Timeseries-lab1

Prudhvi Peddmallu(prupe690),Zhixuan Duan(zhidu838)

11 September 2019

Assignment 1. Computations with simulated data

Assignment 1. Computations with simulated data

a) Generate two time series $x_t = -0.8x_{t-2} + w_t$, where $x_0 = x_1 = 0$ and $x_t = \cos\left(\frac{2\pi t}{5}\right)$ with 100 observations each. Apply a smoothing filter $v_t = 0.2(x_t + x_{t-1} + x_{t-2}+x_{t-3} + x_{t-4})$ to these two series and compare how the filter has affected them.

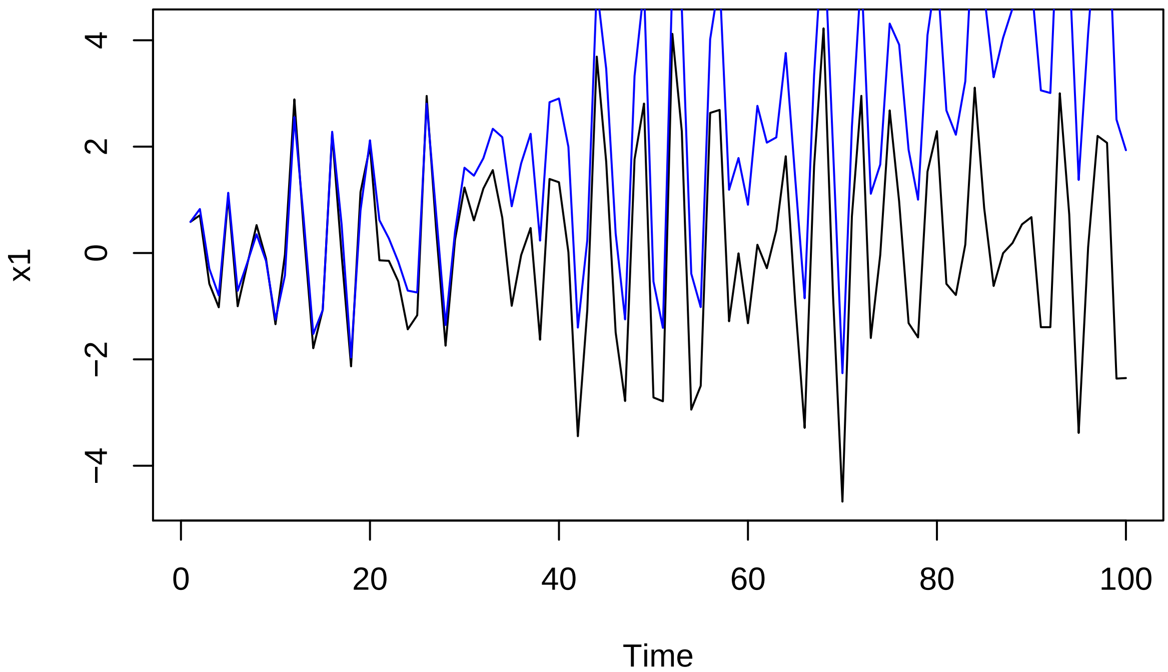
b) Consider time series $x_t - 4x_{t-1} + 2x_{t-2} + x_{t-5} = w_t + 3w_{t-2} + w_{t-4} - 4w_{t-6}$. Write an appropriate R code to investigate whether this time series is casual and invertible.

c) Use built-in R functions to simulate 100 observations from the process $x_t + \frac{3}{4}x_{t-1} = w_t - \frac{1}{9}w_{t-2}$, compute sample ACF and theoretical ACF, use seed 54321. Compare the ACF plots.

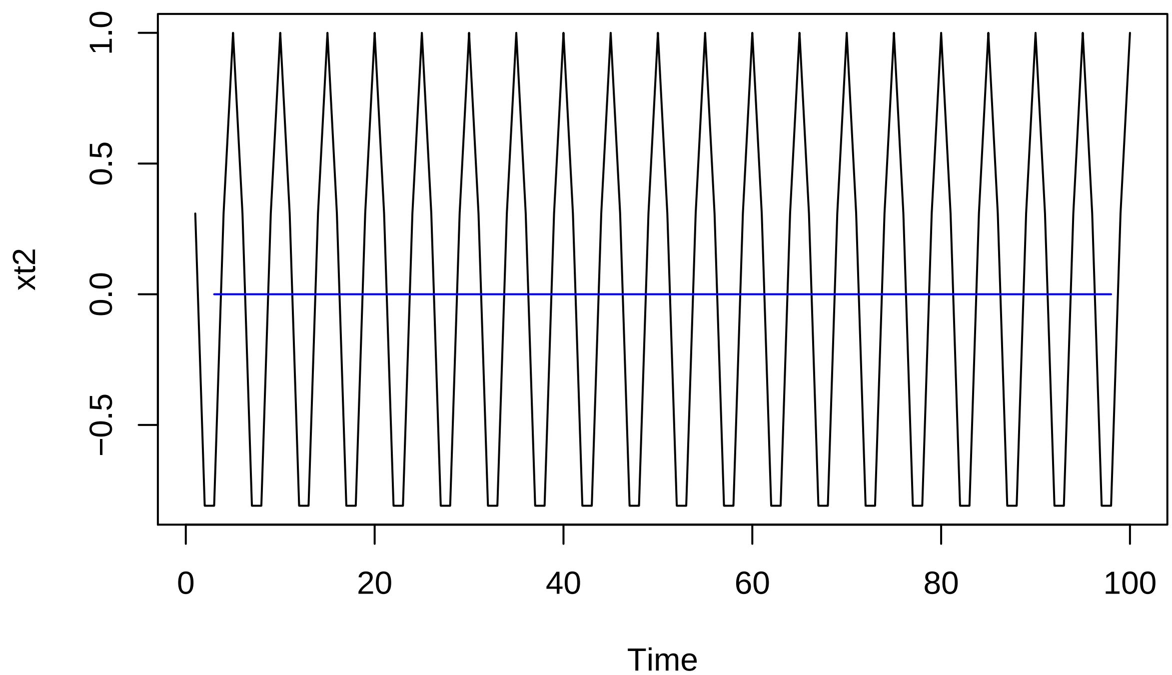
1a)

Analysis:- The filter has made the time series more smooth for time series 1,the series has less variation on the horizontal axis. The cosine time series2, we can see that smoothing filter makes the time series completely horizontal.

```r
set.seed(12345)
wt <- rnorm(100, 0, 1)
x1 <- filter(wt, c(0, -0.8), method = "recursive")
v1 <- filter(x1, rep(0.2, 5), sides = 1, method = "recursive")
plot.ts(x1)
lines(v1, col = "blue")
```




```
xt2<-rep(0,100)#100 observations
#wt is the white noise always r norm
for (i in 1:length(xt2)) {
  xt2[i] <- cos(2*pi*i/5)
}
v2<-filter(xt2,rep(0.2, 5))
plot.ts(xt2)
lines(v2, col = "blue")
```



1b)

Analysis:-The process is not invertible and not causal.

```
xt<-c(1,-4,2,0,0,1)
xt_polyroot<-polyroot(xt)
xt_polyroot
```

```
## [1]  0.2936658+0.000000i -1.6793817+0.000000i  1.0000000-0.000000i
## [4]  0.1928579-1.410842i  0.1928579+1.410842i
```

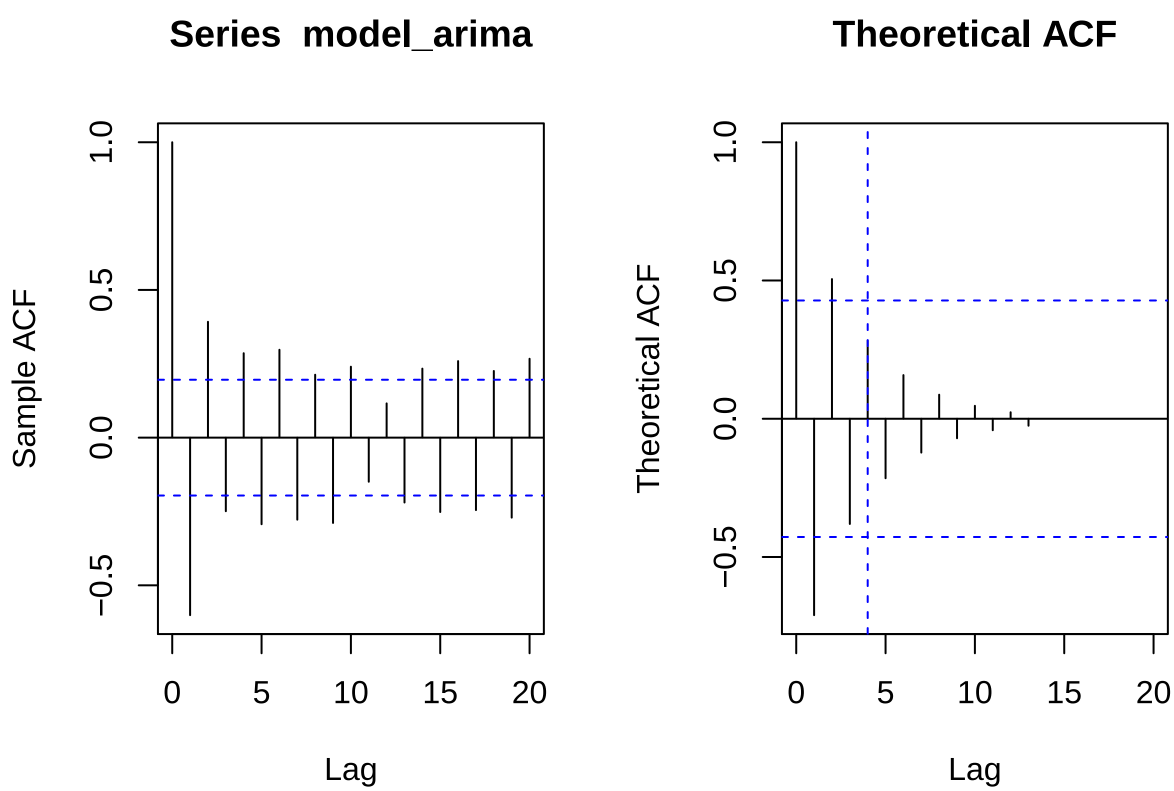
```
wt<-c(1,0,3,0,1,-4)
wt_polyroot<-polyroot(wt)
wt_polyroot
```

```
## [1]  0.0728440+0.5667298i -0.4872808+0.6872411i  0.0728440-0.5667298i
## [4]  1.0788736-0.0000000i -0.4872808-0.6872411i
```

