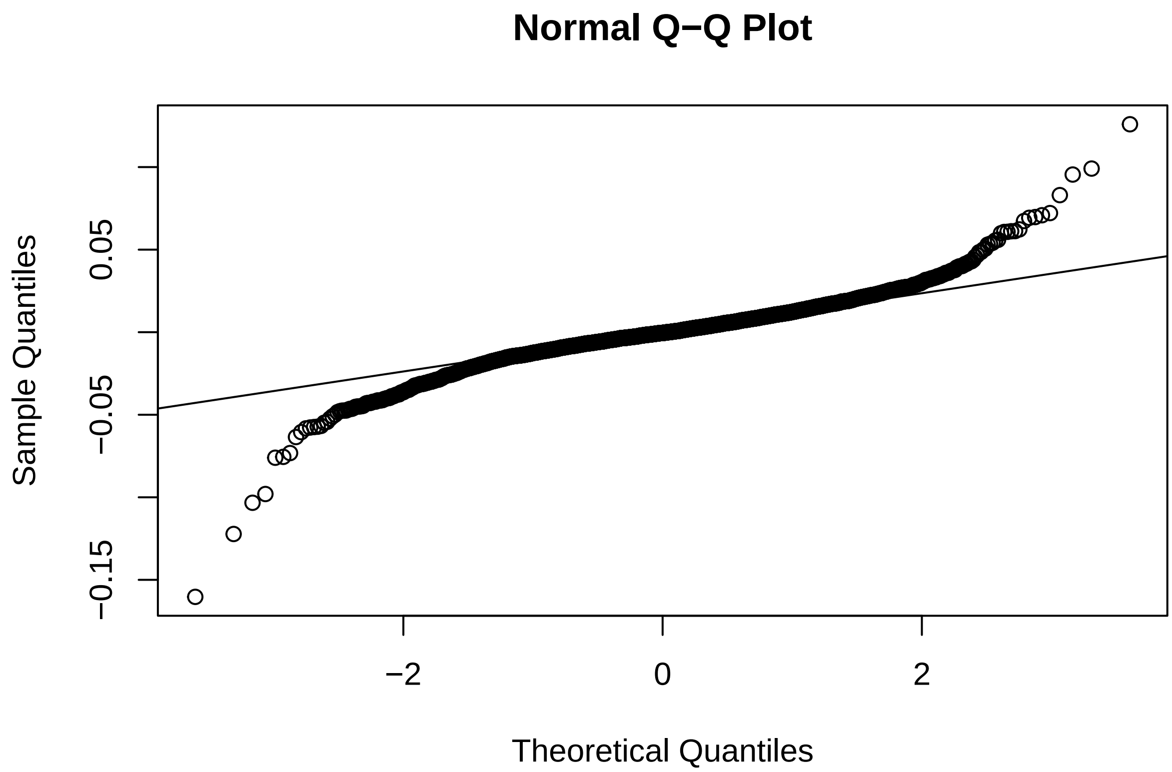
The ACF plot looks fairly stationary.

Stationarity: ACF plot for squared residuals



The ACF plot above for the squared residuals display serial autocorrelation in squared residuals up to lag 16.

Q-Q Plot:



The thick tails in the Q-Q plot my indicate a not-so-well goodness-of-fit.

Lag independence: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 108.25, df = 10, p-value < 2.2e-16
##
## Box-Ljung: implies dependency present over 10 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung test shows the model contains lag dependency and thus serial autocorrelation.

ARCH effects: Box-Ljung ARCH test

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 1263.2, df = 10, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 10 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung ARCH(m) test results in a p-value $< 0{,}05$ for the model, implying non-constant variance, possible variance clustering, which results with ARCH effects present.

Mean Zero: T-Test

```
##
##  One Sample t-test
##
## data:  data
## t = -1.4606, df = 3206, p-value = 0.1442
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0009718938  0.0001420514
## sample estimates:
##     mean of x
## -0.0004149212
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains zero, therefore the mean is statistically 0 and the linear trend is removed.

Normalcy: Skewness

```
##       skew     lwr.ci     upr.ci
## -0.3328748 -0.3661577 -0.3196728
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model has exhibits left skewness, showing non-normalcy in respect to a Gaussian PDF.

Normalcy: (excess) Kurtosis

```
##     kurt   lwr.ci   upr.ci
## 8.680675 8.734588 9.052796
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model has exhibits tall (excess) Kurtosis, showing non-normalcy in respect to a Gaussian PDF, as well as a not-so-well goodness-of-fit.

Stationarity: ADF test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -10.7705
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:05 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test shows the model is stationary.

Stationarity: KPSS

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.1153
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test shows the model is stationary.

1.4. Build an ARMA-GARCH model with Student-t innovations ($cond.dist = "std"$) for the log return series. Perform model checking and write the model to be fitted.

We create the following mean model based on the EDA above using AR(1) and MA(1):

```r
m <- garchFit(~arma(1,1)+garch(1,1),data=msft,trace=F,cond.dist="std")
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(1, 1) + garch(1, 1), data = msft, cond.dist = "std",
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(1, 1) + garch(1, 1)
## <environment: 0x0000018ba9853258>
##  [data = msft]
##
## Conditional Distribution:
##  std
##
## Coefficient(s):
##          mu          ar1          ma1        omega       alpha1        beta1
##  2.0951e-04   8.0916e-01  -8.4953e-01   8.8933e-06   1.1353e-01   8.5878e-01
##       shape
##  4.5915e+00
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##          Estimate  Std. Error  t value Pr(>|t|)
## mu      2.095e-04   8.707e-05    2.406 0.016123 *
## ar1     8.092e-01   7.485e-02   10.811  < 2e-16 ***
## ma1    -8.495e-01   6.668e-02  -12.741  < 2e-16 ***
## omega   8.893e-06   2.502e-06    3.555 0.000378 ***
## alpha1  1.135e-01   1.925e-02    5.899 3.65e-09 ***
## beta1   8.588e-01   2.347e-02   36.590  < 2e-16 ***
## shape   4.592e+00   3.727e-01   12.319  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  9179.268    normalized:  2.86226
##
## Description:
##  Tue Mar 28 19:56:05 2023 by user: Reed
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  6532.776  0
##  Shapiro-Wilk Test  R    W      0.9473975 0
##  Ljung-Box Test     R    Q(10)  11.19926  0.342206
##  Ljung-Box Test     R    Q(15)  19.69824  0.1838166
##  Ljung-Box Test     R    Q(20)  21.54496  0.3657037
##  Ljung-Box Test     R^2  Q(10)  6.873134  0.7373651
##  Ljung-Box Test     R^2  Q(15)  9.3266    0.8598308
##  Ljung-Box Test     R^2  Q(20)  11.78147  0.9233923
```

```
##  LM Arch Test       R    TR^2   7.567096  0.8179801
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -5.720154 -5.706899 -5.720164 -5.715403
```

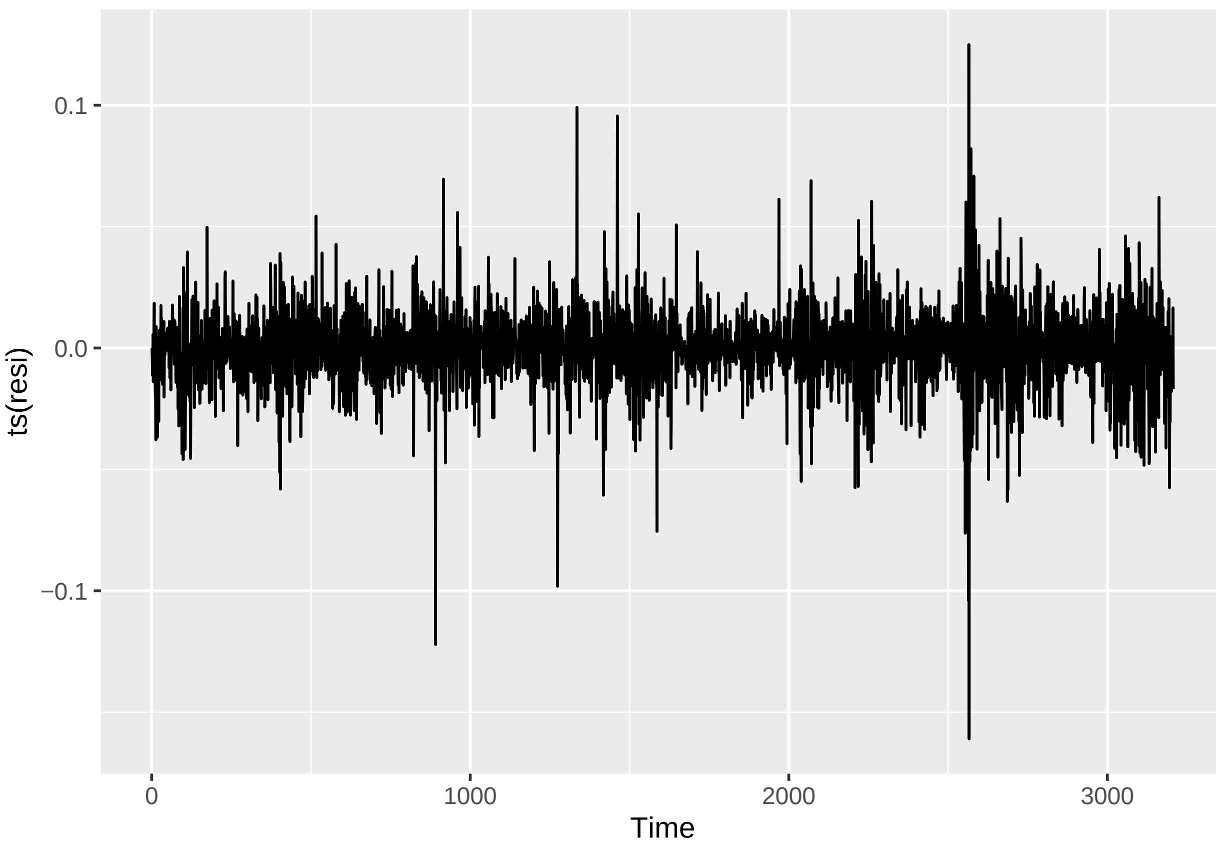
Adding the α and β coefficients together:

```
sum(coef(m)[(length(coef(m))-1):(length(coef(m))-2)]) # 0.9084379 - slow to revert back to mean
```

```
## [1] 0.9723147
```

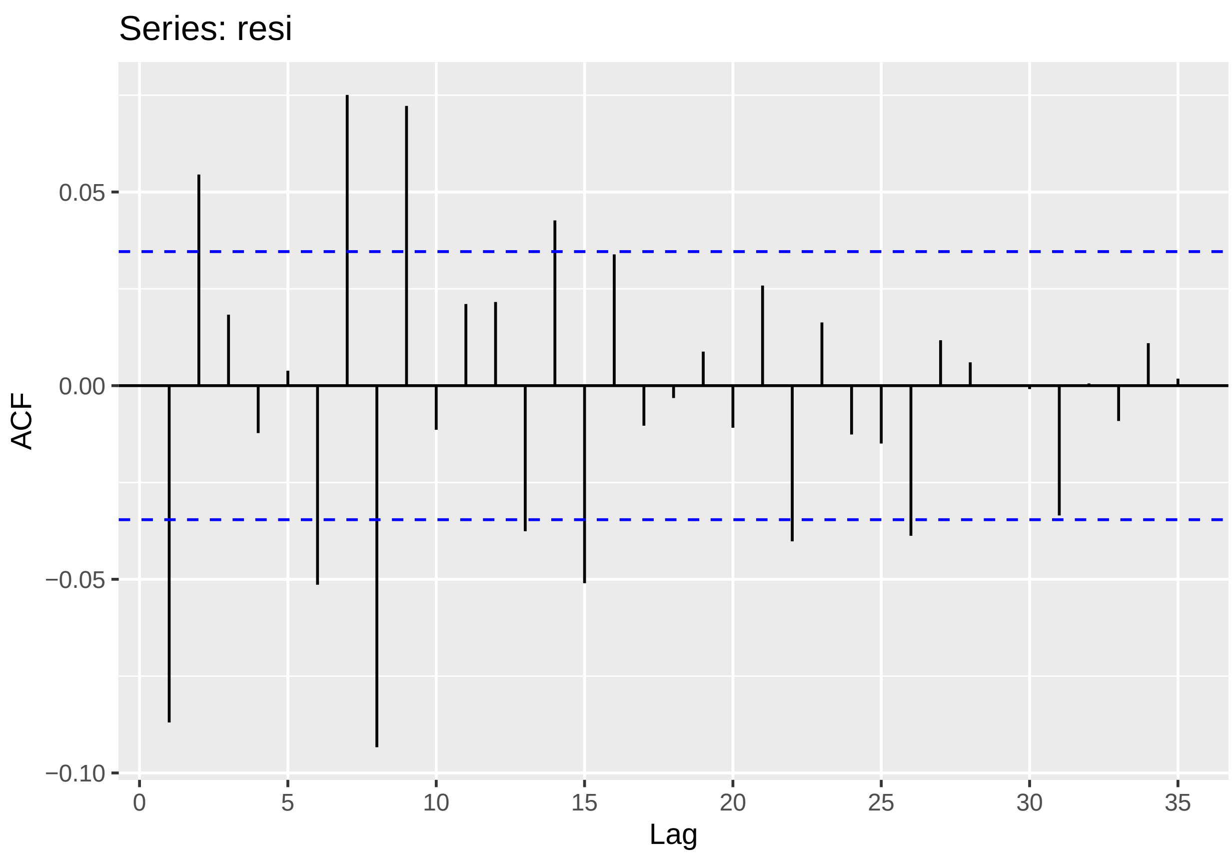
The combined value is 0.9723, which is even closer to 1, which makes the model's volatility even slower to revert back to the mean.

Let us perform model diagnostics.



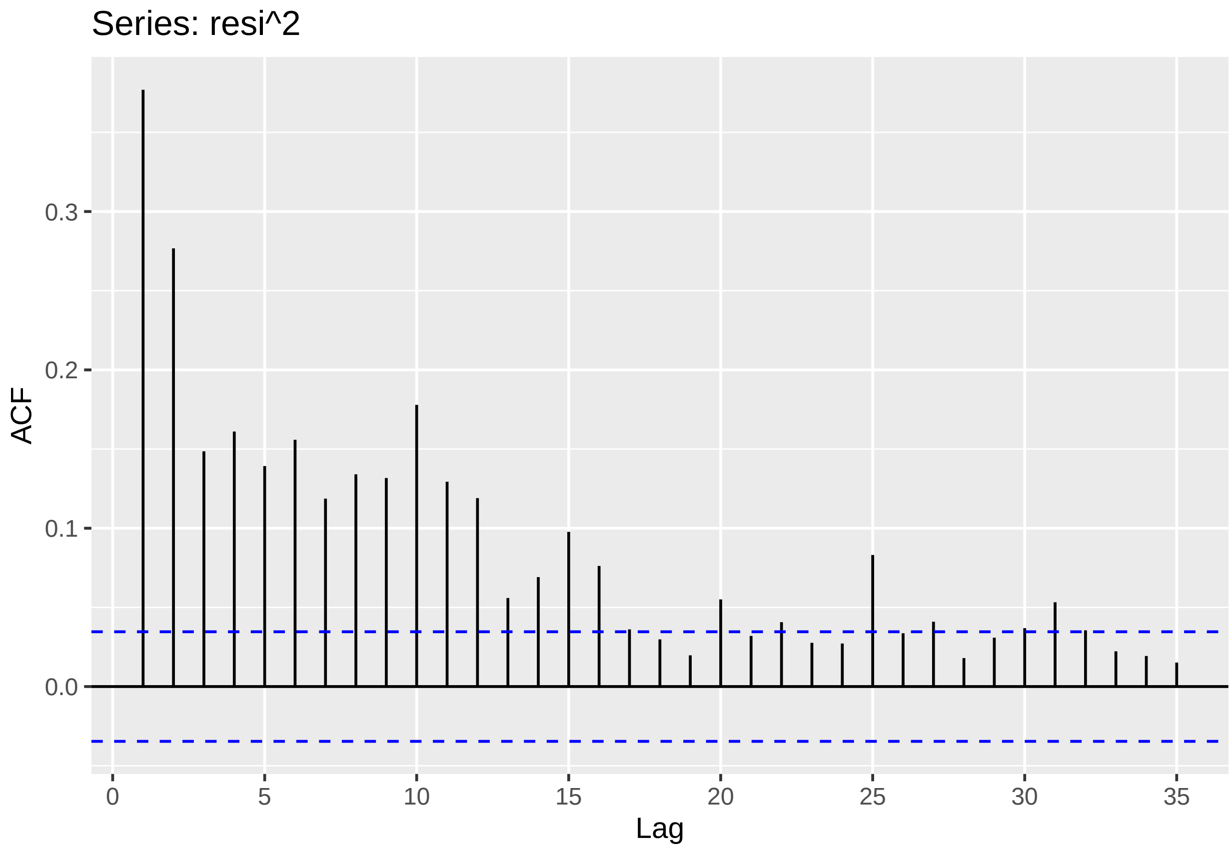
The plot may exhibit mean 0 and possible non-constant variance.

Stationarity: ACF plot for residuals



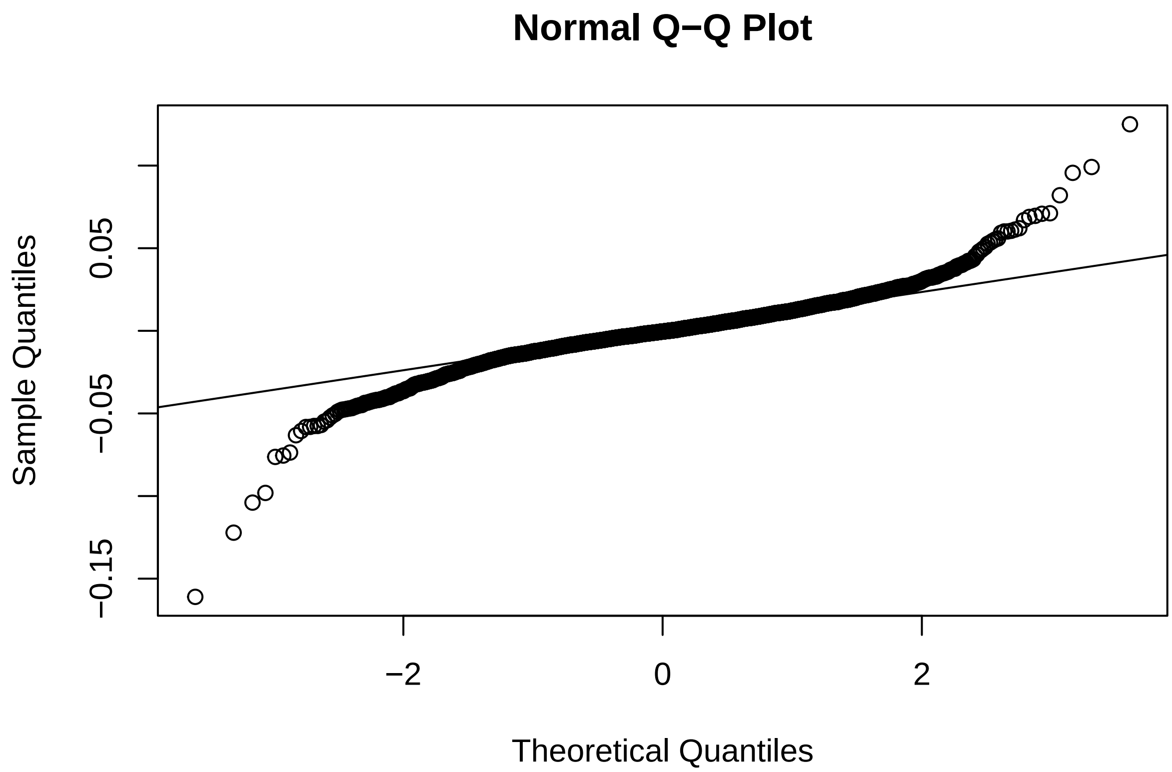
The ACF plot looks fairly stationary.

Stationarity: ACF plot for squared residuals



The ACF plot above for the squared residuals display serial autocorrelation in squared residuals up to lag 16.

Q-Q Plot:



The thick tails in the Q-Q plot my indicate a not-so-well goodness-of-fit.

Lag independence: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 107.3, df = 10, p-value < 2.2e-16
##
## Box-Ljung: implies dependency present over 10 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung test shows the model contains lag dependency and thus serial autocorrelation.

ARCH effects: Box-Ljung ARCH test

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 1257.4, df = 10, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 10 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung ARCH(m) test results in a p-value < 0,05 for the model, implying non-constant variance, possible variance clustering, which results with ARCH effects present.

Mean Zero: T-Test

