Time Series Analysis

Final Exam

Submitted by Sruthy Annie Santhosh, 312213

Problem 1.

Let Z_t be a sequence of independent normal random variables, each with mean 0 and variance $\sigma^2 > 0$.

(1) Which, if any, of the following processes are causal?

We know that from the slides,

Theorem 5.6

An AR(p) process $(X_t)_{t\in\mathbb{Z}}$ is causal if

$$\Phi(x) := 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p$$

has no root inside the complex unit circle. It then has a representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

where $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and the coefficients are defined recursively by $\psi_k = 0$ for $k < 0$, $\psi_0 = 1$ and

$$\psi_j = \sum_{i=1}^{p} \phi_i \psi_{j-i}, \qquad j \geq 1.$$

Prof. Dr. Sebastian Mentemeier Time Series Analysis (5284) 24.05.2022 19

Institut für Mathematik und Angewandte Informatik Universität Hildesheim

A sufficient condition for an AR(p) process to be causal is

$$\sum_{k=1}^{p} |\phi_k| < 1.$$

(a) $X_t = \frac{5}{6} X_{t-1} - \frac{1}{6} X_{t-2} + Z_t - 2 Z_{t-1}$

Writing in polynomial form,

$\Phi(x) = 1 - \frac{5}{6} x + \frac{1}{6} x^2$

The solutions of the polynomial $\Phi(x) = 0$ are x1 = 2.0 + 0i and x2 = 3.0+0i

The roots of the function can be found using the polyroot() function in R.

```
> polyroot(c(1, -5/6, 1/6))
[1] 2+0i 3+0i
>
```

Here the roots are greater than 1 i.e, are outside the complex unit circle. Hence the process {Xt} is causal.

(b) $X_t = 8/3\ X_{t-1} - 4/3\ X_{t-2} + Z_t - 1/2\ Z_{t-1}$

Writing in polynomial form,

$\Phi(x) = 1 - 8/3\ x + 4/3\ x^2$

The solutions of the polynomial $\Phi(x) = 0$ are x1 = 0.5 - 0i and x2 = 1.5 + 0i

The roots of the function can be found using the polyroot() function in R.

```
> polyroot(c(1,-8/3,4/3))
[1] 0.5-0i 1.5+0i
>
```

Here the root x1 is less than 1 i.e, is inside the complex unit circle. Hence the process {Xt} is not causal.

(2) . Which, if any, of the following processes are invertible?

Recall that an ARMA process,

$\Phi(B)\ X_t = \Theta(B)\ Z_t$ is called invertible,

if Θ has no root inside the complex unit circle.

(a)$X_t = \frac{5}{6}\ X_{t-1} - \frac{1}{6}\ X_{t-2} + Z_t - 2\ Z_{t-1}$

Writing in polynomial form,

$\Theta(x) = 1 - 2\ x$

The solution of the polynomial $\Theta(x) = 0$ is x = 0.5 +0i

The root of the function can be found using the polyroot() function in R.

```
> polyroot(c(1,-2))
[1] 0.5+0i
>
```

Here the root is less than 1, i.e., it is inside the complex unit circle. Hence the process $\{X_t\}$ is not invertible.

(b) $X_t = 8/3\ X_{t-1} - 4/3\ X_{t-2} + Z_t - 1/2\ Z_{t-1}$

Writing in polynomial form,

$\Theta(x) = 1 - 1/2\ x$

The solution of the polynomial $\Theta(x) = 0$ is $x = 2 + 0i$

The root of the function can be found using the polyroot() function in R.

```
> polyroot(c(1,-1/2))
[1] 2+0i
>
```

Here the root is greater than 1, i.e., it is outside the complex unit circle. Hence the process $\{X_t\}$ is invertible.

(3) Identify each of the following as an ARIMA model by giving the values of (p, d, q).

We know that from the slides, an ARIMA(p,d,q) process has the form

$$\Phi(B)(1-B)^d X_t = \Theta(B) Z_t,$$

where $\Phi(B)$ and $\Theta(B)$ are polynomials of orders p and q, respectively.

