Time Series 732A62

Lab 1

Eric Herwin
erihe068

Albin Västerlund
albva223

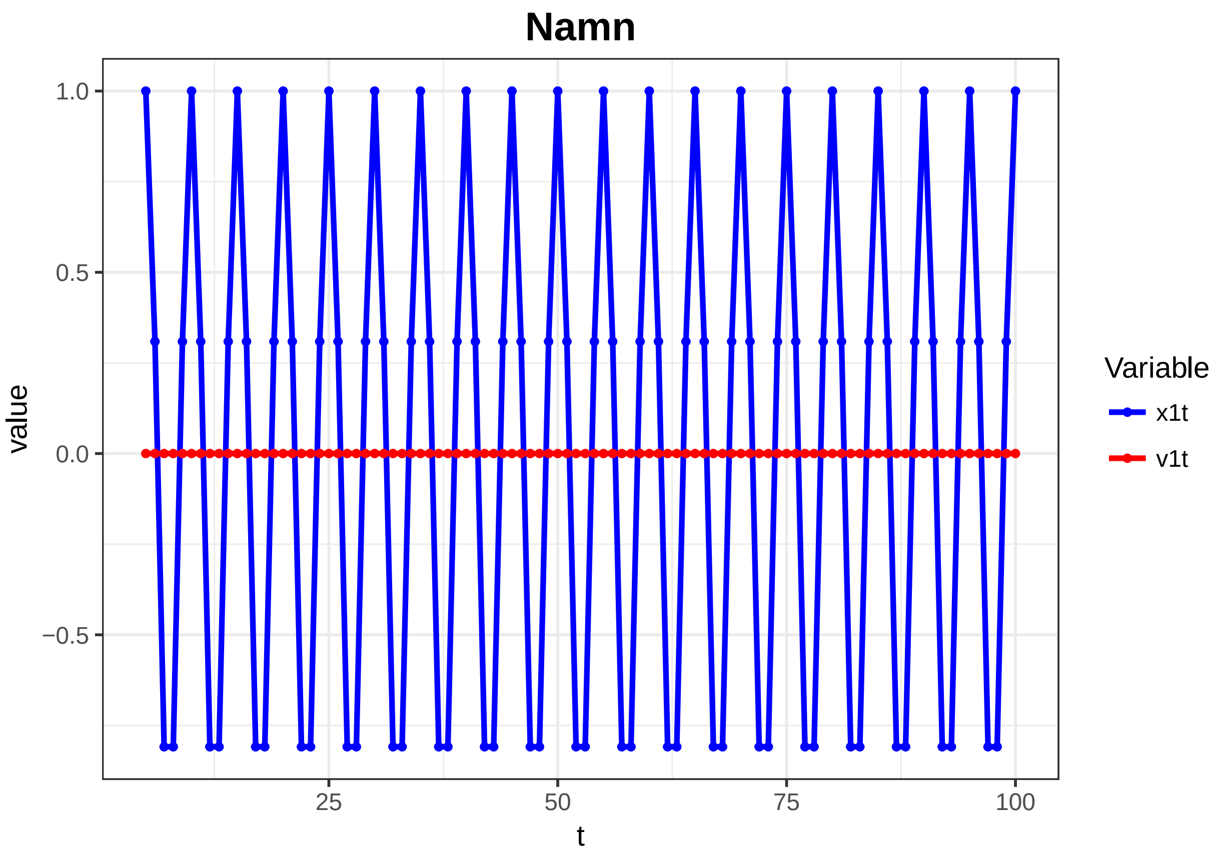
Contents

Assignment 1. Computations with simulated data

a)

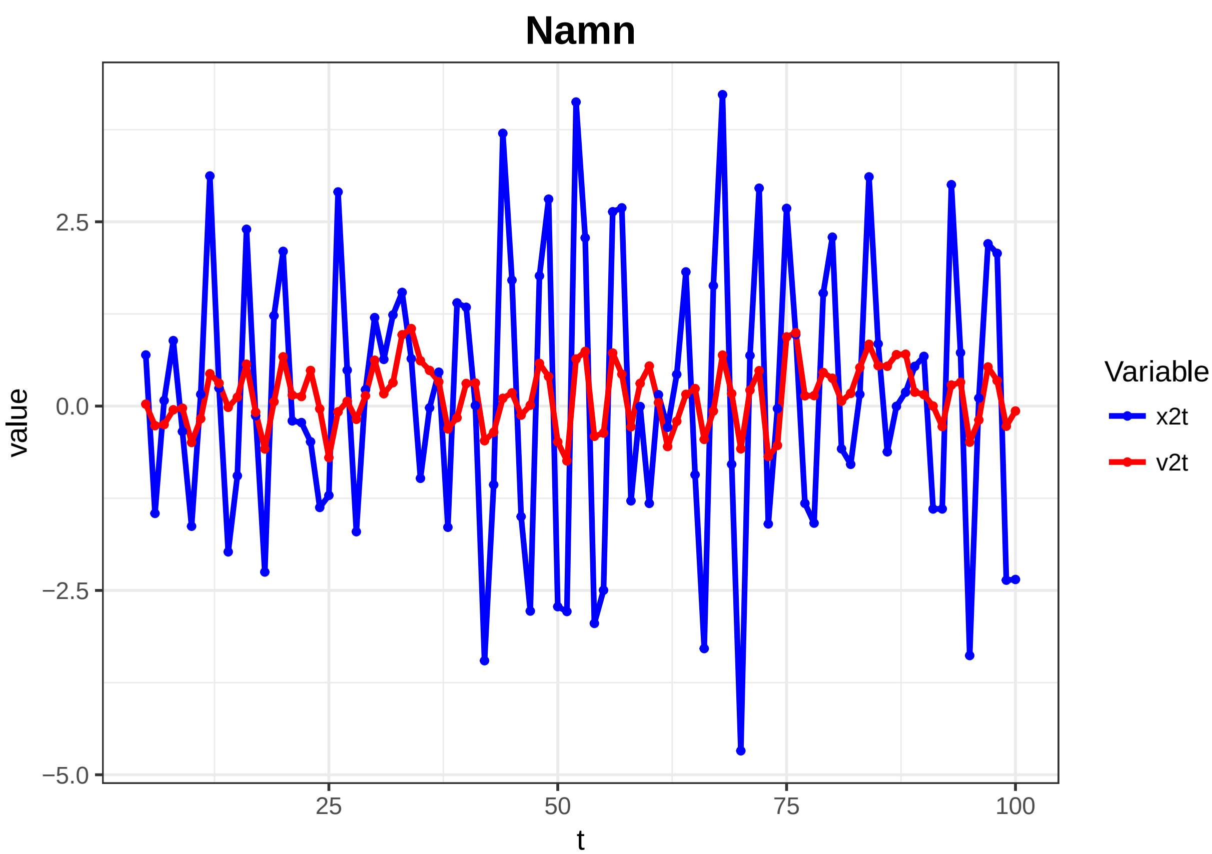
In this assigment we shall generate two variables, $x_{1t} = cos((2\pi t)/5)$ and $x_{2t} = -0.8x_{t-2} + w_t$, and then apply a smoothing filter to generate the variables $v_{it} = 0.2(x_t + x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}+)$.