1c)

Analysis:-The main difference between the theoretical and the sample ACF is that the autocorrelations in the theoretical ACF plot are given by a recursive filter when our lag > 3 in the plot 2.

```
#Set seed Given in question
set.seed(54321)
#generate arima model
model_arima <- arima.sim(n = 100, list(ar = c(-3/4), ma = c(0,-1/9)))
#plots
par(mfrow = c(1,2))
acf_test <- acf(model_arima, ylab = "Sample ACF")
acf(ARMAacf(ar = c(-3/4), ma = c(0,-1/9), lag.max = 20), ylab = "Theoretical ACF",
    main = "Theoretical ACF", xlim = c(0,20))
abline(v = 4, col = "blue", lty = 2)
```



Assignment 2. Visualization, detrending and residual analysis of Rhine data

The data set **Rhine.csv** contains monthly concentrations of total nitrogen in the Rhine River in the period 1989-2002.

a) Import the data to R, convert it appropriately to *ts* object (use function *ts()*) and explore it by plotting the time series, creating scatter plots of x_t against $x_{t-1}, \dots x_{t-12}$. Analyze the time series plot and the scatter plots: Are there any trends, linear or seasonal, in the time series? When during the year is the concentration highest? Are there any special patterns in the data or scatterplots? Does the variance seem to change over time? Which variables in the scatterplots seem to have a significant relation to each other?

b) Eliminate the trend by fitting a linear model with respect to t to the time series. Is there a significant time trend? Look at the residual pattern and the sample ACF of the residuals and comment how this pattern might be related to seasonality of the series.

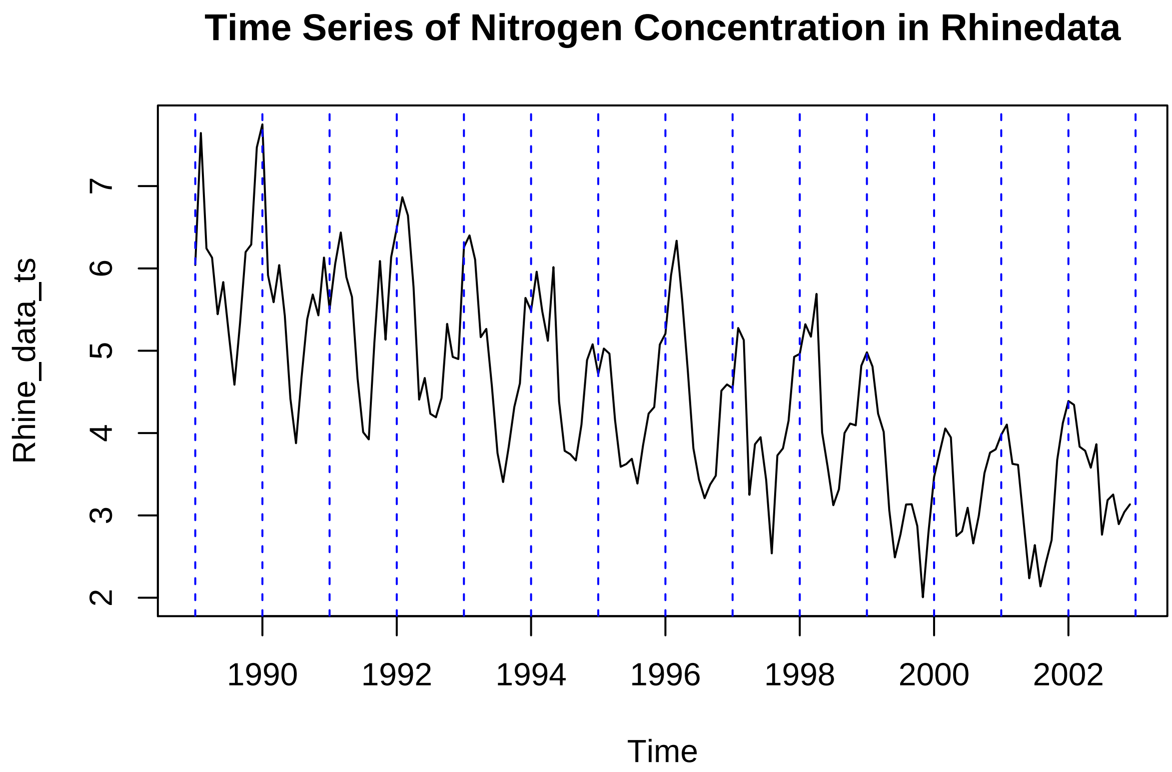
c) Eliminate the trend by fitting a kernel smoother with respect to t to the time series (choose a reasonable bandwidth yourself so the fit looks reasonable). Analyze the residual pattern and the sample ACF of the residuals and compare it to the ACF from step b). Conclusions? Do residuals seem to represent a stationary series?

d) Eliminate the trend by fitting the following so-called seasonal means model:

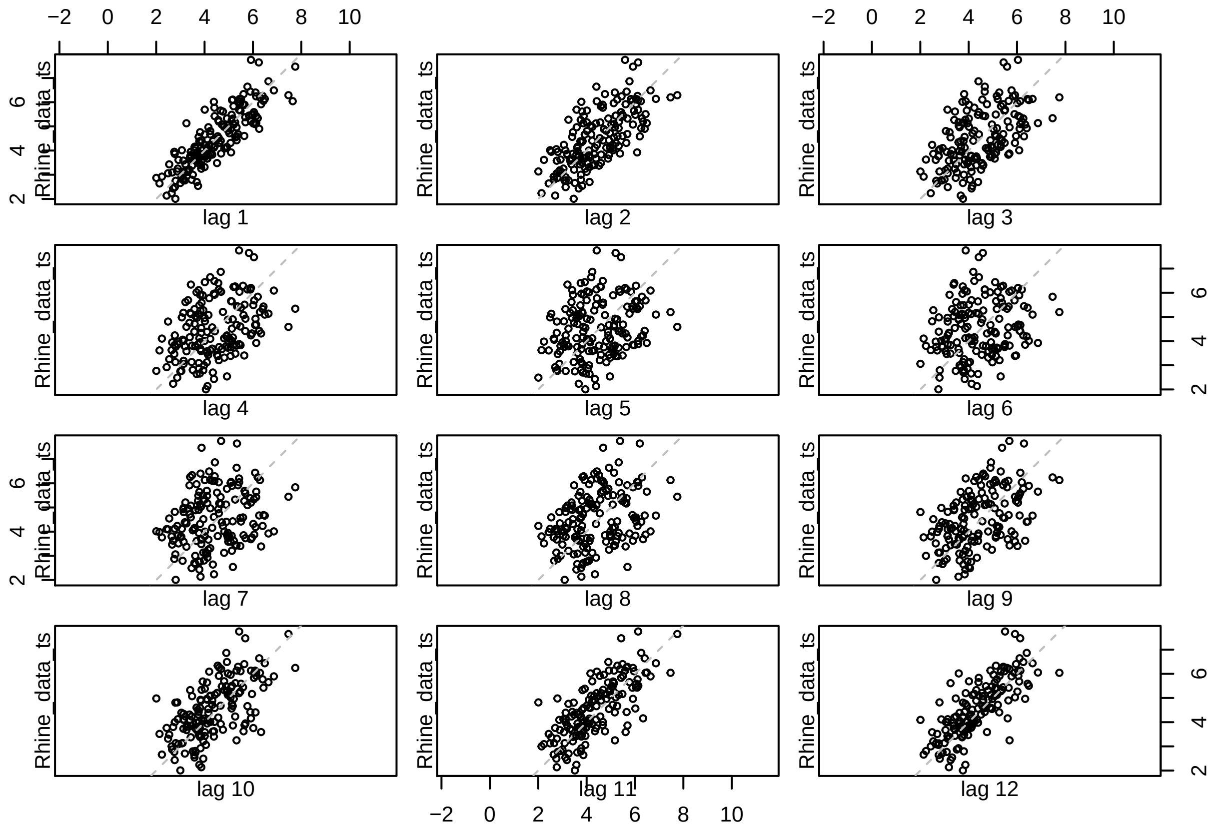
$$x_t = \alpha_0 + \alpha_1 t + \beta_1 I(month = 2) + \cdots + \beta_{12} I(month = 12) + w_t,$$

2a)

