732A62 Lab 1

Emil K Svensson & Rasmus Holm

2017-09-21

Assignment 1

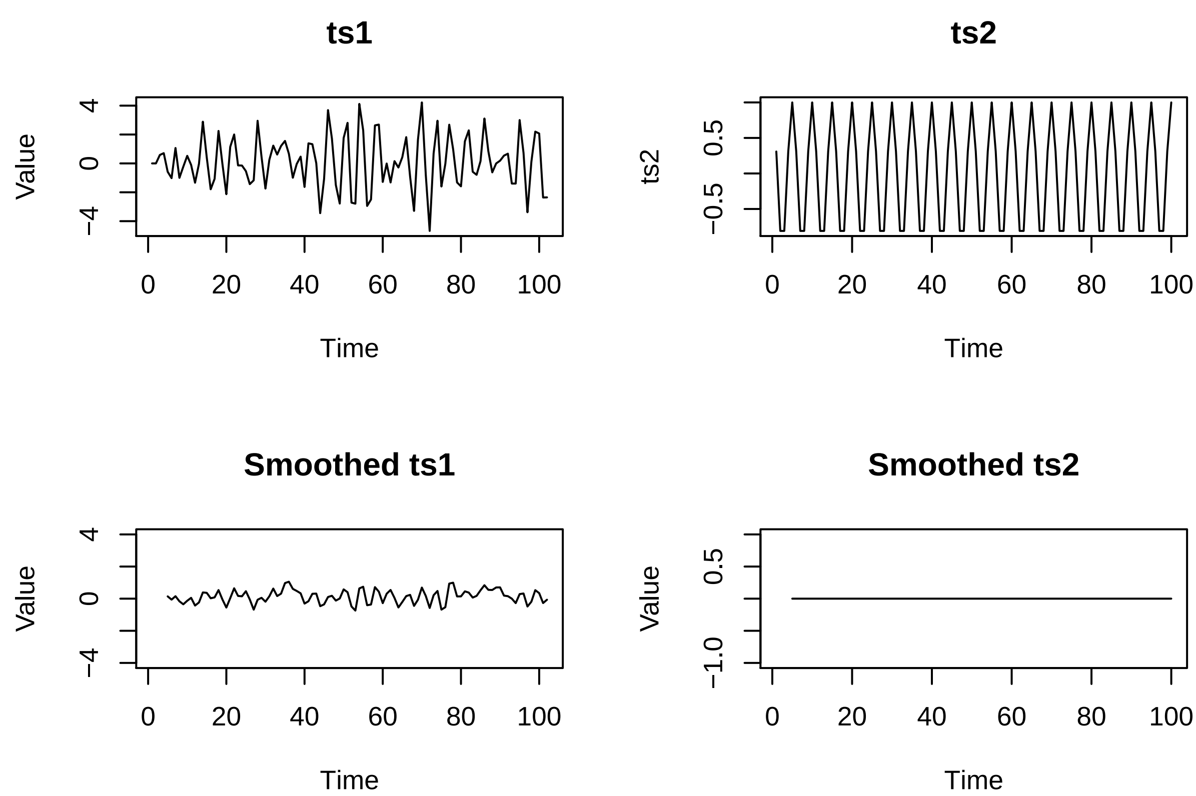
a)

```r
set.seed(12345)
t1 <- c(0, 0, rnorm(100, 0, 1))
ts1 <- filter(t1, filter = c(0, -0.8),
              method = "recursive", sides = 1)

t2 <- 1:100
ts2 <- cos(2 * pi * t2 / 5)

vt1 <- filter(ts1, filter = c(rep(0.2, times = 5)),
              method = "convolution", sides = 1)

vt2 <- filter(ts2, filter = c(rep(0.2, times = 5)),
              method = "convolution", sides = 1)
```



Time series 1 (ts1) show no noticeable change in its random pattern but the scale has been reduced. Time series 2 (ts2) is flattened by the smoothing filter and all values are now basically 0. This is because the average of ts2 lies around zero because the model is a non-translated cosine function. It is then reasonable to expect that the moving average smoother would give the same same result since the period is rather small.

b)

```r
leftside <- c(1, -4, 2, 0, 0, 1) # the x's
rightside <- c(1, 0, 3, 0, 1, 0, -4) # The w's

causal <- polyroot(leftside) #Not causal
invertible <- polyroot(rightside) #Non invertible

complex_dist <- function(x) {
    sqrt(Re(x)^2 + Im(x)^2)
}

print("The causal")
```

```
## [1] "The causal"
```

```r
sapply(causal, complex_dist)
```

```
## [1] 0.2936658 1.6793817 1.0000000 1.4239626 1.4239626
```

```r
print("The invertible")
```

```
## [1] "The invertible"
```

```r
sapply(invertible, complex_dist)
```