We start with x_{1t} and v_{1t}. Here is a graph of the variables:



The filter v_{1t} applied on x_{1t} is presented in this graph and we can see that it has evened out the whole serie so every value of v_{1t} has been 0.

Now we look at x_{2t} and v_{2t}. The graph is presented below:



We can see that the filter v_{2t} has the same pattern as x_{2t}. The big difference between the seies is that v_{2t} has a lower variance then x_{2t}.

b)

In this task we will see if this the time series $x_t - 4x_{t-1} + 2x_{t-2} + x_{t-5} = w_t + 3w_{t-2} + w_{t-4} - 4w_{t-6}$ are casual or/and invertible.

We need to get the real roots and check if $|z| > 1$ or not.

First we check if it is causial:

```
## [1] 0.2936658 1.6793817 1.0000000 1.2179842 1.6037000
```

The real roots are under 1 and we can see that the series is not causial.

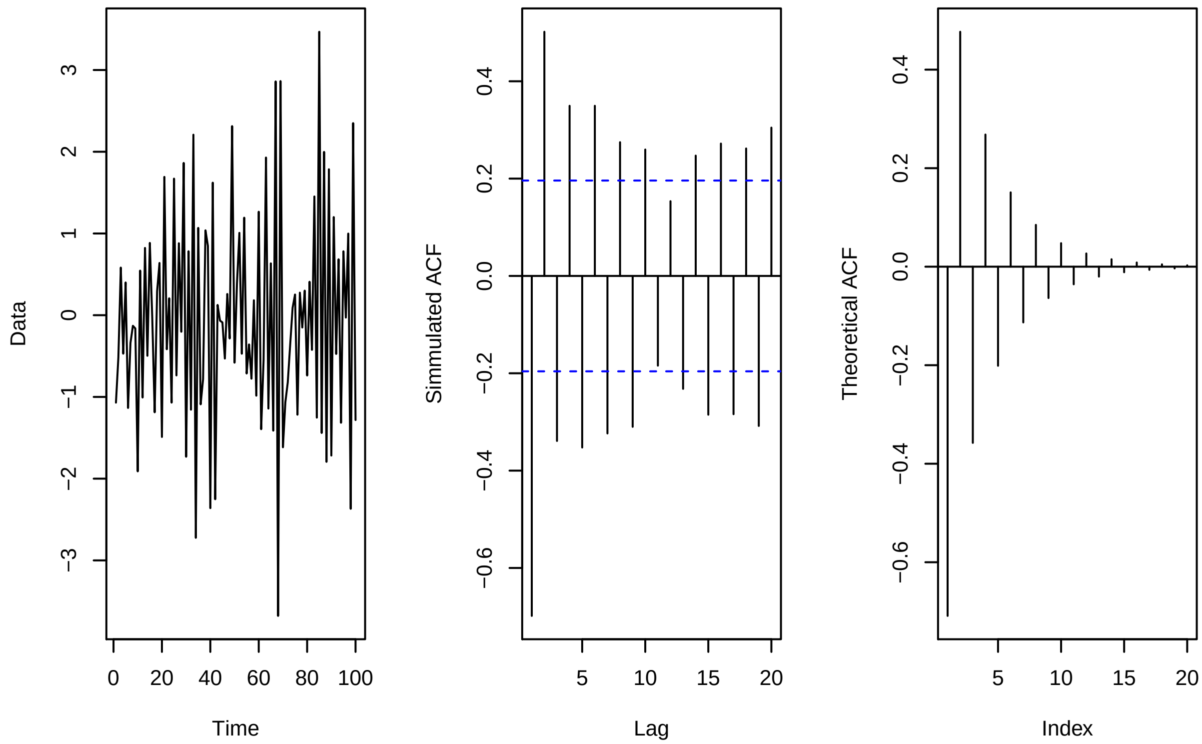
We also check if the serie is inverible:

```
## [1] 0.6874372 0.6874372 0.6874372 0.6874372 1.0580446 1.0580446
```

Because the real roots, z, are not above 1 we can assume that this series in not invertible.

c)

With built-in R-funtions we will simmulate a 100 observation time series and compute a sample ACF and theoretical ACF.