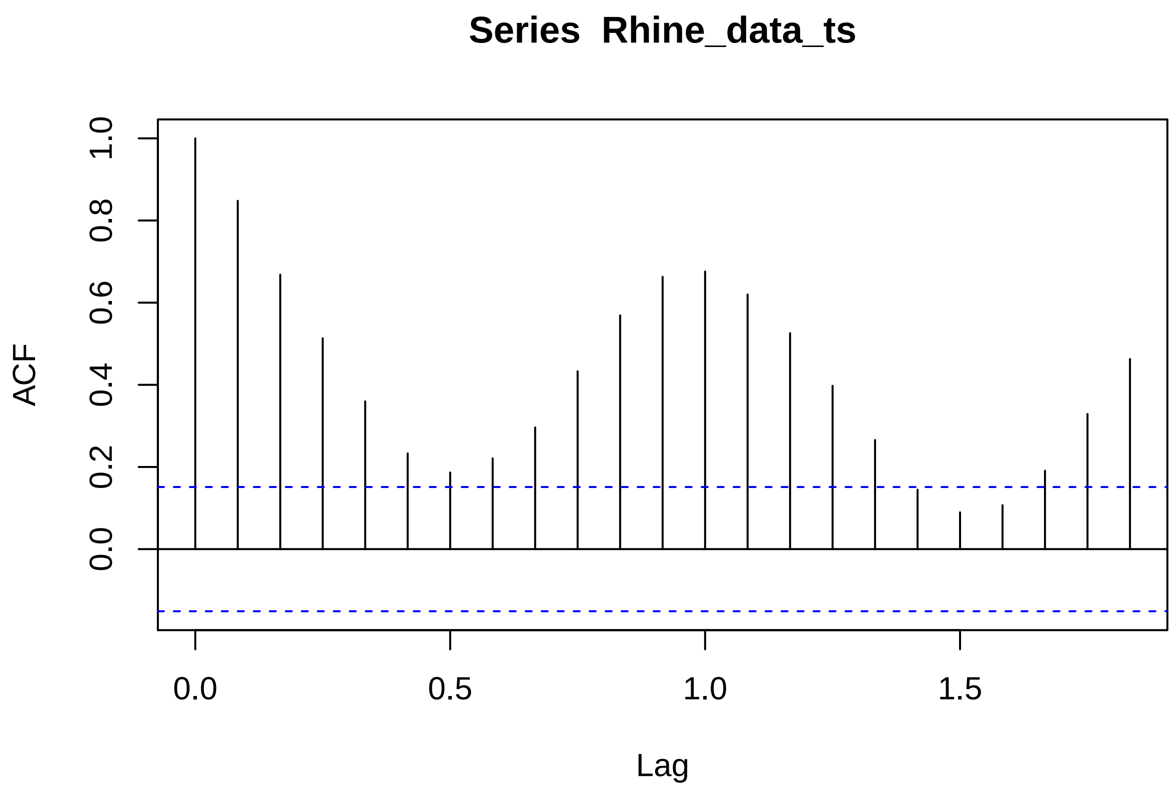
```
#a
set.seed(12345)
Rhine_data <- read.csv2("Rhine.csv")
Rhine_data_ts <- ts(data = Rhine_data$TotN_conc,start = c(1989,1),frequency = 12)#frequency
#is xt-1...xt-12
plot.ts(Rhine_data_ts, main="Time Series of Nitrogen Concentration in Rhinedata")
abline(v = c(1989:2004), col = "blue", lty = 2)
```




```r
lag.plot(Rhine_data_ts,lags = 12)#lags are 12 because xt-1 ...xt-12
```




```
acf(Rhine_data_ts)
```



Analysis:-From the plot 1 we can see that there is a diminishing pattern and there appear to be some sort of regular impact since a similar pattern happens each year. We can likewise see that the pattern is diminishing start from January every year and expanding during the part of the bargain. From the plot 2 where we can see that the Nitrogen Concentration has a progressively direct pattern in the first and a months ago of the year.

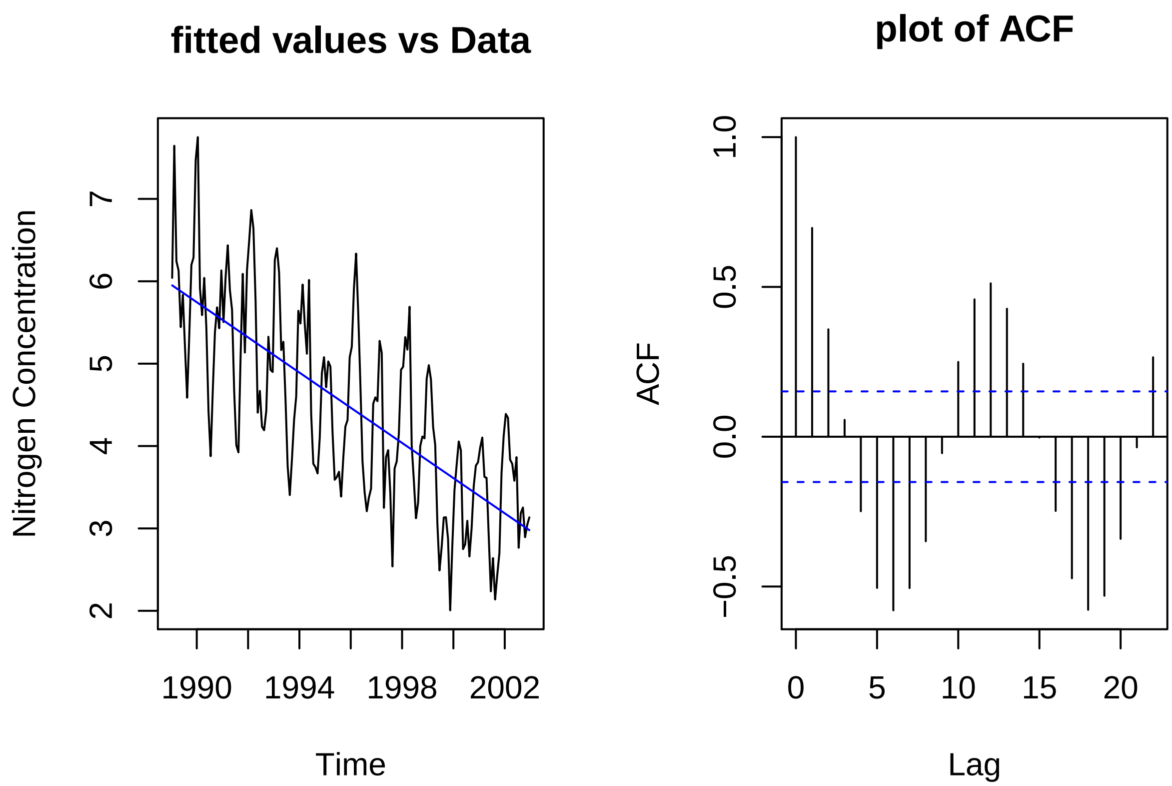
2b)

```r
#fitting a linear model to eliminate the trend
linear_model <- lm(TotN_conc ~ Time, data = Rhine_data)#result list of 12 objects
summary(linear_model)
```

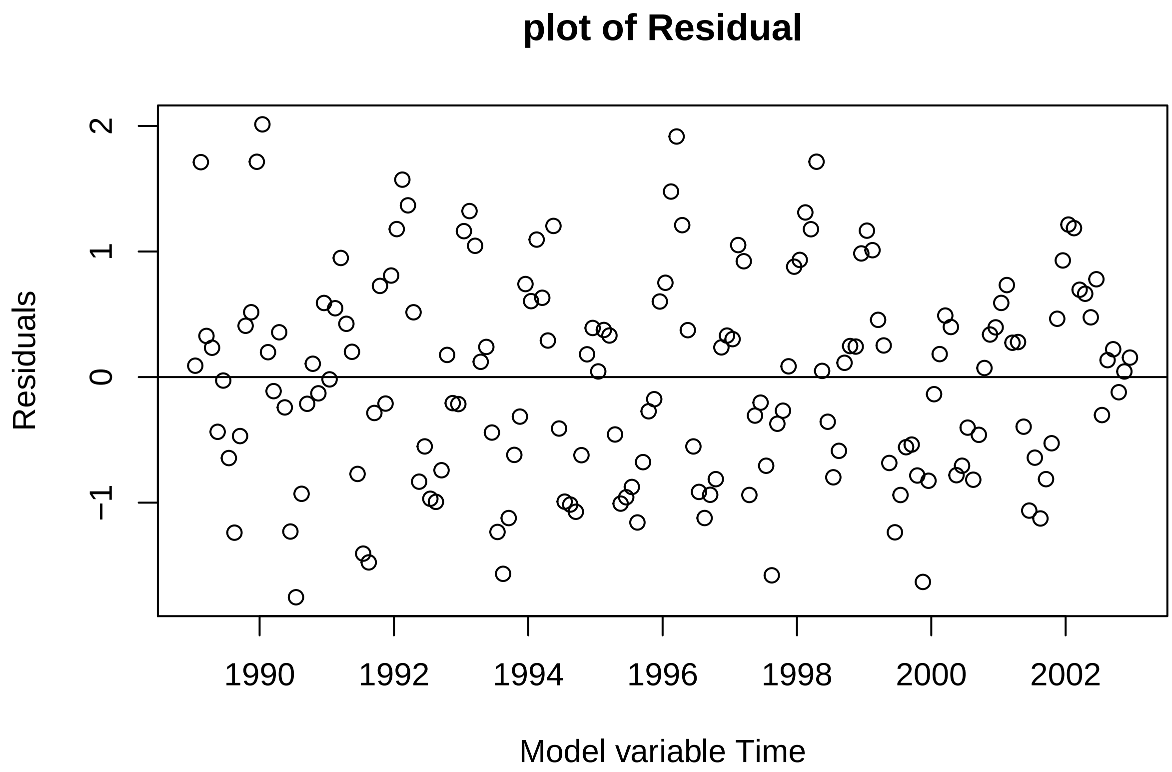
```
##
## Call:
## lm(formula = TotN_conc ~ Time, data = Rhine_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.75325 -0.65296  0.06071  0.52453  2.01276
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 430.70725   31.26570   13.78   <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
```

```
#Plot fitted values vs data
par(mfrow = c(1,2))
plot(y = Rhine_data$TotN_conc, x = Rhine_data$Time, type = "l",
main = "fitted values vs Data",
ylab = "Nitrogen Concentration",
xlab = "Time")
lines(y = linear_model$fitted.values, x = linear_model$model$Time,
col = "blue")#in the list linear model one variable is fitted values
#ACF plot
acf(linear_model$residuals, main = "plot of ACF")
```



