Analysis of Time Series 1MS014
Assignment 2

Viktor Ågren

May 2022

1 Problem 1

Let $x_t = w_t + 0.4w_{t-1}$, $y_t = 0.3x_{t-1} + v_t$ where w_t and v_t are independent white noise processes with variances $\sigma_w^2 = 2, \sigma_v^2 = 4$.

(a) Calculate the cross covariance function $\gamma_{yx}(h) = \mathrm{Cov}(y_{t+h}, x_t)$ for all integer h.

(b) Calculate the cross spectral density $f_{yx}(\omega)$.

(c) Calculate the squared coherence $\rho^2_{y\cdot x}(\omega)$ for $\omega = 0.2$.

Solution (a) Per definition the cross-covariance function between two time series y_t and x_t is

$$\gamma_{yx}(h) = \mathrm{Cov}(y_{t+h}, x_t) = \mathrm{E}[(y_{t+h} - \mu_y)(x_t - \mu_y)].$$

It is easy to see that both x_t and y_t have zero mean, i.e.

$$\mu_x = \mathrm{E}[x_t] = \mathrm{E}[w_t + 0.4w_{t-1}] = \mathrm{E}[w_t] + 0.4\mathrm{E}[w_{t-1}] = 0,$$

and

$$\begin{aligned}\mu_y = \mathrm{E}[y_t] &= \mathrm{E}[0.3x_{t-1} + v_t] = \mathrm{E}[0.3w_{t-1} + 0.12w_{t-2} + v_t] \\ &= 0.3\mathrm{E}[w_{t-1}] + 0.12E[w_{t-2}] + \mathrm{E}[v_t] = 0.\end{aligned}$$

Hence, the cross-covariance function is

$$\begin{aligned}\gamma_{yx}(h) &= \mathrm{E}[y_{t+h}x_t] = E[(0.3w_{t+h-1} + 0.12w_{t+h-2} + v_{t+h})(w_t + 0.4w_{t-1})] \\ &= 0.3\mathrm{E}[w_{t+h-1}w_t] + 0.12\mathrm{E}[w_{t+h-1}w_{t-1}] + 0.12\mathrm{E}[w_{t+h-2}w_t] + 0.048\mathrm{E}[w_{t+h-2}w_{t-1}] \\ &\quad + \mathrm{E}[v_{t+h}w_t] + 0.4\mathrm{E}[v_{t+h}w_{t-1}] \\ &= 0.3\gamma_w(h-1) + 0.12\gamma_w(h) + 0.12\gamma_w(h-2) + 0.048\gamma_w(h-1) \\ &= 0.12\gamma_w(h) + 0.348\gamma_w(h-1) + 0.12\gamma_w(h-2).\end{aligned}$$

Thus the cross-covariance function for all integers h is

$$\gamma_{yx}(h) \begin{cases} 0.12\sigma_w^2 & \text{if } h = 0, \\ 0.348\sigma_w^2 & \text{if } h = 1, \\ 0.12\sigma_w^2 & \text{if } h = 2, \\ 0 & \text{otherwise.} \end{cases}$$

The exact solution to the cross-covariance function is

$$\gamma_{yx}(h) \begin{cases} 0.24 & \text{if } h = 0, \\ 0.696 & \text{if } h = 1, \\ 0.24 & \text{if } h = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Solution (b) Per definition we solve the cross-spectral density as

$$\begin{aligned} f_{yx}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{yx}(h) e^{-2\pi i\omega h} \\ &= 0.24 + 0.696 e^{-2\pi i\omega} + 0.24 e^{-4\pi i\omega} \end{aligned}$$

Solution (c) The squared coherence formula is

$$\rho^2_{y\cdot x}(\omega) = \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega) f_{yy}(\omega)},$$

where $f_{xx}(\omega)$ and $f_{yy}(\omega)$ are the individual spectra of the x_t and y_t, respectively. The first step then is to calculate the auto-covariance functions of x_t and y_t. For the time series x_t we have

$$\begin{aligned} \gamma_{xx}(h) &= \mathrm{Cov}(x_{t+h}, x_t) = \mathrm{E}[x_{t+h} xt] \\ &= E[(w_{t+h} + 0.4 w_{t+h-1})(w_t + 0.4 w_{t-1})] \\ &= 1.16\gamma_w(h) + 0.4\gamma_w(h+1) + 0.4\gamma_w(h-1), \end{aligned}$$

i.e.

$$\gamma_{xx}(h) \begin{cases} 1.16\sigma_w^2 & \text{if } h = 0, \\ 0.4\sigma_w^2 & \text{if } |h| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the spectra is

$$\begin{aligned} f_{xx}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{xx}(h) e^{-2\pi i\omega h} \\ &= \sigma_w^2 [1.16 + 0.4(e^{-2\pi i\omega} + e^{2\pi i\omega})] \\ &= \sigma_w^2 [1.16 + 0.8\cos(2\pi\omega)] \\ &= 2.32 + 1.6\cos(2\pi\omega). \end{aligned}$$

The same steps for y_t gives

$$\begin{aligned}\gamma_{yy}(\omega) &= \mathrm{Cov}(y_{t+h}, y_t) = \mathrm{E}[y_{t+h}y_t] \\ &= E[(0.3w_{t+h-1} + 0.12w_{t+h-2} + v_{t+h})(0.3w_{t-1} + 0.12w_{t-2} + v_t)] \\ &= 0.1044\sigma_w^2 + \sigma_v^2 + 0.036(\gamma_w(h+1) + \gamma_w(h-1))\end{aligned}$$

i.e

$$\gamma_{yy}(h) \begin{cases} 0.1044\sigma_w^2 + \sigma_v^2 & \text{if } h = 0, \\ 0.036\sigma_w^2 & \text{if } |h| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

and the spectral density

$$\begin{aligned}f_{yy}(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_{yy}(h) e^{-2\pi i \omega h} \\ &= 4.2088 + 0.144\cos(2\pi\omega).\end{aligned}$$

Lastly, we compute the squared coherence

$$\begin{aligned}\rho^2_{y\cdot x}(\omega) &= \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega) f_{yy}(\omega)} \\ &= \frac{|0.24 + 0.696e^{-2\pi i\omega} + 0.24e^{-4\pi i\omega}|^2}{(2.32 + 1.6\cos(2\pi\omega))(4.2088 + 0.144\cos(2\pi\omega))} \\ &= \frac{0.576 + 2\cdot 0.696^2 \cdot\ cos\,(2\pi\cdot 0.2) + 2\cdot 0.24^2\cdot\ cos\,(4\pi\cdot 0.2)}{(2.32 + 1.6\cos(2\pi\cdot 0.2))(4.2088 + 0.144\cos(2\pi\cdot 0.2))},\end{aligned}$$

for $\omega = 0.2$

$$\rho^2_{y\cdot x}(0.2) \approx 0.06534.$$

2 Problem 2

The file carsmon752203.dat at Studium contains monthly data for the number of cars in traffic in Sweden starting 1975 and ending in March 2022.

(a) Plot the series.

(b) Estimate the spectral density non parametrically and parametrically, and plot your results.

(c) Do the same as in (a)-(b) for the differenced series.

