Time Series

Computer Lab A

Andreas Charitos(andch552),Ruben Muñoz (rubmu773)

2019-09-16

Contents

1 Computations with simulate data

1.1 A)

Generate two time series $x_t = -0.8x_{t-2} + w_t$ where $x_0 = x_1 = 0$ and $x_t = cos(\frac{2\pi t}{5})$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.

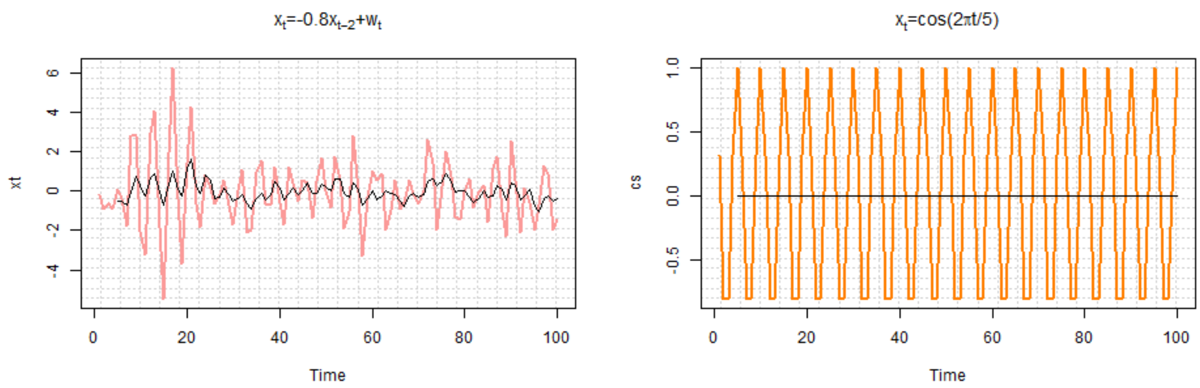

Figure 1: *Comparation of ACF plots.*

1.2 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 appropiate R code to investigate whether this time series is casual and invertible.

```
## [1] "Not casual or invertible "
```

1.3 C)

Use built-in R funcitons 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.

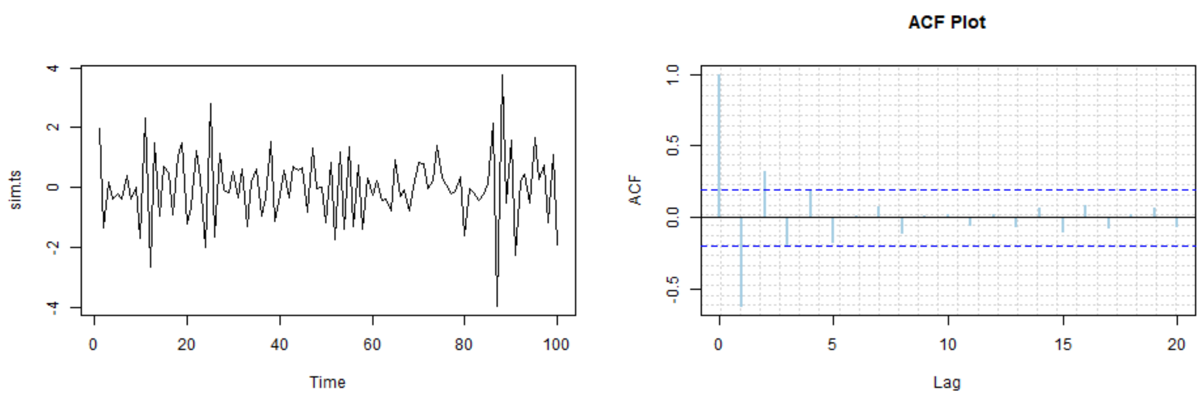

Figure 2: *Comparation of ACF plots.*

2 Visualization, detrending and residuals analysis of Rhine data

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

2.1 A)

Import the data to R, convert it appropriately to *ts* object (use function `ts()`) and explore it by plotting the time series, creating scatterplots of x_t against $x_{t-1}, ...x_{t-12}$. Analyze the time series plot and scatter plots: Are there any trends, linear or seasonal, in the time series? When suring the year is the concentration highest? Are there any special patters 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?