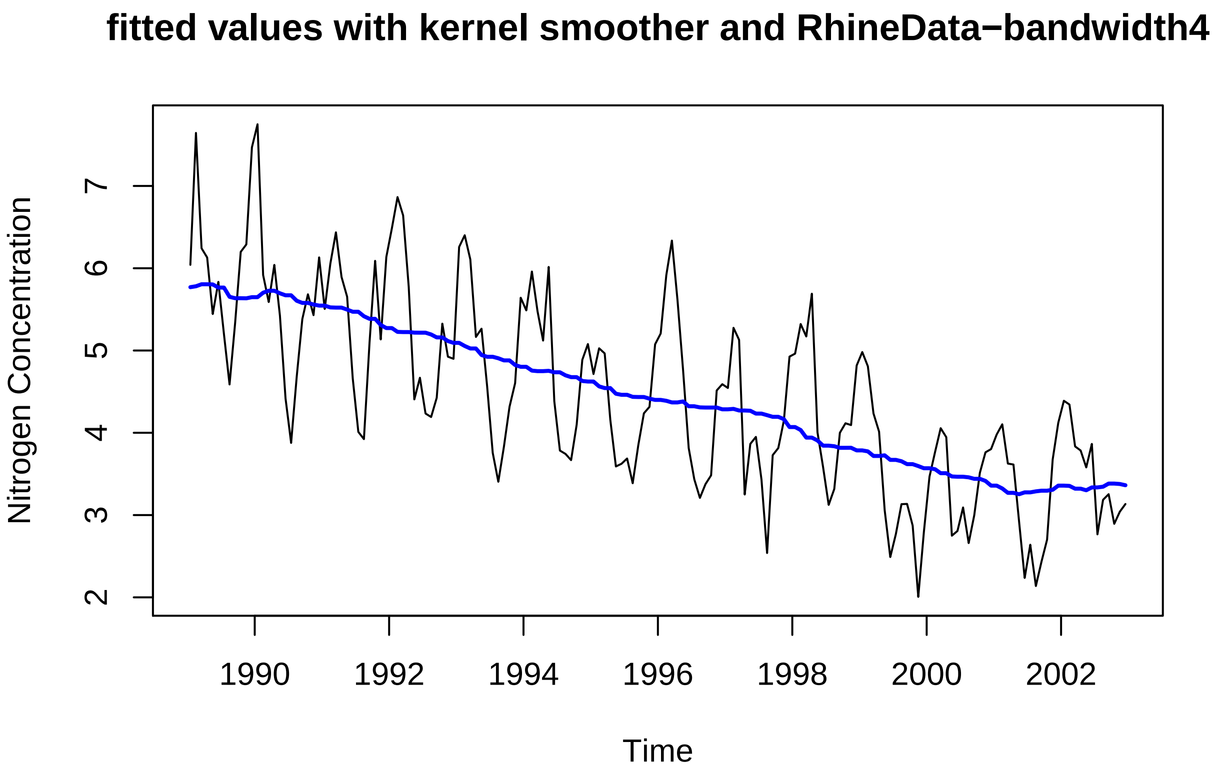

```
par(mfrow = c(1,1))
#residuals plot
plot(linear_model$residuals, main = "plot of Residual",
x = linear_model$model$Time, ylab = "Residuals",
xlab = "Model variable Time", type = "p")
# result list of linear modelcontain  variables are residuals
#,model& model contain variable time
abline(h = 0)
```



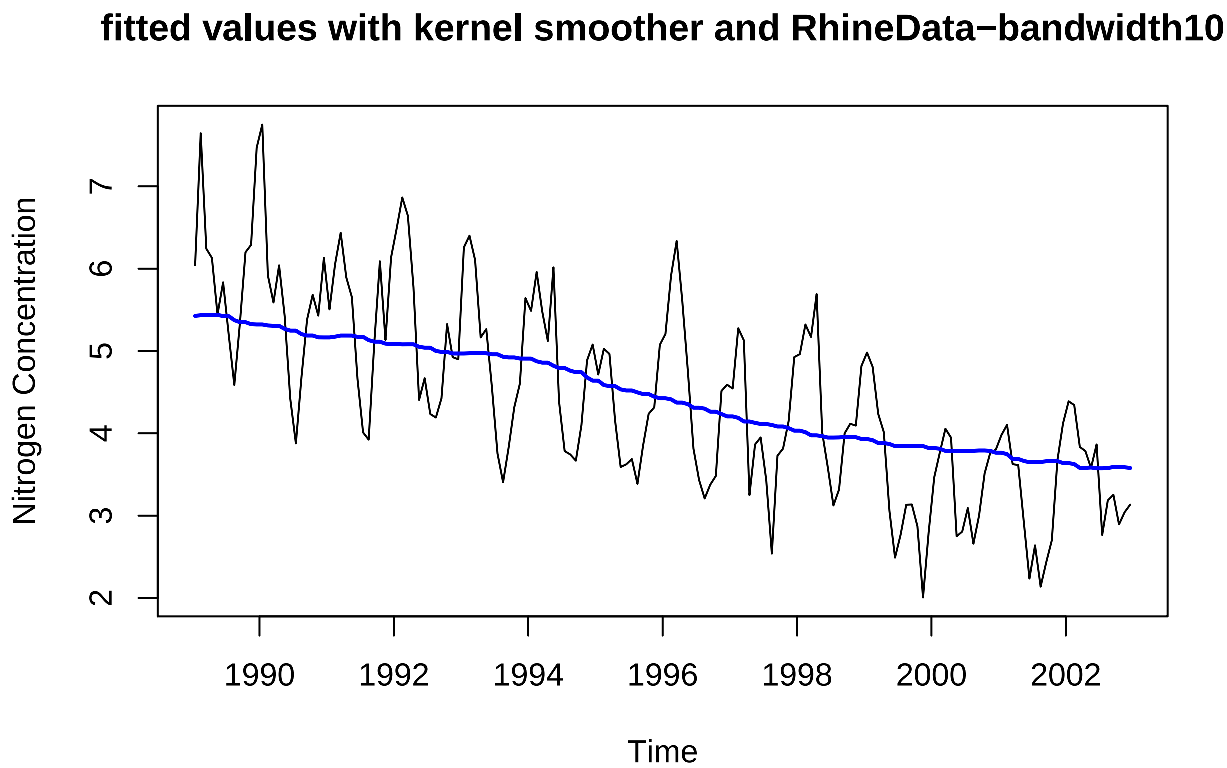
Analysis:-From the plot we can observe that there is a decreasing trend in Nitrogen Concentration over time.The ACF we can see that the autocorrelation has a reappering pattern every year which can be a sign of seasonalityand.

2c)

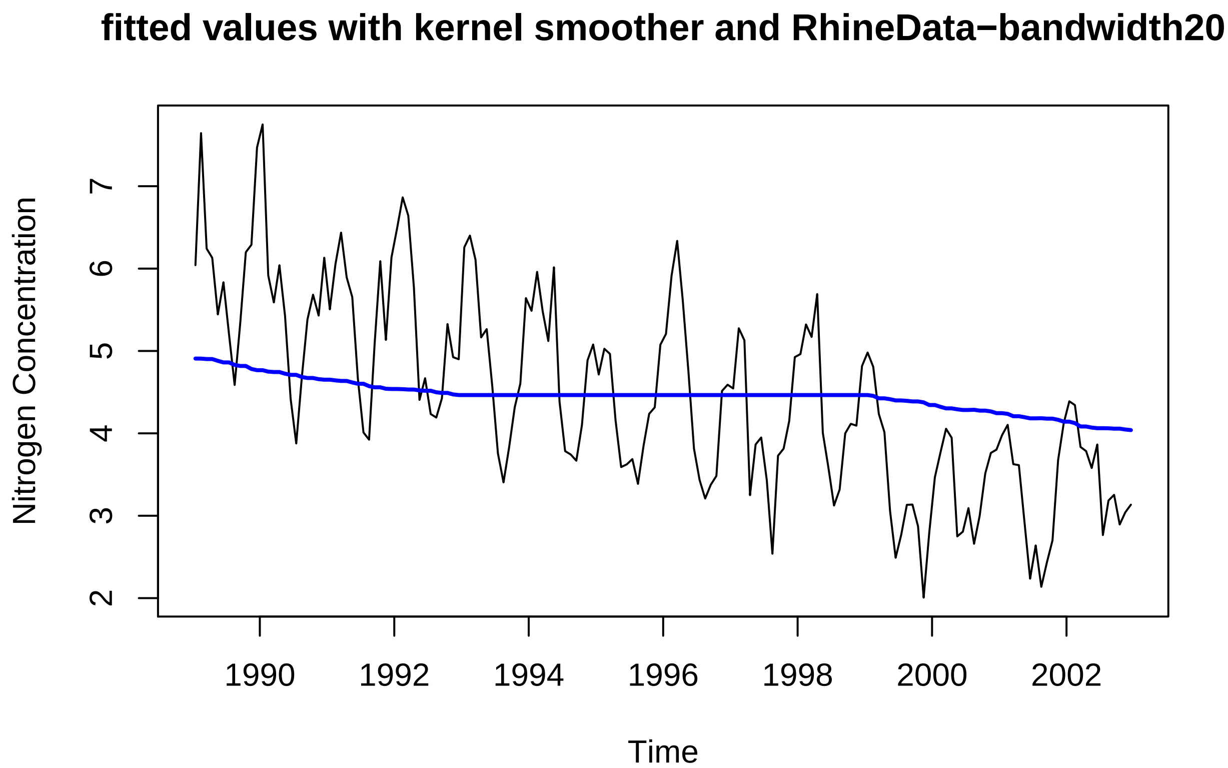
```
#Eliminating the trend by fitting a kernel smoother
kernel_model4 <- ksmooth(x = Rhine_data$Time, y =Rhine_data$TotN_conc, bandwidth=4)
kernel_model10 <- ksmooth(x = Rhine_data$Time, y =Rhine_data$TotN_conc, bandwidth=10)
kernel_model20 <- ksmooth(x = Rhine_data$Time, y =Rhine_data$TotN_conc, bandwidth=20)

residuals_kernel4 <- Rhine_data$TotN_conc - kernel_model4$y
residuals_kernel10 <- Rhine_data$TotN_conc - kernel_model10$y
residuals_kernel20 <- Rhine_data$TotN_conc - kernel_model20$y
#Plot
plot(x = Rhine_data$Time, y =Rhine_data$TotN_conc, type = "l",
main = "fitted values with kernel smoother and RhineData-bandwidth4",
ylab = "Nitrogen Concentration",
xlab = "Time")
lines(kernel_model4, lwd=2, col="blue")
```



