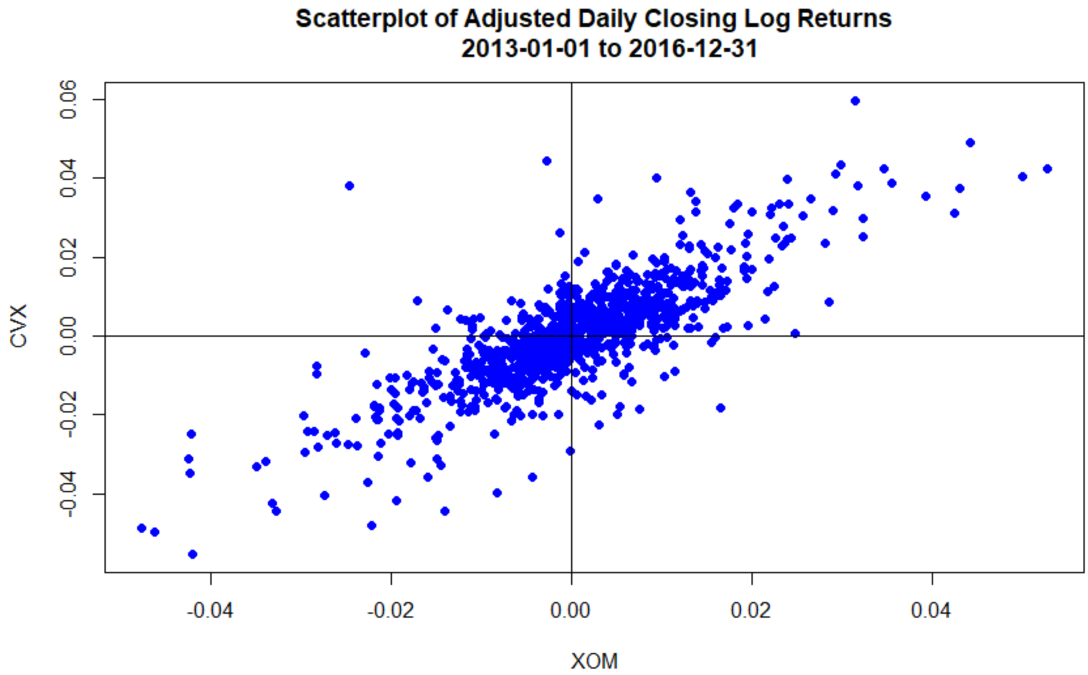
Ji Jung

June 5, 2017

Predict 413 Predictive Modeling II Spring 2017, Northwestern University

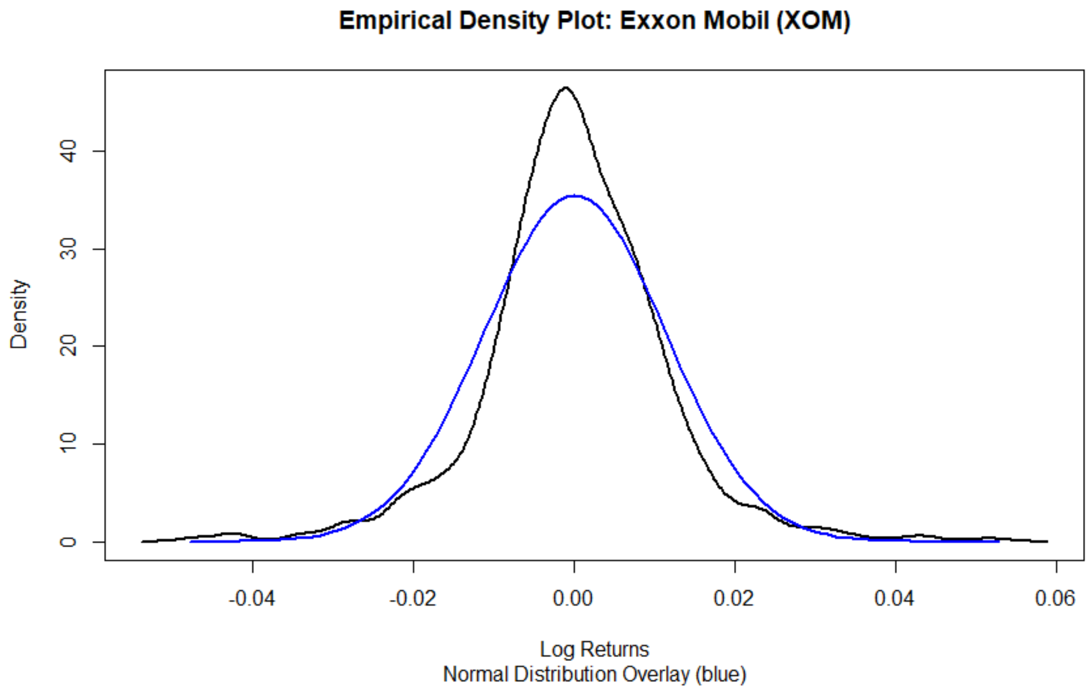
1 Time Series Model Construction

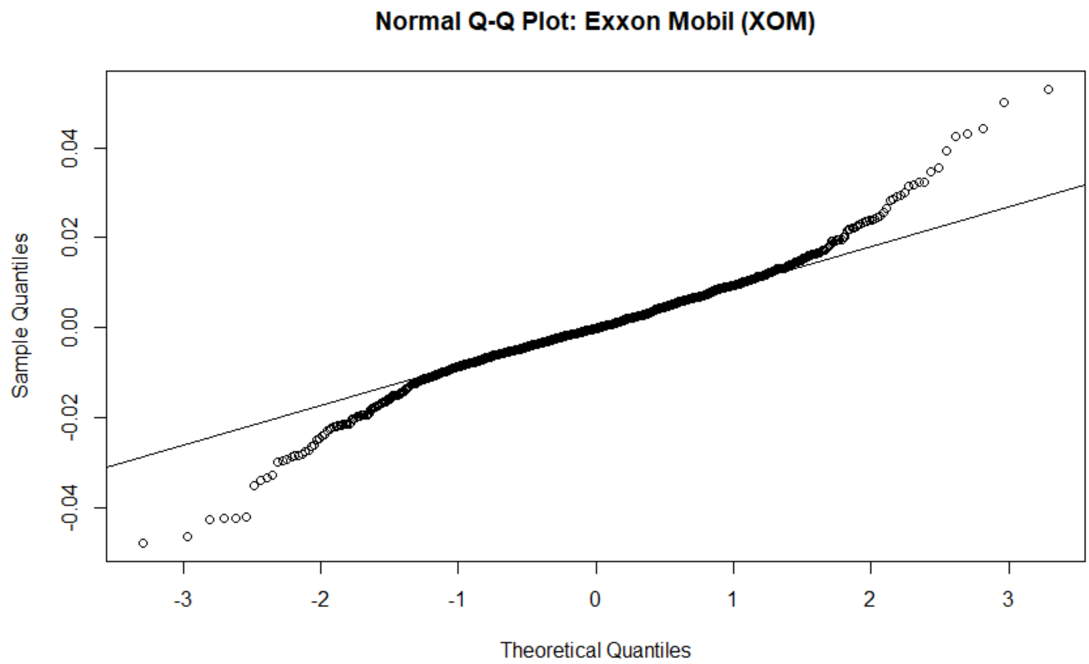
1.1 EDA

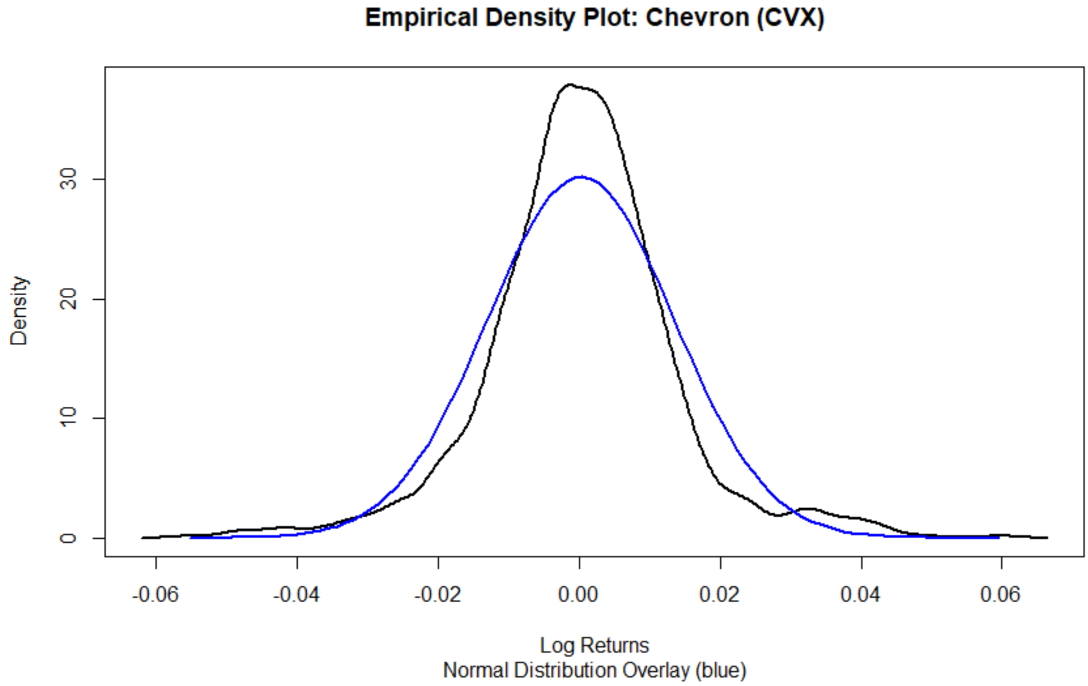


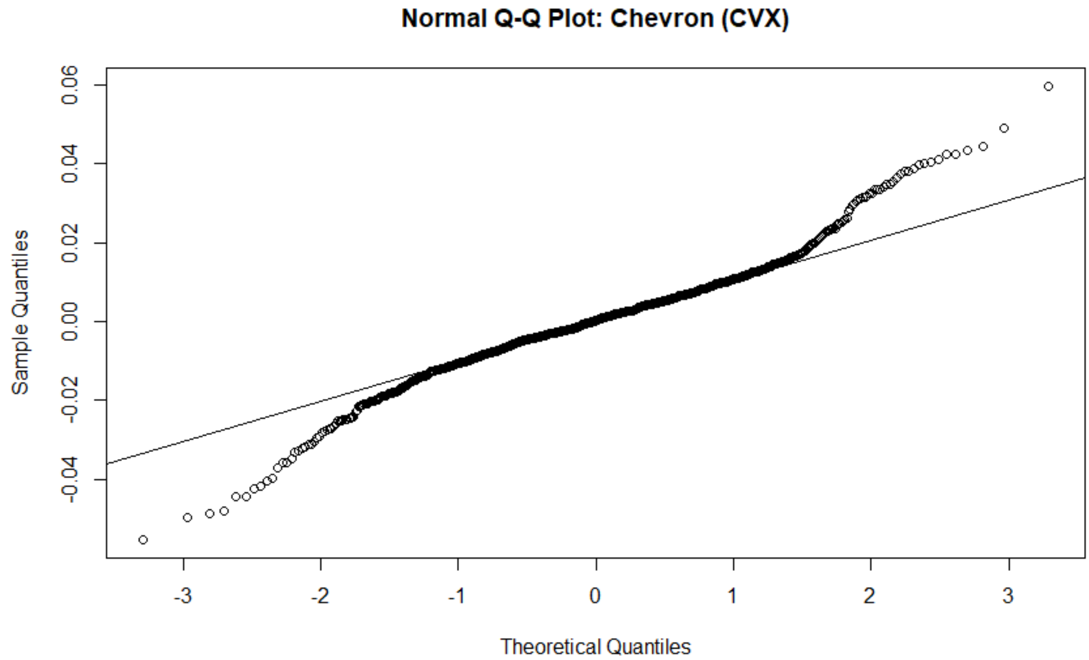
	XOM	CVX
nobs	1007.000000	1007.000000
NAs	0.000000	0.000000
Minimum	-0.047709	-0.055185
Maximum	0.052894	0.059539
1. Quartile	-0.005595	-0.006786
3. Quartile	0.006331	0.006975
Mean	0.000138	0.000216
Median	-0.000202	0.000162
Sum	0.138969	0.217343
SE Mean	0.000354	0.000417
LCL Mean	-0.000558	-0.000602
UCL Mean	0.000834	0.001034
Variance	0.000127	0.000175
Stdev	0.011248	0.013226
Skewness	0.019916	0.041104
Kurtosis	2.785387	2.269472

Both XOM and CVX are fit for Gaussian distribution.


Empirical Density Plot: Exxon Mobil (XOM)
Density
0
10
20
30
40
-0.04
-0.02
0.00
0.02
0.04
0.06
Log Returns
Normal Distribution Overlay (blue)


Normal Q-Q Plot: Exxon Mobil (XOM)
Sample Quantiles
-0.04
-0.02
0.00
0.02
0.04
-3
-2
-1
0
1
2
3
Theoretical Quantiles





1.2 Tests

```
        One Sample t-test

data:  returns.XOM
t = 0.38932, df = 1006, p-value = 0.6971
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0005575806  0.0008335871
```

```
sample estimates:
   mean of x
0.0001380033

	One Sample t-test

data:  returns.cvx
t = 0.51783, df = 1006, p-value = 0.6047
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0006020694  0.0010337337
sample estimates:
   mean of x
0.0002158322
```

Third Moment: Skewness	XOM	CVX
Skewness	0.258402	0.5332947
P-Value	0.7960967	0.5938296

Fourth Moment: Kurtosis	XOM	CVX
Kurtosis	18.11695	14.76845
P-Value	0	0

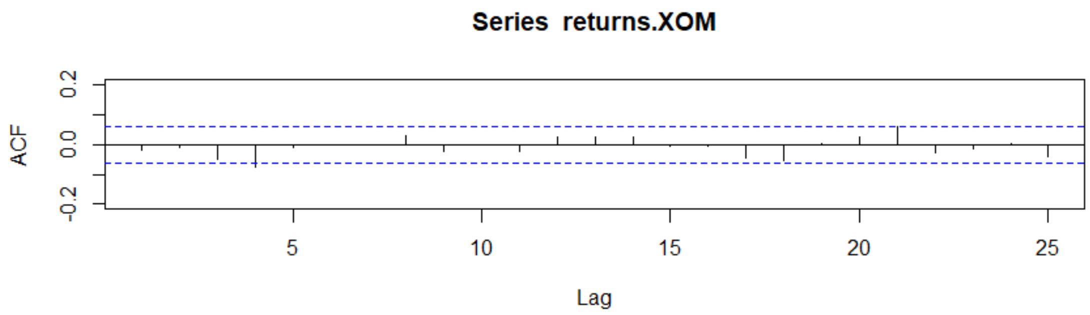
1.3 Box-Jenkins

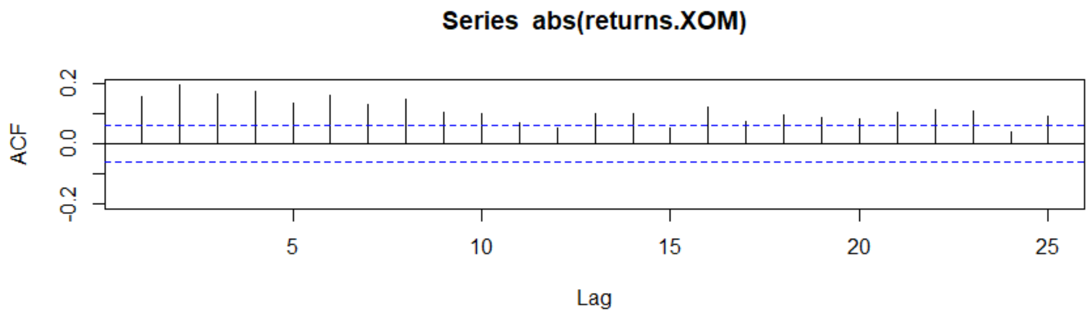
XOM

Applying an ADF test - check for unit-root nonstationarity

```
	Augmented Dickey-Fuller Test

data:  returns.XOM
Dickey-Fuller = -10.342, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

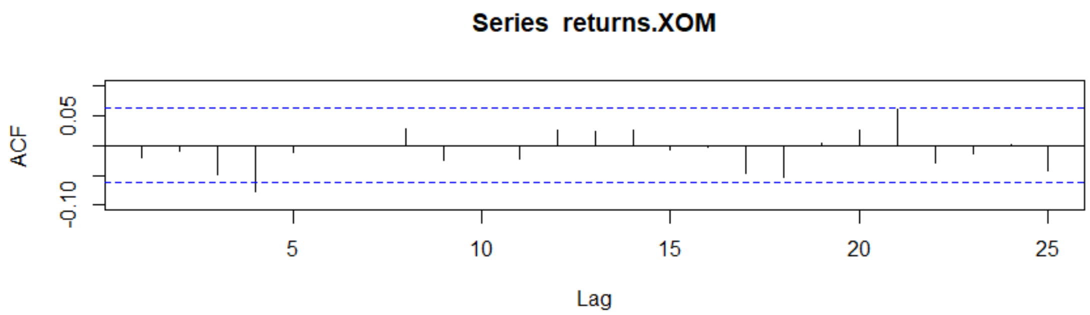
Checking for serial correlations in the log return.

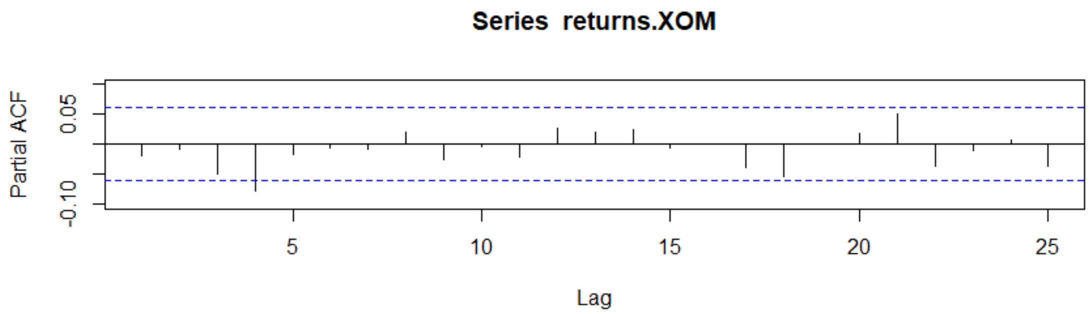





```
        Box-Ljung test

data:  returns.XOM
X-squared = 10.508, df = 10, p-value = 0.3971
```

First lags removed ACF and PACF





Based on AIC, ARMA(2,2) is suggested.