Figure 3 shows the ploted data as a time series giving a sence of having a descending trend and also s possible sence of peiodicity. As instructed in this lab, by using the scatterplots ith lag, we can see in the begining as lag 1 that there is a significant level of correlation.

This in itself, is a signal that indeed there is a semblance of seasonality to the data with that lag. As e move long the Scatterplot in fugure 4 both left and right, it shows that ahen compared ith lag 5 or 6 the correlation is almost non existant, but then again, proving its periodicity as e move further in the lag upt to 12 then again e can see a high correlation of the data with itself, thus proving its seasonality.

Is also worth to mention that this seasonality seems to appear yearly, given the lag being 1 and 12 as well as the meaning of the unit being 1 month.

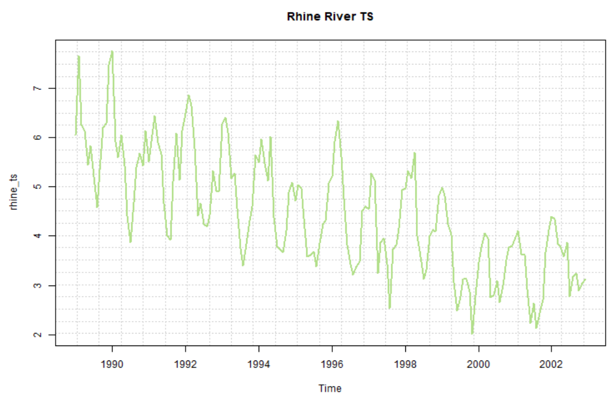

Figure 3: Time series plot

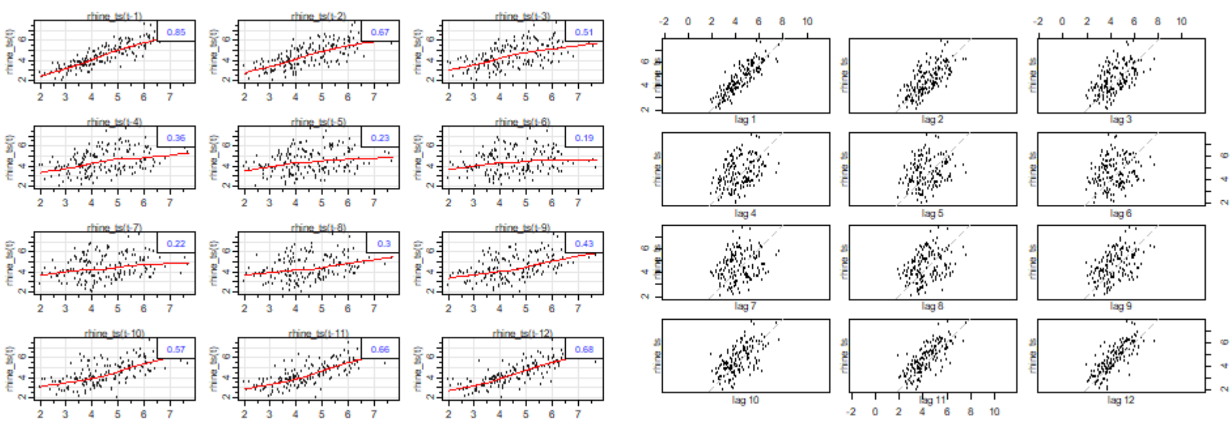

Figure 4: *Ploted data as a time series plot with lag1() left and lag() right.*

2.2 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 AFC of the residuals and comment on this pattern might be related to the seasonality series.