(d) Take a yearly moving average of the series (i.e. over 12 consecutive months) and do the same as in (a)-(b).

(f) Compare and discuss your results.

Solution (a) We plot the time series in figure 1.

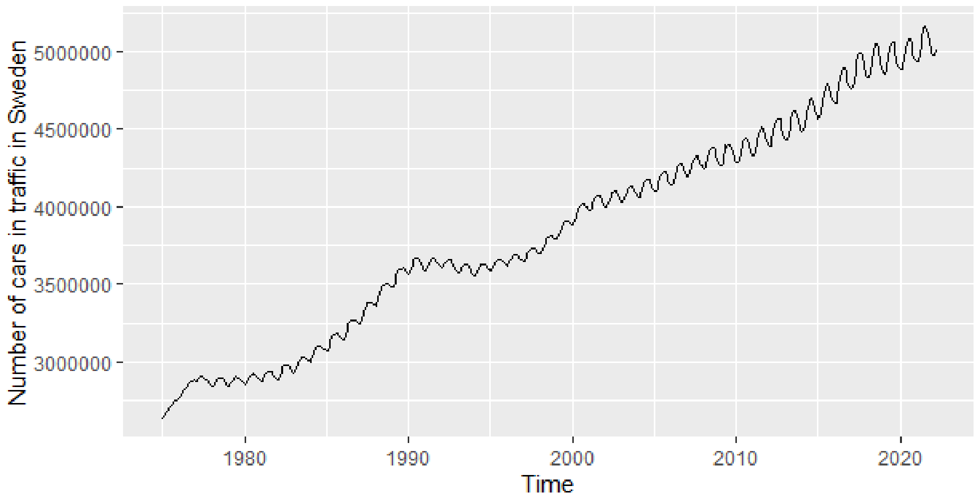

Figure 1: Time series of monthly data.

Solution (b) We can see from figure 1 that the series has an upward trend throughout the entire series. The series has a total of 567 data points over a 47 year period. Estimating the spectral density with no defined spans we get figure 2. This is a raw periodogram and as expected not very smooth because it is not a consistent estimator. The blue line in the figure represents the 95% confidence interval. The periodogram has an expected peak at frequency equal to one as well as frequency equal to two and three. We expect the peaks to be a lot less pronounced as we apply spans to the estimation. Additionally, we can infer that the series is non-stationary (trend) because of the large peak at frequency zero. In figure 3 we apply some mild smoothing and we can see that we still have the same peaks as before at one, two and three, as well as the trend peak at zero and the residuals peak at six. So far the result is as expected and we continue with larger spans in figure 4 which gives an even smoother estimate. We can now infer some signs of over smoothing, since the peak at one is showing signs of plateauing. We continue with figures for larger spans shown in figures 5, 6 and 7. It seems like the best choice would be with spans=10 or spans=12 for this estimation. The last figure with spans=14 carries some large signs of over smoothing and plateauing.

The Parametric spectral estimation is presented in figure 8. The only peak in the figure is the trend at frequency equals to zero. This can be due to the business cycle or non-stationarity. The estimation spectrum in this case is an AR(2) model.

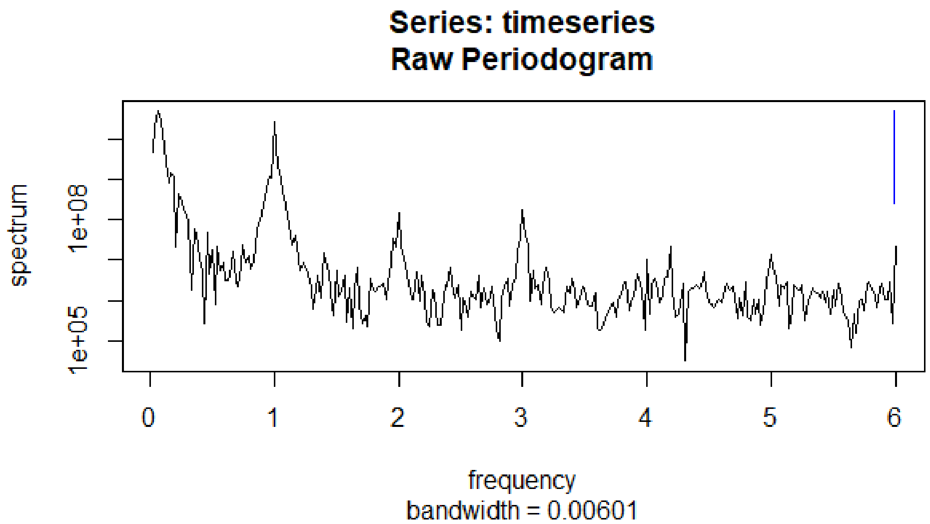

Figure 2: Estimation of spectral density

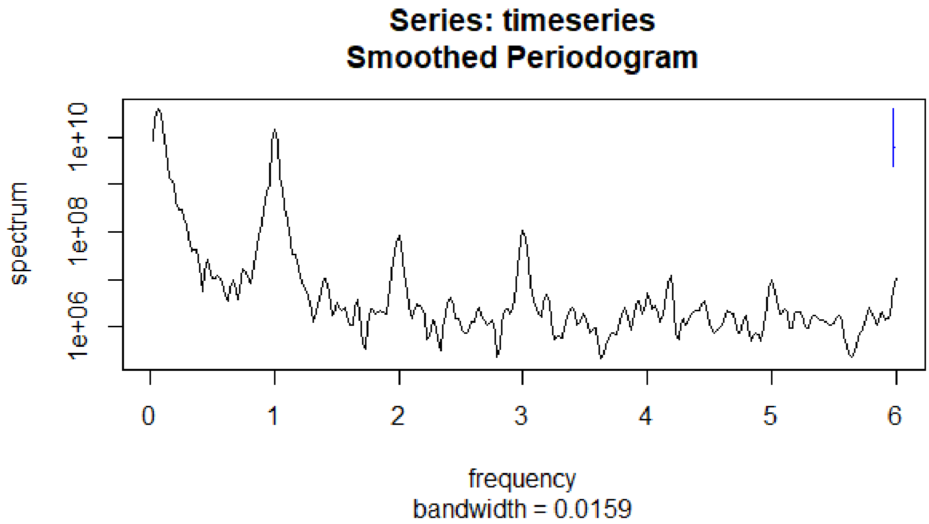

Figure 3: Estimation of spectral density (spans=2)

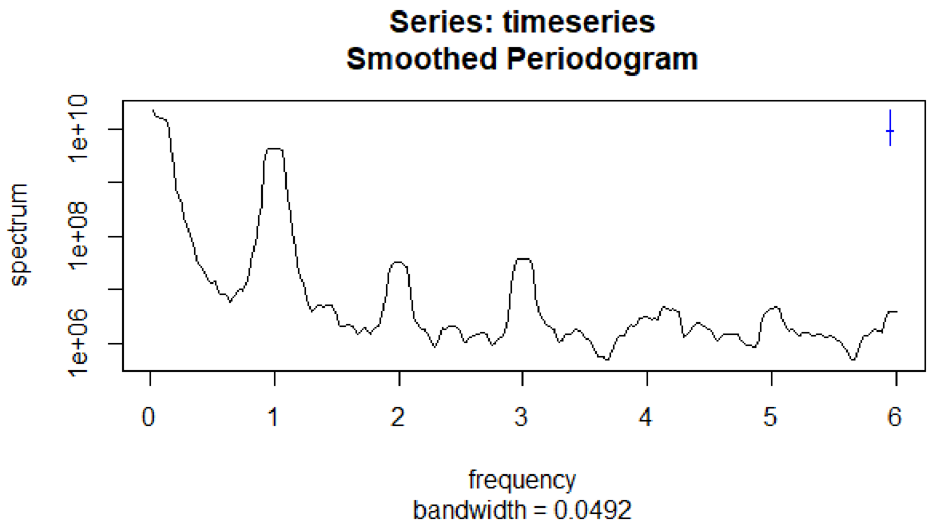

Figure 4: Estimation of spectral density (spans=8)

