Assignment 4

2022-12-13

Question 1

a) We have $\mathrm{E}(X_t) = \mathrm{E}(Y - 2Z_{t-1} + Z_t)$ and since Y is independent of $\{Z_t\}$, it follows that

$$\mathrm{E}(X_t) = \mathrm{E}(Y - 2Z_{t-1} + Z_t) = \mathrm{E}(Y) - 2\,\mathrm{E}(Z_{t-1} + \mathrm{E}(Z_t)) = 1.$$

Furthermore,

$$\mathrm{E}(X_t X_{t+h}) = \mathrm{E}((Y - 2Z_{t-1} + Z_t)(Y - 2Z_{t+h-1} + Z_{t+h})) = \mathrm{E}(Y^2) = \mathrm{Var}(Y) - \mathrm{E}(Y)^2 = 0$$

because in the expansion, any term with Z_{t+k} will have 0 expectation. Thus, $\{X_t\}$ is stationary. We can see that $\{Y_t\}$ is not stationary because X_t depends on t and thus, $\mathrm{E}(Y_t)$ depends on t.

b) We have $\nabla_d(W_t + 1) = W_t + 1 - W_{t-d} - 1 = W_t - W_{t-d}$ so it follows that

$$\begin{aligned}
\nabla_d^2(W_t + 1) &= \nabla_d(W_t - W_{t-d}) \\
&= W_t - W_{t-d} - W_{t-d} + W_{t-2d} \\
&= W_t - 2W_{t-d} + W_{t-2d} \\
&= at^2 + bts_t + X_t - 2(a(t-d)^2 + b(t-d)s_t + X_{t-d}) + a(t-2d)^2 + b(t-2d)s_t + X_{t-2d} \\
&= a(t^2 - 2t^2 + 4dt - 2d^2 + t^2 - 4dt + 4d^2) + bs_t(t - 2t + 2d + t - 2d) + X_t - 2X_{t-d} + X_{t-2d} \\
&= ad^2 + bs_t 0 + X_t - 2X_{t-d} + X_{t-2d} = ad^2 + X_t - 2X_{t-d} + X_{t-2d}.
\end{aligned}$$

If $\{X_t\}$ is stationary then $Y_t = ad^2 + (X_t - X_{t-d}) - (X_{t-d} - X_{t-2d})$ is also stationary because $\{(X_t - X_{t-d})\}$ and $\{(X_{t-d} - X_{t-2d})\}$ are stationary.

Question 2

a) Note that $aY_{t-1} = 2aX_{t-1} - aX_{t-2}$ so

$$\begin{aligned}
aY_t &= 2aX_t - aX_{t-1} \\
&= 2a(0.2X_{t-1} + Z_t - Z_{t-1}) - a(0.2X_{t-2} + Z_{t-1} - Z_{t-2}) \\
&= 0.4aX_{t-1} - 0.2aX_{t-2} - aZ_{t-1} + aZ_{t-2} + 2aZ_t \\
&= \frac{1}{5}(2aX_{t-1} - aX_{t-2}) - aZ_{t-1} + aZ_{t-2} + 2aZ_t \\
&= \frac{1}{5}aY_{t-1} - aZ_{t-1} + aZ_{t-2} + 2aZ_t.
\end{aligned}$$

Let $a = 1/2$, we have

$$aY_t = \frac{1}{5}aY_{t-1} - aZ_{t-1} + aZ_{t-2} + Z_t$$

so $\{aY_t\}$ follows an ARMA(p, q) model for $a = 1/2$, $\alpha_1 = 1/5$, $\beta_1 = \beta_2 = a$, $p = 1$, and $q = 2$.

b) Assuming that $\{X_t\}$ is stationary,

$$\text{Cov}(X_t, Z_t) = \text{Cov}(0.2X_{t-1} + Z_t - Z_{t-1}, Z_t) = 0.2\text{Cov}(X_{t-1}, Z_t) + 1 = 1.$$

We have $\mathrm{E}(X_t) = \mathrm{E}(0.2X_{t-1} + Z_t + Z_{t-1}) = 0.2\,\mathrm{E}(X_{t-1}) = 0.2\,\mathrm{E}(X_t)$, thus it follows that $\mathrm{E}(X_t) = 0$. Also,

$$\text{Var}(X_t) = \text{Var}(0.2X_{t-1} + Z_t - Z_{t-1}) = 0.04\text{Var}(X_{t-1}) + 2 = 0.04\text{Var}(X_t) + 2$$

then $\text{Var}(X_t) = \frac{2}{0.96}$. We have

$$\begin{aligned}
\gamma_X(1) &= \text{Cov}(X_t, X_{t+1}) \\
&= \text{Cov}(X_t, 0.2X_t + Z_{t+1} - Z_t) \\
&= \text{Cov}(X_t, 0.2X_t) + \text{Cov}(X_t, Z_{t-1}) - \text{Cov}(X_t, Z_t)) \\
&= 0.2\text{Var}(X_t) - 1 \\
&= 0.2\frac{2}{0.96} - 1 = -0.583.
\end{aligned}$$

Question 3

a) The autocovariance function

$$\gamma_Z(h) = \begin{cases} \sigma^2 & \text{for } h = 0 \\ 0 & \text{for } h \neq 0 \end{cases}$$

then the spectral density

$$f_Z(\lambda) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma_Z(h)e^{-ih\lambda} = \frac{1}{2\pi}\sigma^2.$$

b) $\{X_t\}$ is an MA(q) time series by introducing new WN $Z_t^{'} = Z_{t+1}$ We have

$$\mathrm{E}(X_t) = \mathrm{E}\left(\frac{1}{2}Z_{t+1} + Z_t - \frac{1}{2}Z_{t-1}\right) = \frac{1}{2}\,\mathrm{E}(Z_{t+1}) + \mathrm{E}(Z_t) - \frac{1}{2}\,\mathrm{E}(Z_{t-1}) = \sigma^2$$

not dependent on t. Doing the same thing for $\mathrm{E}(X_t X_{t+h})$ to see that the expansion only contains Z_t terms so it is also independent of t. Thus, $\{X_t\}$ is stationary.

The autocovariance function

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{i=0}^{2-h} \beta_i\beta_{i+h} & h = 0, 1, \ldots, 2, \\ \gamma_X(-h) & h = -1, \ldots, -2, \\ 0, & \text{otherwise} \end{cases}$$

and the spectral density

$$\begin{aligned}
f_X(\lambda) &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\gamma_X(h)e^{-ih\lambda} \\
&= \frac{1}{2\pi}\sum_{h=-2}^{2}\gamma_X(h)e^{-ih\lambda} \\
&= \frac{1}{2\pi}\left(\sum_{h=-2}^{-1}\gamma_X(h)e^{-ih\lambda}+\sum_{h=0}^{2}\gamma_X(h)e^{-ih\lambda}\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(2)e^{i2\lambda}+\gamma_X(1)e^{i\lambda}+\gamma_X(0)+\gamma_X(1)e^{-i\lambda}+\gamma_X(2)e^{-i2\lambda}\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(2)e^{i2\lambda}+\gamma_X(1)e^{i\lambda}+\gamma_X(0)+\gamma_X(1)e^{-i\lambda}+\gamma_X(2)e^{-i2\lambda}\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(2)(e^{i2\lambda}+e^{-i2\lambda}+\gamma_X(0))+\gamma_X(1)(e^{i\lambda}+e^{-i\lambda})\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(2)(e^{i2\lambda}+e^{-i2\lambda})+\gamma_X(0)+\gamma_X(1)(e^{i\lambda}+e^{-i\lambda})\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(2)\left(e^{i2\lambda}+e^{-i2\lambda}\right)+\frac{3}{2}\sigma^2\right) \\
&= \frac{1}{2\pi}\sigma^2\left(\frac{-1}{2}\cos 2\lambda+\frac{3}{2}\right)
\end{aligned}$$

c) $\{X_t\}$ is a linear transformation of $\{Z_t\}$ because $X_t = \frac{1}{2}Z_t' + Z_{t-1}' - \frac{1}{2}Z_{t-2}'$ which is a linear combination of $\{Z_t'\}$ by letting $Z_t' = Z_{t+1}$. The corresponding filter coefficients $\psi_{-1} = 1/2$, $\psi_0 = 1$, $\psi_1 = -1/2$, and $\psi_k = 0$ for $k \notin \{0, 1, 2\}$. The transfer function

$$\psi(\lambda) = \sum_j \psi_j e^{-ij\lambda} = \frac{1}{2}e^{i\lambda} + 1 - \frac{1}{2}e^{-\lambda i}.$$