```
ARIMA(2,0,2) with zero mean     : -6175.533
```

```
 ARIMA(0,0,0) with zero mean     : -6178.975
 ARIMA(1,0,0) with zero mean     : -6177.354
 ARIMA(0,0,1) with zero mean     : -6177.362
 ARIMA(1,0,1) with zero mean     : -6175.346

 Best model: ARIMA(0,0,0) with zero mean

Series: returns.XOM
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 0.0001264:  log likelihood=3090.49
AIC=-6178.98   AICc=-6178.97   BIC=-6174.06
```

Model 3

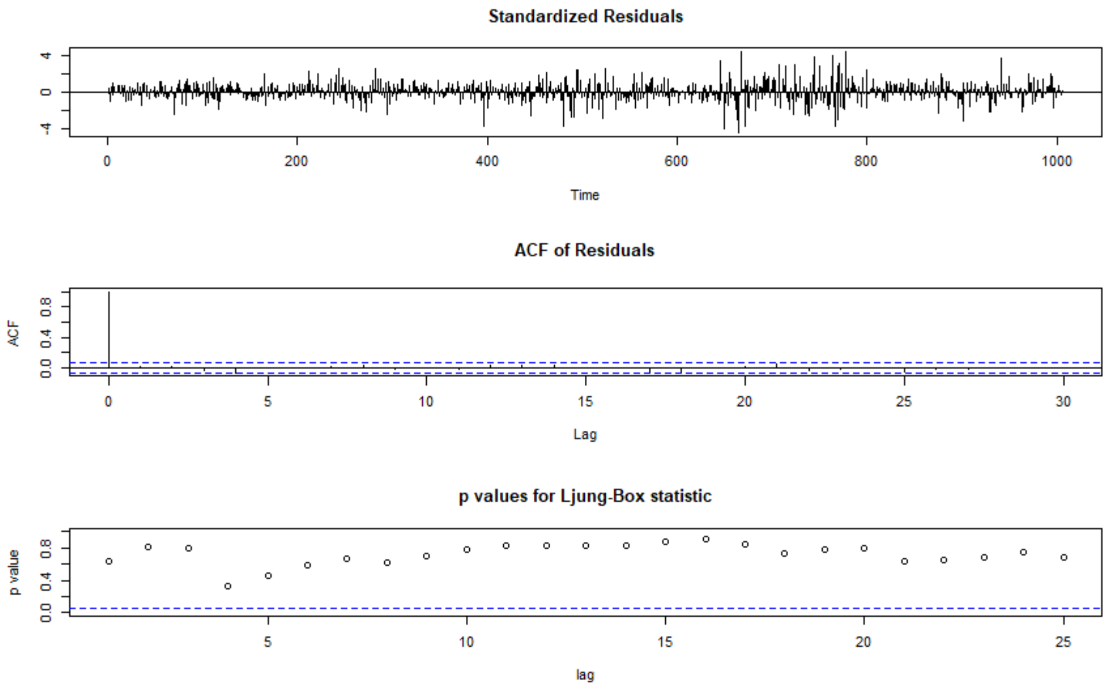
```
ARIMA(2,0,2) with zero mean

Coefficients:
          ar1     ar2     ma1      ma2
      -0.0483  0.7500  0.0094  -0.7788
s.e.   0.2954  0.2674  0.2878   0.2649

sigma^2 estimated as 0.0001263:  log likelihood=3092.8
AIC=-6175.59   AICc=-6175.53   BIC=-6151.02
```

Standard of Error

```
ar1       ar2       ma1       ma2
0.2953717 0.2673597 0.2877983 0.2649023
```

T-Ratio

```
ar1        ar2        ma1        ma2
0.16365687 2.80525082 0.03274678 2.93977798
```



Box-Ljung test with 10 lags

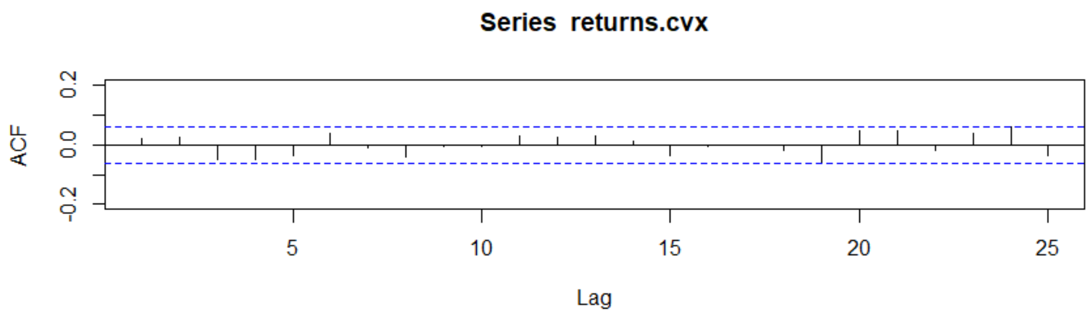
```
data:  residuals(returns.XOM.m3)
X-squared = 6.3955, df = 6, p-value = 0.3804

        Box-Ljung test with 20 lags

data:  residuals(returns.XOM.m3)
X-squared = 14.589, df = 16, p-value = 0.5549
```

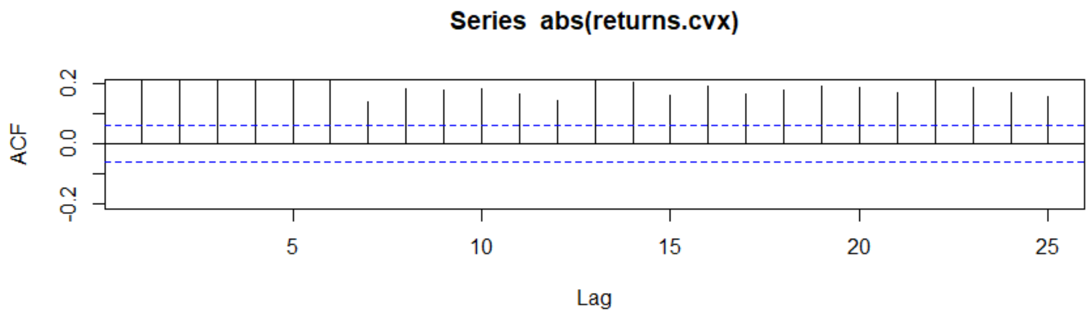
CVX

Applying an ADF test - check for unit-root nonstationarity

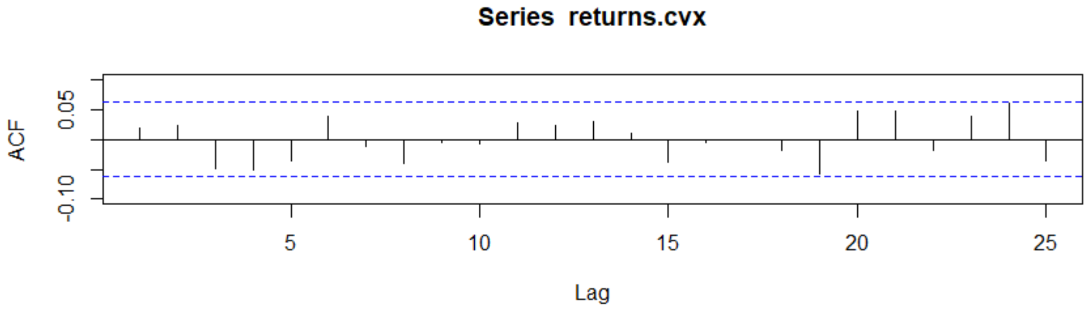
```
        Augmented Dickey-Fuller Test

data:  returns.cvx
Dickey-Fuller = -9.905, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```

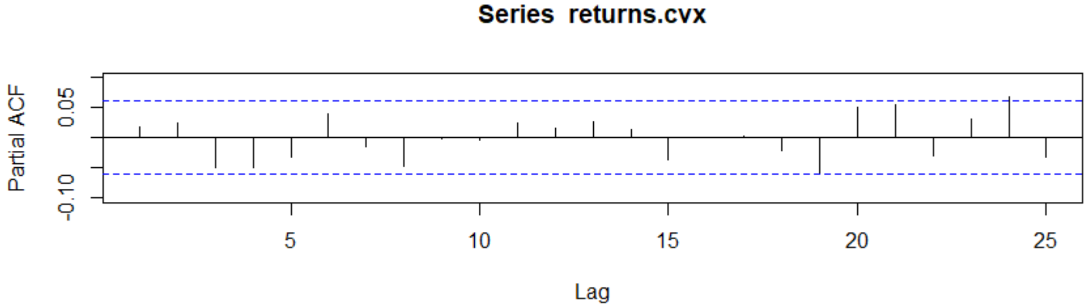
Checking for serial correlations in the log return.






```
        Box-Ljung test

data:  returns.cvx
X-squared = 10.628, df = 10, p-value = 0.3872
```





Based on AIC, ARMA(3, 2) is suggested

```
 ARIMA(2,0,2) with zero mean     : -5849.555
 ARIMA(0,0,0) with zero mean     : -5852.61
 ARIMA(1,0,0) with zero mean     : -5850.951
 ARIMA(0,0,1) with zero mean     : -5850.934
 ARIMA(1,0,1) with zero mean     : -5849.019

 Best model: ARIMA(0,0,0) with zero mean

Series: returns.cvx
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 0.0001748:  log likelihood=2927.31
AIC=-5852.61   AICc=-5852.61   BIC=-5847.7
```

Model 3

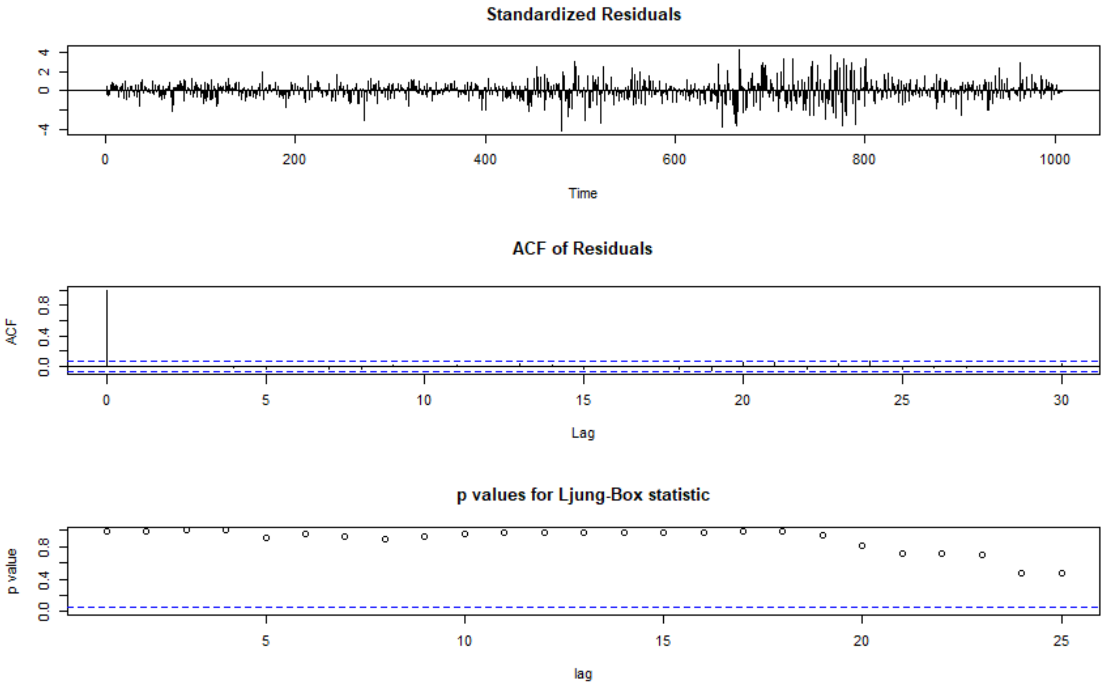
```
Series: returns.cvx
ARIMA(3,0,2) with zero mean

Coefficients:
         ar1      ar2      ar3      ma1     ma2
      0.5806  -0.6408  -0.0506  -0.5639  0.6652
s.e.  0.2700   0.2020   0.0357   0.2697  0.1854

sigma^2 estimated as 0.0001744:  log likelihood=2931.12
AIC=-5850.23   AICc=-5850.15   BIC=-5820.74
```

Standard of Error

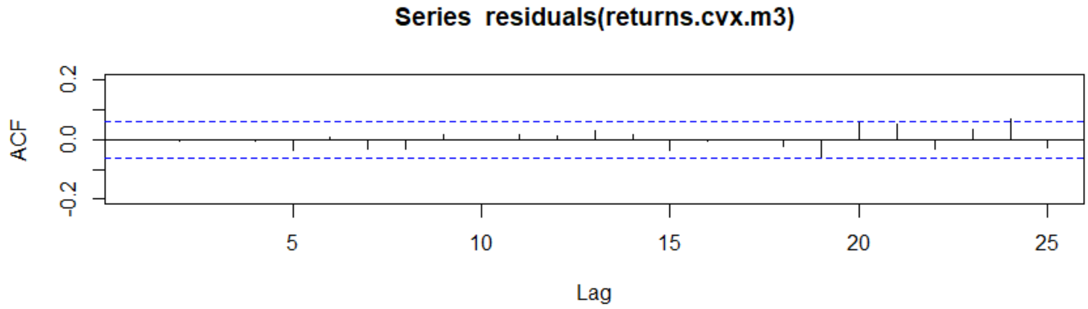
```
ar1       ar2       ma1       ma2
0.2953717 0.2673597 0.2877983 0.2649023
```

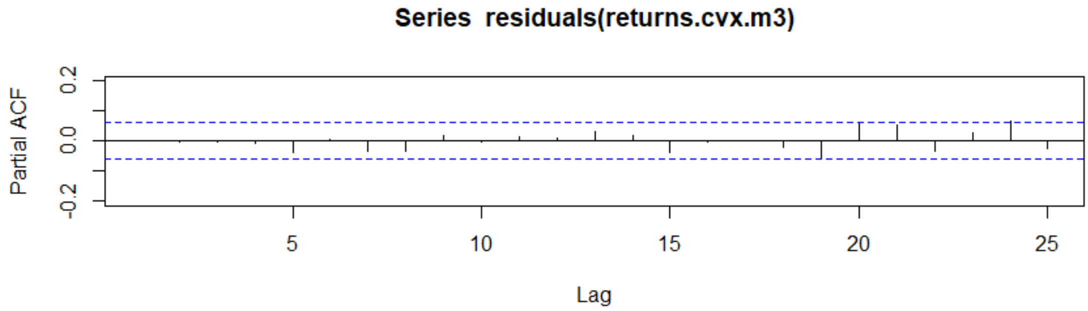
T-Ratio

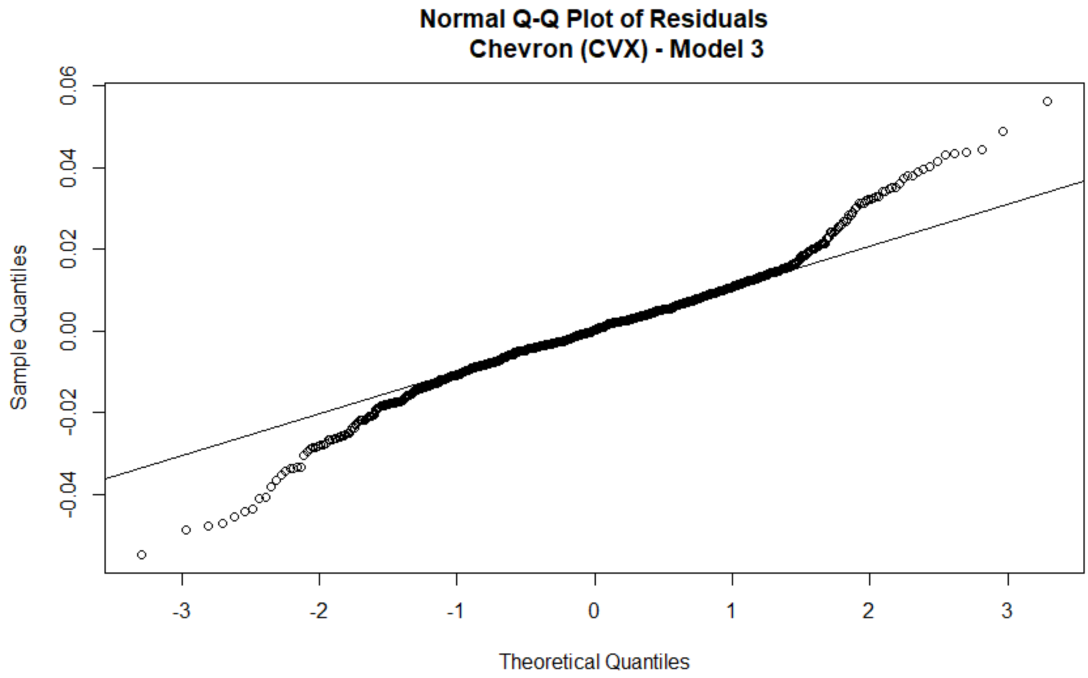
```
ar1      ar2      ar3      ma1      ma2
2.150496 3.172070 1.416026 2.090551 3.587158
```




```
       Box-Ljung test with 10 lags

data:  residuals(returns.cvx.m3)
X-squared = 3.9148, df = 5, p-value = 0.5617

       Box-Ljung test with 20 lags

data:  residuals(returns.cvx.m3)
X-squared = 14.247, df = 15, p-value = 0.5069
```







