NASDAQ Composite

Trabajo Final - Modelos de Volatilidad - Financial Statistics

Alexander J Ohrt

22 mars, 2022

Contents

1 Abstract

The mean and conditional variance of the NASDAQ Composite index are modelized. The work shows that . . .

MAKE REFERENCES BETWEEN PLOTS (NUMBERS AND CITATIONS IN TEXT, bookdown is necessary maybe?) LATER.

2 Introduction

Describe

- Scenario and objective of the work. What will be analyzed.
- Precise description of variable (NASDAQ Composite) used in the analysis and description of where the data is gathered from (Yahoo Finance)
- Summary of structure of the work (description of what is done in each part)

The data is gathered directly from Yahoo Finance and can be found here.

3 Empirical Application

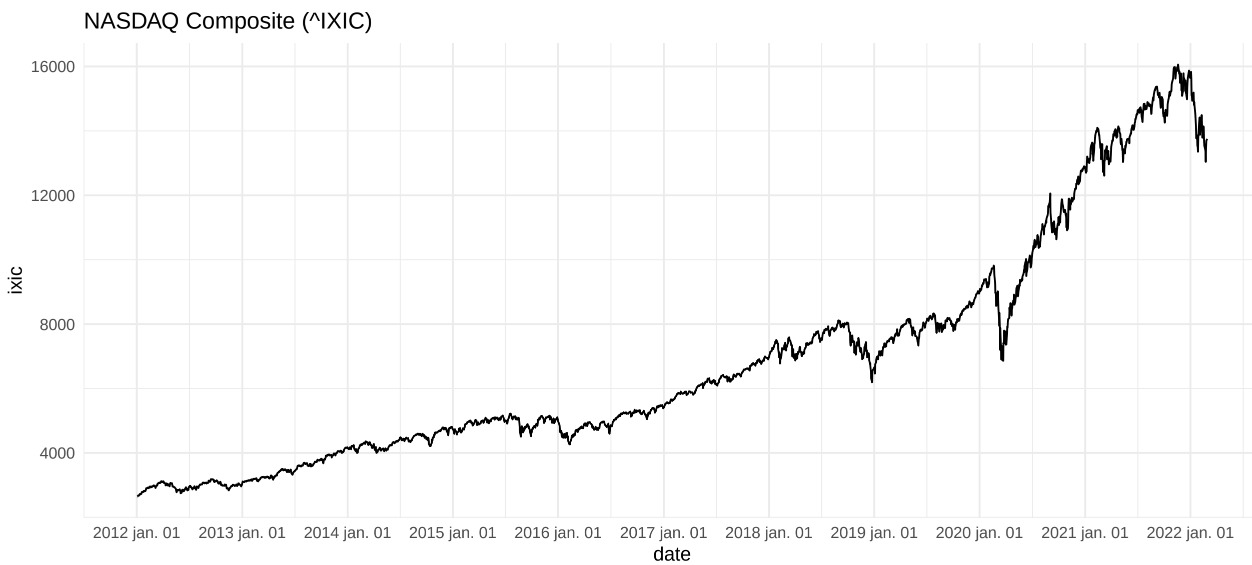
3.1 Load Data

First, we load the NASDAQ Composite data from Yahoo Finance. Note that I downloaded the data in csv format instead of loading directly via the `getSymbols` API, just be to be sure that I always have access to the data.

```
#> [1] 2556    7
```

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

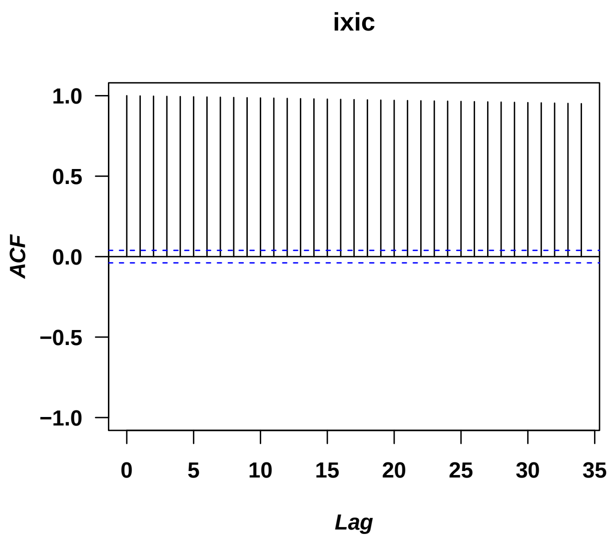
The data does not have any NA values (Weekends and holidays have been removed already), we can start working with the data directly.

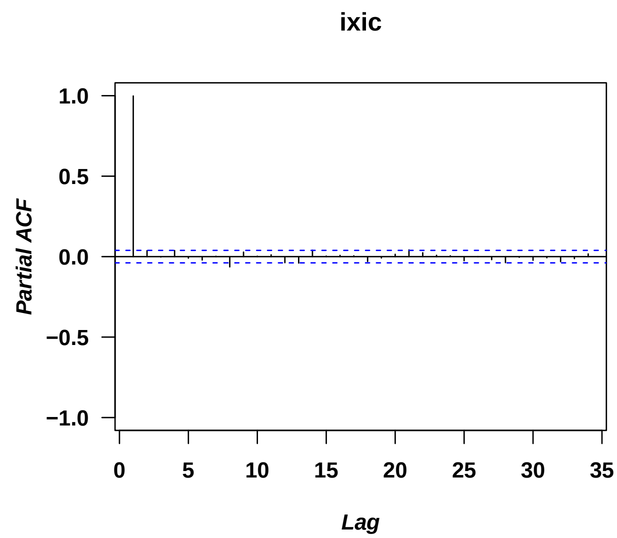


COMMENT: DRAW SOME HAPPENINGS IN THE SERIES (Covid March 2020 and Russia-Ukraine in Feb 2022 + leading up to Feb in the beginning of 2022). Did something happen in Jan 2019 in the US (with tech-companies?) Did something happen in Jan 2016 (small regression).

3.2 Analysis of Stationarity

In order to see if the series is stationary, we will employ both informal and formal tests. Immediately, by looking at the plot above (reference later), the series does not look stationary, since the mean of the process looks to change quite dramatically with time. Some more informal tests are done. The function of autocorrelation and partial autocorrelation (empirical) for the series are plotted below.





As is seen from the function of autocorrelation (ACF), the coefficients decrease slowly. This suggests that the time series is non-stationary, since a stationary series would show exponentially decreasing coefficients in the ACF.

SJEKK AT ALT DETTE GIR MENING (OG BRUKER KORREKTE BEGREPER) SENERE! (TIL SLUTT)

Next, some Ljung-Box tests are done. Here we are testing the joint hypothesis that all m of the correlation coefficients are simultaneously equal to zero. Below we are testing for $m \in \{1, 5, 10, 15, 20\}$.

MAKE A SUMMARY-TABLE HERE INSTEAD LATER! JUST SHOW THE LAGS AND THE P-VALUES.

```
#>
#>  Box-Ljung test
#>
#> data:  ixic
#> X-squared = 2551.4, df = 1, p-value < 2.2e-16
```

```
#>
#>  Box-Ljung test
#>
#> data:  ixic
#> X-squared = 12698, df = 5, p-value < 2.2e-16
```

```
#>
#>  Box-Ljung test
#>
#> data:  ixic
#> X-squared = 25246, df = 10, p-value < 2.2e-16
```

```
#>
#>  Box-Ljung test
#>
#> data:  ixic
#> X-squared = 37633, df = 15, p-value < 2.2e-16
```

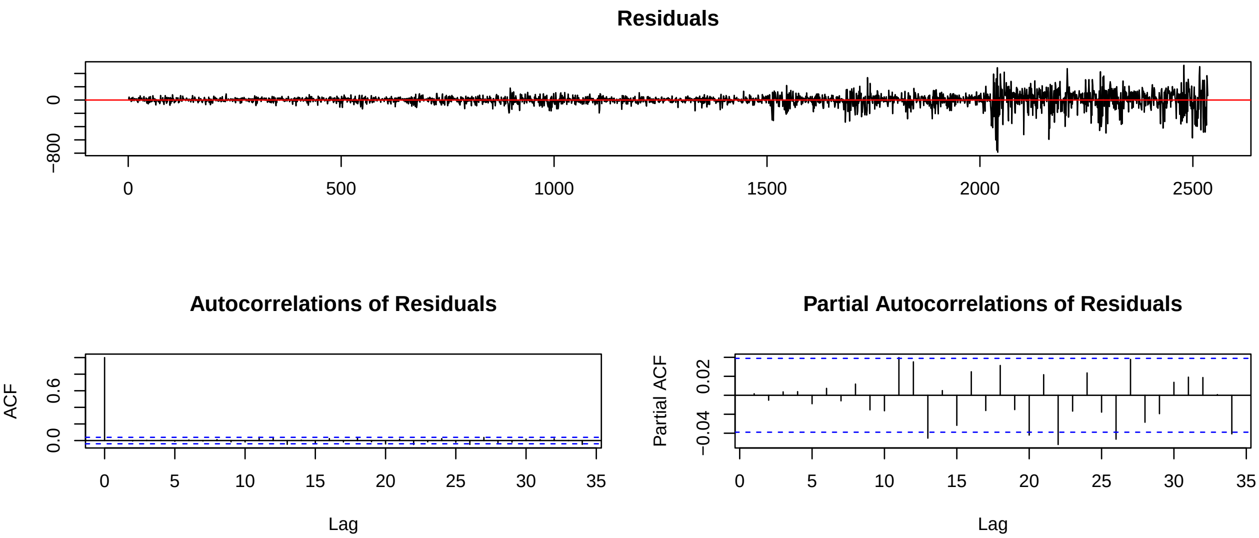
```
#>
#>  Box-Ljung test
#>
#> data:  ixic
#> X-squared = 49855, df = 20, p-value < 2.2e-16
```

All the p-values from the Ljung-Box tests are low, which means that we would reject the null hypothesis that all m correlation coefficients are simultaneously equal to zero. This further suggests that the series is non-stationary.

Next, some formal tests are done to check stationarity of the series. First, the Augmented-Dickey-Fuller (ADF) unit root test is done. The null hypothesis for this case states that the series is integrated of order 1, i.e. that it is non-stationary. Below, the ADF test is done assuming both a stochastic and deterministic trend in the data. The maximum number of lags considered are 20 and the number of lags used are chosen by BIC.

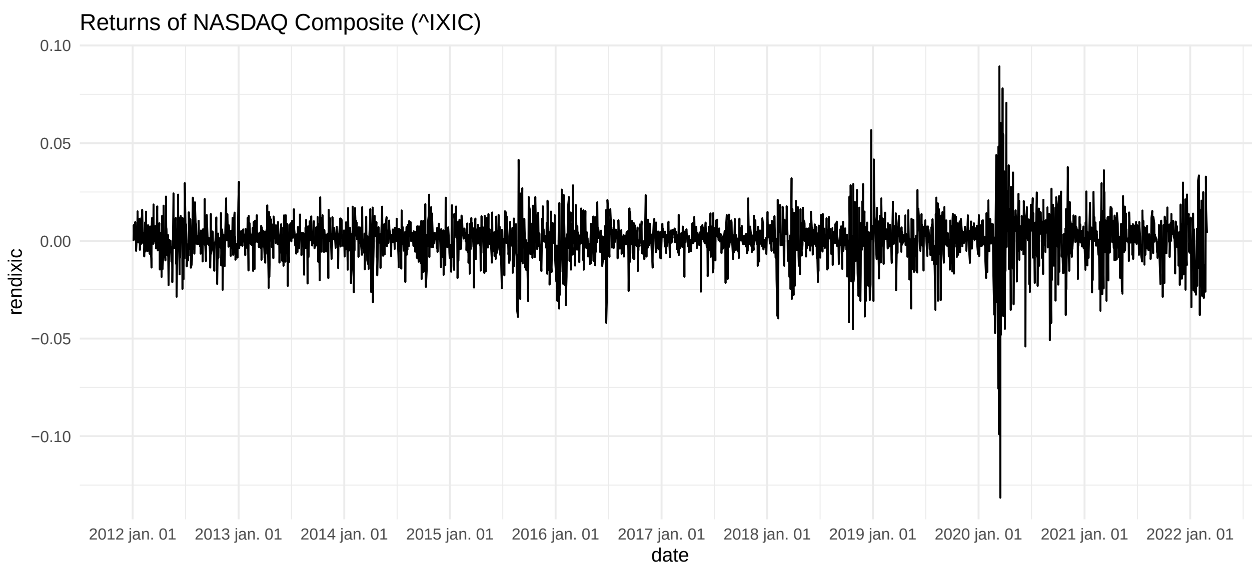
```
#>
#> ###############################################
#> # Augmented Dickey-Fuller Test Unit Root Test #
#> ###############################################
#>
#> Test regression trend
#>
#>
#> Call:
#> lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -785.57  -28.42    3.81   35.46  523.61
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)  4.013754   4.319321   0.929  0.35285
#> z.lag.1     -0.002490   0.001457  -1.709  0.08751 .
#> tt           0.013739   0.006845   2.007  0.04482 *
#> z.diff.lag1 -0.090268   0.019879  -4.541 5.87e-06 ***
#> z.diff.lag2  0.051188   0.019870   2.576  0.01005 *
#> z.diff.lag3 -0.004620   0.019903  -0.232  0.81646
#> z.diff.lag4 -0.062694   0.019939  -3.144  0.00168 **
#> z.diff.lag5  0.007115   0.019994   0.356  0.72197
#> z.diff.lag6 -0.026554   0.019982  -1.329  0.18401
#> z.diff.lag7  0.084723   0.020069   4.222 2.51e-05 ***
#> z.diff.lag8 -0.104673   0.020111  -5.205 2.10e-07 ***
#> z.diff.lag9  0.062169   0.020173   3.082  0.00208 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 97.42 on 2523 degrees of freedom
#> Multiple R-squared:  0.05128,    Adjusted R-squared:  0.04715
#> F-statistic:  12.4 on 11 and 2523 DF,  p-value: < 2.2e-16
#>
#>
#> Value of test-statistic is: -1.7093 3.3031 2.0816
#>
#> Critical values for test statistics:
#>       1pct  5pct 10pct
#> tau3 -3.96 -3.41 -3.12
#> phi2  6.09  4.68  4.03
#> phi3  8.27  6.25  5.34
```

From the output it is apparent that BIC chooses 9 lags in the DF test. Moreover, the value of the test-statistic clearly suggests that we cannot reject the null-hypothesis, since the value is much larger than the critical values for this left-sided test. Thus, we would conclude that the series is non-stationary. Note that the test leads to the same conclusion when assuming no trends and when assuming only a drift. Moreover, the same amount of lags are chosen for all three variants. Below, the residuals and the autocorrelation functions of the residuals are plotted, in order to check if the number of lags chosen via BIC is satisfactory.



The autocorrelation function of the residuals has no significant coefficients, which leads us to conclude that the amount of lags chosen via BIC is satisfactory.

Next, we check if the returns (rendimientos) are stationary. Below we calculate the returns and remove the first difference, since it is not a numerical value.