We then have that

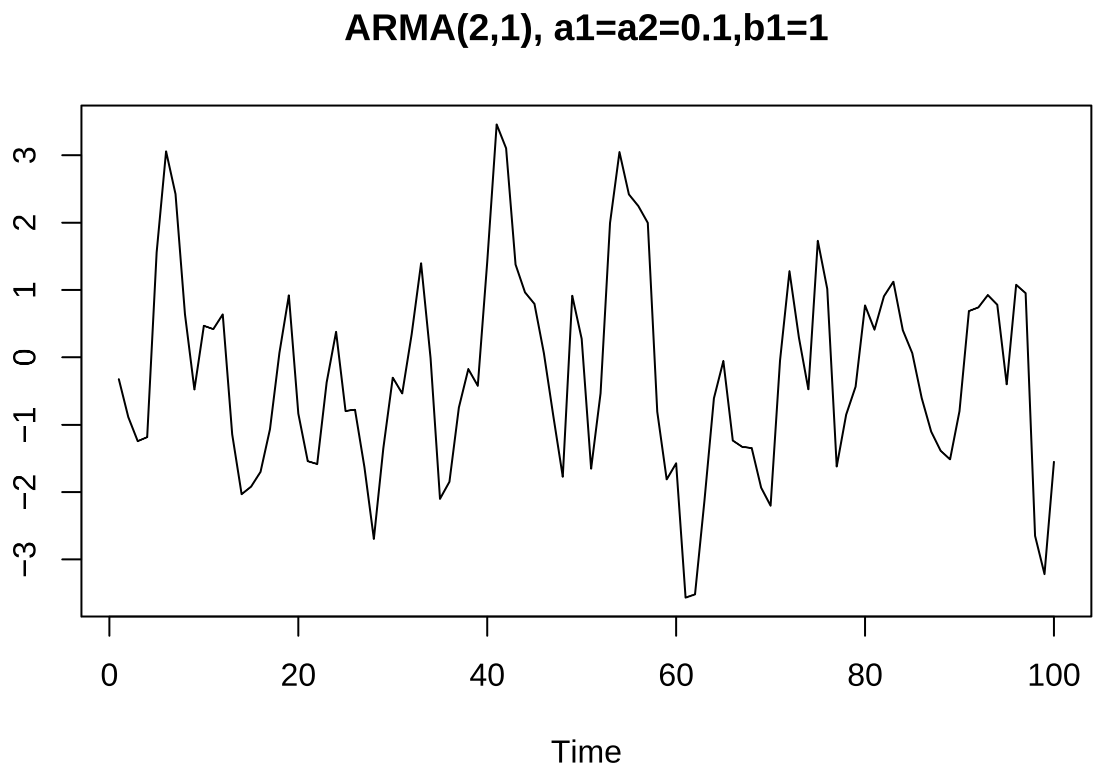
$$\begin{aligned}
|\psi(\lambda)|^2 &= \left|1+\frac{1}{2}\cos\lambda + \mathrm{i}\,\frac{1}{2}\sin\lambda - \frac{1}{2}\cos\lambda + \mathrm{i}\,\frac{1}{2}\sin\lambda\right|^2 \\
&= \frac{3}{2} - \frac{\cos 2\lambda}{2}
\end{aligned}$$

Thus, it follows that $f_X(\lambda) = |\psi(\lambda)|^2 f_Z(\lambda)$.

Question 4

a) Let $n = 100$, $p = 2$, and $q = 1$ with $\alpha_1 = \alpha_2 = 0.1$, $\beta_1 = 1$ then the plot of the time series is

```
arma1=arima.sim(100,model=list(ar=c(0.1,0.1),ma=1))
plot(arma1,ylab="",main="ARMA(2,1), a1=a2=0.1,b1=1")
```



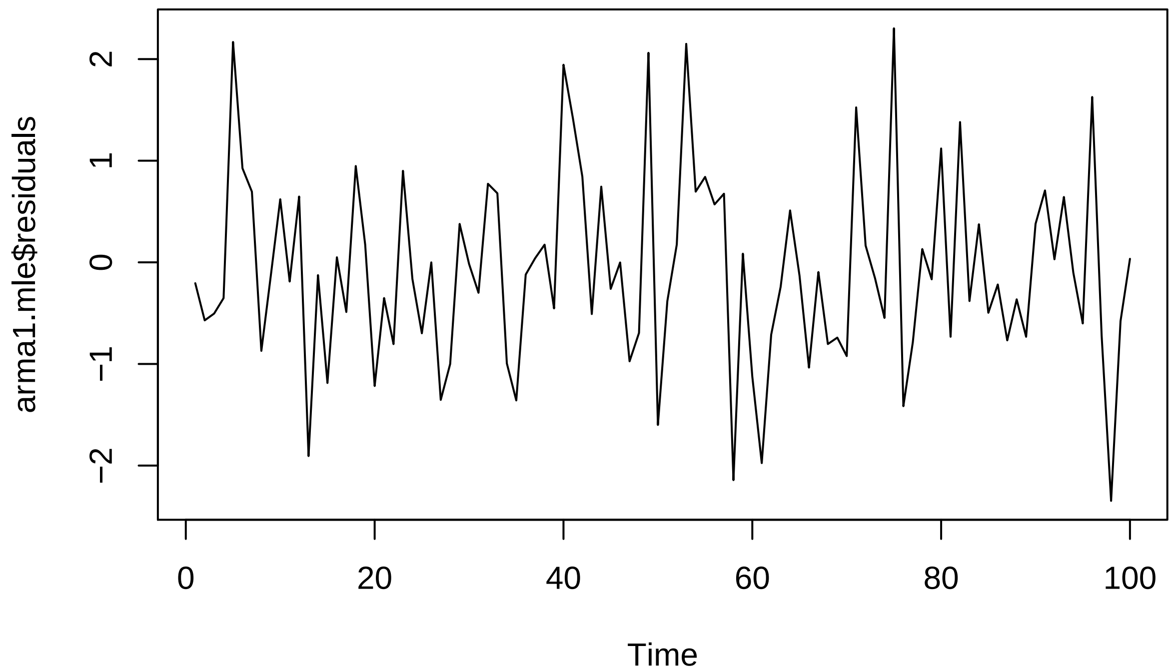
We then find the estimates of the model's parameters.

```
arma1.mle = arima(arma1,order=c(2,0,1),method="ML",include.mean=F);
arma1.mle
```

```
##
## Call:
## arima(x = arma1, order = c(2, 0, 1), include.mean = F, method = "ML")
##
## Coefficients:
##          ar1     ar2     ma1
##       0.2395  0.1117  0.8990
## s.e.  0.1382  0.1347  0.0935
##
## sigma^2 estimated as 0.8818:  log likelihood = -136.6,  aic = 281.2
```