1.4 Forecasts

XOM – 20 days

```
$pred
Time Series:
Start = 1008
End = 1027
Frequency = 1
```

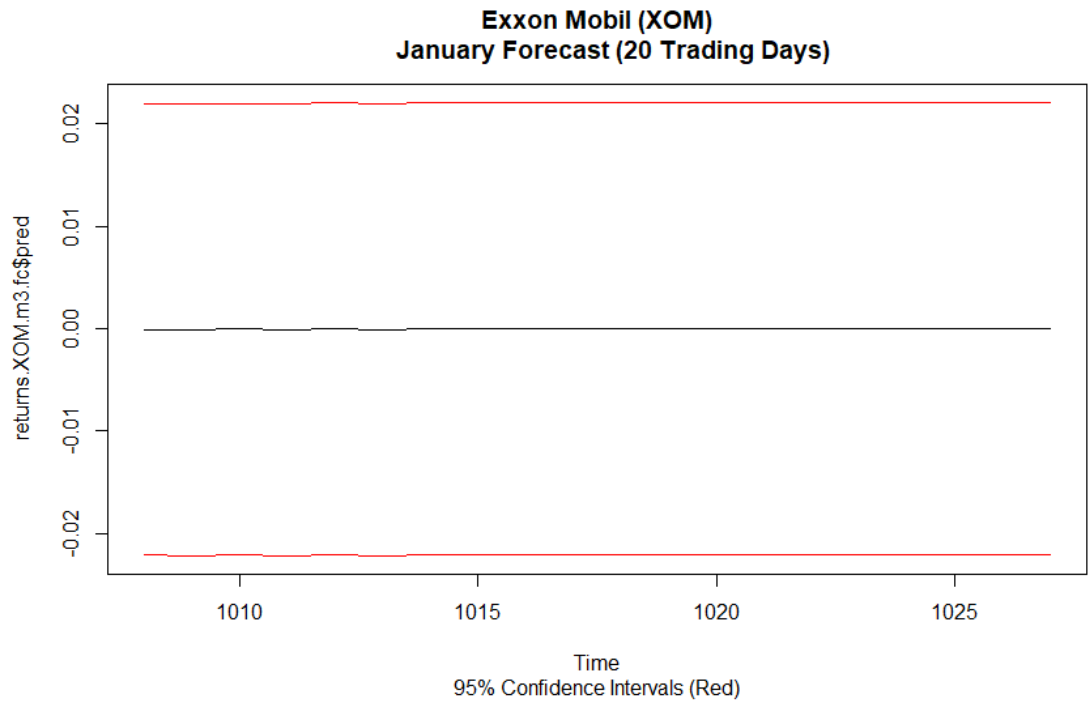
```
 [1] -6.568394e-05 -1.075056e-04 -4.406690e-05 -7.850019e-05 -
2.925599e-05 -5.746179e-05 -1.916464e-05 -4.217056e-05
 [9] -1.233518e-05 -3.103211e-05 -7.751440e-06 -2.289972e-05 -
4.706702e-06 -1.694752e-05 -2.710842e-06 -1.257979e-05
[17] -1.425059e-06 -9.366091e-06 -6.160568e-07 -6.994891e-06

$se
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1] 0.01124023 0.01124874 0.01125279 0.01125716 0.01125914
0.01126139 0.01126235 0.01126351 0.01126398 0.01126458
[11] 0.01126481 0.01126512 0.01126523 0.01126539 0.01126544
0.01126553 0.01126555 0.01126560 0.01126561 0.01126563
```

Lower confidence level

```
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1] -0.02209614 -0.02215463 -0.02209914 -0.02214213 -0.02209677
-0.02212938 -0.02209297 -0.02211825 -0.02208934
[10] -0.02210921 -0.02208637 -0.02210213 -0.02208415 -0.02209671
-0.02208257 -0.02209261 -0.02208150 -0.02208953
[19] -0.02208080 -0.02208723
```

Upper confidence level

```
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1] 0.02196477 0.02193962 0.02201100 0.02198513 0.02203826
0.02201445 0.02205465 0.02203391 0.02206467 0.02204714
[11] 0.02207087 0.02205633 0.02207474 0.02206281 0.02207715
0.02206745 0.02207865 0.02207080 0.02207957 0.02207324
```



CVX – 20 days

```
$pred
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1]  4.143730e-04  3.438333e-04 -1.474836e-05 -2.498543e-04 -
1.530248e-04  7.199532e-05  1.525024e-04  5.015984e-05
 [9] -7.223884e-05 -8.180358e-05 -3.747689e-06  5.389761e-05
3.783626e-05 -1.237754e-05 -3.415908e-05 -1.381759e-05
[17]  1.449173e-05  1.899713e-05  2.443749e-06 -1.148738e-05

$se
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1] 0.01320437 0.01320623 0.01321391 0.01322529 0.01323974
0.01323977 0.01324601 0.01324911 0.01324943 0.01325194
[11] 0.01325236 0.01325281 0.01325358 0.01325360 0.01325388
0.01325406 0.01325407 0.01325419 0.01325422 0.01325424
```

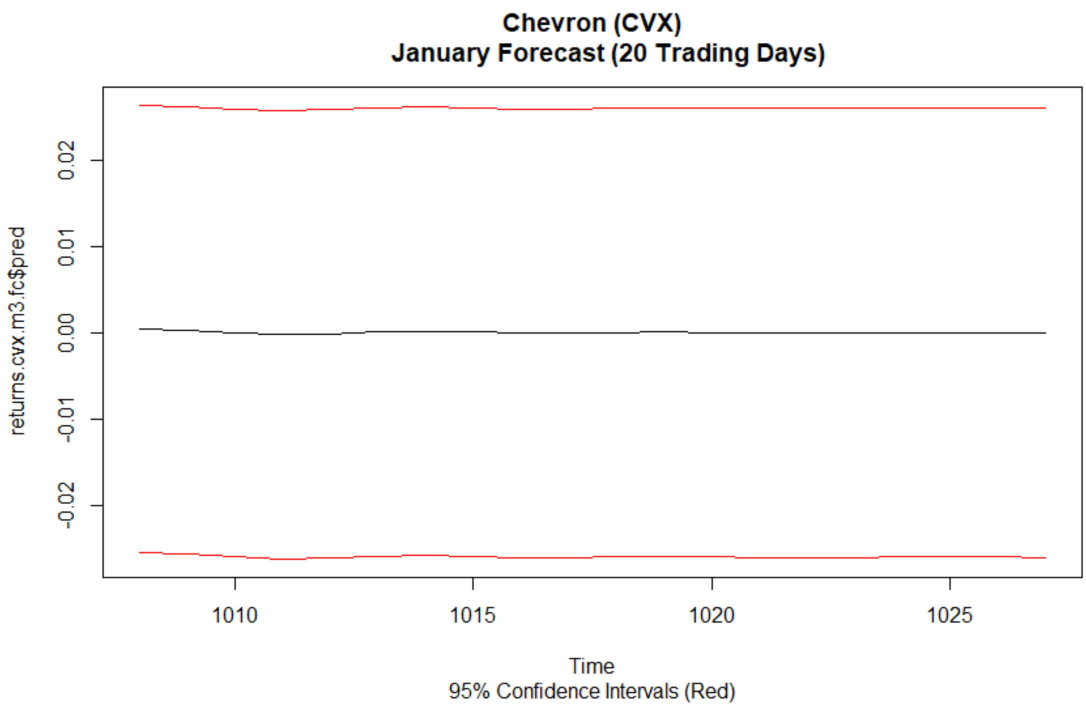
Lower confidence level

```
Time Series:
Start = 1008
End = 1027
Frequency = 1
```

```
 [1] -0.02546572 -0.02553990 -0.02591353 -0.02617095 -0.02610243
-0.02587748 -0.02580919 -0.02591762 -0.02604065
[10] -0.02605513 -0.02597789 -0.02592112 -0.02593871 -0.02598895
-0.02601129 -0.02599130 -0.02596301 -0.02595875
[19] -0.02597536 -0.02598932
```

Upper confidence level

```
Time Series:
Start = 1008
End = 1027
Frequency = 1
 [1] 0.02629447 0.02622756 0.02588404 0.02567124 0.02579638
0.02602147 0.02611420 0.02601794 0.02589617 0.02589152
[11] 0.02597040 0.02602892 0.02601438 0.02596420 0.02594297
0.02596367 0.02599199 0.02599674 0.02598024 0.02596635
```

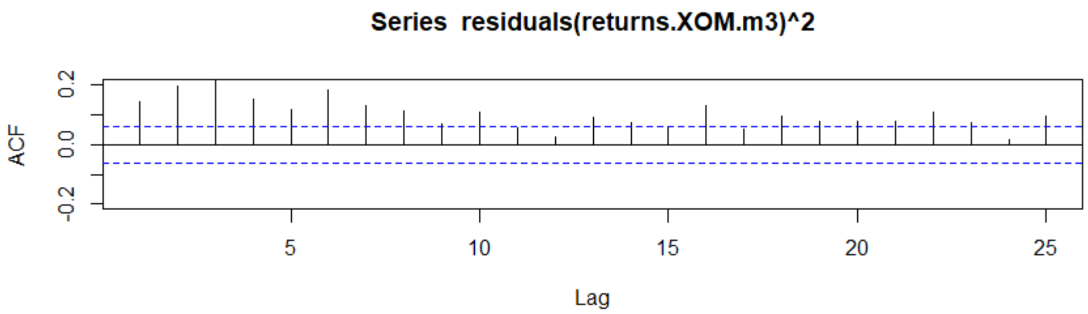


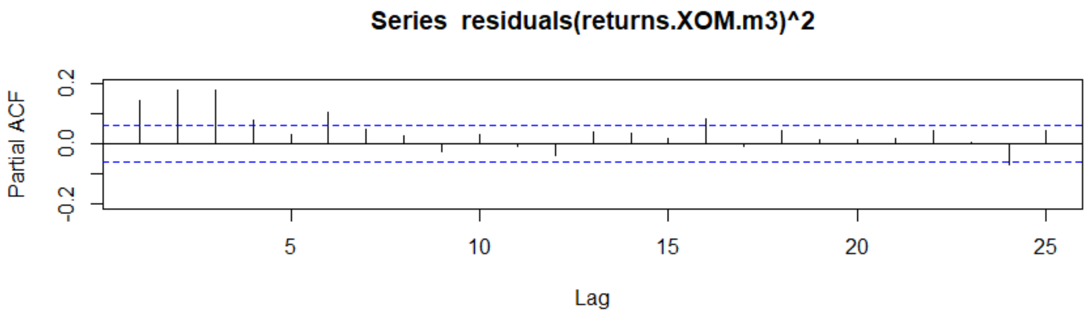
1.5 ARMA-GARCH

XOM – Model 3

```
        Box-Ljung test

data:  residuals(returns.XOM.m3)^2
X-squared = 225.91, df = 10, p-value < 2.2e-16
```

The p-value test of the squared log returns is significant, so the null is rejected. This shows that ARCH effects may be present in the log returns.

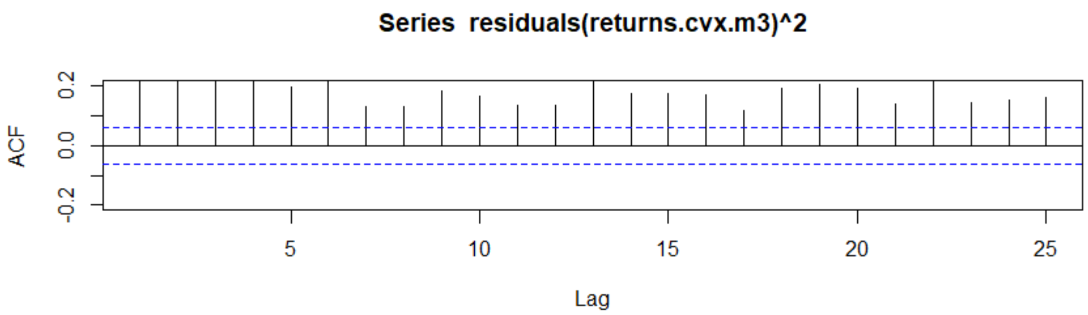


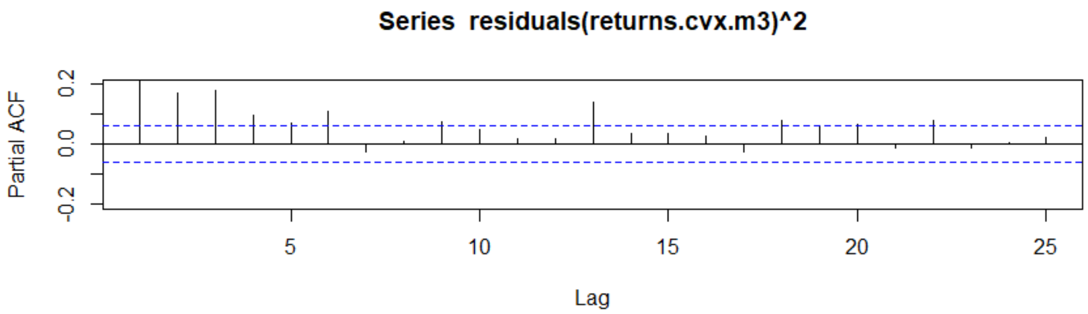


CVX – Model 3

```
        Box-Ljung test

data:  residuals(returns.cvx.m3)^2
X-squared = 445.21, df = 10, p-value < 2.2e-16
```

The p-value test of the squared log returns is significant, so the null is rejected. This shows that ARCH effects may be present in the log returns.



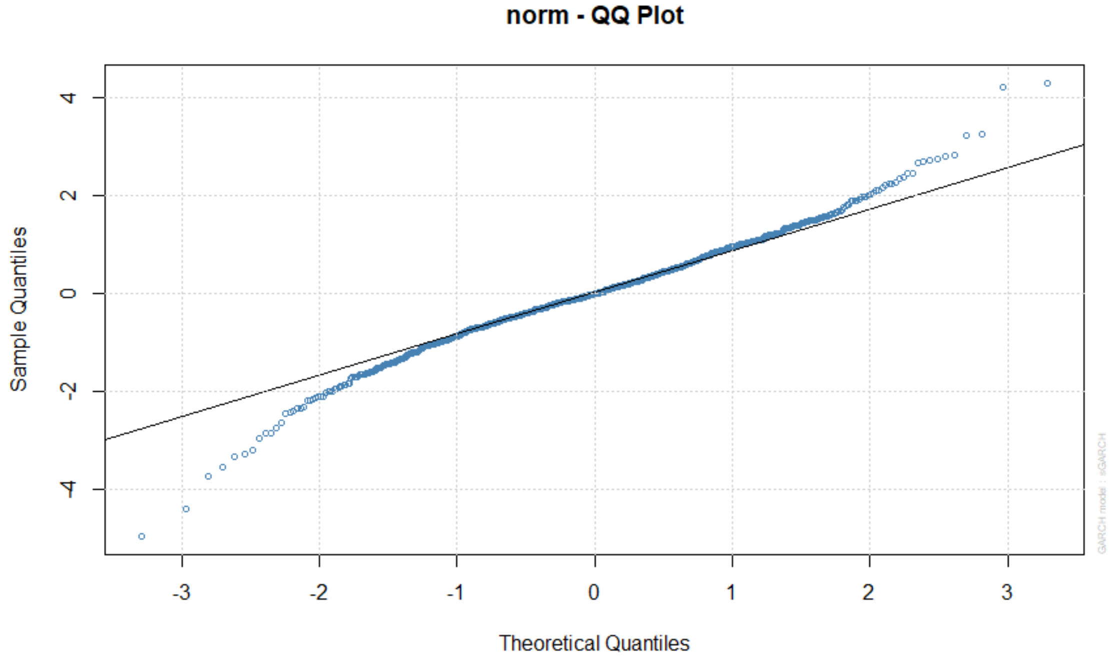


XOM Gaussian ARMA-GARCH

Examining the model fit including Ljung-Box results for standardized residuals - adequacy of model mean equation and standardized residuals squared - adequacy of model variance equation.

