TIME SERIES PROJECT

November 16th, 2020

Baltzi, Sofia
sbaltzi@aueb.gr

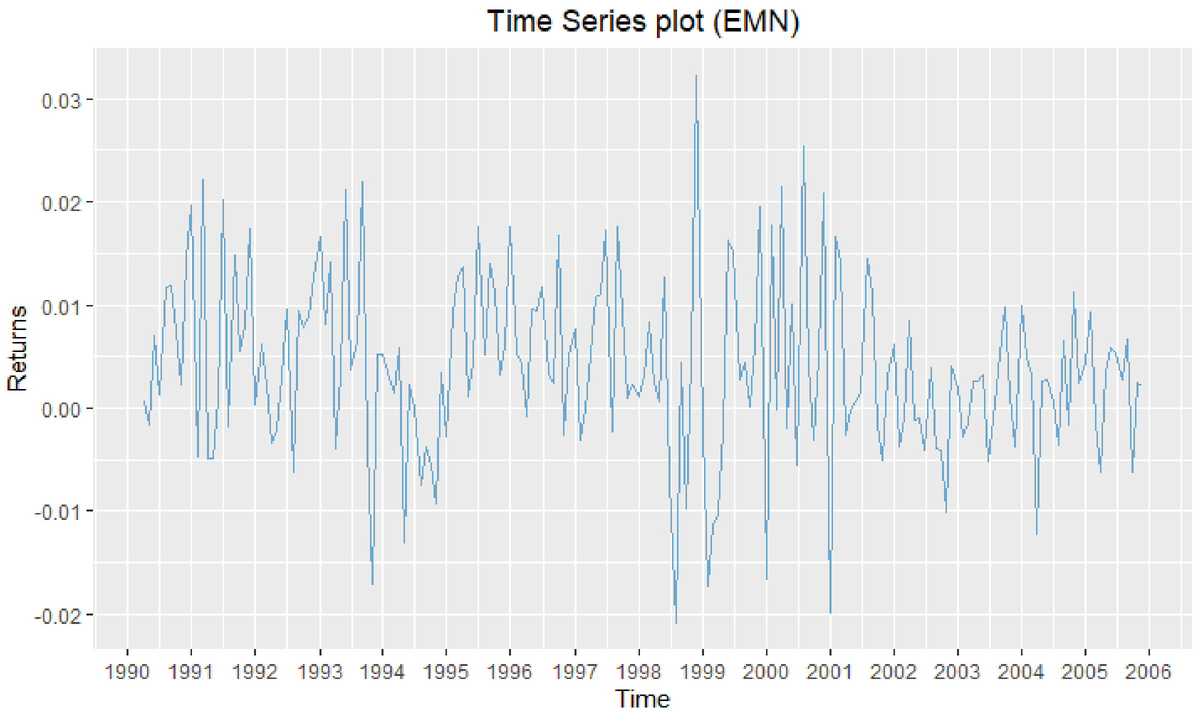
Contents

Stock EMN

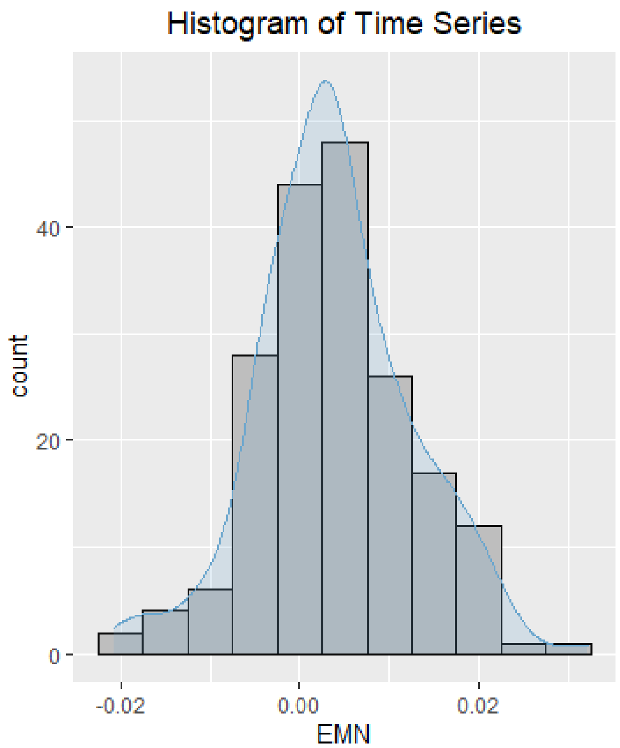
Question 1

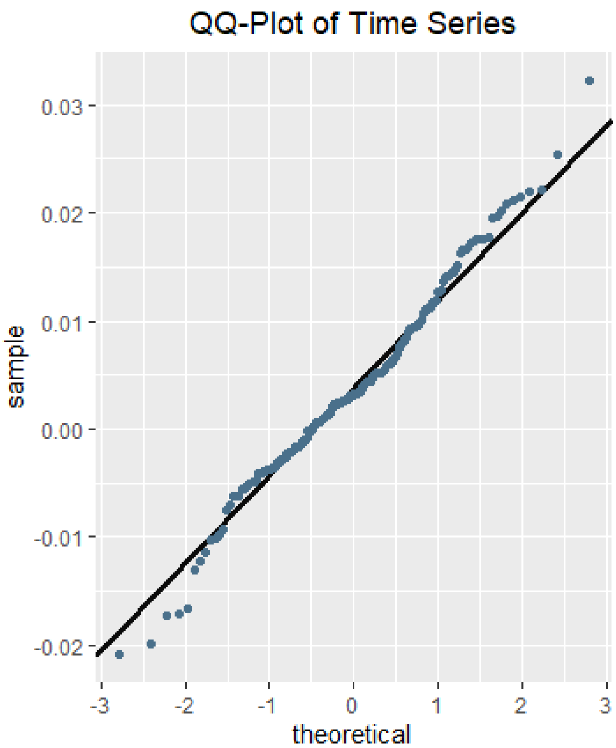
Identification Step

The time series analyzed in this section are the Returns of EMN Stock. As it can be observed, the series seem stationary but with a problem of heteroscedasticity. There does seem to be a drift but no trend.



The histogram of the observations shows that they, somewhat, follow normal distribution. Adding to that, is the QQ-plot, which also shows normal observations with existence of outliers.





In order to verify our claims a series of tests are performed:

a. *Shapiro-Wilk test for Normality*

The Null Hypothesis for this test is that the observations are drawn from a normal distribution (against the alternative Hypothesis that they are not drawn from a normal distribution). Since p-value = 0.063 which is larger than the confidence level a = 0.05, we fail to reject the Null Hypothesis and thus accept the **normality of the observations**.

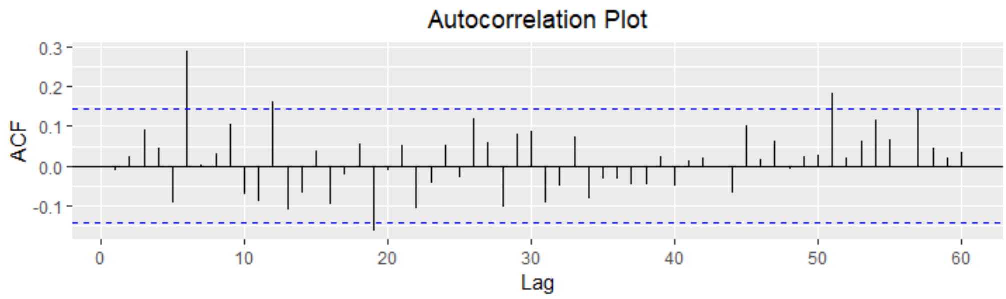
```
> shapiro.test(ts1$EMN)

        Shapiro-Wilk normality test

data:  ts1$EMN
W = 0.98632, p-value = 0.06377
```

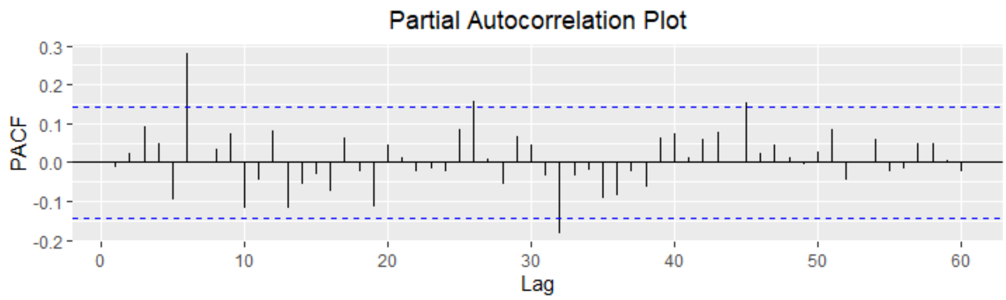
b. *Box-Pierce / Ljung-Box for Autocorrelation*