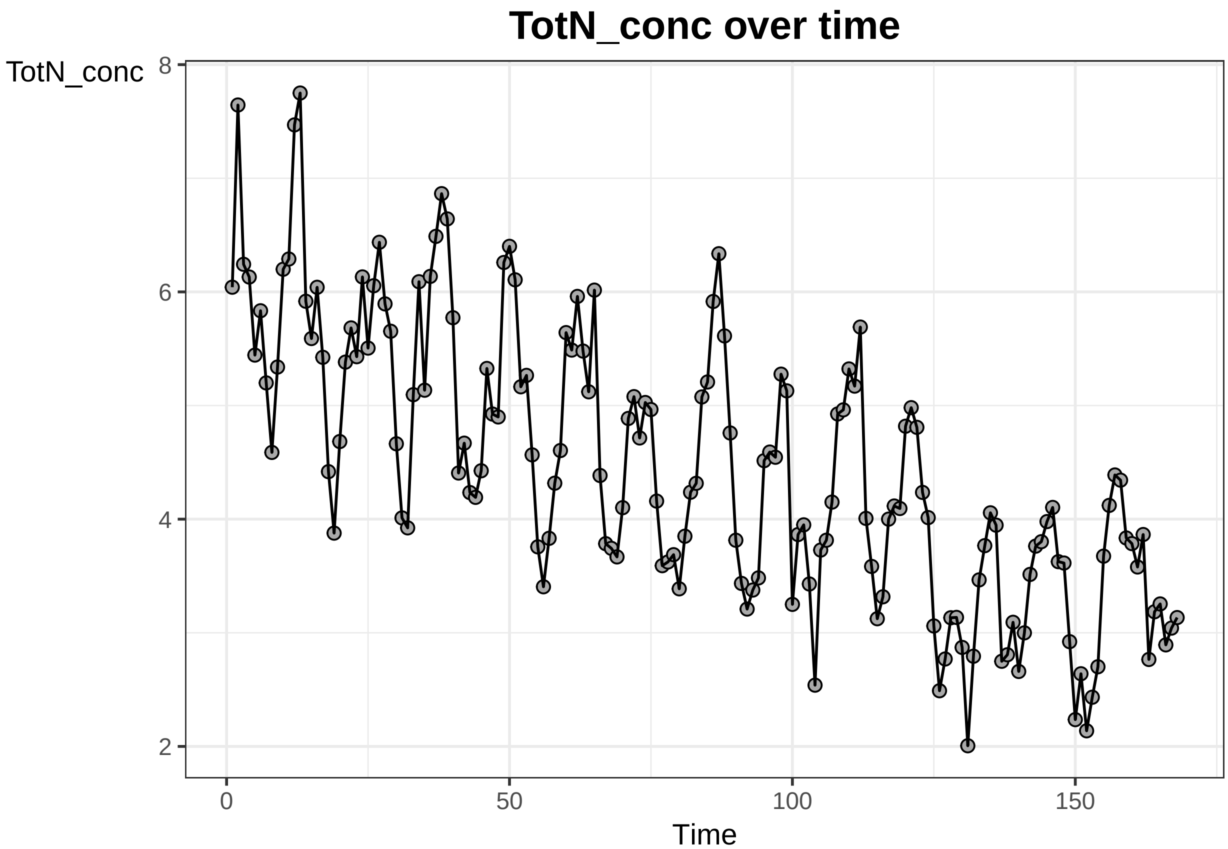


We can see that the sample ACF has more correlation over time than the theoretical ACF. Though you have the same pattern of correlation with both ACFs which means that the expected correlation converge to the theoretical. The larger the amount of observations there is, the better the correlation estimation.

Assignment 2. Visualization, detrending and residualanalysis of Rhine data

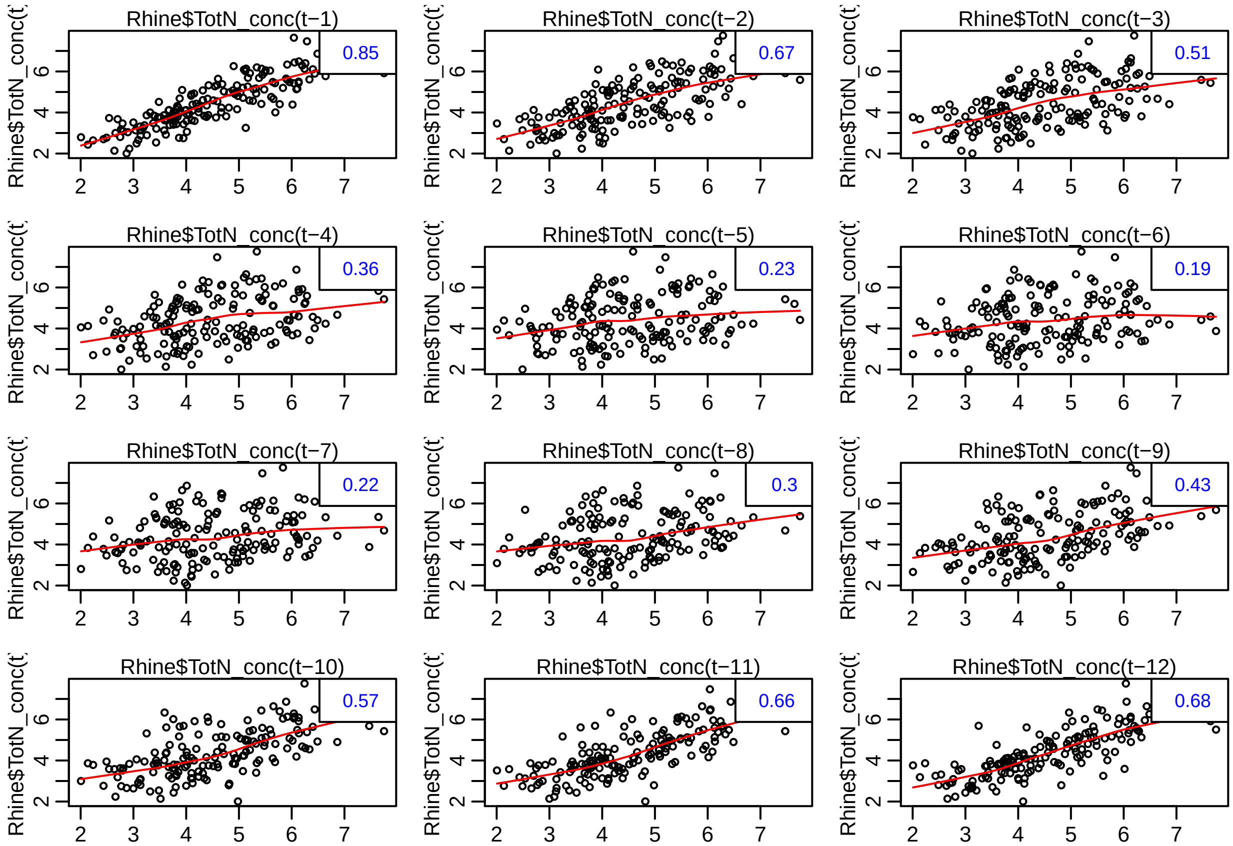
a)

In this task we will use the data Rhine. We start to explore the data by ploting TotN_conc (x_t) over time.



The plot seem to show a downward linear trend and some seasonal pattern occuring eact month with biggest values in february and lowest values in august. The variance does not seem to change that much over time.

We also explore the data by ploting the time series x_t against it's lag, up to 12.



When we observe the scatterplots we see that, for example, one lag had a high correlation and 12 lags as well. This means that first of all that it is dependent of it's previous values and that it has a high correlation every 12 months.

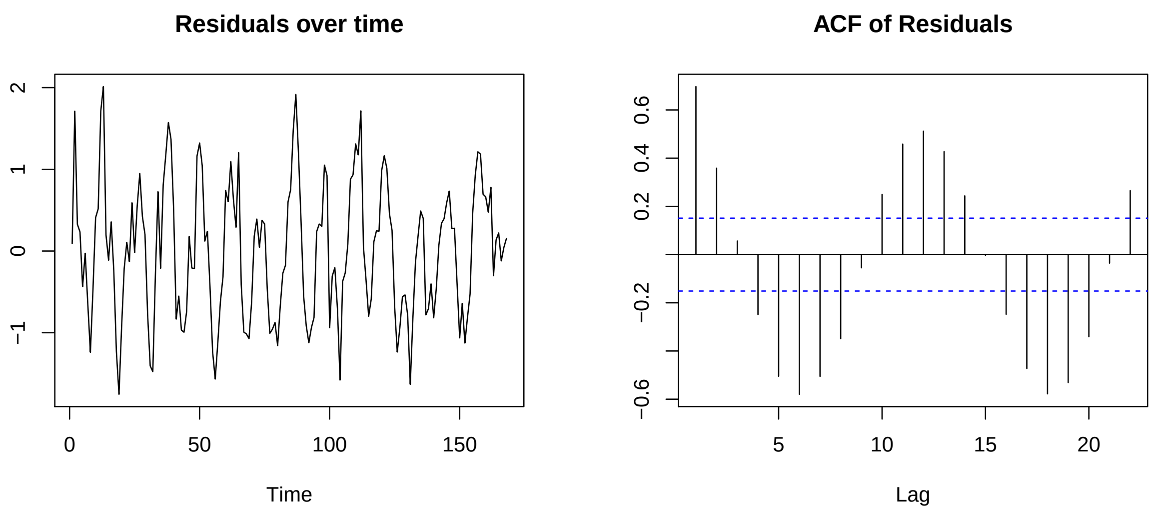
b)