The estimated parameters are $\hat{\alpha_1} = 0.16$, $\hat{\alpha_2} = 0.088$, and $\hat{\beta}_1 = 0.947$ which are pretty different from the true ones. The residuals plot is

```
plot(arma1.mle$residuals)
```



There is no visible pattern to the residual plot which suggests that the estimated coefficients are appropriate. We use the Portmanteau test to evaluate the quality of the fit

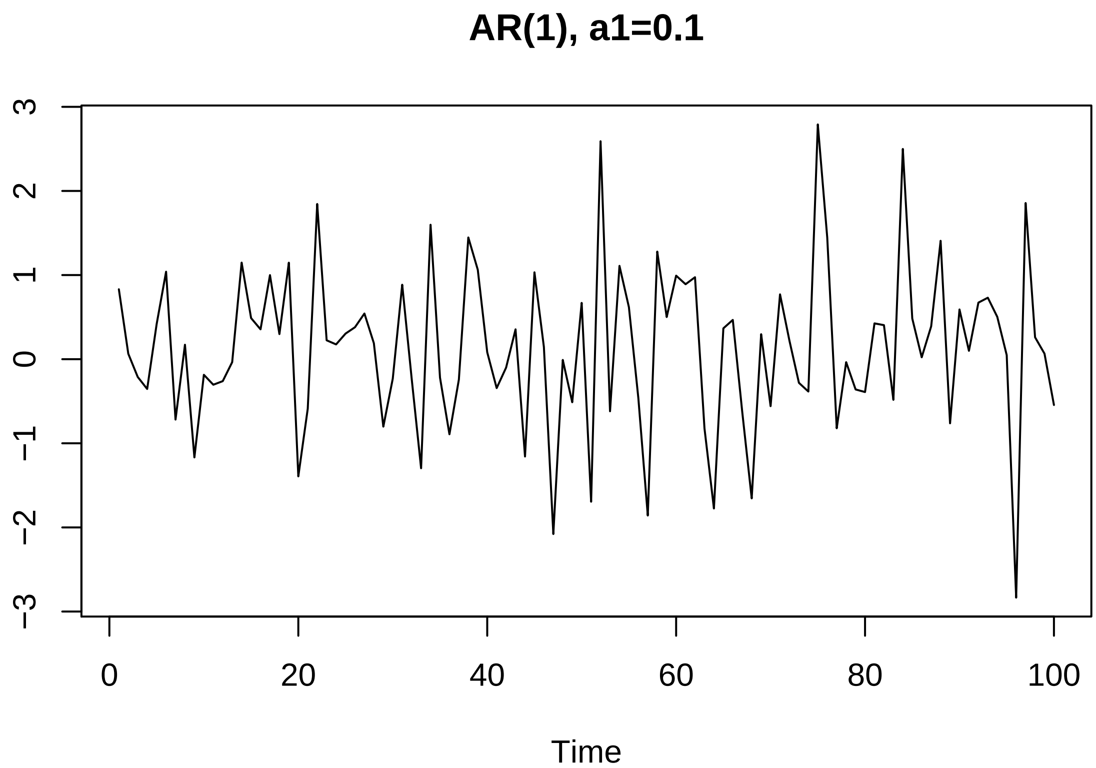
```
Box.test(resid(arma1.mle),type="Box-Pierce")$p.value
```

```
## [1] 0.9021885
```

The returned p-value is 0.792 which suggests an ok fit.

b) We first generate an AR(p) times series of length 100 for $p = 1$ with $\alpha_1 = 0.1$.

```
ar1=arima.sim(100,model=list(ar=0.1))
plot(ar1,ylab="",main="AR(1), a1=0.1")
```



The estimated parameter of the model using Yule-Walker is $\hat{\alpha}_1 = 0.245$. The p-value resulting from the Portmanteau test is 0.7 which is larger than 0.05 signifying that this is a good fit.

```
ar.yw(ar1, order.max=1)
```

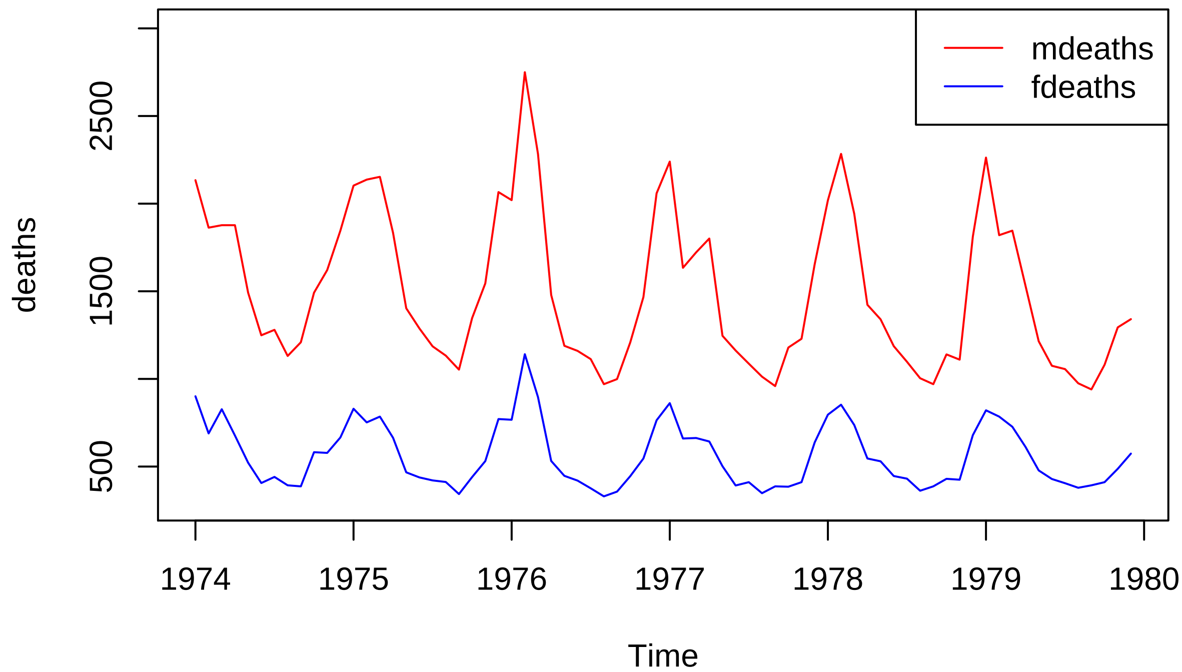
```
##
## Call:
## ar.yw.default(x = ar1, order.max = 1)
##
## Coefficients:
##       1
## -0.1435
##
## Order selected 1  sigma^2 estimated as  0.9292
```

```
Box.test(resid(arima(ar1,order=c(1, 0, 0),method="ML",include.mean=F),type="Box-Pierce"))$p.value
```

```
## [1] 0.6933295
```