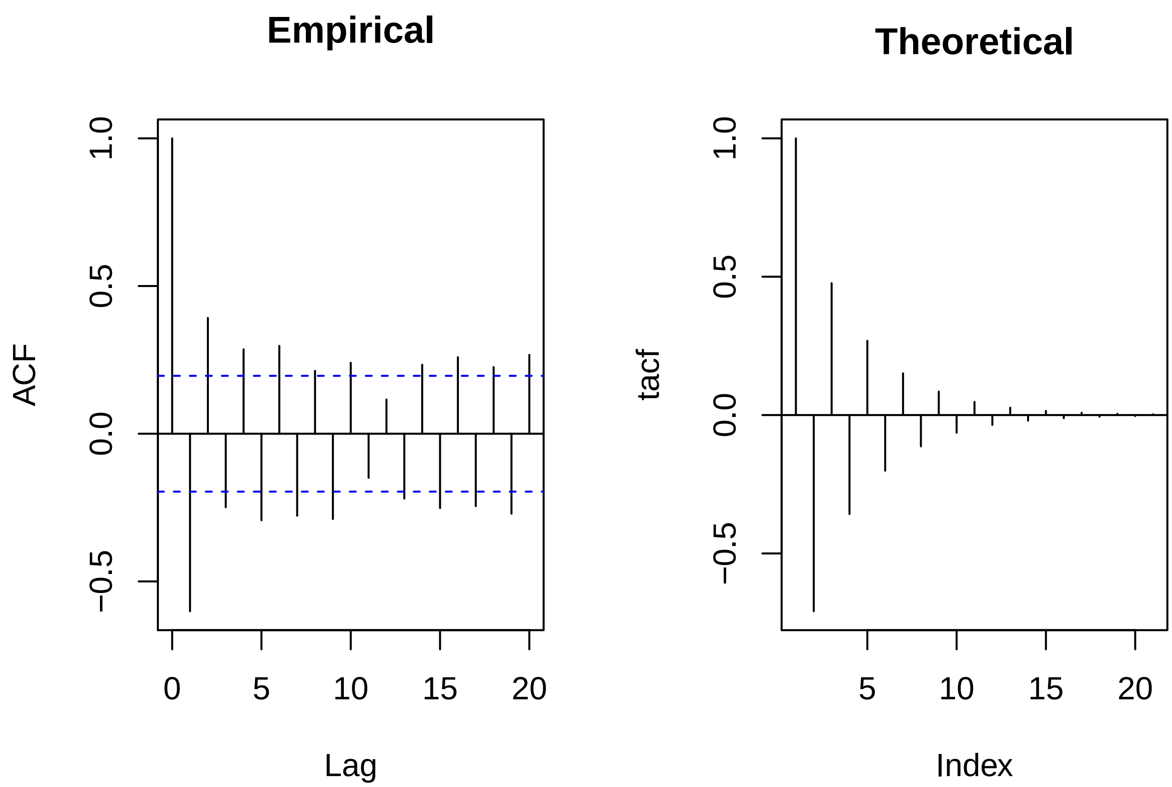
```
## [1] 0.6874372 0.6874372 0.6874372 0.6874372 1.0580446 1.0580446
```

Since both parts contains values below 1 they are inside the unit circle and the time series is therefore not causal nor invertible.

c)

```r
set.seed(54321)
model1c <- arima.sim(n = 100, list(ar = c(-3 / 4), ma = c(0, -1 / 9)))
tacf <- ARMAacf(ar=c(-3 / 4), ma=c(0, -1/9), lag.max=20)

old <- par(mfrow=c(1 , 2))
acf(model1c, main="Empirical")
plot(tacf, type="n", main="Theoretical")
segments(1:length(tacf), rep(0, length(tacf)), 1:length(tacf), tacf)
abline(h=0)
```




```r
par(old)
```

The ACF in the first plot shows a high correlation between all the lags except lag 11 and lag 12. Lag 2 has a larger correlation with the initial observation than the other lags. There seems to be no tendencies of the correlation dying down as the lags increases.

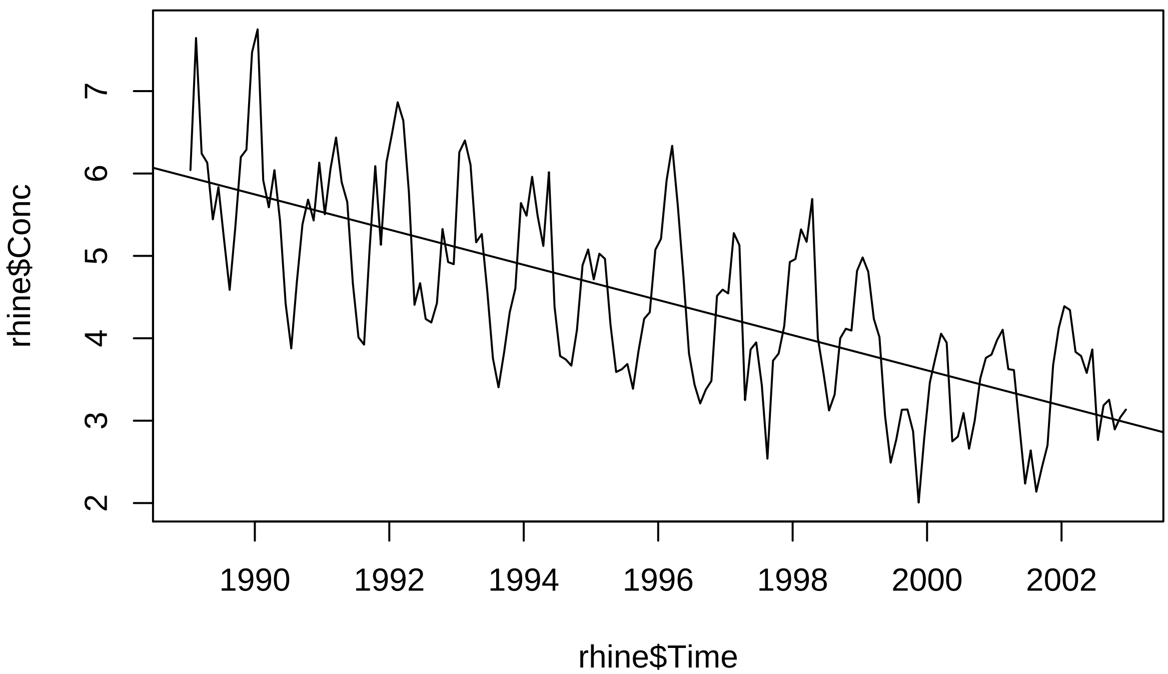
In the theoretical ACF it seems to indicate that the correlation should be dying down in a positive to negative correlations between the lags. This would indicate that the empirical autocorrelation does not always approximate the theoretical very well which is important to know in practical applications.