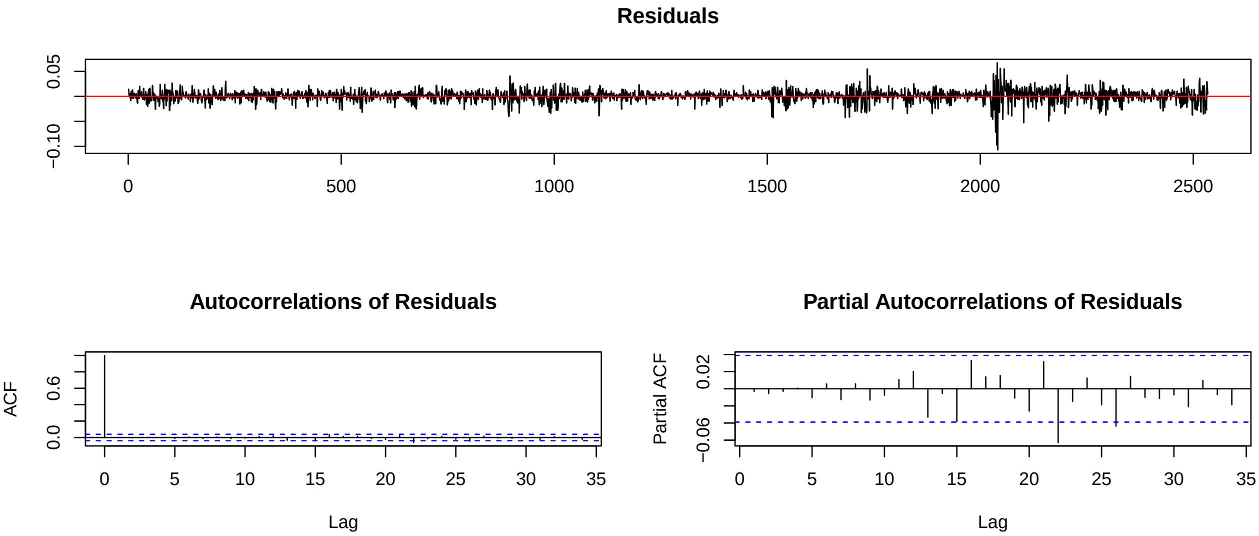
Then, the ADF test is calculated without trends, since there does not look to be any trends in the plot of the returns. Note that, as earlier, the conclusion of the test and the amount of lags that are chosen via BIC are the same when assuming a drift or both types of trends.

```
#>
#> ###############################################
#> # Augmented Dickey-Fuller Test Unit Root Test #
#> ###############################################
#>
#> Test regression none
```

```
#>
#>
#> Call:
#> lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
#>
#> Residuals:
#>       Min        1Q    Median        3Q       Max
#> -0.107159 -0.004275  0.001424  0.006875  0.067084
#>
#> Coefficients:
#>             Estimate Std. Error t value Pr(>|t|)
#> z.lag.1     -1.07192    0.06437 -16.651  < 2e-16 ***
#> z.diff.lag1 -0.02421    0.06026  -0.402 0.687881
#> z.diff.lag2  0.02301    0.05659   0.407 0.684281
#> z.diff.lag3  0.02650    0.05193   0.510 0.609830
#> z.diff.lag4 -0.02609    0.04723  -0.552 0.580728
#> z.diff.lag5 -0.01914    0.04188  -0.457 0.647707
#> z.diff.lag6 -0.06599    0.03630  -1.818 0.069208 .
#> z.diff.lag7  0.02827    0.02966   0.953 0.340668
#> z.diff.lag8 -0.06861    0.01996  -3.437 0.000597 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.01167 on 2525 degrees of freedom
#> Multiple R-squared:  0.5826, Adjusted R-squared:  0.5811
#> F-statistic: 391.6 on 9 and 2525 DF,  p-value: < 2.2e-16
#>
#>
#> Value of test-statistic is: -16.6512
#>
#> Critical values for test statistics:
#>       1pct  5pct 10pct
#> tau1 -2.58 -1.95 -1.62
```

It is apparent that 8 lags are chosen. Moreover, from the test-statistic above we would reject the null-hypothesis, which means that we have found evidence against the hypothesis that the returns are $I(1)$, i.e. evidence against the hypothesis that the original series is $I(0)$. Thus, we conclude that the returns are $I(0)$ or the original series is $I(1)$. This means that, through the results from this test, the original series is not stationary (which we have seen earlier), but the returns are stationary and can be used further in the analysis. SJEKK AT DET JEG SKRIVER HER STEMMER, TROR DET GJØR DET! HAR NOEN NOTATER FRA ET EKSEMPEL PÅ TAVLA PÅ DETTE!

As earlier, the plot below shows that the amount of lags for the ADF test chosen via BIC is satisfactory.



For completeness, we also use the Philips-Perron (PP) test to check stationarity of the series. This test defines the same null-hypothesis as the ADF test, which means that this is a left-tailed test as well. COULD MAKE A TABLE FROM THESE TWO LAST TESTS (WITH THE MOST IMPORTANT STATISTICS), TO SAVE SOME ROOM IF NEEDED, SINCE THE CONCLUSIONS ARE THE SAME AS EARLIER (AS EXPECTED).

```
#>
#> ###################################
#> # Phillips-Perron Unit Root Test #
#> ###################################
#>
#> Test regression with intercept and trend
#>
#>
#> Call:
#> lm(formula = y ~ y.l1 + trend)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -983.02  -26.63    2.68   34.73  658.52
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept) 23.065091  10.220234   2.257   0.0241 *
#> y.l1         0.997252   0.001472 677.449   <2e-16 ***
#> trend        0.014404   0.006891   2.090   0.0367 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 99.37 on 2552 degrees of freedom
#> Multiple R-squared:  0.9992, Adjusted R-squared:  0.9992
#> F-statistic: 1.543e+06 on 2 and 2552 DF,  p-value: < 2.2e-16
#>
#>
#> Value of test-statistic, type: Z-tau  is: -1.669
#>
#>            aux. Z statistics
```

```
#> Z-tau-mu              2.9272
#> Z-tau-beta            1.9503
#>
#> Critical values for Z statistics:
#>                      1pct      5pct     10pct
#> critical values -3.967077 -3.414184 -3.128848
```

All combinations of trend assumptions and/or long or short lags yield the same conclusions as from the output above; namely that we have not found sufficient evidence to reject the null-hypothesis of non-stationarity of the series. Below the PP-test is done with the returns.

```
#>
#> ##################################
#> # Phillips-Perron Unit Root Test #
#> ##################################
#>
#> Test regression with intercept
#>
#>
#> Call:
#> lm(formula = y ~ y.l1)
#>
#> Residuals:
#>       Min        1Q    Median        3Q       Max
#> -0.120202 -0.004652  0.000702  0.006097  0.076985
#>
#> Coefficients:
#>               Estimate Std. Error t value Pr(>|t|)
#> (Intercept)  0.0007315  0.0002344   3.121  0.00182 **
#> y.l1        -0.1345383  0.0196155  -6.859 8.68e-12 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.01183 on 2552 degrees of freedom
#> Multiple R-squared:  0.0181, Adjusted R-squared:  0.01772
#> F-statistic: 47.04 on 1 and 2552 DF,  p-value: 8.683e-12
#>
#>
#> Value of test-statistic, type: Z-tau  is: -57.7836
#>
#>          aux. Z statistics
#> Z-tau-mu            3.1176
#>
#> Critical values for Z statistics:
#>                      1pct      5pct     10pct
#> critical values -3.435853 -2.863173 -2.567664
```

When referring to the returns, the conclusion is the same as for the ADF test; the returns are stationary while the original series is not.

Finally, we use the KPSS test to check stationarity of the series. The null hypothesis for this test states that the series is stationary. In the test below we have chosen to assume the deterministic component as a constant with a linear trend, and we have used short lags. Note that the conclusion is the same with all different variations of assumptions for the test.

```
#>
#> #######################
```

```
#> # KPSS Unit Root Test #
#> #######################
#>
#> Test is of type: tau with 8 lags.
#>
#> Value of test-statistic is: 4.8701
#>
#> Critical value for a significance level of:
#>                 10pct  5pct 2.5pct  1pct
#> critical values 0.119 0.146  0.176 0.216
```

Since this is a right-tailed test, the test-statistic is clearly sufficiently large to reject the null-hypothesis to the lowest significance level shown (0.01). Thus, we conclude that the series is non-stationary, as expected. The test below shows that the returns are stationary, in line with what we have concluded earlier, since we cannot find strong evidence against the null-hypothesis.

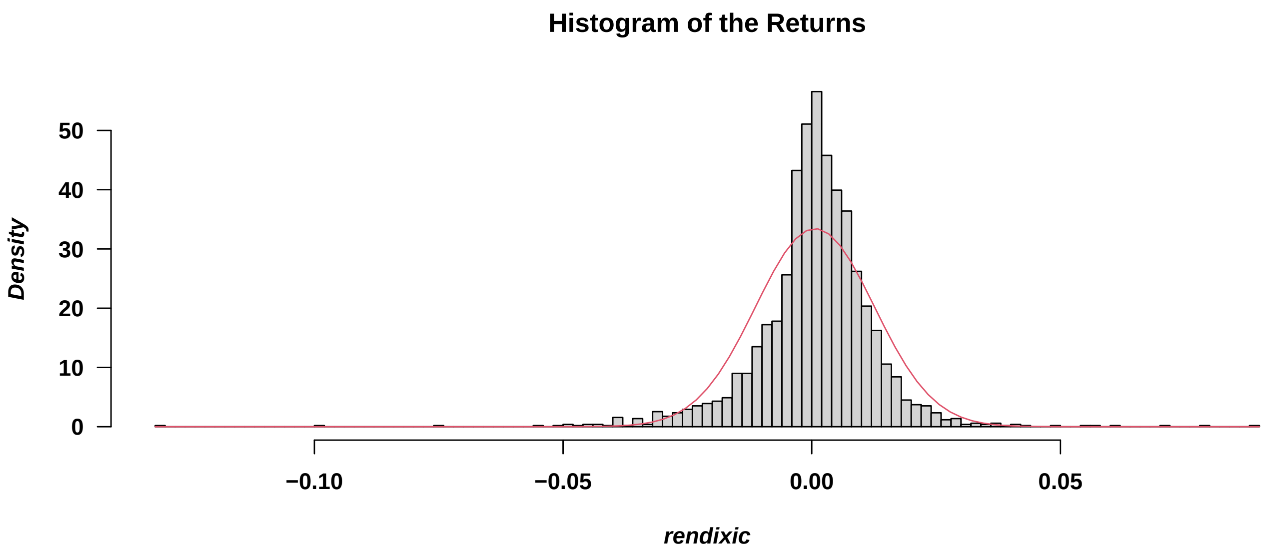
```
#>
#> #######################
#> # KPSS Unit Root Test #
#> #######################
#>
#> Test is of type: mu with 8 lags.
#>
#> Value of test-statistic is: 0.0313
#>
#> Critical value for a significance level of:
#>                 10pct  5pct 2.5pct  1pct
#> critical values 0.347 0.463  0.574 0.739
```

Conclusively, the original time series is not stationary, but the returns are stationary, which means that the returns will be used in the following analysis. We can be relatively certain that this is the case, since all three formal tests, as well as the informal tests, point to this conclusion.

3.3 Basic Statistical Properties of the Stationary Series

Some basic statistical properties of the stationary series, the returns, are shown below.

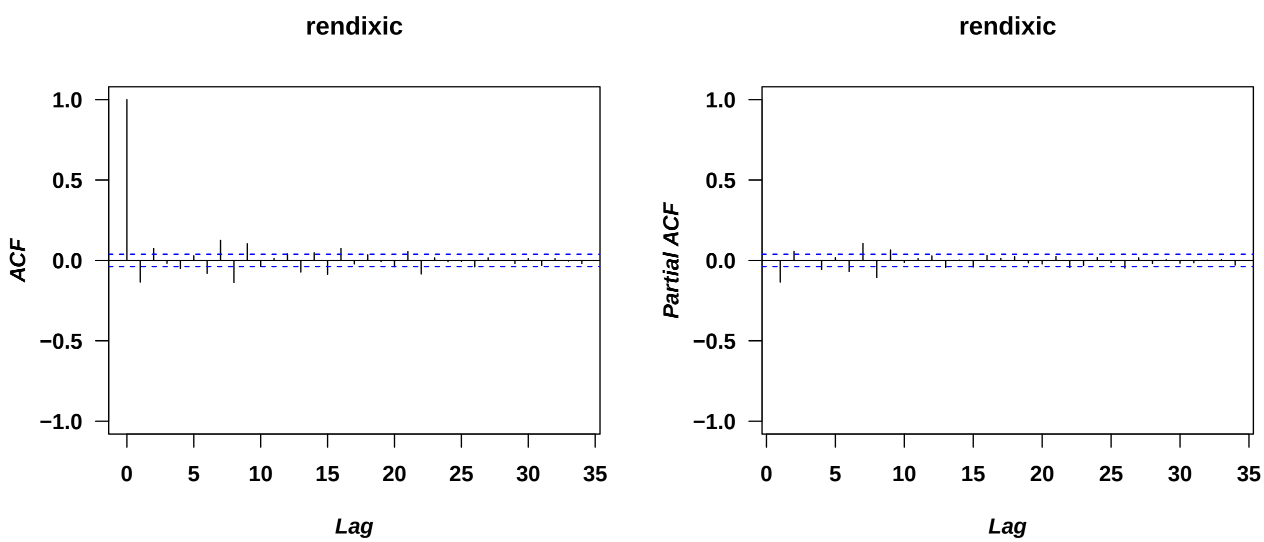
```
#>                rendixic
#> nobs        2555.000000
#> NAs            0.000000
#> Minimum       -0.131492
#> Maximum        0.089347
#> 1. Quartile   -0.003980
#> 3. Quartile    0.006759
#> Mean           0.000645
#> Median         0.001093
#> Sum            1.647064
#> SE Mean        0.000236
#> LCL Mean       0.000182
#> UCL Mean       0.001108
#> Variance       0.000142
#> Stdev          0.011933
#> Skewness      -0.841975
#> Kurtosis      12.309021
```



We can see that the series is Leptokurtic, both by the kurtosis value and from the histogram above. The superposed red curve is a Gaussian distribution with empirical mean and standard error according to the returns of the NASDAQ Composite series. Moreover, the skewness is negative, which means that the distribution of the returns are heavy-tailed in the left tail. This is also apparent from the histogram above.

3.4 Identification, Estimation and Diagnostics of a Model for the Mean

Plotting the autocorrelation functions of the returns.



Note that the third coefficient of both ACF and PACF seems to be non-significant, which might be a hint to what order of model would be fitting. Notice also that both the ACF and the PACF have significant coefficients after the third lag; 6, 7, 8 and 9 seem to be significant. An ARMA of order 6, 7, 8 or 9 seems like a too large order of model to estimate, so I will try with smaller models instead, noting that the third coefficient is non-significant. The table below shows the BIC and the AIC for different orders of ARMA-models. The largest model I have considered is ARMA(3,2) (or ARMA(2,3)), since an ARMA(3,3) yields NaNs in the estimates.

Note that all models we have estimated here have significant coefficient estimates to a predetermined significance level of $\alpha = 0.05$.