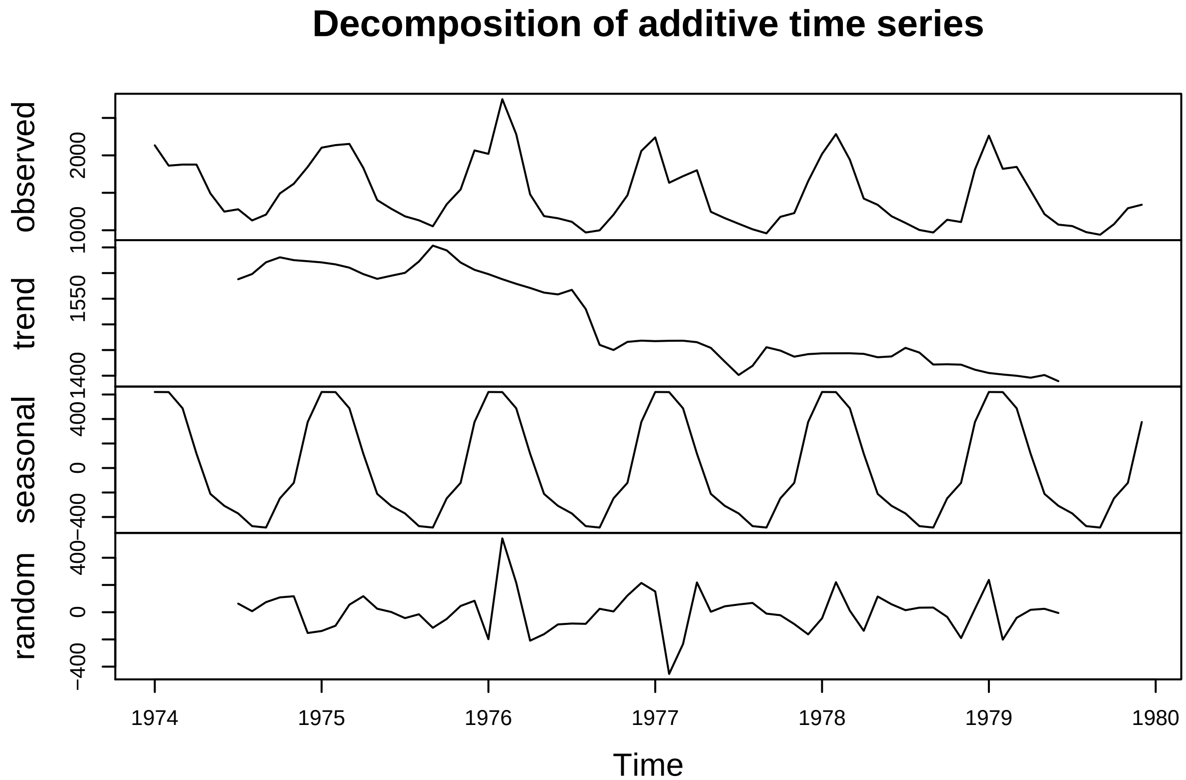
The estimated parameter from Yule-Walker for this model is not as good as the ones from a) because 0.16 and 0.088 are much closer to 0.1 than 0.245. 0.245 is much closer to 0.1 than ### Question 5 #### a) The plot of $mdeaths$ and $fdeaths$ is

```
plot.ts(mdeaths, type='l', col="red", ylim=c(300, 3000), ylab="deaths")
lines(fdeaths, type="l", col="blue")
legend("topright", legend=c("mdeaths", "fdeaths"),
       col=c("red", "blue"),
       lty=1)
```



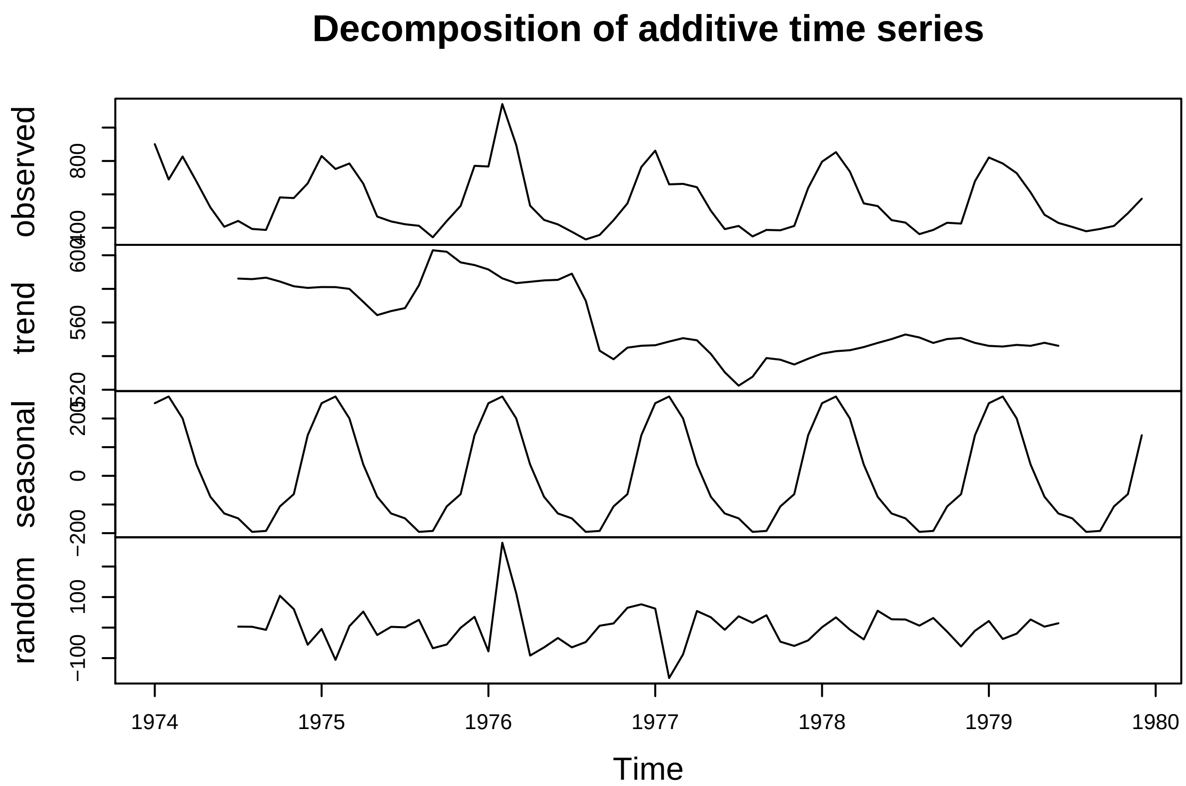
We can also decompose the time series into its components. The decomposition for *mdeaths* is:

```
decompm=decompose(mdeaths);
plot(decompm)
```