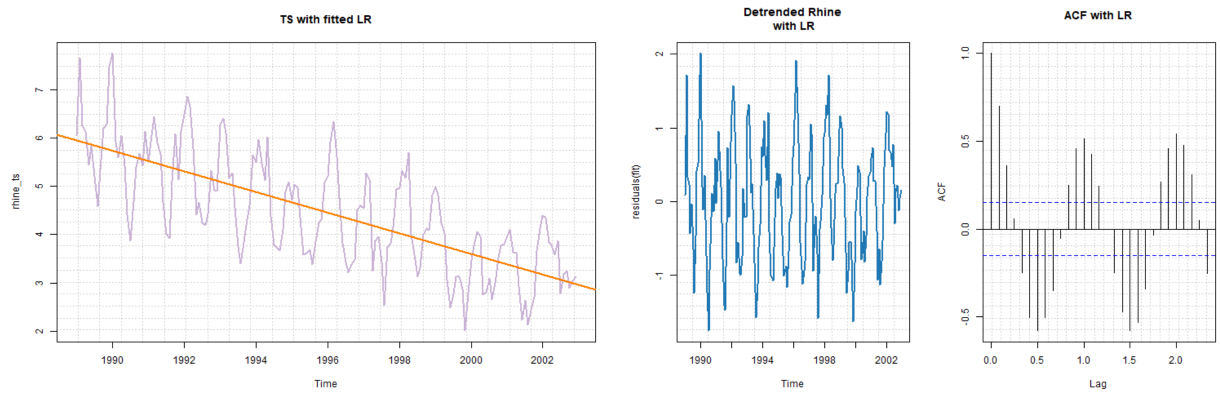

Figure 5: *Analisys of the TS with a fitted LR at the left and its residuals analisys..*

Figure 5, first plot, shows the TS fitted with a Linear Regression, this is showing a decreasing trend. In the other half of the figure, e can see the residual pattern showing an interesting um and donwn movement, or simmilar to seasonal, remanising of what the lag scatterplots showed.

Finally in the ACF of the residuals it's showing what could be reffered to a beautiful seadonality, thus also related to the lag scatterplots from before.

2.3 C)

Eliminate the trend by fitting a kernel smooter 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?

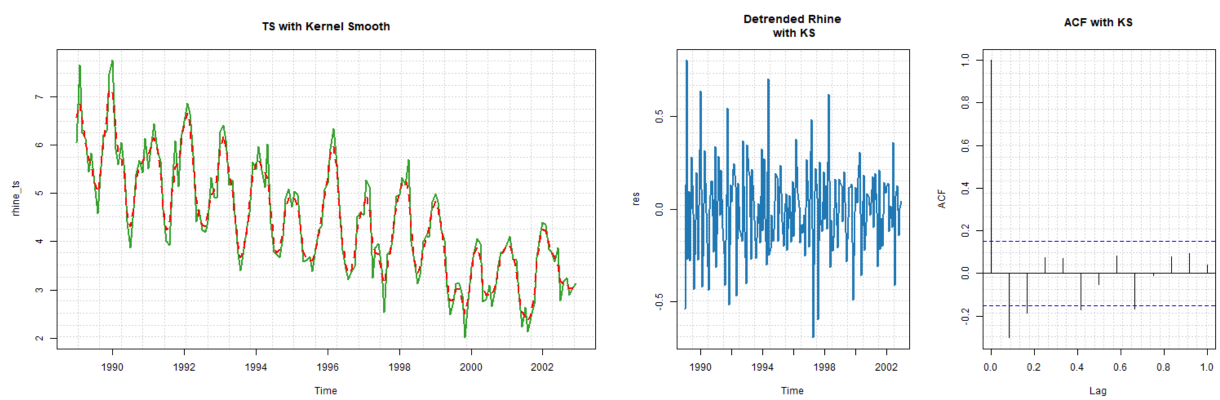

Figure 6: *Analisys of the TS with a smoothing kernel at the left and its residuals analisys.*

Figure 6 has both the visual result of what the Kernel smooth does to the TS data. The smoother in a way, tries to reduce w, white noise, thus its expected that the seasonality before observed in the data with a linear regression will also be less present. The ressiduals do seem to represent the behavior closer to a stationary series.

2.4 D)

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

$$x_t = \alpha_0 + \alpha_1 t + \beta_1 I(month = 2) + ... + \beta_1 2I(month = 12) + w_t$$

where $I(x) = 1$ if x is true and 0 otherwise. Fitting of this model will require you to augument data with a categorical variable shoing the current month, and then fitting a usual linear regression. Analyze the residual plattern and the ACT of residuals.