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model    : sGARCH(1,1)
Mean Model     : ARFIMA(2,0,2)
Distribution   : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
ar1    -1.307991    0.000360 -3634.1486 0.000000
ar2    -0.310043    0.000145 -2139.5458 0.000000
ma1     1.244026    0.000164  7593.1051 0.000000
ma2     0.238998    0.000062  3869.6487 0.000000
omega   0.000004    0.000002     2.5925 0.009527
alpha1  0.081774    0.025727     3.1785 0.001480
beta1   0.884560    0.013897    63.6509 0.000000

Robust Standard Errors:
        Estimate  Std. Error     t value Pr(>|t|)
ar1    -1.307991    0.002172  -602.12310  0.00000
ar2    -0.310043    0.000260 -1190.36245  0.00000
ma1     1.244026    0.000730  1703.87056  0.00000
ma2     0.238998    0.000242   987.01228  0.00000
omega   0.000004    0.000010     0.40483  0.68560
alpha1  0.081774    0.145502     0.56201  0.57411
beta1   0.884560    0.059937    14.75815  0.00000

LogLikelihood : 3173.905

Information Criteria
------------------------------------

Akaike       -6.2898
Bayes        -6.2556
Shibata      -6.2899
Hannan-Quinn -6.2768

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                         statistic p-value
Lag[1]                       1.080  0.2986
Lag[2*(p+q)+(p+q)-1][11]     5.637  0.7195
Lag[4*(p+q)+(p+q)-1][19]     8.639  0.6979
d.o.f=4
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.6584  0.4171
Lag[2*(p+q)+(p+q)-1][5]    3.1773  0.3756
```

```
Lag[4*(p+q)+(p+q)-1][9]    4.5597  0.4968
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     2.650 0.500 2.000  0.1035
ARCH Lag[5]     2.890 1.440 1.667  0.3062
ARCH Lag[7]     3.749 2.315 1.543  0.3842

Nyblom stability test
------------------------------------
Joint Statistic:  3.441
Individual Statistics:
ar1    0.01687
ar2    0.01666
ma1    0.02260
ma2    0.02240
omega  0.07298
alpha1 0.28167
beta1  0.21006

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:       1.69 1.9 2.35
Individual Statistic:        0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias          1.02325 0.3064
Negative Sign Bias 0.23772 0.8121
Positive Sign Bias 0.07864 0.9373
Joint Effect       1.60733 0.6577


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     40.57     0.002755
2    30     57.42     0.001279
3    40     67.12     0.003390
4    50     72.79     0.015299
```

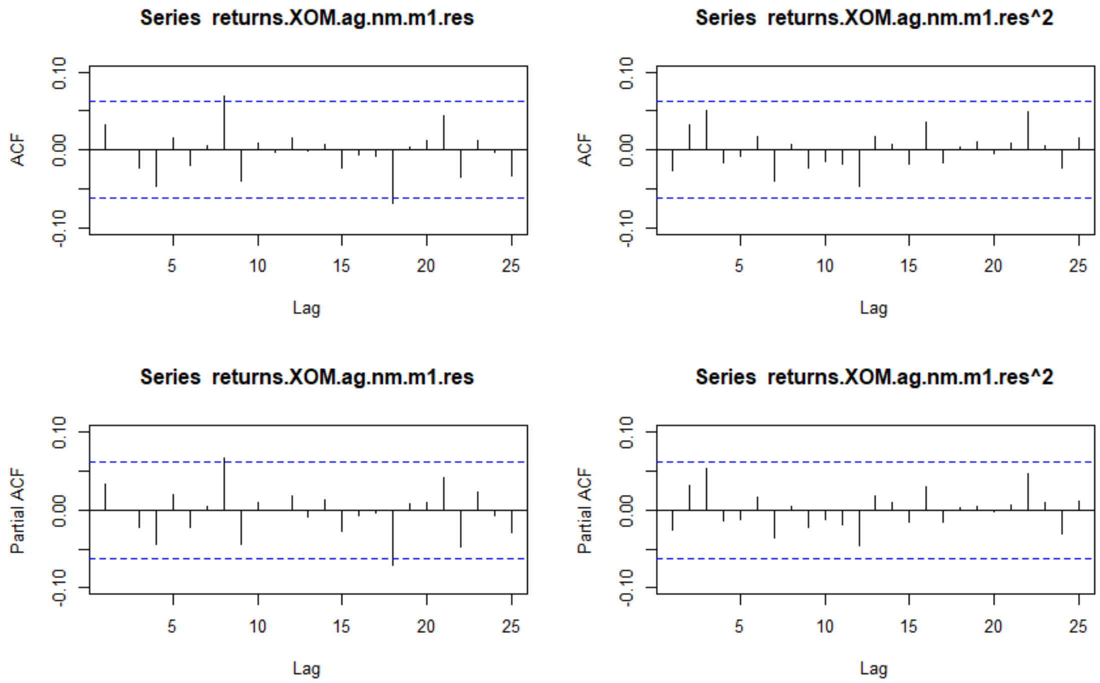


ACF & PACF

Standardized residuals - adequacy of model mean equation

Standardized residuals squared - adequacy of model variance equation

First lag removed

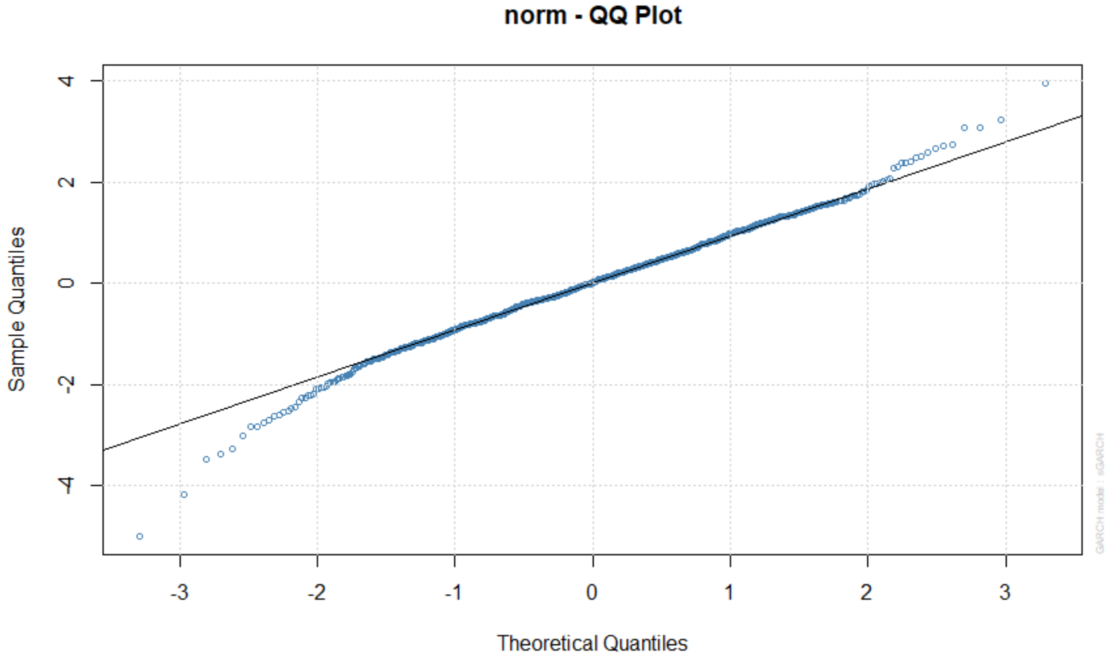


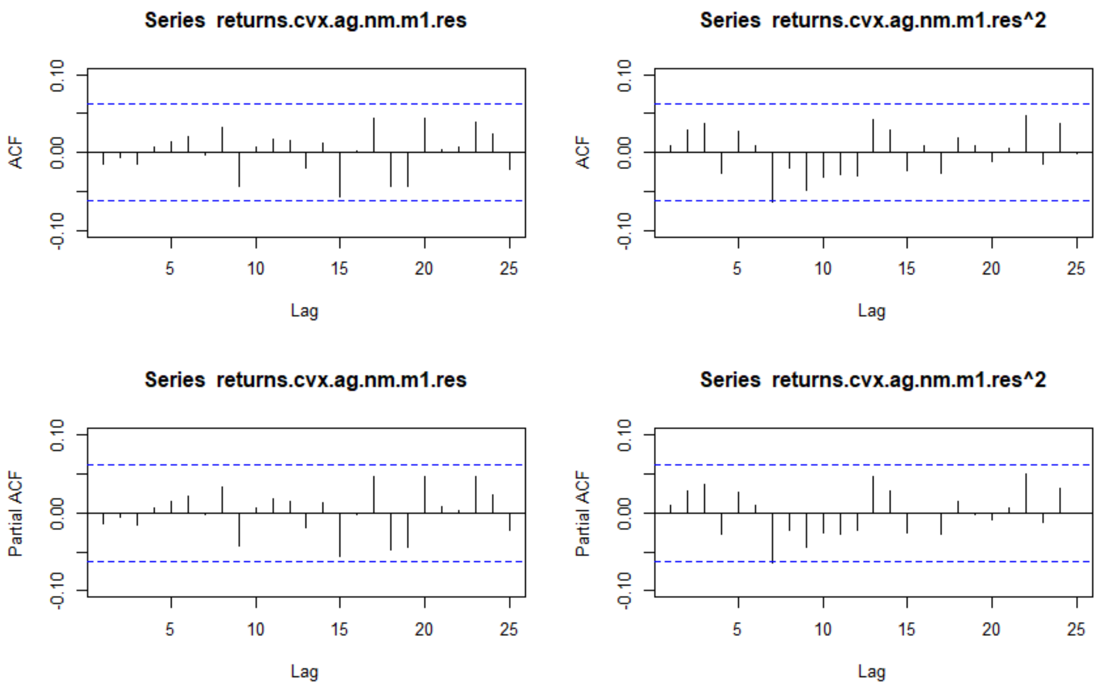
CVX Gaussian ARMA-GARCH

Checking for model fit and Ljung-Box test

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model    : sGARCH(1,1)
Mean Model     : ARFIMA(3,0,2)
Distribution   : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.000000          NA       NA       NA
ar2    -0.583687    0.262056 -2.22734 0.025925
ar3     0.010041    0.026558  0.37807 0.705379
ma1     0.000000          NA       NA       NA
ma2     0.632504    0.248889  2.54131 0.011044
omega   0.000003    0.000002  1.19701 0.231305
alpha1  0.083416    0.020693  4.03109 0.000056
beta1   0.900812    0.023167 38.88374 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.000000          NA       NA       NA
ar2    -0.583687    0.207917 -2.80731 0.004996
ar3     0.010041    0.027150  0.36983 0.711511
ma1     0.000000          NA       NA       NA
ma2     0.632504    0.197731  3.19881 0.001380
omega   0.000003    0.000009  0.29822 0.765536
alpha1  0.083416    0.055385  1.50612 0.132036
beta1   0.900812    0.072280 12.46276 0.000000

LogLikelihood : 3059.249

Information Criteria
------------------------------------

Akaike       -6.0640
Bayes        -6.0348
Shibata      -6.0641
Hannan-Quinn -6.0529

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                         statistic p-value
Lag[1]                      0.1951  0.6587
Lag[2*(p+q)+(p+q)-1][14]    2.4934  1.0000
Lag[4*(p+q)+(p+q)-1][24]    7.3988  0.9862
d.o.f=5
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
```

```
                            statistic p-value
Lag[1]                        0.08992  0.7643
Lag[2*(p+q)+(p+q)-1][5]       2.02323  0.6129
Lag[4*(p+q)+(p+q)-1][9]       4.56492  0.4960
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     1.375 0.500 2.000  0.2409
ARCH Lag[5]     2.380 1.440 1.667  0.3932
ARCH Lag[7]     4.308 2.315 1.543  0.3044

Nyblom stability test
------------------------------------
Joint Statistic:  16.3659
Individual Statistics:
ar2    0.1290
ar3    0.1517
ma2    0.1248
omega  2.8064
alpha1 0.2304
beta1  0.1820

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:       1.49 1.68 2.12
Individual Statistic:       0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.5690 0.5695
Negative Sign Bias  1.0972 0.2728
Positive Sign Bias  0.1434 0.8860
Joint Effect        1.2347 0.7447


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     35.13      0.01349
2    30     47.17      0.01789
3    40     56.08      0.03752
4    50     73.59      0.01308
```





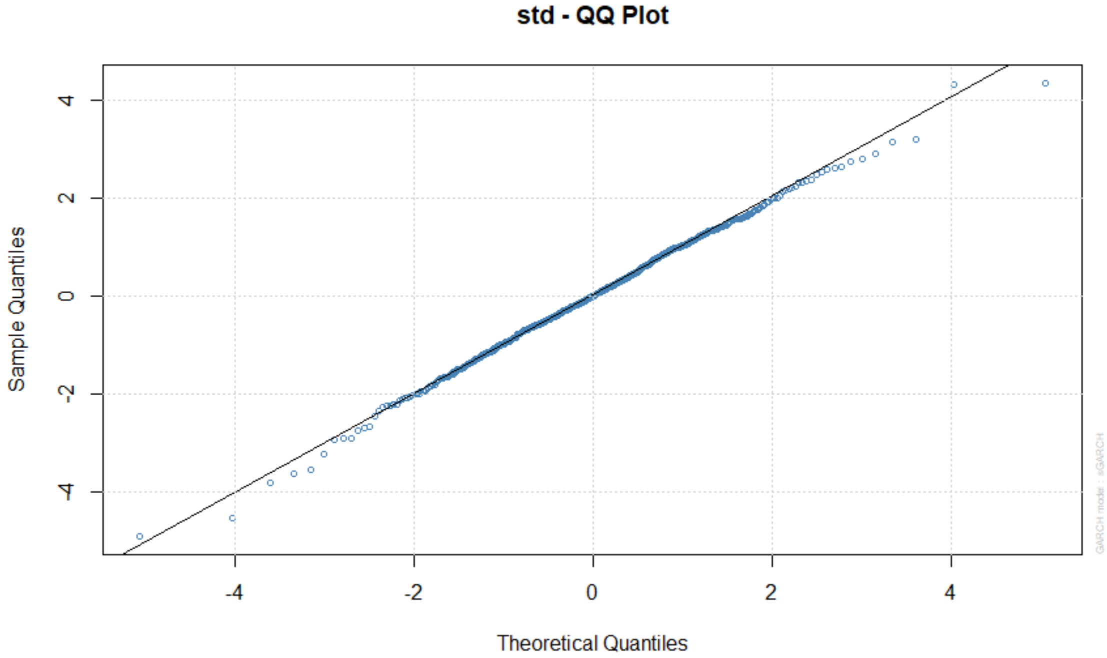
XOM ARMA-GARCH with Student-t

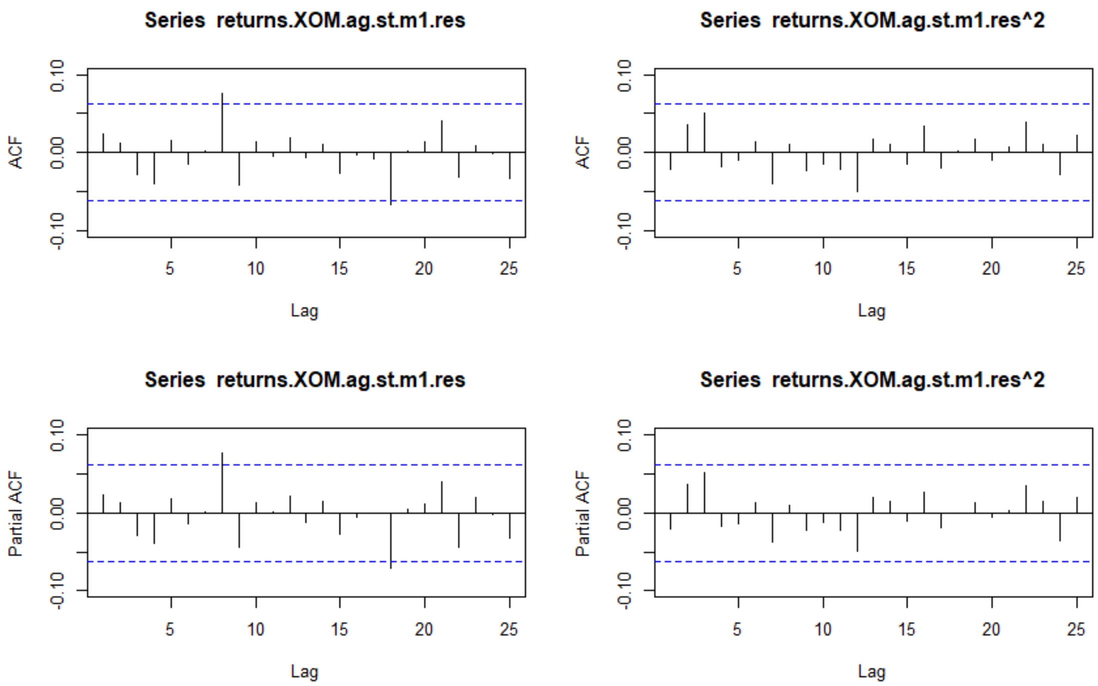
Checking model-fit

```
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|)
ar1     0.716930    0.029463   24.3329 0.000000
ar2     0.215451    0.034221    6.2958 0.000000
ma1    -0.783910    0.005509 -142.2954 0.000000
ma2    -0.156110    0.018099   -8.6255 0.000000
omega   0.000003    0.000003    1.1554 0.247938
alpha1  0.074141    0.020349    3.6434 0.000269
beta1   0.901481    0.026573   33.9247 0.000000
shape   5.525404    0.963017    5.7376 0.000000

Robust Standard Errors:
        Estimate  Std. Error    t value Pr(>|t|)
ar1     0.716930    0.028763   24.92534 0.000000
ar2     0.215451    0.030499    7.06424 0.000000
ma1    -0.783910    0.003889 -201.58634 0.000000
ma2    -0.156110    0.002410  -64.78366 0.000000
omega   0.000003    0.000008    0.43003 0.667174
alpha1  0.074141    0.030900    2.39934 0.016425
beta1   0.901481    0.055060   16.37268 0.000000
shape   5.525404    0.890285    6.20633 0.000000

LogLikelihood : 3196.959

Information Criteria
------------------------------------

Akaike       -6.3336
Bayes        -6.2945
Shibata      -6.3337
Hannan-Quinn -6.3187

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                         statistic p-value
Lag[1]                      0.5623  0.4533
Lag[2*(p+q)+(p+q)-1][11]    5.4196  0.8319
Lag[4*(p+q)+(p+q)-1][19]    8.8001  0.6717
d.o.f=4
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.4431  0.5057
Lag[2*(p+q)+(p+q)-1][5]    3.2172  0.3688
Lag[4*(p+q)+(p+q)-1][9]    4.6204  0.4872
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
```

```
ARCH Lag[3]     2.594 0.500 2.000  0.1073
ARCH Lag[5]     2.891 1.440 1.667  0.3060
ARCH Lag[7]     3.707 2.315 1.543  0.3908

Nyblom stability test
------------------------------------
Joint Statistic:  11.7716
Individual Statistics:
ar1    0.03188
ar2    0.04328
ma1    0.03171
ma2    0.04035
omega  1.43348
alpha1 0.21378
beta1  0.13913
shape  0.11241

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           1.3746 0.1696
Negative Sign Bias  0.3259 0.7446
Positive Sign Bias  0.1057 0.9158
Joint Effect        2.8813 0.4103


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     18.48       0.4905
2    30     22.21       0.8116
3    40     36.30       0.5938
4    50     42.50       0.7322
```