And the decomposition of $fdeaths$ is

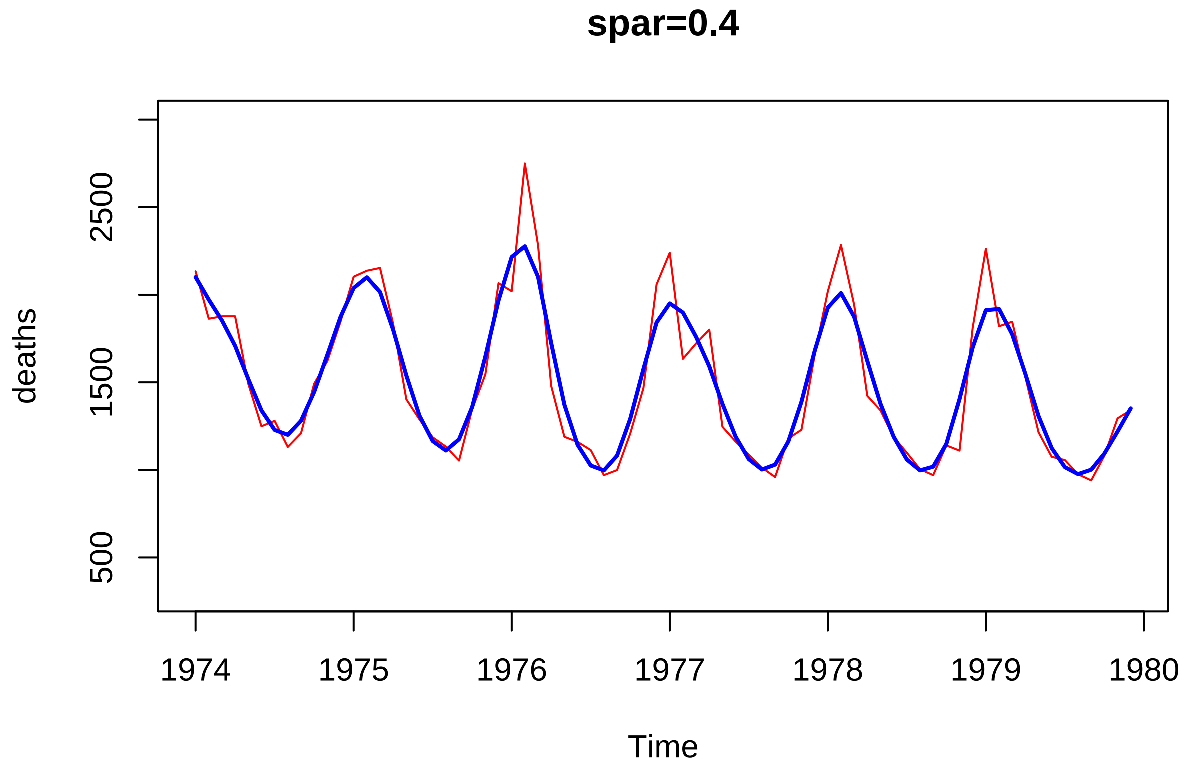
```
decompf=decompose(fdeaths);
plot(decompf)
```



It can be seen that there exists seasonality in the two datasets since the number of deaths of both genders spike every year at the beginning of the year. One obvious difference between the two datasets is that the *mdeaths* is far higher than *fdeaths* (male oppression). Furthermore, for both datasets, there seems to be a faint downward trend of the number of deaths but this trend is stronger in *mdeaths* than *fdeaths*.

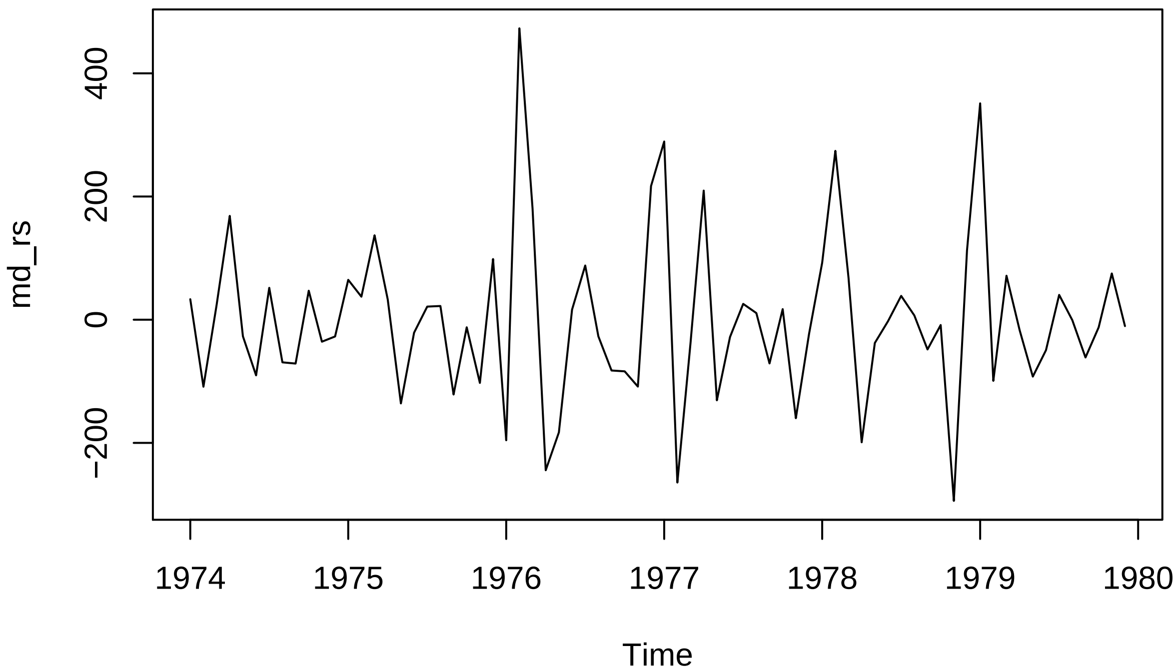
b) For this part, we only consider *mdeaths*, the plot of the time series *mdeaths* with its smoothing spline is

```
plot.ts(mdeaths, type='l', col="red", ylim=c(300, 3000),
        ylab="deaths", main="spar=0.4")
sm = smooth.spline(x=mdeaths, spar=0.4)
lines(sm, col="blue", lwd=2)
```



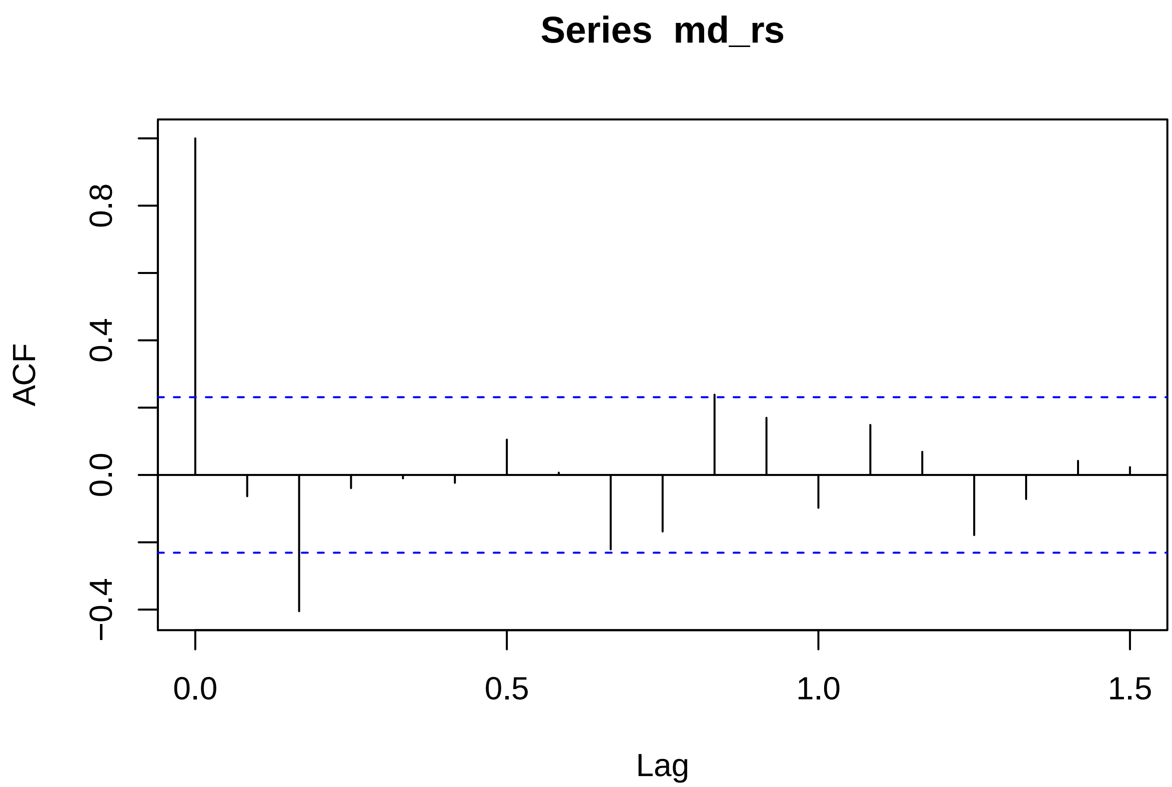
We then subtract the spline from the original time series to obtain a time series with no trend or seasonality.

```
md_rs = mdeaths - sm$y;
plot.ts(md_rs)
```