```
##
##  One Sample t-test
##
## data:  data
## t = -1.6802, df = 3206, p-value = 0.09302
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -1.034179e-03  7.967973e-05
## sample estimates:
##     mean of x
## -0.0004772497
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains zero, therefore the mean is statistically 0 and the linear trend is removed.

Normalcy: Skewness

```
##       skew     lwr.ci     upr.ci
## -0.3541866 -0.3862334 -0.3407519
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model has exhibits left skewness, showing non-normalcy in respect to a Gaussian PDF.

Normalcy: (excess) Kurtosis

```
##     kurt   lwr.ci   upr.ci
## 8.698179 8.717137 9.039336
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model has exhibits tall (excess) Kurtosis, showing non-normalcy in respect to a Gaussian PDF, as well as a not-so-well goodness-of-fit.

Stationarity: ADF test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -10.5794
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:06 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test shows the model is stationary.

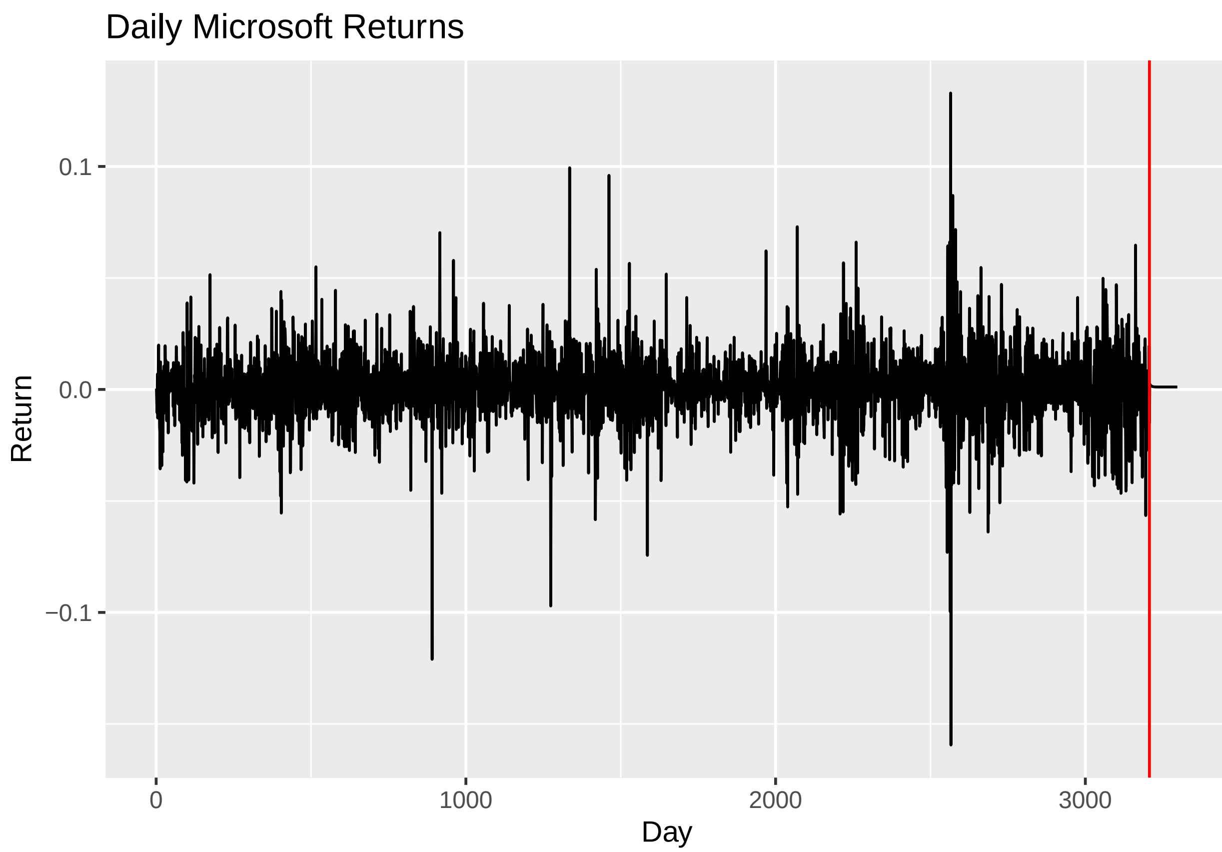
Stationarity: KPSS

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.1237
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test shows the model is stationary.

1.5. Obtain 1-step to 90-step ahead mean and volatility forecasts using the fitted ARMA-GARCH model with Student-t innovations.



The predicted values for redisuals of the differenced log(msft) data doesn't match the volatility shown in the origial residuals plot. The predicted values lie just above the statistically zero mean.

1.6. As the estimated coefficient of the mean equation is small, we may ignore the mean equation; i.e., use the mean equation $r_t = a_t$. Fit an IGARCH(1) model to the log returns. Write the model to be ftted.

```
m <- Igarch(msft)
```

```
## Estimates:  0.9633717
## Maximized log-likehood:  -8892.529
##
## Coefficient(s):
##        Estimate  Std. Error  t value   Pr(>|t|)
## beta 0.96337169  0.00321591  299.565 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The beta component's coefficient of the IGARCH(1) model is 0.9634, near 1, which indicates the model's volatility might be slower to reach the mean.

```
##              Length Class  Mode
## par            1    -none- numeric
## volatility 3207   ts     numeric
```

Based on the summary above, the equation is:

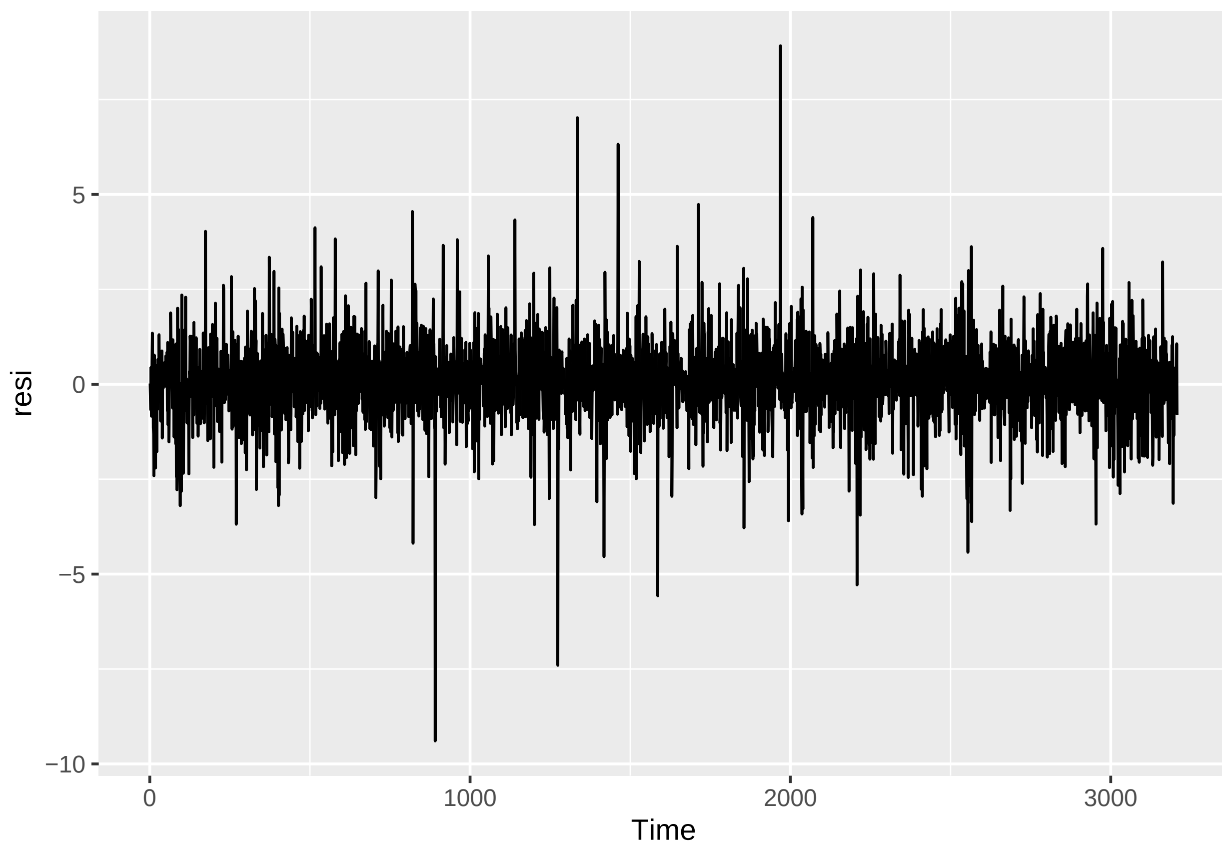
$$\sigma^2_{t|t-1} = \beta\sigma^2_{t-1|t-2} + (1-\beta)^2_{t-1}$$

1.7. Let σ_t be the fitted volatility of the IGARCH(1) model. Define the standardized residuals as $\epsilon_t = r_t/\sigma_t$, where rt is the daily log return. Is serial correlation in the standardized residuals present? Persistent? Why?

Calculate the standardized residuals:

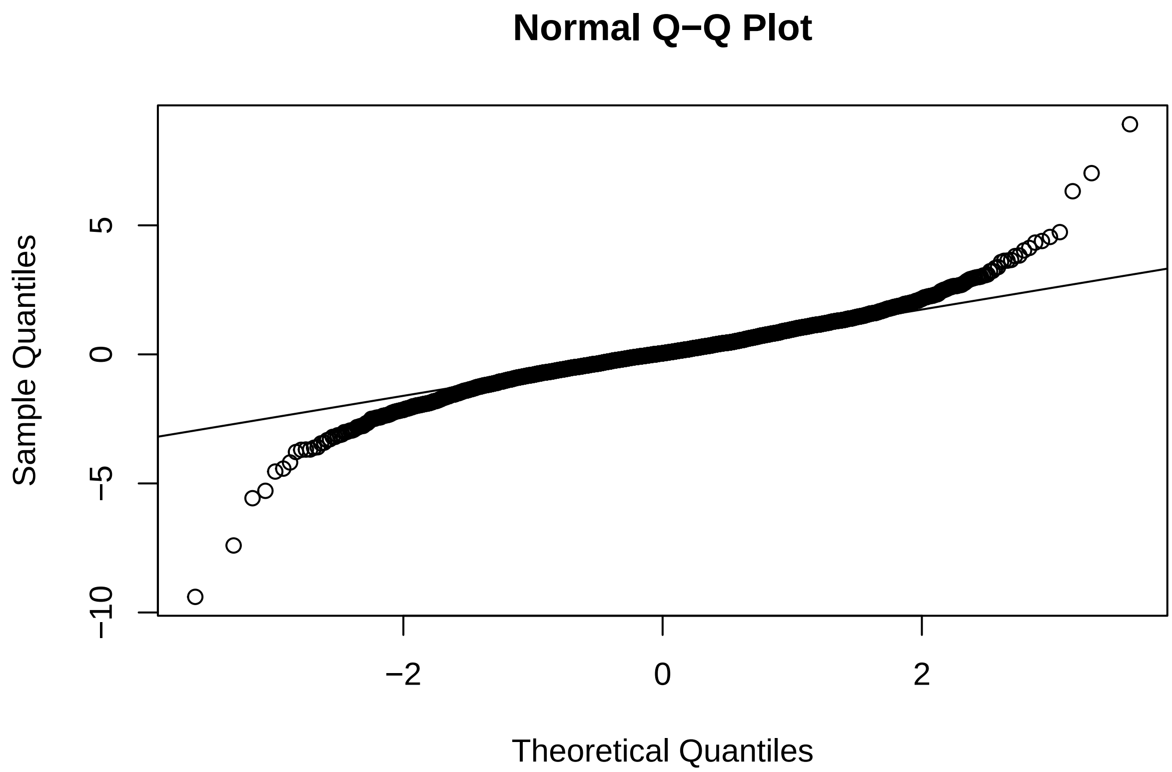
```
sigma.t <- m$volatility
resi <- msft/sigma.t
```

Plotting the σ_t residuals:

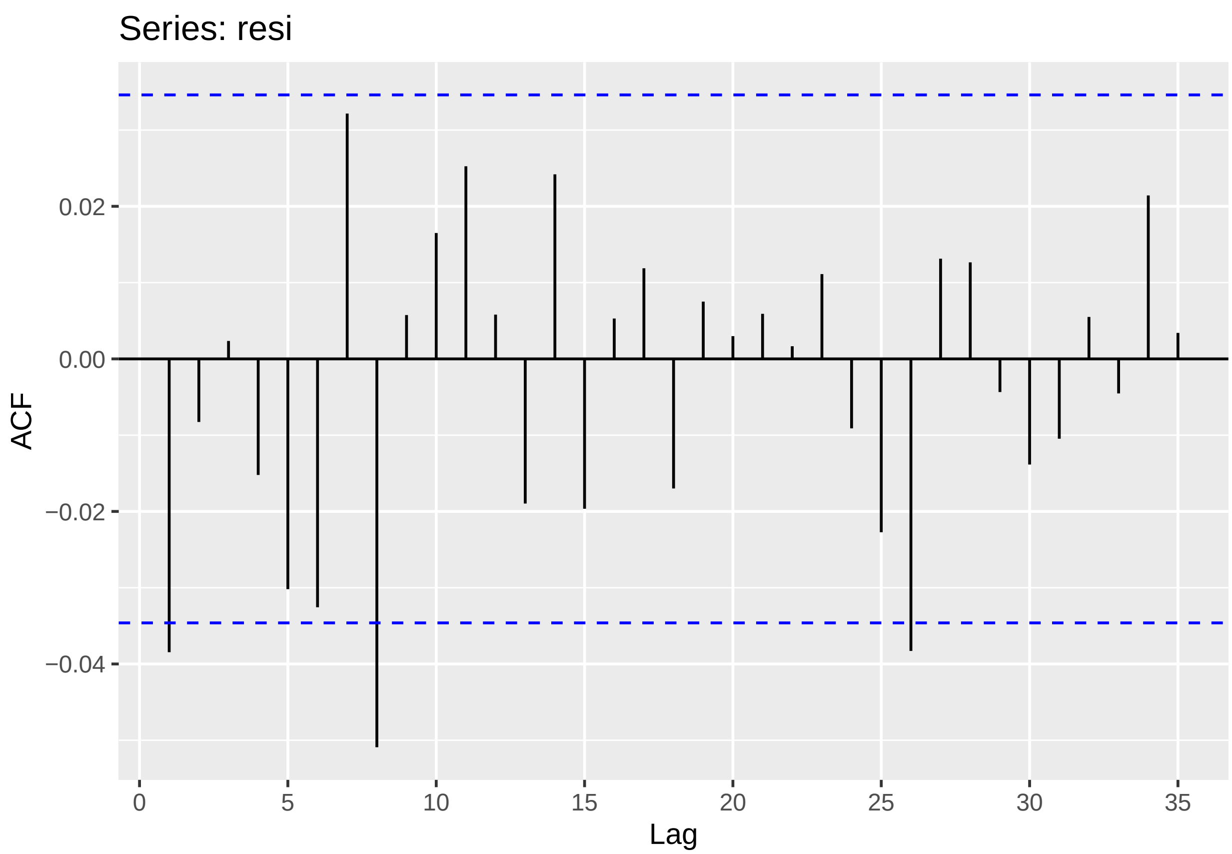


From the plot we see a zero mean, and due to the outliers, non-constant variance.

Q-Q Plot:



The thick tails of the IGARCH(1,1) indicate a not a not-too-great goodness-of-fit.



The ACF plot shows the IGARCH(1) model of the standardized residuals is fairly stationary.

Serial Autocorrelation: Box-Ljung test

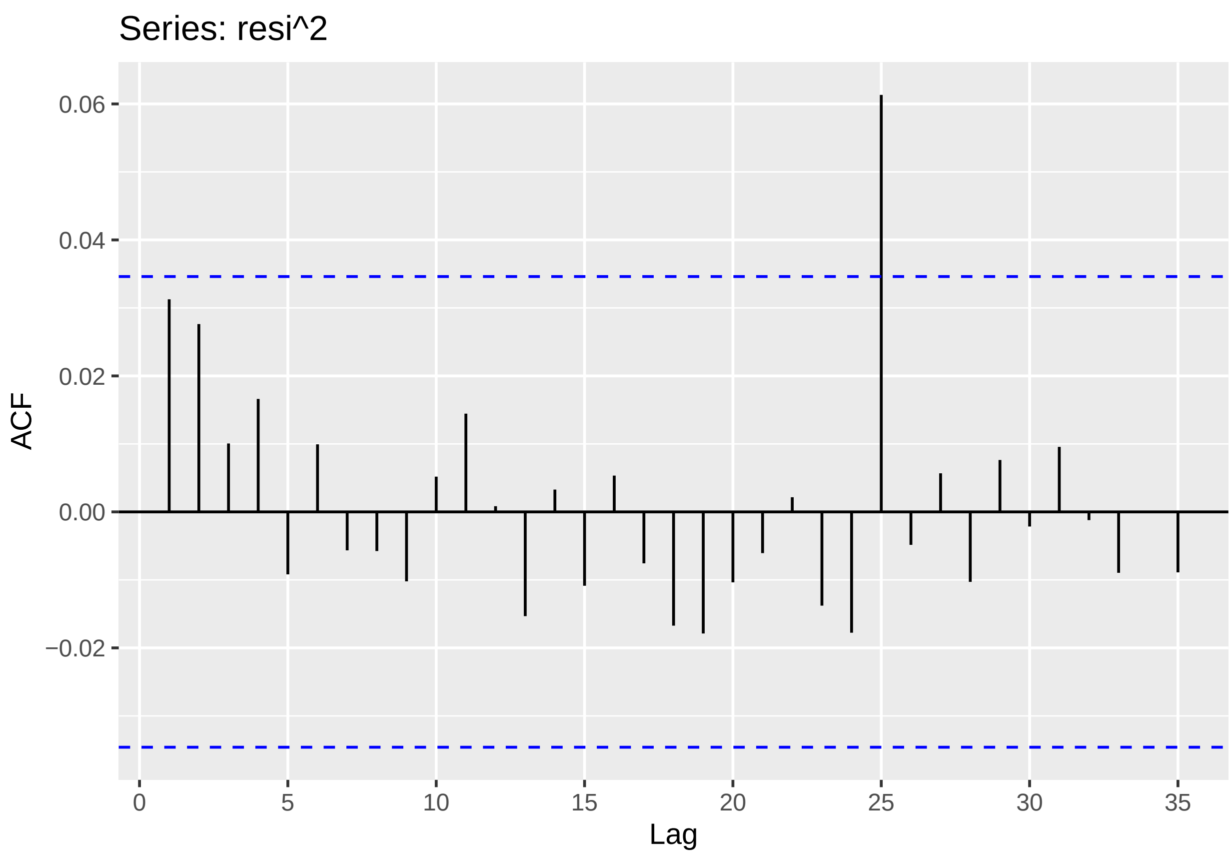
```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 24.715, df = 10, p-value = 0.005914
##
## Box-Ljung: implies dependency present over 10 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung test indicates lag dependency in the standardized residuals of the IGARCH(1) model, therefore serial correlation is also present.