Table 1: AIC and BIC of different estimated models for the returns of NASDAQ Composite

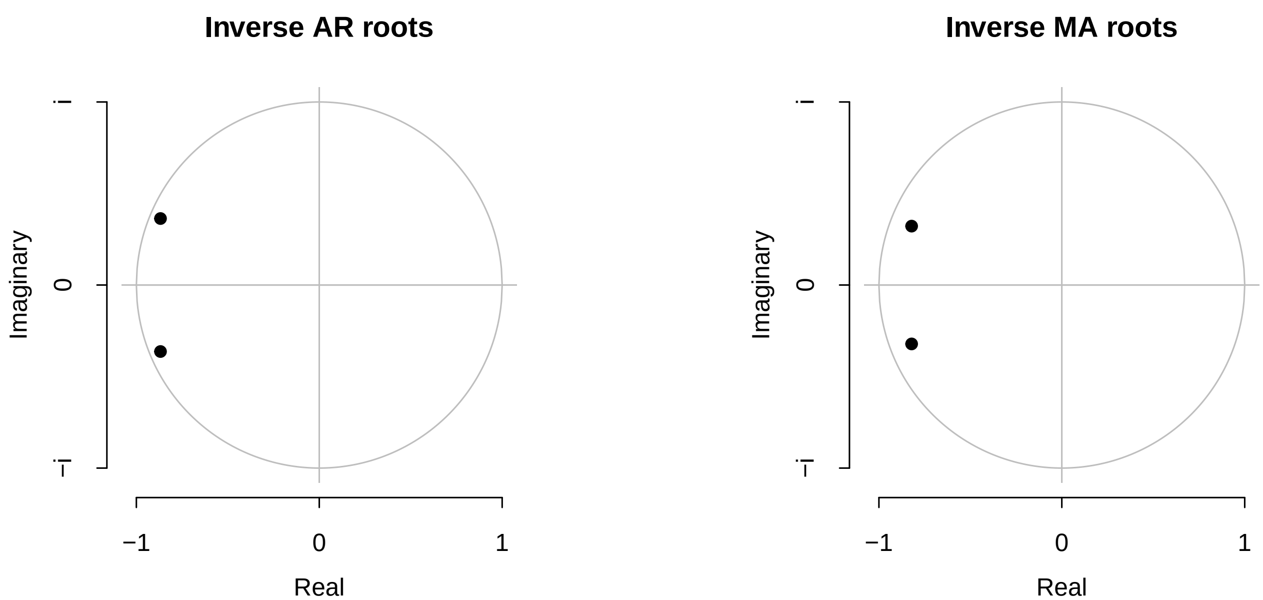
Model	BIC	AIC
AR(1)	-15402.7116191511	-15420.249041659
MA(1)	-15397.436369982	-15414.9737924899
ARMA(1,1)	-15400.9107843157	-15424.2940143262
AR(2)	-15403.1541772126	-15426.5374072231
MA(2)	-15402.9296919165	-15426.312921927
ARMA(2,2)	-15470.3358282124	-15505.4106732282
AR(3)	-15395.308473024	-15424.5375105372
MA(3)	-15397.4156685398	-15426.644706053
ARMA(3,2)	-15386.1619828535	-15427.082635372

The table above clearly shows that ARMA(2,2) yields the lowest AIC and BIC. Moreover, as noted, the estimated coefficients of the model are significant to a level of 0.05. The estimated model ARMA(2,2) is shown below

```
#>
#> Call:
#> arima(x = rendixic, order = c(2, 0, 2), include.mean = TRUE)
#>
#> Coefficients:
#>           ar1      ar2     ma1    ma2  intercept
#>       -1.7362  -0.8856  1.6425  0.778      6e-04
#> s.e.   0.0241   0.0226  0.0326  0.030      2e-04
#>
#> sigma^2 estimated as 0.0001349:  log likelihood = 7758.71,  aic = -15505.41
```

```
#>           ar1           ar2           ma1           ma2     intercept
#>  0.000000e+00  0.000000e+00  0.000000e+00 7.280516e-149  1.595550e-03
```

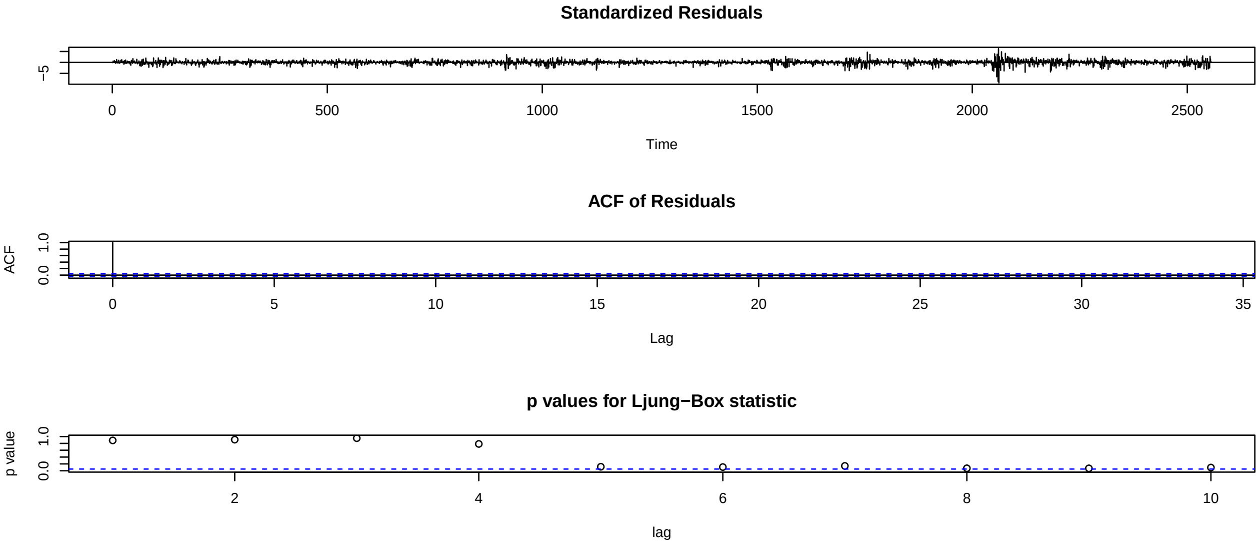
Some model diagnostics have to be done to check if the model is adequate. We must check if the model is stationary.



The inverse roots of the characteristic polynomial of AR and MA are shows above. DO WE WANT ALL

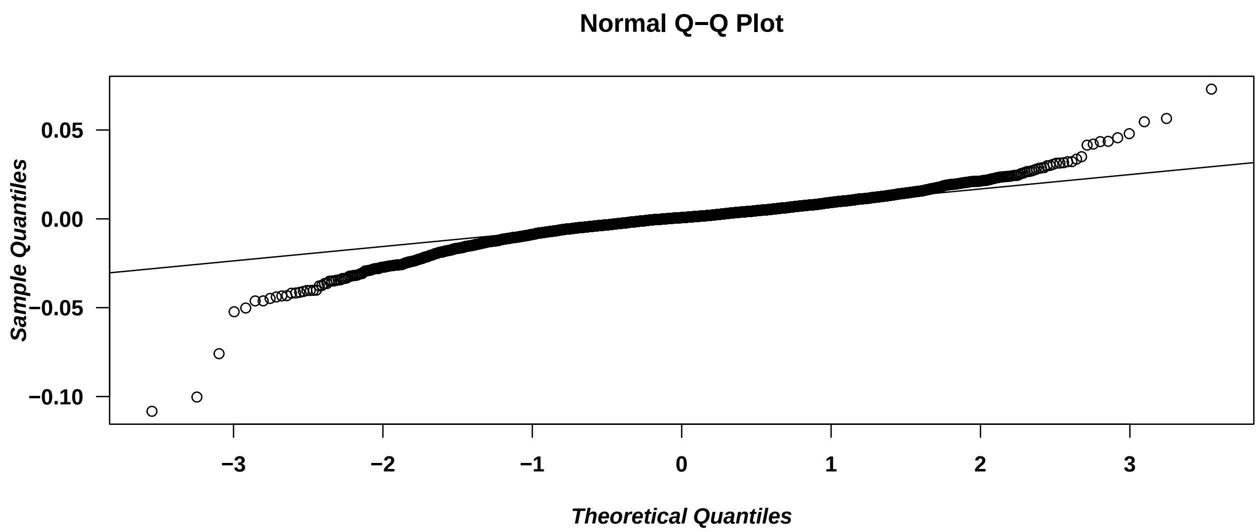
OF THE INVERSE ROOTS TO FALL INSIDE, FOR BOTH THE AR AND THE MA PROCESS? The stationarity condition for the AR-process is satisfied, since the roots have absolute values greater than one. Moreover, the invertibility condition holds for the MA process, since the roots of this process also have absolute values greater than one. Thus, the model is stationary.

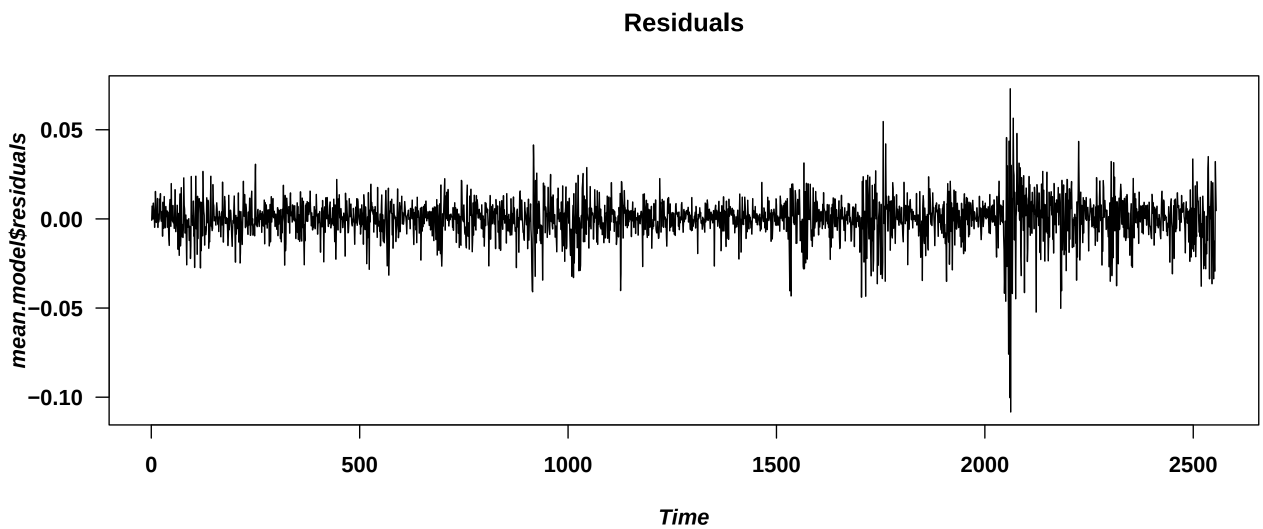
The residuals of the model are analyzed below.



There are no significant coefficients in the autocorrelation function above, which suggests that the model has adequately captured the information in the data. Moreover, the Ljung-Box statistic p-values are all relatively large, which means that we will not reject the Ljung-Box null hypothesis that all m of the correlation coefficients are simultaneously equal to zero. Thus, this further suggests that the residuals are not correlated and we have found a model that seems reasonable in this regard.

Next, a QQ-plot of the residuals, and the residuals themselves (not standardized), is plotted.



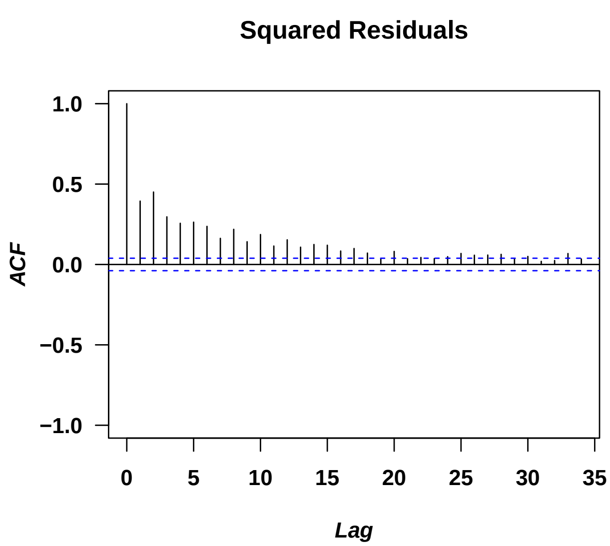


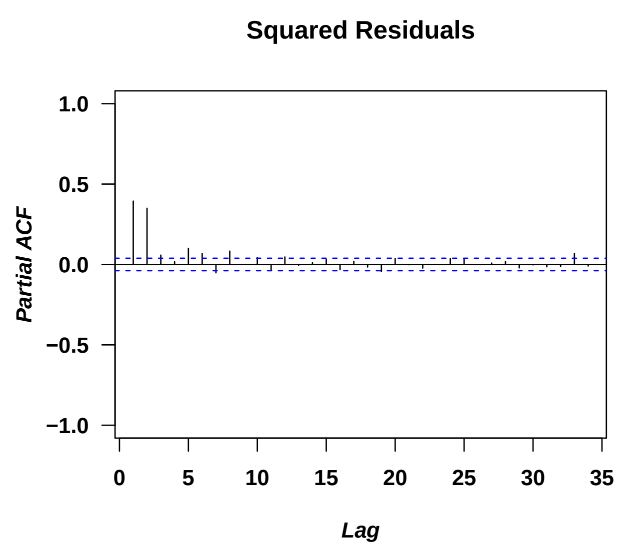

```
#>
#> Title:
#>  Jarque - Bera Normalality Test
#>
#> Test Results:
#>   STATISTIC:
#>     X-squared: 7472.6779
#>   P VALUE:
#>     Asymptotic p Value: < 2.2e-16
#>
#> Description:
#>  Tue Mar 22 18:13:20 2022 by user: ajo
```

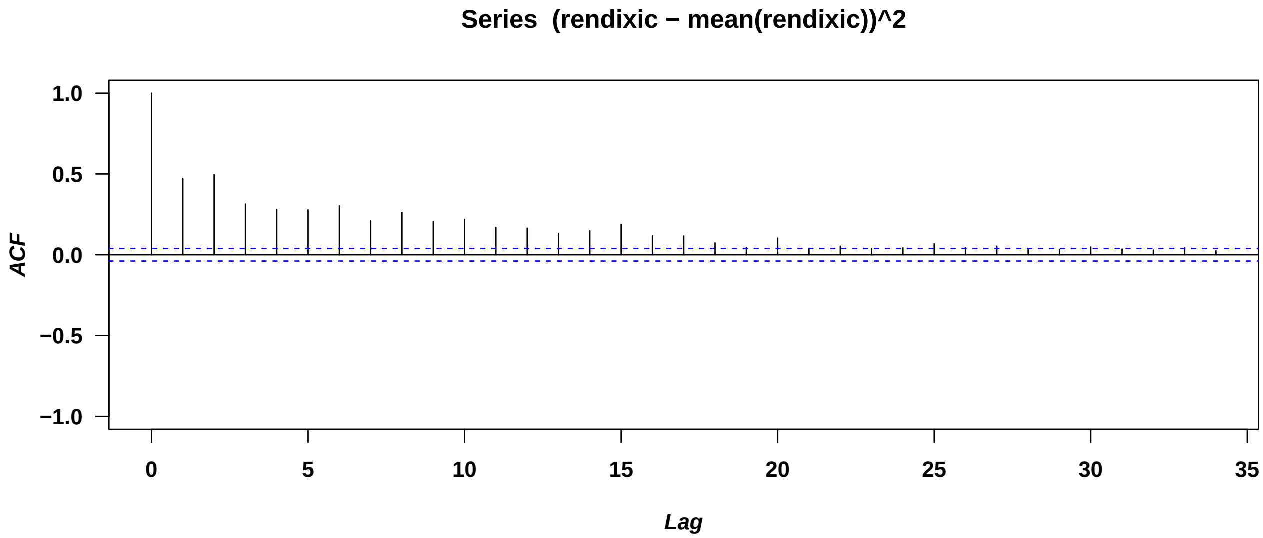
It is apparent that the residuals have heavy tails. It is not reasonable to assume normality of the residuals, an argument that the Jarque-Bera Normality test further substantiates because its null hypothesis of normality is rejected following the very small p-value.

3.5 Identification, Estimation and Diagnostics of a Model for the Variance

First we test for ARCH effects using the residuals of the mean model.








```
#>
#>  Box-Ljung test
#>
#> data:  residuals2
#> X-squared = 397.36, df = 1, p-value < 2.2e-16
```

```
#>
#>  Box-Ljung test
#>
#> data:  residuals2
#> X-squared = 1487.7, df = 5, p-value < 2.2e-16
```

```
#>
#>  Box-Ljung test
#>
#> data:  residuals2
#> X-squared = 2163.8, df = 15, p-value < 2.2e-16
```

As seen in the ACF and PACF of the squared residuals, they are clearly presenting autocorrelation, i.e. there are ARCH effects present. This argument is further substantiated by the Ljung-Box tests above, since they lead to the conclusion that the squared residuals are correlated. Thus, it is relevant to identify and estimate a model for the volatility. Joint estimation of the mean and volatility equations, for different types of models, is done in the following.