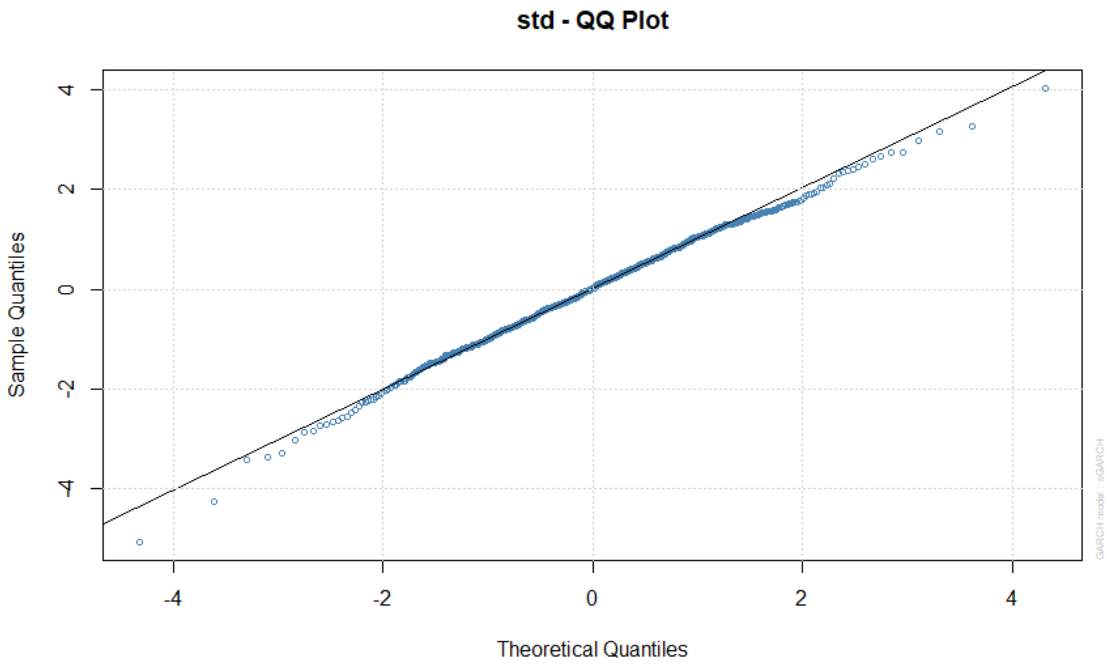


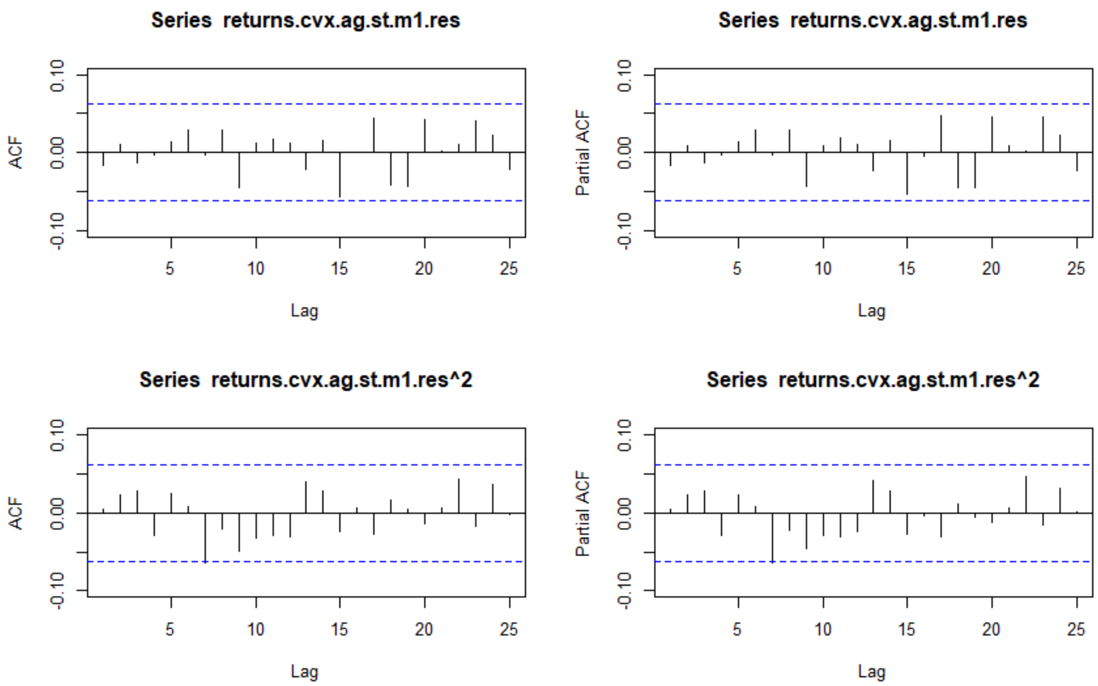
CVX ARMA-GARCH with Student-t

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model    : sGARCH(1,1)
Mean Model     : ARFIMA(3,0,2)
Distribution   : std
```

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.000000          NA       NA       NA
ar2    -0.539478    0.382588 -1.41007 0.158518
ar3     0.009453    0.027016  0.34989 0.726419
ma1     0.000000          NA       NA       NA
ma2     0.571637    0.370981  1.54088 0.123346
omega   0.000003    0.000005  0.57413 0.565878
alpha1  0.091699    0.048649  1.88491 0.059442
beta1   0.893458    0.055136 16.20457 0.000000
shape   8.318781    2.169568  3.83430 0.000126

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.000000          NA       NA       NA
ar2    -0.539478    0.307028 -1.75710 0.078901
ar3     0.009453    0.026977  0.35040 0.726038
ma1     0.000000          NA       NA       NA
ma2     0.571637    0.292528  1.95413 0.050686
omega   0.000003    0.000026  0.10540 0.916059
alpha1  0.091699    0.243740  0.37622 0.706754
beta1   0.893458    0.283809  3.14810 0.001643
shape   8.318781    5.681693  1.46414 0.143156

LogLikelihood : 3073.747

Information Criteria
------------------------------------

Akaike       -6.0909
Bayes        -6.0567
Shibata      -6.0910
Hannan-Quinn -6.0779

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                         statistic p-value
Lag[1]                      0.2487  0.6180
Lag[2*(p+q)+(p+q)-1][14]    2.7423  1.0000
Lag[4*(p+q)+(p+q)-1][24]    7.6184  0.9815
d.o.f=5
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                    0.01602  0.8993
Lag[2*(p+q)+(p+q)-1][5]   1.39571  0.7655
Lag[4*(p+q)+(p+q)-1][9]   3.82912  0.6179
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.8112 0.500 2.000  0.3678
ARCH Lag[5]    1.8420 1.440 1.667  0.5072
```

```
ARCH Lag[7]    3.7777 2.315 1.543  0.3798

Nyblom stability test
------------------------------------
Joint Statistic:  20.4362
Individual Statistics:
ar2    0.1095
ar3    0.1622
ma2    0.1037
omega  4.0255
alpha1 0.3444
beta1  0.2479
shape  0.1808

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:       1.69 1.9 2.35
Individual Statistic:       0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.5275 0.5980
Negative Sign Bias  0.8943 0.3714
Positive Sign Bias  0.2860 0.7749
Joint Effect        0.8943 0.8268


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     19.40      0.43176
2    30     43.06      0.04495
3    40     44.08      0.26533
4    50     58.00      0.17755
```



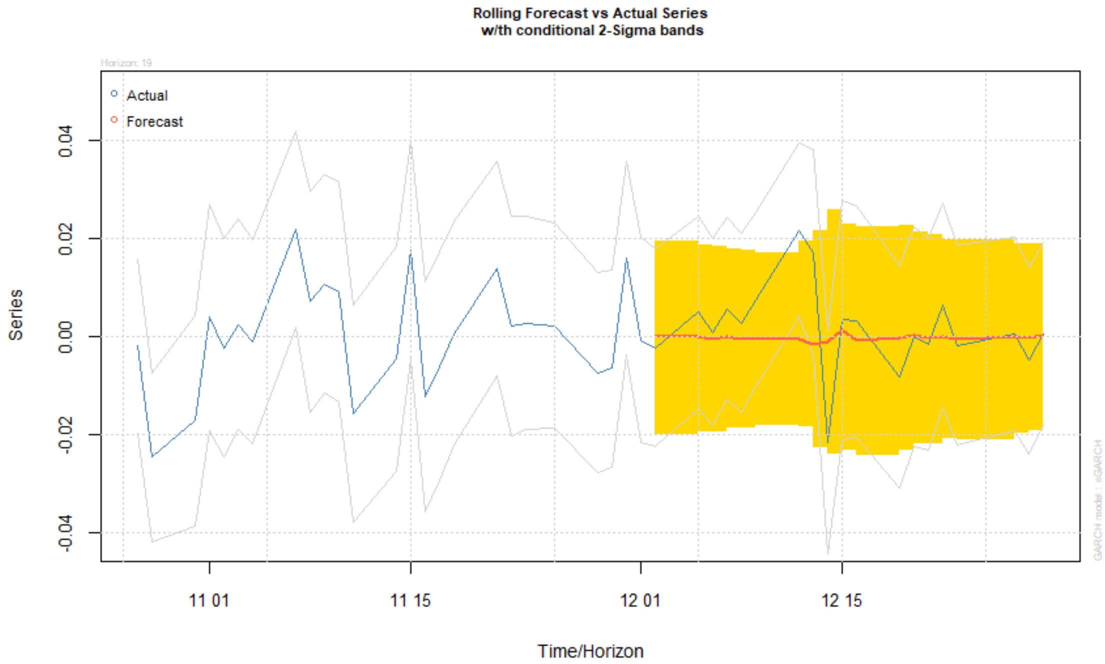


1.6 1-step ahead

<u>XOM</u>

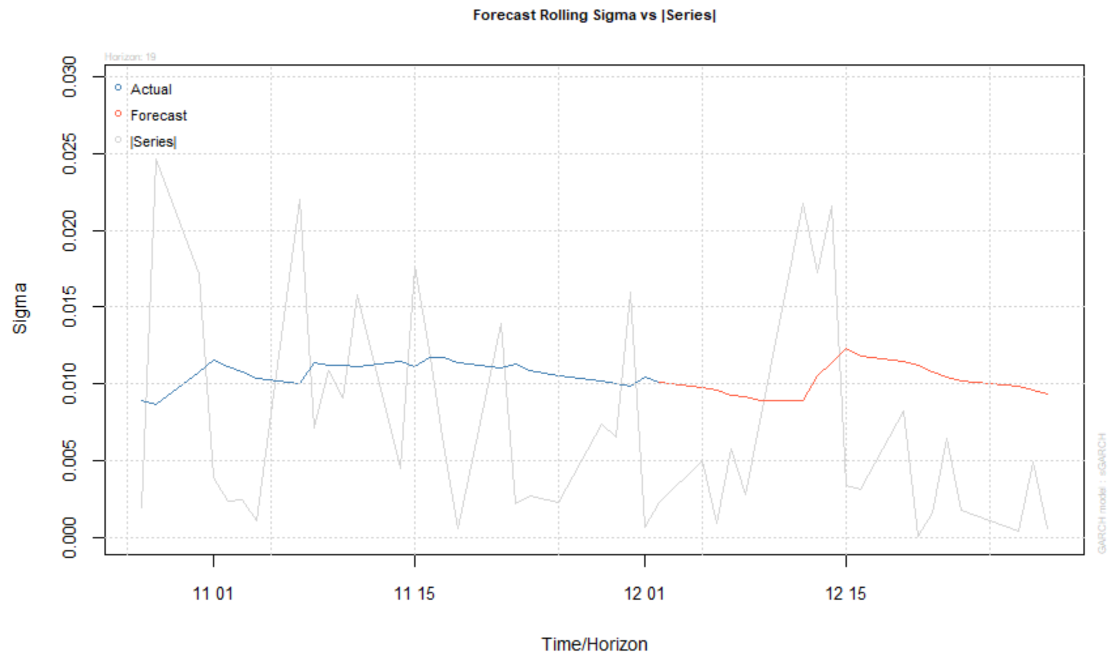
Returns Forecast

2016-12-01 2016-12-02 2016-12-05 2016-12-06 2016-12-07
2016-12-08 2016-12-09 2016-12-12
T+1 0.0001781358 3.043068e-05 -0.0004231653 -7.615014e-05 -
0.0004776686 -0.0002225285 -0.0006110808 -0.001546231
2016-12-13 2016-12-14 2016-12-15 2016-12-16 2016-12-
19 2016-12-20 2016-12-21 2016-12-22
T+1 -0.001161702 0.001359747 -0.0007204318 -0.0003024275
0.000397644 -0.0002604175 -1.67004e-06 -0.0005919288
2016-12-23 2016-12-27 2016-12-28 2016-12-29
T+1 6.074736e-05 -0.0001989777 0.0002268491 -0.0001984879



Standard Deviation Forecast

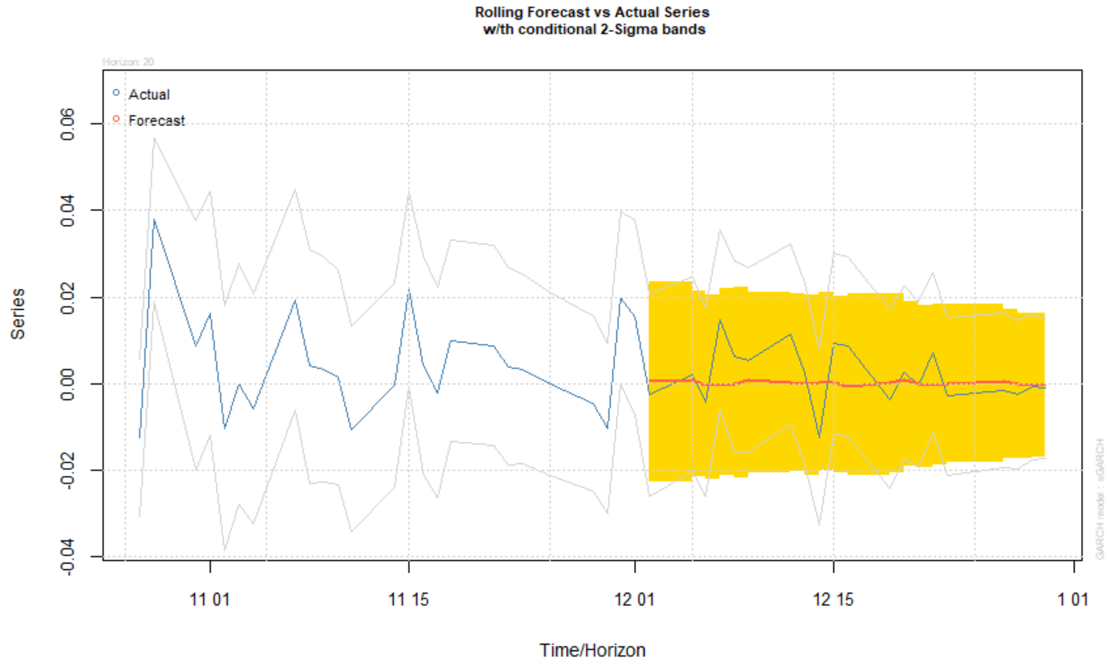
2016-12-01 2016-12-02 2016-12-05 2016-12-06 2016-12-07 2016-
12-08 2016-12-09 2016-12-12 2016-12-13 2016-12-14
T+1 0.01010693 0.009794167 0.0095735 0.009280448 0.009136271
0.008909965 0.008908097 0.01055334 0.01138303 0.01227789
2016-12-15 2016-12-16 2016-12-19 2016-12-20 2016-12-21 2016-
12-22 2016-12-23 2016-12-27 2016-12-28 2016-12-29
T+1 0.01182031 0.01142569 0.01121602 0.01081087 0.01043744
0.01023014 0.009891962 0.009571455 0.009358713 0.00907356