The Null Hypothesis for these tests is that the observations are correlated (against the alternative Hypothesis that they are not). The tests are performed for different values of lag (6, 10) and all give p-values bellow the confidence level a = 0.05, leading us to reject Null Hypothesis and assume **autocorrelation between observations**. The ACF and PACF plot also show possible autocorrelations.

```
> Box.test(ts1$EMN, 6, type="Box-Pierce")

        Box-Pierce test

data:  ts1$EMN
X-squared = 19.376, df = 6, p-value = 0.003574

> Box.test(ts1$EMN, 6, type="Ljung-Box")

        Box-Ljung test

data:  ts1$EMN
X-squared = 20.181, df = 6, p-value = 0.002571
```

```
> Box.test(ts1$EMN, 10, type="Box-Pierce")

        Box-Pierce test

data:  ts1$EMN
X-squared = 22.59, df = 10, p-value = 0.01237

> Box.test(ts1$EMN, 10, type="Ljung-Box")

        Box-Ljung test

data:  ts1$EMN
X-squared = 23.596, df = 10, p-value = 0.008748
```





* ACF, PACF plots do not show correlation with lag 0

c. Augmented Dickey-Fuller test for Stationarity

The Null Hypothesis for this test is that the series is not stationary (against the alternative that it is). The test was performed assuming drift of the series and setting the lags to 5. Since the test statistic (-4.01) is below tau2 at confidence level a = 0.05 (-2.88) we reject the Null Hypothesis and thus, we can consider the series **stationary**.

```
Value of test-statistic is: -4.0124 8.065

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

Estimation Step

Based on the AIC criterion the model that was found to best fit the data (to give the smallest value of AIC) is a **constrained AR(6)** model:

$$y_t = 0.0039 + 0.2854 y_{t-6} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

```
> ar6

Call:
arima(x = tser, order = c(6, 0, 0), fixed = c(0, 0, 0, 0, 0, NA, NA))

Coefficients:
        ar1  ar2  ar3  ar4  ar5     ar6  intercept
          0    0    0    0    0  0.2854     0.0039
s.e.      0    0    0    0    0  0.0690     0.0008

sigma^2 estimated as 6.85e-05:  log likelihood = 637.7,  aic = -1269.4
```

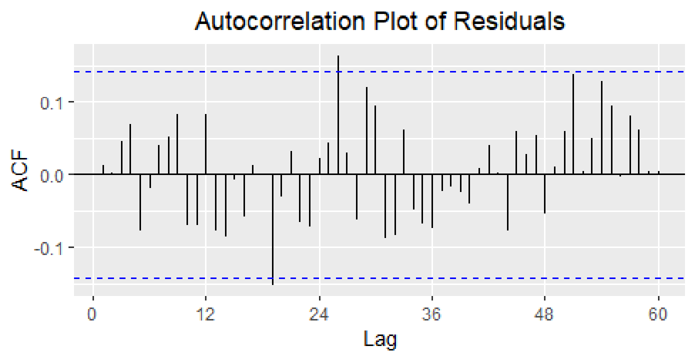
Diagnostic Plots

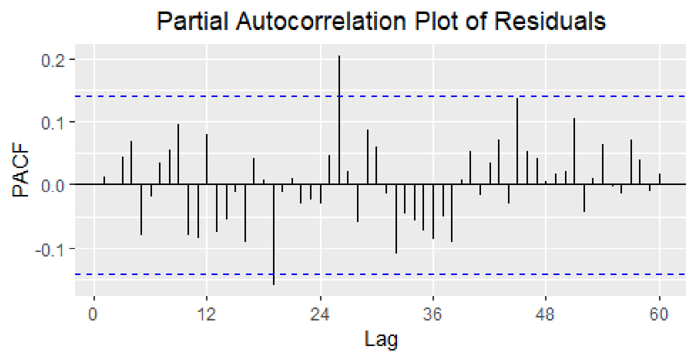
From the ACF and PACF plots of residuals we can see that the problem of **autocorrelation for the first lags is resolved**. There are still some spikes out of the boundaries but since they are on large lags (26) they are ignored.

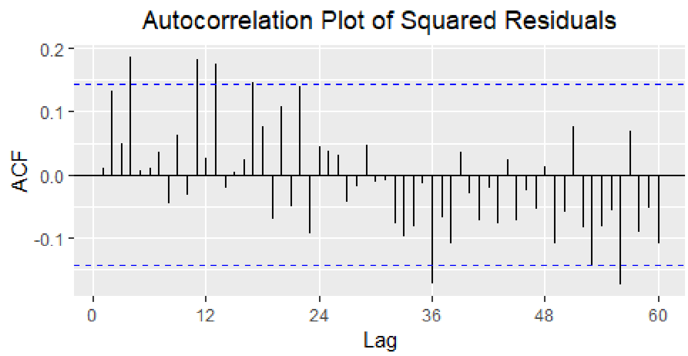
On the other hand, the ACF and PACF of squared residuals show quite a lot of spikes outside the boundaries leading us to assume **heteroscedasticity problems**.

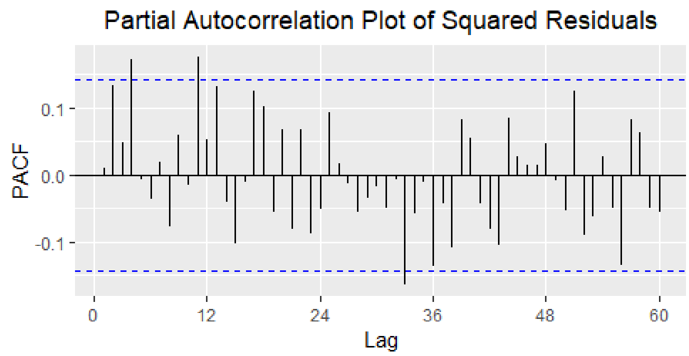
From the QQ-plot we can see that **residuals are mostly normal** with few outliers.

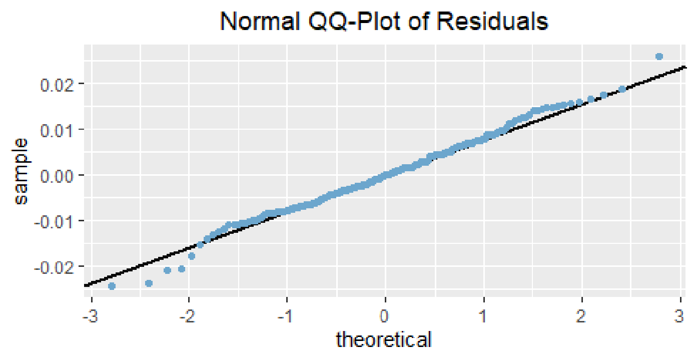
Finally, we can see the forecast for the next 20 instances. Obviously after forecasting for 6 instances the model draws the observation it needs (y_{t-6}) from our forecasted values.

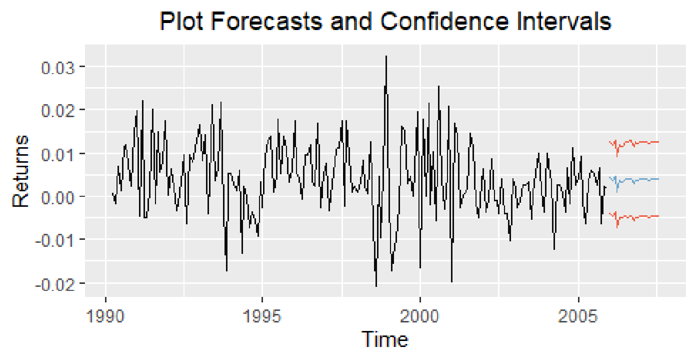












The claims made are verified from the same tests for normality of residuals (Shapiro-Wilk), autocorrelation of residuals (Ljung-Box) and autocorrelation of squared residuals (heteroscedasticity).

As expected, we do not reject Null Hypothesis for normality and autocorrelation (meaning there is not autocorrelation for 60 lags).

We do reject the Null Hypothesis for heteroscedasticity though.

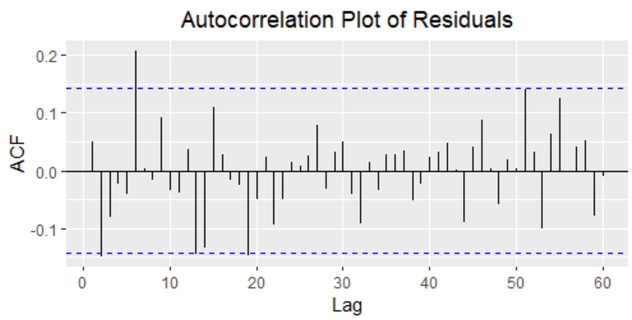
```
> shapiro.test(residuals)

        Shapiro-Wilk normality test

data:  residuals
W = 0.99199, p-value = 0.3835

> # Autocorrelation
> Box.test(residuals, 60, type="Ljung-Box")