Following this method, in figure 7,we can see that in the detrended plot with seasonal mean the seasonal behavior is little bit more vissible but still not quite there. Although the ACF does show a more understandable seasonality with the lag unit being a montth.

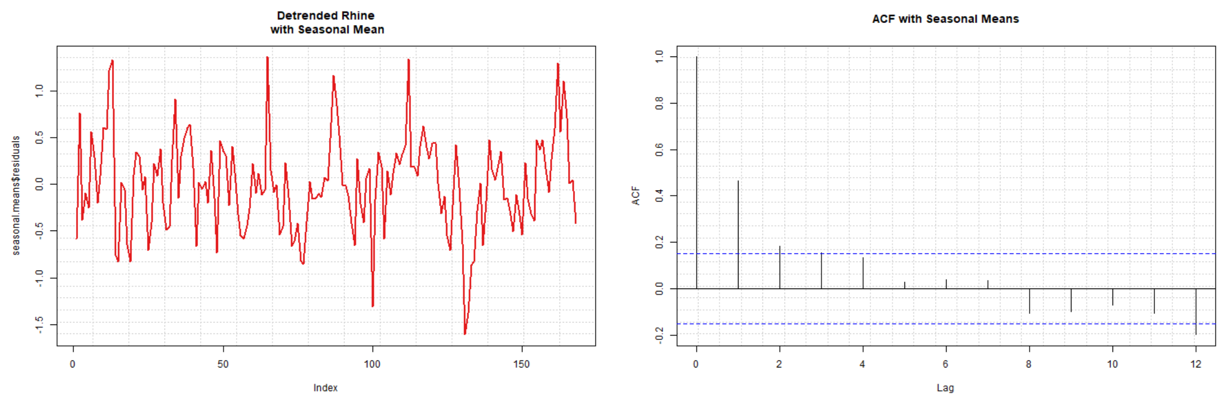

Figure 7: *Time series after trend elimination ith a seasonal means sodel.*

2.5 E)

Perform stepwise variable selection in model from step d). Which model gives you the lower AIC value? Which variables are left in the model?

[1] -202.0227

Start: AIC=-202.02 ## TotN_conc ~ month_enc + Time ## ## Df Sum of Sq RSS AIC ## <none> 43.237 -202.023 ## - month_enc 11 68.524 111.761 -64.477 ## - Time 1 118.387 161.624 17.499

[1] -202.0227

Accoding to the shown results both seem to perform equally well with the same respective AIC score.

3 Analisys 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.

3.1 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.

Figure 8 (left plot) shows what we would describe visually as something that is not a stationary series.

3.2 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?

Figure 8 (right) does shows that the applied transformation made the data easier at least for visual analisys. This is thanks to the 2d reduction that a log transformation brings to greater numbers hen compared to relative small ones, allowing a clearer comparison loosing less detail of the smaller movements in the TS.

Figure 8: *Time series before (to the left) and after (to the right) the agumentation of a log-transformation.*

3.3 C)

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

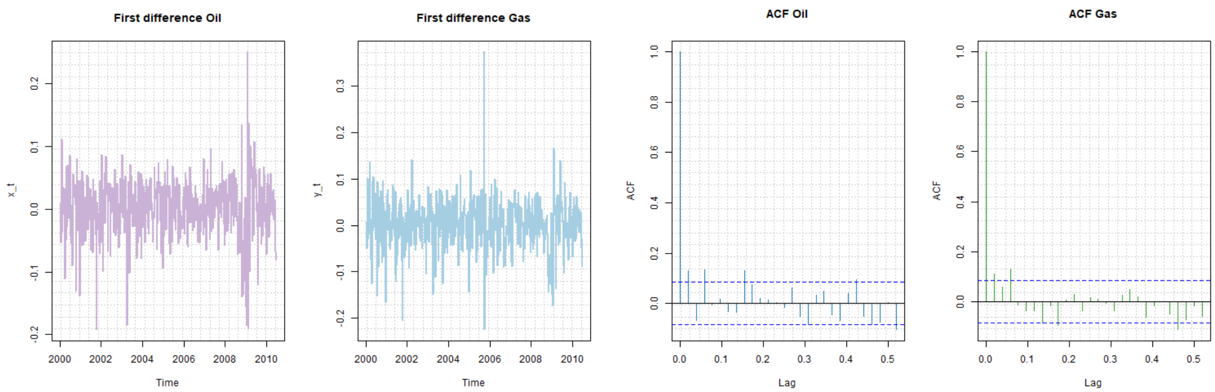

Figure 9: *Time series analysis of Oil and Gas First difference and ACF respectively.*

3.4 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 putliers? Are the relationships linear? Are there changes in the trend?