CVX

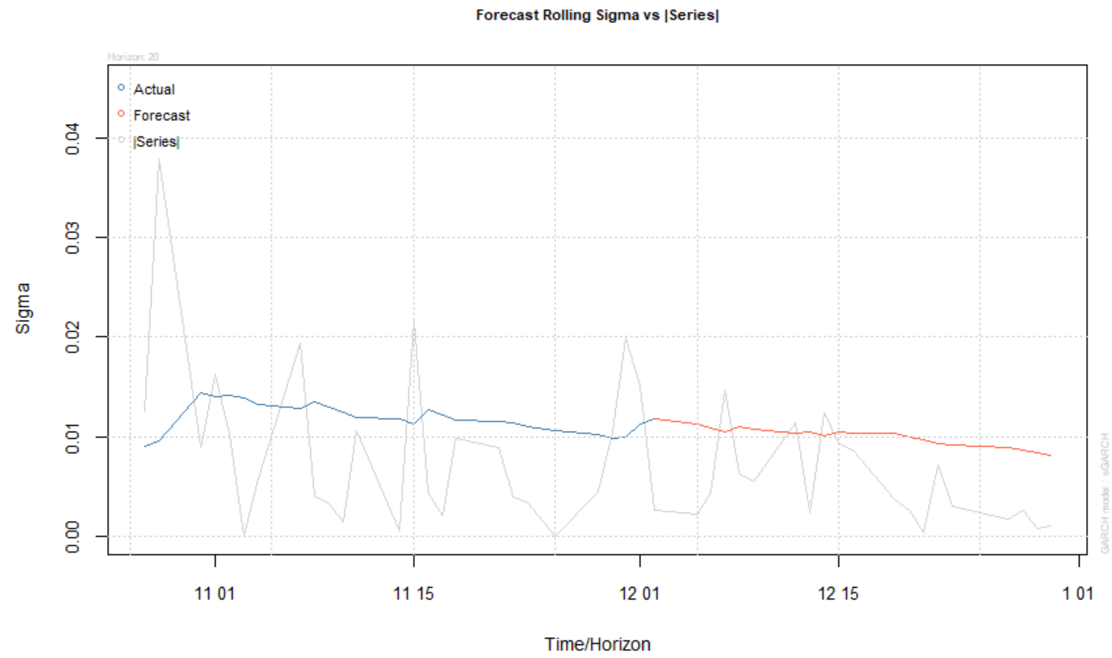
Returns Forecast

```
2016-12-01    2016-12-02     2016-12-05     2016-12-06    2016-12-07
2016-12-08    2016-12-09     2016-12-12
T+1 0.000674497 0.0009096267 -0.0003683026 -0.0004457999
6.334256e-05 0.0007576452 0.0002936765 -0.0001804874
      2016-12-13    2016-12-14     2016-12-15    2016-12-16    2016-12-
19     2016-12-20     2016-12-21    2016-12-22
T+1 0.0002902772 0.0002610344 -0.0006015486 0.0001105895
0.0007169806 -0.0001361864 -0.0003329547 8.081024e-05
      2016-12-23     2016-12-27     2016-12-28    2016-12-29    2016-
12-30
T+1 0.0004480239 -0.0001041342 -0.0003317101 -4.83271e-05
0.0001404408
```



Standard Deviation Forecast

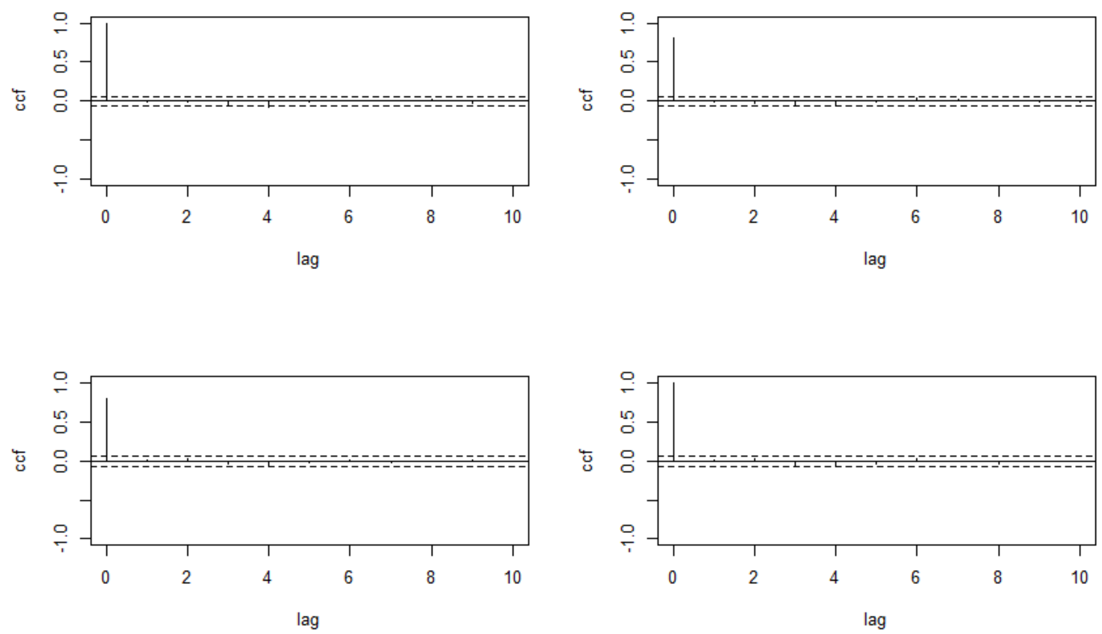
2016-12-01 2016-12-02 2016-12-05 2016-12-06 2016-12-07 2016-12-08 2016-12-09 2016-12-12 2016-12-13 2016-12-14
T+1 0.01174145 0.01128477 0.01082295 0.01044933 0.01099307 0.010705 0.01036949 0.01049313 0.01010188 0.01041718
    2016-12-15 2016-12-16 2016-12-19 2016-12-20 2016-12-21 2016-12-22 2016-12-23 2016-12-27 2016-12-28
T+1 0.01035694 0.01032402 0.009978941 0.009609869 0.009250572 0.009192765 0.008906573 0.008618344 0.008360259
     2016-12-29 2016-12-30
T+1 0.008088489 0.00784065



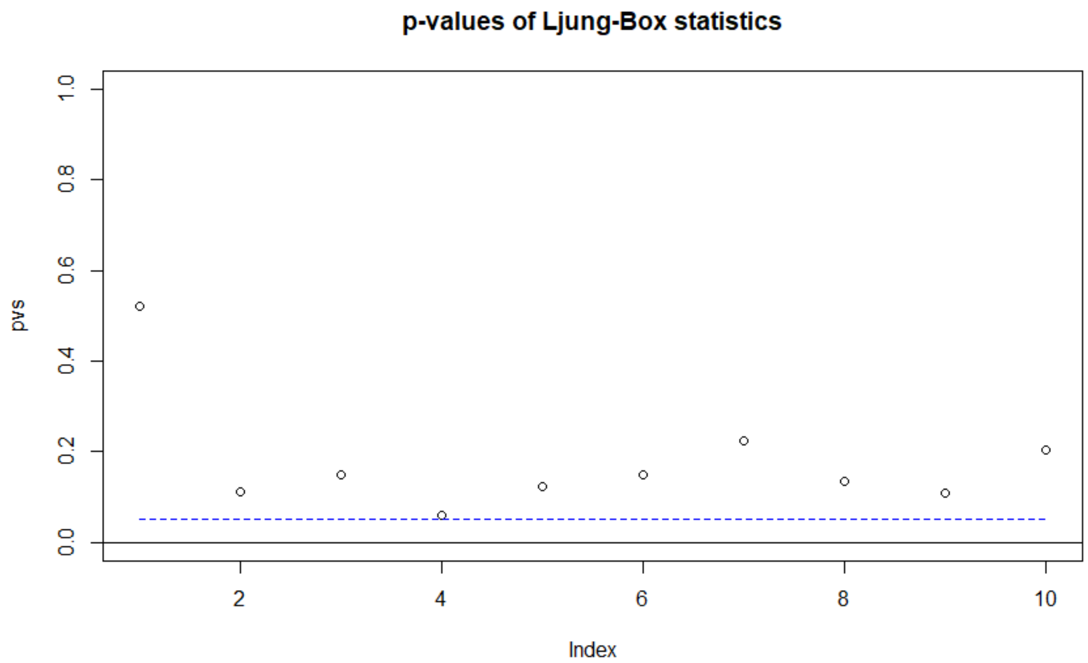
1.7 Cross-correlation

Obtaining the lags of the sample cross-correlation matrices of series using level = T will output values and simplified notation.

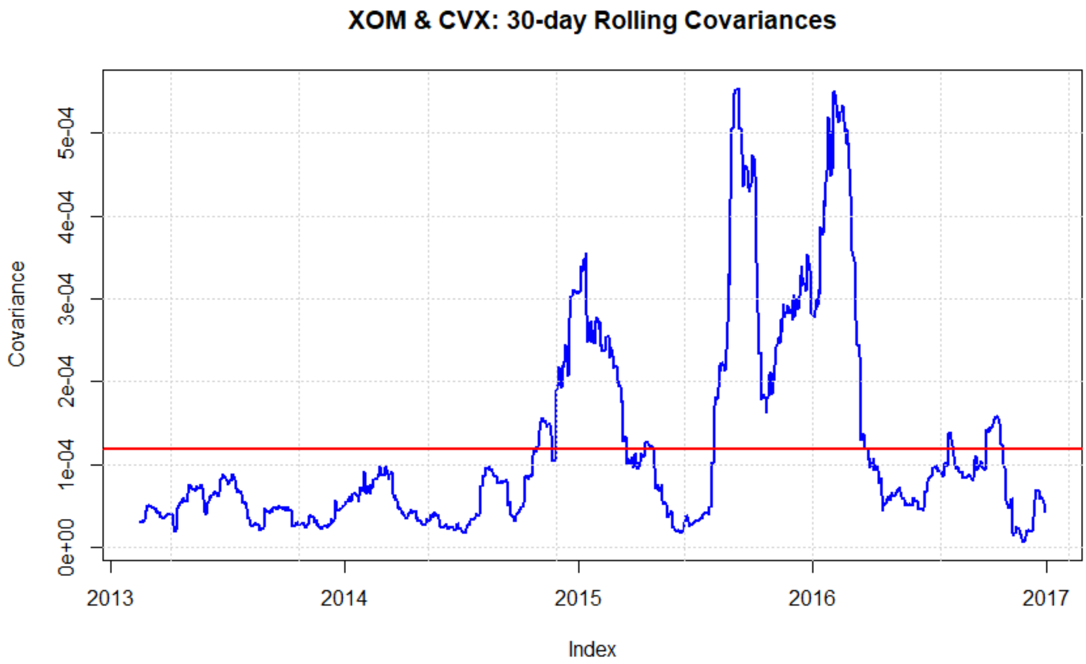
ACFs are on primary diagonal and CCFs are on off diagonal

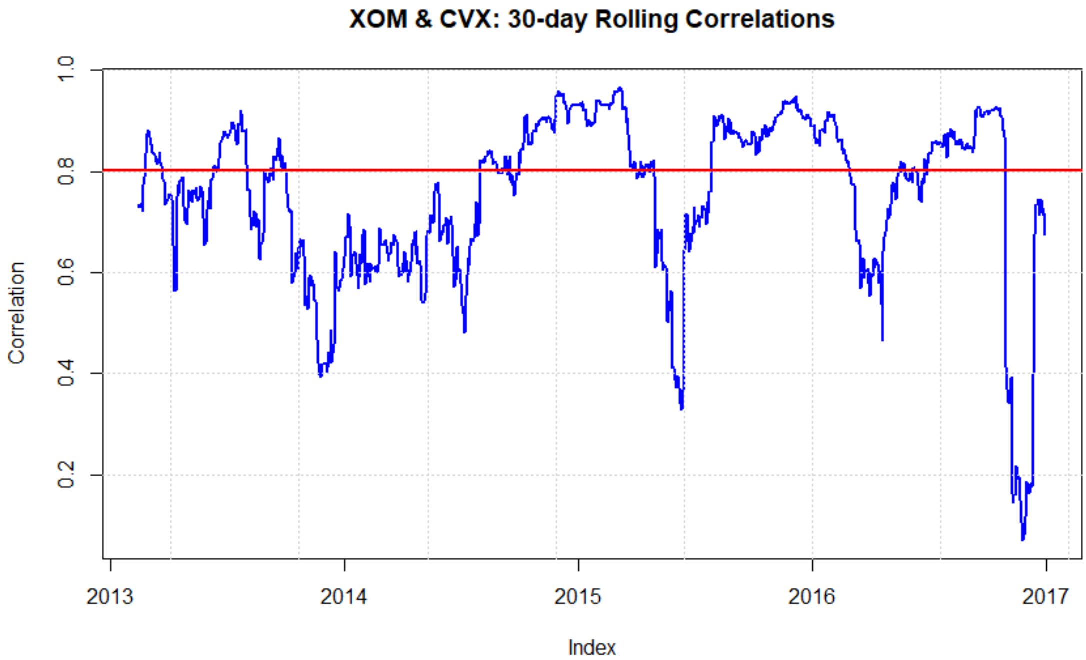


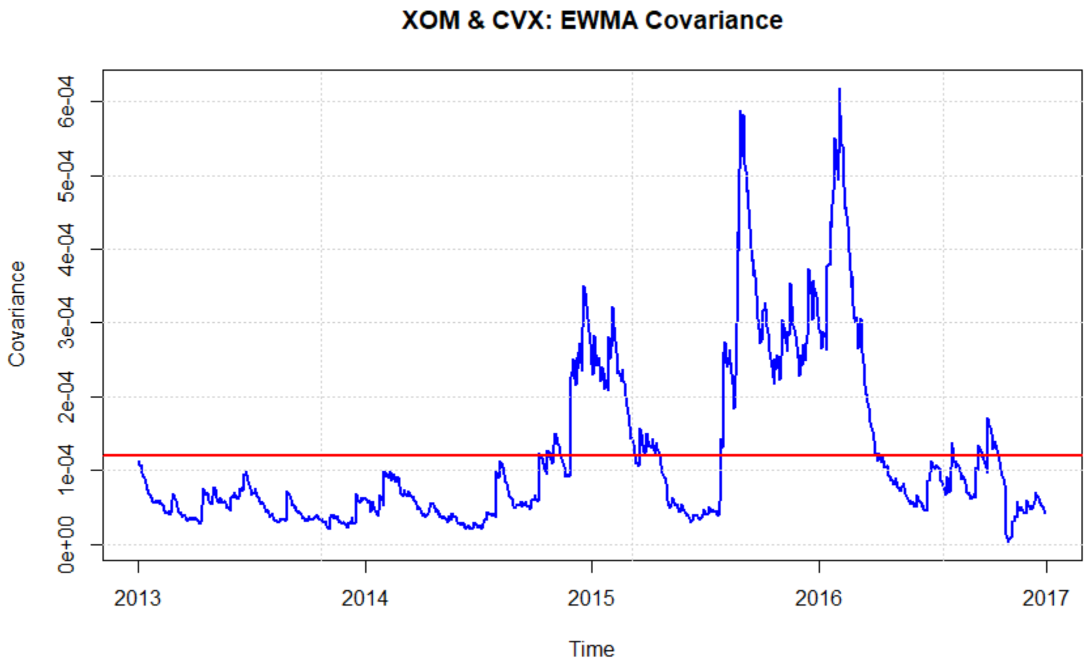
```
m,  Q(m)    and p-value:

[1] 1.00000 3.22994 0.52011
[1]  2.00000 13.01221  0.11143
[1]  3.00000 16.98059  0.15033
[1]  4.000000 25.645223  0.059229
[1]  5.00000 27.46108  0.12278
[1]  6.00000 31.14604  0.14962
[1]  7.0000 33.3443  0.2233
[1]  8.00000 40.94478  0.13353
[1]  9.00000 46.70983  0.10899
[1] 10.0000 47.1654  0.2029
```

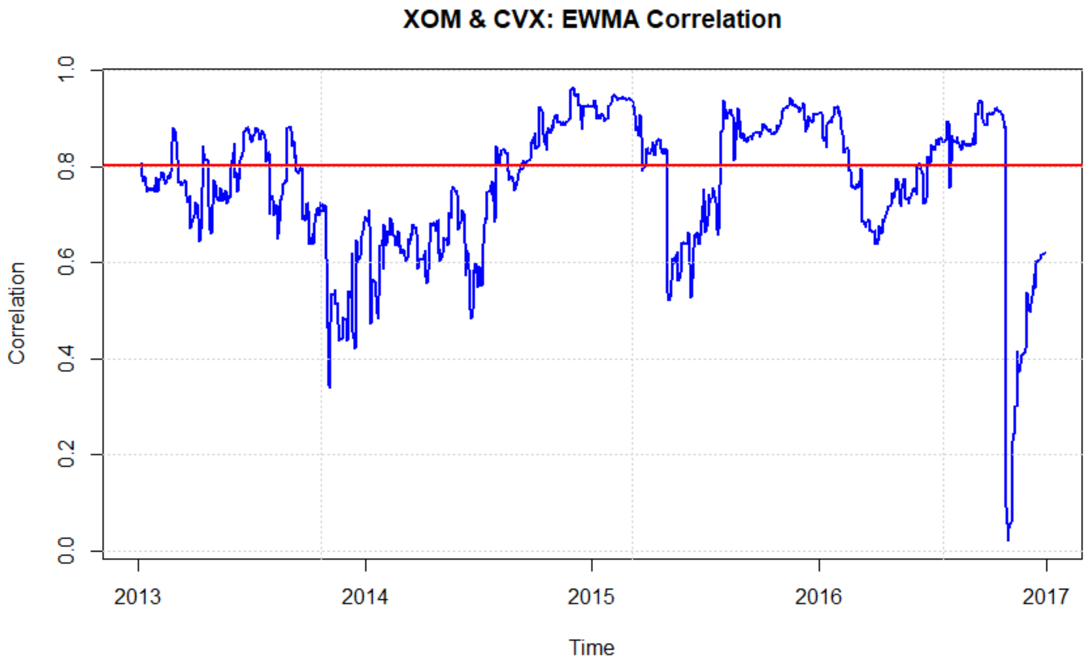


No, the Ljung-Box statistics of the residuals give Q(10) = 47.1654 with p-value 0.2029. Therefore, the null hypothesis of zero cross-correlation matrices cannot be rejected at the 5% level.