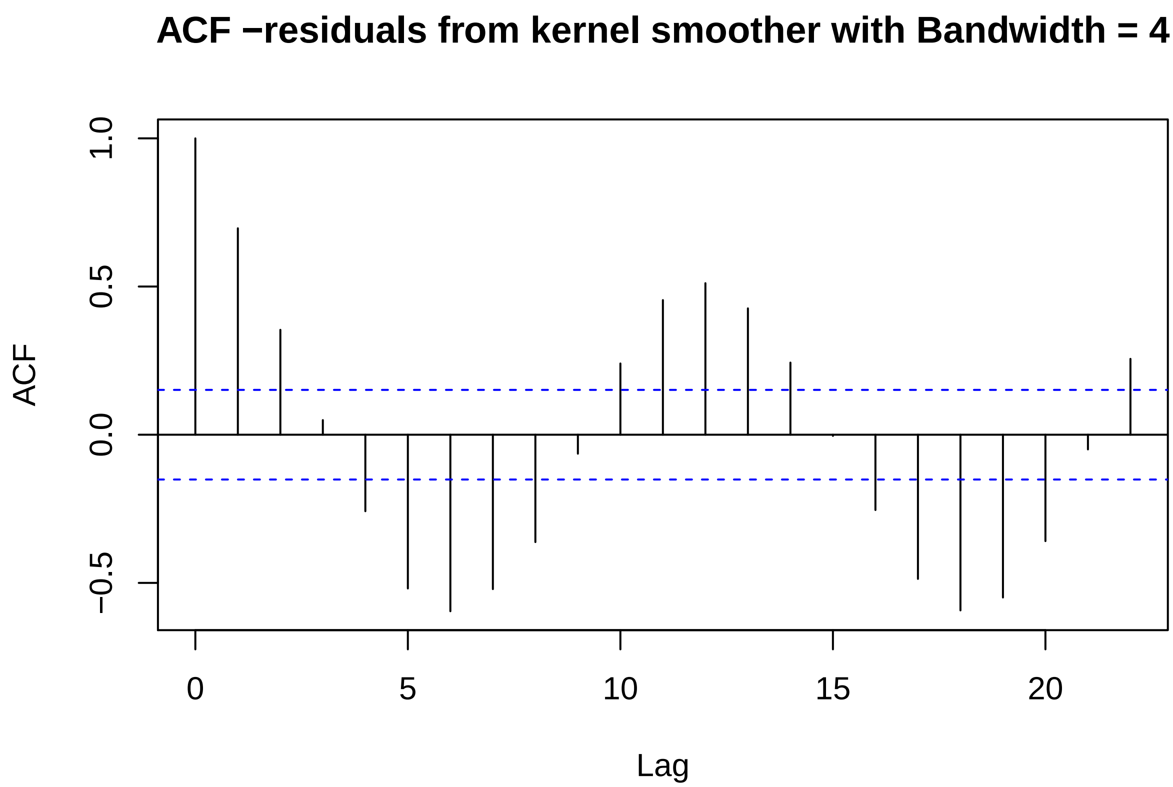

```r
plot(x = Rhine_data$Time, y =Rhine_data$TotN_conc, type = "l",
main = "fitted values with kernel smoother and RhineData-bandwidth10",
ylab = "Nitrogen Concentration",
xlab = "Time")
lines(kernel_model10, lwd=2, col="blue")
```



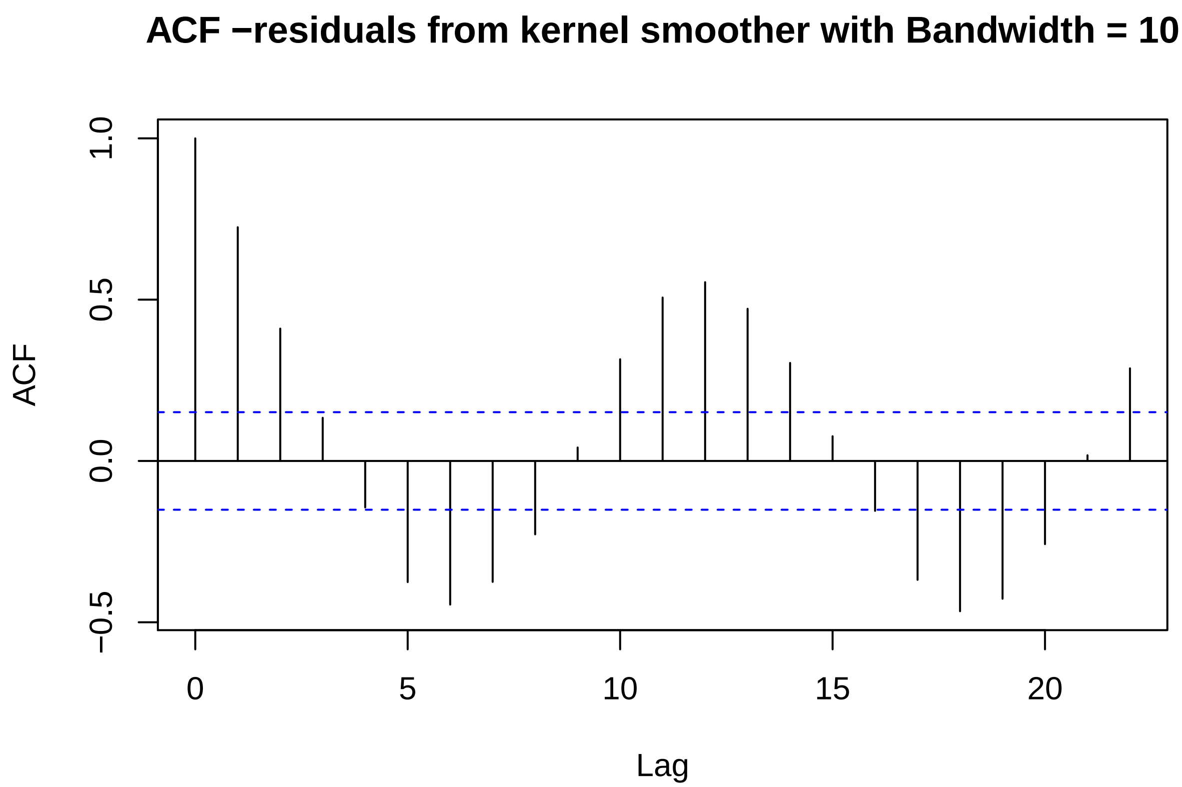

```
plot(x = Rhine_data$Time, y =Rhine_data$TotN_conc, type = "l",
main = "fitted values with kernel smoother and RhineData-bandwidth20",
ylab = "Nitrogen Concentration",
xlab = "Time")
lines(kernel_model20, lwd=2, col="blue")
```



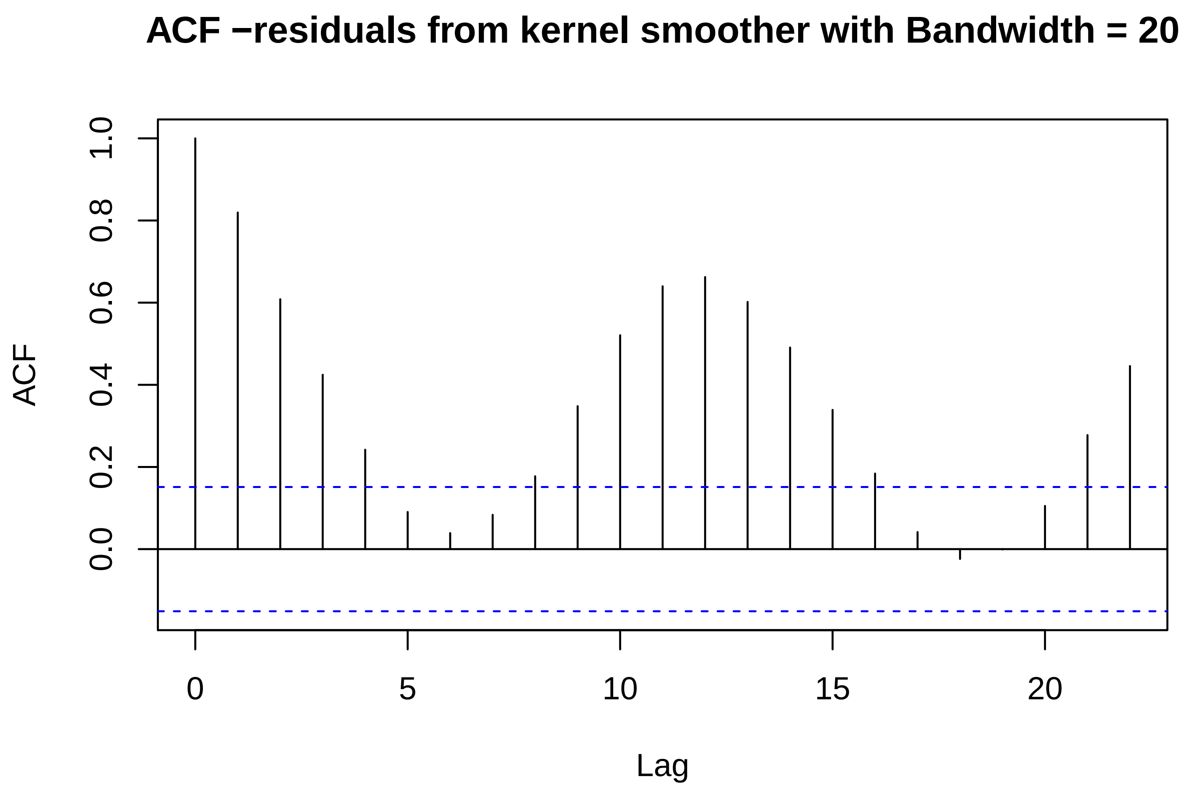

```r
#Plot ACF- residuals of kernel smoother
acf(residuals_kernel4, main = "ACF -residuals from kernel smoother with Bandwidth = 4")
```




```
acf(residuals_kernel10, main = "ACF -residuals from kernel smoother with Bandwidth = 10")
```