        Box-Ljung test

data:  residuals
X-squared = 60.653, df = 60, p-value = 0.4521

> # Heteroscedasticity
> Box.test(residuals^2, 60, type="Ljung-Box")

        Box-Ljung test

data:  residuals^2
X-squared = 96.914, df = 60, p-value = 0.001794
```

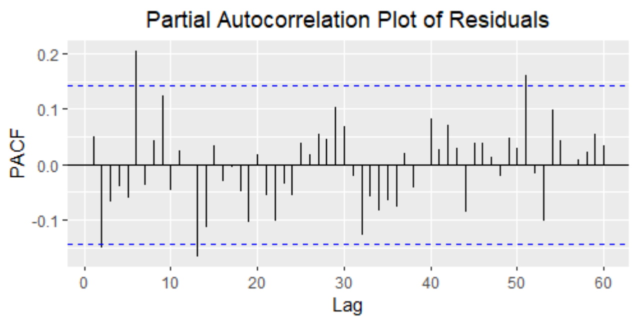
Question 2

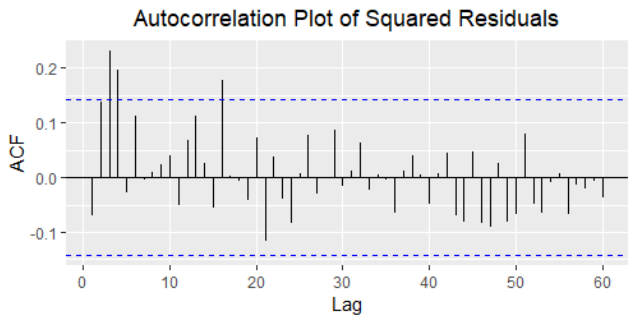
For this question, a regression model was applied to the data using all independent variables (x_1-x_{15}). Using the AIC criterion this model was gradually reduced to having only x_1, x_5, x_6, x_7, x_8, x_{11}, x_{13}, x_{15} to predict the returns of EMN Stock:

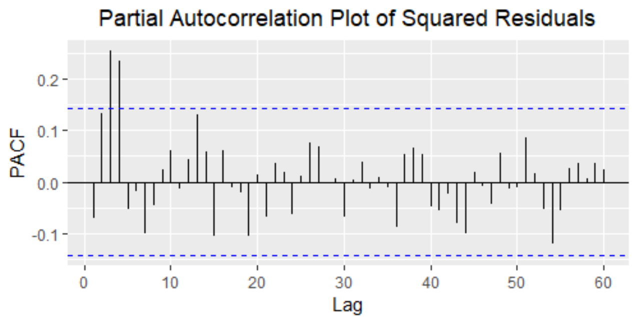
$$y_t = -0.001342 + 0.058918x_{1,t} + 0.080913x_{5,t} + 0.041066x_{6,t} + 0.071914x_{7,t} + 0.074959x_{8,t} \\ + 0.729843x_{11,t} + 0.013942x_{13,t} + 1.152034x_{15,t} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

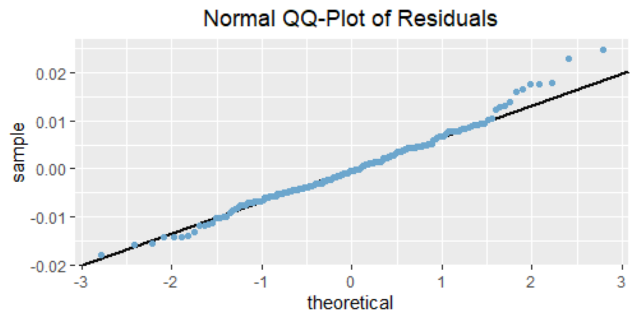
```
> model$anova
Stepwise Model Path 
Analysis of Deviance Table

Initial Model:
y1 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + 
    x12 + x13 + x14 + x15

Final Model:
y1 ~ x1 + x5 + x6 + x7 + x8 + x11 + x13 + x15


   Step Df     Deviance Resid. Df  Resid. Dev       AIC
1                             173 0.009688440 -1835.050
2 - x12  1 5.251822e-07       174 0.009688965 -1837.039
3  - x3  1 2.152096e-06       175 0.009691117 -1838.997
4 - x10  1 3.540703e-06       176 0.009694658 -1840.928
5  - x9  1 7.204559e-06       177 0.009701862 -1842.788
6 - x14  1 9.739279e-06       178 0.009711601 -1844.598
7  - x2  1 7.400387e-05       179 0.009785605 -1845.163
8  - x4  1 8.889627e-05       180 0.009874502 -1845.454
> model

Call:
lm(formula = y1 ~ x1 + x5 + x6 + x7 + x8 + x11 + x13 + x15, data = data)

Coefficients:
(Intercept)           x1           x5           x6           x7           x8          x11          x13  
  -0.001342     0.058918     0.080913     0.041066     0.071914     0.074959     0.729843     0.013942  
        x15  
   1.152034  
```











The ACF, PACF plots of residuals show some spikes. The ACF, PACF plots of squared residuals show more spikes. Residuals seem normal with outliers.

In order to verify what we observed in the diagrams we use statistical tests. Shapiro-Wilk hypothesis for normality is rejected meaning the residuals are not normal, but this is probably due to outliers. Box-Pierce and Ljung-Box hypothesis for residuals were rejected for small values of lag (6), but not rejected for larger values (60), meaning that we probably have **correlated residuals**. Finally, Box-Pierce and Ljung-Box hypothesis for squared residuals were also rejected for small values of lag (6), meaning that we have **problem with heteroscedasticity**.

```
> shapiro.test(residuals) # Rejected - no noramlity

        Shapiro-Wilk normality test

data:  residuals
W = 0.98426, p-value = 0.03249

> Box.test(residuals, 6, type="Ljung-Box") # Rejected - There is autocorrelation

        Box-Ljung test

data:  residuals
X-squared = 14.933, df = 6, p-value = 0.02079

> Box.test(residuals, 60, type="Ljung-Box") # Not Rejected - There is no autocorrelation

        Box-Ljung test

data:  residuals
X-squared = 65.433, df = 60, p-value = 0.2938

> Box.test(residuals^2, 6, type="Ljung-Box") # Rejected - There is heteroscedasticity

        Box-Ljung test

data:  residuals^2
X-squared = 25.049, df = 6, p-value = 0.0003345

> Box.test(residuals^2, 60, type="Ljung-Box") # Not Rejected - There is no heteroscedasticity

        Box-Ljung test

data:  residuals^2
X-squared = 65.684, df = 60, p-value = 0.2865
```

Question 3

Regression + ARMA

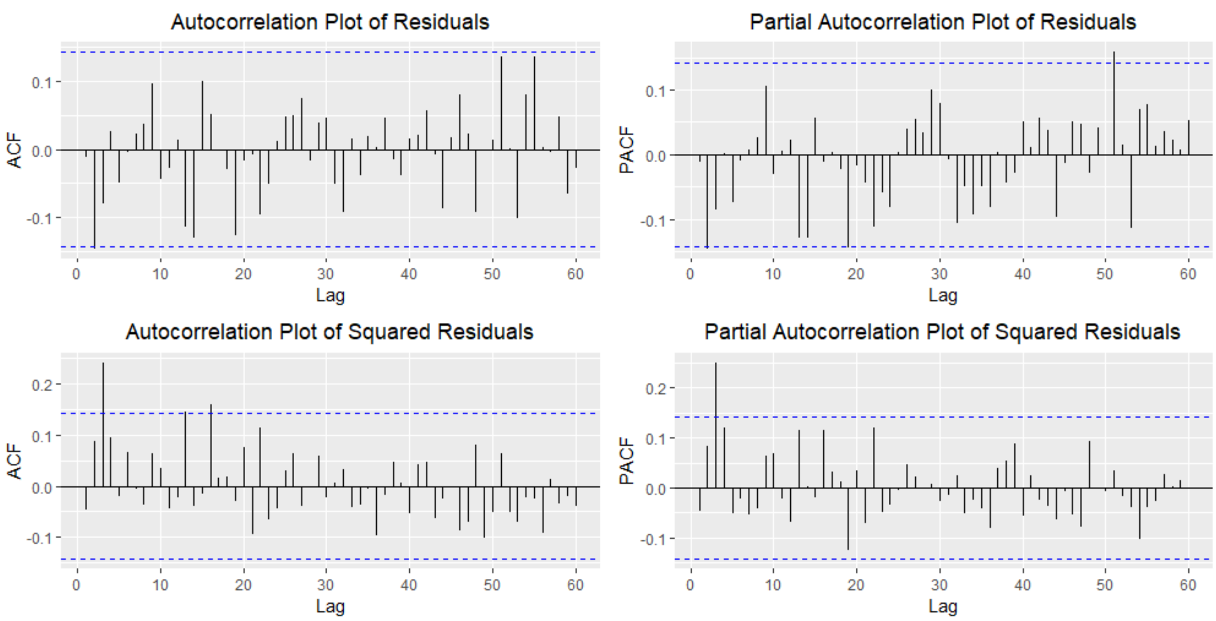
For this question the residuals of the regression model (z) we presented before were extracted and model with a **constrained ARMA(6,1)** model. This model was the one to give the smallest value for AIC criterion.

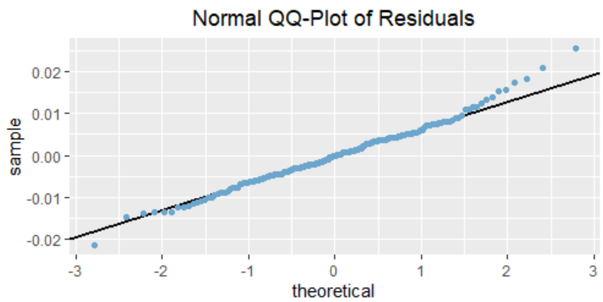
```
> arma61