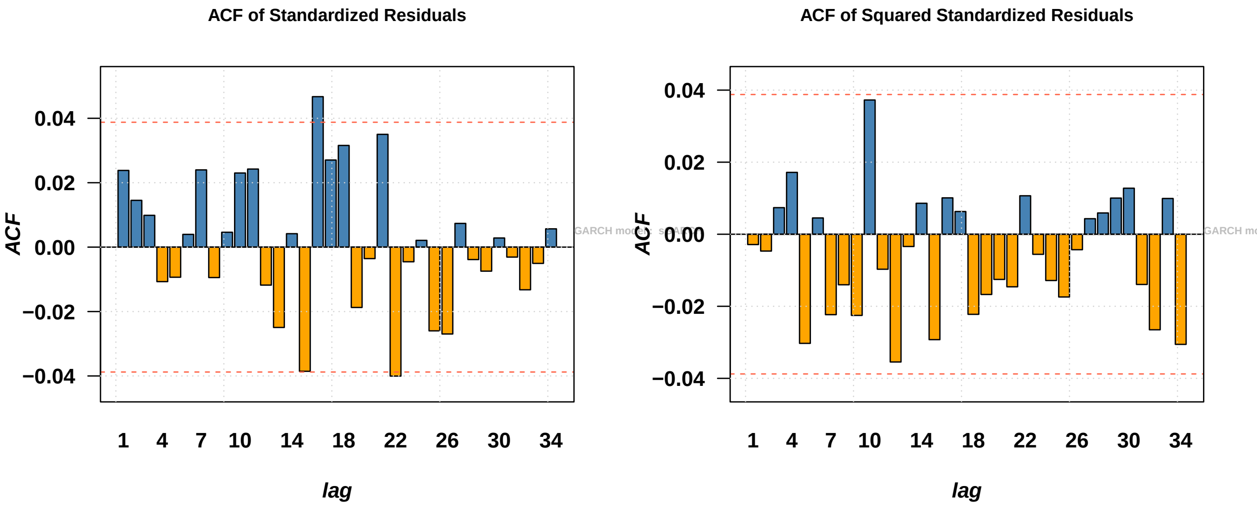
Now over to estimation of GARCH models for the variance of the returns. First we estimate a ARMA(2,2)-GARCH(1,1) with a t-student distribution. WHY CHOSE A T-STUDENT OVER A NORMAL (OR VICE VERSA)?

```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
#>
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model  : sGARCH(1,1)
#> Mean Model   : ARFIMA(2,0,2)
#> Distribution : std
#>
```

```
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.001230    0.000085  14.4538 0.000000
#> ar1     0.411215    0.010896  37.7392 0.000000
#> ar2     0.468396    0.002407 194.5656 0.000000
#> ma1    -0.467394    0.010241 -45.6404 0.000000
#> ma2    -0.455194    0.009398 -48.4364 0.000000
#> omega   0.000005    0.000002   2.2329 0.025555
#> alpha1  0.163393    0.020408   8.0062 0.000000
#> beta1   0.813486    0.023304  34.9081 0.000000
#> shape   5.240364    0.590405   8.8759 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error   t value Pr(>|t|)
#> mu      0.001230    0.000101  12.15400  0.00000
#> ar1     0.411215    0.026259  15.65987  0.00000
#> ar2     0.468396    0.026987  17.35607  0.00000
#> ma1    -0.467394    0.016152 -28.93808  0.00000
#> ma2    -0.455194    0.015014 -30.31821  0.00000
#> omega   0.000005    0.000005   0.95667  0.33873
#> alpha1  0.163393    0.020752   7.87359  0.00000
#> beta1   0.813486    0.039750  20.46509  0.00000
#> shape   5.240364    0.726690   7.21127  0.00000
#>
#> LogLikelihood : 8262.277
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.4605
#> Bayes        -6.4399
#> Shibata      -6.4605
#> Hannan-Quinn -6.4530
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                           statistic p-value
#> Lag[1]                        1.450  0.2286
#> Lag[2*(p+q)+(p+q)-1][11]      3.676  1.0000
#> Lag[4*(p+q)+(p+q)-1][19]      8.810  0.6701
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                     0.02101  0.8847
#> Lag[2*(p+q)+(p+q)-1][5]    0.92104  0.8773
#> Lag[4*(p+q)+(p+q)-1][9]    2.69363  0.8083
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
```

```
#>             Statistic Shape Scale P-Value
#> ARCH Lag[3]    0.1391 0.500 2.000  0.7092
#> ARCH Lag[5]    2.1473 1.440 1.667  0.4396
#> ARCH Lag[7]    3.0338 2.315 1.543  0.5072
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  1.8626
#> Individual Statistics:
#> mu     0.37609
#> ar1    0.06996
#> ar2    0.12329
#> ma1    0.07412
#> ma2    0.11176
#> omega  0.24242
#> alpha1 0.64422
#> beta1  0.45667
#> shape  0.47015
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          2.1 2.32 2.82
#> Individual Statistic:     0.35 0.47 0.75
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value     prob sig
#> Sign Bias           2.3681 0.017953  **
#> Negative Sign Bias  0.1143 0.909045
#> Positive Sign Bias  0.9489 0.342756
#> Joint Effect       15.5615 0.001395 ***
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     66.58    3.371e-07
#> 2    30     89.80    3.904e-08
#> 3    40     96.76    8.200e-07
#> 4    50    118.17    1.208e-07
#>
#>
#> Elapsed time : 0.5604644
```

We observe that all the parameter estimates are significant to a 5% significance level. Moreover, we note that the condition of positivity holds, because $\hat{\alpha}_1 > 0$ and $\hat{\beta}_1 > 0$, where we follow the standard statistical notation of a hat indicating an estimate. Also, we note that the condition of stationarity holds, because $\hat{\alpha}_1 + \hat{\beta}_1 < 1$. We record the AIC of this first model in order to compare to other models later. The residual plots below show that the residuals do not present any autocorrelation (before moving to around 15 lags, which is a large number of lags), which indicates that this model has modeled the data in a sufficient or reasonable way. However, note that the p-values of the `Weighted Ljung-Box Test on Residuals` are large, which means we cannot reject the null hypothesis of no serial correlation for the different lags. We will keep this is mind as a downfall of this first proposed model.



Next we will fit a ARMA(2,2)-GJR-GARCH(1,1) model, assuming a t-distribution.

```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
#>
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model  : gjrGARCH(1,1)
#> Mean Model   : ARFIMA(2,0,2)
#> Distribution : std
#>
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error   t value Pr(>|t|)
#> mu      0.001047    0.000138  7.606420 0.000000
#> ar1     0.275227    0.319606  0.861146 0.389157
#> ar2     0.470193    0.205718  2.285615 0.022277
#> ma1    -0.320370    0.321325 -0.997029 0.318750
#> ma2    -0.455726    0.214102 -2.128549 0.033292
#> omega   0.000005    0.000000 15.301373 0.000000
#> alpha1  0.000000    0.009195  0.000002 0.999999
#> beta1   0.823974    0.014252 57.815933 0.000000
#> gamma1  0.260477    0.031890  8.168122 0.000000
#> shape   5.594916    0.591457  9.459543 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error   t value Pr(>|t|)
#> mu      0.001047    0.000130  8.037487 0.000000
#> ar1     0.275227    0.196628  1.399736 0.161593
#> ar2     0.470193    0.119749  3.926475 0.000086
#> ma1    -0.320370    0.198050 -1.617626 0.105743
#> ma2    -0.455726    0.122730 -3.713240 0.000205
#> omega   0.000005    0.000001  9.596872 0.000000
#> alpha1  0.000000    0.012423  0.000001 0.999999
#> beta1   0.823974    0.013038 63.195840 0.000000
#> gamma1  0.260477    0.035268  7.385755 0.000000
```

```
#> shape   5.594916    0.550880 10.156318 0.000000
#>
#> LogLikelihood : 8303.976
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.4923
#> Bayes        -6.4695
#> Shibata      -6.4924
#> Hannan-Quinn -6.4841
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                      0.2822  0.5953
#> Lag[2*(p+q)+(p+q)-1][11]    3.0139  1.0000
#> Lag[4*(p+q)+(p+q)-1][19]    8.8996  0.6552
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
#> ------------------------------------
#>                         statistic p-value
#> Lag[1]                     0.2543  0.6141
#> Lag[2*(p+q)+(p+q)-1][5]    0.5938  0.9423
#> Lag[4*(p+q)+(p+q)-1][9]    1.6828  0.9393
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
#>             Statistic Shape Scale P-Value
#> ARCH Lag[3]   0.06952 0.500 2.000  0.7920
#> ARCH Lag[5]   0.85359 1.440 1.667  0.7768
#> ARCH Lag[7]   1.59629 2.315 1.543  0.8019
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  17.2754
#> Individual Statistics:
#> mu     0.7414
#> ar1    0.1534
#> ar2    0.3645
#> ma1    0.1807
#> ma2    0.3876
#> omega  1.3411
#> alpha1 2.3979
#> beta1  0.9204
#> gamma1 0.8213
#> shape  0.3773
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          2.29 2.54 3.05
#> Individual Statistic:     0.35 0.47 0.75
```

```
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value   prob sig
#> Sign Bias           1.2277 0.2197
#> Negative Sign Bias  1.2552 0.2095
#> Positive Sign Bias  0.2208 0.8253
#> Joint Effect        2.7117 0.4383
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     70.26    8.326e-08
#> 2    30     85.76    1.611e-07
#> 3    40    109.44    1.324e-08
#> 4    50    129.36    3.569e-09
#>
#>
#> Elapsed time : 0.8586736
```

The stationarity conditions holds, since $\hat{\alpha}_1 + \hat{\beta}_1 + \frac{1}{2}\hat{\gamma} \approx 0.954 < 1$. However, the positivity condition does not seem to hold, since $\hat{\alpha} = 0 \ngtr 0$. Thus, the GJR-GARCH based model will not be used, even though the AIC is lower and the residuals do not present any autocorrelation according to plots. Moreover, the several of the ARMA-parameter coefficient estimates are not significant to a reasonable level, which is the case no matter what `armaOrder` is use in the estimation.

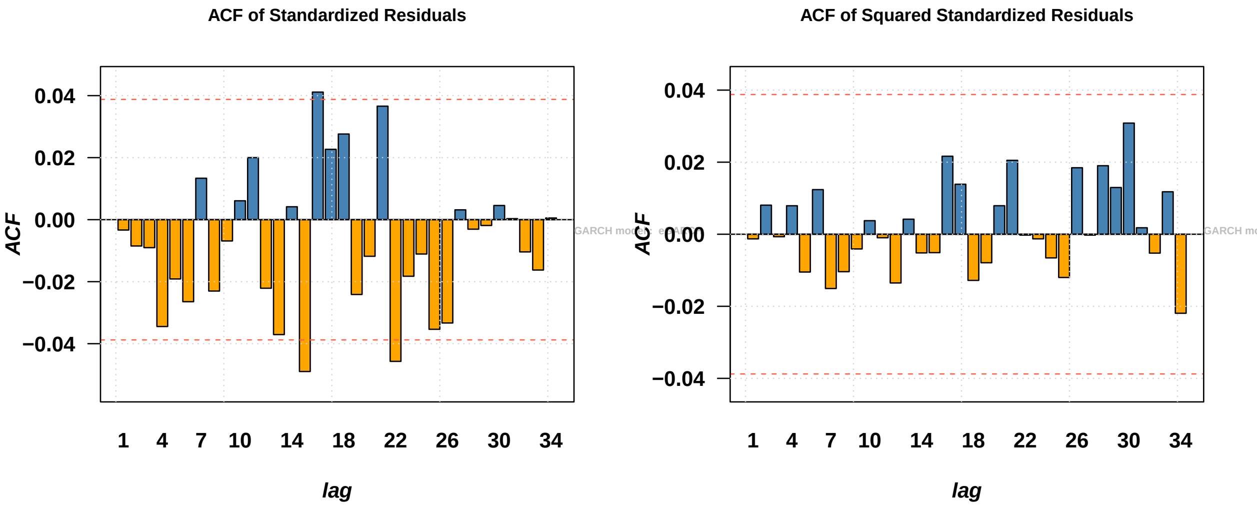
Next, we will fit an ARMA(2,2)-EGARCH model.

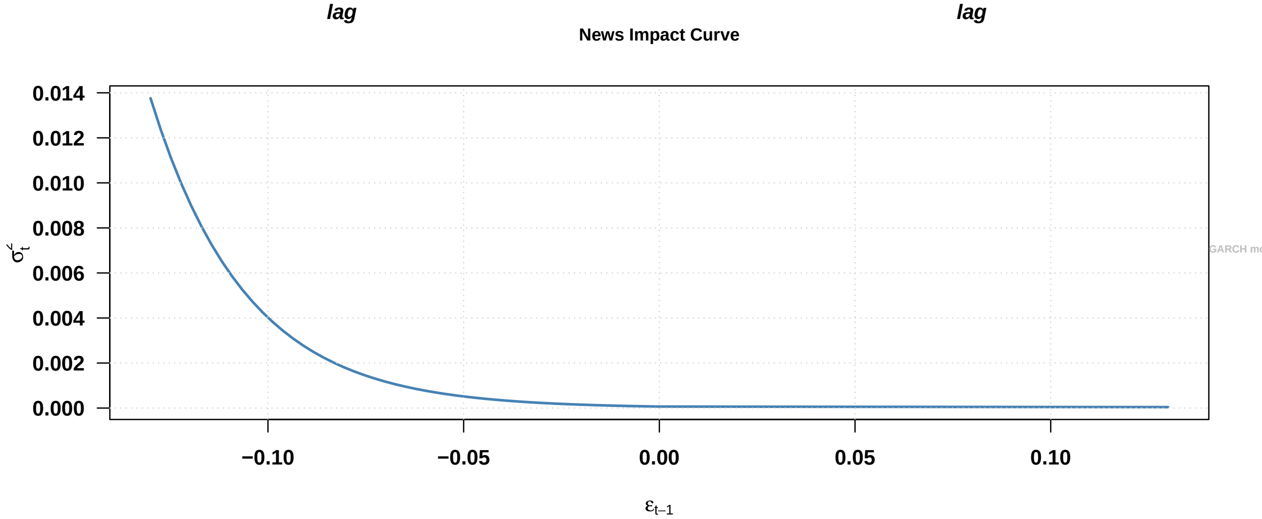
```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
#>
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model  : eGARCH(1,1)
#> Mean Model   : ARFIMA(2,0,2)
#> Distribution : std
#>
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.000945    0.000155   6.1066 0.000000
#> ar1     0.137188    0.047974   2.8596 0.004241
#> ar2     0.310470    0.022509  13.7932 0.000000
#> ma1    -0.178777    0.045824  -3.9014 0.000096
#> ma2    -0.292323    0.022317 -13.0986 0.000000
#> omega  -0.389642    0.013992 -27.8483 0.000000
#> alpha1 -0.193182    0.019053 -10.1394 0.000000
#> beta1   0.958980    0.001550 618.8427 0.000000
#> gamma1  0.161978    0.021992   7.3654 0.000000
#> shape   5.708617    0.654727   8.7191 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error  t value Pr(>|t|)
```

```
#> mu      0.000945    0.000154   6.1427        0
#> ar1     0.137188    0.006823  20.1069        0
#> ar2     0.310470    0.010004  31.0345        0
#> ma1    -0.178777    0.016029 -11.1534        0
#> ma2    -0.292323    0.009662 -30.2545        0
#> omega  -0.389642    0.008250 -47.2320        0
#> alpha1 -0.193182    0.024672  -7.8300        0
#> beta1   0.958980    0.001166 822.5532        0
#> gamma1  0.161978    0.025847   6.2667        0
#> shape   5.708617    0.685084   8.3327        0
#>
#> LogLikelihood : 8308.361
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.4958
#> Bayes        -6.4729
#> Shibata      -6.4958
#> Hannan-Quinn -6.4875
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                      0.0286  0.8657
#> Lag[2*(p+q)+(p+q)-1][11]    4.9973  0.9583
#> Lag[4*(p+q)+(p+q)-1][19]   11.6665  0.2290
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
#> ------------------------------------
#>                         statistic p-value
#> Lag[1]                   0.004198  0.9483
#> Lag[2*(p+q)+(p+q)-1][5]  0.258762  0.9877
#> Lag[4*(p+q)+(p+q)-1][9]  0.851583  0.9916
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
#>             Statistic Shape Scale P-Value
#> ARCH Lag[3]  0.001218 0.500 2.000  0.9722
#> ARCH Lag[5]  0.297744 1.440 1.667  0.9409
#> ARCH Lag[7]  0.811170 2.315 1.543  0.9422
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  4.2243
#> Individual Statistics:
#> mu     1.04997
#> ar1    0.07622
#> ar2    0.30588
#> ma1    0.08127
#> ma2    0.32174
```

```
#> omega  1.37628
#> alpha1 0.12155
#> beta1  1.23031
#> gamma1 0.94587
#> shape  0.16952
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          2.29 2.54 3.05
#> Individual Statistic:     0.35 0.47 0.75
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value   prob sig
#> Sign Bias          0.32686 0.7438
#> Negative Sign Bias 0.48021 0.6311
#> Positive Sign Bias 0.06459 0.9485
#> Joint Effect       0.24566 0.9699
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     78.12    3.914e-09
#> 2    30     97.88    2.130e-09
#> 3    40    100.98    2.135e-07
#> 4    50    118.33    1.151e-07
#>
#>
#> Elapsed time : 0.5763299
```

All estimated parameters are significant to a level of $\alpha = 0.05$. For this model, we do not require positivity of the GARCH parameter estimates. WHAT ABOUT STATIONARITY, DO WE REQUIRE THIS? The residual plots below do not show any autocorrelation before moving to a large number of lags, which is a good sign that the model has modelized the data adequately. Moreover, the AIC is smaller for this model compared to the first proposed model. Thus, I would prefer this model over the first model. Note that $\gamma > 0$, which should mean that the positive news have a larger effect on the news compared to the negative news, which does not make sense here, when looking at the news impact curve. THIS IS STRANGE, I DO NOT UNDERSTAND THIS!? ASK PROFE! DETTE ER MOTSATT DEFINERT I R VIRKER DET SOM! EN POSITIV GAMMA ER DET SAMME SOM EN NEGATIV I LIGNINGENE, I.E. EN POSITIV GAMMA GIR STØRRE EFFEKT FOR DE NEGATIVE NYHETENE ENN DE POSITIVE, SOM VI SER AV PLOTTET!