XOM & CVX: 30-day Rolling Covariances
Covariance
0e+00
1e-04
2e-04
3e-04
4e-04
5e-04
2013
2014
2015
2016
2017
Index


XOM & CVX: 30-day Rolling Correlations
Correlation
0.2
0.4
0.6
0.8
1.0
2013
2014
2015
2016
2017
Index





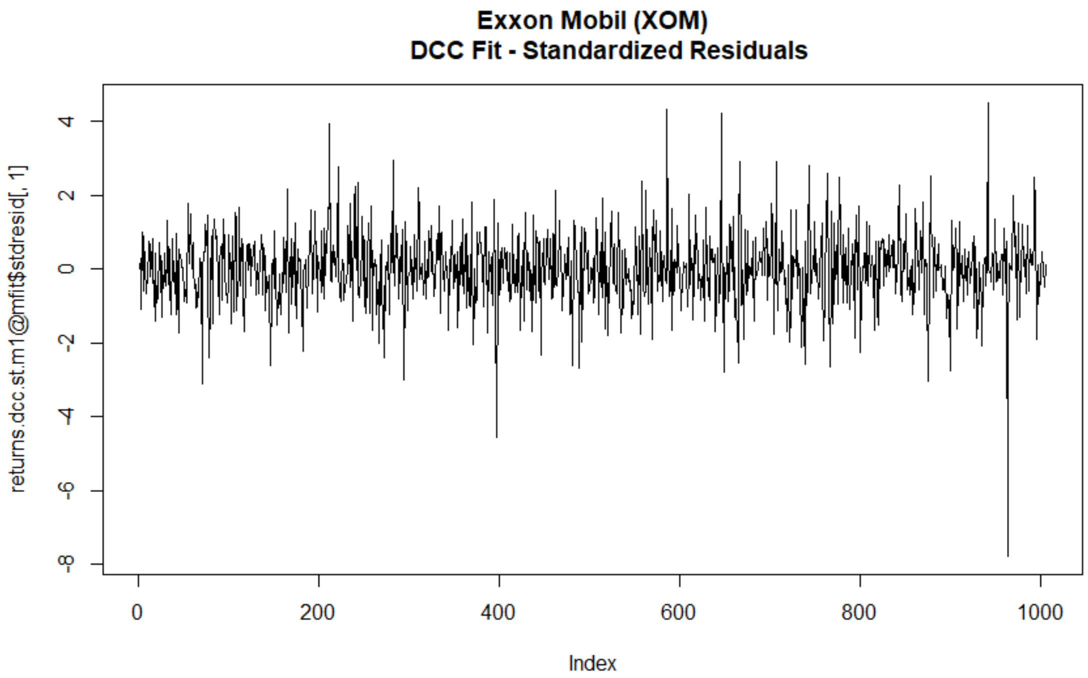
1.8 Normal-DCC(1,1) model

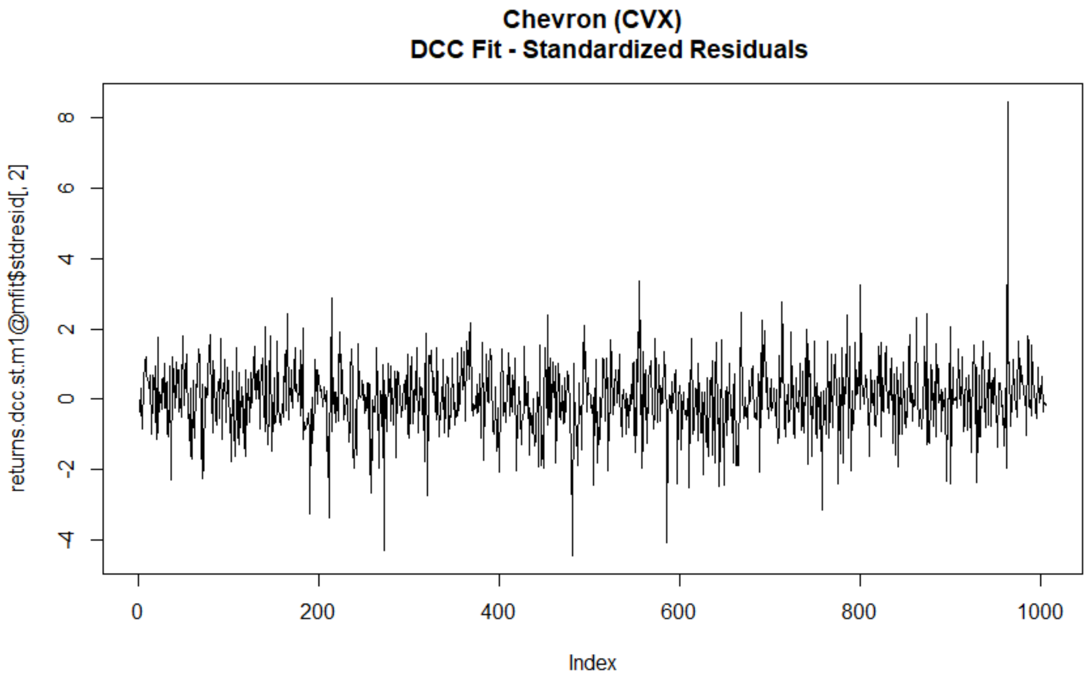
Examining model fit

```
Distribution          :  mvt
Model                 :  DCC(1,1)
No. Parameters        :  19
```

```
[VAR GARCH DCC UncQ] : [0+15+3+1]
No. Series           :  2
No. Obs.             :  1007
Log-Likelihood       :  6780.4
Av.Log-Likelihood    :  6.73

Optimal Parameters
-----------------------------------
               Estimate  Std. Error    t value Pr(>|t|)
[XOM].ar2     -0.336902    0.142078   -2.37124 0.017728
[XOM].ar3     -0.004375    0.027601   -0.15850 0.874060
[XOM].ma2      0.365645    0.136716    2.67449 0.007484
[XOM].omega    0.000003    0.000003    1.20258 0.229140
[XOM].alpha1   0.074980    0.021074    3.55801 0.000374
[XOM].beta1    0.900196    0.027726   32.46793 0.000000
[XOM].shape    5.564082    0.852299    6.52832 0.000000
[CVX].ar1      0.354861    0.008405   42.22138 0.000000
[CVX].ar2     -0.968072    0.007141 -135.56693 0.000000
[CVX].ma1     -0.357550    0.002083 -171.63859 0.000000
[CVX].ma2      0.991944    0.000222 4470.16334 0.000000
[CVX].omega    0.000003    0.000008    0.35756 0.720669
[CVX].alpha1   0.092002    0.083411    1.10299 0.270029
[CVX].beta1    0.892032    0.096508    9.24307 0.000000
[CVX].shape    7.929880    2.155822    3.67836 0.000235
[Joint]dcca1   0.018296    0.006907    2.64883 0.008077
[Joint]dccb1   0.969707    0.013185   73.54412 0.000000
[Joint]mshape  5.477195    0.532513   10.28556 0.000000

Information Criteria
---------------------

Akaike       -13.429
Bayes        -13.336
Shibata      -13.429
Hannan-Quinn -13.394
```

Plotting of standardized residuals

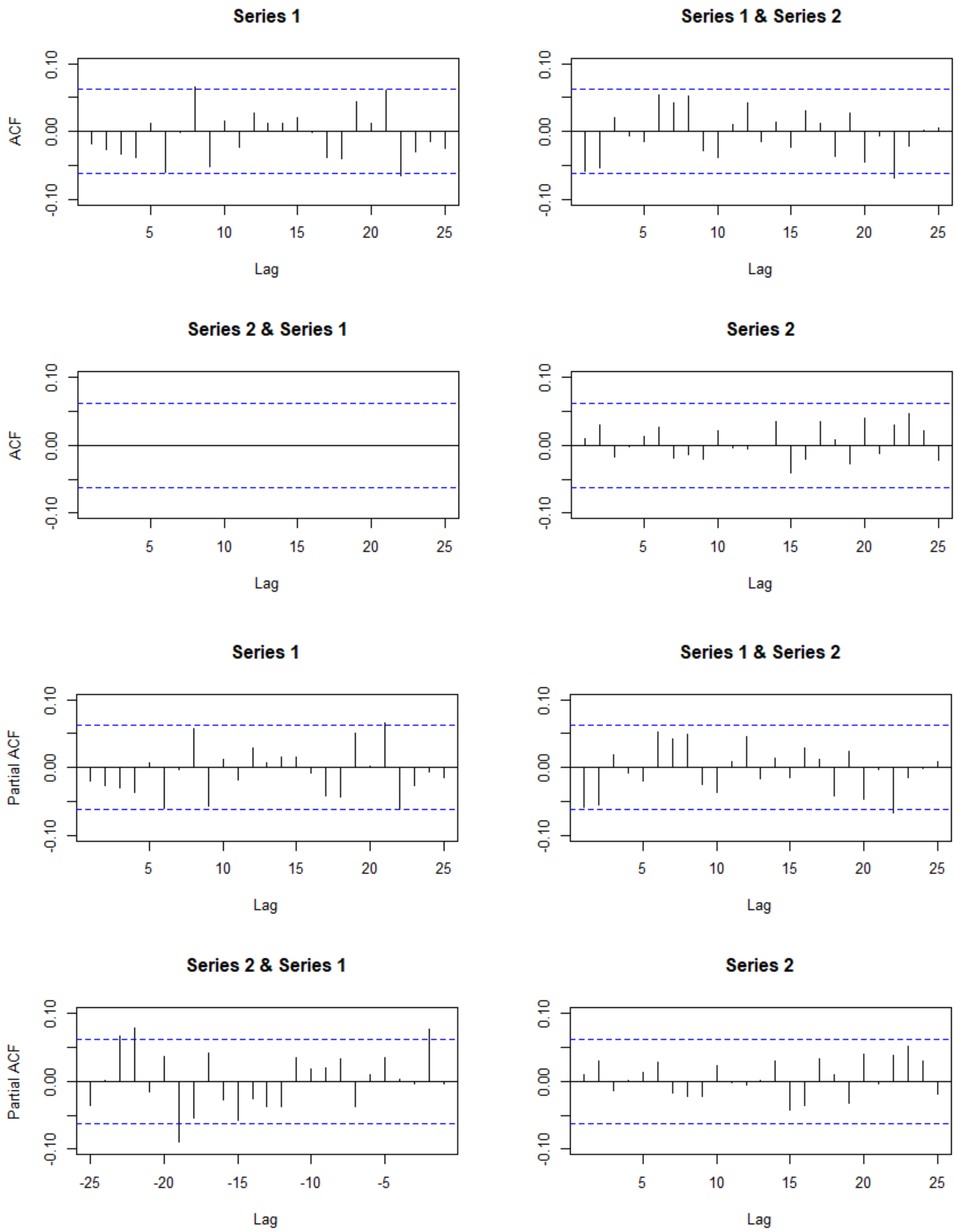


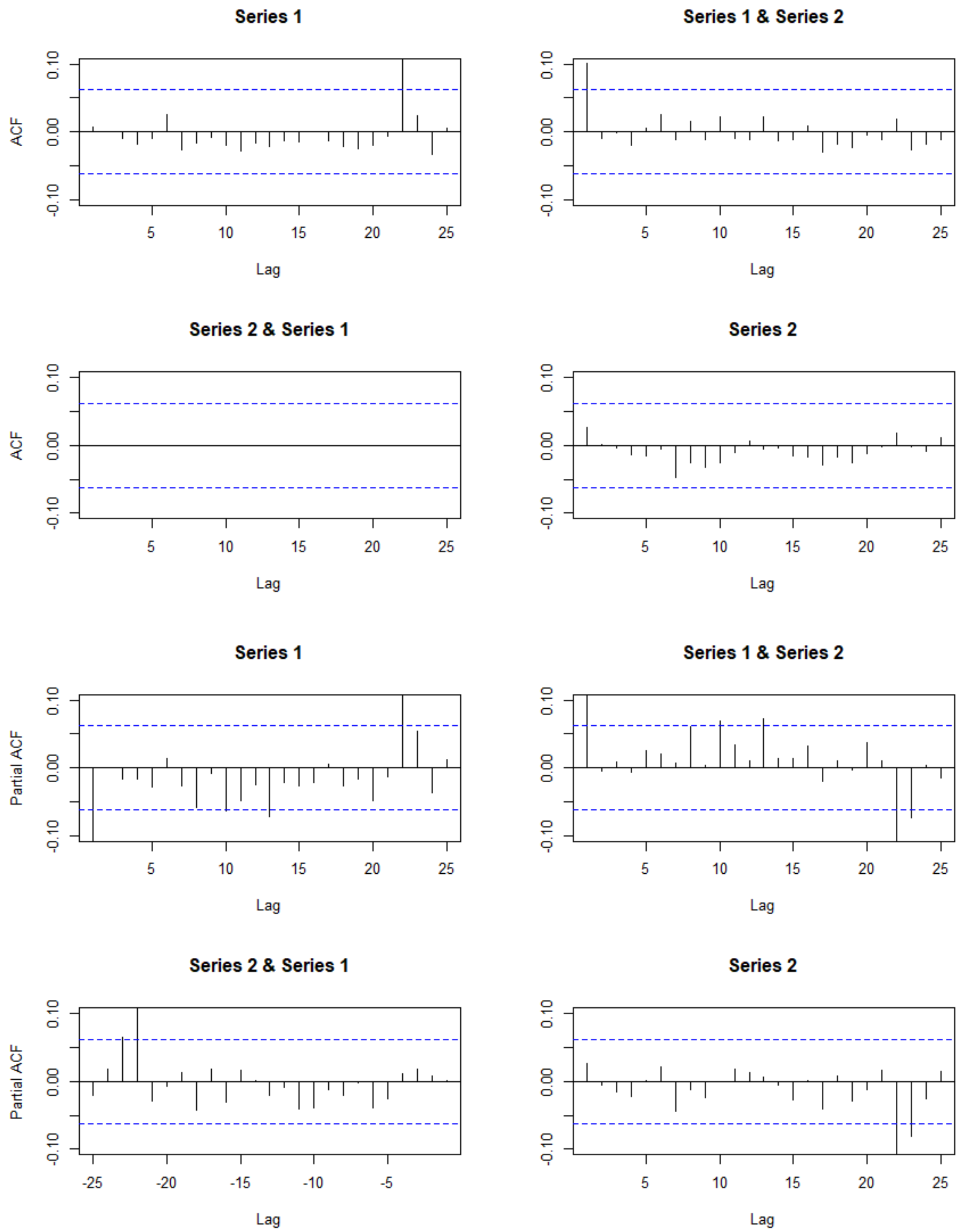


ACF & PACF

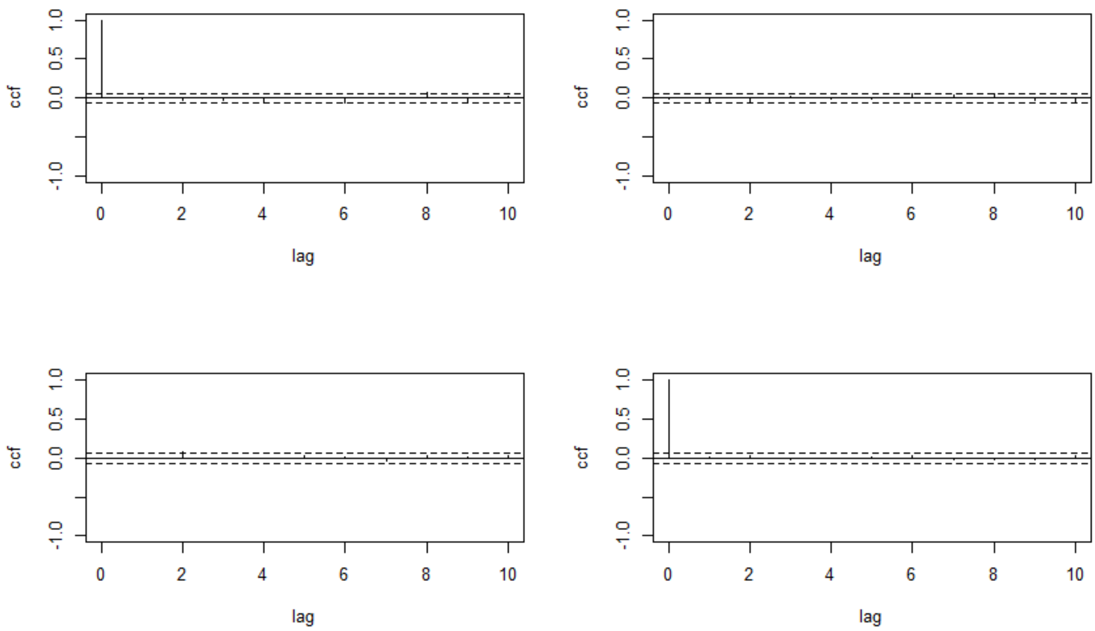
Standardized residuals – adequacy of model mean equation

Standardized residuals squared - adequacy of model variance equation





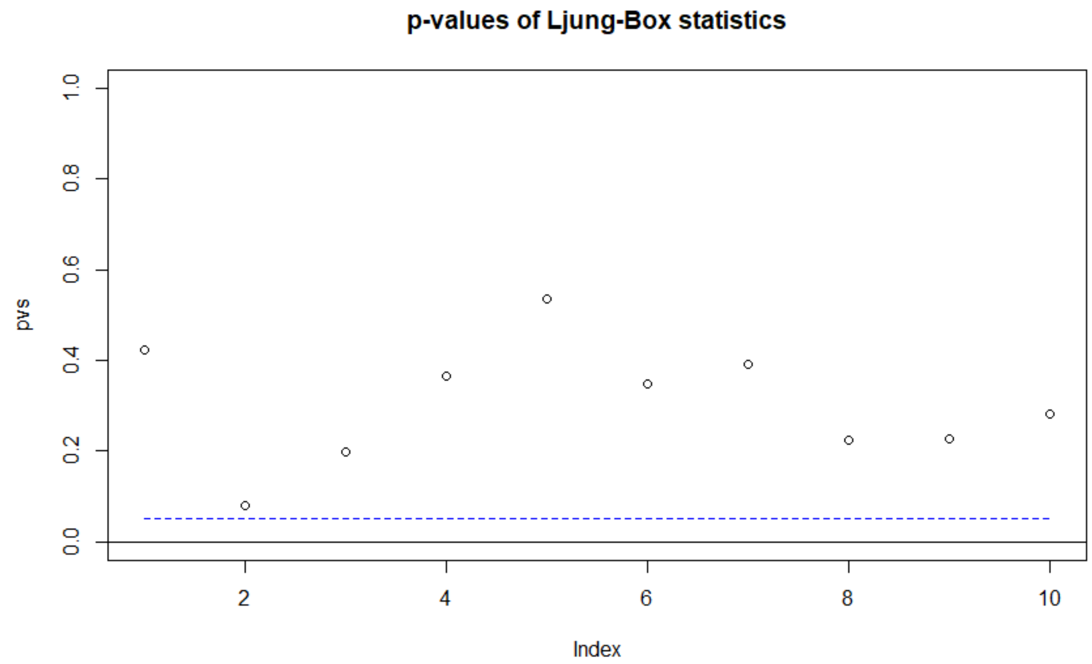
Obtain the lags of the sample cross-correlation matrices of series using level = T will output values and simplified notation. ACFs are on primary diagonal and CCFs are on off diagonal.