Figure 10 does show an slight decrease of the linear relationship as the lead increases.

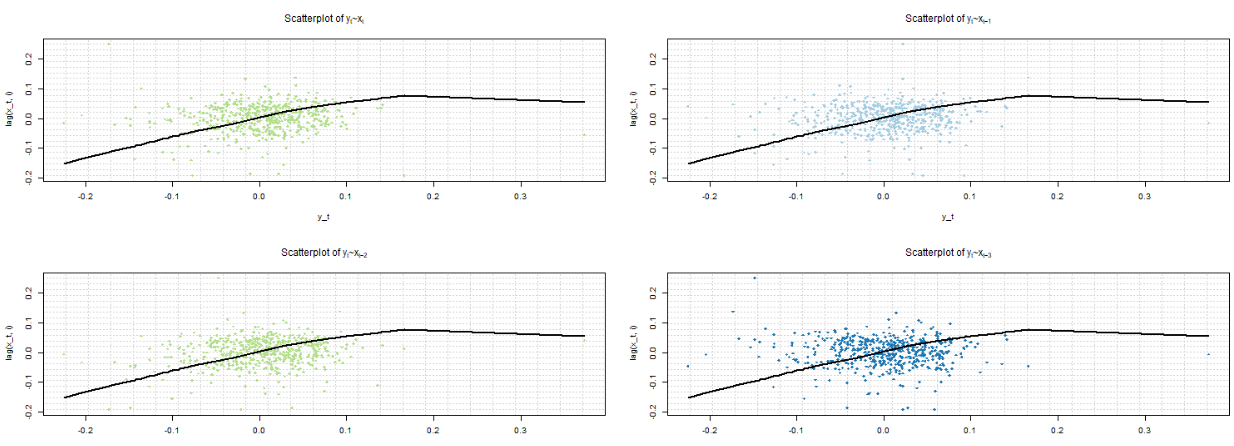

Figure 10: *Time series analysis scatterplot with 0-3 weeks of lead time.*

3.5 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 interpretted? Analyze the residual pattern and the ACF of the residuals.

```
## (Intercept) time(rhine_ts) ## TRUE TRUE
```