Assignment 2

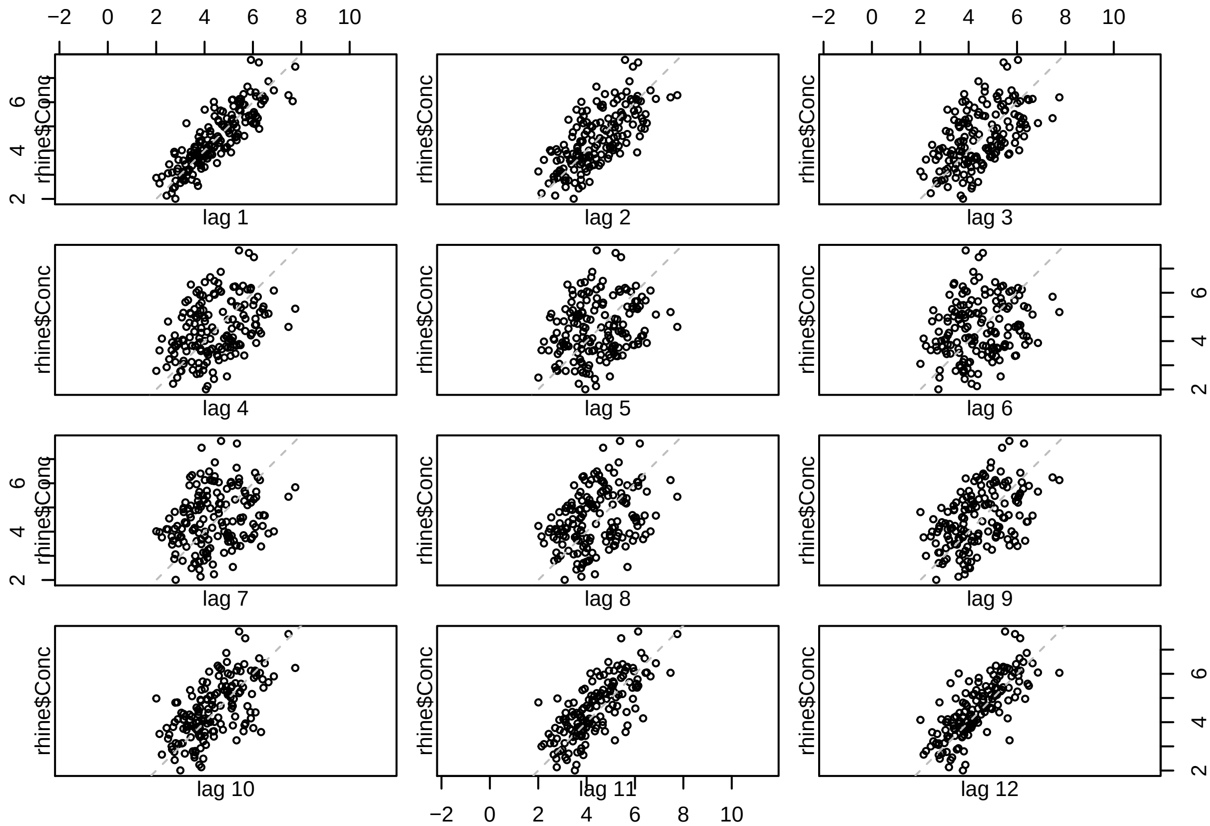
a)

```
rhine <- read.csv2("../data/Rhine.csv")
colnames(rhine)[4] <- "Conc"

lmobj <- lm(Conc ~ Time, data = rhine)
predobj <- predict(lmobj, se.fit = TRUE)
```



In the first plot there seems a monthly variation over all years with a linear decreasing trend with what could be considered a constant variance. It could be debated that it might be a lower variance in the end series.



There seems to be a strong correlation between lags 1, 11, and 12 and the original variable. A seasonal trend with a clear winter and summer period seems to be present.

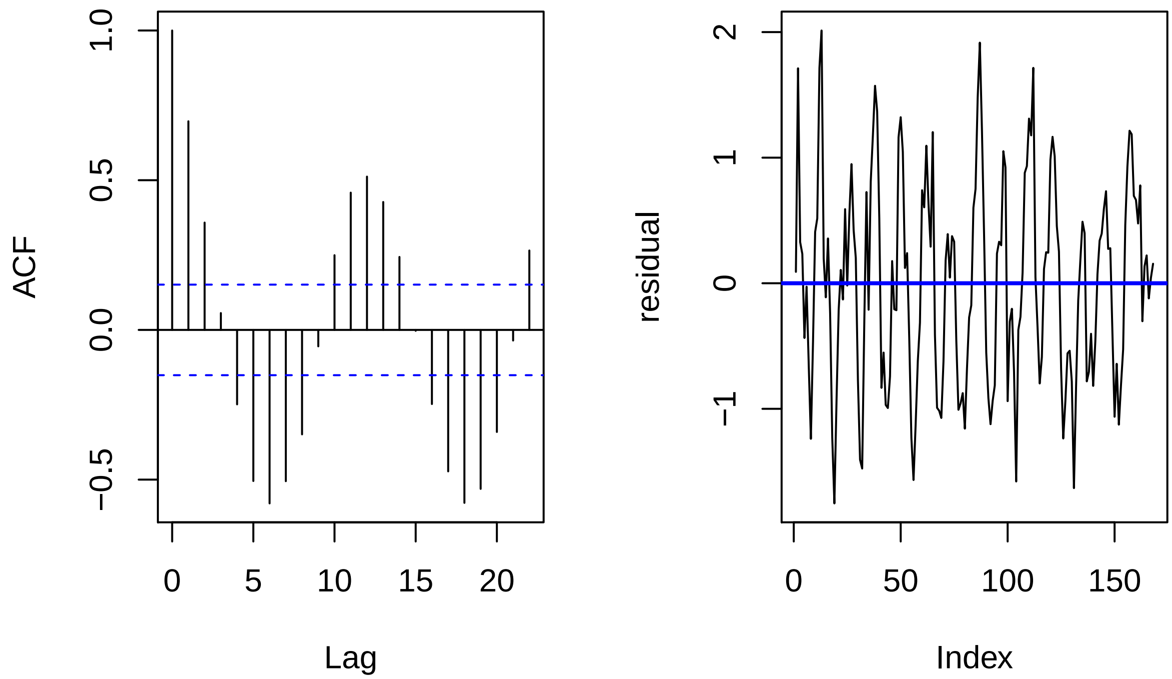
b)

```r
res1 <- residuals(lmobj)
summary(lmobj)
```

```
##
## Call:
## lm(formula = Conc ~ Time, 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) 430.70725   31.26570   13.78   <2e-16 ***
## Time         -0.21355    0.01566  -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
```