Call:
arima(x = residuals_reg, order = c(6, 0, 1), fixed = c(0, 0, 0, 0, 0, NA, NA,
    NA))

Coefficients:
      ar1  ar2  ar3  ar4  ar5     ar6     ma1  intercept
        0    0    0    0    0  0.2160  0.0880      0e+00
s.e.    0    0    0    0    0  0.0717  0.0842      7e-04

sigma^2 estimated as 4.961e-05:  log likelihood = 668.3,  aic = -1328.6
```

That leaves us with the model below:

$$y_t = -0.001342 + 0.058918x_{1,t} + 0.080913x_{5,t} + 0.041066x_{6,t} + 0.071914x_{7,t} + 0.074959x_{8,t} + 0.729843x_{11,t} + 0.013942x_{13,t} + 1.152034x_{15,t} + z_t$$

$$z_t = 0 + 0.2160\varepsilon_{t-6} + 0.0880\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$





The ACF, PACF plots of residuals show that autocorrelation problem is resolved but the ACF, PACF plots of squared residuals show more spikes. Residuals seem normal with outliers.

In order to verify what we observed in the diagrams we use statistical tests. Shapiro-Wilk hypothesis for normality is not rejected meaning the **residuals are normal**. Box-Pierce and Ljung-Box hypothesis for residuals were not rejected for either small values of lag (6) or for larger values (60), meaning that we have **uncorrelated residuals**. Box-Pierce and Ljung-Box hypothesis for squared residuals were rejected for small values of lag (6), meaning that we have **problem with heteroscedasticity**.

```
> shapiro.test(residuals_arma) # Not Rejected - Noramlity

	Shapiro-Wilk normality test

data:  residuals_arma
W = 0.98793, p-value = 0.1082

> Box.test(residuals_arma, 6, type="Ljung-Box") # Not Rejected - There is not autocorrelation

	Box-Ljung test

data:  residuals_arma
X-squared = 5.9711, df = 6, p-value = 0.4264

> Box.test(residuals_arma, 60, type="Ljung-Box") # Not Rejected - There is not autocorrelation

	Box-Ljung test

data:  residuals_arma
X-squared = 54.025, df = 60, p-value = 0.6926

> Box.test(residuals_arma^2, 6, type="Ljung-Box") # Rejected - There is heteroscedasticity

	Box-Ljung test

data:  residuals_arma^2
X-squared = 15.694, df = 6, p-value = 0.01549

> Box.test(residuals_arma^2, 60, type="Ljung-Box") # Not Rejected - There is no heteroscedasticity

	Box-Ljung test

data:  residuals_arma^2
X-squared = 58.628, df = 60, p-value = 0.526
```

Regression + ARMA + GARCH

In this section, in order to fix the problem of heteroscedasticity, we fit an **ARCH(6)** model to the residuals of ARMA model we presented before.

```
Coefficient(s):
         mu        omega       alpha1       alpha2       alpha3       alpha4       alpha5       alpha6
-4.6794e-06   1.4829e-05   1.0000e-08   1.6512e-01   2.3833e-01   2.6373e-02   6.6614e-02   2.6116e-01
```

That leaves us with the model below:

$$y_t = -0.001342 + 0.058918x_{1,t} + 0.080913x_{5,t} + 0.041066x_{6,t} + 0.071914x_{7,t} + 0.074959x_{8,t} + 0.729843x_{11,t} + 0.013942x_{13,t} + 1.152034x_{15,t} + z_t$$

$$z_t = 0 + 0.2160\varepsilon_{t-6} + 0.0880\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0 + 0.0165\varepsilon_{t-2}^2 + 0.2383\varepsilon_{t-3}^2 + 0.0264\varepsilon_{t-4}^2 + 0.0666\varepsilon_{t-5}^2 + 0.2611\varepsilon_{t-6}^2$$

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  0.4172318 0.811707
 Shapiro-Wilk Test  R    W      0.9955478 0.8535027
 Ljung-Box Test     R    Q(10)  7.51774   0.6758283
 Ljung-Box Test     R    Q(15)  12.80108  0.6176595
 Ljung-Box Test     R    Q(20)  18.76258  0.537303
 Ljung-Box Test     R^2  Q(10)  2.54287   0.9902362
 Ljung-Box Test     R^2  Q(15)  4.575495  0.9951512
 Ljung-Box Test     R^2  Q(20)  6.511711  0.9980059
 LM Arch Test       R    TR^2   6.035608  0.9142765
```

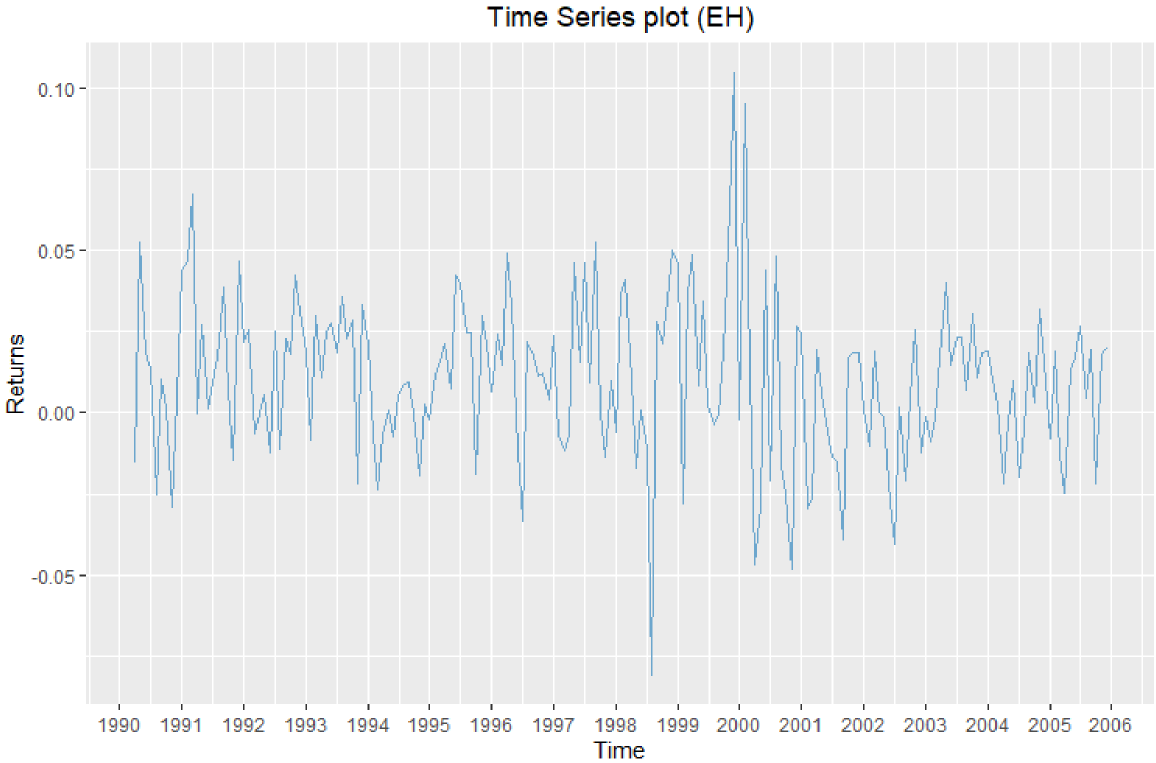
Observing the test statistics, we see that **autocorrelation problem is remaining resolved**. **Heteroscedasticity problem is also resolved**. Residuals are **normal**.