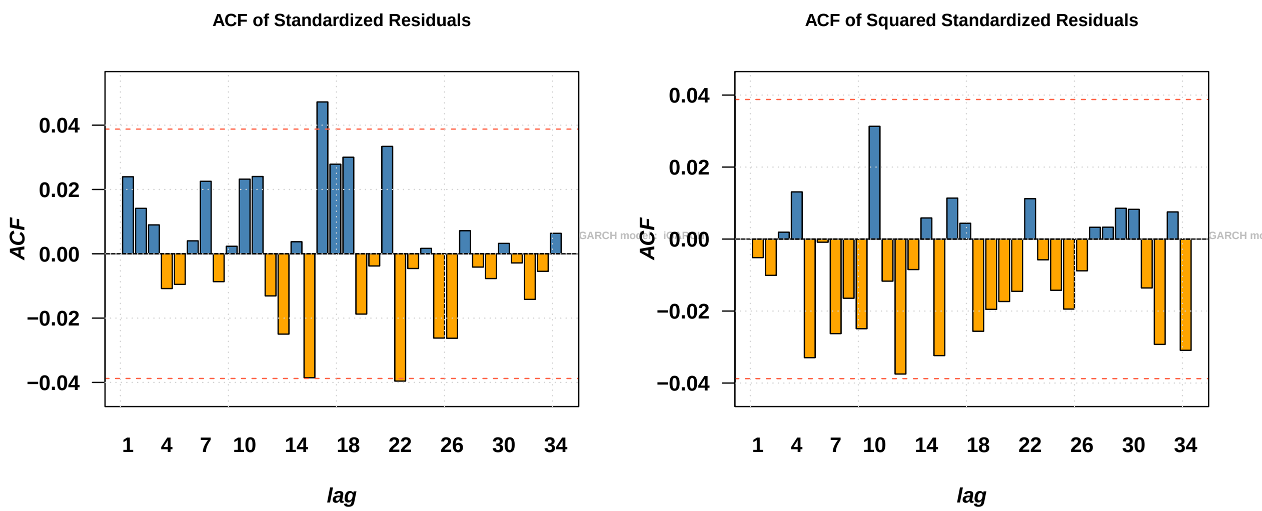
Next, we fit and ARMA(2,2)-IGARCH model.

```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
#>
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model   : iGARCH(1,1)
#> Mean Model    : ARFIMA(2,0,2)
#> Distribution : std
#>
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.001238    0.000084  14.7029 0.000000
#> ar1     0.411677    0.010557  38.9943 0.000000
#> ar2     0.467050    0.002286 204.2796 0.000000
#> ma1    -0.468876    0.010466 -44.8003 0.000000
#> ma2    -0.453161    0.009587 -47.2689 0.000000
#> omega   0.000004    0.000002   2.2337 0.025503
#> alpha1  0.181978    0.024164   7.5310 0.000000
```

```
#> beta1   0.818022          NA       NA       NA
#> shape   4.750542    0.428166  11.0951 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error   t value Pr(>|t|)
#> mu      0.001238    0.000107  11.53516 0.000000
#> ar1     0.411677    0.027459  14.99226 0.000000
#> ar2     0.467050    0.028844  16.19242 0.000000
#> ma1    -0.468876    0.017833 -26.29308 0.000000
#> ma2    -0.453161    0.016554 -27.37489 0.000000
#> omega   0.000004    0.000004   0.94899 0.342627
#> alpha1  0.181978    0.043129   4.21934 0.000025
#> beta1   0.818022          NA        NA       NA
#> shape   4.750542    0.511909   9.28005 0.000000
#>
#> LogLikelihood : 8260.896
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.4602
#> Bayes        -6.4419
#> Shibata      -6.4602
#> Hannan-Quinn -6.4536
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                       1.467  0.2258
#> Lag[2*(p+q)+(p+q)-1][11]     3.540  1.0000
#> Lag[4*(p+q)+(p+q)-1][19]     8.661  0.6944
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
#> ------------------------------------
#>                         statistic p-value
#> Lag[1]                    0.06793  0.7944
#> Lag[2*(p+q)+(p+q)-1][5]   1.01394  0.8565
#> Lag[4*(p+q)+(p+q)-1][9]   3.06639  0.7480
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
#>              Statistic Shape Scale P-Value
#> ARCH Lag[3]  0.009302 0.500 2.000  0.9232
#> ARCH Lag[5]  2.024476 1.440 1.667  0.4659
#> ARCH Lag[7]  3.123161 2.315 1.543  0.4906
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  4.8042
#> Individual Statistics:
#> mu     0.36519
```

```
#> ar1    0.07278
#> ar2    0.12984
#> ma1    0.07754
#> ma2    0.11797
#> omega  1.56210
#> alpha1 0.13294
#> shape  0.42921
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          1.89 2.11 2.59
#> Individual Statistic:     0.35 0.47 0.75
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value     prob sig
#> Sign Bias           2.3282 0.019982  **
#> Negative Sign Bias  0.4695 0.638737
#> Positive Sign Bias  1.2485 0.211975
#> Joint Effect       15.9491 0.001162 ***
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     64.44    7.537e-07
#> 2    30     85.41    1.820e-07
#> 3    40     98.48    4.757e-07
#> 4    50    121.14    4.823e-08
#>
#>
#> Elapsed time : 0.4661305
```

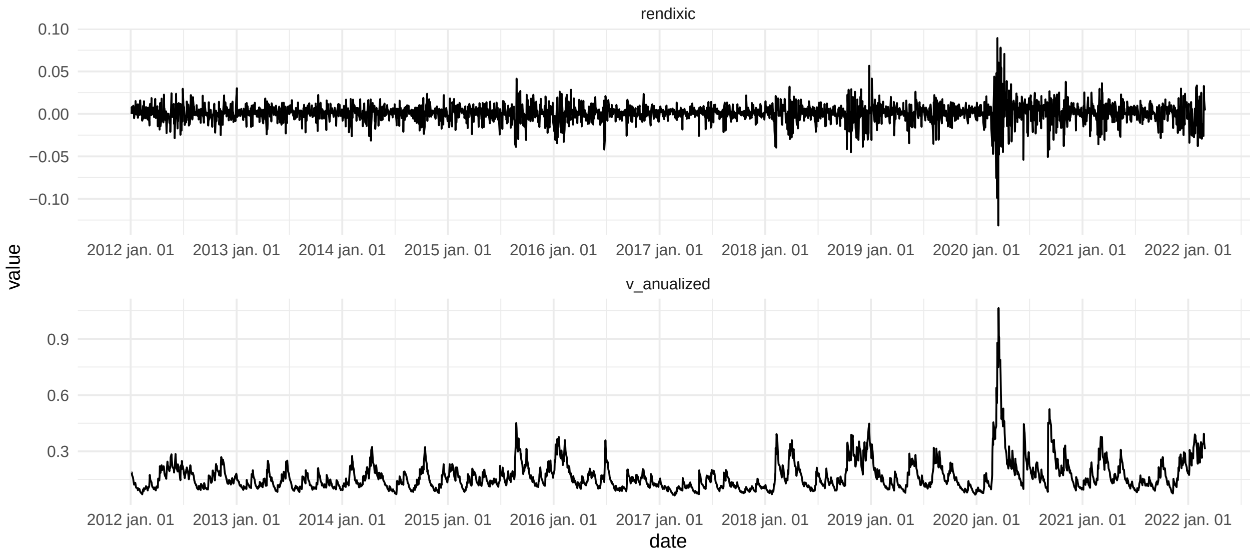
The residuals for the iGARCH look alright, but the AIC is lower for the EGARCH based model.



Thus, I would conclude that the best model out of the three fitted is the EGARCH based model.

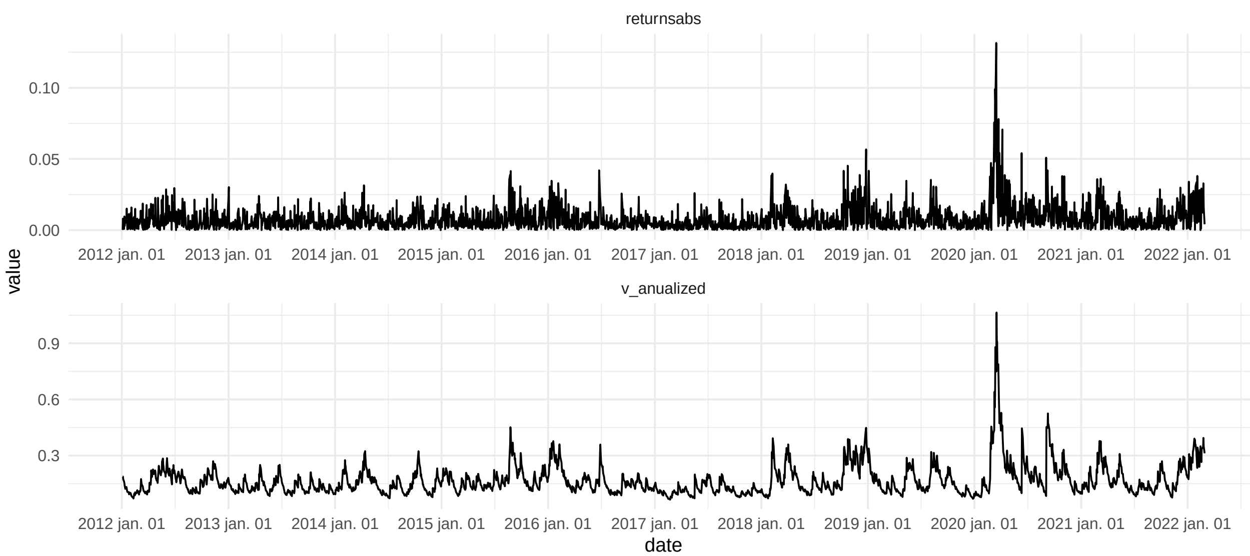
4 Grafic and Interpretation of the Estimated Series of Volatility

The estimated (anualized) series of volatility for the ARMA(2,2)-EGARCH model is shown below, together with the returns.

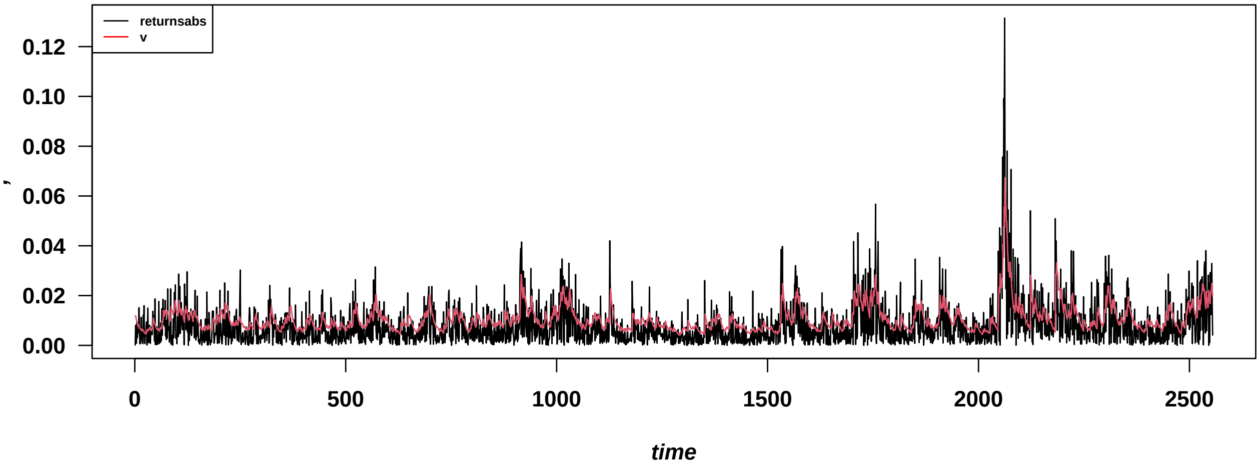


The general behaviour seems to match relatively well, i.e. the movements in the two plots coincide relatively well. Days with larger (absolute) returns go together with the days with larger estimated volatilities.

A comparison of the estimated volatilities and the absolute value of the returns is given below.



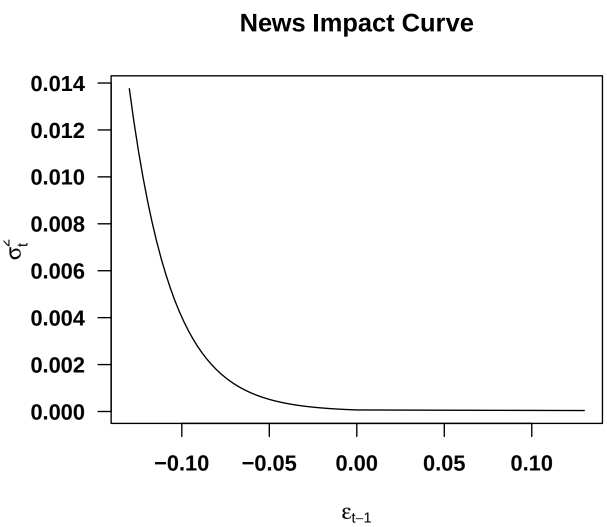
They all seem to follow a similar pattern, which is a good sign, even though we cannot compare the absolute values of the data shown in the plots.

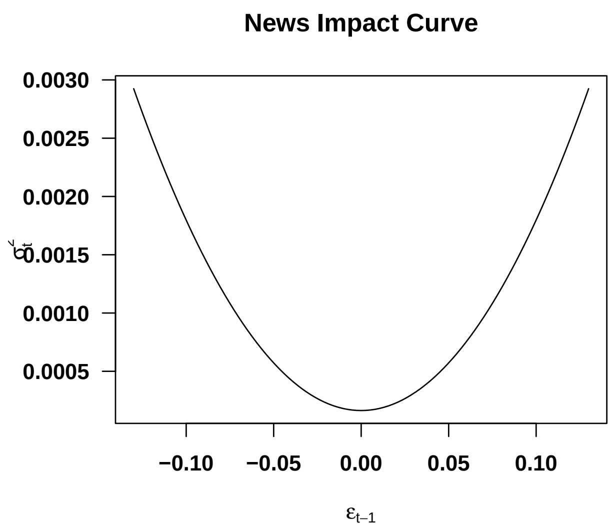


THESE ARE ALL JUST DIFFERENT WAYS OF SEEING THE SAME RESULTS, REMOVE SOME OF THEM LATER!

5 Grafic and Interpretation of the News Impact Curve

The news impact curve for our chosen model is shown below. Since the standard GARCH model does not take the leverage effect into effect, the news impact curve is symmetric.





The EGARCH based model takes the leverage effect into effect, which is the news impact curve is non-symmetric. This curve indicates that the volatility is impacted to a higher degree by negative news compared to the impact on the volatility following positive news, which seems to be increasingly small as the positivity of the news increases THIS IS VERY STRANGE!!