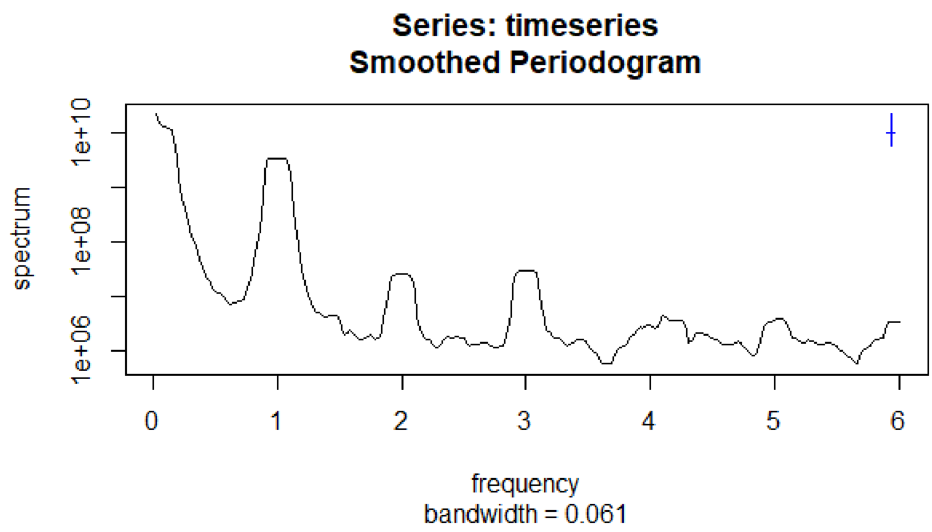

Figure 5: Estimation of spectral density (spans=10)

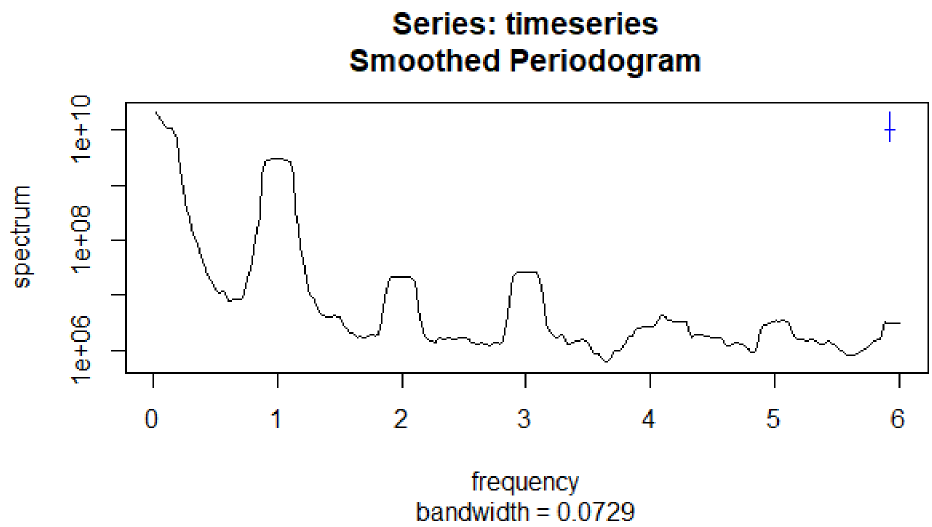

Figure 6: Estimation of spectral density (spans=12)

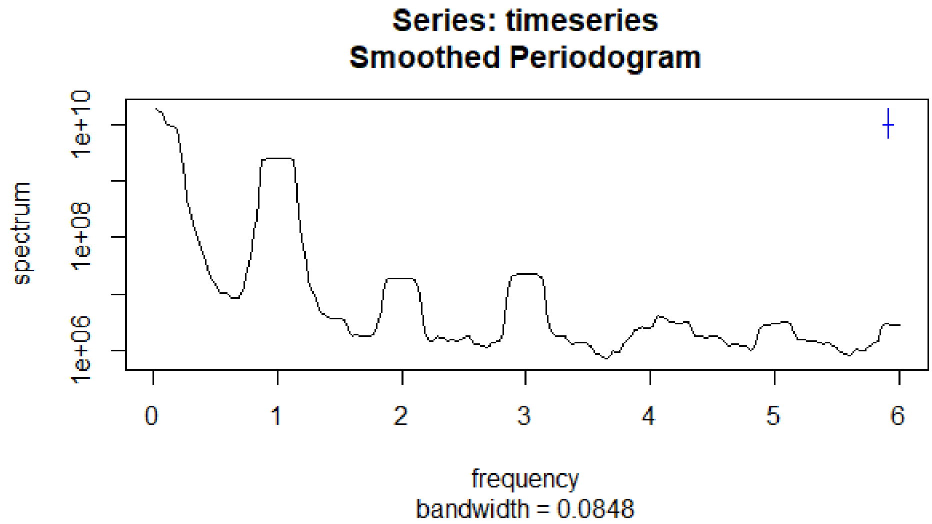

Figure 7: Estimation of spectral density (spans=14)

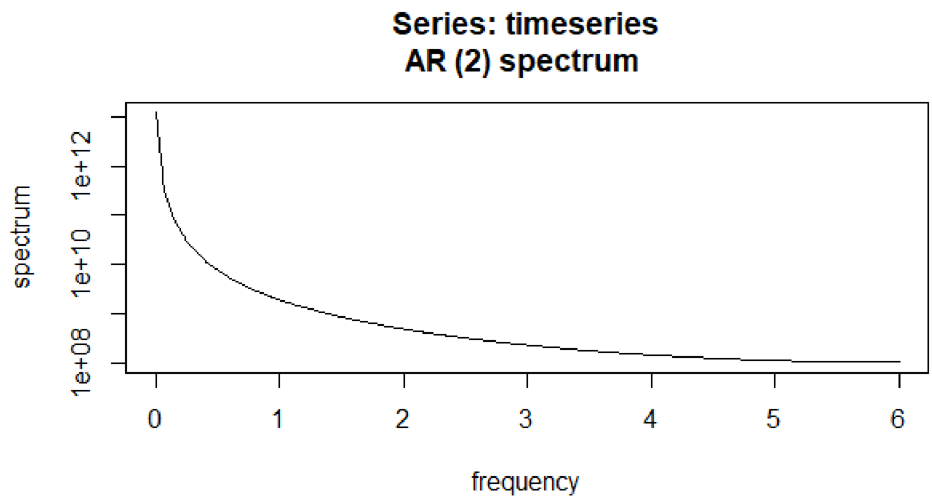

Figure 8: Parametric estimation of spectral density

Solution (c) Following the same procedure as in (a) and (b) we see the differenced series in figure 9. In this plot we see that the series is now stationary and it is oscillating around zero mean. The spectral estimation seems to have the best choice again around spans=10. This time we still see some residual peaks, but the (trend) peak is a lot less predominant.