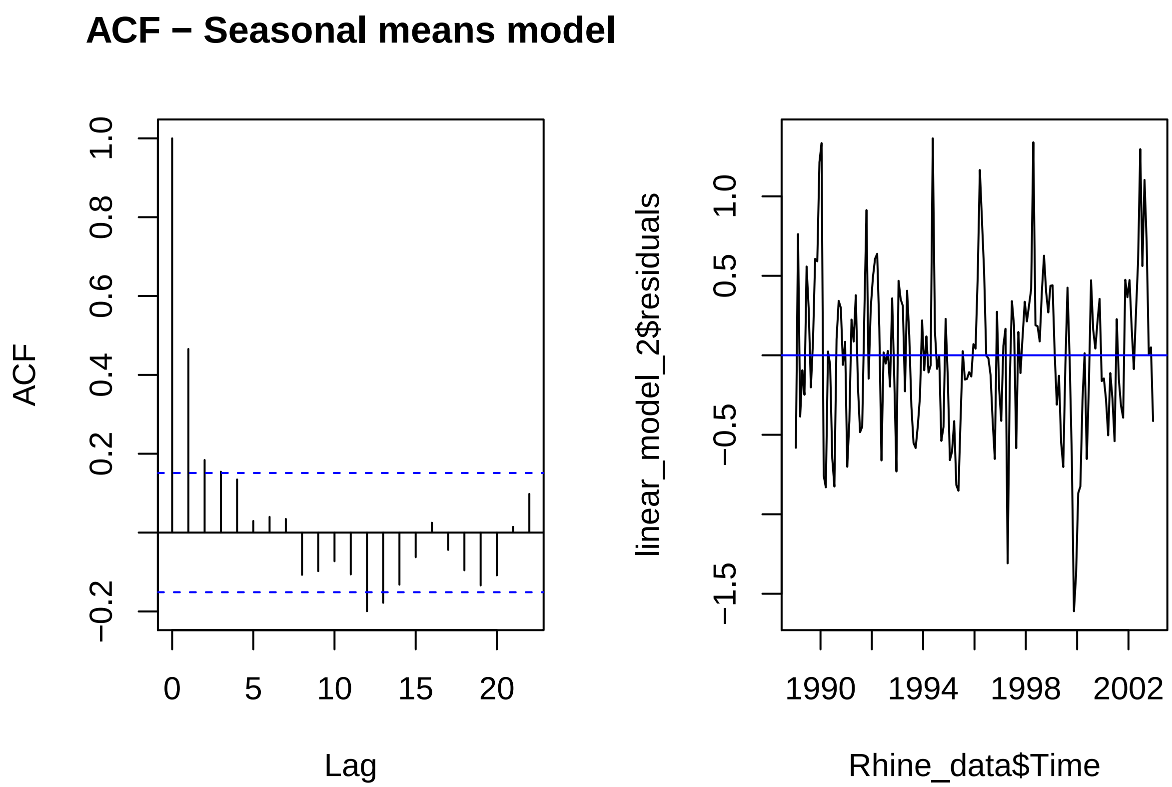

```
acf(residuals_kernel20, main = "ACF -residuals from kernel smoother with Bandwidth = 20")
```



Analysis:-The kernel smoother with bandwith 4,kernel smoother with bandwith 10 and kernel smoother with bandwith 20 . As we increase the bandwith the fitted line becomes underfitted and the residual pattern changes. The residuals does not seem to be stationary in the models since they seem to follow a pattern that is dependent on time this can be interpreted as a seasonal effect.

2d)

```
linear_model_2 <- lm(Rhine_data$TotN_conc ~ Rhine_data$Time + as.factor(Rhine_data$Month))
par(mfrow = c(1,2))
acf(linear_model_2$residuals,
main = "ACF - Seasonal means model")
plot(y = linear_model_2$residuals, x = Rhine_data$Time,
type = "l")
abline(h=0, col = "blue")
```




```r
par(mfrow = c(1,1))
```

Analysis:-The residual plot does not look dependent on time anymore and this looks more like noise and we could visually confirm that its stationary. The ACF plot we can see that the seasonality is disapearing.

2e)

```r
#library
library(MASS)
#AIC  model  stepwise selection
stepwise_aic <- step(linear_model_2, direction = "backward")
```

```
## Start:  AIC=-202.02
## Rhine_data$TotN_conc ~ Rhine_data$Time + as.factor(Rhine_data$Month)
##
##                               Df Sum of Sq     RSS      AIC
## <none>                                      43.237 -202.023
## - as.factor(Rhine_data$Month) 11    68.524 111.761  -64.477
## - Rhine_data$Time              1   118.387 161.624   17.499
```

```r
stepwiseAIC <- summary(stepwise_aic)
```

Analysis:- Since the variable month is a factor the stepwise won't evacuate any components even in spite of the fact that its not huge. We can find in the rundown that several variables are not critical but rather the best model given AIC is a similar model,we get a similar model as the best one since we just think about two factors, time and month to nitrogen.

Assignment 3. Analysis of oil and gas time series.

Weekly time series *oil* and *gas* present in the package *astsa* show the oil prices in dollars per barrel and gas prices in cents per dollar.

a) Plot the given time series in the same graph. Do they look like stationary series? Do the processes seem to be related to each other? Motivate your answer.

b) Apply log-transform to the time series and plot the transformed data. In what respect did this transformation made the data easier for the analysis?

c) To eliminate trend, compute the first difference of the transformed data, plot the detrended series, check their ACFs and analyze the obtained plots. Denote the data obtained here as x_t (oil) and y_t (gas).

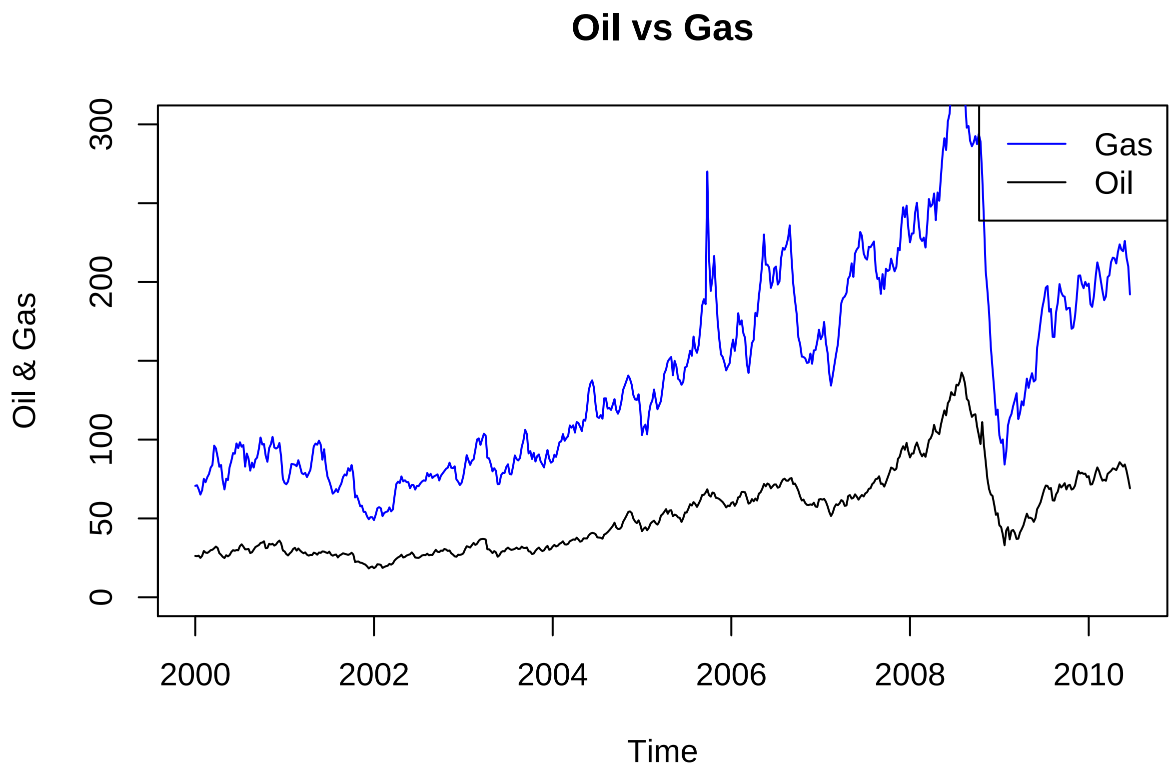
d) Exhibit scatterplots of x_t and y_t for up to three weeks of lead time of x_t; include a nonparametric smoother in each plot and comment the results: are there outliers? Are the relationships linear? Are there changes in the trend?

e) Fit the following model: $y_t = \alpha_0 + \alpha_1 I(x_t > 0) + \beta_1 x_t + \beta_2 x_{t-1} + w_t$ and check which coefficients seem to be significant. How can this be interpreted? Analyze the residual pattern and the ACF of the residuals.