1.8. Is serial correlation in the squares of the standardized residuals present? Why?

ACF Plot: squares of standardized residuals



The ACF plot of the squares of the standardized residuals show the model is stationary.

Box-Ljung ARCH Test:

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 8.0205, df = 10, p-value = 0.6268
##
## Box-Ljung ARCH(m): implies constant variance over 10 lags, *NO* ARCH effects present,
##  *NO* variance clustering, *NO* volatility, homoscedastic -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The Box-Ljung ARCH test indicates no volatility of the squares of the standardized residuals of the IGARCH(1) model, therefore serial correlation is not present. This is due to the model's high p-value > 0.05.

Mean test for 0:

```
##
##  One Sample t-test
##
## data:  data
## t = 2.9542, df = 3206, p-value = 0.003157
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01848458 0.09144312
## sample estimates:
##  mean of x
## 0.05496385
##
## T-Test: mean *NOT* statistically zero, linear trend present ->
## reject H0
```

```
## [1] FALSE
```

While the 95% CI does not contain zero, the upper and lower bounds are values that are small and close to zero.

Nirmacly: Skewness

```
##        skew      lwr.ci      upr.ci
## -0.09963946 -0.11243673 -0.07147882
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The IGARCH(1) model contains left skewness, thus not normal.

Normalcy: Kurtosis

```
##     kurt   lwr.ci   upr.ci
## 6.720370 6.755596 7.007163
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The IGARCH(1) model has tall Kurtosis, thus not normal. But this calculated Kurtosis seems to be smaller than the other models tested so far, maybe implying a better goodness-of-fit.

Stationarity: ADF Test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -10.0875
##   P VALUE:
##     0.01
##
## Description:
```

```
##  Tue Mar 28 19:56:08 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The IGARCH(1) model is stationary based on the ADF test.

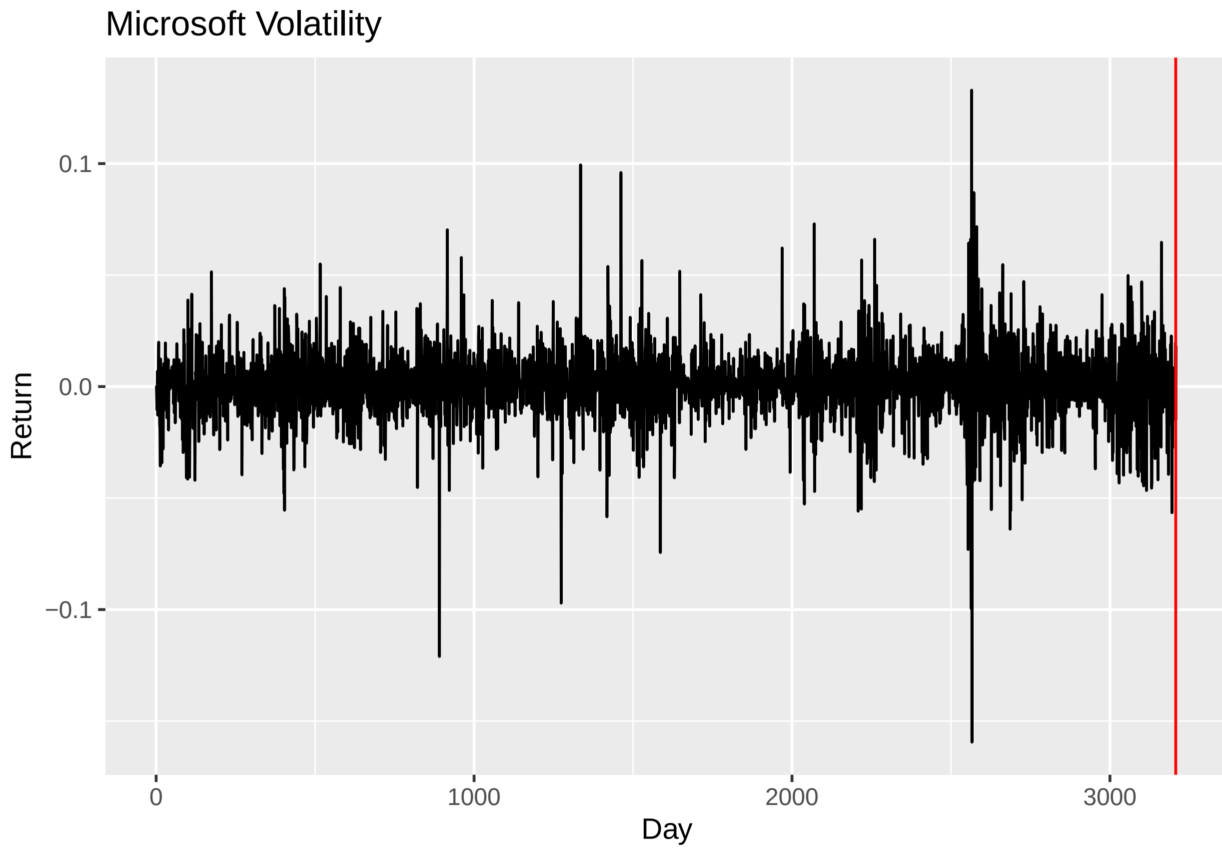
Stationarity: KPSS Test

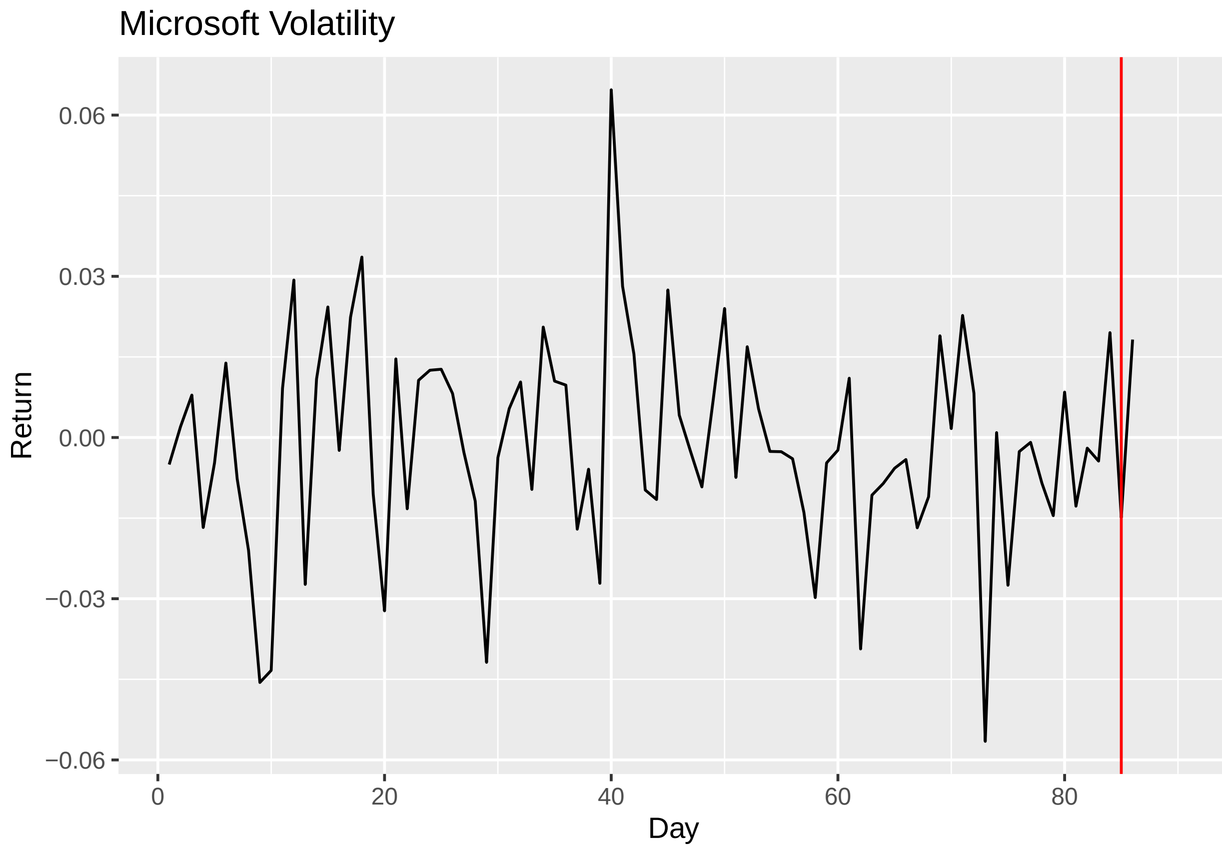
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.1272
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The IGARCH(1) model is stationary based on the ADF test.

1.9. Based on the model checking, is the IGARCH model adequate? If yes, obtain 1-step to 5-step ahead volatility forecasts for the log return series (forecast origin is the last data point).





The IGARCH(1) seems to be an adequate model, due to its ability to show some short-term activity similar to the pre-existing trends, unlike what the other models were able to do. Based on its lower kurtosis it could be considered to have a better goodness-of-fit.

1.10. Which model do you recommend and why?

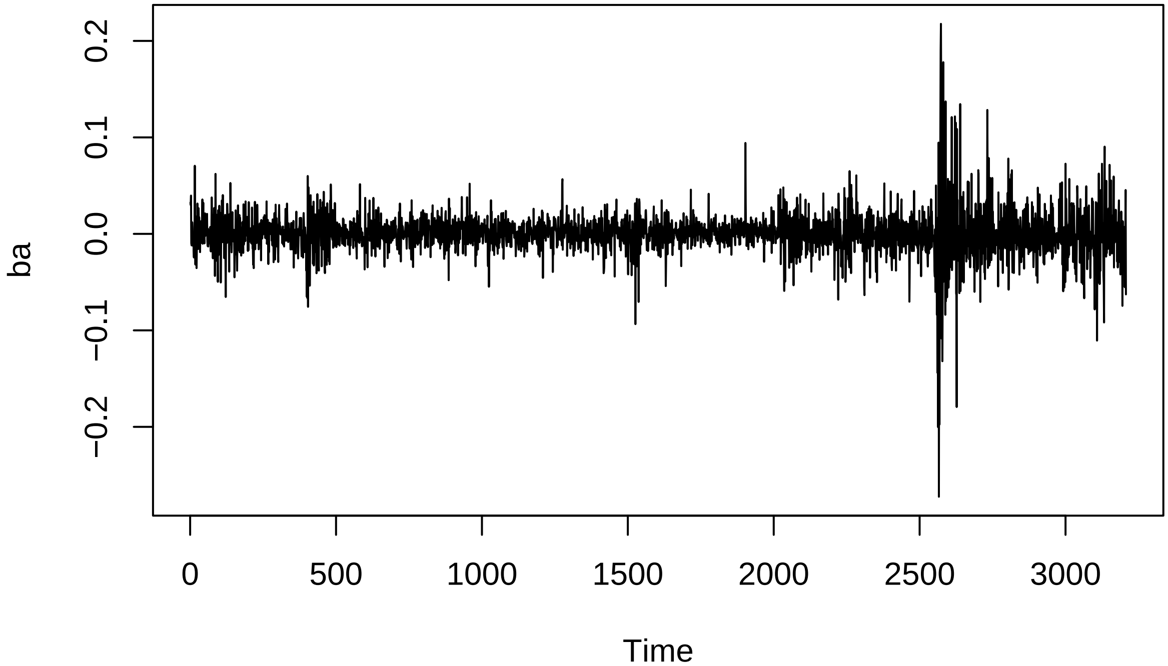
The high alpha + beta values near 1 of the ARMA/GARMA models suggest that we should try to fit the data to an IGARCH(1) model. When testing the IGARCH(1) model we find it to have a better goodness of fit due to its lower Kurtosis, and its forecast plot seems to mimic some of the previous MSFT stock activity after volatility has been considered.

2. Boeing returns (30 points)

Consider the monthly returns of Boeing (*ba*) stock.