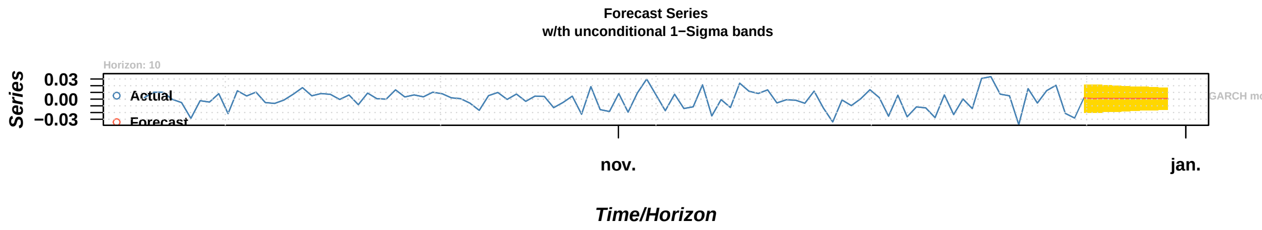
6 Volatility Predictions and Interpretations

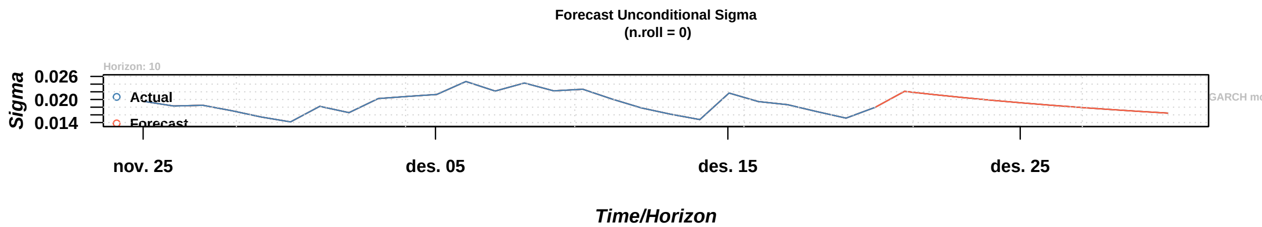
Below volatility are predicted while leaving out the last 10 observations when estimating the ARMA(2,2)-EGARCH model. The prediction is done 10 steps ahead into the future, first statically.

```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
```

```
#>
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model  : eGARCH(1,1)
#> Mean Model   : ARFIMA(2,0,2)
#> Distribution : std
#>
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.000948    0.000148   6.3869 0.000000
#> ar1     0.155652    0.051076   3.0474 0.002308
#> ar2     0.307442    0.027053  11.3644 0.000000
#> ma1    -0.198667    0.050248  -3.9537 0.000077
#> ma2    -0.286833    0.027142 -10.5680 0.000000
#> omega  -0.391038    0.013548 -28.8640 0.000000
#> alpha1 -0.193163    0.018411 -10.4920 0.000000
#> beta1   0.958837    0.001504 637.3978 0.000000
#> gamma1  0.162456    0.021326   7.6178 0.000000
#> shape   5.672263    0.628296   9.0280 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error   t value Pr(>|t|)
#> mu      0.000948    0.000147    6.4711        0
#> ar1     0.155652    0.006593   23.6086        0
#> ar2     0.307442    0.009750   31.5322        0
#> ma1    -0.198667    0.011283  -17.6078        0
#> ma2    -0.286833    0.010276  -27.9119        0
#> omega  -0.391038    0.007671  -50.9775        0
#> alpha1 -0.193163    0.022357   -8.6401        0
#> beta1   0.958837    0.000907 1057.4728        0
#> gamma1  0.162456    0.023734    6.8448        0
#> shape   5.672263    0.663446    8.5497        0
#>
#> LogLikelihood : 8283.935
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.5021
#> Bayes        -6.4792
#> Shibata      -6.5021
#> Hannan-Quinn -6.4938
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                         statistic p-value
#> Lag[1]                    0.02091  0.8850
#> Lag[2*(p+q)+(p+q)-1][11]   4.95941  0.9643
#> Lag[4*(p+q)+(p+q)-1][19]  11.45257  0.2544
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
```

```
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                    0.003375  0.9537
#> Lag[2*(p+q)+(p+q)-1][5]   0.248025  0.9887
#> Lag[4*(p+q)+(p+q)-1][9]   0.833143  0.9921
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
#>             Statistic Shape Scale P-Value
#> ARCH Lag[3] 0.0007562 0.500 2.000  0.9781
#> ARCH Lag[5] 0.2874882 1.440 1.667  0.9436
#> ARCH Lag[7] 0.7940757 2.315 1.543  0.9446
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  4.2827
#> Individual Statistics:
#> mu     1.05339
#> ar1    0.07059
#> ar2    0.34955
#> ma1    0.07642
#> ma2    0.36645
#> omega  1.36438
#> alpha1 0.12781
#> beta1  1.22056
#> gamma1 0.96366
#> shape  0.15798
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          2.29 2.54 3.05
#> Individual Statistic:     0.35 0.47 0.75
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value   prob sig
#> Sign Bias          0.29443 0.7685
#> Negative Sign Bias 0.43279 0.6652
#> Positive Sign Bias 0.09461 0.9246
#> Joint Effect       0.20103 0.9774
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     78.06    3.999e-09
#> 2    30     95.65    4.802e-09
#> 3    40    103.25    1.024e-07
#> 4    50    116.83    1.819e-07
#>
#>
#> Elapsed time : 0.572068
```

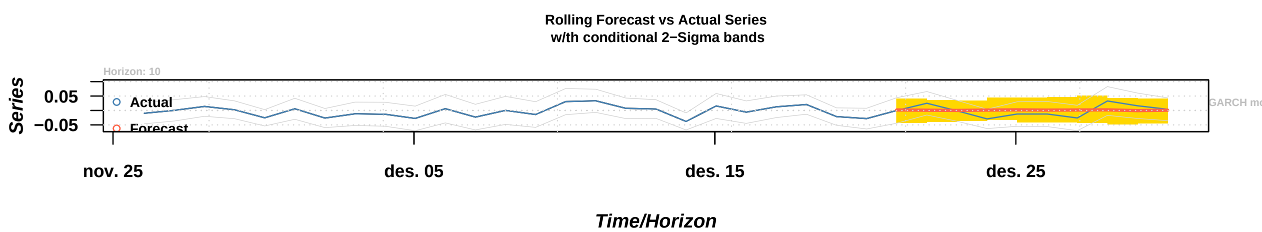


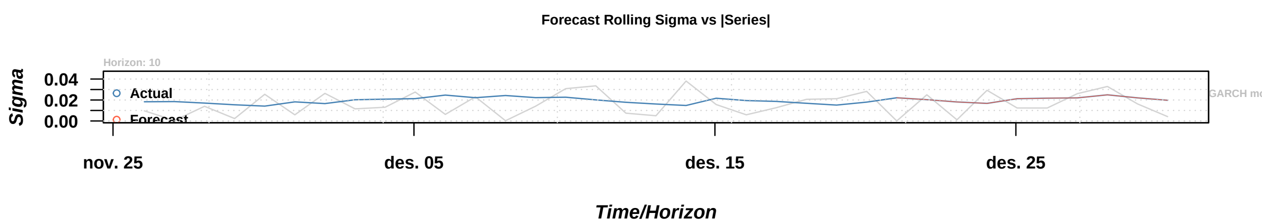



```
#> [1] 0.008652676
```

As we can see from the uppermost plot, these static predictions (for the mean) 10 steps ahead are relatively useless. ALSO LOOKS LIKE THE FORECAST OF THE VARIANCE WILL GO TOWARDS THE UNCONDITIONAL VARIANCE. NOT SURE IN WHAT SITUATIONS THIS SHOULD HAPPEN THOUGH, ASK PROFE PERHAPS!

Next, let us predict 10 steps into the future with a rolling window. We reestimate the model at each time step and estimate one step into the future after each reestimation. After doing this 10 times, we have effectively predicted 10 days into the future.





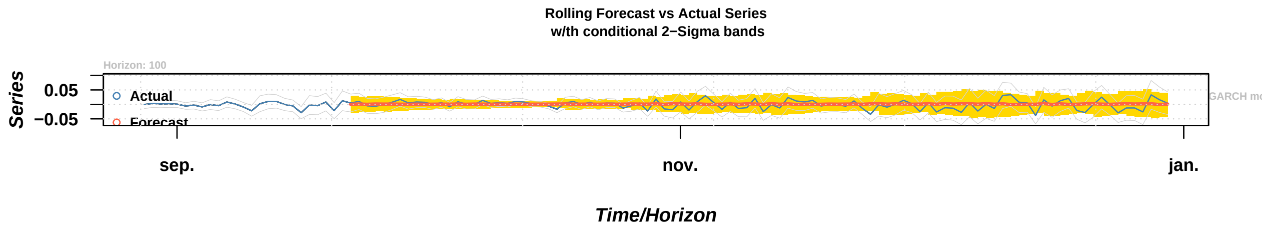
The predictions are still lousy, as can be seen from the predictions of the mean in the uppermost plot. However, from the second plot, it looks like the predictions of the variance are somewhat following the same movements as the absolute value of the series; when the absolute value of the series hits a spike, the predictions of the volatility increase as well.

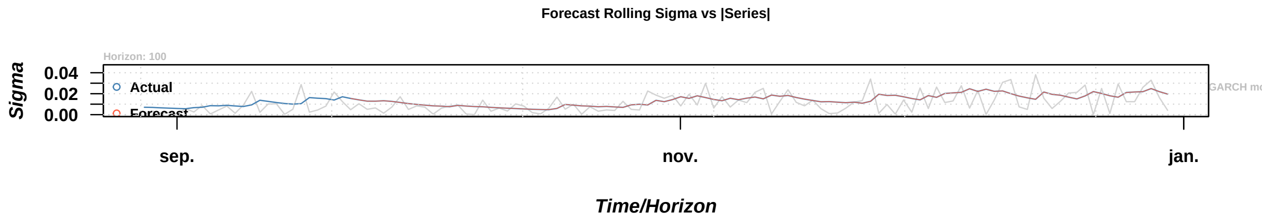
The same type of movement can be seen when predicting with a rolling window 100 steps into the future, where the results of this prediction is shown below.

```
#>
#> *---------------------------------*
#> *          GARCH Model Fit        *
#> *---------------------------------*
#>
```

```
#> Conditional Variance Dynamics
#> -----------------------------------
#> GARCH Model  : eGARCH(1,1)
#> Mean Model   : ARFIMA(2,0,2)
#> Distribution : std
#>
#> Optimal Parameters
#> ------------------------------------
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.000962    0.000160   6.0272 0.000000
#> ar1     0.048676    0.031073   1.5665 0.117230
#> ar2     0.354829    0.052201   6.7973 0.000000
#> ma1    -0.093784    0.034073  -2.7524 0.005916
#> ma2    -0.340675    0.051374  -6.6313 0.000000
#> omega  -0.406135    0.025430 -15.9704 0.000000
#> alpha1 -0.194287    0.015071 -12.8915 0.000000
#> beta1   0.957333    0.003192 299.9545 0.000000
#> gamma1  0.163769    0.044924   3.6455 0.000267
#> shape   5.522329    0.611122   9.0364 0.000000
#>
#> Robust Standard Errors:
#>         Estimate  Std. Error  t value Pr(>|t|)
#> mu      0.000962    0.000168   5.7166 0.000000
#> ar1     0.048676    0.034059   1.4292 0.152949
#> ar2     0.354829    0.032361  10.9648 0.000000
#> ma1    -0.093784    0.024981  -3.7542 0.000174
#> ma2    -0.340675    0.014981 -22.7404 0.000000
#> omega  -0.406135    0.079000  -5.1410 0.000000
#> alpha1 -0.194287    0.057930  -3.3538 0.000797
#> beta1   0.957333    0.009373 102.1392 0.000000
#> gamma1  0.163769    0.124739   1.3129 0.189220
#> shape   5.522329    0.690047   8.0028 0.000000
#>
#> LogLikelihood : 8025.953
#>
#> Information Criteria
#> ------------------------------------
#>
#> Akaike       -6.5303
#> Bayes        -6.5067
#> Shibata      -6.5303
#> Hannan-Quinn -6.5217
#>
#> Weighted Ljung-Box Test on Standardized Residuals
#> ------------------------------------
#>                          statistic p-value
#> Lag[1]                      0.0455  0.8311
#> Lag[2*(p+q)+(p+q)-1][11]    4.6141  0.9932
#> Lag[4*(p+q)+(p+q)-1][19]   11.5902  0.2379
#> d.o.f=4
#> H0 : No serial correlation
#>
#> Weighted Ljung-Box Test on Standardized Squared Residuals
#> ------------------------------------
```

```
#>                          statistic p-value
#> Lag[1]                   3.056e-05  0.9956
#> Lag[2*(p+q)+(p+q)-1][5] 2.679e-01  0.9869
#> Lag[4*(p+q)+(p+q)-1][9] 8.595e-01  0.9913
#> d.o.f=2
#>
#> Weighted ARCH LM Tests
#> ------------------------------------
#>             Statistic Shape Scale P-Value
#> ARCH Lag[3] 0.0002621 0.500 2.000  0.9871
#> ARCH Lag[5] 0.2775122 1.440 1.667  0.9462
#> ARCH Lag[7] 0.7881588 2.315 1.543  0.9454
#>
#> Nyblom stability test
#> ------------------------------------
#> Joint Statistic:  4.1736
#> Individual Statistics:
#> mu     0.91848
#> ar1    0.03933
#> ar2    0.28956
#> ma1    0.04143
#> ma2    0.30905
#> omega  1.15684
#> alpha1 0.15415
#> beta1  1.02769
#> gamma1 1.13095
#> shape  0.10918
#>
#> Asymptotic Critical Values (10% 5% 1%)
#> Joint Statistic:          2.29 2.54 3.05
#> Individual Statistic:     0.35 0.47 0.75
#>
#> Sign Bias Test
#> ------------------------------------
#>                    t-value   prob sig
#> Sign Bias           0.3186 0.7500
#> Negative Sign Bias  0.3536 0.7237
#> Positive Sign Bias  0.1932 0.8468
#> Joint Effect        0.1696 0.9823
#>
#>
#> Adjusted Pearson Goodness-of-Fit Test:
#> ------------------------------------
#>   group statistic p-value(g-1)
#> 1    20     76.77    6.652e-09
#> 2    30     90.23    3.350e-08
#> 3    40     96.02    1.033e-06
#> 4    50    112.03    7.687e-07
#>
#>
#> Elapsed time : 0.5820739
```