Stock EH

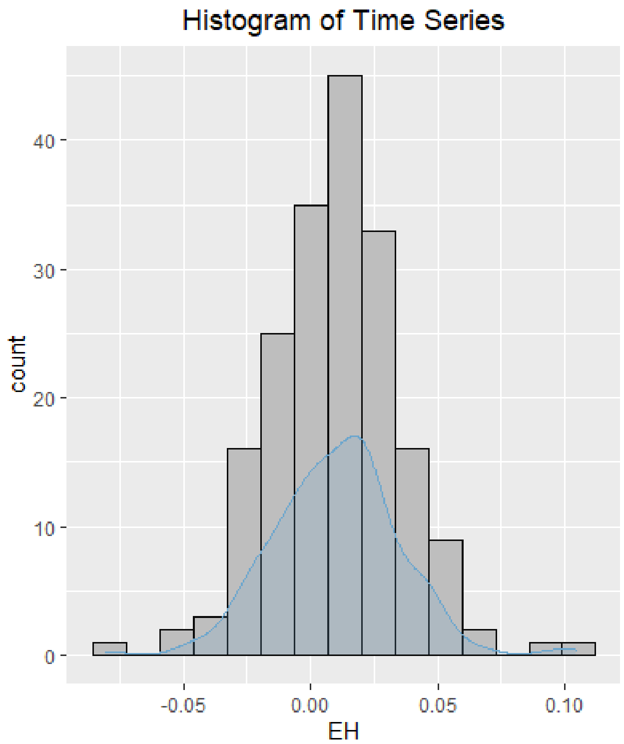
Question 1

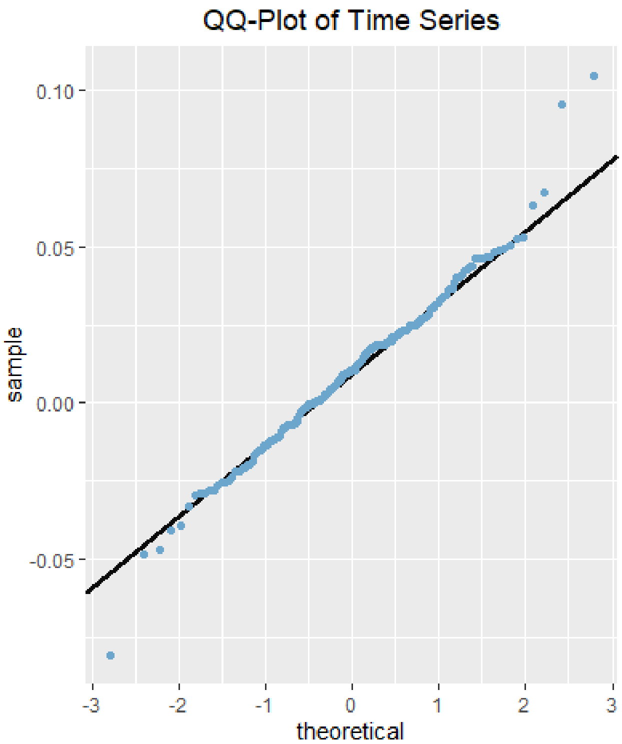
Identification Step

The time series analyzed in this section are the Returns of EH Stock. As it can be observed, the series seem stationary but with a problem of heteroscedasticity. There does seem to be a drift but no trend.

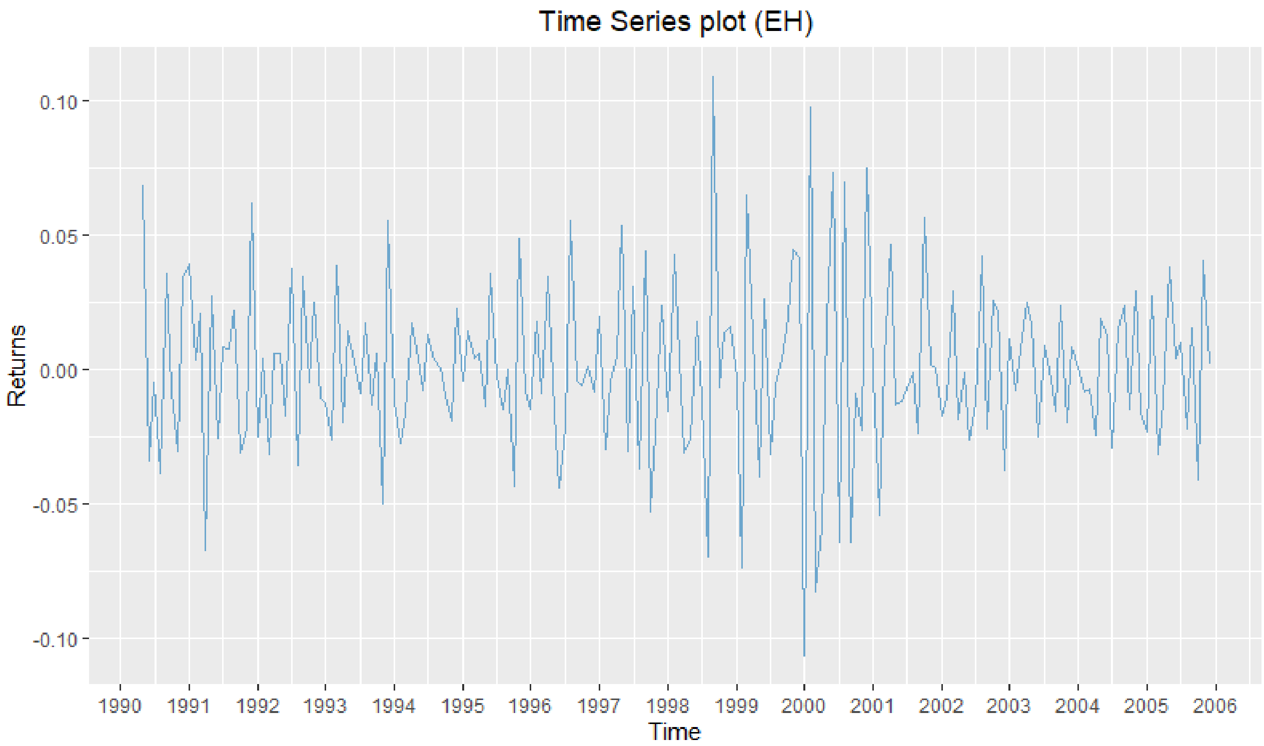


The histogram of the observations shows that they do not follow normal distribution. There are quite a lot of observations in the tails, something that we can verify by also looking at the QQ-plot.

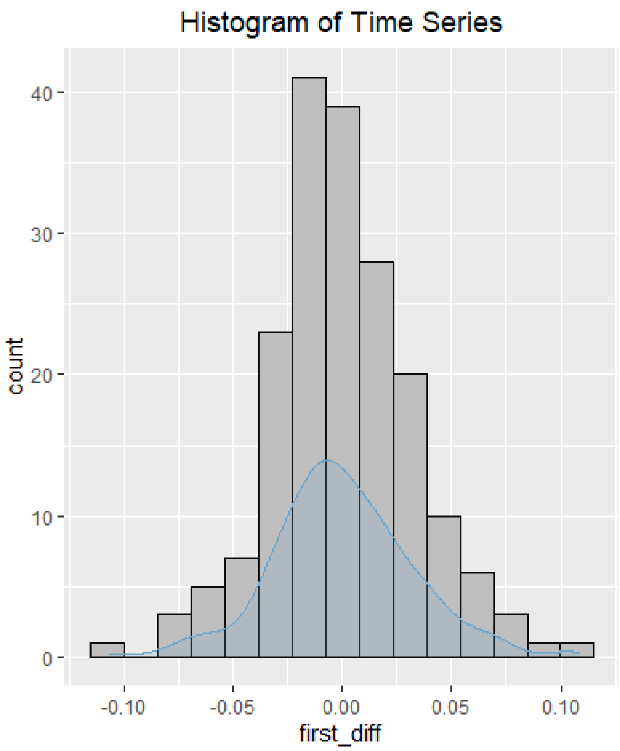


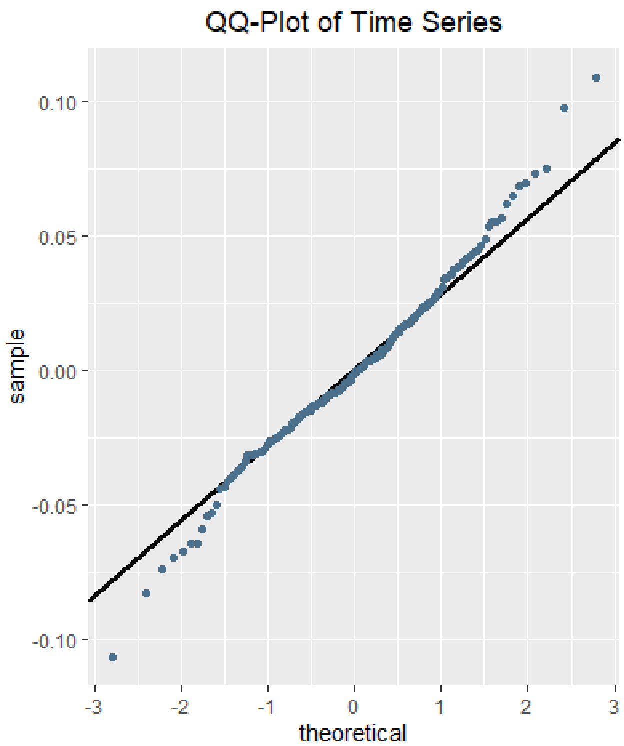


Let's try to fix the problem of normality by taking the **first differences** for the returns. That means that we lose an observation. Indeed, the series seem smoother.