3a)

Analysis:-A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time according to this defination ,No the plot does not look like stationary series,In the plot gas variance is dependent on time.The two series are increasing trend from the year 2008,So the processes seem to be related to each other

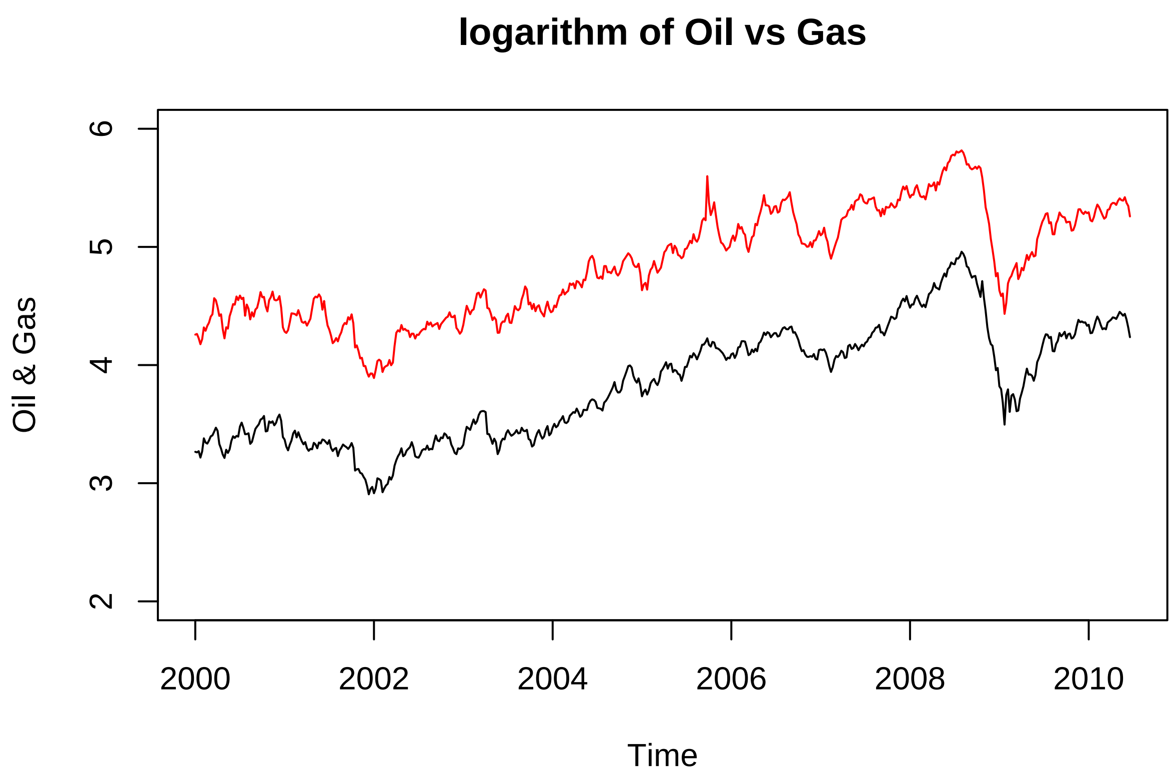
```
#Weekly time series oil and gas present in the package library astsa
library(astsa)
plot(oil, ylab = "Oil & Gas",xlab = "Time", main = "Oil vs Gas",
ylim = c(0,300))
lines(gas, col = "blue")
legend("topright", legend = c("Gas", "Oil"), lty = c(1,1), col = c("blue", "black"))
```



3b)

Analysis:-In time series analysis, log transformation is often considered to stabilize the variance of a series. Forecasts based on the original series are compared to forecasts based on logs.So it is easier to compare the two different time series when log stabilizes the variance.

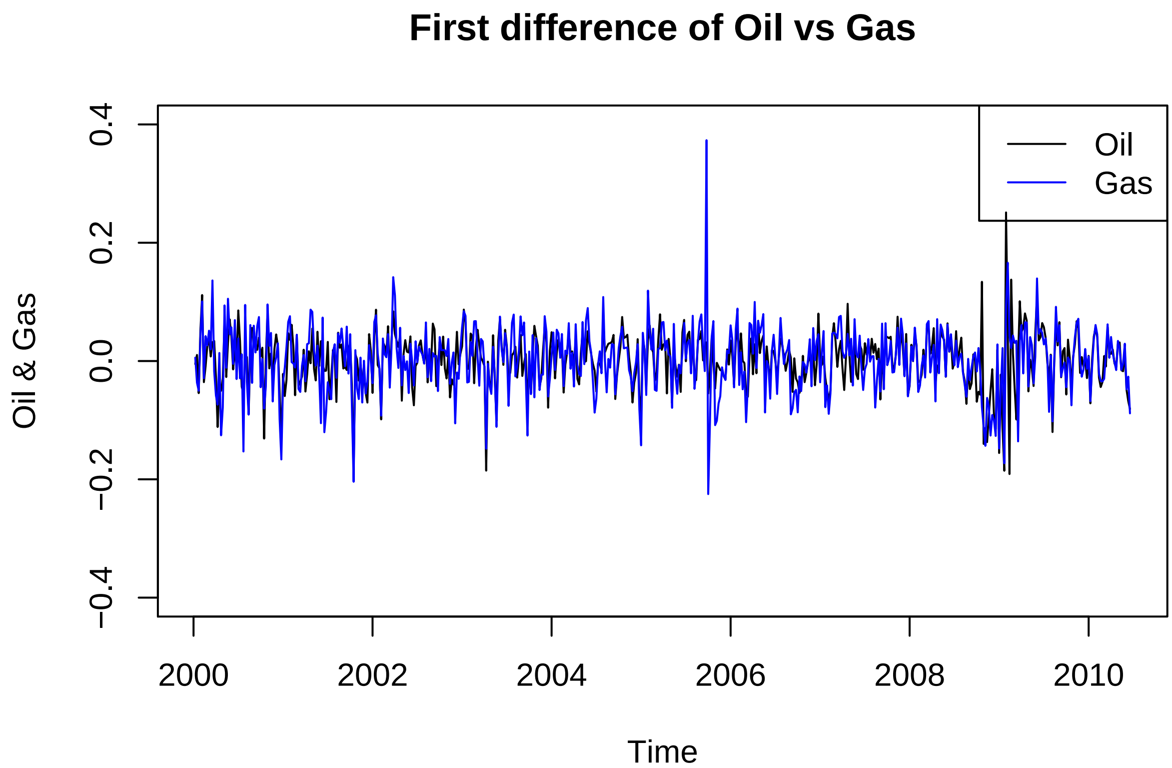
```r
#Apply log-transform to the time series
logOil <- log(oil)# log-transform
logGas <- log(gas)
plot(logOil, ylab = "Oil & Gas",
xlab = "Time", main = "logarithm of Oil vs Gas",
ylim = c(2,6))
lines(logGas, col = "red")
```



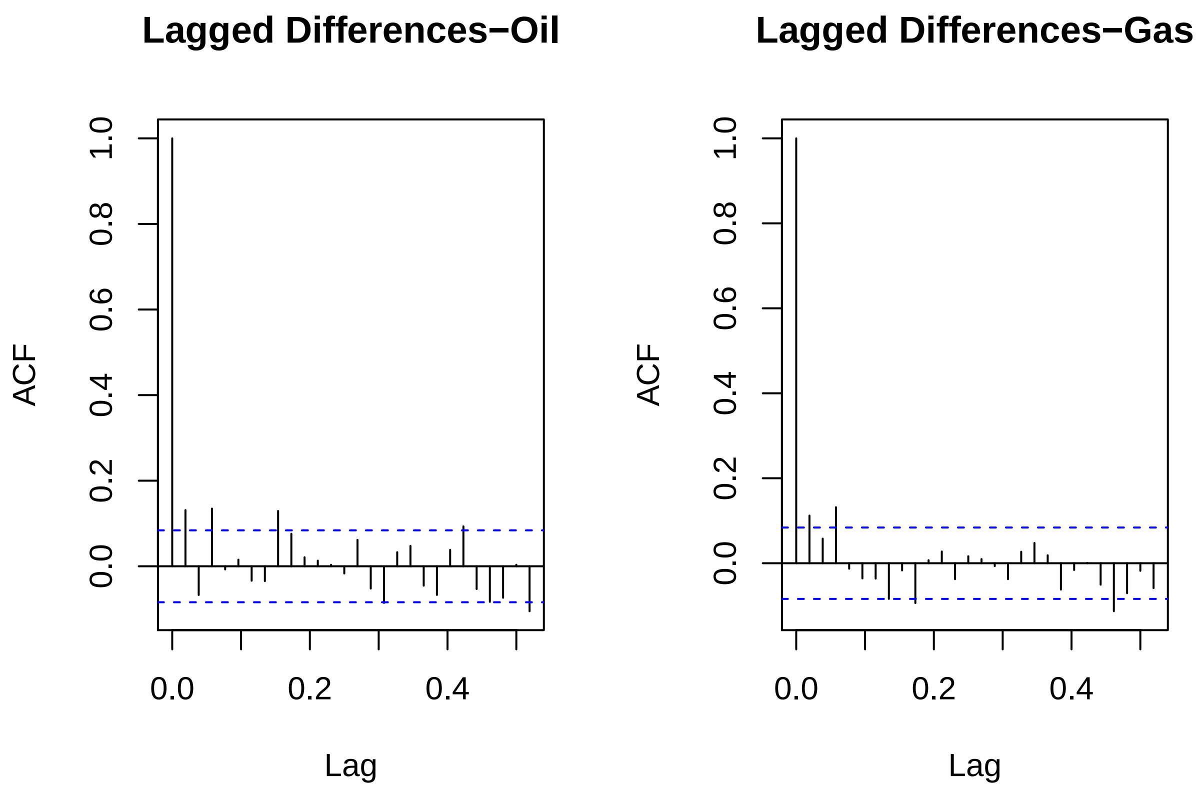
3c)