The parametric spectral estimation for the differenced time series is shown in figure 8. In this case we see similar peaks to the non parametrical estimation at frequency equals to 1, 2 and 3. Most of the same properties can be observed when compared to the estimation shown in figure 7.

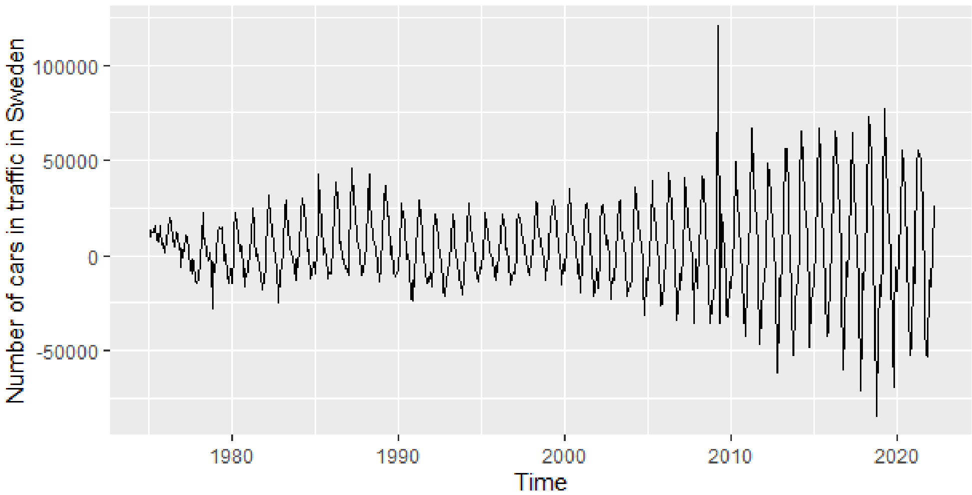

Figure 9: Differenced time series of monthly data.

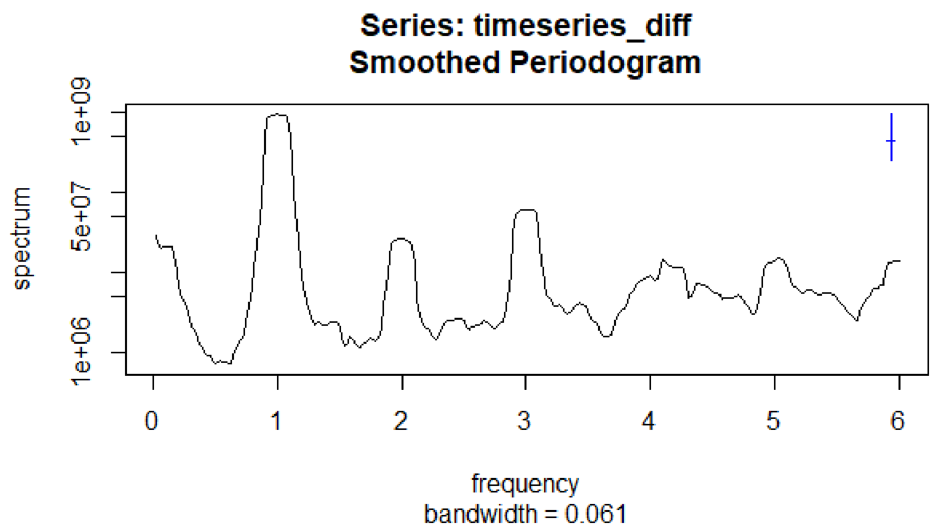

Figure 10: Estimation of spectral density (spans=10)

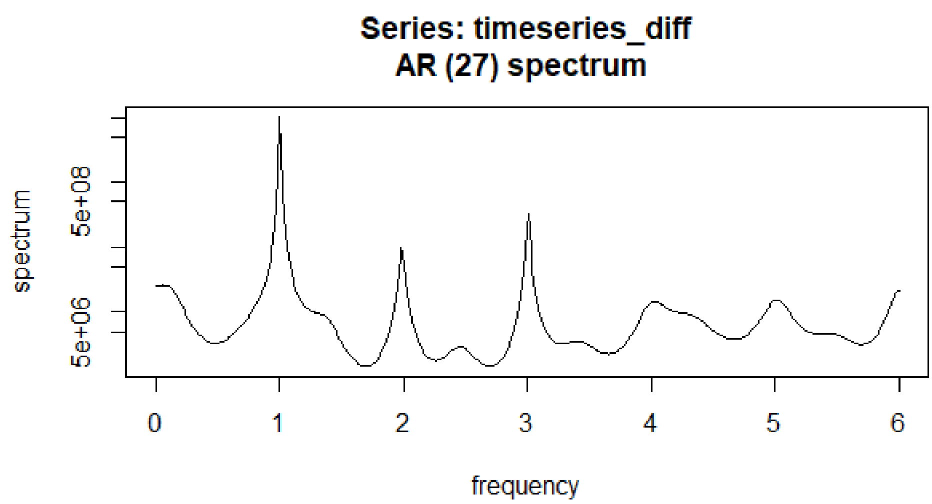

Figure 11: Parametric estimation of spectral density

Solution (d) The yearly moving average is shown in figure 12. Similarly, the non parametrically spectral estimate is shown in figure 13. In this case, the plateauing started earlier so spans equal to 8 or 10 provide some what smooth estimation with less predominant peaks.