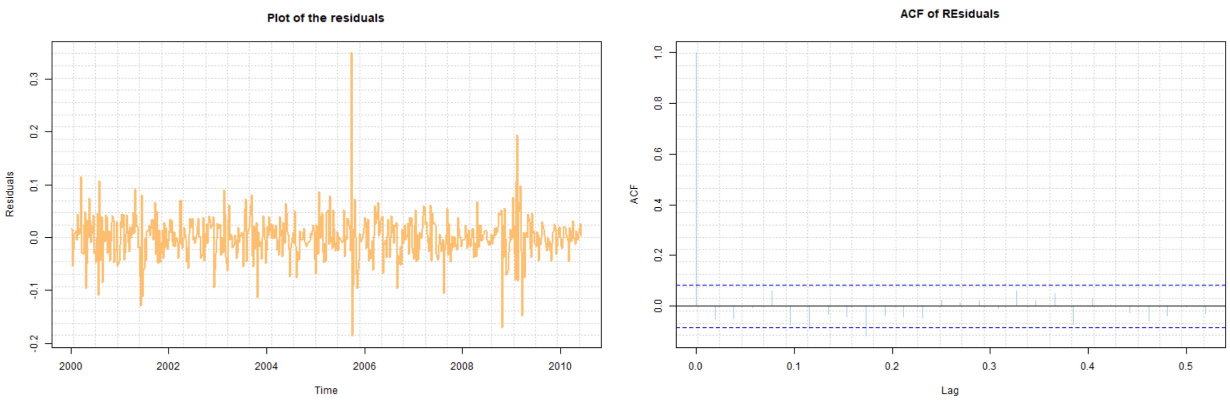

Figure 11: *Time series analysis scatterplot with 0-3 weeks of lead time.*

4 Appendix

4.1 Code

4.1.1 Code used for Computations with simulate data

```
library(ggplot2)
library(ggfortify)
# autoplot(USAccDeaths)
library(forecast)
set.seed(54321)
require(RColorBrewer)
pal <- brewer.pal(9, "Paired")
xt = arima.sim(list(order = c(2, 0, 0), ar = c(0, -0.8)),
    n = 100, start.innov = c(0, 0), n.start = 2)
t <- 1:100
cs = cos((2 * pi * t)/5)
# apply the filter
f_coefs = rep(0.2, 5)
xt_filtered = filter(xt, filter = f_coefs, sides = 1)
cs_filtered = filter(cs, filter = f_coefs, sides = 1)
col1 <- sample(pal, 1)
col2 <- sample(pal, 1)
check_causality <- function(z) {
    return(sqrt(Re(z)^2 + Im(z)^2))
}
inv_caus_func <- function(AR_operator, MA_operator) {
    n1 <- length(AR_operator) - 1
    n2 <- length(MA_operator) - 1
    res <- polyroot(AR_operator)
    res2 <- polyroot(MA_operator)
    casuality <- sapply(res, function(y) {
        check_causality(y)
    })
    # print(casuality>1)
    invert <- sapply(res2, function(y) {
        check_causality(y)
    })
    # print(invert>1) print(sum(casuality>1))
    # print(sum(invert>1))
    if ((sum(casuality > 1) == n1) & (sum(invert > 1) ==
        n2)) {
        print("The series is invertible and casual")
    } else if (sum(casuality > 1) == n1) {
        print("The series is casal only!")
    } else if (sum(invert > 1) == n2) {
        print("The series is invertible only !")
    } else {
        print("Not casual or invertible ")
    }
}
ar_operator = c(1, -4, 2, 0, 0, 1)
ma_operator = c(1, 0, 3, 0, 1, 0, -1)
inv_caus_func(ar_operator, ma_operator)
```

4.1.2 Code used for Visualization, detrending and residuals analysis of Rhine data

```r
library(astsa)
rhine <- read.csv2("data/Rhine.csv", sep = ";")
rhine_ts <- ts(rhine[, 4], start = c(rhine[1, 1], 1), end = c(2002,
    12), frequency = 12)
fit <- lm(rhine_ts ~ time(rhine_ts), na.action = NULL)
# summary(fit)
col3 <- sample(pal, 1)
col4 <- sample(pal, 1)
plot.ts(rhine_ts, col = col3, main = "TS with fitted LR",
    panel.first = grid(25, 25), lwd = 2)
abline(fit, col = col4, lwd = 2)
legend(130, 7, legend = c("TS Rhine", "LR"), col = c(col3,
    col4), lty = 1, cex = 0.8, text.font = 4, bg = "white",
    lwd = 2)
par(mfrow = c(1, 2))
plot(residuals(fit), main = "Detrended Rhine\n with LR",
    col = sample(pal, 1), panel.first = grid(15, 25), lwd = 2)
acf(residuals(fit), 28, main = "ACF with LR", panel.first = grid(25,
    25))
kernelSmooth <- ksmooth(time(rhine_ts), rhine_ts, "normal",
    bandwidth = 0.2)
col5 <- sample(pal, 1)
plot.ts(rhine_ts, col = col5, main = "TS with Kernel Smooth",
    panel.first = grid(25, 25), lwd = 2)
lines(kernelSmooth, col = "red", lwd = 2, lty = 2)
legend(130, 7, legend = c("TS Rhine", "KS"), col = c(col5,
    "red"), lty = 1, cex = 0.8, text.font = 4, bg = "white",
    lwd = 2)
res <- (rhine_ts - kernelSmooth$y)
par(mfrow = c(1, 2))
plot(res, main = "Detrended Rhine\n with KS", col = sample(pal,
    1), panel.first = grid(15, 25), lwd = 2)
acf(res, 12, main = "ACF with KS", panel.first = grid(25,
    25))
new_rhine <- cbind(rhine, month_enc = c("January", "February",
    "March", "April", "May", "June", "July", "August", "September",
    "October", "November", "December"))
new_rhine$month_enc <- as.factor(new_rhine$month_enc)
str(new_rhine)
seasonal.means <- lm(TotN_conc ~ month_enc + Time, data = new_rhine)
# plot(seasonal.means$)
par(mfrow = c(1, 2))
plot(seasonal.means$residuals, main = "Detrended Rhine\n with Seasonal Mean",
    col = sample(pal, 1), panel.first = grid(15, 25), lwd = 2,
    type = "l")
acf(seasonal.means$residuals, 12, main = "ACF with Seasonal Means",
    panel.first = grid(25, 25))
temp <- step(seasonal.means, direction = "both", trace = 0,
    steps = 1)
temp$anova$AIC
summary(temp)
library(MASS)
```

```
temp1 <- stepAIC(seasonal.means, direction = "both")
temp1$effects
```