The overall shape of the distribution looks more like the one of a normal distribution but form the QQ-plot we see that we have outliers.





In order to verify our claims a series of tests are performed:

a. *Shapiro-Wilk test for Normality*

The Null Hypothesis for this test is that the observations are drawn from a normal distribution (against the alternative Hypothesis that they are not drawn from a normal distribution). Since p-value = 0.1344 which is larger than the confidence level a = 0.05, we fail to reject the Null Hypothesis, thus we assume that the **distribution is normal**.

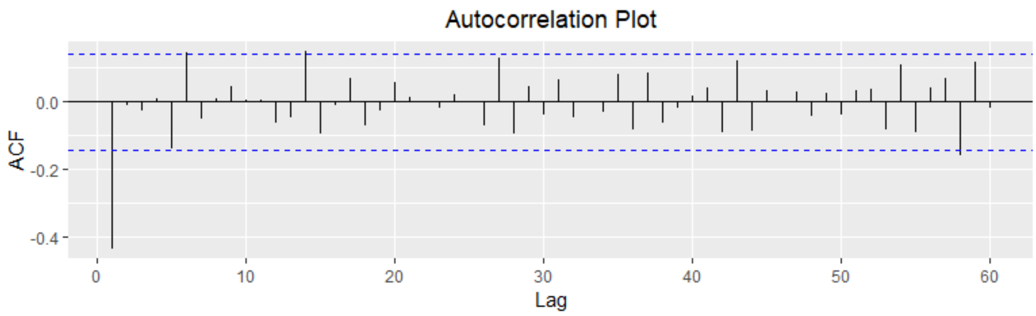
```
> shapiro.test(dts2$first_diff)

        Shapiro-Wilk normality test

data:  dts2$first_diff
W = 0.98855, p-value = 0.1344
```

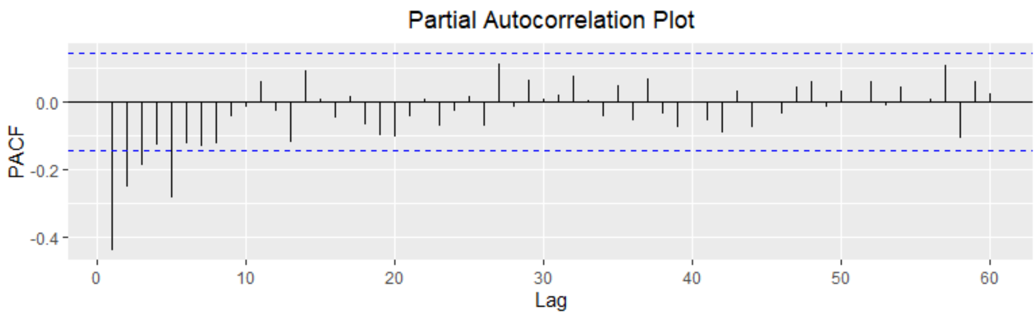
b. *Box-Pierce / Ljung-Box for Autocorrelation*

The Null Hypothesis for these tests is that the observations are correlated (against the alternative Hypothesis that they are not). The tests are performed for different values of lag (6, 10) and all give p-values bellow the confidence level a = 0.05, leading us to reject Null Hypothesis and assume **autocorrelation between observations**. The ACF and PACF plot also show possible autocorrelations, where there are peaks outside of horizontal blue lines.

```
> Box.test(dts2$first_diff, 6, type="Box-Pierce")

        Box-Pierce test

data:  dts2$first_diff
X-squared = 43.842, df = 6, p-value = 0.00000007944

> Box.test(dts2$first_diff, 6, type="Ljung-Box")

        Box-Ljung test

data:  dts2$first_diff
X-squared = 44.738, df = 6, p-value = 0.00000005277
```

```
> Box.test(dts2$first_diff, 10, type="Box-Pierce")

        Box-Pierce test

data:  dts2$first_diff
X-squared = 44.727, df = 10, p-value = 0.000002436

> Box.test(dts2$first_diff, 10, type="Ljung-Box")

        Box-Ljung test

data:  dts2$first_diff
X-squared = 45.671, df = 10, p-value = 0.000001645
```





* ACF, PACF plots do not show correlation with lag 0

c. Augmented Dickey-Fuller test for Stationarity

The Null Hypothesis for this test is that the series is not stationary (against the alternative that it is). The test was performed assuming drift of the series and setting the lags to 7. Since the test statistic (-8.56) is below tau1 for confidence level a = 0.05 (-1.95) we reject the Null Hypothesis and thus, we can consider the series **stationary**.

```
Value of test-statistic is: -8.5661

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Estimation Step

Based on the AIC criterion the model that was found to best fit the data (to give the smallest value of AIC) is a **constrained ARMA(5,1)** model:

```
> arma51

Call:
arima(x = tser, order = c(5, 0, 1), fixed = c(NA, NA, NA, 0, NA, NA, NA))

Coefficients:
         ar1     ar2      ar3  ar4      ar5      ma1  intercept
      0.0914  0.0084  -0.0527    0  -0.1264  -0.8501     0.1009
s.e.  0.1096  0.0950   0.0903    0   0.0914   0.0900     0.0003

sigma^2 estimated as 0.0005328:  log likelihood = 441.06,  aic = -868.13
```

$$y_t = 0.1009 + 0.0914y_{t-1} + 0.0084y_{t-2} - 0.0527y_{t-3} - 0.1264y_{t-5} - 0.8501\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

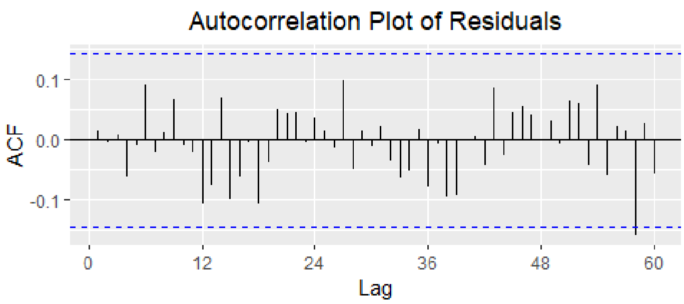
Diagnostic Plots

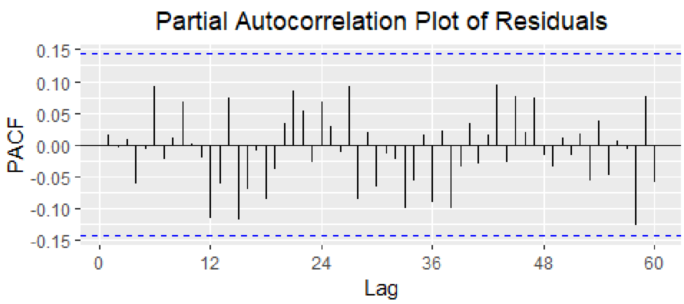
From the ACF and PACF plots of residuals we can see that the problem of **autocorrelation for the first lags is resolved**. The only spike that is outside the limits is close to lag 60 and will be ignored.

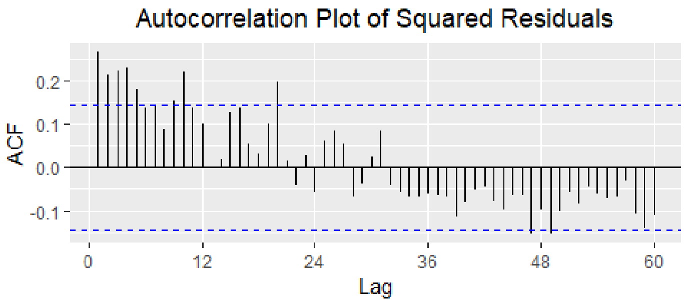
On the other hand, the ACF and PACF of squared residuals show quite a lot of spikes outside the boundaries leading us to assume **heteroscedasticity problems**.

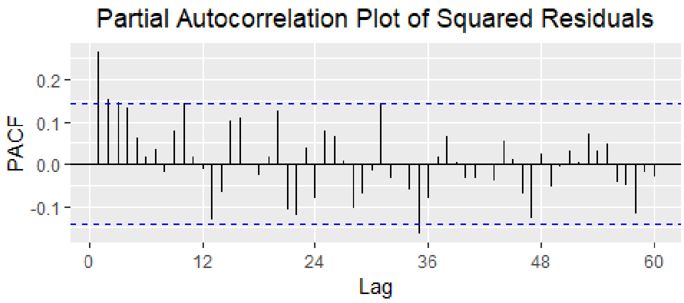
From the QQ-plot we can see that **residuals are mostly normal** with few outliers.

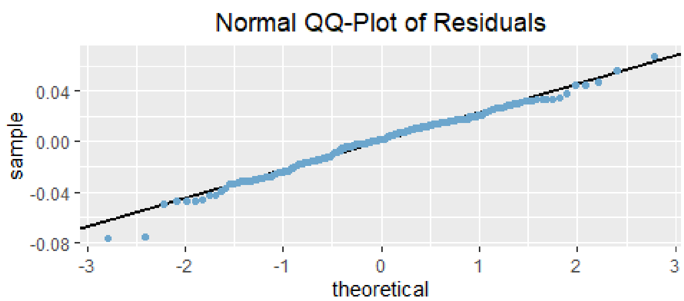
Finally, we can see the forecast for the next 20 instances. Obviously after forecasting for 5 instances the model draws the observation it needs (y_{t-5}) from our forecasted values.

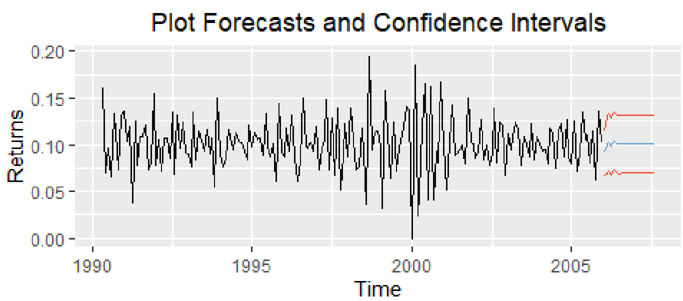