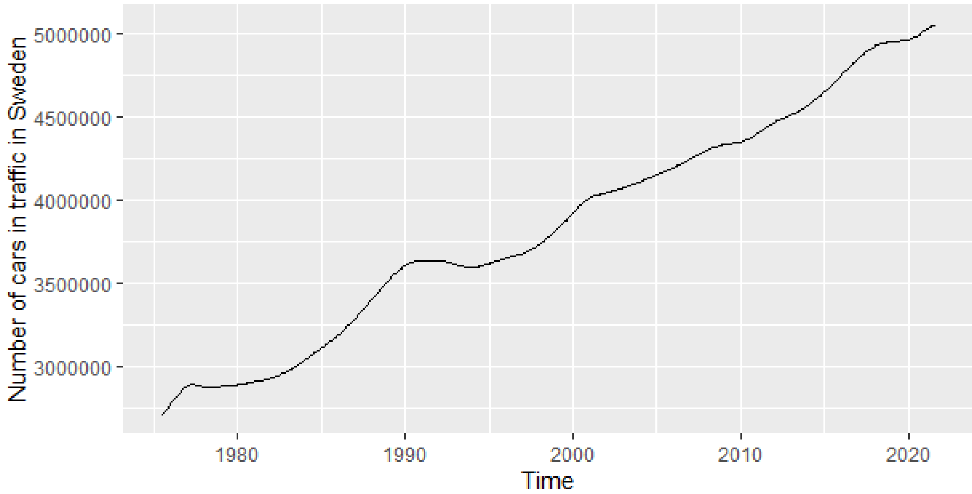

Figure 12: Yearly moving average time series

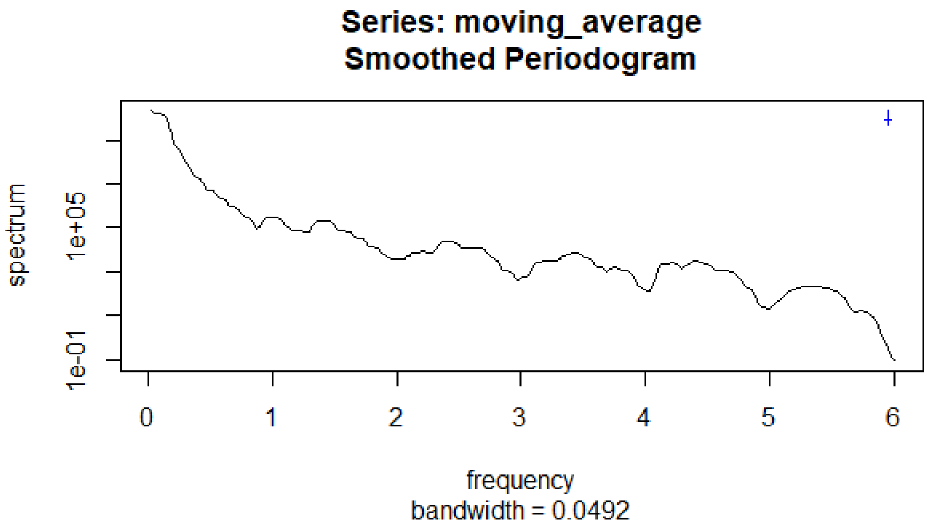

Figure 13: Estimation of spectral density (spans=8)

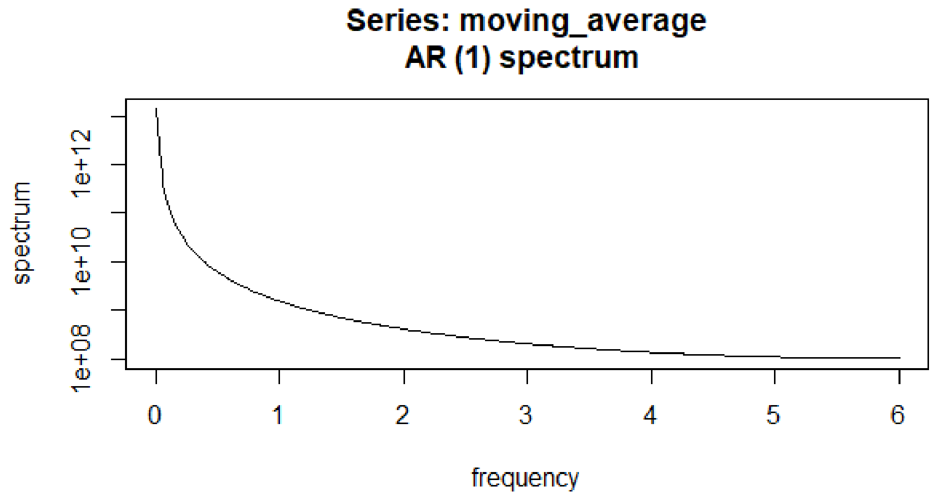

Figure 14: Parametric estimation of spectral density

Solution (f) Considering the non-parametric estimations above, the original time-series is comparable to the differenced one with the exception of the trend at frequency around zero. While the moving average model notably eliminated most of the peaks and gave a much more smooth estimation. For the parametric estimations we saw about the same in the original and moving average estimates, with no peaks and them being on the AR(2) and AR(1) spectrum, respectively. The differenced time series produced an AR(27) spectrum which is a lot more similar to the non-parametric estimation.

3 Problem 3

The files storliensnow.dat and storlienperc.dat at Studium contain daily data on snow depth in meters and the percipitation in mm in Storlien, Sweden, for December 2, 2021 until April 11, 2022.
Find a suitable transfer function model to describe these data

Solution We denote the snow depth in meters as y_t in figure 15 and the precipitation in mm as x_t in figure 16. The data set contains 131 data points over the period from December 2, 2021 to April 11, 2022.

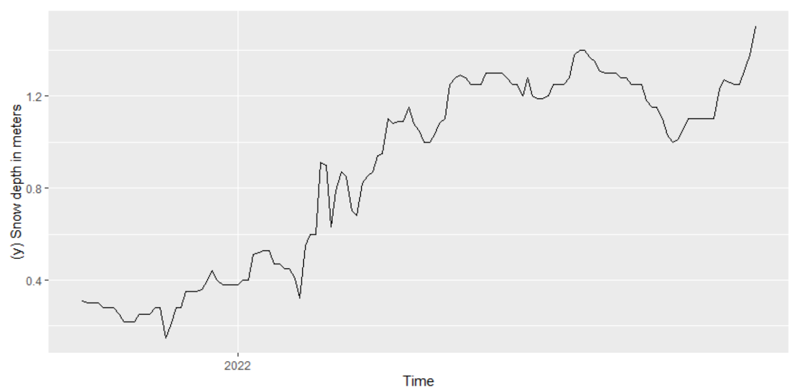

Figure 15: Time series y_t

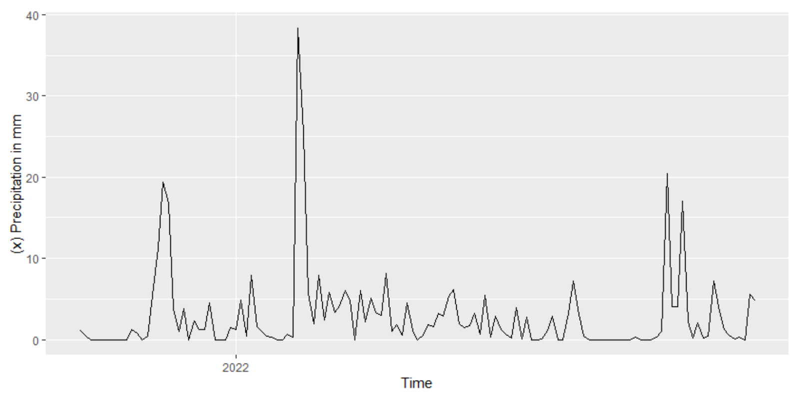

Figure 16: Time series x_t

At first, we have to make sure that the x_t series is stationary and have mean zero.

The stationary tests we apply to the x_t series are the Augmented Dickey-Fuller test and Kwiatkowski–Phillips–Schmidt–Shin test. We start with examining the figure for the time series and see that there seems to be no trend or seasonality. For the ADF test we get a p value of less than 0.01, i.e. the p-value from the test is less than some significance level, then we can reject the null hypothesis and conclude that the time series is stationary. The KPSS test provides the same result; Hence, we will not apply any difference to the x_t series. The series x_t however does not have mean zero, therefore we subtract the mean. The series y_t is non-stationary, this is because the obvious upward trend. The statistical tests (ADF and KPSS) confirm this and we applied a difference to the series.