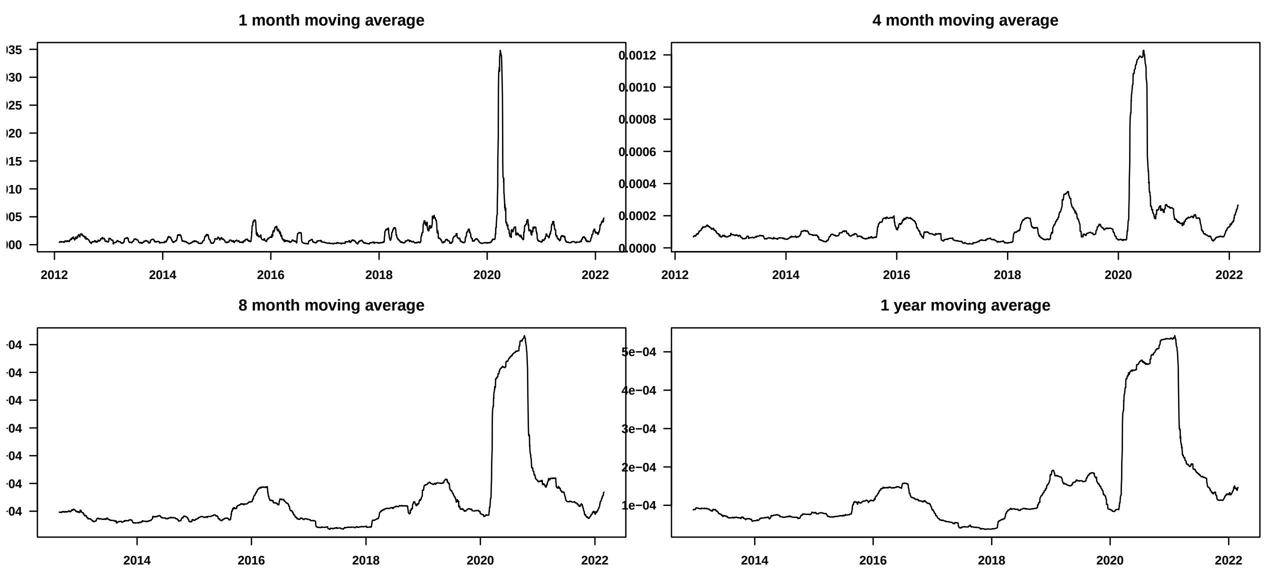




7 Calculations via Historical Volatility and EWMA

Historical volatility is calculated below. The historical volatility has been calculated using Simple Moving Average (SMA) over different time periods

$$\sigma_t^2 = \frac{1}{k}\sum_{i=1}^{k} r_{t-1}^2.$$

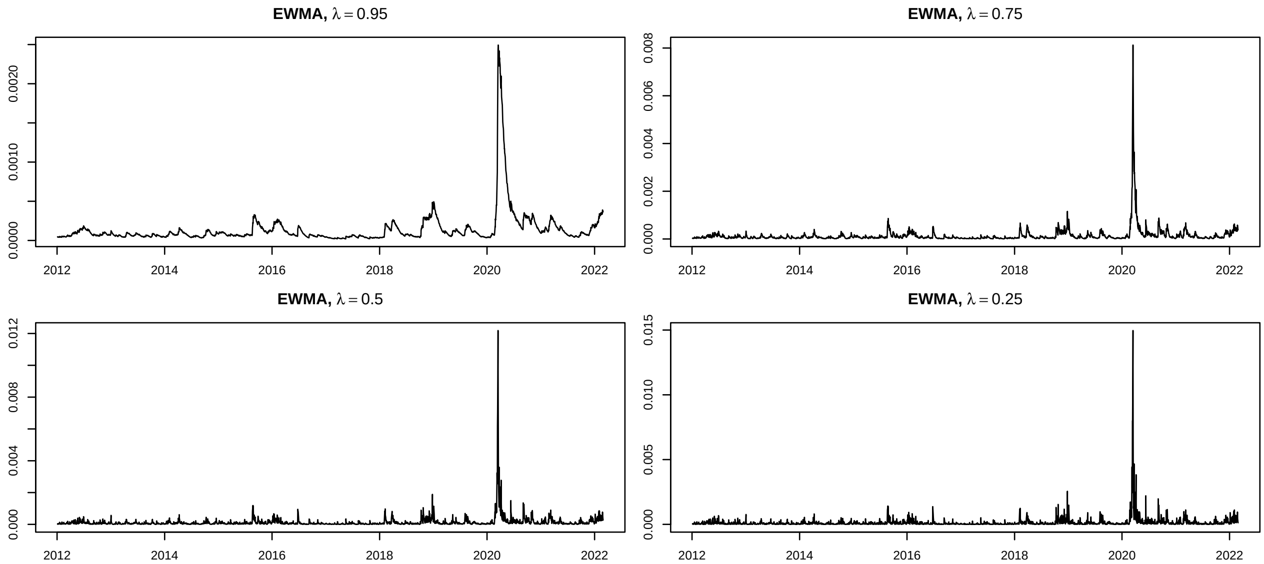


As is apparent from the plots, the volatility pattern is highly dependent on the k, i.e. the number of observations used to calculate the moving average. Moreover, we can see that the results are greatly affected by extreme values, especially when k is small, which is clearly seen in the results for the 1 month moving average. The volatility pattern is smoother when k is larger. Which of these values for k gives the "best" results? This is difficult to answer.

The Exponentially Weighted Moving Average (EWMA) model is used to calculate volatility

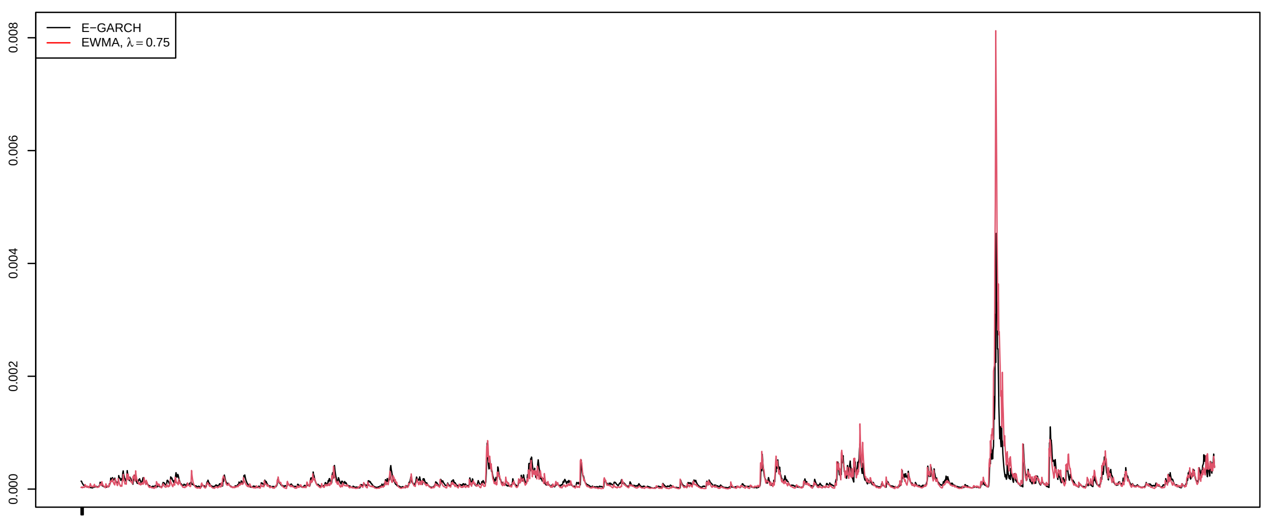
$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2 = (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}r_{t-1}^2, \quad 0 < \lambda < 1.$$

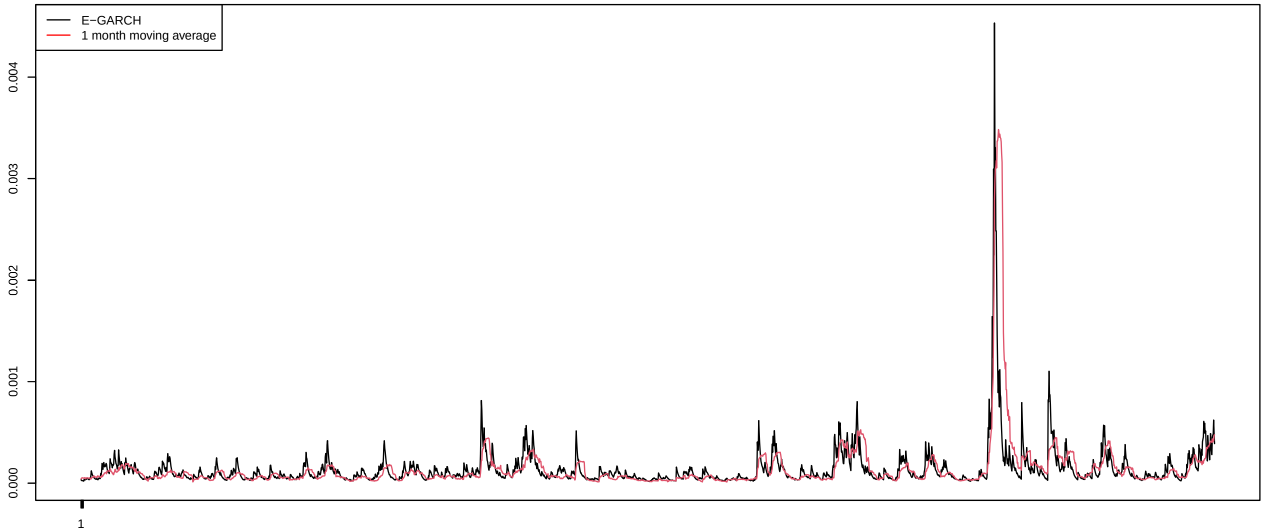
Different values of the parameter λ are used in order to see how the results depend on it. From the theoretical point of view, we know that the term $(1-\lambda)r_{t-1}^2$ determines the reaction of volatility to market events, i.e. the larger the term $(1-\lambda)$ the larger the reaction in the volatility stemming from yesterday's return. Moreover, the term $\lambda\sigma_{t-1}^2$ determines the persistence in volatility. In other terms, it decides how much of yesterday's volatility is allowed to persist to today's volatility: A larger value of λ gives larger persistence. Thus, the EWMA model gives a trade-off between persistence and reaction in the volatility, depending on the value of λ.



As we can see from the plots above, the larger values of λ give smoother plots, since the persistence is larger, while the smaller values of λ give a more reactive or non-smooth volatility pattern, since the persistence of the volatility is much lower in these cases. Comparing to the results obtained when using the historical volatility, all the volatility patterns obtained with EWMA are more non-smooth than the former, being most similar to the 1 month moving average. Note also that the choice of λ seems somewhat arbitrary in this case (similar to the choice of k for historical volatility), as it is difficult to be certain about the best choice of the parameter.

Doing a quick comparison between these two models and the results from the EGARCH model, it looks like the EWMA model with $\lambda = 0.75$ gives a relatively similar volatility pattern, whereas the 1 month moving average (which is the one among the four models that is most similar to the results from EGARCH) is lacking in comparison.





NOTE THAT THE TIMES ON THE X-AXIS ARE FUCKED UP. TRY TO FIX THIS LATER!

8 Calculation and Interpretation of VaR

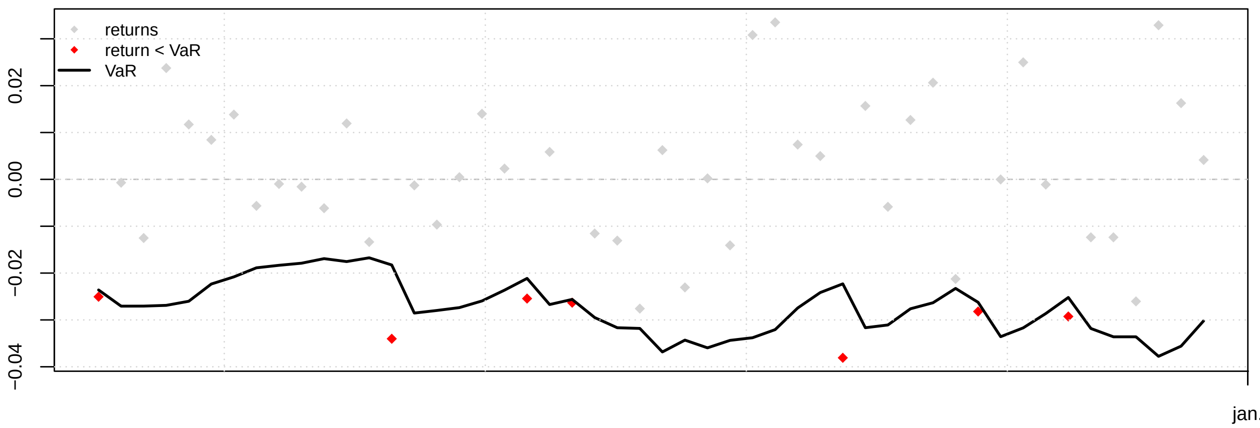
Here we will calculate and interpret the Value at Risk (VaR) using estimated volatilities from several different models. First we use the variance-covariance method, calculating the VaR with a static forecast 1 ahead, using the EGARCH model. Remember that this was the first type of forecast we did in section 6.

```
#>
#> *------------------------------------*
#> *       GARCH Model Forecast         *
#> *------------------------------------*
#> Model: eGARCH
#> Horizon: 1
#> Roll Steps: 0
#> Out of Sample: 0
#>
#> 0-roll forecast [T0=1976-12-30 01:00:00]:
#>        Series   Sigma
#> T+1 0.0006739 0.01791
```

```
#> [1] -0.02945933
```

This value means that, with a confidence level of 95%, the largest expected loss for tomorrow in our index is $\approx$ 2.95%. In other terms, the probability of the return tomorrow being lower than -2.95% is 5%.

Next we calculate the VaR with a rolling window dynamic forecast, using the EGARCH model, with a significance level of 5%.

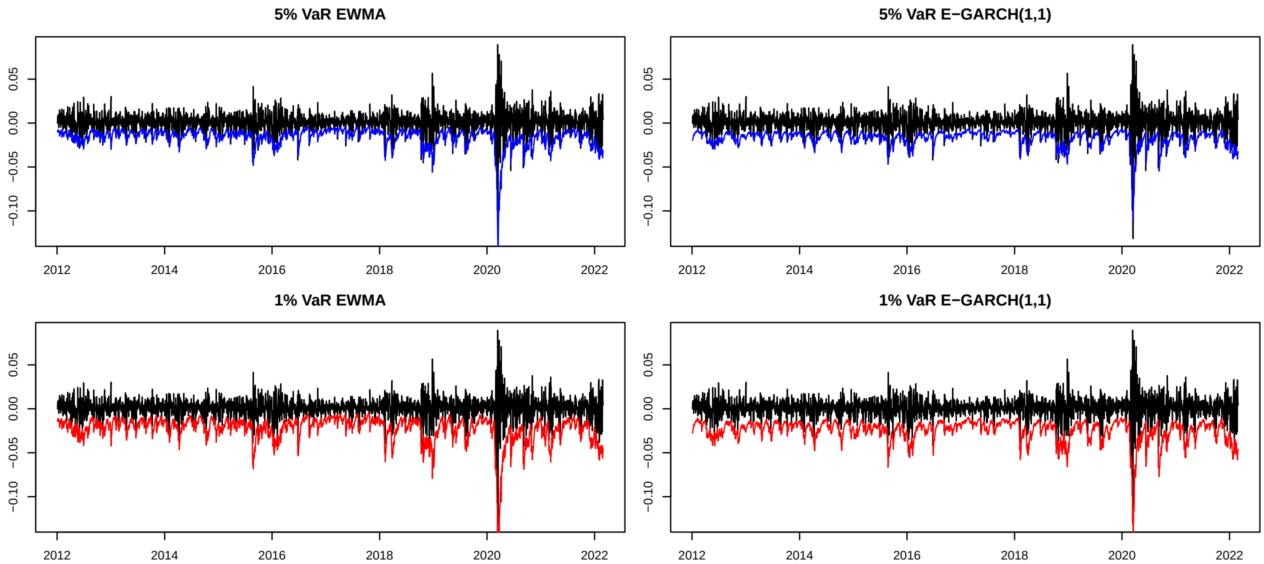



```
#> VaR Backtest Report
#> ===========================================
#> Model:                 eGARCH-std
#> Backtest Length: 50
#> Data:
#>
#> ==========================================
#> alpha:                 5%
#> Expected Exceed: 2.5
#> Actual VaR Exceed:   7
#> Actual %:              14%
#>
#> Unconditional Coverage (Kupiec)
#> Null-Hypothesis: Correct Exceedances
#> LR.uc Statistic: 5.855
#> LR.uc Critical:      3.841
#> LR.uc p-value:       0.016
#> Reject Null:     YES
#>
#> Conditional Coverage (Christoffersen)
#> Null-Hypothesis: Correct Exceedances and
#>                  Independence of Failures
#> LR.cc Statistic: 7.839
#> LR.cc Critical:      5.991
#> LR.cc p-value:       0.02
#> Reject Null:     YES
```

The report above shows that our set level of 5% significance is not kept, i.e. that the largest expected loss cannot be quantified at the 5% significance level. Instead, the VaR is estimated to be 14%, which means that the probability of the return the next day being lower than the VaR is $\approx 14\%$ instead of 5%. In practice, this means that the company should set aside more funds than expected, in order to cover the significance level of 5%.

LITT USIKKER PÅ DENNE TOLKNINGEN!

Next, we calculate the VaR using an EWMA model with $\lambda = 0.75$.



HVA ER DETTE?

HVORFOR IKKE FORECASTS SOM BLIR BRUKT HER?

9 Multivariate DCC GARCH

In order to solve this problem I have chosen the stock of Stratus Properties Inc. (STRS), which is one of the top 30 components of the NASDAQ Composite Index. Similarly to the IXIC-data, this data has been downloaded in a csv-file directly from the Yahoo Finance website. The adjusted close of the time series is plotted below.