A linear model has been made to eliminate the trend, and the model is provided in the following R-output:

```
##
## Call:
## lm(formula = TotN_conc ~ Trend, data = Rhine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.75325 -0.65296  0.06071  0.52453  2.01276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.968486   0.127177   46.93   <2e-16 ***
## Trend       -0.017796   0.001305  -13.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8205 on 166 degrees of freedom
## Multiple R-squared:  0.5282, Adjusted R-squared:  0.5254
## F-statistic: 185.9 on 1 and 166 DF,  p-value: < 2.2e-16
```

The trend is significant.

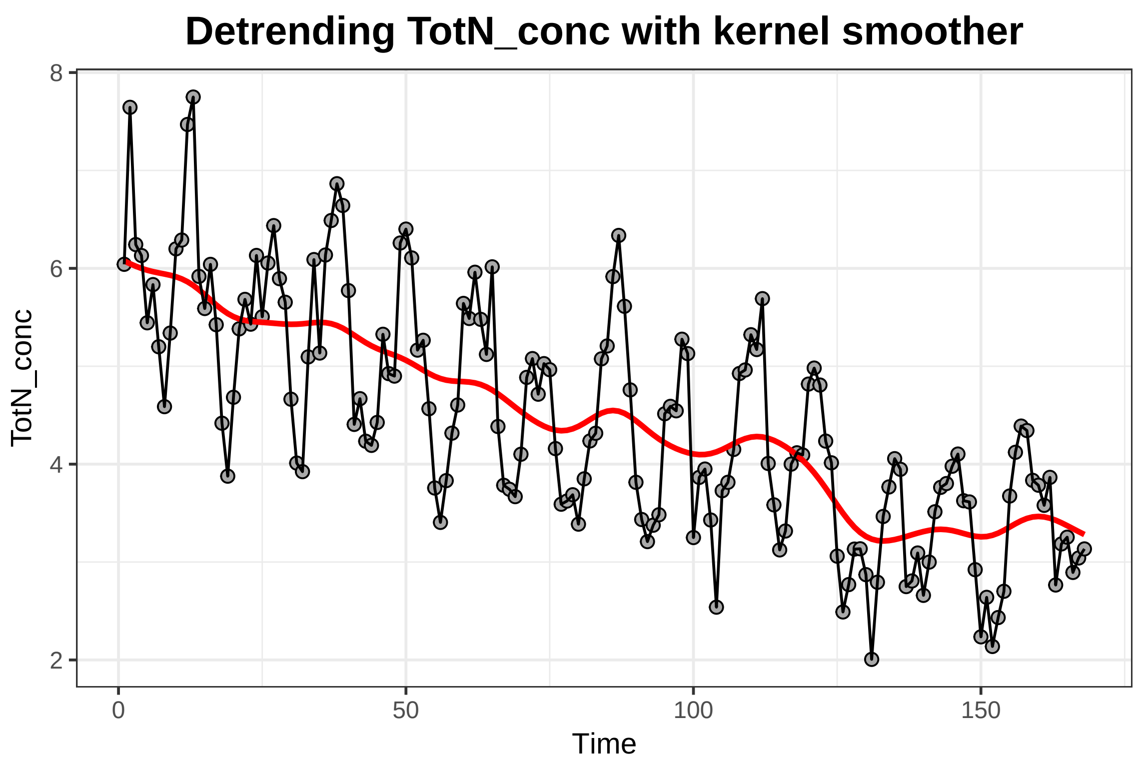
The residuals are plotted in a scatterplot and a ACF:



As we assumed in the scatterplotts with x_t and its lags, x_t has a high correlation with $t-1$ and $t-12$. Which can be obtained in the ACF, there is big spikes on lag 1 and 12.

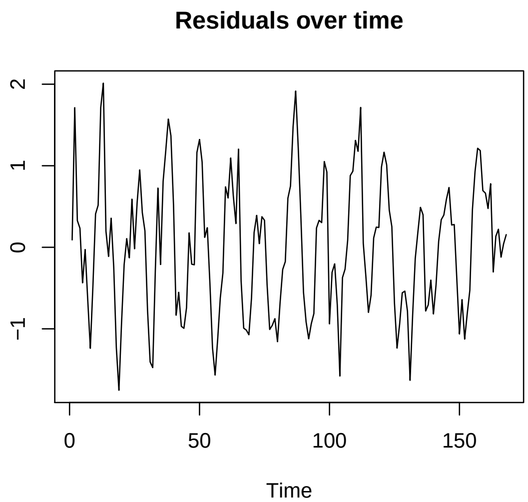
c)

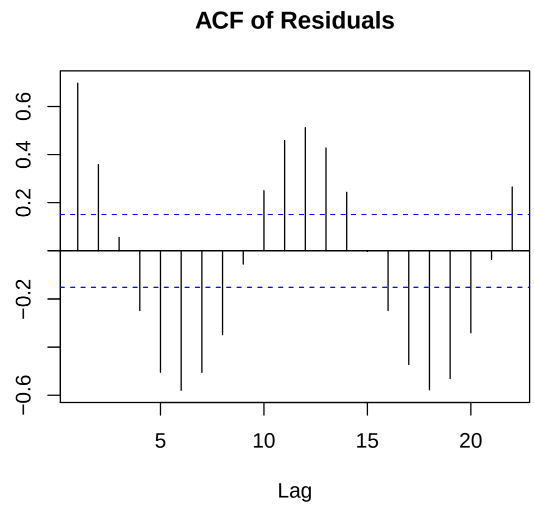
A kernel smoother is used to eleminate the trend.