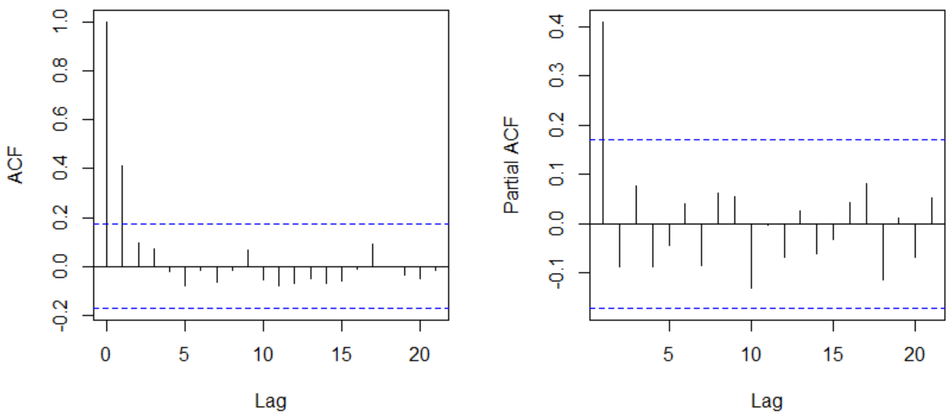

Figure 17: ACF and PACF for x_t

Answer Examining the ACF and PACF in figure 17, we see that the ACF cuts of after the strong correlation at lag 1 and then oscillates around 0 with no significant lags after that. The PACF shows no significant lags and tails off. This suggest an MA(q) model and after trails for several models, the MA(1) model has the lowest AIC.

```
ARIMA(0,0,1) with zero mean

Coefficients:
         ma1
      0.4327
s.e.  0.0746

sigma^2 = 22.34:  log likelihood = -388.95
AIC=781.89   AICc=781.99   BIC=787.64
```

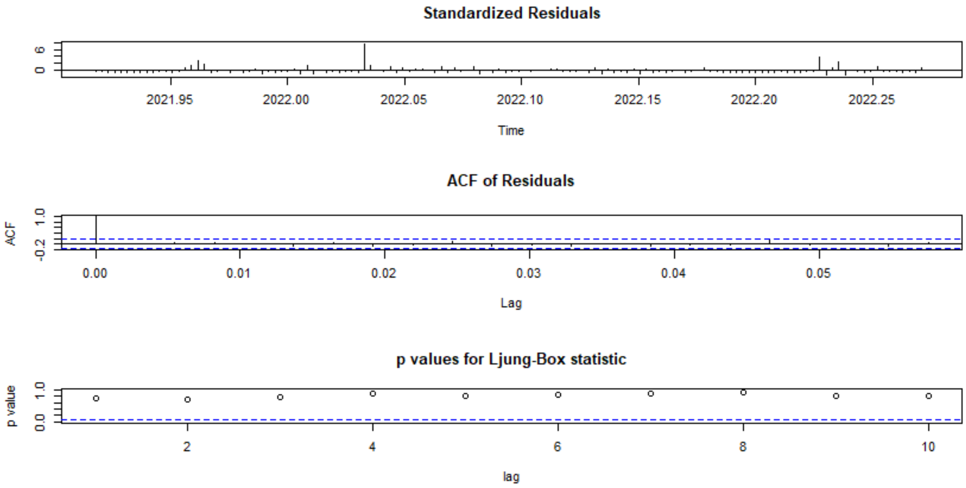

Figure 18: Residuals of the ARIMA(0,0,1) model

In figure 18, the residuals of the model are presented. Analyzing the standardized residuals we see the expected random pattern, they have zero mean and they seem uncorrelated. The lack of correlation in the ACF suggests the model utilized is sufficient. Lastly, the p-values in the Ljung-Box statistic are the not significant i.e., the p-values are relatively large compared to the significance level. Thus, we can conclude that the residuals are not distinguishable from a white noise series.

We can represent the ARIMA model as $\phi(B)xt = \theta(B)w_t$, where w_t is white noise. The estimated model for input:

$$x_t = w_t + 0.4327w_{t-1}$$

i.e. $x_t = (1 + 0.4327B)w_t$. Hence, the prewhitening of input is represented as $y_t = \alpha(B)x_t + \eta_t = \alpha(B)\frac{\theta(B)}{\phi(B)}w_t + \eta_t$, i.e.

$$y_t = \alpha(B)(1 + 0.4327B)w_t + \eta_t.$$

Prewhitening of output:

$$\begin{aligned}\tilde{y}_t &= \frac{\phi(B)}{\theta(B)}y_t = \alpha(B)w_t + \tilde{\eta}_t,\\ &= \frac{1}{(1 + 0.4327B)}w_t + \tilde{\eta}_t.\end{aligned}$$

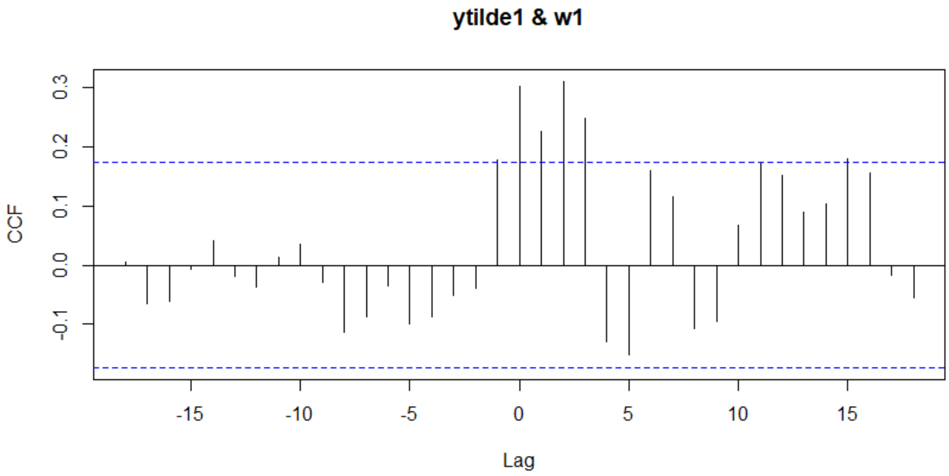

Figure 19: CCF

The *Cross-Correlation Function* (CCF) is shown in figure 19, calculated by

$$\gamma_{\tilde{y}w}(h) = \mathrm{E}(\tilde{y}_{t+h}w_t) = \sigma_w^2 \alpha_h.$$

The most significant cross correlations occur between lags $h = 0$ and $h = 3$. No other spikes occur at any other lag at a significant levels. There are a lot of models that we could try based on the CCF for the regression model. We want to identify d, s and r from

$$\alpha(B) = \frac{\delta(B)}{\omega(B)} B^d,$$

where

$$\delta(B) = \delta_0 + \delta_1 B + \cdots + \delta_s B^s,$$
$$\omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_r B^r.$$

Out of the models tested the best approximation came from $d = 1, s = 0$ and $r = 3$ with zero AR coefficients for lag 1, i.e.