```
> # Normality
> shapiro.test(residuals)

        Shapiro-Wilk normality test

data:  residuals
W = 0.98684, p-value = 0.07722

> # Autocorrelation
> Box.test(residuals, 60, type="Ljung-Box")

        Box-Ljung test

data:  residuals
X-squared = 44.038, df = 60, p-value = 0.9393

> # Heteroscedasticity
> Box.test(residuals^2, 60, type="Ljung-Box")

        Box-Ljung test

data:  residuals^2
X-squared = 158.08, df = 60, p-value = 0.00000000009077
```

The claims made are verified from the same tests for normality of residuals (Shapiro-Wilk), autocorrelation of residuals (Ljung-Box) and autocorrelation of squared residuals (heteroscedasticity).

As expected, we do not reject Null Hypothesis for normality and autocorrelation (meaning there is not autocorrelation for 60 lags).

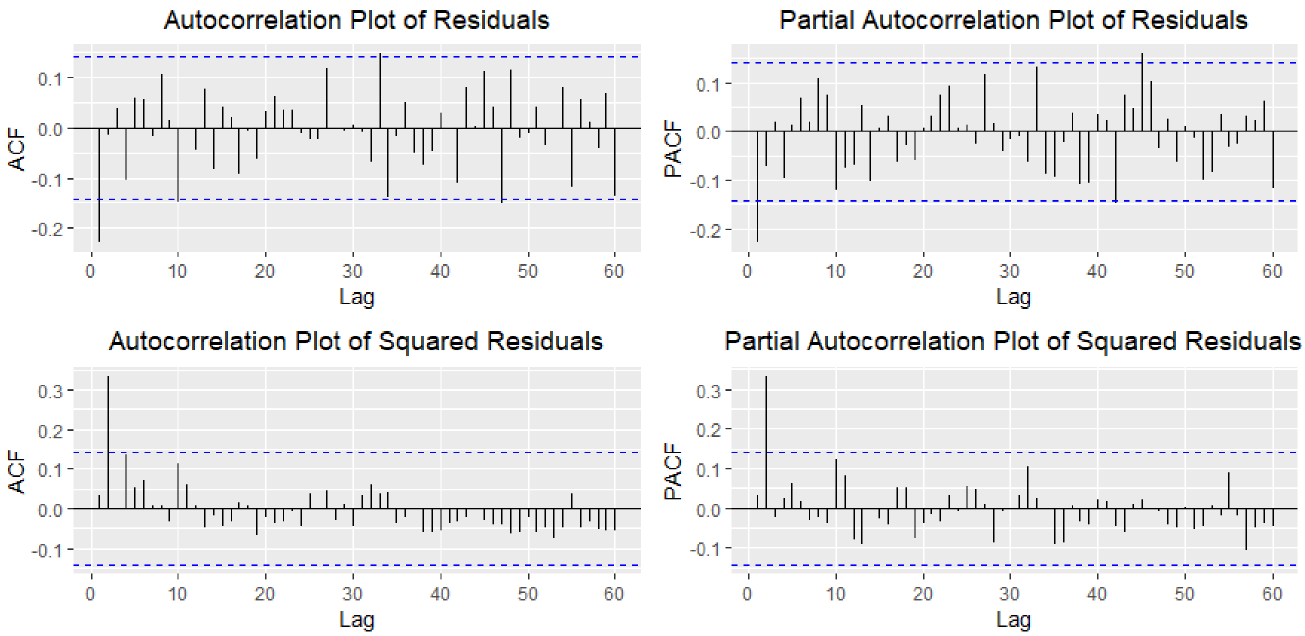
We do reject the Null Hypothesis for heteroscedasticity though.

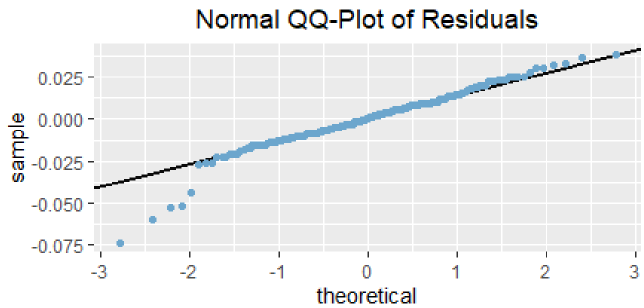
Question 2

For this question, a regression model was applied to the data using all independent variables (x_1-x_{15}). Using the AIC criterion this model was gradually reduced to having only x_1, x_2, x_5, x_7, x_8, x_{13} to predict the returns of EH Stock:

$$y_t = 0.09884 + 0.38182x_{1,t} - 0.34345x_{2,t} + 0.33069x_{5,t} + 0.08100x_{7,t} + 0.15253x_{8,t} + 0.05026x_{13,t} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

```
> model$anova
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
y2 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 +
    x12 + x13 + x14 + x15

Final Model:
y2 ~ x1 + x2 + x5 + x7 + x8 + x13


    Step Df          Deviance Resid. Df Resid. Dev       AIC
1                                   172 0.04821872 -1522.469
2    - x9  1 0.0000003397649       173 0.04821906 -1524.467
3   - x10  1 0.0000021007709       174 0.04822116 -1526.459
4   - x14  1 0.0000535362298       175 0.04827469 -1528.250
5   - x12  1 0.0000555216760       176 0.04833022 -1530.034
6    - x6  1 0.0001167189727       177 0.04844693 -1531.581
7   - x11  1 0.0001967001507       178 0.04864363 -1532.819
8    - x4  1 0.0003169358486       179 0.04896057 -1533.598
9   - x15  1 0.0004481550313       180 0.04940873 -1533.885
10   - x3  1 0.0005220997709       181 0.04993083 -1533.909
> model

Call:
lm(formula = y2 ~ x1 + x2 + x5 + x7 + x8 + x13, data = xs)

Coefficients:
(Intercept)           x1           x2           x5           x7           x8          x13
    0.09884      0.38182     -0.34345      0.33069      0.08100      0.15253      0.05026
```





The ACF, PACF plots of residuals show some spikes. The ACF, PACF plots of squared residuals also show spikes. Residuals seem normal with outliers.

In order to verify what we observed in the diagrams we use statistical tests. Shapiro-Wilk hypothesis for normality is rejected meaning the residuals are not normal. Box-Pierce and Ljung-Box hypothesis for residuals were rejected for small values of lag (6), but not rejected for larger values (60), meaning that we probably have **correlated residuals**. Finally, Box-Pierce and Ljung-Box hypothesis for squared residuals were also rejected for small values of lag (6), meaning that we have **problem with heteroscedasticity**.