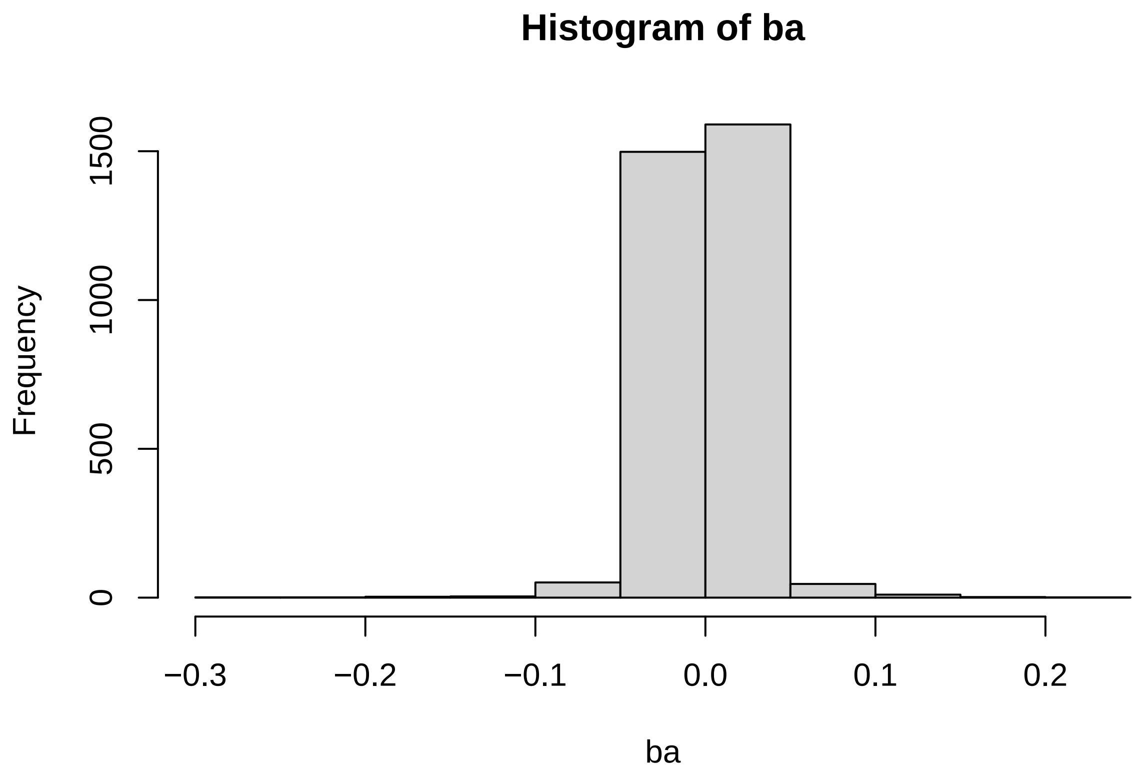
2.1. Use EDA on the BA monthly closing prices. Is the expected ba log return zero? Why? Is there serial correlation in the log returns? Why? Is there any ARCH effect in the log returns? Why?

Plot of log(ba):



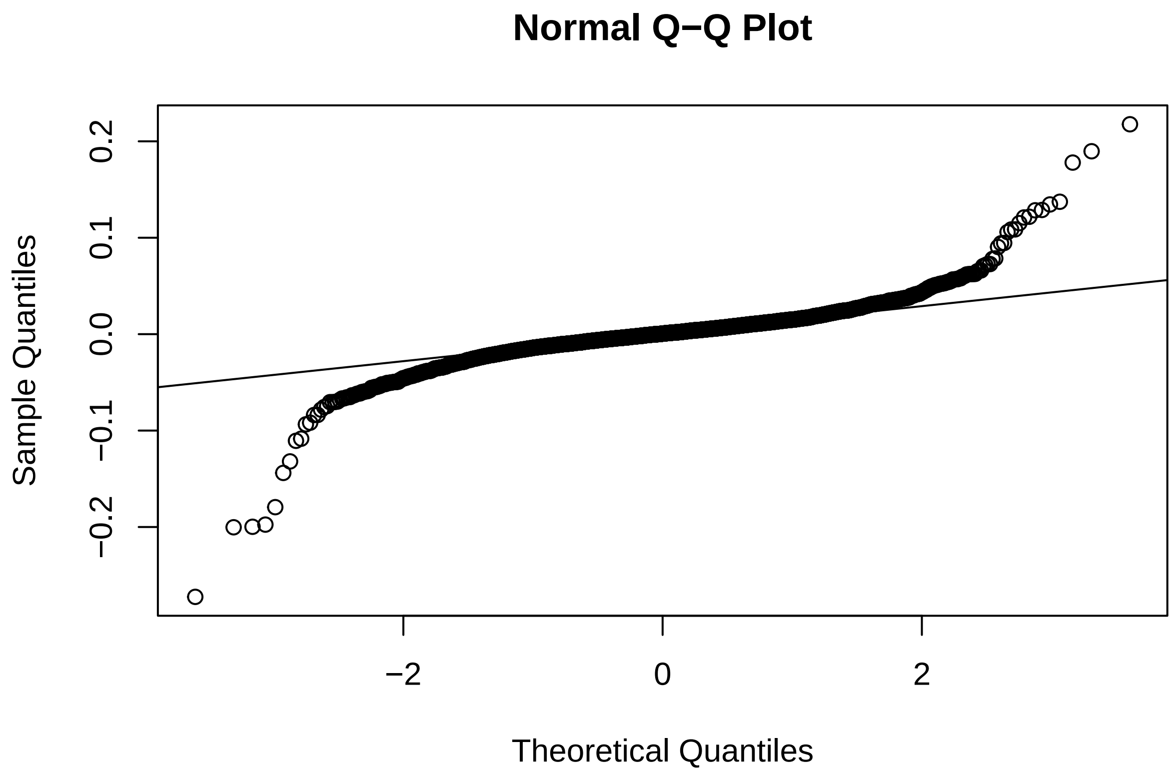
The plot of log(ba) looks to have mean 0 but non-constant variance.

Histogram: log(ba)



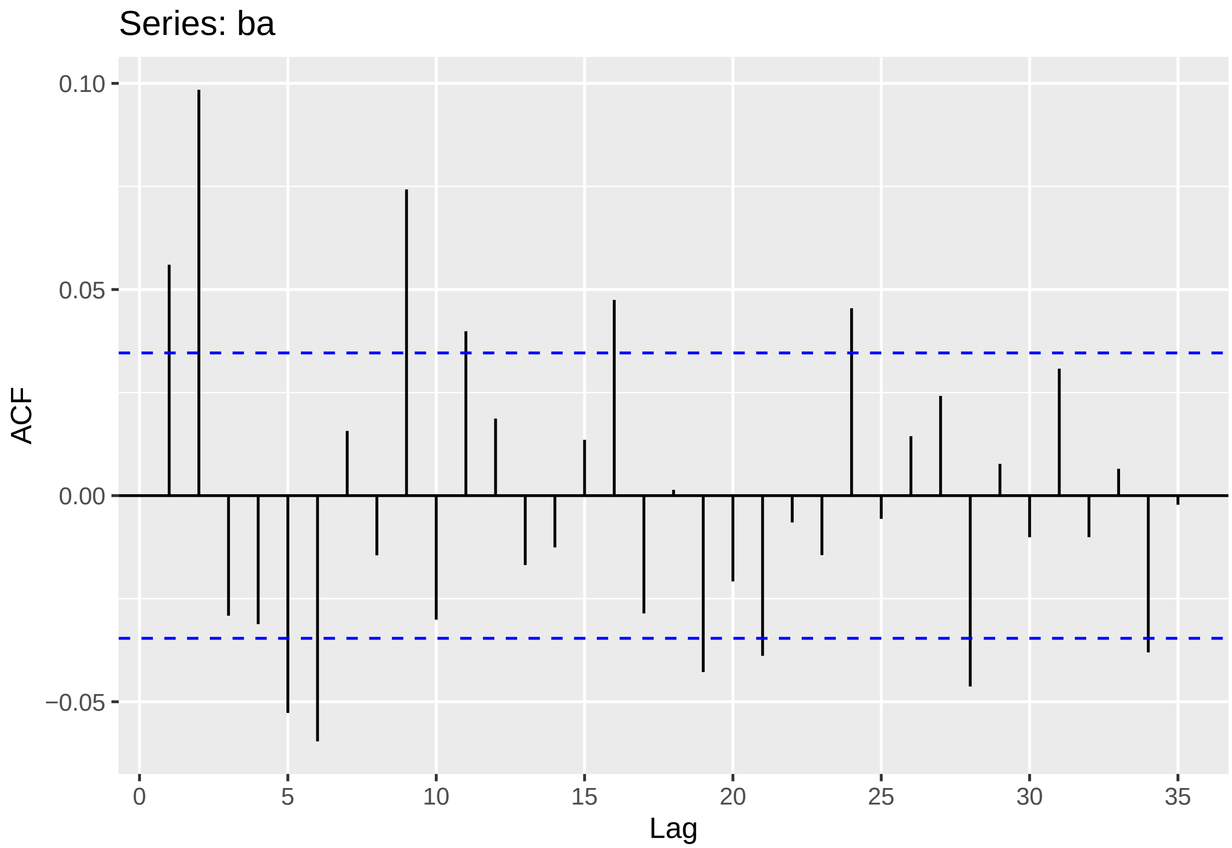
The ditribution looks tall and slightly left skewed, thus not normal.

Q-Q plot: log(ba)



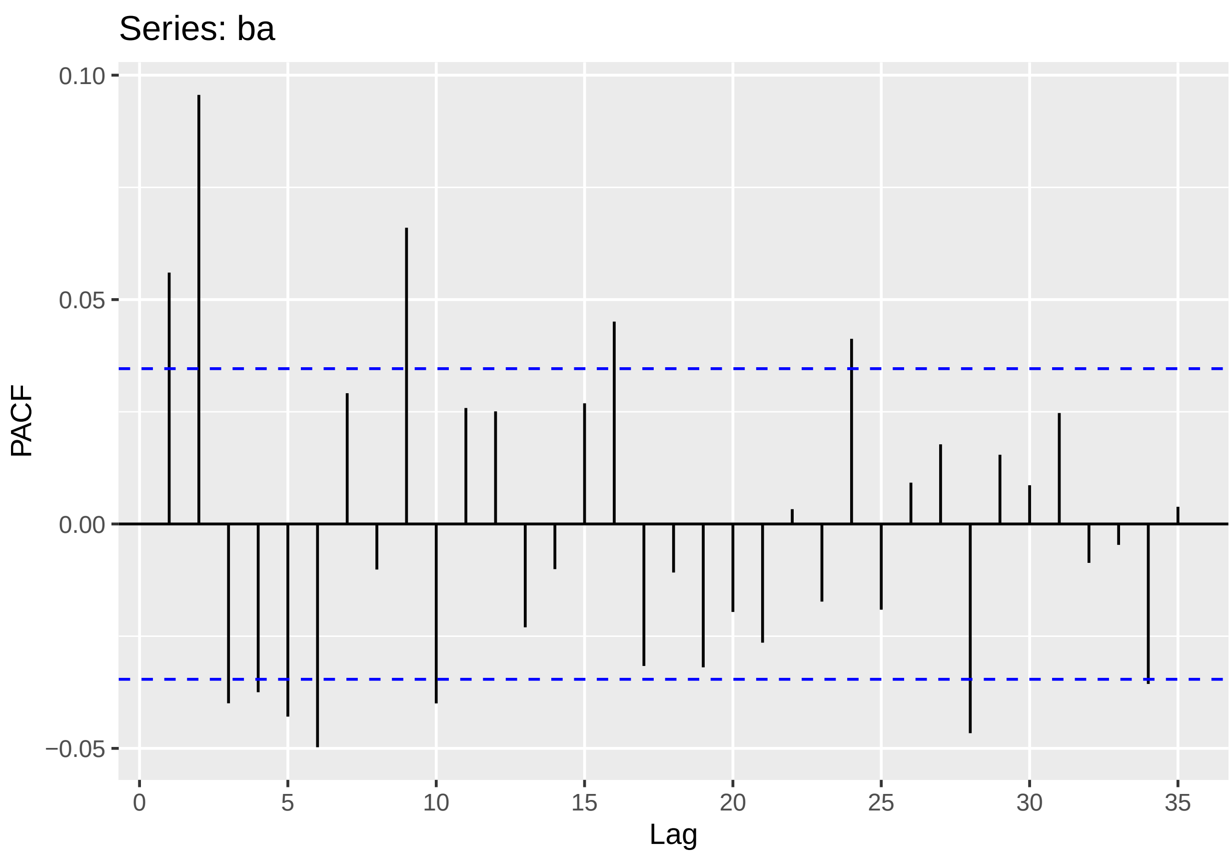
The thick tails indicate tall Kurtosis, therefore not normal.

ACF Plot: log(ba)



The ACF plot suggests to use a MA(2) component in an ARMA model.

ACF Plot: log(ba)



The PACF plot suggests to use a AR(2) component in an ARMA model.

T-Test for mean 0: log(ba)

```
##
##  One Sample t-test
##
## data:  data
## t = 0.809, df = 3206, p-value = 0.4186
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0004669204  0.0011228786
## sample estimates:
##    mean of x
## 0.0003279791
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains zero, therefore the mean is statistically 0 and ther linear trend removed.

Normalcy: Skewness - log(ba)

```
##       skew     lwr.ci     upr.ci
```

```
## -0.5234508 -0.6194119 -0.5355653
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The skewness of log(ba) is left-skewed, thus not normal.

Normalcy: Kurtosis - log(ba)

```
##     kurt   lwr.ci   upr.ci
## 21.12742 21.60890 22.16531
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The distribution of log(ba) has tall Kurtosis, thus not normal.

Constant Variance: Breusch-Pagan Test - log(ba)

```
##
##  studentized Breusch-Pagan test
##
## data:  lm(data ~ seq(1, length(data)))
## BP = 55.286, df = 1, p-value = 1.042e-13
##
## Breusch-Pagan: *NON*-constant variance, possible clustering,
## heteroscedastic -> reject H0
```

```
##    BP
## FALSE
```

As we've seen in the plot, the log(ba) data has non-constant variance via the Breusch-Panan test.

Serial Correlation: Box-Ljung test - log(ba)

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 137.49, df = 30, p-value = 9.992e-16
##
## Box-Ljung: implies dependency present over 30 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung test implies lag dependency, thus the log(ba) data has serial correlation.

Box-Ljung ARCH test: log(ba)

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 3956.1, df = 12, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 12 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung ARCH test suggests that ARCH effects are present in the log(ba) data due to its small p-value < 0.05.

2.2. Build a GARCH model with Gaussian innovations for the log return series. Perform model checking and write the model to be fitted.

We create the following GARCH model.

```r
m <- garchFit(~garch(1,1),data=ba,trace=F)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~garch(1, 1), data = ba, trace = F)
##
## Mean and Variance Equation:
##  data ~ garch(1, 1)
## <environment: 0x0000018ba9f1d5b0>
##  [data = ba]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu       omega      alpha1       beta1
## 9.1382e-04  8.1509e-06  8.9476e-02  8.8989e-01
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     9.138e-04   2.714e-04    3.367 0.000761 ***
## omega  8.151e-06   1.392e-06    5.857  4.7e-09 ***
## alpha1 8.948e-02   9.773e-03    9.156  < 2e-16 ***
## beta1  8.899e-01   1.130e-02   78.730  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Log Likelihood:
##  8427.656    normalized:  2.627894
##
## Description:
##  Tue Mar 28 19:56:09 2023 by user: Reed
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  1989.845  0
##  Shapiro-Wilk Test  R    W      0.9694177 0
##  Ljung-Box Test     R    Q(10)  12.23429  0.2696825
##  Ljung-Box Test     R    Q(15)  13.62309  0.5542851
##  Ljung-Box Test     R    Q(20)  17.5601   0.6163615
##  Ljung-Box Test     R^2  Q(10)  2.929826  0.9830276
##  Ljung-Box Test     R^2  Q(15)  3.178975  0.9994251
##  Ljung-Box Test     R^2  Q(20)  6.077029  0.9987894
##  LM Arch Test       R    TR^2   3.001689  0.9955321
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -5.253294 -5.245719 -5.253297 -5.250578
```

Based on the summary above, we have the following model:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1} + \mu$$

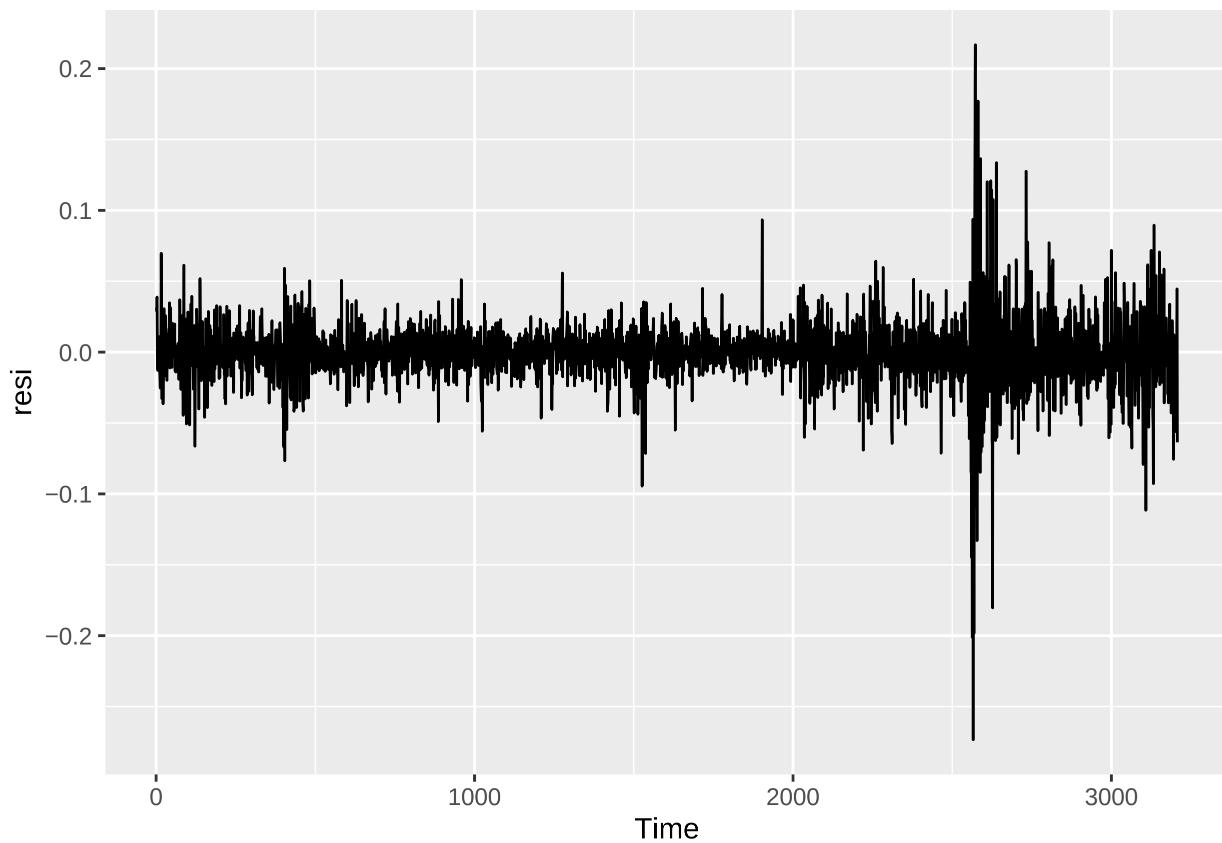
Adding the alpha and beta coefficients give us the following:

```
## [1] 0.9793683
```

The combined value of 0.9793 suggests the model volatility would be slow to revert back to the mean, but overall the volatility is persistent.

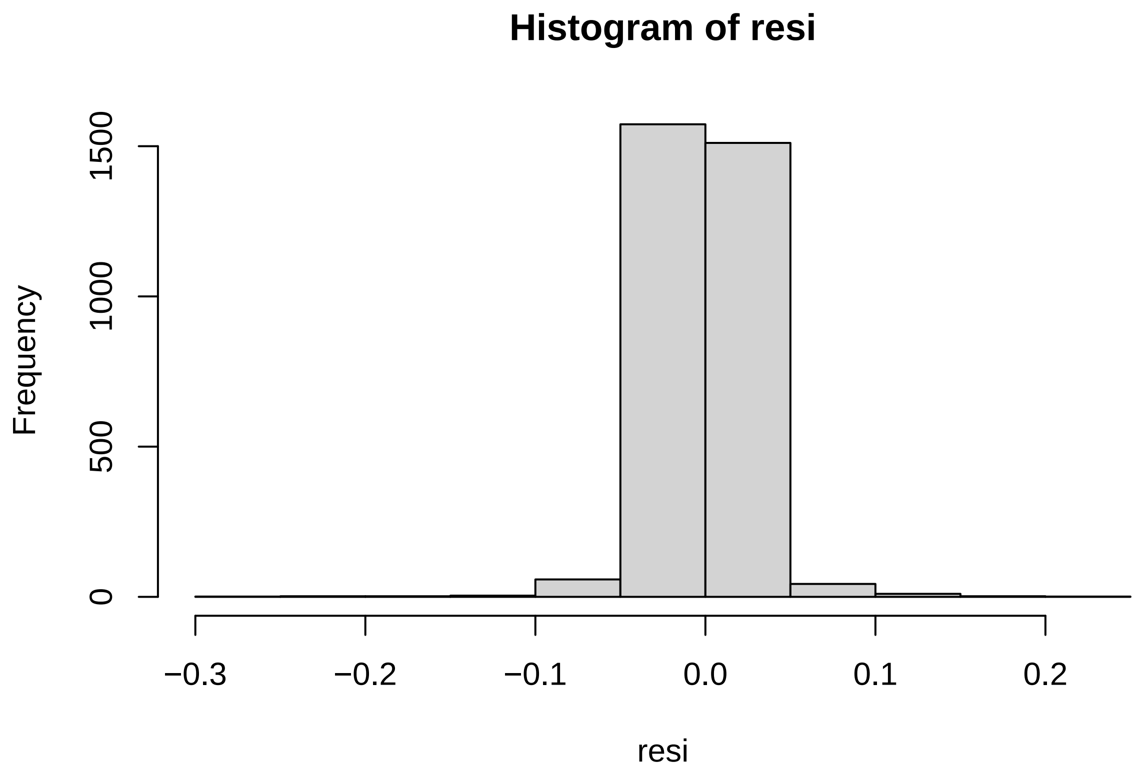
We will conduct model diagnostics.

Plot:



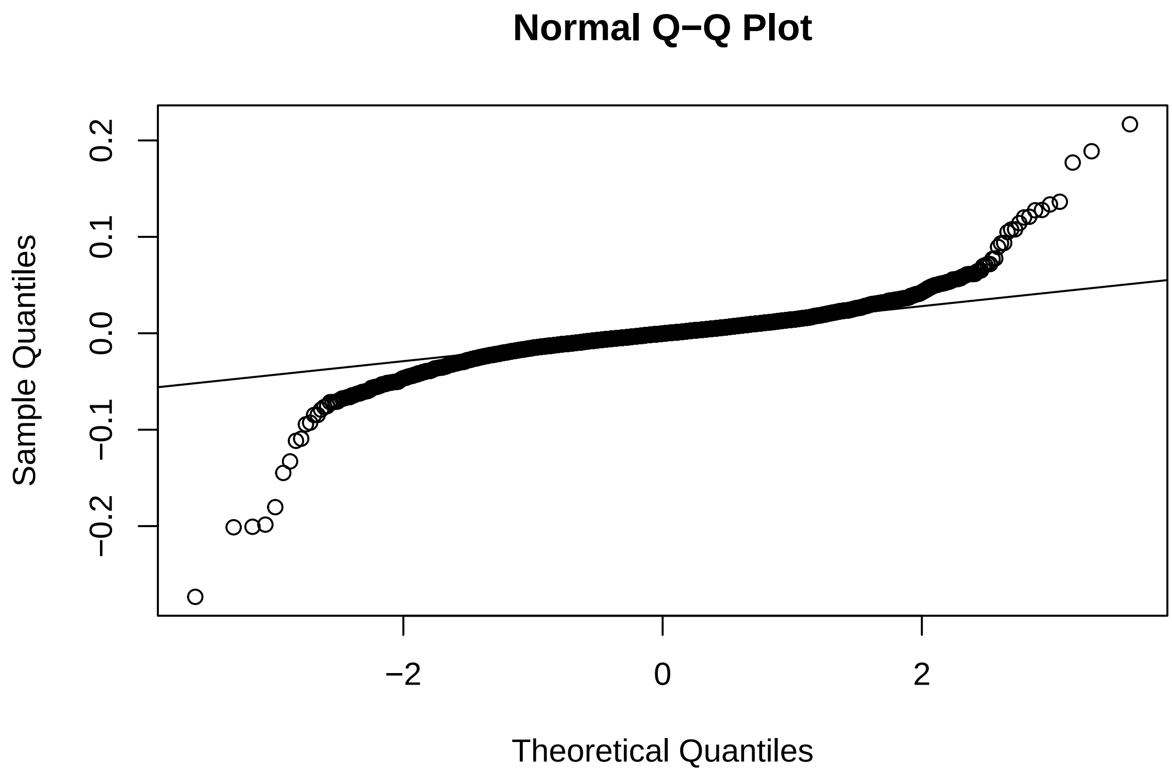
We notice mean zero but also non-constant variance.

Histogram:



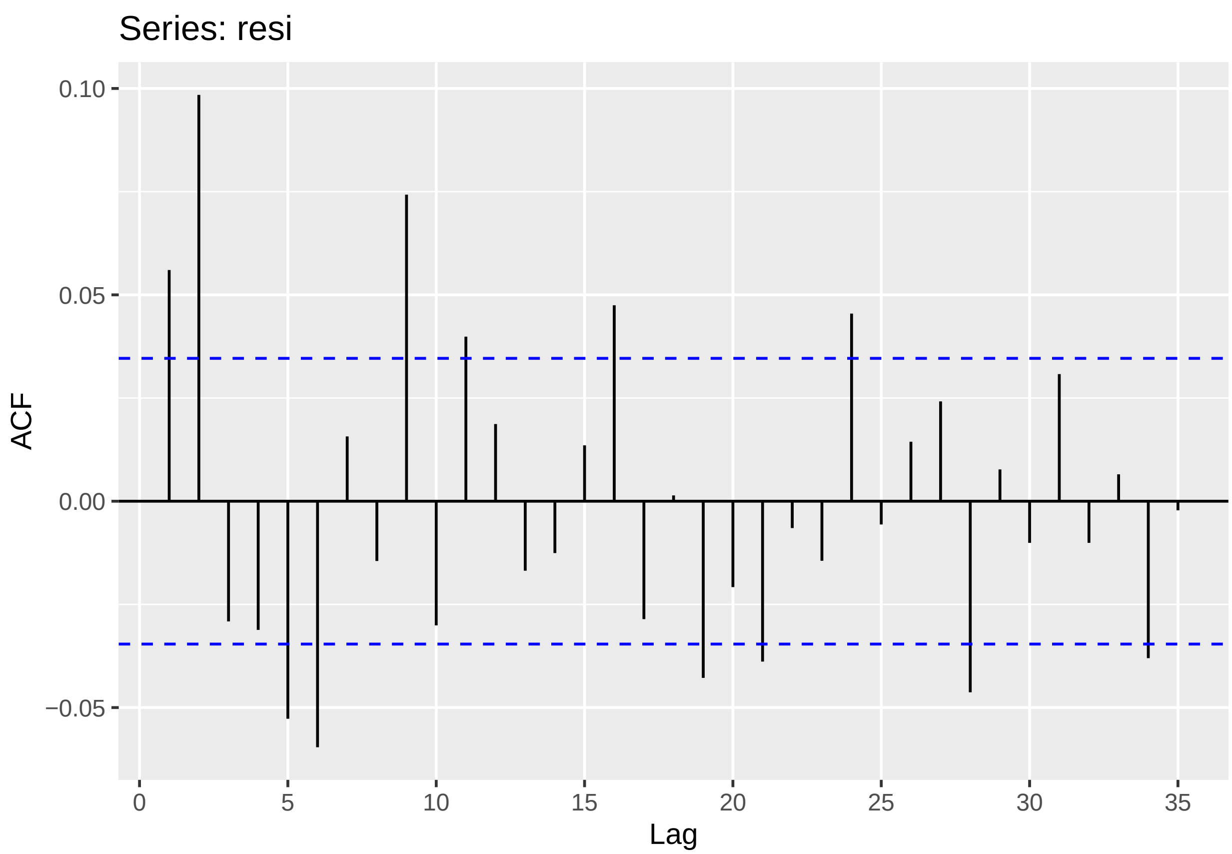
We can see tall, left skewness, suggesting that the distribution is not normal.

Q-Q plot:



We can see thick tails at the ends of the Q-Q plot, indiciating a not-so-good fit of the data to the model.

ACF Plot: standard residuals



The model looks fairly stationary.

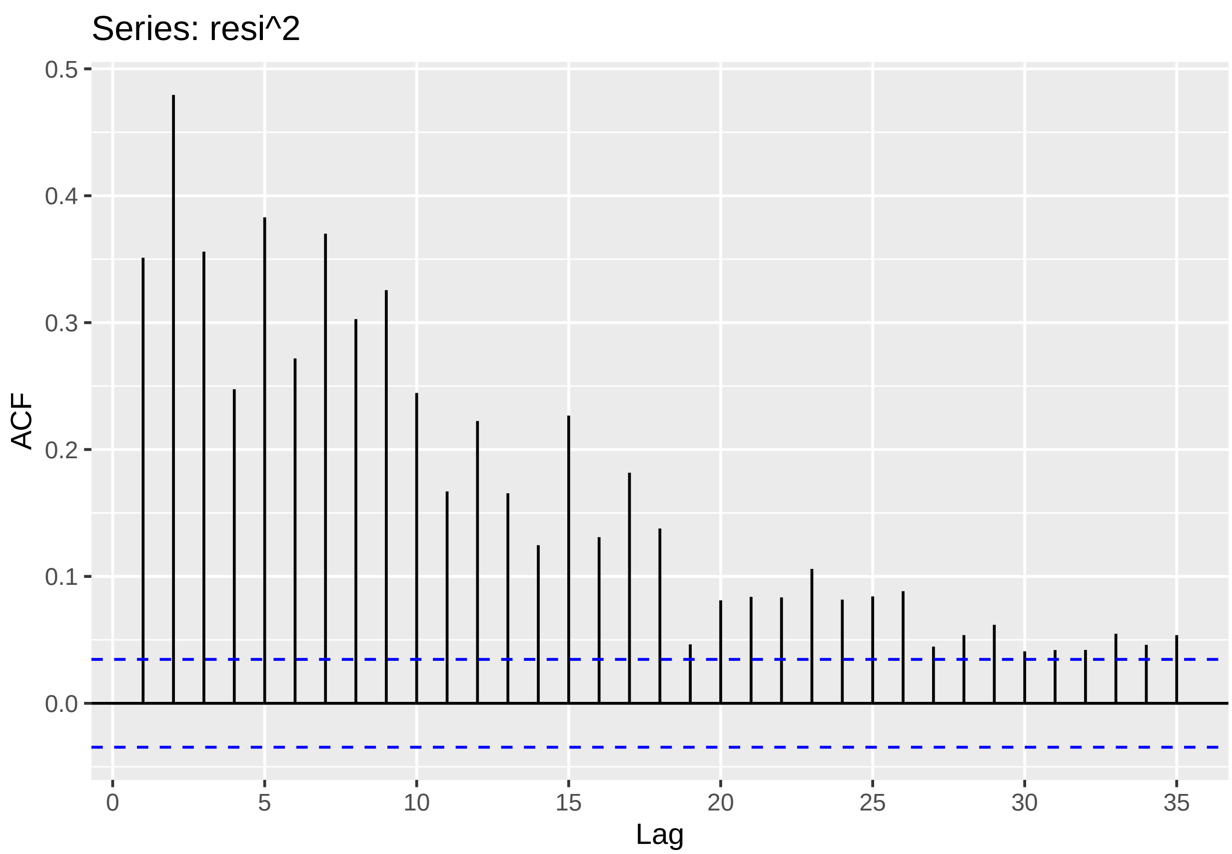
Lad dependence: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 95.767, df = 12, p-value = 3.775e-15
##
## Box-Ljung: implies dependency present over 12 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The test confirms the model residuals does shows non-constant variance.

ACF Plot: squares of the residuals



The ACF plot of the squares of the residuals show quite a bit of serial correlation with most of its lags are out the of the CI threshold.

Box-Ljung test: squares of the residuals

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 3964.7, df = 12, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 12 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The Box-Ljung test's p-value is lower than 0.05, and shows there are ARCH effects and suggests serial autocorrelation.

T-test for 0

```
##
##  One Sample t-test
##
## data:  data
## t = -1.445, df = 3206, p-value = 0.1485
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0013807447  0.0002090543
## sample estimates:
##     mean of x
## -0.0005858452
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains 0, therefore the mean is statistically 0.

Normalcy: skewness

```
##       skew     lwr.ci     upr.ci
## -0.5234508 -0.5772398 -0.4882573
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model residual distribution contains left skewness, thus is not normal.

Normalcy: Kurtosis

```
##     kurt   lwr.ci   upr.ci
## 21.12742 21.53215 22.09370
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The residuals distribution contain thick-tailed Kurtosis, indicating a non-normal distribution.

Stationarity: ADF Test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -9.881
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:10 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test supports the model is stationary

Stationarity: KPSS Test

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.0737
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test supports the model is stationary.