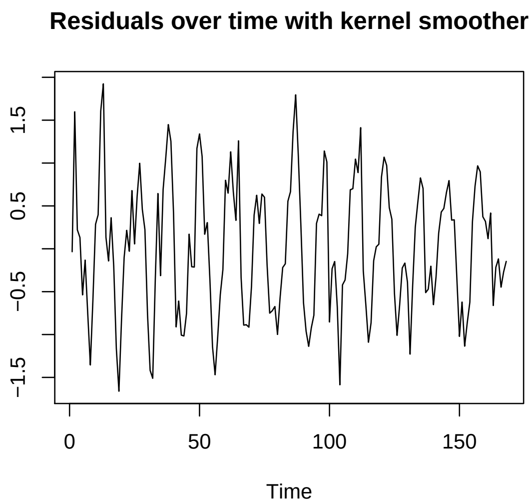


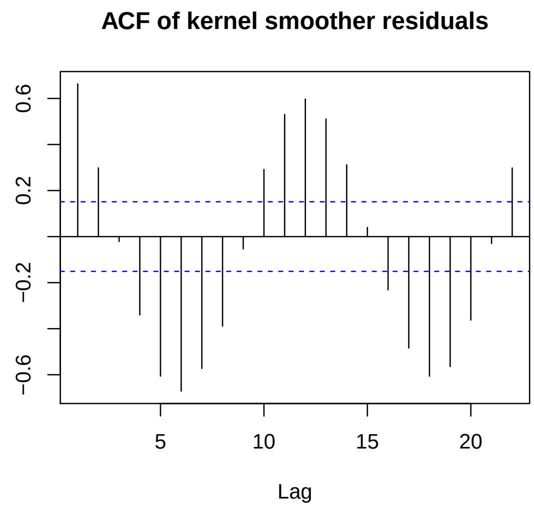
We set the bandwidth to 15. If a lower bandwidth is set, the fitted values pick up the season in the data. We just wanted the fitted values to pick up the trend.

Now we want to obtain the ACF and a plot of the residuals and time and compere the results with assignment 2.









Same pattern in ACF and the time series of the residuals is still the same.

d)

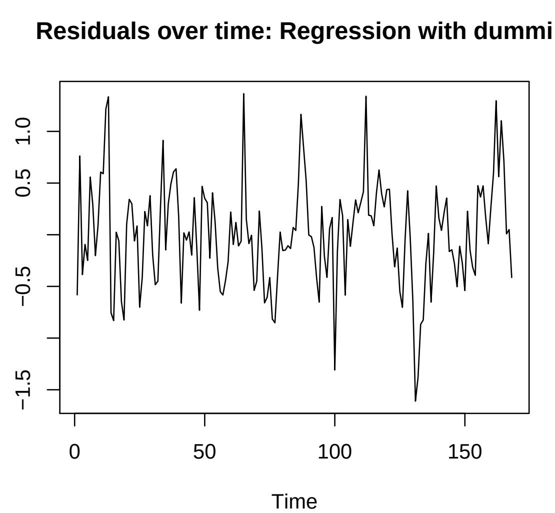
In this task we will make a cathegorial variable that indicate which month it is. Then we will use this variable in a linear regression to handle the seasonality. The regression is presented here:

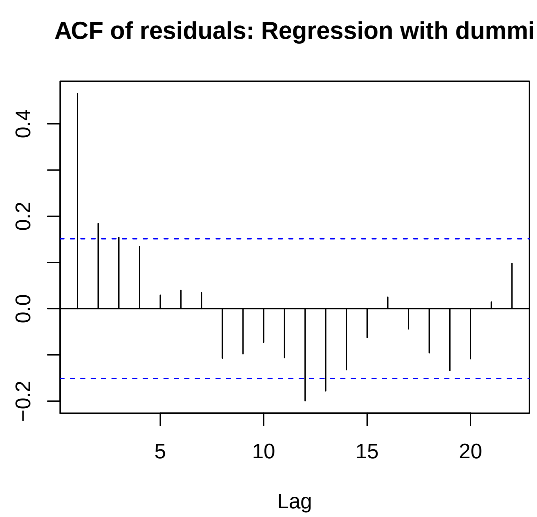
```
##
## Call:
## glm(formula = TotN_conc ~ Trend + Dummi, data = Rhine)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6101  -0.3119   0.0082   0.3115   1.3646
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.6405394  0.1560548  42.553  < 2e-16 ***
## Trend       -0.0173536  0.0008424 -20.601  < 2e-16 ***
## Dummi2       0.2765862  0.1996248   1.386  0.16788
## Dummi3       0.0400625  0.1996301   0.201  0.84121
## Dummi4      -0.3464274  0.1996390  -1.735  0.08468 .
## Dummi5      -0.8616526  0.1996514  -4.316 2.82e-05 ***
## Dummi6      -1.2611407  0.1996674  -6.316 2.70e-09 ***
## Dummi7      -1.6080759  0.1996870  -8.053 2.00e-13 ***
## Dummi8      -1.7124154  0.1997101  -8.575 9.56e-15 ***
## Dummi9      -1.2366876  0.1997367  -6.192 5.11e-09 ***
## Dummi10     -0.8744578  0.1997669  -4.377 2.20e-05 ***
## Dummi11     -0.7512727  0.1998007  -3.760  0.00024 ***
## Dummi12     -0.1774468  0.1998379  -0.888  0.37594
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2789454)
##
##     Null deviance: 236.891  on 167  degrees of freedom
## Residual deviance:  43.237  on 155  degrees of freedom
## AIC: 276.74
##
## Number of Fisher Scoring iterations: 2
```

As we saw in assignment 2a the biggest values is obtained in february and the lowest in august which also the modell proves.

The residuals are plotted as a time series and in a ACF:





The big spike at lag 12 have now turnd negative and it's half the size. We still have a big spike at lag 1. Which still indicate on autocorrelation.

e)

A stepwise model selection is performed to see if the seasonal variable or a trend variable should be in the model.

The models are:

```
#Trend and season
summary(modell)$aic #Best modell if we look at aic
```

```
## [1] 276.7406
```

```
#
#
#Only trend
summary(glm(data=Rhine,formula = TotN_conc~Trend))$aic
```

```
## [1] 414.2863
```

```
#
#
#Only season
summary(glm(data=Rhine,formula = TotN_conc~Dummi))$aic
```

```
## [1] 496.2627
```

```
#
#
#Only intercept
summary(glm(data=Rhine,formula = TotN_conc~1))$aic
```