(a) $X_t = \frac{5}{6} X_{t-1} - \frac{1}{6} X_{t-2} + Z_t - \frac{1}{4} Z_{t-2}$

Considering the polynomial,

$\Phi(x) = 1 - \frac{5}{6} x + \frac{1}{6} x^2$

The solutions of the polynomial $\Phi(x) = 0$ are $x1 = 2+0i$, $x2 = 3+0i$

The roots of the function can be found using the polyroot() function in R.

```
> polyroot(c(1,-5/6,1/6))
[1] 2+0i 3+0i
>
```

Here both the roots are greater than 1, ie, they are outside the complex unit circle. And no root is equal to 1.

Hence Φ(B) is the polynomial of order p = 2, and d = 0.

Now considering the polynomial , $\Theta(x) = 1 - \frac{1}{4} x^2$. The solutions of the polynomial are x1 = 2+0i , x2 = -2 + 0i . Here both the roots are outside the complex unit circle.

```
> polyroot(c(1,0,-1/4))
[1]  2+0i -2+0i
>
```

Hence Θ(B) is the polynomial (moving average part) of order q = 2.

Thus the process $\{X_t\}$ is an ARIMA(2,0,2) model.

(b) $X_t = \frac{3}{4} X_{t-2} + \frac{1}{4} X_{t-3} + Z_t - Z_{t-1} + \frac{1}{4} Z_{t-2}$

Considering the polynomial,

$$\Phi(x) = 1 - \frac{3}{4} x^2 - \frac{1}{4} x^3$$

The solutions of the polynomial Φ(x) = 0 are x1 = 1-0i , x2 = -2-0i , x3 = -2+0i

The roots of the function can be found using the polyroot() function in R.

```
> polyroot(c(1,0,-3/4,-1/4))
[1]  1-0i -2-0i -2+0i
>
```

Here root x1 is on the unit complex unit circle while roots x2 and x3 are outside the unit circle as x1 is equal to 1 and the absolute values of x2 and x3 are greater than 1.

Hence Φ(B) is the polynomial of order p = 2, and d = 1.

Now considering the polynomial , $\Theta(x) = 1 - 1x + \frac{1}{4} x^2$. The solutions of the polynomial are x1 = 2+0i , x2 = 2 + 0i . Here both the roots are outside the complex unit circle.

```
> polyroot(c(1,-1,1/4))
[1] 2+0i 2-0i
>
```

Hence Θ(B) is the polynomial (moving average part) of order q = 2.

Thus the process $\{X_t\}$ is an ARIMA(2,1,2) model.

Problem 2.

Below, you find the realizations, the empirical ACF and PACF of four time series. Based on these plots, decide what model should be fitted: AR, MA, ARIMA(p, 1, q) or FARIMA(0, d, 0) (with 0 < d < 1 2). Explain your answer.

Process1 :

From the graphs it looks like Process 1 is non- stationary, as the decay is gradual in ACF and PACF. The autocorrelations are positive for many numbers of lags, till 15, hence differencing is needed (d=1). Thus **ARIMA(p,1,q)** model is best fitting.

Process2 :

From the graphs it looks like Process 2 is stationary as the tail cuts off after 2 in ACF . All the rest of the lags are not significant (within the confidence interval) .Even though some spikes are on the blue line, we require the simplified model. There is a tail off at PACF (decreases gradually). Hence the best fitted model is **MA(2)**.

Process3 :

From the graphs we can observe that Tail off is observed in the ACF plot (gradual decrease - geometric trend) . Hence it is an AR model. From PACF we can see that cut off happens at lag 2. Thus the order is 2. Also the process looks to be stationary as the decrease is exponential in ACF. So the best fitted model is **AR(2).**

Process4 :

From the graphs we can observe that the process is not stationary as the decay is gradual in ACF. The lags are significant till 30 and have heavy tails. Also there is a positive correlation for all lags in ACF. From PACF, partial correlation is almost 0 after 4 lags. Hence the best fitted model is **FARIMA(0,d,0)** ($0<d<½$)

Problem 3

Historical daily exchange rate EUR/ USD (1 EUR equals X US Dollar) from 2003- 12-01 to 2022-07-04 is given in the data file named "EURUSD=X.csv" (source: yahoo.com). The latest closing exchange rate EUR/ USD as of 2022-07-04 is 1.043623.