2.3. Fit a GARCH model with skew-Student-t innovations ($cond.dist = "sstd"$) to the log return series. Perform model checking and write the model to be fitted. Based on the fitted model, is the monthly log returns of ba stock skewed? Why?

```
m <- garchFit(~garch(1,1),data=ba,trace=F,cond.dist="sstd")
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~garch(1, 1), data = ba, cond.dist = "sstd",
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ garch(1, 1)
## <environment: 0x0000018ba6c1ac28>
##  [data = ba]
##
## Conditional Distribution:
##  sstd
##
## Coefficient(s):
##         mu       omega      alpha1       beta1        skew       shape
## 0.00091804  0.00000611  0.10235057  0.88842116  0.99603434  4.86601113
```

```
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     9.180e-04   2.591e-04    3.543 0.000395 ***
## omega  6.110e-06   1.648e-06    3.708 0.000209 ***
## alpha1 1.024e-01   1.503e-02    6.809 9.85e-12 ***
## beta1  8.884e-01   1.507e-02   58.962  < 2e-16 ***
## skew   9.960e-01   2.428e-02   41.015  < 2e-16 ***
## shape  4.866e+00   4.172e-01   11.664  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  8573.741    normalized:  2.673446
##
## Description:
##  Tue Mar 28 19:56:11 2023 by user: Reed
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  2787.375  0
##  Shapiro-Wilk Test  R    W      0.9669383 0
##  Ljung-Box Test     R    Q(10)  11.58828  0.3135558
##  Ljung-Box Test     R    Q(15)  12.93293  0.6074784
##  Ljung-Box Test     R    Q(20)  16.3228   0.6964071
##  Ljung-Box Test     R^2  Q(10)  3.3011    0.9734244
##  Ljung-Box Test     R^2  Q(15)  3.516623  0.9989401
##  Ljung-Box Test     R^2  Q(20)  6.828703  0.9972102
##  LM Arch Test       R    TR^2   3.312451  0.9929072
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -5.343150 -5.331788 -5.343157 -5.339077
```

Based on the summary above, we have the following model:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1} + \mu$$

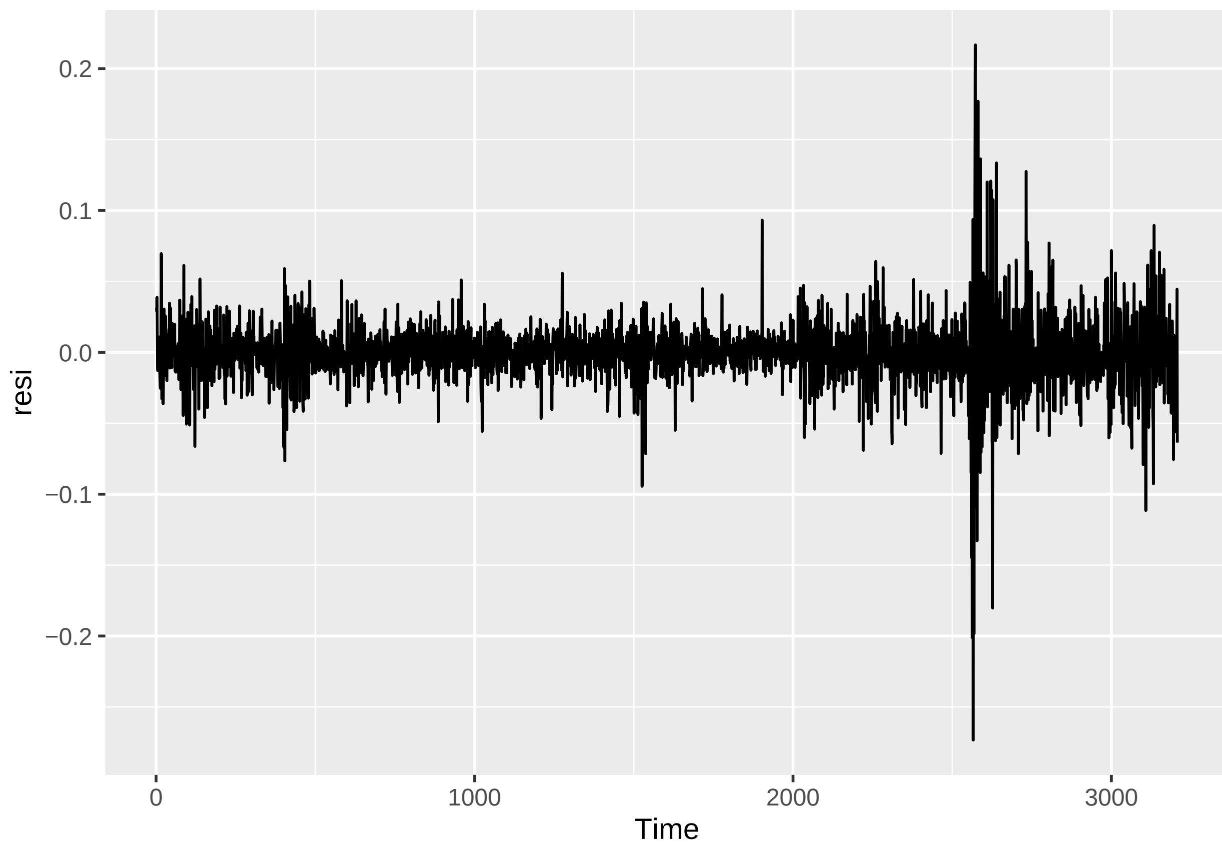
Adding the alpha and beta coeffficients gives us the following:

```r
sum(coef(m)[(length(coef(m))-3):(length(coef(m))-2)])
```

```
## [1] 0.9907717
```

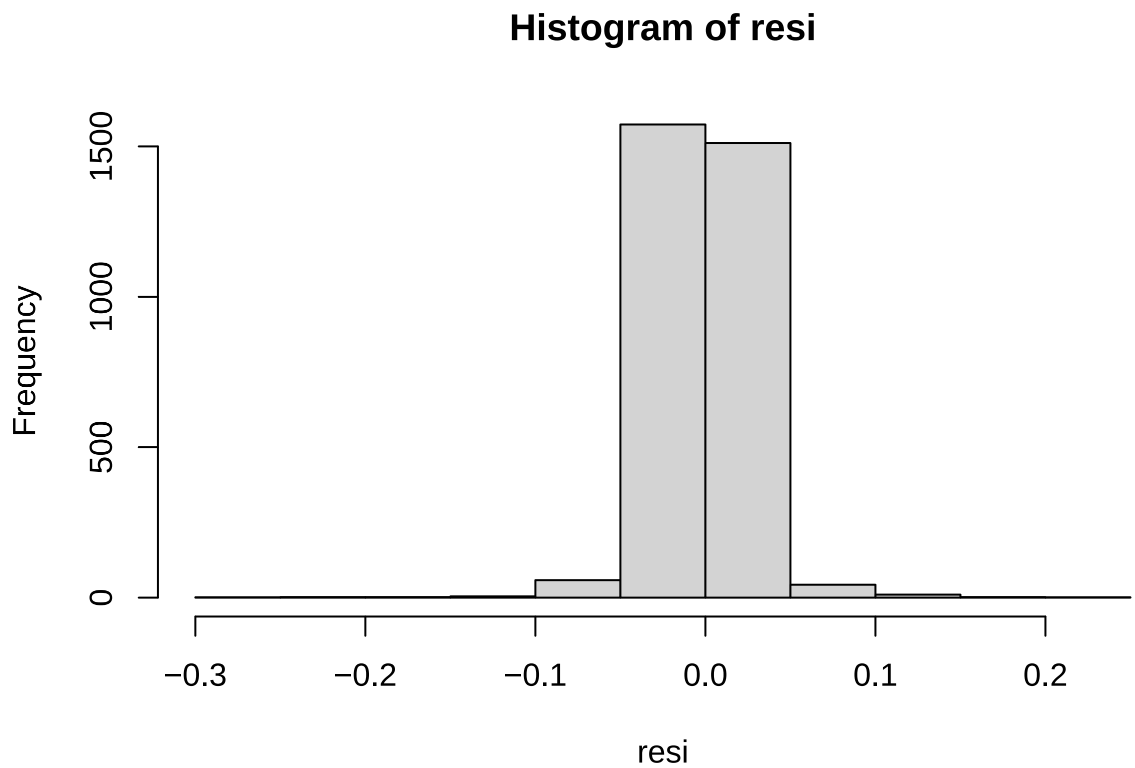
The combined value of 0.9908 is very close to 1 meaning the model's volatility is very slow to revert back to the mean, but overall the volatility is still persistent.

We will conduct model diagnostics.



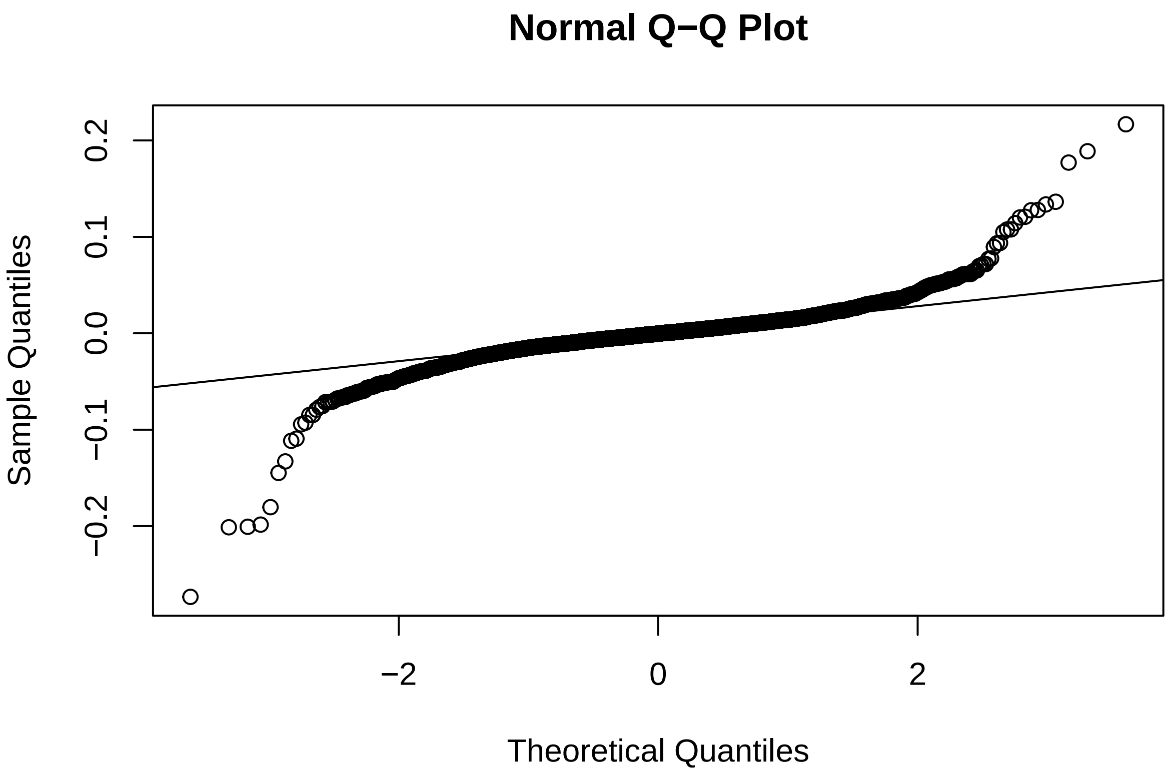
From the plot we can see mean 0 and non-constant variance.

Histogram:



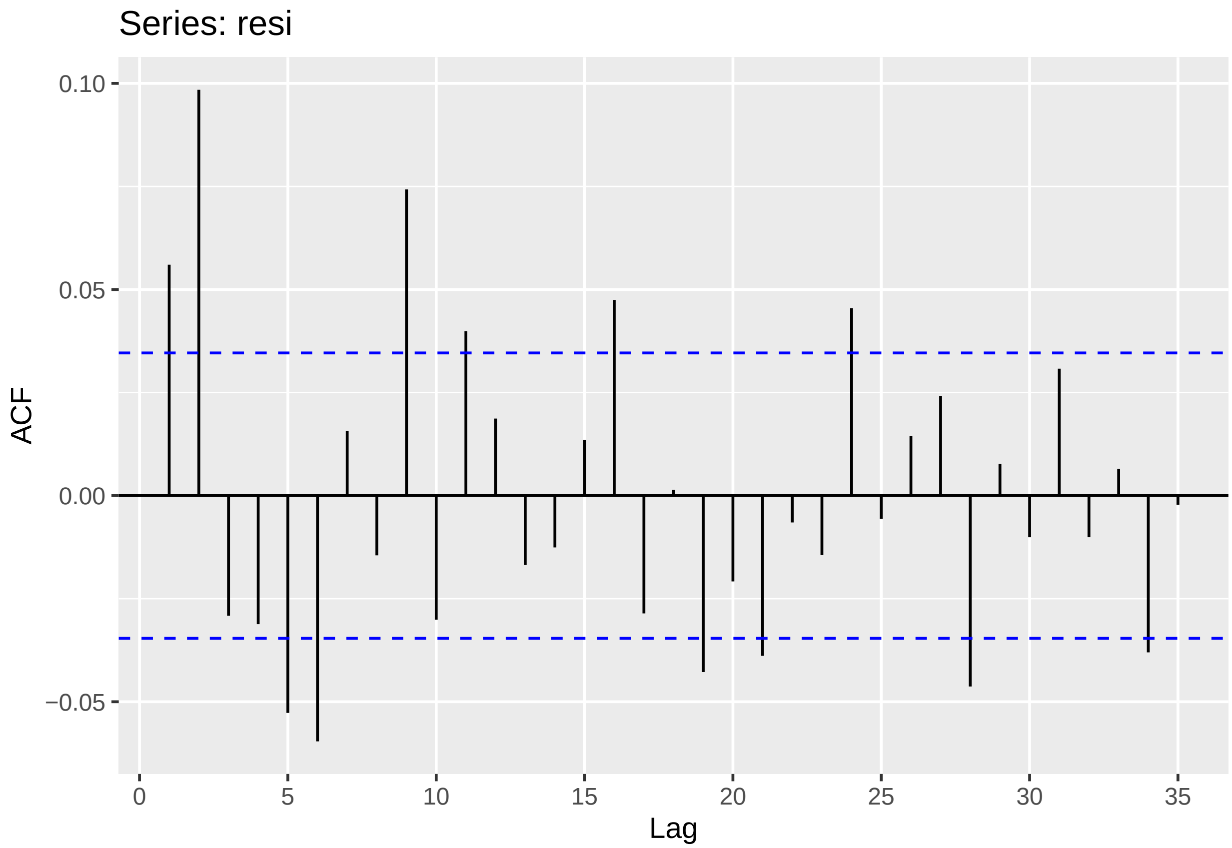
We can observe left skewness and very tall kurtosis, showing the distribution of the model residuals is not normal.

Q-Q Plot:



The thick tails at the end show tall kurtosis, as well as signs of not-so-well goodness-of-fit.

Serial correlation: ACF plot



The ACF plot is not very stationary.

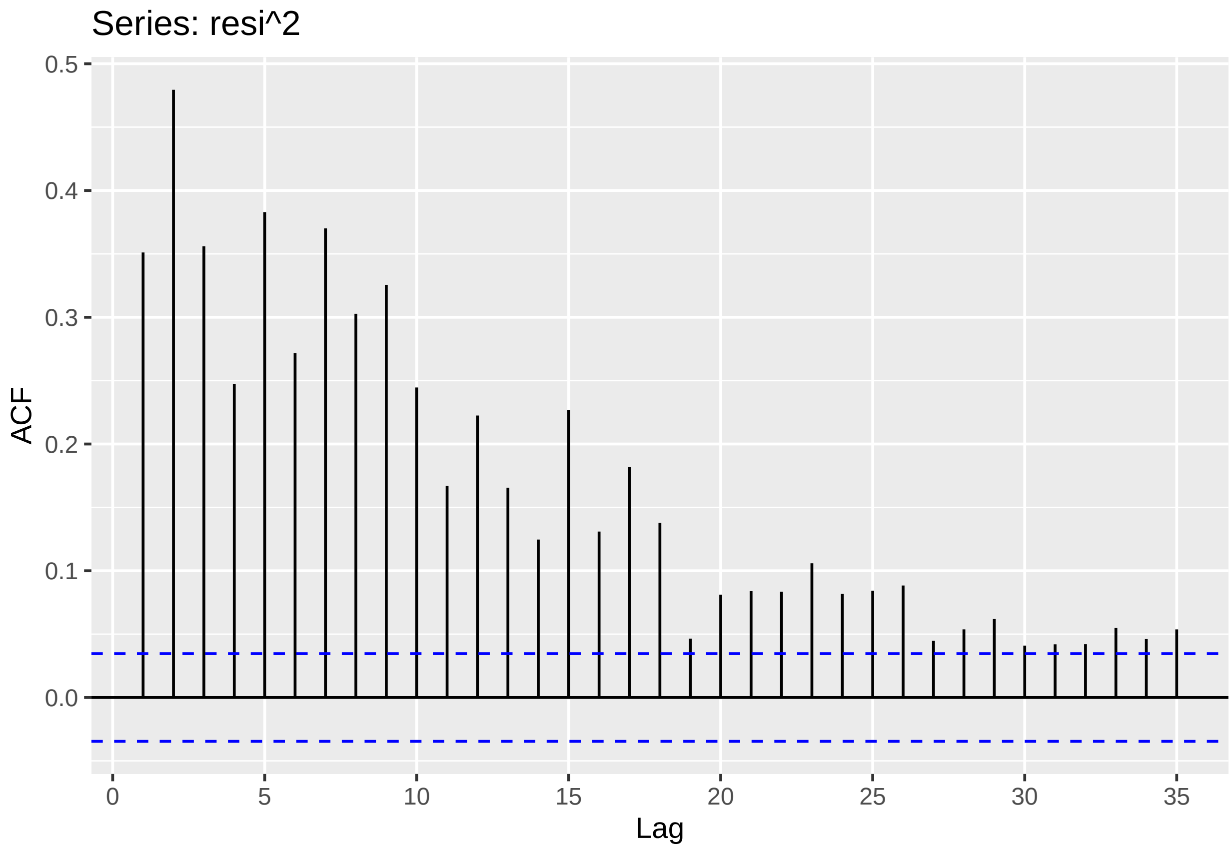
Correlation, lag independende: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 95.767, df = 12, p-value = 3.775e-15
##
## Box-Ljung: implies dependency present over 12 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The test's low p-value < 0.05 indicates lag dependency and serial correlation.

Serial correlation: ACF plot - squares of standardized residuals



The plot suggests much serial correlation between lags.

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 3964.8, df = 12, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 12 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The box-Ljung test for square of serialized residuals

T-test for mean 0:

```
##
##  One Sample t-test
##
## data:  data
## t = -1.4554, df = 3206, p-value = 0.1456
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0013849559  0.0002048431
## sample estimates:
```

```
##     mean of x
## -0.0005900564
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains zero, therefore the mean of the model residuals is statistically 0.

Skewness:

```
##       skew     lwr.ci     upr.ci
## -0.5234508 -0.5567968 -0.4696878
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

As we've seen in the distribution plot, the log(ba) stock returns is skewed to the left.