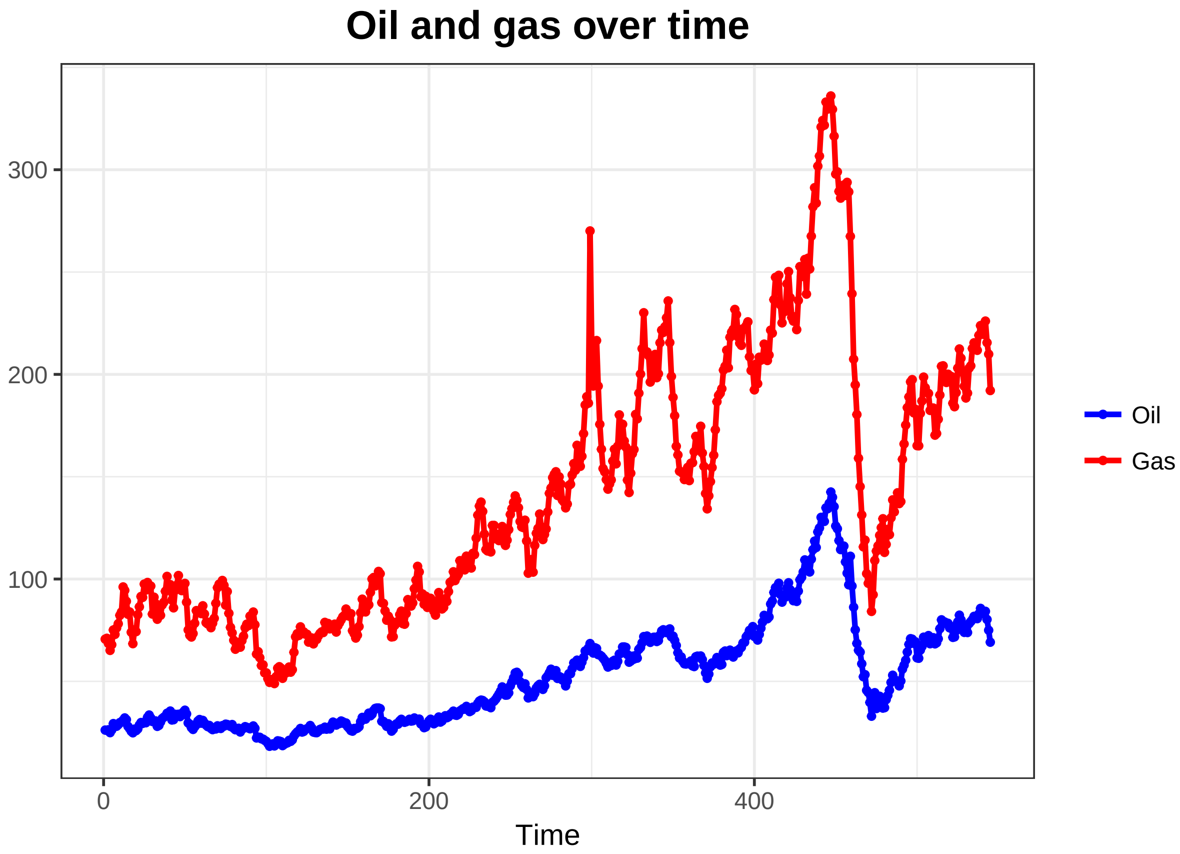
```
## [1] 538.4942
```

By looking at the AIC the model how have both season and trend in it is the best model. That is because of that model have the smalest AIC value.

Assignment 3. Analysis of oil and gas time series.

a)

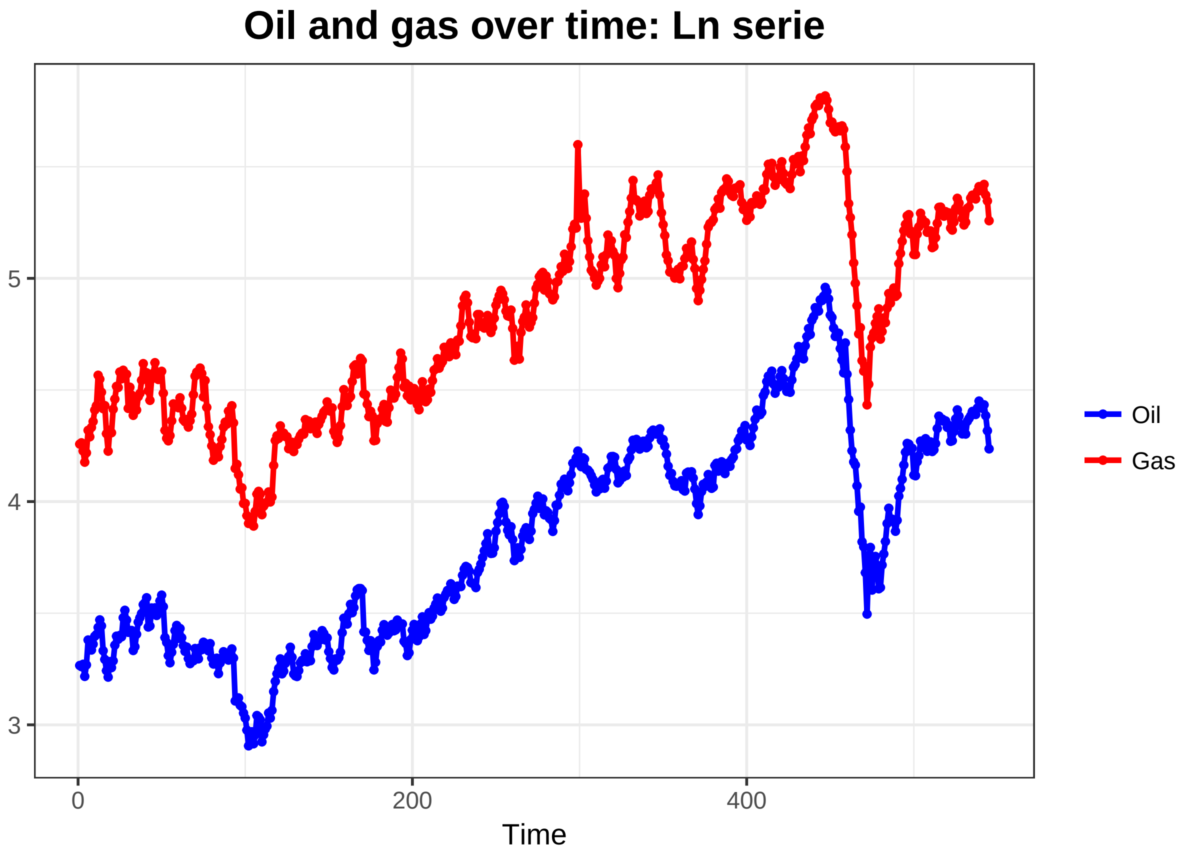
In this assignment we will investigate if the vaiables oil and gas in the package astsa is stationary and if the variables seem to be related to each other.