$$y_t = \frac{\delta_0}{1 - \omega_2 B^2 - \omega_3 B^3} B x_t + \eta_t.$$

We estimate the regression $(1 - \omega_2 B^2 - \omega_3 B^3) y_t = \delta_0 B x_t + u_t$, where $u_t = (1 - \omega_2 B^2 - \omega_3 B^3)\eta_t$. The residuals produced are

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.212811 -0.017744  0.006384  0.031759  0.277800

Coefficients:
```

```
    Estimate Std. Error t value Pr(>|t|)
x0  0.002977   0.001036   2.874  0.00477 **
y2 -0.179677   0.085427  -2.103  0.03746 *
y3  0.190710   0.085444   2.232  0.02741 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '␣' 1

Residual standard error: 0.06101 on 124 degrees of freedom
Multiple R-squared:  0.1336,    Adjusted R-squared:  0.1127
F-statistic: 6.375 on 3 and 124 DF,  p-value: 0.000471
```

The coefficients are all within the significance at twice the standard error and the probabilities seem large enough to be included. An issue with this approximation is that the R-squared value is only about 0.133%, this would suggest the model does not capture a lot of the variance of the data. The estimated Regression

$$y_t = 0.0030x_t - 0.180y_{t-2} + 0.190y_{t-3} + u_t$$

where $u_t = (1 + 0.180B^2 - 0.190B^3)\eta_t$. Then we construct $\eta_t = (1 + 0.180B^2 - 0.190B^3)^{-1}u_t$ and fit an ARMA model $\phi_\eta(B)\eta_t = \theta_\eta(B)z_t$ where z_t is white noise.

The ACF and PACF in figure 21 looks decent with a couple significant lags at smaller values. The next step is to fit an ARIMA model to the dataset for η_t. The ACF and PACF suggests an AR(p) process since the PACF signifies a cuts off after lag 2. The model could include an MA(q) process since there are significant lags, although it could also represent geometric decay. Trail and error suggest an ARIMA(2,0,0) model. This model produces the ACF and PACF in figure 21, the histogram and normal Q-Q in figure 22, and the residuals in figure 23. The final model

$$\phi_\eta(B)\omega(B)y_t = \phi_\eta(B)\delta(B)B^d x_t + \omega(B)\theta_\eta(B)z_t$$

is estimated as

$$\begin{aligned}&(1 + 0.0226B + 0.2411B^2)(1 + 0.180B^2 - 0.190B^3)y_t\\&= (1 + 0.0226B + 0.2411B^2)0.0030Bx_t + (1 + 0.180B^2 - 0.190B^3)z_t,\end{aligned}$$

where $z_t = (1 + 0.0226B + 0.2411B^2)\eta_t$ is white noise

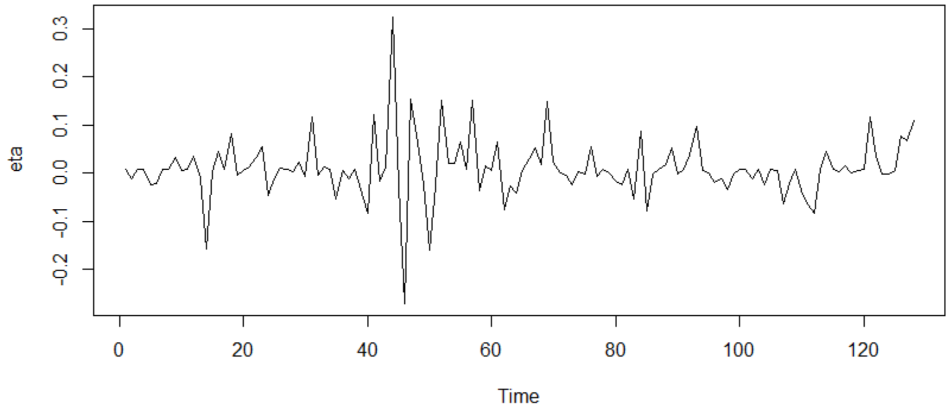

Figure 20: Plot of η_t

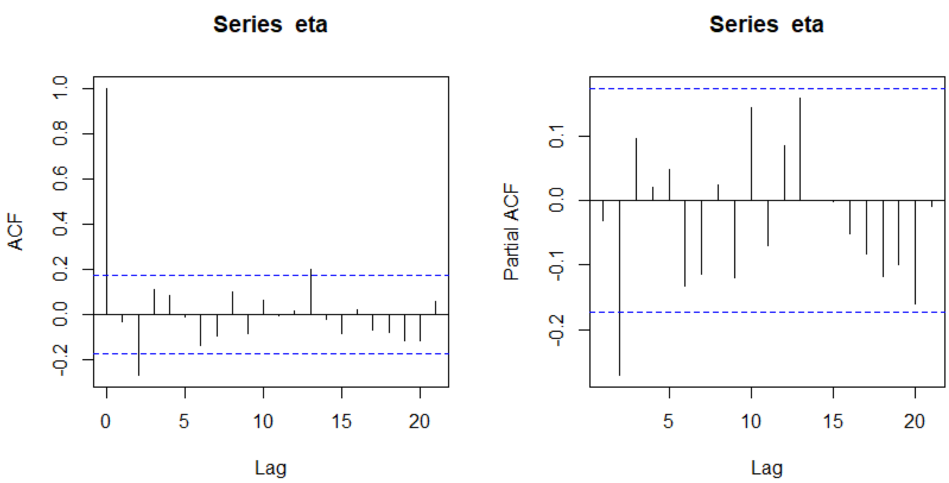

Figure 21: ACF and PACF of η_t

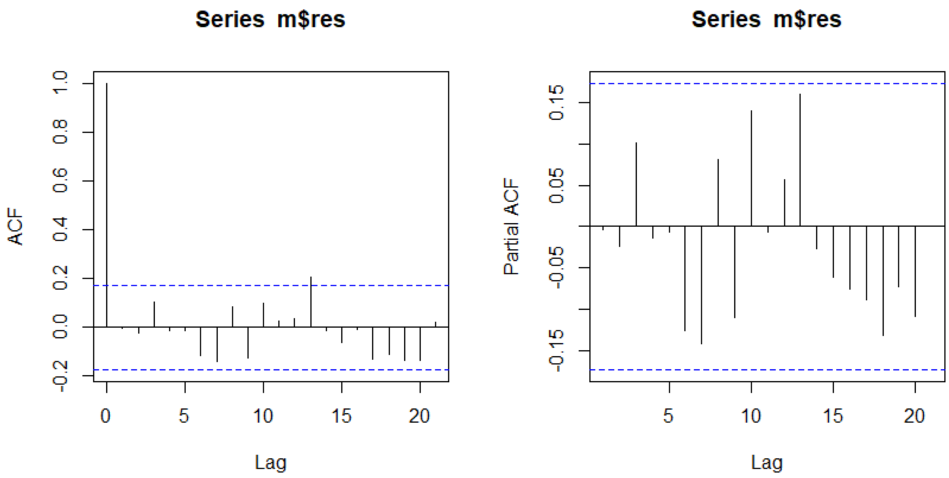

Figure 22: ACF and PACF of η_t model

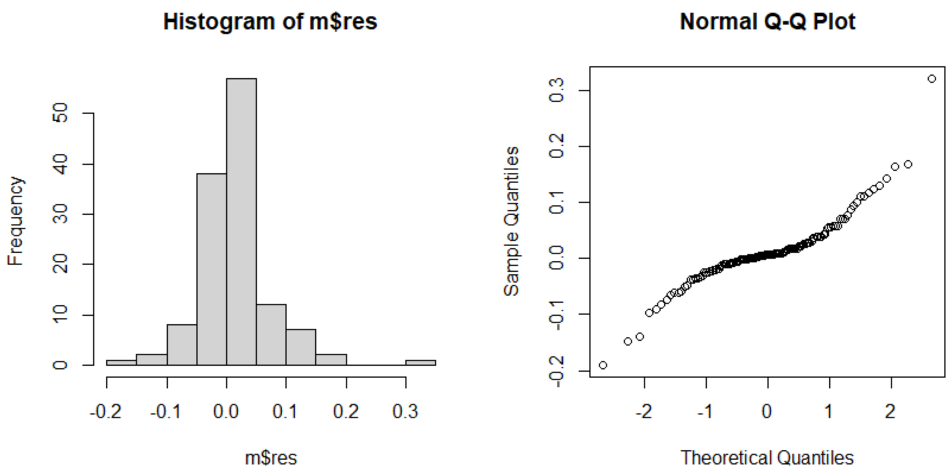

Figure 23: Histogram and Normal Q-Q Plot of η_t model

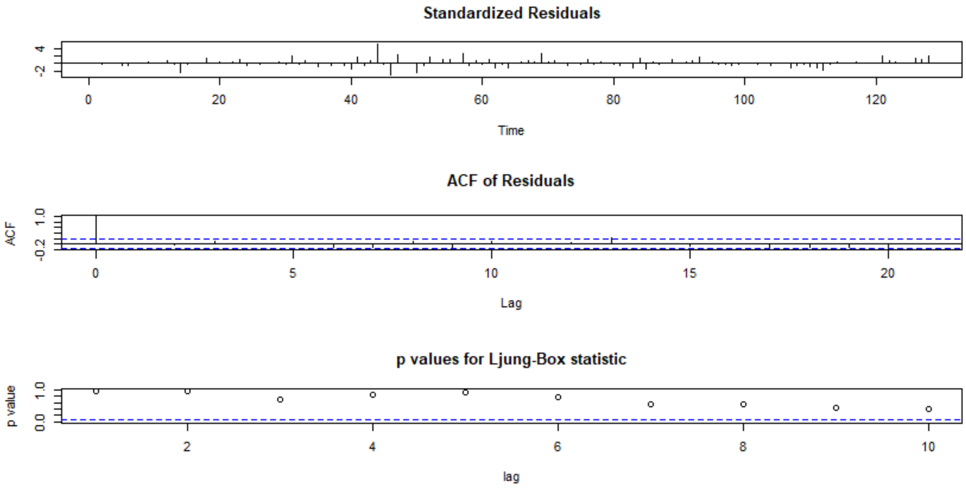

Figure 24: Residuals of η_t model

A R code

A.1 Problem 2