Kurtosis:

```
##     kurt   lwr.ci   upr.ci
## 21.12742 21.43511 21.98951
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The Kurtosis test confirms the tall distribution we've seen in the distribution plot.

Stationarity: ADF

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -9.8796
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:12 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The KPSS test confirms the model is stationary. Stationarity: KPSS

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.0737
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test confirms the model is stationary.

2.4. Fit a GARCH-M model to the monthly log returns. Write the model to be fitted. Is the risk premium statistically significant? Persistent? Why?

We create the following GARCH-M model:

```r
m <- garchM(ba)
```

```
## Maximized log-likehood:  8426.135
##
## Coefficient(s):
##           Estimate   Std. Error  t value   Pr(>|t|)
## mu     6.23631e-04  3.78731e-04  1.64663   0.099633 .
## gamma -5.90418e-01  1.06209e+00 -0.55590   0.578279
## omega  8.14068e-06  1.42144e-06  5.72706 1.0219e-08 ***
## alpha  8.94764e-02  9.87788e-03  9.05826 < 2.22e-16 ***
## beta   8.89892e-01  1.14793e-02 77.52148 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##           Length Class  Mode
## residuals 3207   -none- numeric
## sigma.t   3207   -none- numeric
```

Based on the summary above, we have the following model:

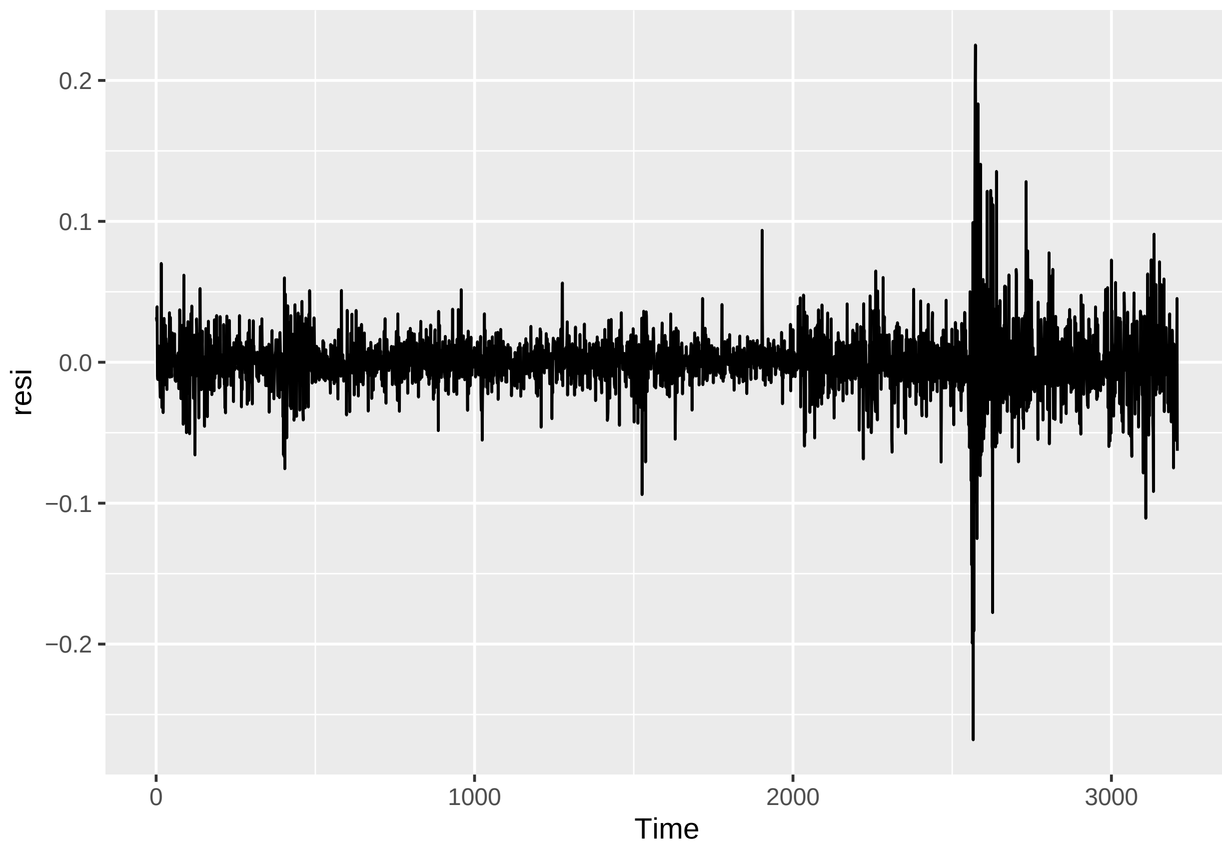
$$x_i = \mu + \gamma\sigma_i^2 + a_i,\ \sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1\sigma_{t-1}^2,\ a_t = \sigma_i\epsilon_t$$

In this particular case, the γ symbol represents the risk premium parameter, and based on the summary above, is not statistically significant compared to the other components of the GARCH-M equation. Hypothetically speaking, though, if the risk premium parameter was statistically significant, it would have been negatively related to its volatility, due to its negative value.

The sum of the alpha and beta coefficients is 0.9794 which means the model's volatility is slow to revert back to the mean but overall the volatility is persistent.

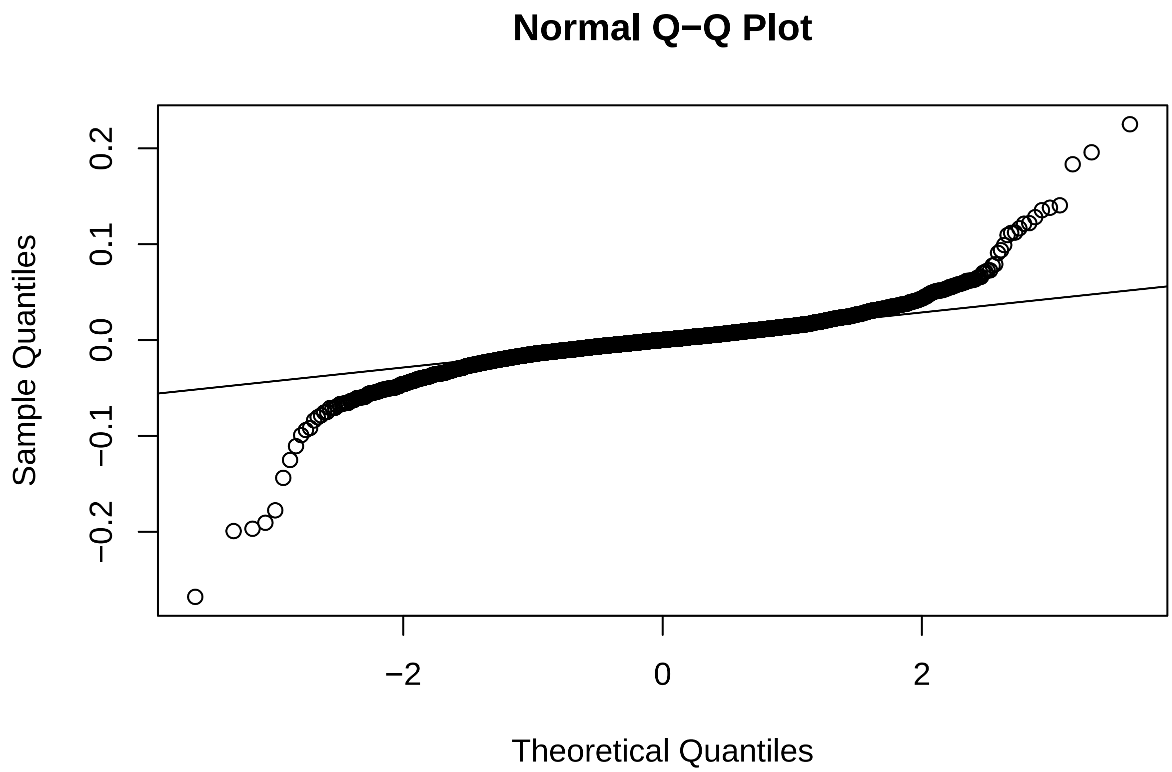
We will conduct model diagnostics.

Plot:



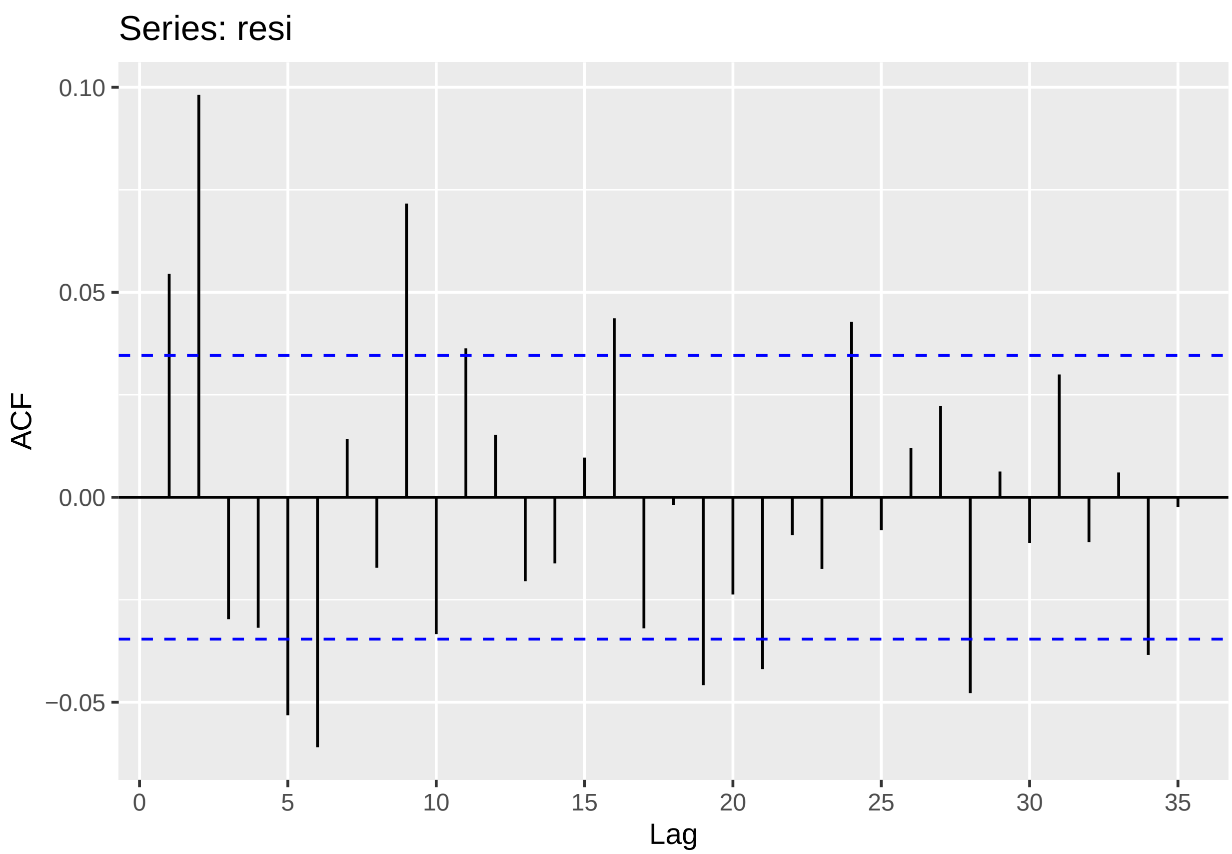
We notice mean 0 and non-constant variance.

Q-Q Plot:



Thick end tails from Kurtosis show a not-so-well goodness-of-fit.

Stationarity: ACF Plot

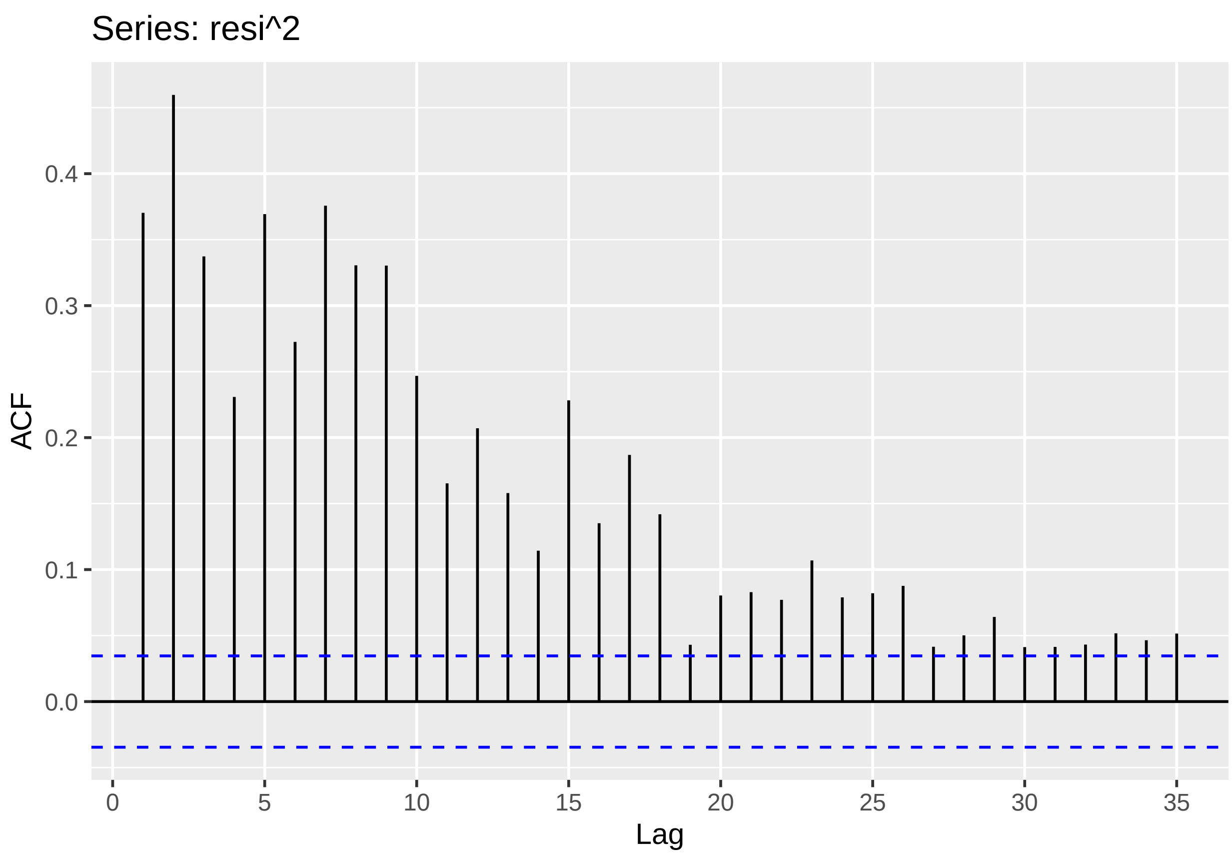


The ACF plots makes the model look somewhat stationary.

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 94.304, df = 12, p-value = 7.216e-15
##
## Box-Ljung: implies dependency present over 12 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The model residuals contain serial autocorrelation.



The ACF plot of the squared residuals show serial correlation among many lags well above the CI threshold.

Lag Independence: Box-Ljung arch test for squared residuals

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 3911.4, df = 12, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 12 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

The test indicates an ARCH effect is present.

T-test for mean 0:

```
##
##  One Sample t-test
##
## data:  data
## t = -0.0010829, df = 3206, p-value = 0.9991
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.0007949523  0.0007940747
## sample estimates:
##     mean of x
## -4.388049e-07
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI includes 0, therefore the mean of the model residuals is statistically 0.

Skewness:

```
##       skew     lwr.ci     upr.ci
## -0.3130276 -0.4262960 -0.3391772
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model residuals has left skewness, indicating that the distribution is not normal.

Kurtosis:

```
##     kurt   lwr.ci   upr.ci
## 21.13865 21.60664 22.17109
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model residuals distribution has tall Kurtosis, indicating a not normal distribution.

Stationarity: ADF test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -10.3758
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:39 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test confirm the model is stationary.

Stationarity: KPSS test

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.0558
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

2.5. Fit a TGARCH(1,1) model to the monthly log returns. Write the model to be fitted. Is the leverage effect statistically significant? Persistent? Why?