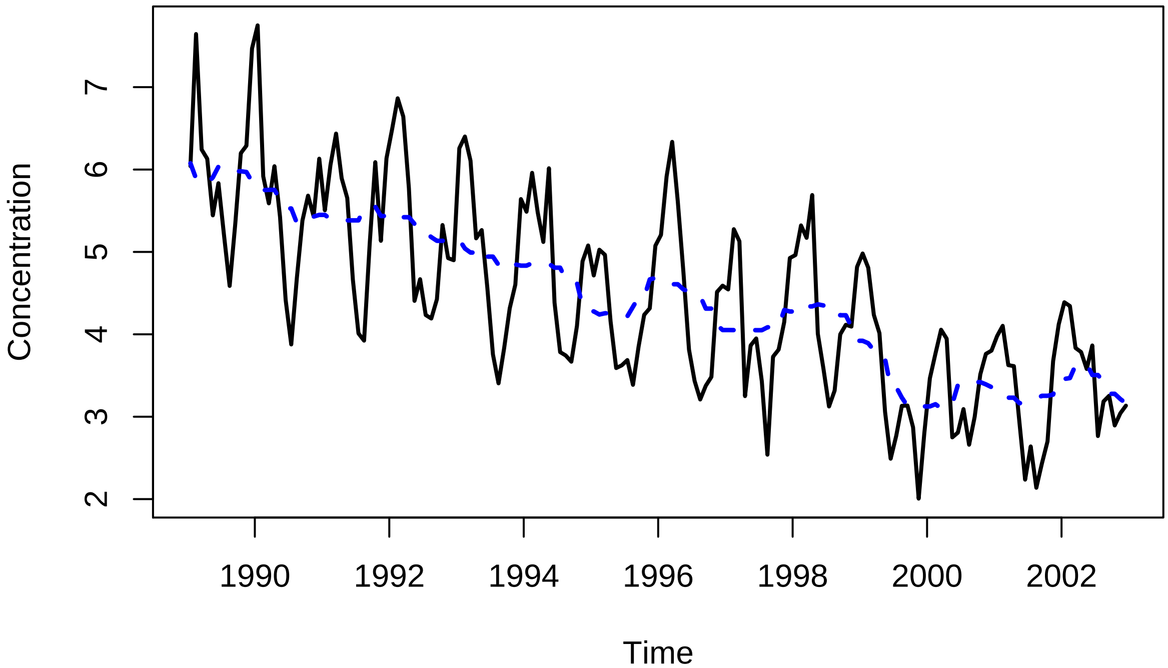
We can see from the linear fit that the concentration is reduced over time which is statistical significant.



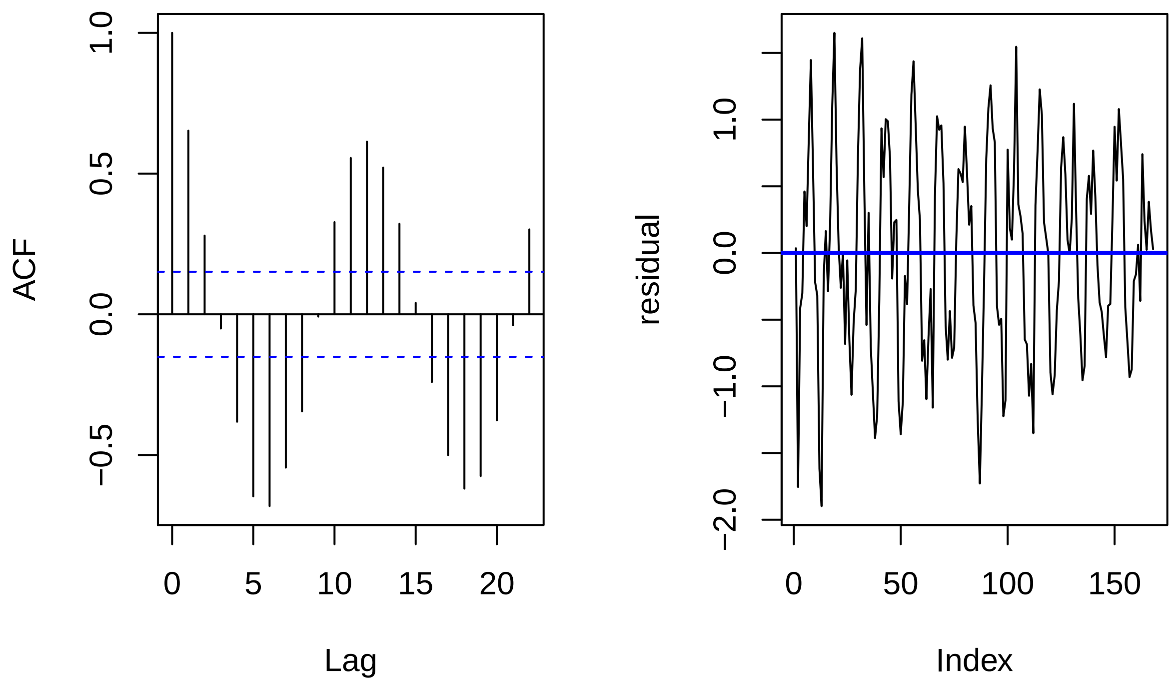
There is clearly a seasonality trend in the data that repeat every 12 or so months. The concentration increases during the winter months and decreases during the summer months. The same trend can be seen in the residuals.

c)

```
kersmo <- ksmooth(y = rhine$Conc, x = rhine$Time, bandwidth = 1)
kerres <- kersmo$y - rhine$Conc
```



We fitted a kernel smoother close to a linear model since the task given is to eliminate the trend, as seen in the first plot the fit seems to follow the general trend nicely.