We then plot an estimate of the autocorrelation function on the de-trended and de-seasonalized time series.

```
acf(md_rs)
```



Since the ACF is mostly small, this confirms the fact that we are dealing with a de-trended and de-seasonalized time series. For $p \in \{0, \ldots, 5\}$ and $q \in \{0, \ldots, 5\}$, we fit different ARMA(p, q) models to the de-trended and de-seasonalized time series and obtain the best one. Different combinations of p and q and their corresponding ARMA(p, q) model's AIC are listed below.

```
for (p in c(1:5)) {
  for (q in c(1:5)) {

    print(c(p, q, arima(md_rs,order=c(p,0,q),method="ML",include.mean=F)$aic))
  }
}
```

```
## [1]   1.0000   1.0000 889.9842
## [1]   1.0000   2.0000 883.3127
## [1]   1.0000   3.0000 885.3632
## [1]   1.0000   4.0000 873.0483
## [1]   1.0000   5.0000 874.8177
## [1]   2.000   1.000 882.351
## [1]   2.0000   2.0000 883.9786
## [1]   2.0000   3.0000 871.3379
## [1]   2.0000   4.0000 856.2432
## [1]   2.0000   5.0000 877.0378
## [1]   3.0000   1.0000 884.3122
## [1]   3.0000   2.0000 885.6045
## [1]   3.0000   3.0000 872.1019
## [1]   3.0000   4.0000 856.4304
## [1]   3.0000   5.0000 860.2427
```

```
## [1]   4.0000   1.0000 882.9752
## [1]   4.0000   2.0000 866.0423
```

```
## Warning in log(s2): NaNs produced
```

```
## [1]   4.0000   3.0000 848.0102
## [1]   4.0000   4.0000 874.8433
## [1]   4.0000   5.0000 842.2854
## [1]   5.0000   1.0000 884.3229
## [1]   5.0000   2.0000 866.2353
## [1]   5.0000   3.0000 869.9696
## [1]   5.0000   4.0000 888.0419
```

```
## Warning in arima(md_rs, order = c(p, 0, q), method = "ML", include.mean = F):
## possible convergence problem: optim gave code = 1
```

```
## [1]   5.0000   5.0000 875.6489
```

It can be seen that the (p, q) value with the minimum AIC score is $(4, 5)$ with an AIC score of 842.

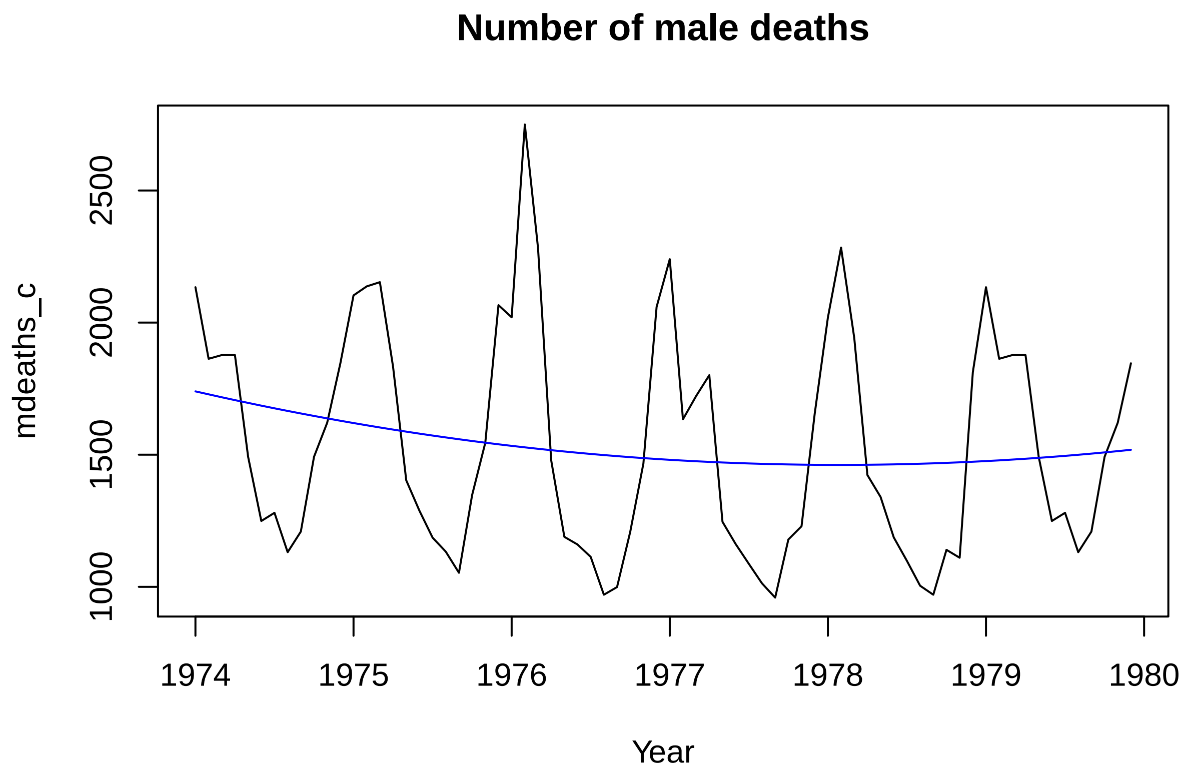
```
mdeaths_c = ts(first(mdeaths, "60 months"), start=c(1974, 1), end=c(1979, 12), frequency=12);
year = index(mdeaths_c);
yearsq=year^2;
```

d) Using the (fractional) year as time, we estimate the trend by a quadratic function of time. The coefficients of the quadratic function are $(-2.33e7, 2.36e4, -6)$. The plot of *mdeaths* up to the last 12 months with its trend as quadratic function shown in blue is below.

```
trend_md = lm(mdeaths_c~year+yearsq);
trend_md$coef
```