4.1.3 Code used for Analisys of oil and gas time series

```
library(astsa)
rhine <- read.csv2("data/Rhine.csv", sep = ";")
rhine_ts <- ts(rhine[, 4], start = c(rhine[1, 1], 1), end = c(2002,
    12), frequency = 12)
fit <- lm(rhine_ts ~ time(rhine_ts), na.action = NULL)
# summary(fit)
col3 <- sample(pal, 1)
col4 <- sample(pal, 1)
plot.ts(rhine_ts, col = col3, main = "TS with fitted LR",
    panel.first = grid(25, 25), lwd = 2)
abline(fit, col = col4, lwd = 2)
legend(130, 7, legend = c("TS Rhine", "LR"), col = c(col3,
    col4), lty = 1, cex = 0.8, text.font = 4, bg = "white",
    lwd = 2)
par(mfrow = c(1, 2))
plot(residuals(fit), main = "Detrended Rhine\n with LR",
    col = sample(pal, 1), panel.first = grid(15, 25), lwd = 2)
acf(residuals(fit), 28, main = "ACF with LR", panel.first = grid(25,
    25))
kernelSmooth <- ksmooth(time(rhine_ts), rhine_ts, "normal",
    bandwidth = 0.2)
col5 <- sample(pal, 1)
plot.ts(rhine_ts, col = col5, main = "TS with Kernel Smooth",
    panel.first = grid(25, 25), lwd = 2)
lines(kernelSmooth, col = "red", lwd = 2, lty = 2)
legend(130, 7, legend = c("TS Rhine", "KS"), col = c(col5,
    "red"), lty = 1, cex = 0.8, text.font = 4, bg = "white",
    lwd = 2)
res <- (rhine_ts - kernelSmooth$y)
par(mfrow = c(1, 2))
plot(res, main = "Detrended Rhine\n with KS", col = sample(pal,
    1), panel.first = grid(15, 25), lwd = 2)
acf(res, 12, main = "ACF with KS", panel.first = grid(25,
    25))
new_rhine <- cbind(rhine, month_enc = c("January", "February",
    "March", "April", "May", "June", "July", "August", "September",
    "October", "November", "December"))
new_rhine$month_enc <- as.factor(new_rhine$month_enc)
str(new_rhine)
seasonal.means <- lm(TotN_conc ~ month_enc + Time, data = new_rhine)
# plot(seasonal.means$)
par(mfrow = c(1, 2))
plot(seasonal.means$residuals, main = "Detrended Rhine\n with Seasonal Mean",
    col = sample(pal, 1), panel.first = grid(15, 25), lwd = 2,
    type = "l")
acf(seasonal.means$residuals, 12, main = "ACF with Seasonal Means",
    panel.first = grid(25, 25))
temp <- step(seasonal.means, direction = "both", trace = 0,
```

```
    steps = 1)
temp$anova$AIC
summary(temp)
library(MASS)
temp1 <- stepAIC(seasonal.means, direction = "both")
temp1$effects
```