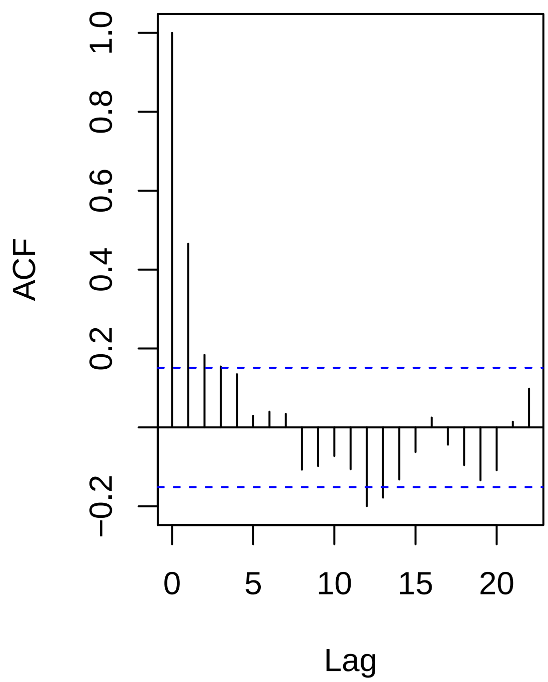


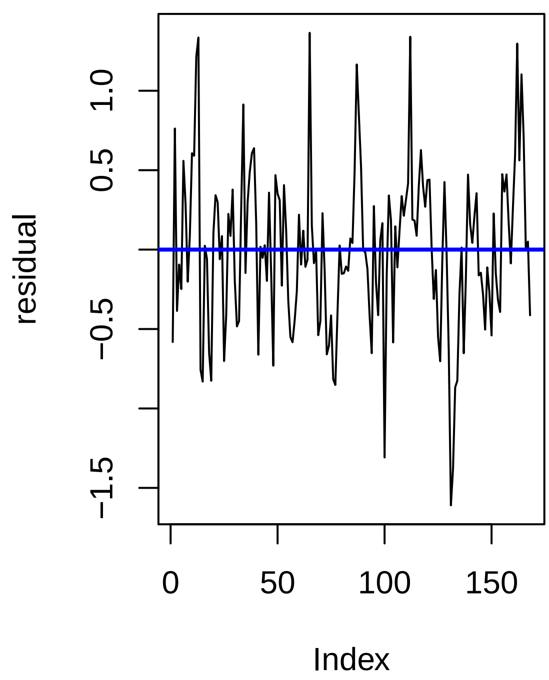
The residuals show a very high time dependence between each other which is not that big of suprise seeing how the model is fitted straight through the fluctuations between winters and summers.

The AFC confirms what has been seen in the previous plots, a strong seasonal pattern presents itself in the correlations between the lags.

d)

```
rhine$Month.f <- as.factor(rhine$Month)
dumlm <- lm(Conc ~ Time + Month.f, rhine)
```





The model with dummy-variables seems to have reduced variance of the residuals and there seem to be less time-dependency between residuals.

Judging from the AFC we come to a similar conclusion where most of the correlation between lags have disappeared, some correlation seem to remain in the second lag.

e)

```r
library(MASS)
stepAIC(dumlm, direction = "backward", steps = 1000)
```

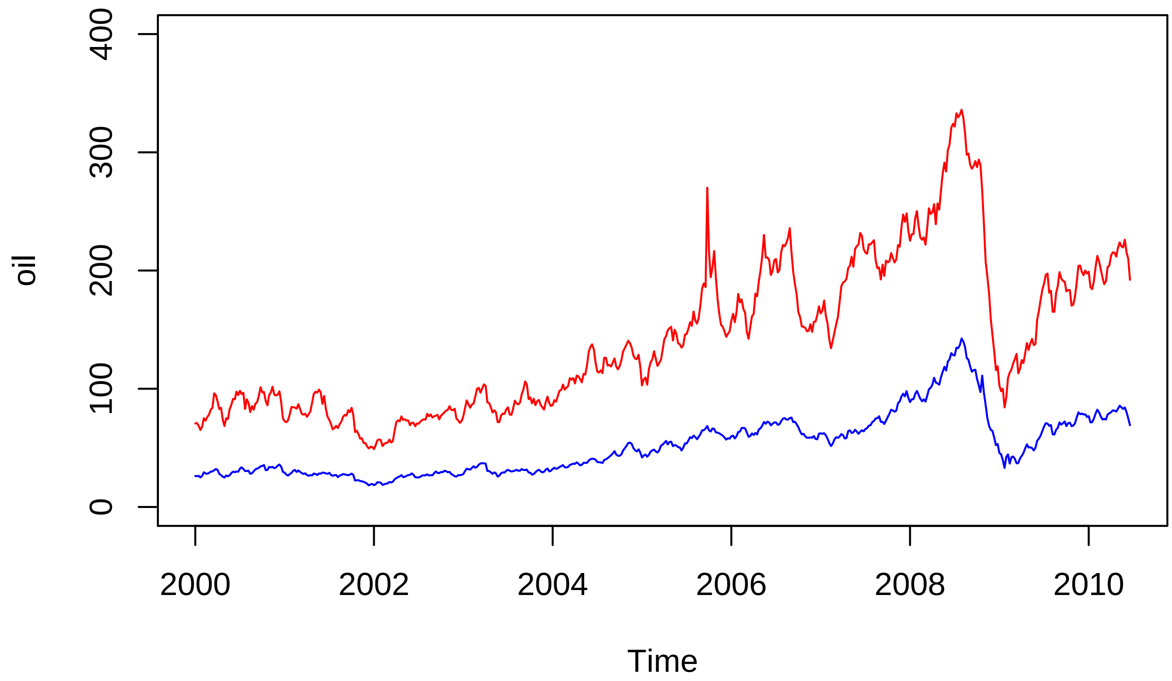
```
## Start:  AIC=-202.02
## Conc ~ Time + Month.f
##
##           Df Sum of Sq     RSS      AIC
## <none>                  43.237 -202.023
## - Month.f 11    68.524 111.761  -64.477
## - Time     1   118.387 161.624   17.499
##
## Call:
## lm(formula = Conc ~ Time + Month.f, data = rhine)
##
## Coefficients:
## (Intercept)         Time     Month.f2     Month.f3     Month.f4
##   420.82746     -0.20824      0.27659      0.04006     -0.34643
```

```
##    Month.f5    Month.f6    Month.f7    Month.f8    Month.f9
##    -0.86165    -1.26114    -1.60808    -1.71242    -1.23669
##   Month.f10   Month.f11   Month.f12
##    -0.87446    -0.75127    -0.17745
```