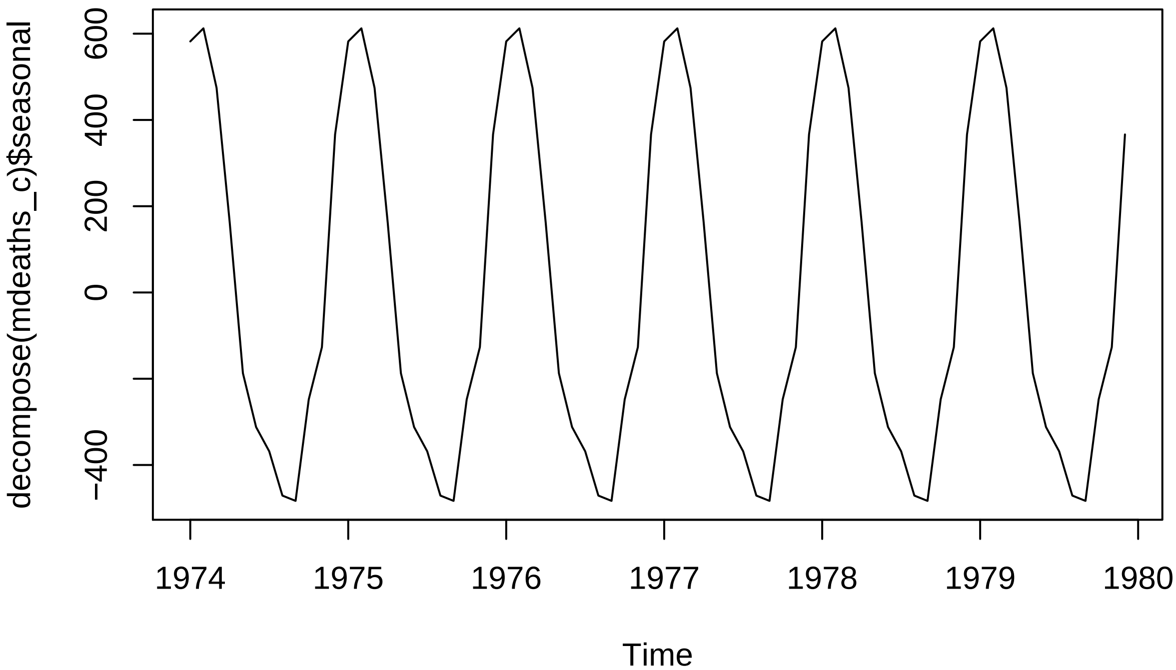
```
##   (Intercept)          year        yearsq
##  6.569602e+07 -6.642283e+04  1.678979e+01
```

```
plot(mdeaths_c, xlab="Year", main="Number of male deaths")
lines(year,trend_md$coef[1]+trend_md$coef[2]*year+trend_md$coef[3]*yearsq,col="blue")
```



The seasonal components can be plotted as

```
plot.ts(decompose(mdeaths_c)$seasonal);
```



Subtracting the trend and seasonality from the data to obtain the random component.

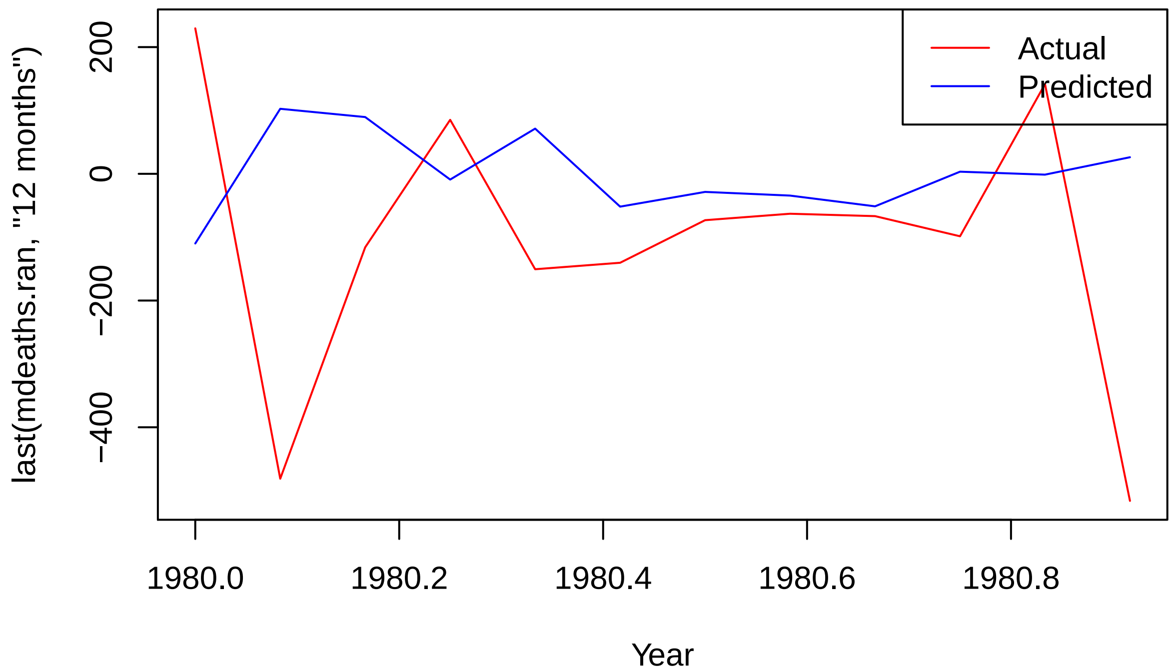
```
mdeaths_c.dtr = mdeaths_c - (trend_md$coef[1]+trend_md$coef[2]*year+trend_md$coef[3]*yearsq);
mdeaths_c.ran = diff(mdeaths_c.dtr, lag=12);
```

We then fit the ARMA(4,5) model from b) to the random component to predict the next 12 months.

```
mdeaths_c.mle = arima(mdeaths_c.ran,order=c(4,0,5),method="ML",include.mean=F);
pred=predict(mdeaths_c.mle,n.ahead=12);
plot(pred$pred, xlab="Year");
```

We compare the predictions against the actual (de-trended and de-seasonalized) data. data of the next 12 months.

```
ya = index(mdeaths);
yaq = ya^2;
mdeaths.dtr = mdeaths - (trend_md$coef[1]+trend_md$coef[2]*ya+trend_md$coef[3]*yaq);
mdeaths.ran = diff(mdeaths.dtr, lag=12);
plot(x=index(pred$pred), y=last(mdeaths.ran, "12 months"), col="red", type="l",
     xlab="Year")
lines(pred$pred, col="blue")
legend("topright", legend=c("Actual", "Predicted"),
       col=c("red", "blue"),
       lty=1)
```



It can be seen that there are times where the differences between the actual and predicted data differ quite a bit. ### Question 6 #### a)

The first step to get use the MACD method is to obtain the data. It is important to note that we are requested MACD and Signal lines to be defined for two specific dates. Hence, we need the data starting from earlier than the desired date. By trial and error we checked that we need our data to start on "2021-11-19".

```
library('TTR')
library(quantmod)
getSymbols("AAPL",from="2021-11-19",to="2022-11-19")
```

```
## [1] "AAPL"
```

```
stock=AAPL
```

Moving on, we now choose our parameters for MACD: $S = 12$, $L = 26$ and $K = 9$. We can also check the numnber of NA values in the column Signal, for this column is the one that will have more NA's. A manual check has sufficed to check that indeed the first day for which both columns are defined is the desired date.

```
macd<-MACD(Cl(stock),nFast=12,nSlow=26,nSig=9,percent=FALSE)
n.na=sum(is.na(macd$signal))
n.na # OK
```

```
## [1] 33
```

b) Before visualizing the data, we first will create a new table, for which we will have no NA values and, more importantly, we only have the data we care about. Afterwards, we will visualize a plot of the stock's price:

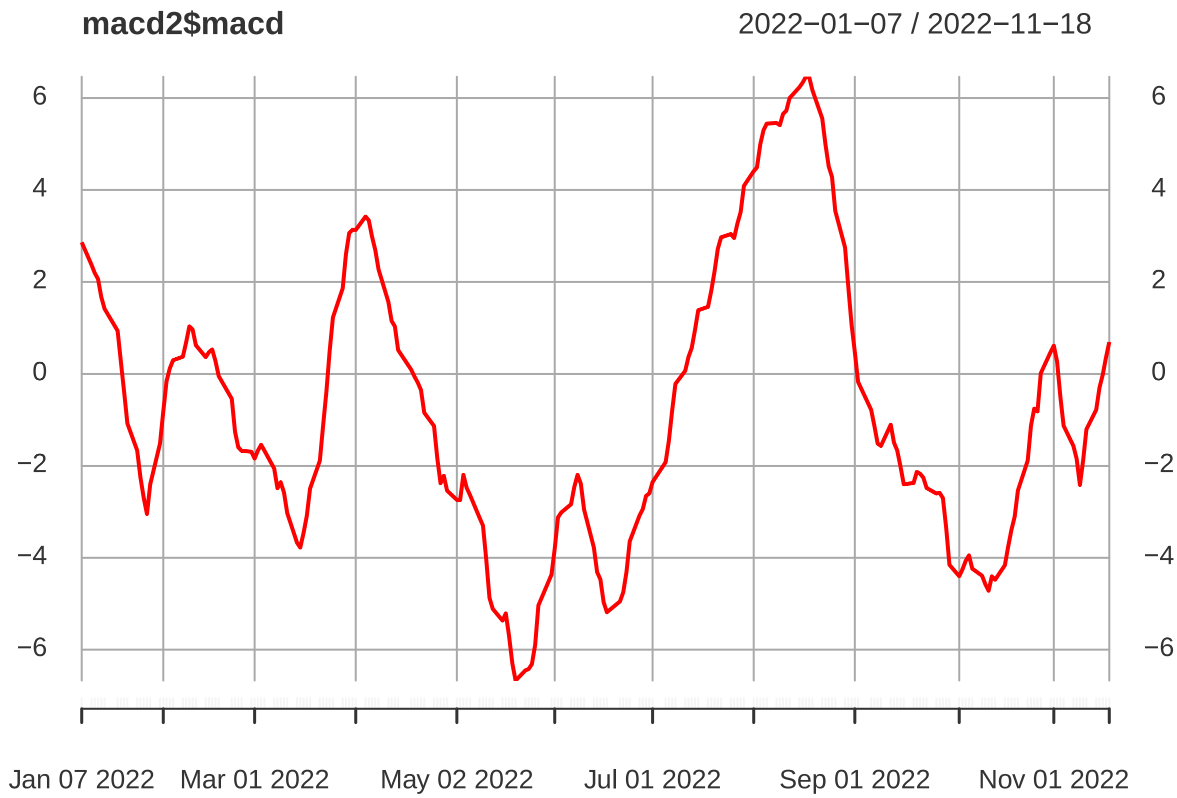
```
# trim
t1=time(macd[n.na+1,]); t1 #=time(macd[nSlow+nSig-1,]) # row 34=26+9-1
```

```
## [1] "2022-01-07"
```

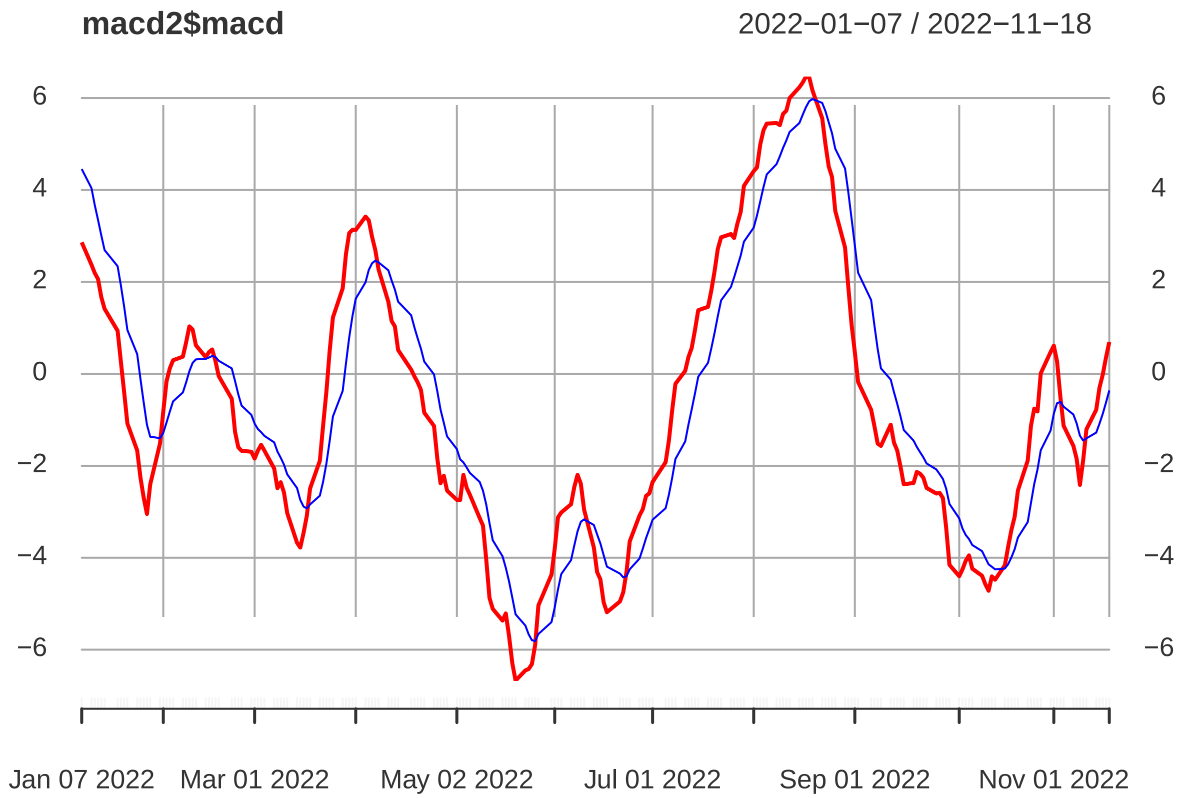
```
macd2=macd[time(macd)>=t1] # macd2 starting from the date t1

plot(Cl(stock))  # plot of the closing prices of the stock
```

Also, we can visualize in a same plot our valiues for MACD (red) and Signal (blue), paying special attention to the points where the two lines cross each other.

```
plot(macd2$macd, col ="red") # plot of MACD line
```




```
lines(macd2$signal,col="blue") # plot of Signal line
```



c) We now have to simulate how much benefit or loss we would have obtained by trading the stock on the crossings between the MACD line and Signal line. We start with an initial capital of 172.17, but note that we do not buy the stock inmediately. In fact, we wait untill the Signal line crosses the MACD line from above to below. After that, we keep on following the strategy and, in the end, we sell the stock in the last day.

By trading on the line crossings, we managed to obtain a benefit of 30.36 and we still have the stock, which is worth 151.29. This adds up to a total final capital of 181.65. Of course, we still need to our initial capital, and we will see we have a final gain of 9.48.

```
price<-as.numeric(Cl(stock[time(stock)>=t1]))
n=length(price)
gain1=price[n]-price[1]; gain1 # gain without using any trading strategy
```

```
## [1] -20.88
```

```
buy<-as.numeric(ifelse(macd2$macd < macd2$signal,1,-1))

gain = price[1] # however, note we still don't buy it
for (i in 1:218) {
  j = i+1
  if (buy[i] == 1 && buy[j] == -1) {
    gain = gain - price[j] #buy stock
  } else if (buy[i] == -1 && buy[j] == 1) {
    gain = gain + price[j] #sell stock
  }
}
gain; price[n]
```

```
## [1] 30.36002

## [1] 151.29
```

```
final_capital = gain + price[n]
final_gain = final_capital - price[1]

final_capital; final_gain
```

```
## [1] 181.65

## [1] 9.48001
```

d) Very easily, we can check that the trivial strategy of buying, holding and selling does not seem to be beneficial for this stock at this particular time. In fact, one would have lost -20.88, which is worse than not investing at all.

```
gain1=price[n]-price[1]; gain1 # gain without using any trading strategy
```

```
## [1] -20.88
```

It is important to remark that here we are not taking into account the fees and comissions that apply to every time we buy or sell an asset. This would make the final gain be even lower and, probably, negative.