By looking at the graph we can se that the series seems to have the same pattern. If oil gets a large value at time point t then will gas probably also have a large value at time point t. The time series is not stationary. For example the criteria for $\mu_t = \mu$ is not fulfilled for the series. The graph also indicates of a difference of variance over time for the time serie gas.

b)

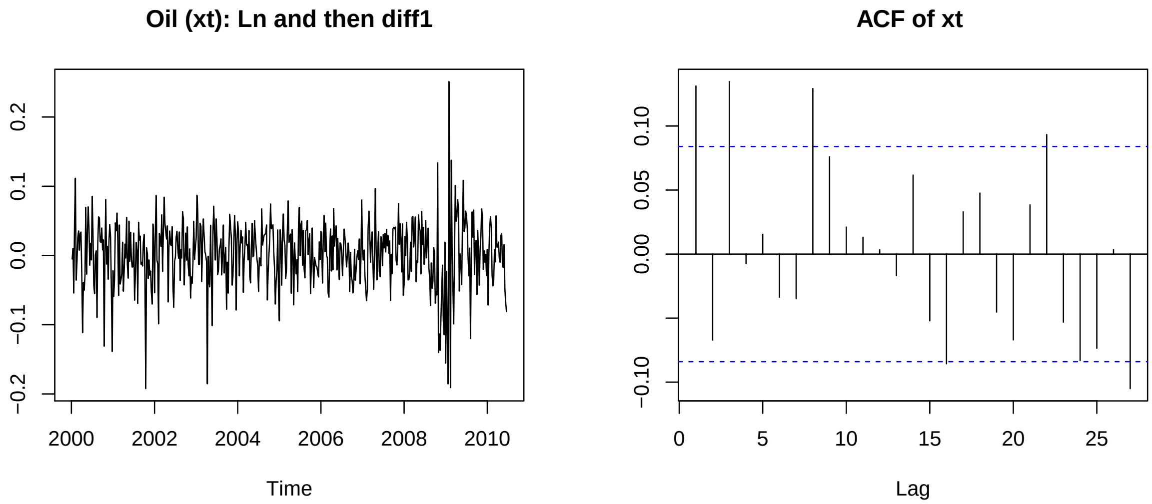
Now we will transform our data by using the log function. We use the natural base e.

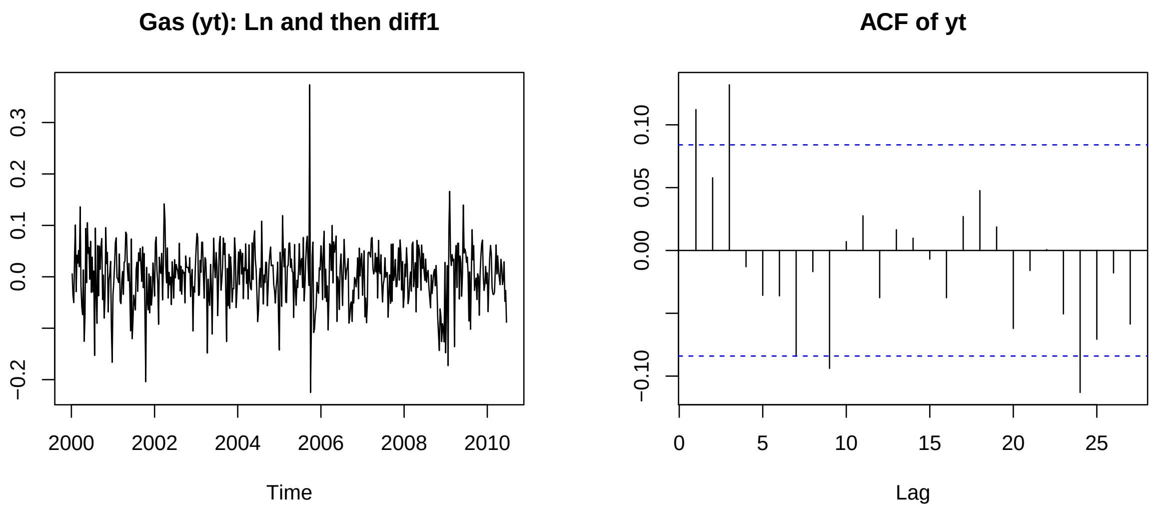


By using the log function on the data to compensate the difference of variance over time. Which means that we try to make the same variance over the hole time series.

c)

In this assignment we shall compute the first difference of the transformed data and then analyse the new data by ploting the series over time and look at the ACF.

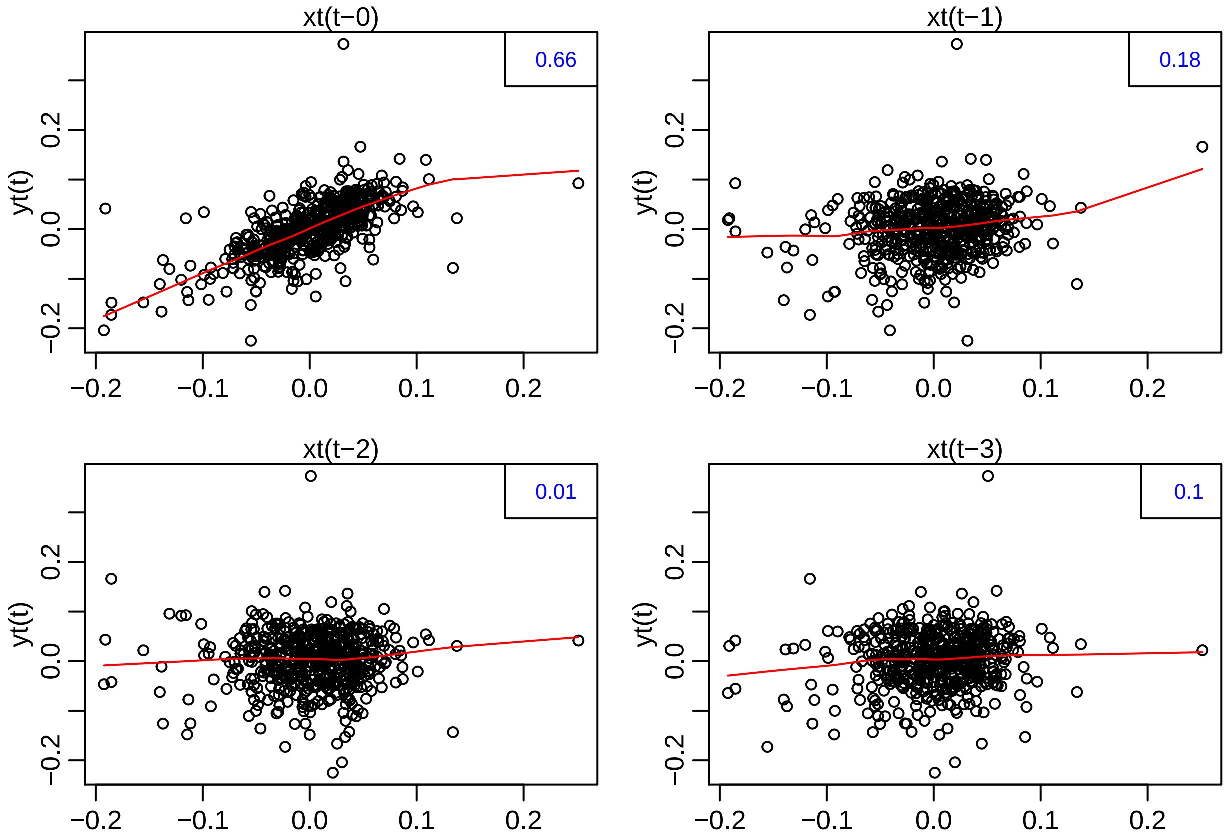




Both x_t and y_t seems to be stationary. We don't see any extreme spike at any lag for both series in the ACF.