Analysis:- A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time,So according to the defination it seem to be stationary even though the plot contain some peaks.Not much autocorrelation we can see from the ACF plots.

```
oil_diff <- diff(logOil)#diff for Lagged Differences
gas_diff <- diff(logGas)
plot(oil_diff, ylab = "Oil & Gas",
xlab = "Time", main = "First difference of Oil vs Gas",
ylim = c(-0.4,0.4))
lines(gas_diff, col = "blue")
legend("topright", legend = c("Oil","Gas"), col = c("black", "blue"),
lty = c(1,1))
```




```r
#ACF of laged differences of oil and gas
par(mfrow = c(1,2))
acf(oil_diff, main = "Lagged Differences-Oil")
acf(gas_diff, main = "Lagged Differences-Gas")
```

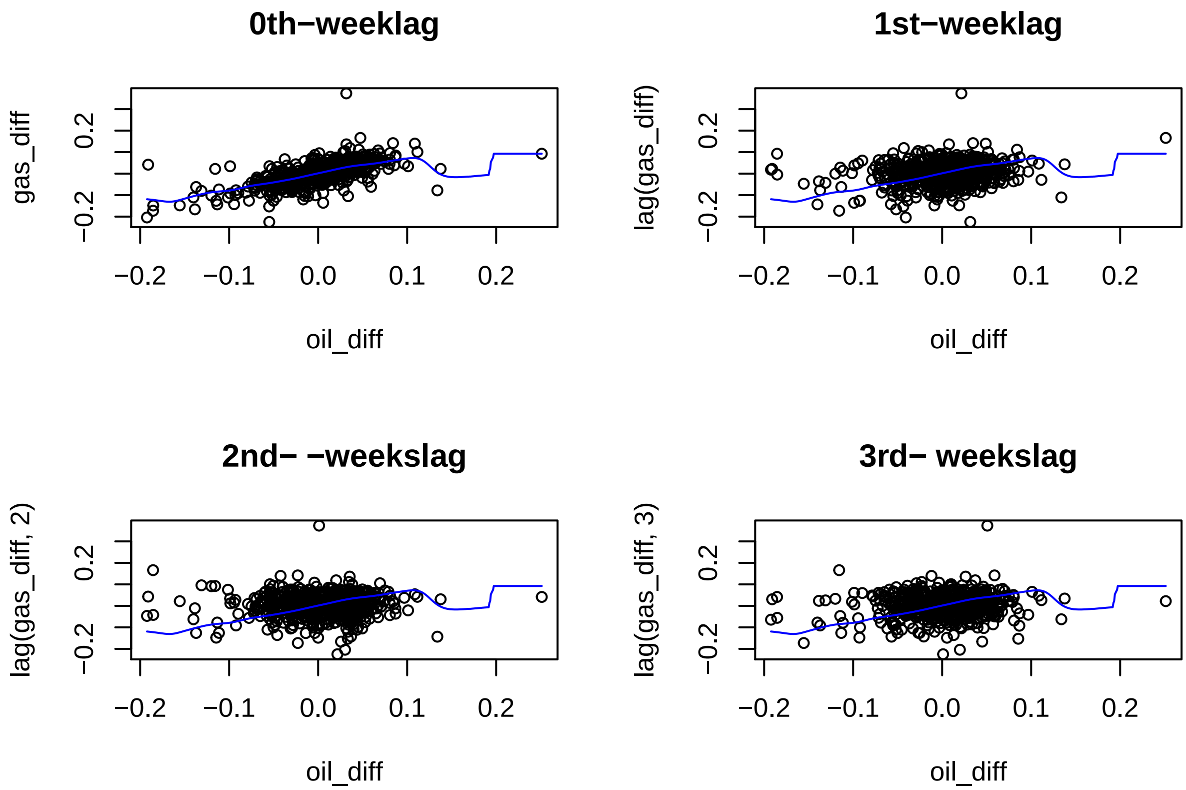



```
par(mfrow = c(1,1))
```

3d)

Analysis:-The relationships are linear,No changes in the trend,and we have a couple of outliers in all the plots.

```
par(mfrow = c(2,2))
plot(x = oil_diff , y = gas_diff, main = "0th-weeklag")
lines(ksmooth(x = oil_diff, y = gas_diff, bandwidth = 0.04, kernel = "normal"), col = "blue")
plot(x = oil_diff, y = lag(gas_diff), main = "1st-weeklag")
lines(ksmooth(x = oil_diff, y = lag(gas_diff), bandwidth = 0.04, kernel = "normal"), col = "blue")
plot(x = oil_diff, y = lag(gas_diff, 2), main = "2nd- -weekslag")
lines(ksmooth(x = oil_diff, y = lag(gas_diff,2), bandwidth = 0.04, kernel = "normal"), col = "blue")
plot(x =oil_diff, y = lag(gas_diff, 3), main = "3rd- weekslag")
lines(ksmooth(x = oil_diff, y = lag(gas_diff,3), bandwidth = 0.04, kernel = "normal") ,col = "blue")
```



3e)

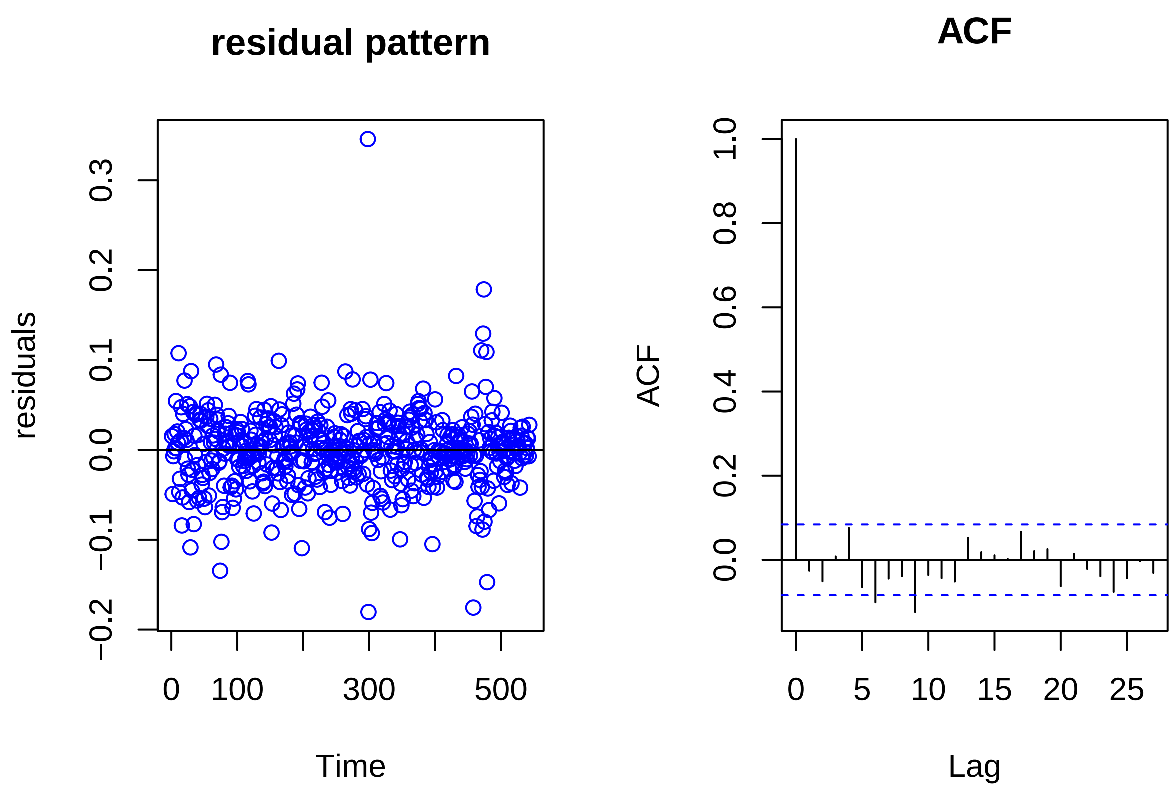
Analysis:-According to the summary the binary variable I(xt > 0) seem to be significant , and the lag variable does not seem to be significant.The residuals plot seem to be constant over time with some outliers.

```
intersect_ts <- ts.intersect(y = gas_diff, xt = oil_diff,xt1 = lag(oil_diff,1),xtbin = oil_diff>0)
model_lm <- lm(y ~ xt + xt1 + xtbin, data = intersect_ts )
summary <- summary(model_lm)
summary
```

```
##
## Call:
## lm(formula = y ~ xt + xt1 + xtbin, data = intersect_ts)
##
## 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 ***
## xt1          0.012660   0.038729   0.327   0.7439
## xtbin        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
```

```r
par(mfrow = c(1,2))
plot(residuals(model_lm), ylab = "residuals",xlab = "Time", main = "residual pattern",col = "blue")
abline(h = 0)
#acf plot
acf(residuals(model_lm), main = "ACF")
```




```r
par(mfrow = c(1,1))
```