We have the following TGarch(1,1) model:

```
m <- Tgarch11(ba)
```

```
## Log likelihood at MLEs:
## [1] 8445.41
##
## Coefficient(s):
##          Estimate  Std. Error  t value   Pr(>|t|)
## mu    5.90760e-04 2.76072e-04  2.13988   0.032365 *
## omega 8.13490e-06 1.40283e-06  5.79891 6.6747e-09 ***
## alpha 4.00255e-02 9.28344e-03  4.31150 1.6215e-05 ***
## gam1  7.73935e-02 1.35681e-02  5.70409 1.1696e-08 ***
## beta  8.99031e-01 1.21698e-02 73.87369 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##             Length Class  Mode
## residuals   3207   ts     numeric
## volatility  3207   ts     numeric
## par            5   -none- numeric
```

Based on the summary above, we have the following model:

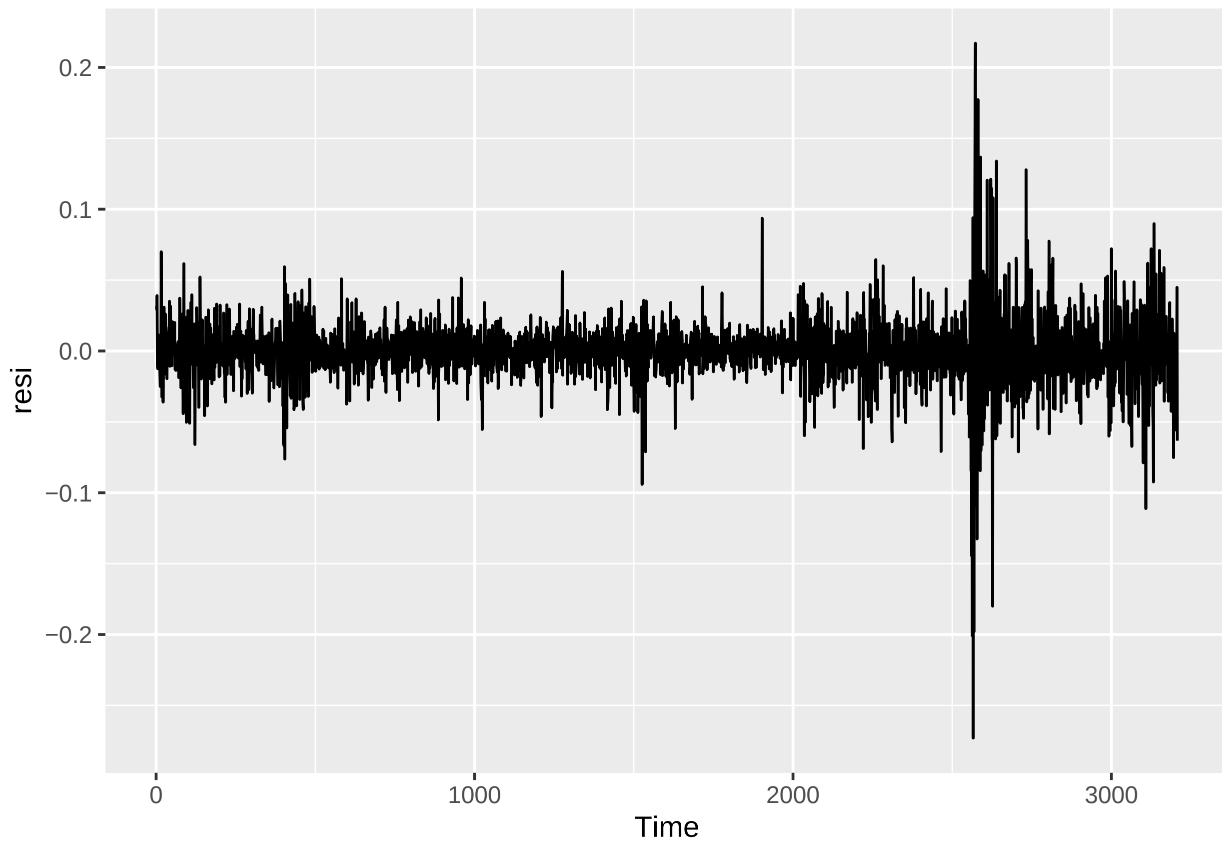
$$a_t = \epsilon_t \sigma_{t|t-1},\ \sigma^2_{t|t-1} = \omega + \alpha_1(|a_{t-1}| + \gamma a_{t-1})^2 + \beta_1 \sigma^2_{t-1},\ for\ -1 < \gamma < 1$$

The sum of the alpha and beta coefficients is 0.9390563 which means the model's volatility is slow to revert back to the mean but overall the volatility is persistent.

In regards to the leverage effect, the γ parameter is statistically significant. With a positive value, it should have a positive leverage effect to the model.

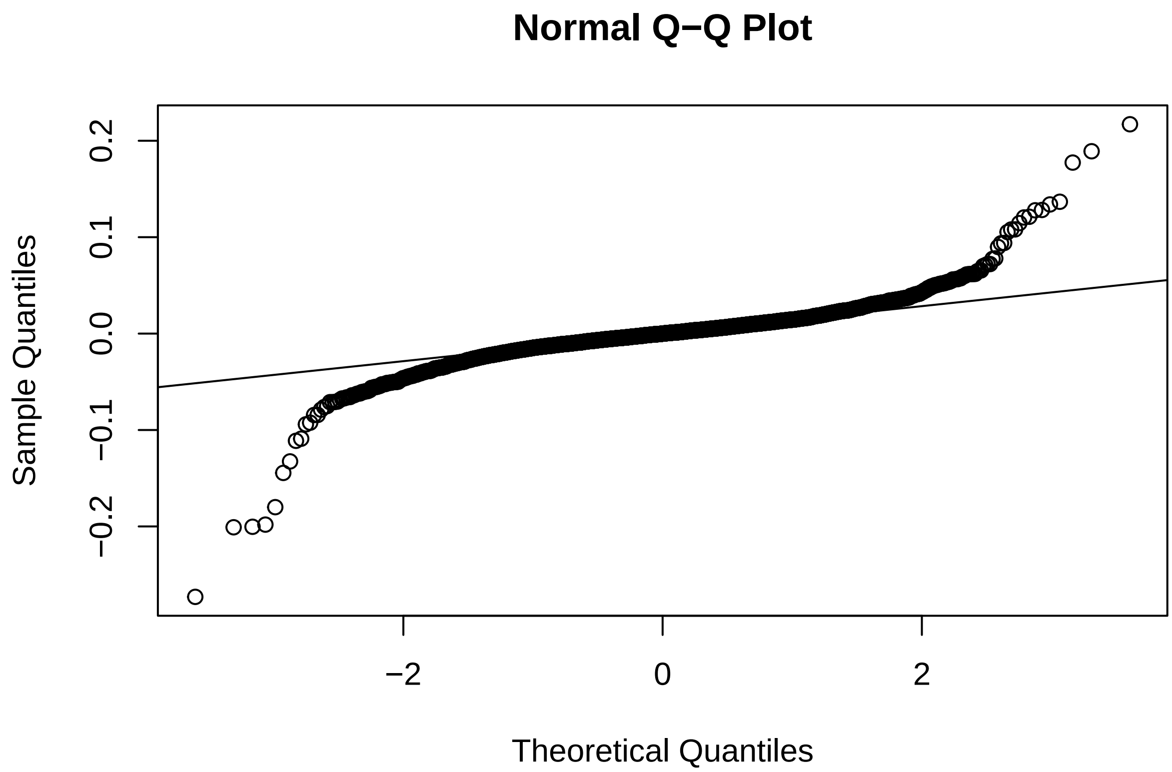
We will conduct model diagnostics.

Plot:



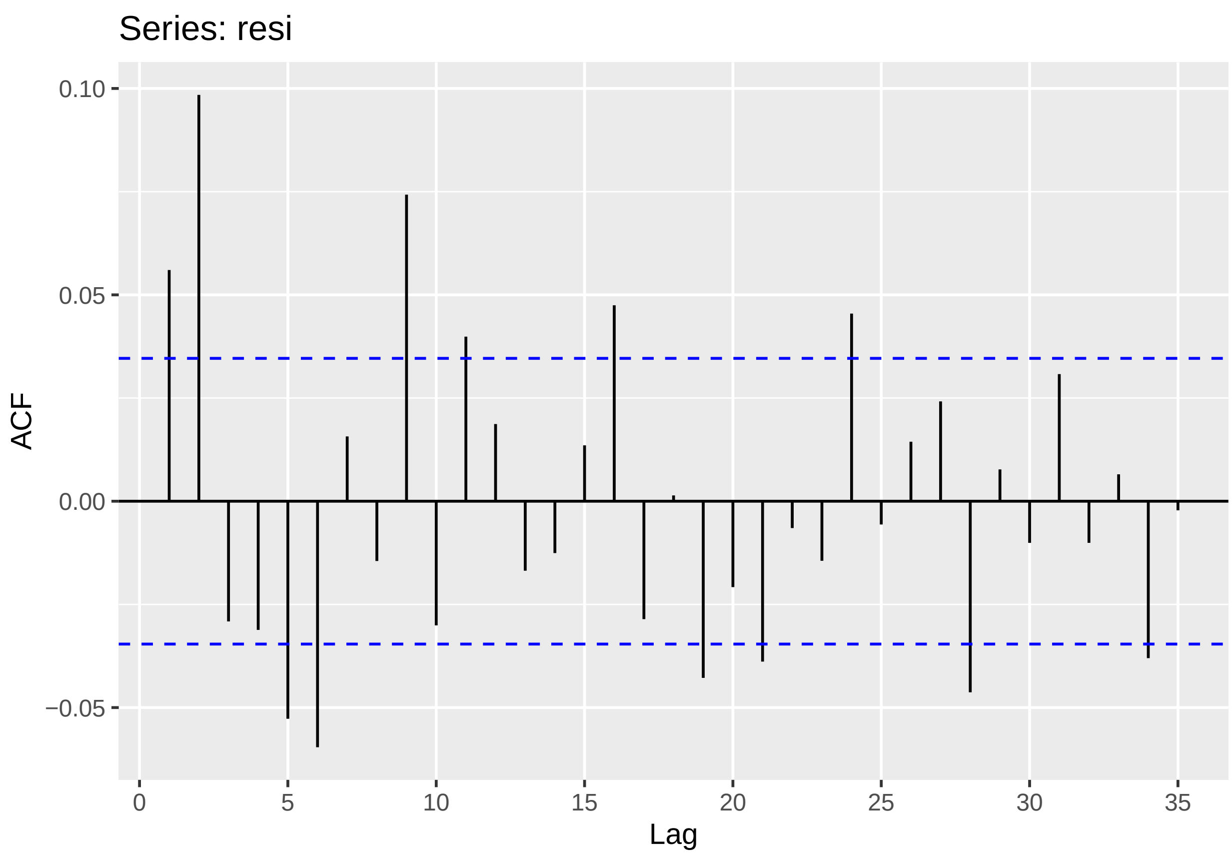
From the plot we can see mean 0 and non-constant variance.

Q-Q Plot:



Due to the thick end tails from Kurtosis, we see a not-so-great goodness-of-fit.

ACF plot:



The ACF plot looks relatively stationary.

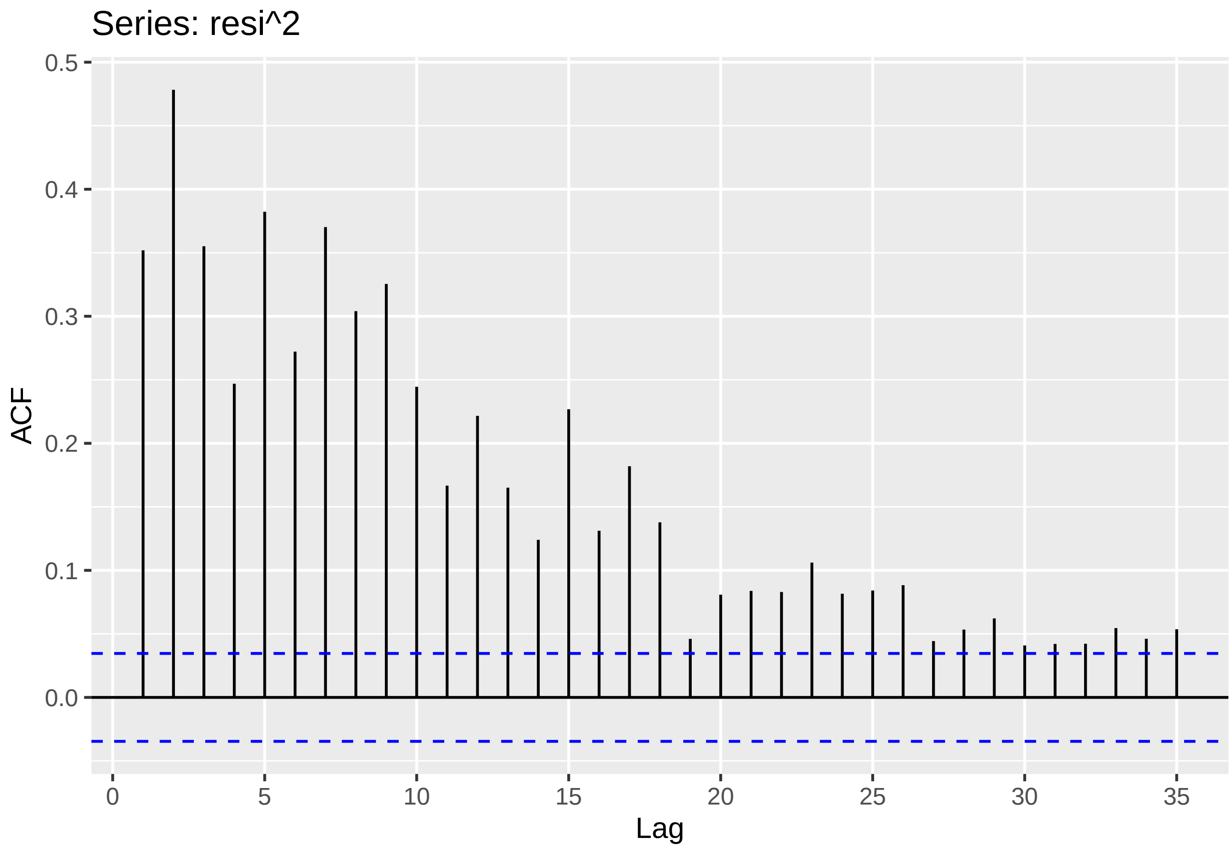
Correlation, lad indepencence: Box-Ljung test

```
##
##  Box-Ljung test
##
## data:  data
## X-squared = 95.767, df = 12, p-value = 3.775e-15
##
## Box-Ljung: implies dependency present over 12 lags,
## autocorrelation present -> reject H0
```

```
## [1] FALSE
```

The test's low p-value < 0.05 show there is serial correlation in the standard residuals.

ACF plot: square of stardard residuals



We notice much serial correlation between lags in the plot.

ARCH effect: Box-Ljung test for squared standard residuals

```
##
##  Box-Ljung test
##
## data:  data^2
## X-squared = 3960.2, df = 12, p-value < 2.2e-16
##
## Box-Ljung ARCH(m): implies *NON*-constant variance over 12 lags, ARCH effects present,
##  possible variance clustering, volatility, heteroscedastic -> reject H0
```

```
## [1] FALSE
```

We can see ARCH effects present in the squared standard residuals.

T-test for mean 0

```
##
##  One Sample t-test
##
## data:  data
## t = -0.64818, df = 3206, p-value = 0.5169
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.0010576808  0.0005321182
## sample estimates:
##     mean of x
## -0.0002627813
##
## T-Test: mean is statistically zero, linear trend *REMOVED* ->
## *FAIL* to reject H0
```

```
## [1] TRUE
```

The 95% CI contains 0, therefore can consider the mean of the standard residuals to be statistically 0.

Skewness

```
##       skew     lwr.ci     upr.ci
## -0.5234508 -0.5834575 -0.4935484
## Skew: has *LEFT* skewness,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model residual distribution has left skewness, therefore not normal.

Kurtosis:

```
##     kurt   lwr.ci   upr.ci
## 21.12742 21.66389 22.22236
## Kurt: has *TALL thick-tailed* (excess) kurtosis,
## property does *NOT* conform to normality and Gaussian PDF
```

```
## [1] FALSE
```

The model residual distribution has very tall Kurtosis, therefore not normal.

Stationarity: ADF test

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 30
##   STATISTIC:
##     Dickey-Fuller: -9.965
##   P VALUE:
##     0.01
##
## Description:
##  Tue Mar 28 19:56:41 2023 by user: Reed
##
## ADF: contains *NO* unit roots over 30 lags,
## indicates *NO* mean drift & *NO* random walk,
## business cycles *NOT* present, series is stationary -> reject H0
```

```
##
## FALSE
```

The ADF test shows the model residuals are stationary

Stationarity: KPSS test

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: tau with 9 lags.
##
## Value of test-statistic is: 0.0737
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.119 0.146  0.176 0.216
##
## KPSS: *NO* unit roots, series is trend stationary -> *FAIL* to reject H0
```

```
## [1] TRUE
```

The KPSS test shows the model residuals are stationary.

3. Report (20 points)

Write a Boeing returns analysis executive summary.

(Based on the analysis in section 2, we will use the TGARCH model from part 2.5 based on its combined alpha and beta coeffient score of 0.9391, the lowest value avaiable out of all the 4 models. It's admittedly not that great as its volitility would be considered quite slow to revert back to the mean, but that is the best we have available right now.)

We are currently working on developing a model that can predict Boeing's short-term daily stock returns from the average return, using our adjusted Boeing stock data from the past 85 trading days up to the end of last month (September 2022). Our development is based on the prototyping of several financial models, and selected the best-performing model using ARCH and GARCH-based modelling tools we currently have available.

The model we chose is based on how well it could dampen everyday volatility from the average mean of returns. With the volatility dampened, we are able to observe a possible general trend and maybe identify critical parameters such as the risk premium, which can determine overall positive or negative returns based on the volatility. With these risk premiums we then can assess the overall leverage effect the volatility can have on the model.

It is still a work-in-progress in regards to developing a functioning forecasting model to present at this time. While we are able to develop short-term predictions using simpler GARCH-based time series models, we feel that they do not fully represent a proper prediction compared to the more complex models we would prefer to use instead. It is a learning process, and as our knowledge base expands we will continue to develop this model to hopefully provide short term predictions on daily average returns.