The backward-elimination procedure ends up with the full model with all dummy variables as the optimal model.

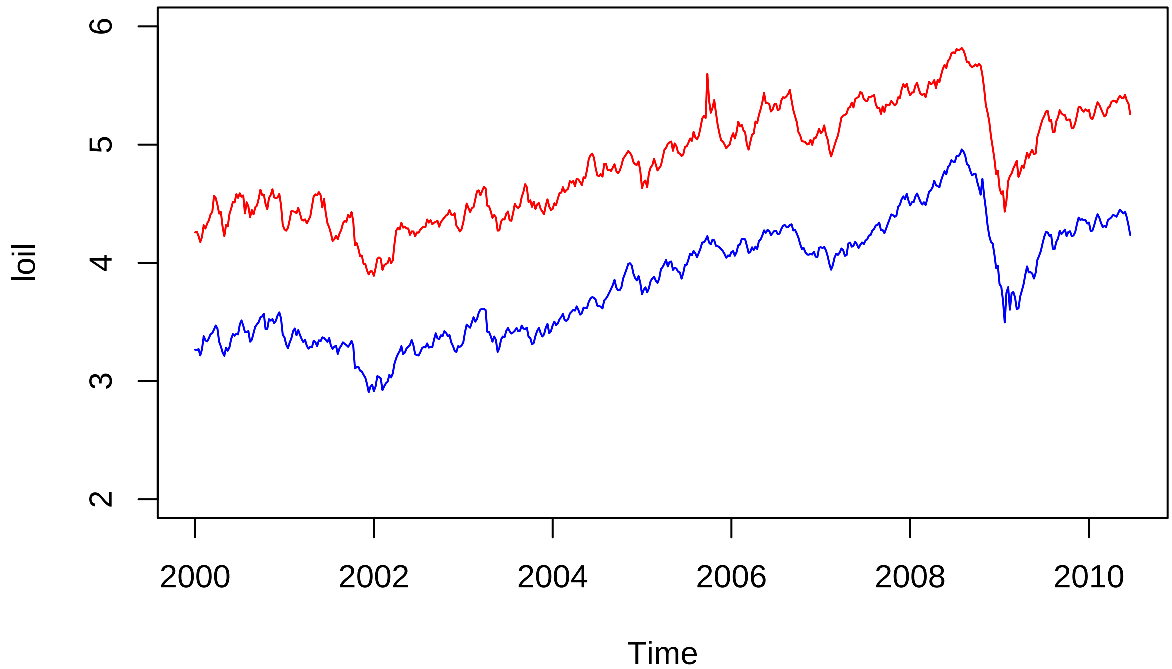
Assignment 3

a)



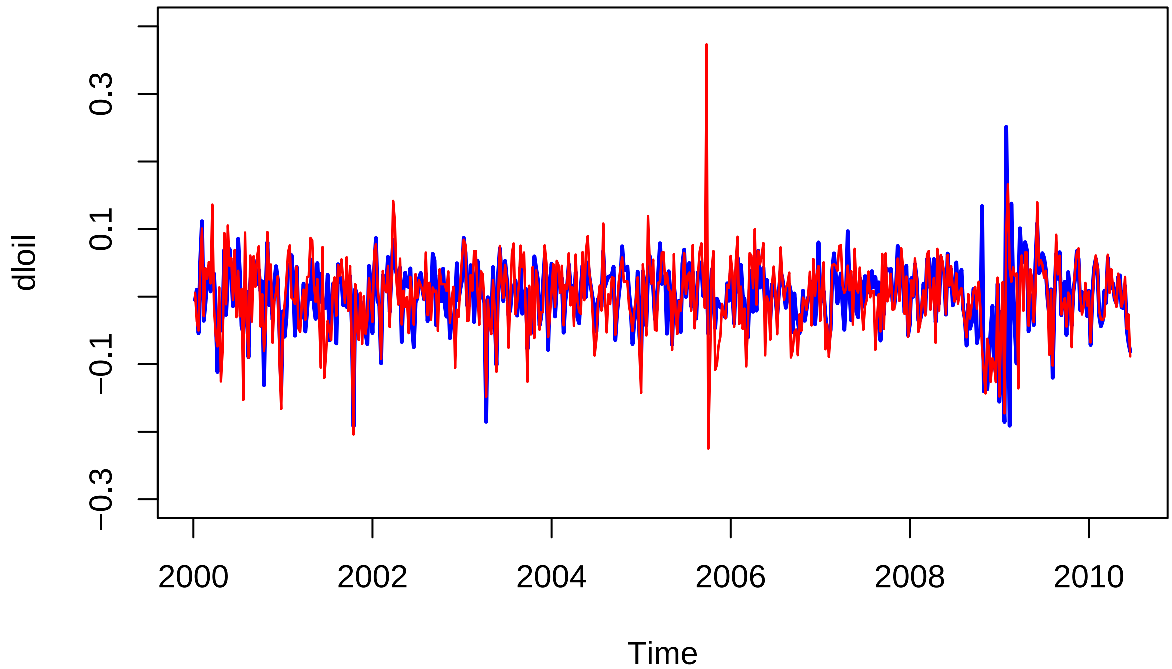
The time series do not look stationary, the mean value seem to change depending on time, and there is no particular pattern happening over time. However, the two time series do have similar shapes indicating that they are related to each other.