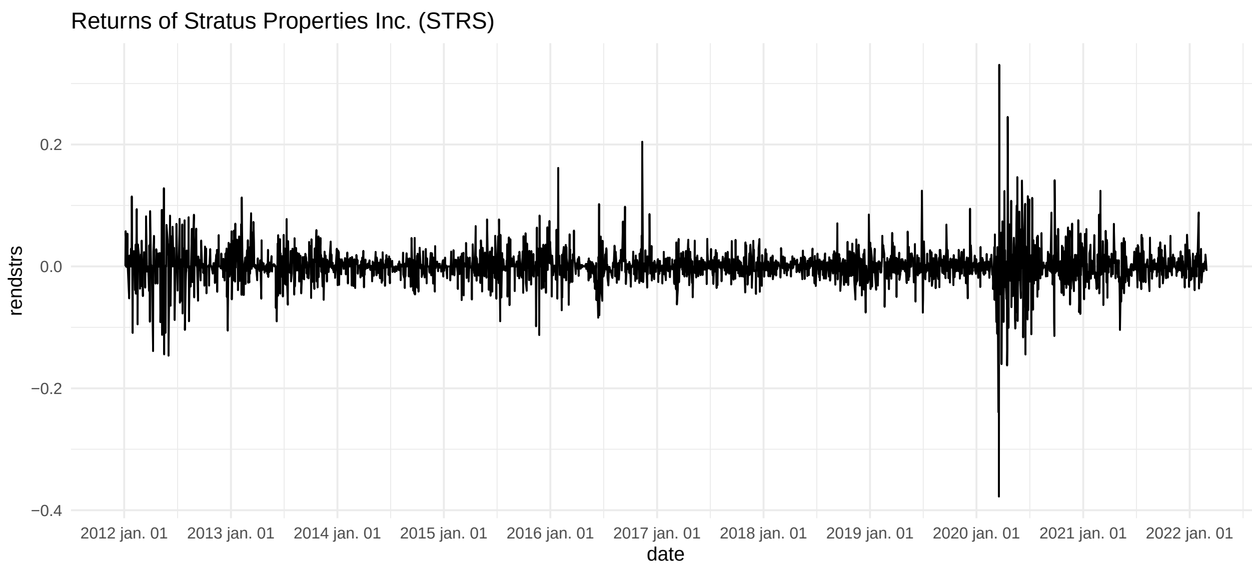
```
#>         Date Open High  Low Close Adj.Close Volume
#> 1 2012-01-03 7.58 7.90 7.58  7.90  7.617352   8200
#> 2 2012-01-04 7.90 7.90 7.90  7.90  7.617352      0
#> 3 2012-01-05 7.90 7.90 7.90  7.90  7.617352      0
#> 4 2012-01-06 7.87 9.89 7.87  8.37  8.070537   1800
#> 5 2012-01-09 8.40 8.65 8.40  8.65  8.340519    600
#> 6 2012-01-10 8.70 8.86 8.64  8.64  8.330877   4500
```

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

```
#> 'data.frame':    2556 obs. of  7 variables:
#>  $ Date     : chr  "2012-01-03" "2012-01-04" "2012-01-05" "2012-01-06" ...
#>  $ Open     : num  7.58 7.9 7.9 7.87 8.4 8.7 8.88 8.79 9.18 9.05 ...
#>  $ High     : num  7.9 7.9 7.9 9.89 8.65 8.86 8.88 9.18 9.18 9.05 ...
#>  $ Low      : num  7.58 7.9 7.9 7.87 8.4 8.64 8.7 8.79 9.18 8.58 ...
#>  $ Close    : num  7.9 7.9 7.9 8.37 8.65 8.64 8.7 9.18 9.18 8.8 ...
#>  $ Adj.Close: num  7.62 7.62 7.62 8.07 8.34 ...
#>  $ Volume   : int  8200 0 0 1800 600 4500 2300 2000 0 2200 ...
```

```
#> [1] 7.617352 7.617352 7.617352 8.070537 8.340519 8.330877
```

The returns of STRS are calculated and plotted below.



Analysis of stationarity shows that this series is integrated of order 1 as well, similarly to IXIC. Thus, we work with the returns of the series instead of the series itself.

After doing a similar analysis of this series, I have come to the conclusion that an MA(1)-EGARCH is the best model to estimate its volatility. This model is used, together with the ARMA(2,2)-EGARCH from IXIC, to estimate the multivariate DCC GARCH model for these two series (both of which are estimated on the logarithmic returns, not on the series themselves).

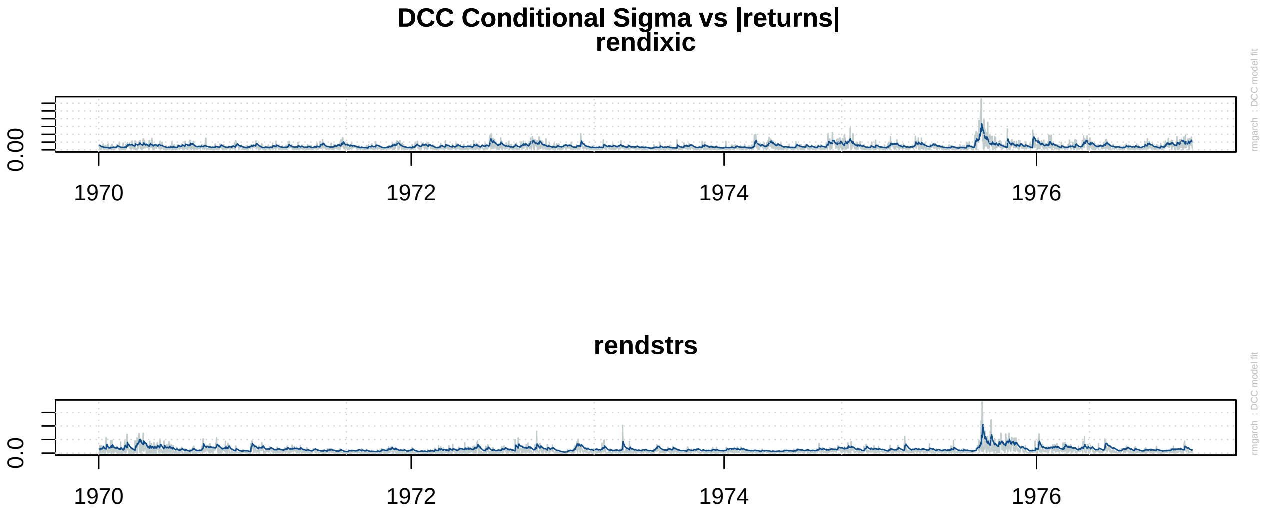
```
#>
#> *---------------------------------*
#> *          DCC GARCH Fit          *
#> *---------------------------------*
#>
#> Distribution         :  mvnorm
#> Model                :  DCC(1,1)
#> No. Parameters       :  20
#> [VAR GARCH DCC UncQ] : [0+17+2+1]
#> No. Series           :  2
```

```
#> No. Obs.             :  2555
#> Log-Likelihood       :  14075.25
#> Av.Log-Likelihood    :  5.51
#>
#> Optimal Parameters
#> -----------------------------------
#>                    Estimate  Std. Error   t value Pr(>|t|)
#> [rendixic].mu      0.000945    0.000168   5.63015 0.000000
#> [rendixic].ar1     0.137188    0.008107  16.92259 0.000000
#> [rendixic].ar2     0.310470    0.011440  27.13847 0.000000
#> [rendixic].ma1    -0.178777    0.017057 -10.48116 0.000000
#> [rendixic].ma2    -0.292323    0.011170 -26.17109 0.000000
#> [rendixic].omega  -0.389642    0.007845 -49.66907 0.000000
#> [rendixic].alpha1 -0.193182    0.023139  -8.34890 0.000000
#> [rendixic].beta1   0.958980    0.001084 884.69593 0.000000
#> [rendixic].gamma1  0.161978    0.022819   7.09833 0.000000
#> [rendixic].shape   5.708617    0.736620   7.74974 0.000000
#> [rendstrs].mu      0.000104    0.000411   0.25408 0.799435
#> [rendstrs].ma1    -0.189850    0.017865 -10.62700 0.000000
#> [rendstrs].omega  -0.271402    0.199469  -1.36062 0.173634
#> [rendstrs].alpha1 -0.064225    0.035459  -1.81126 0.070101
#> [rendstrs].beta1   0.961041    0.028377  33.86663 0.000000
#> [rendstrs].gamma1  0.398593    0.134602   2.96126 0.003064
#> [rendstrs].shape   2.846712    0.207328  13.73046 0.000000
#> [Joint]dcca1       0.016504    0.008918   1.85073 0.064208
#> [Joint]dccb1       0.924780    0.035361  26.15250 0.000000
#>
#> Information Criteria
#> ---------------------
#>
#> Akaike       -11.002
#> Bayes        -10.956
#> Shibata      -11.002
#> Hannan-Quinn -10.986
#>
#>
#> Elapsed time : 3.15439
```

WHAT ASSUMPTIONS NEED TO BE FULFILLED FOR THIS MODEL!?

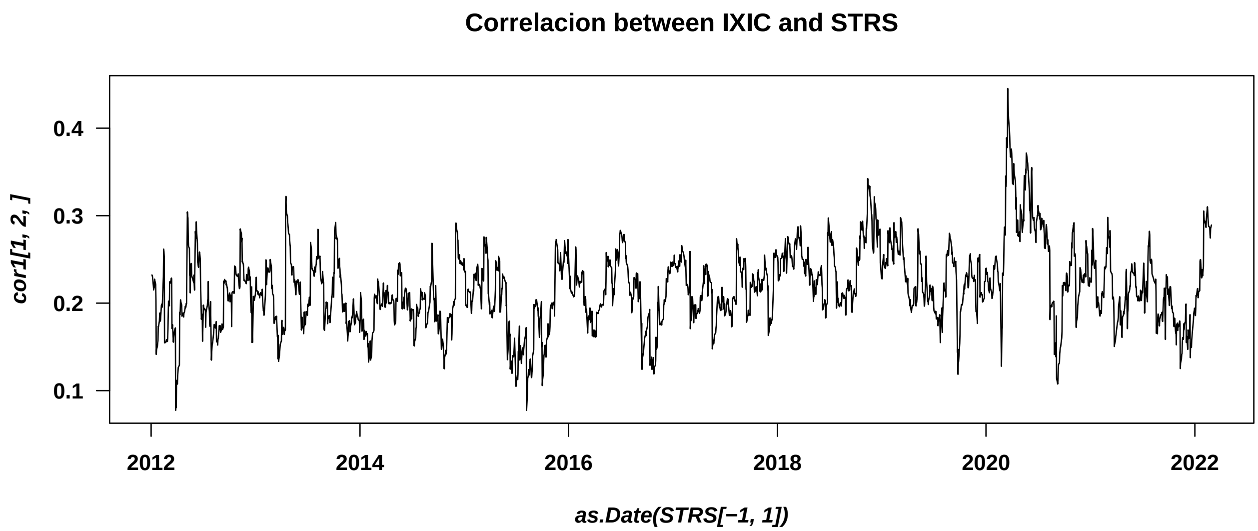
10 Estimated Correlation and News Impact Surface

The plot below shows the conditional standard error estimated from the model (in blue) and the realized absolute returns (in grey).

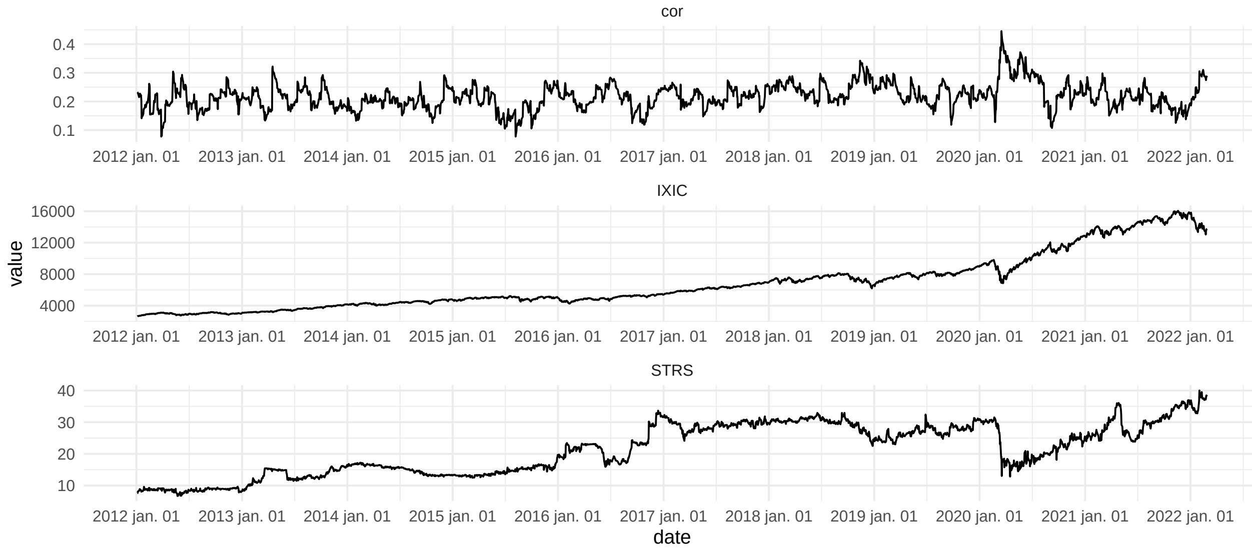


From this plot we can gather that the model has made good estimations of the standard error, because the behaviour of the graphs are similar. Note that we cannot conclude anything based on the absolute values of the quantities; we are only interested in the shape or the behaviour of the quantities.

The plot below shows the conditional correlation estimated from the model.



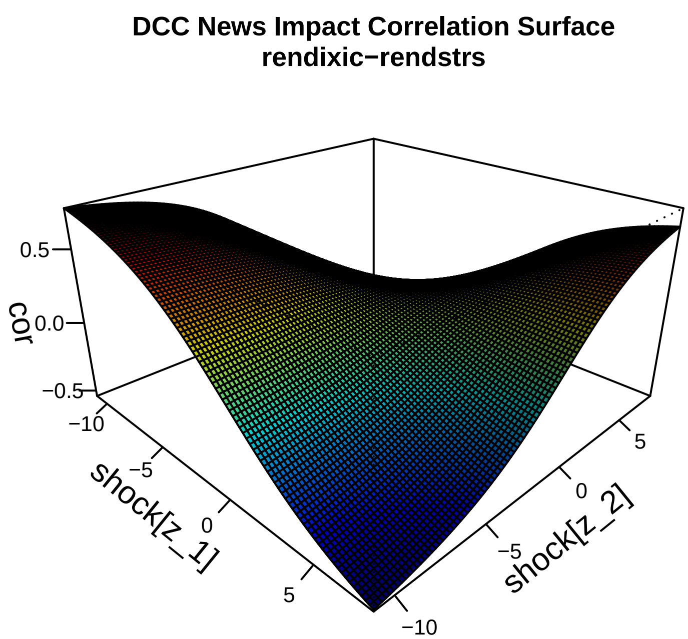
The correlation is plotted alongside the two original time series below.



Just by looking at the two time series side by side, they look to move in a similar fashion. The peak in the correlation between the two series, which corresponds to a correlation of about 0.45, took place in the beginnin of 2020, when both prices fell, most likely because of the COVID-pandemic.

OTHER THAN THAT, I AM NOT QUITE CERTAIN THAT I AM ABLE TO INTERPRET ANYTHING ELSE FROM THE PLOTS.

The news impact surface is plotted below.



From this we can learn that

- Simultaneous negative news in both series lead to the largest increase in correlation ????.
- Simultaneous positive news in both series lead to an increase in correlation ??? as well, but not to the same extent that the negative shocks do. This shows that the leverage effect has been taken into account in the model.
- Negative news in one of the series and positive news in the other leads to a negative correlation ???, which (again) shows that the leverage effect is taken into account, since the negative news clearly get a larger weight than the positive news.

STEMMER DET AT DET ER KORRELASJONEN SOM ENDRES I NEWS IMPACT CURVE OVER, OG IKKE VOLATILITY!?!?!

11 Conclusions

12 QUESTIONS FOR PROFE:

- Positivity and stationarity assumptions for EGARCH? Looks like positivity is not needed! But is stationarity an assumption that needs to be fulfilled in the model?
- GÅ GJENNOM OPPGAVEN FOR Å FINNE FLERE SPØRSMÅL!