d)

In this assignment we will plot y_t against x_t and it's three lag's. We did also include a CCF of the time series.



By looking at the graph's yt and xt seems to be correlated with a correlation of 0.66. It also seems that yt are correlated with x_{t-1}, through the correlation is 0.18. We can use this information for example explain or predict yt. It's looks like there are two outliers. One of the outliers is a outlier in y-axis (large value of y_t) and the other one is a outlier in x-axis (large value of x_t).

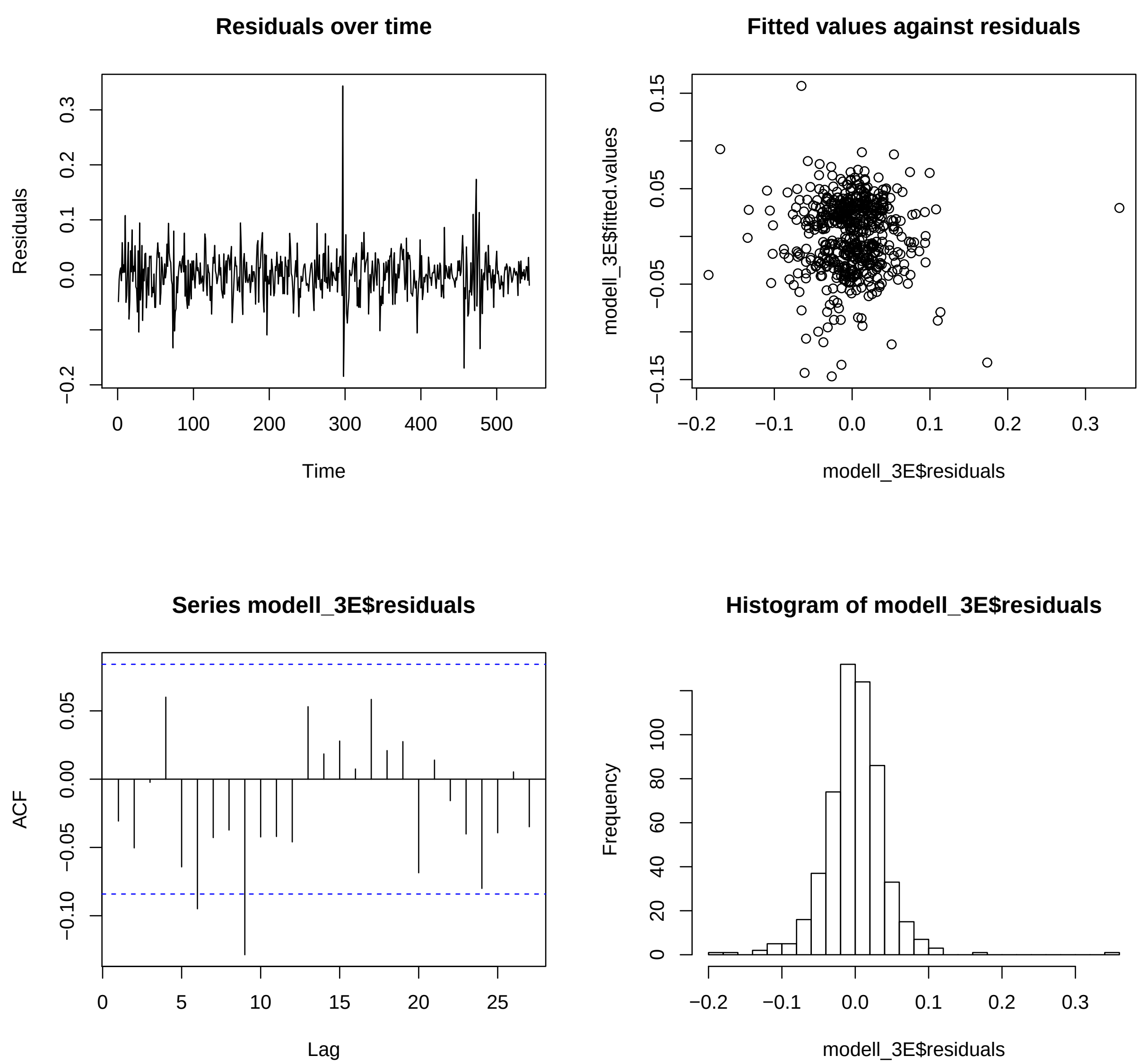
e)

In this assignment we will fit a regression that predict y_t by using the variables x_t, x_{t-1} and a indicator variable. Indicator variable is 1 if $x_t > 0$ and otherwise 0.

```
##
## Call:
## glm(formula = yt ~ xt + xt_lag1 + Indikator, data = Assignment3_data_E)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.18460  -0.02167  -0.00030   0.02176   0.34352
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006110   0.003455  -1.768  0.07759 .
## xt           0.799902   0.068269  11.717  < 2e-16 ***
## xt_lag1     -0.112152   0.038570  -2.908  0.00379 **
## Indikator    0.011785   0.005514   2.137  0.03303 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for gaussian family taken to be 0.001739879)
##
##     Null deviance: 1.72327  on 542  degrees of freedom
## Residual deviance: 0.93779  on 539  degrees of freedom
## AIC: -1903.2
##
## Number of Fisher Scoring iterations: 2
```

The variables x_t, x_{t-1} and the indicator is significant on a significance level $\alpha = 0.05$. The intercept is not significant which seems logical. When we difference the serie we set the $E[y_t] = 0$. Which itself means that no intercept is needed.



When looking at the residuals over time and residuals against fitted values we search if there is a pattern in data. We cant se any specific pattern in the two graph's, which is good.

We also want to check if the residuals are autocorrelated, which we check by look at the ACF. We can't see any big spikes at any lag, so it looks like there is no autocorrelation.

The model is reasonably good by just looking at the residuals.