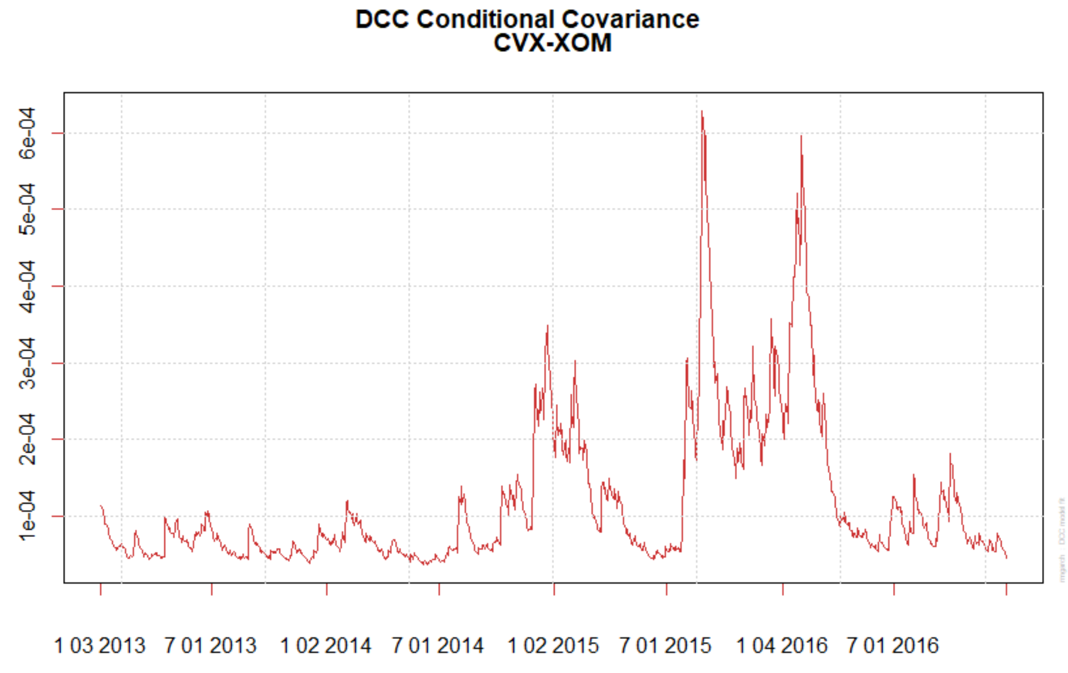
Testing the hypotheses using a 5% significance level;

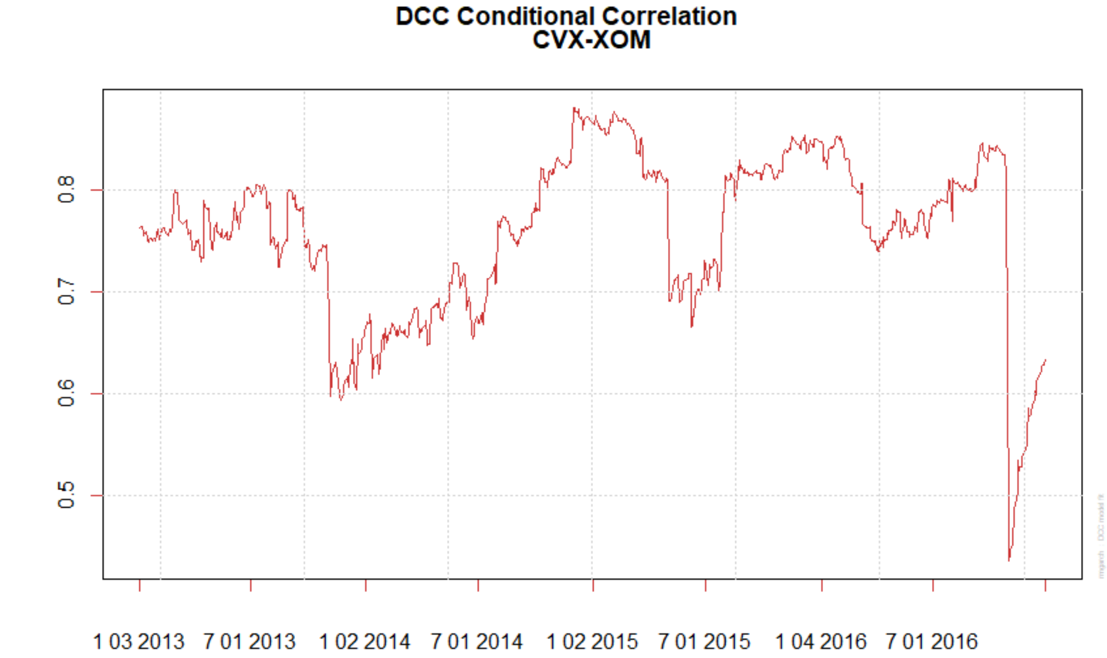
```
m,  Q(m)    and p-value:
[1] 1.00000 3.87512 0.42317
[1]  2.000000 14.090149  0.079446
[1]  3.00000 15.85678  0.19788
[1]  4.00000 17.32905  0.36462
[1]  5.00000 18.77845  0.53627
[1]  6.00000 26.08924  0.34865
[1]  7.00000 29.43764  0.39059
[1]  8.00000 37.74071  0.22334
[1]  9.00000 41.99833  0.22702
[1] 10.00000 44.69971  0.28093
```

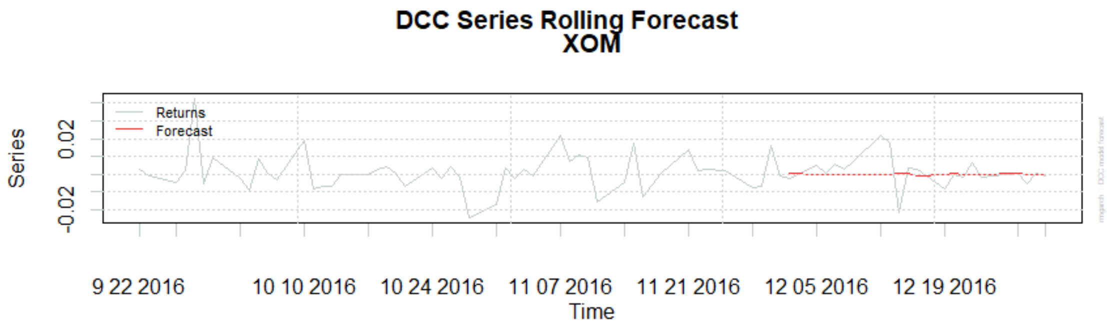
No, the Ljung-Box statistics of the residuals give Q(10) = 44.69971 with p-value 0.28093. Therefore, the null hypothesis of zero cross-correlation matrices cannot be rejected at the 5% level.

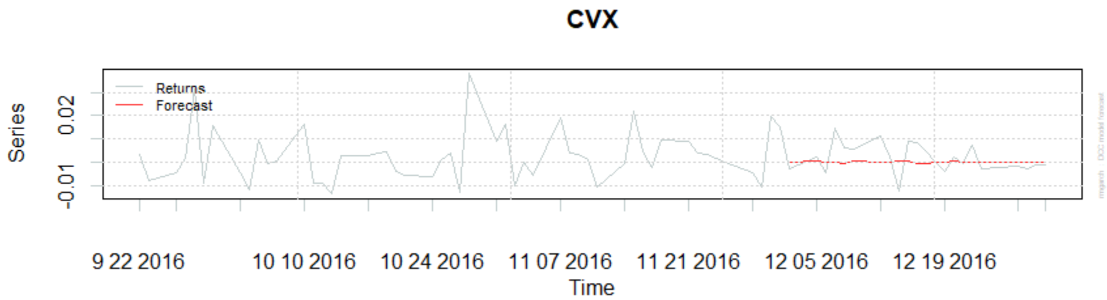


1.9 Conditional covariance and correlation

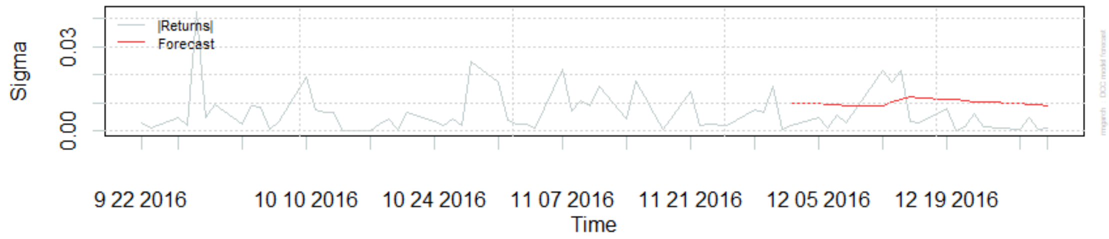



DCC Conditional Correlation
CVX-XOM
0.5
0.6
0.7
0.8
1 03 2013
7 01 2013
1 02 2014
7 01 2014
1 02 2015
7 01 2015
1 04 2016
7 01 2016

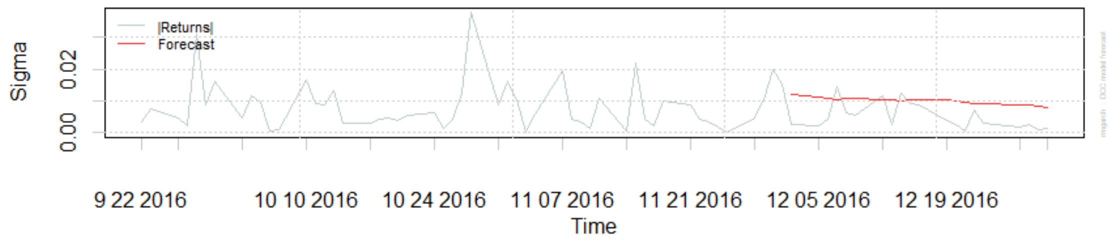

DCC Series Rolling Forecast
XOM
Returns
Forecast
Series
0.02
-0.02
9 22 2016
10 10 2016
10 24 2016
11 07 2016
11 21 2016
12 05 2016
12 19 2016
Time


CVX
Returns
Forecast
Series
0.02
-0.01
9 22 2016
10 10 2016
10 24 2016
11 07 2016
11 21 2016
12 05 2016
12 19 2016
Time

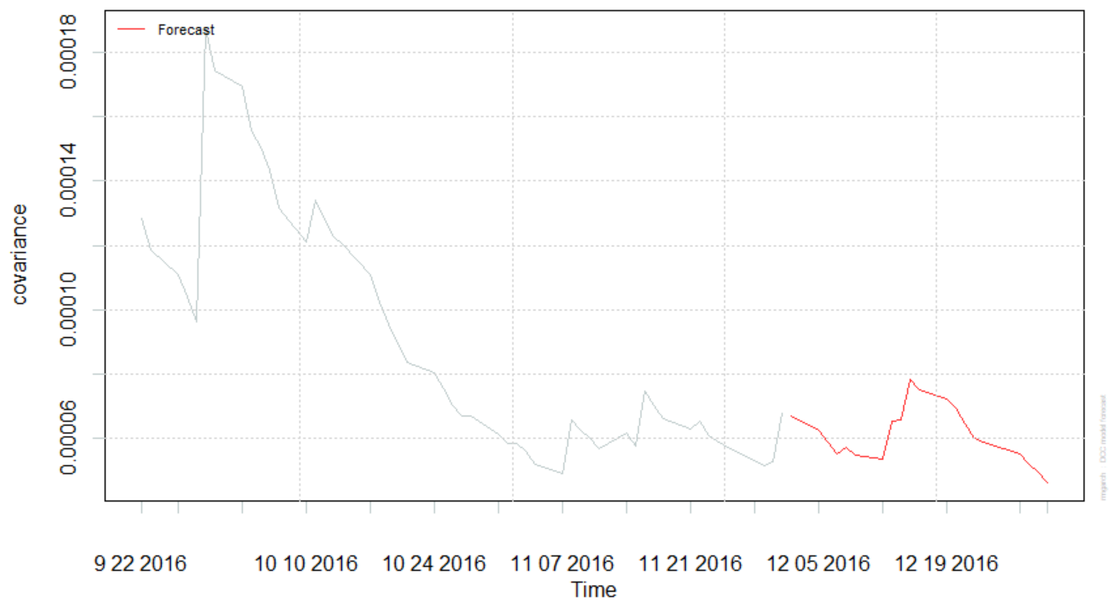
DCC Sigma Rolling Forecast
XOM

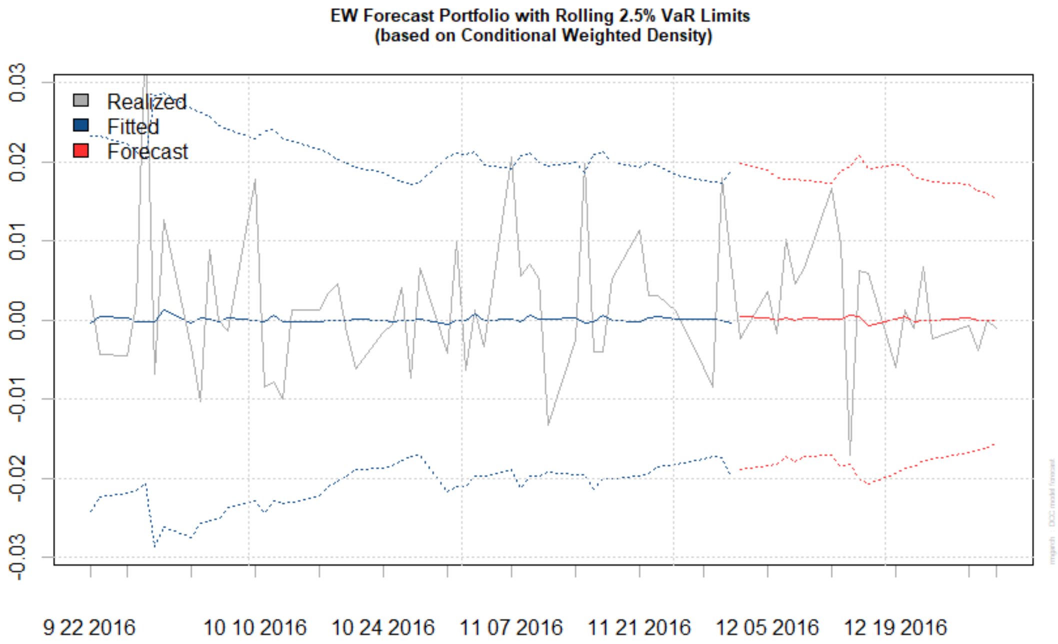


CVX



DCC Covariance Rolling Forecast
CVX-XOM





1.10 Report

XOM

```
MSE          MAE   DAC
1 7.797183e-05 0.005827557 0.55
```

CVX

```
MSE          MAE   DAC
1 4.349736e-05 0.005068049 0.55
```

	XOM Gaussian	**XOM ARMA-Student-t**	**CVX Gaussian**	**CVX ARMA-Student-t**
Akaike	-6.2898	-6.3336	-6.0640	-6.0909
Bayes	-6.2556	-6.2945	-6.0348	-6.0567
Shibata	-6.2899	-6.3337	-6.0641	-6.0910
Hannan-Quinn	-6.2768	-6.3187	-6.0529	-6.0779

For both stocks, ARMA-GARCH model with Student-t has the smallest values in AIC, BIC, Shibata and Hannan-Quinn criteria.

2 Return Surgeries

Introduction explains the motivation behind the analysis and lists the methodology of how the analysis will be studied.

2.2 Data sets can be expanded to include more basic summary of data, including lists of variables and their specifications including mean, standard deviations, minimum and maximum values.

2.2.2 Debrecen Data has a shorter overlapping data. AVVSO and STARA have same lengths of data - approximately three years from May 2010 to July 2013. However, Debrecen data ends in January 2011. Overlapping periods between these three data are barely over six months, from May 2010 to January 2011. The resulting analysis can thus be flawed.

Poisson regression models are limited because they assume events are independent. It may be useful in using generalized event count or GEC. if the Gaussian assumption does not hold, then likelihood function will be off and MLE is not reliable. OLS / ARIMA models use the wrong (Gaussian) distribution.

2.3 Autocorrelation Analysis

2.3.1 Descriptive Analysis

Descriptive analysis is not limited to matching three plots. It is important to identify patterns in correlated data, including trends and seasonal variation.

2.3.2 Autocorrelation Models

Since the first difference has been calculated, augmented Dickey Fuller regression can be applied to determine whether transformation has a unit root to check stationary.

3 Bonus, Analysis Report

4 Bonus, 3D VAR(2) Model

5 Bonus, Normal Distribution

A large number of statistical tests are based on the assumption of normality. This also applies to hypotheses to check confidence level. Another reason in desiring Gaussian distribution is the simplicity of the first order equations and desirable properties of the estimated parameters that are consistent, efficient, and asymptotically normal.

6 Bonus, “Best Model”

If all variables are stationary, then the basic model selection rules apply. That is adjusted r-squared, Akaike Information Criteria, BIC among others. The behavior of dependent variable, whether lagged values should be included in case of ADF models or if granger causality tests are needed between variables. Additionally, misspecification test are required.