b)

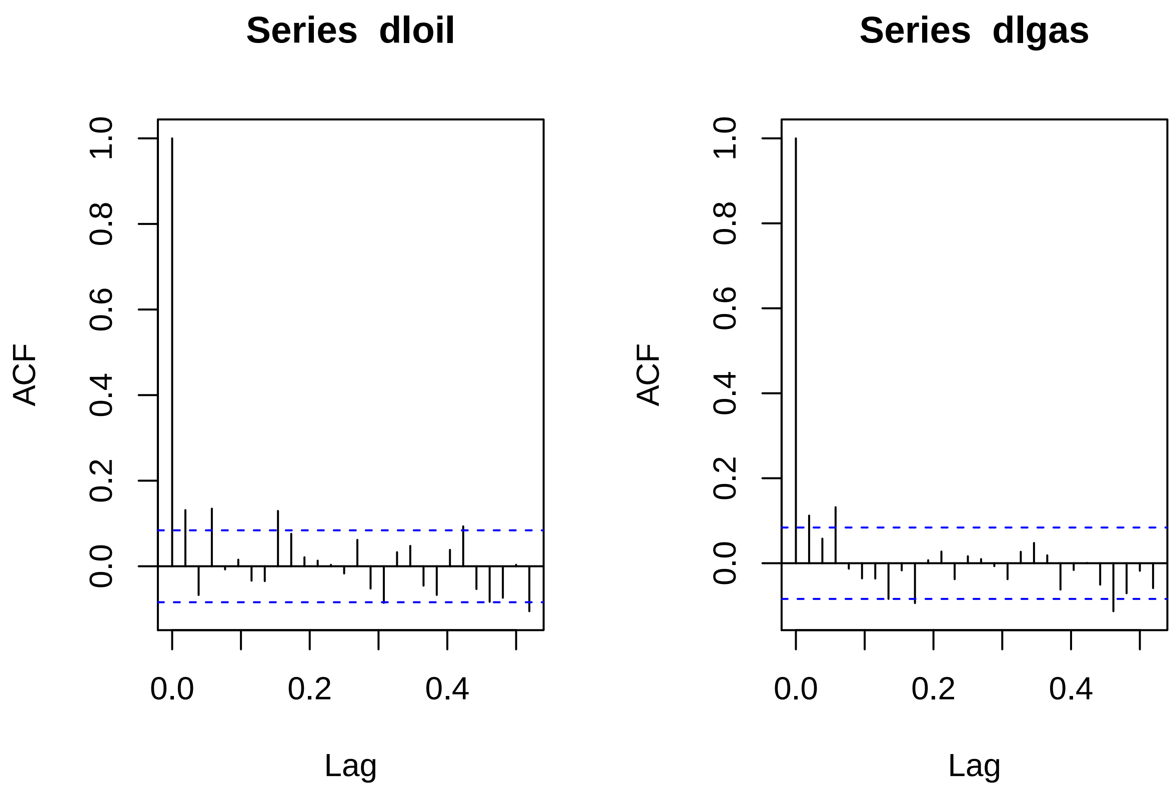


The log-transformation put the two time series in similar scales which makes it easier to compare them in detail. The fluctuations are clearly similar in both time series. Although we can not call the time series stationary.

c)