```
library(forecast)
x<-read.table("carsmon752203.dat")$V1
timeseries<-ts(x, frequency=12,start=c(1975,1))

# (a)
autoplot(timeseries,ylab = "Number␣of␣cars␣in␣traffic␣in␣Sweden")

# (b)
spec.pgram(timeseries)
spec.pgram(timeseries,spans = 2)
spec.pgram(timeseries,spans = 8)
spec.pgram(timeseries,spans = 10)
spec.pgram(timeseries,spans = 12)
spec.pgram(timeseries,spans = 14)
spec.pgram(timeseries,spans = 24)

spec.ar(timeseries)

# (c)
timeseries_diff<-diff(timeseries)
autoplot(timeseries_diff,ylab = "Number␣of␣cars␣in␣traffic␣in␣Sweden")

spec.pgram(timeseries_diff)
spec.pgram(timeseries_diff,spans = 2)
spec.pgram(timeseries_diff,spans = 8)
spec.pgram(timeseries_diff,spans = 10)
spec.pgram(timeseries_diff,spans = 12)
spec.pgram(timeseries_diff,spans = 14)
spec.pgram(timeseries_diff,spans = 24)

spec.ar(timeseries_diff)
# (d)
moving_average<-ma(timeseries,order=12)
moving_average<-na.omit(moving_average)
autoplot(moving_average,ylab = "Number␣of␣cars␣in␣traffic␣in␣Sweden")

spec.pgram(moving_average)
spec.pgram(moving_average,spans = 2)
spec.pgram(moving_average,spans = 8)
spec.pgram(moving_average,spans = 10)
spec.pgram(moving_average,spans = 12)
spec.pgram(moving_average,spans = 14)
spec.pgram(moving_average,spans = 24)
```

```
spec.ar(moving_average)
```

A.2 Problem 3

```
library(forecast)
library(lubridate)
library(tseries)

# read data and create time series
# y is the snow depth in meters and
# x is the precipitation in mm
y <- read.table("storliensnow.dat")$V1
x <- read.table("storlienprec.dat")$V1
y <- ts(y,frequency = 365, start = decimal_date(ymd("2021-12-2")))
x <- ts(x,frequency = 365, start = decimal_date(ymd("2021-12-2")))

# first plot
autoplot(y,xlab="Time",ylab="(y)␣Snow␣depth␣in␣meters")
autoplot(x,xlab="Time",ylab="(x)␣Precipitation␣in␣mm")

# stationary test x
adf.test(x, alternative = c("stationary", "explosive"),
         k = 0)
kpss.test(x)

# stationary test y
adf.test(y, alternative = c("stationary", "explosive"),
         k = 0)
kpss.test(y)

# Create zero mean time series for x and y
x <- x - mean(x)
y <- diff(y)
# ACF and PACF
par(mfrow=c(1,2));acf(x[1:length(x)], main="");pacf(x[1:length(x)],main="")

# ARIMA model
x_model <- arima(x, order=c(0,0,1), include.mean = FALSE)

# diagnostics
tsdiag(x_model)

# Prewhitening of input and output
theta1 <- as.numeric(x_model$coef[1])
ytilde <- filter(y, filter = c(1,-theta1),
                 method = "c", sides = 1)
w <- filter(x, filter = c(1,-theta1),
            method = "c", sides = 1)
```

```r
#CCF
n <- length(y)
ytilde1 <- ytilde[seq(2,n)]
w1 <- w[seq(2,n)]
par(mfrow=c(1,1))
ccf(ytilde1,w1,ylab="CCF")

# Estimated regression
y0 <- y[seq(4,n)]
x0 <- x[seq(4,n)]
y2 <- y[seq(2,n-2)]
y3 <- y[seq(1,n-3)]

r <- lm(y0 ~ x0+y2+y3-1);summary(r)

u <- r$res
omega1 <- as.numeric(r$coef[2])
omega2 <- as.numeric(r$coef[3])
eta <- filter(u, filter = c(0,omega1,omega2), method = "recursive")
plot(eta, type = "l")
par(mfrow = c(1,2));acf(eta);pacf(eta)

# eta model
m <- arima(eta, order = c(2,0,0), include.mean=FALSE)

par(mfrow = c(1,2));acf(m$residuals);pacf(m$residuals)

hist(m$res)
qqnorm(m$res)

tsdiag(m)
```