```
> shapiro.test(residuals) # Rejected - no noramlity

        Shapiro-Wilk normality test

data:  residuals
W = 0.94924, p-value = 0.000003092

> Box.test(residuals, 6, type="Ljung-Box") # Rejected - There is autocorrelation

        Box-Ljung test

data:  residuals
X-squared = 13.732, df = 6, p-value = 0.03278

> Box.test(residuals, 60, type="Ljung-Box") # Not Rejected - There is no autocorrelation

        Box-Ljung test

data:  residuals
X-squared = 77.016, df = 60, p-value = 0.06858

> Box.test(residuals^2, 6, type="Ljung-Box") # Rejected - There is heteroscedasticity

        Box-Ljung test

data:  residuals^2
X-squared = 26.533, df = 6, p-value = 0.0001771

> Box.test(residuals^2, 60, type="Ljung-Box") # Not Rejected - There is no heteroscedasticity

        Box-Ljung test

data:  residuals^2
X-squared = 51.288, df = 60, p-value = 0.781
```

Question 3

Regression + ARMA

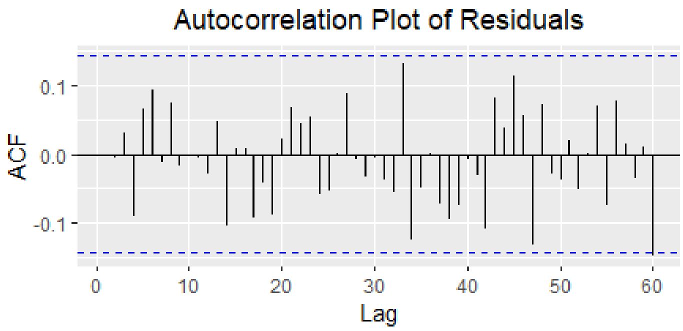
For this question the residuals of the regression model (z) we presented before were extracted and model with a **constrained ARMA(1,10)** model. This model was the one to give the smallest value for AIC criterion.

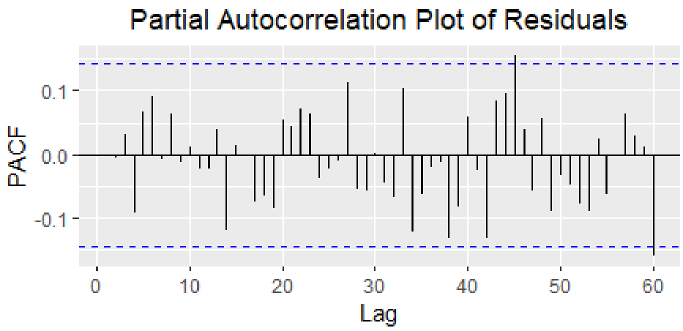
```
> arma110

Call:
arima(x = residuals_reg, order = c(1, 0, 10), fixed = c(NA, NA, 0, 0, 0, 0,
    0, 0, 0, 0, NA, NA))

Coefficients:
         ar1      ma1  ma2  ma3  ma4  ma5  ma6  ma7  ma8  ma9     ma10  intercept
      0.0115  -0.2872    0    0    0    0    0    0    0    0  -0.1714    -0.0001
s.e.  0.2390   0.2275    0    0    0    0    0    0    0    0   0.0714     0.0006

sigma^2 estimated as 0.0002354:  log likelihood = 518.34,  aic = -1026.69
```

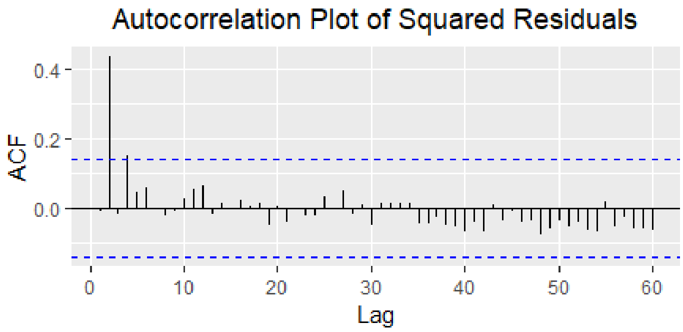
That leaves us with the model below:

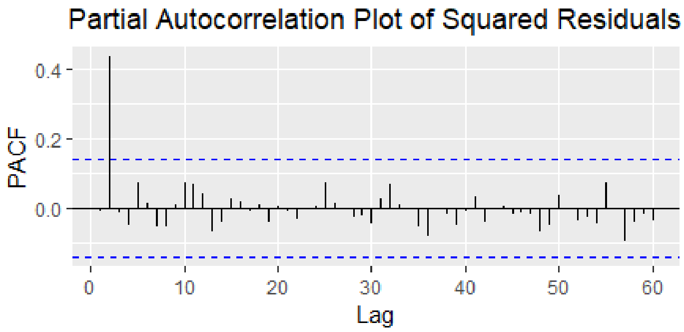
$$y_t = 0.09884 + 0.38182x_{1,t} - 0.34345x_{2,t} + 0.33069x_{5,t} + 0.08100x_{7,t} + 0.15253x_{8,t} + 0.05026x_{13,t} + z_t$$

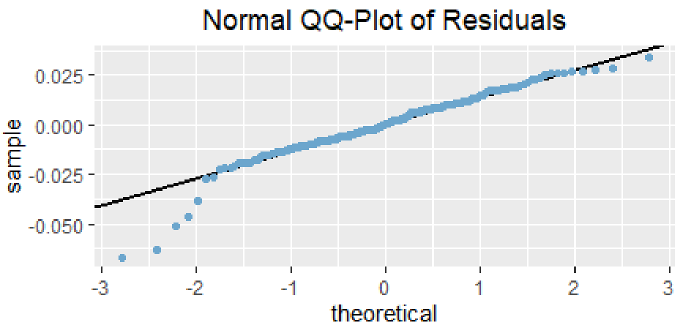
$$z_t = -0.0001 + 0.0115z_{t-1} - 0.2872\varepsilon_{t-1} - 0.1714\varepsilon_{t-10} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$











The ACF, PACF plots of residuals show that autocorrelation problem is resolved but the ACF, PACF plots of squared residuals show more spikes. Residuals seem normal with outliers.

In order to verify what we observed in the diagrams we use statistical tests. Shapiro-Wilk hypothesis for normality is not rejected meaning the **residuals are normal**. Box-Pierce and Ljung-Box hypothesis for residuals were not rejected for either small values of lag (6) or for larger values (60), meaning that we have **uncorrelated residuals**. Box-Pierce and Ljung-Box hypothesis for squared residuals were rejected for small values of lag (6, 20), meaning that we have **problem with heteroscedasticity**.

```
> shapiro.test(residuals_arma) # Not Rejected - Noramlity

        Shapiro-Wilk normality test

data:  residuals_arma
W = 0.943, p-value = 0.0000008496

> Box.test(residuals_arma, 6, type="Ljung-Box") # Not Rejected - There is not autocorrelation

        Box-Ljung test

data:  residuals_arma
X-squared = 4.359, df = 6, p-value = 0.6282

> Box.test(residuals_arma, 60, type="Ljung-Box") # Not Rejected - There is not autocorrelation

        Box-Ljung test

data:  residuals_arma
X-squared = 59.315, df = 60, p-value = 0.5007

> Box.test(residuals_arma^2, 6, type="Ljung-Box") # Rejected - There is heteroscedacity

        Box-Ljung test

data:  residuals_arma^2
X-squared = 42.405, df = 6, p-value = 0.000000153

> Box.test(residuals_arma^2, 20, type="Ljung-Box") # Rejected - There is heteroscedacity

        Box-Ljung test

data:  residuals_arma^2
X-squared = 44.826, df = 20, p-value = 0.001165
```

Regression + ARMA + GARCH

In this section, in order to fix the problem of heteroscedasticity, we fit a **GARCH(4,2)** (some of the coefficients are very close to 0 though) model to the residuals of ARMA model we presented before.

```
Coefficient(s):
         mu        omega       alpha1       alpha2       alpha3       alpha4        beta1        beta2
-0.000222967  0.000040383  0.050838382  0.321437692  0.000000010  0.000000010  0.185969120  0.285767415
```

That leaves us with the model below:

$$y_t = 0.09884 + 0.38182x_{1,t} - 0.34345x_{2,t} + 0.33069x_{5,t} + 0.08100x_{7,t} + 0.15253x_{8,t} + 0.05026x_{13,t} + z_t$$

$$z_t = -0.0001 + 0.0115z_{t-1} - 0.2872\varepsilon_{t-1} - 0.1714\varepsilon_{t-10} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = 0 + 0.051\sigma_{t-1}^2 + 0.321\sigma_{t-2}^2 + 0.186\varepsilon_{t-1}^2 + 0.286\varepsilon_{t-2}^2$$

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  4.660553  0.09726883
 Shapiro-Wilk Test  R    W      0.9879948 0.112504
 Ljung-Box Test     R    Q(10)  2.664293  0.9882592
 Ljung-Box Test     R    Q(15)  4.928177  0.9927172
 Ljung-Box Test     R    Q(20)  9.523699  0.9760097
 Ljung-Box Test     R^2  Q(10)  2.916193  0.983328
 Ljung-Box Test     R^2  Q(15)  4.963957  0.9924272
 Ljung-Box Test     R^2  Q(20)  7.280285  0.9956594
 LM Arch Test       R    TR^2   4.438815  0.9741553
```

Observing the test statistics, we see that **autocorrelation problem is remaining resolved**. **Heteroscedasticity problem is also resolved**. Residuals are **normal**.