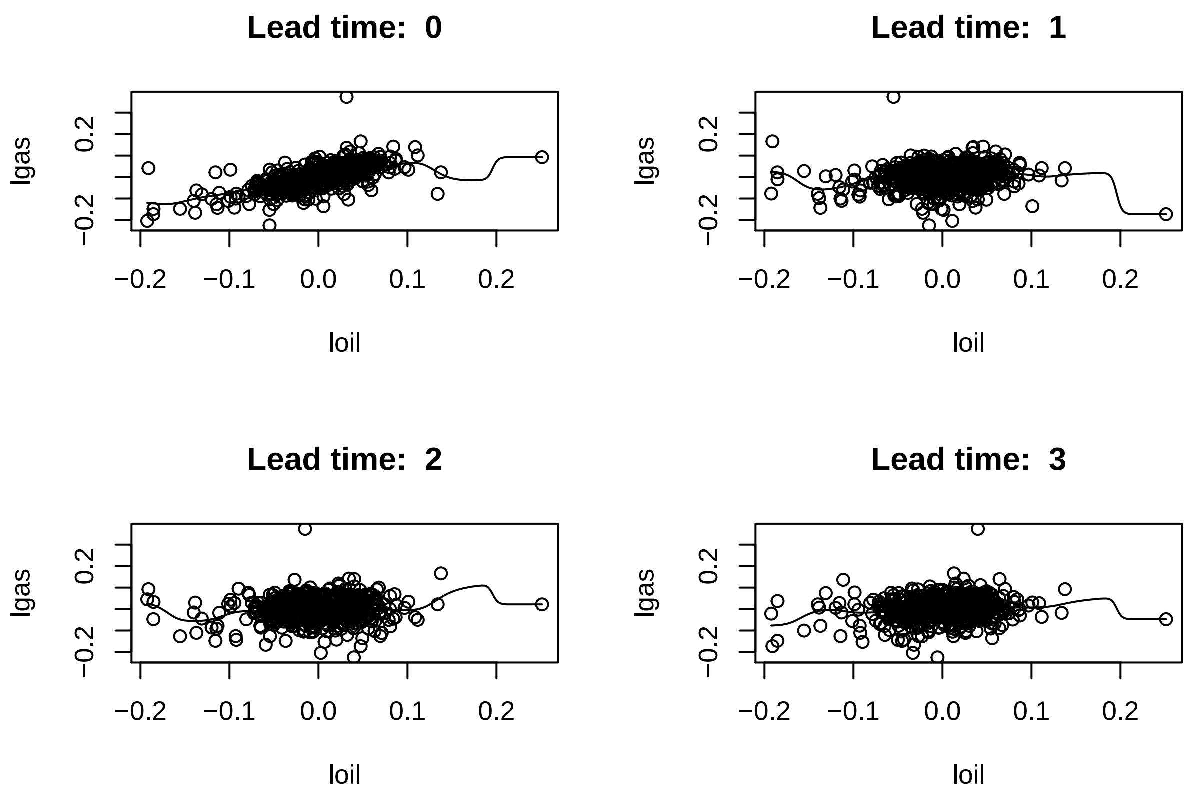


In the plot above we see the two time series after computing the first difference and we can see a resemblance of the two with a few different peaks, but otherwise quite similar. The mean seems more stable except a wild fluctuation around 2009 and could be considerd stationary now.



The autocorrelation plots show that there is not much correlation in the data after transformations.

d)



It does not seem to be any relationship between oil and gas when oil has a lead time. Otherwise they do have a positive correlation as shown in the top-left plot. There are some outliers in the data, especially the two points that have high oil respectively gas values.

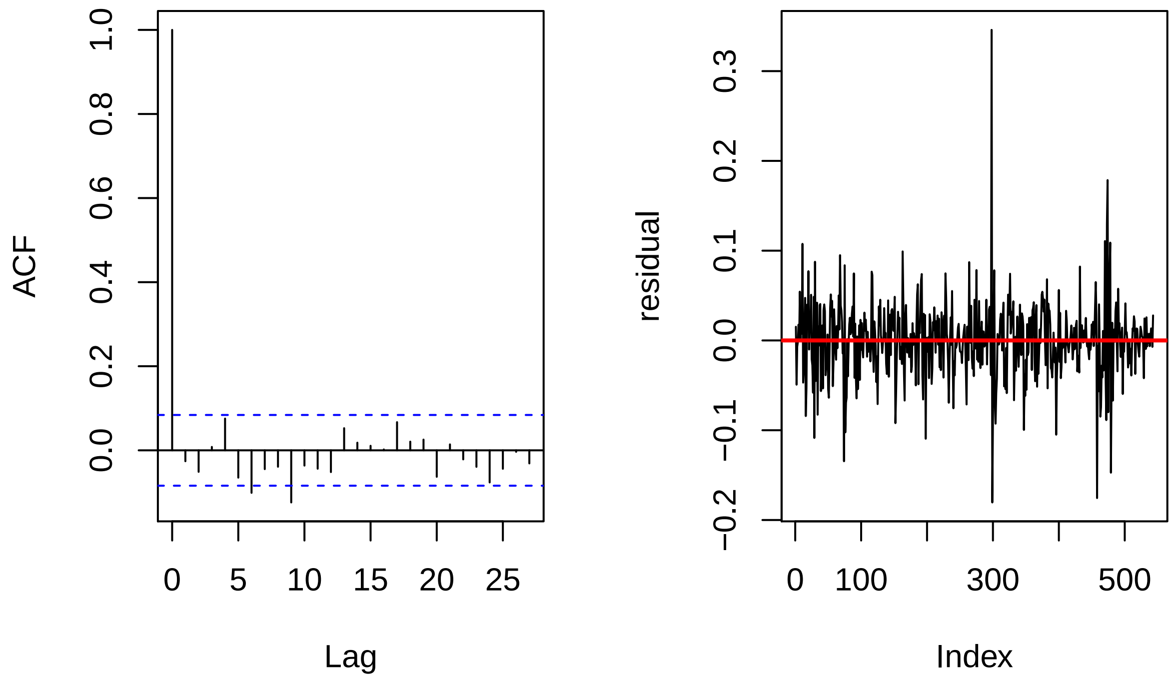
e)

```r
tss <- ts.intersect(yt=yt, xt=xt, lag1xt=lag(xt, 1), dummy=xt > 0)
model3e <- lm(yt ~ xt + lag1xt + dummy, data = tss)
resid3e <- resid(model3e)
summary(model3e)
```

```
##
## Call:
## lm(formula = yt ~ xt + lag1xt + dummy, data = tss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18044 -0.02103  0.00003  0.02170  0.34592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006176   0.003470  -1.780   0.0757 .
## xt           0.694200   0.058898  11.786   <2e-16 ***
```

```
## lag1xt       0.012660   0.038729   0.327   0.7439
## dummy        0.012376   0.005542   2.233   0.0259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04202 on 539 degrees of freedom
## Multiple R-squared:  0.445,  Adjusted R-squared:  0.4419
## F-statistic: 144.1 on 3 and 539 DF,  p-value: < 2.2e-16
```

We can see from the linear fit that lag is not significant which supports our analysis so far that there does not seem be any correlation from previous/future data, only from present. The dummy variable is significant and has a positive coefficient indicating that if there is an increase in oil, there is also an increase in gas.



The residuals do look stationary with a few outlying peaks. There seems to be no dependency neither in short nor long term. The autocorrelations for the residuals also indicate that there is no linear relationship between past/future values.