1. Create a time series containing the "Close" prices, and then plot it.

Load the required libraries and set the correct path. Then we read the given file using read.table() and then create a time series object with the correct start, end and frequency. Then the time series object is plotted.

2. Test for non-stationarity

I am doing three tests for non-stationarity. They are

a) Augmented Dickey-Fuller Test

```
> adf.test(euro.ts)

        Augmented Dickey-Fuller Test

data:  euro.ts
Dickey-Fuller = -2.5295, Lag order = 5, p-value = 0.3564
alternative hypothesis: stationary
```

The null hypothesis of the ADF test is that a unit root is present in the time series. The alternative hypothesis is that the data is stationary.

If this p-value is smaller than 0.05 you can reject the null hypothesis (reject non-stationarity) and accept the alternative hypothesis (stationarity). In this case, since p-value is greater than 0.05, we accept the null hypothesis and hence the data is non-stationary.

b) Phillips-Perron Unit Root Test

```
> pp.test(euro.ts)

	Phillips-Perron Unit Root Test

data:  euro.ts
Dickey-Fuller Z(alpha) = -12.925, Truncation lag parameter = 4, p-value = 0.3696
alternative hypothesis: stationary
```

This test is similar to the ADF test, it is also a Unit root test. Here also the p-value is greater than 0.05, hence we accept the null hypothesis that the data is non-stationary

c) Kwiatkowski-Phillips-Schmidt-Shin(KPSS) Test for Level Stationarity

```
> kpss.test(euro.ts)

	KPSS Test for Level Stationarity

data:  euro.ts
KPSS Level = 1.5033, Truncation lag parameter = 4, p-value = 0.01

Warning message:
In kpss.test(euro.ts) : p-value smaller than printed p-value
```

The null and alternative hypothesis for the KPSS test are opposite that of the ADF test. The Null Hypothesis is that the process is stationary while the alternative hypothesis is that the series has unit root (series is not stationary). Here we see that the p-value is less than 0.05, hence we reject the null hypothesis and accept the alternative hypothesis. Thus the series is found to be non-stationary.

Hence by all three tests, the time series is found to be non-stationary.

3. Fit an ARIMA(1,1,0)-model. Explain the meaning of this model (what would be an interpretation of the p = 1 and d = 1?).

ARIMA(1,1,0) is fitted to the time series object. The AIC is also found

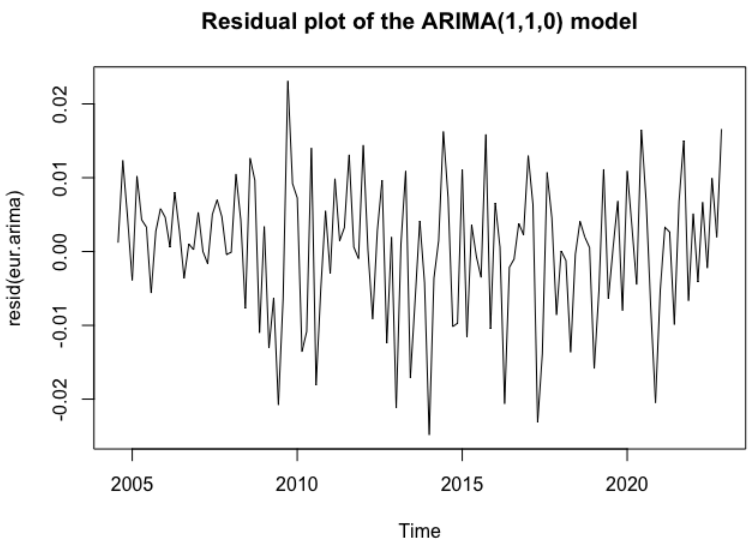
```
> eur.arima

Call:
arima(x = euro.ts, order = c(1, 1, 0))

Coefficients:
          ar1
      -0.0943
s.e.   0.0893

sigma^2 estimated as 8.821e-05:  log likelihood = 415.86,  aic = -827.73
> AIC(eur.arima)
[1] -827.7287
```

From the tests of non-stationarity, we had found the data to be non-stationary. We can convert the data to stationary data by differencing method. The 'd' in ARIMA model identifies the degree of differencing used to transform the data. The value 'p' in ARIMA model refers to the order of autoregression. This indicates the number of immediately preceding values in the series which are used for predicting the value for current time.

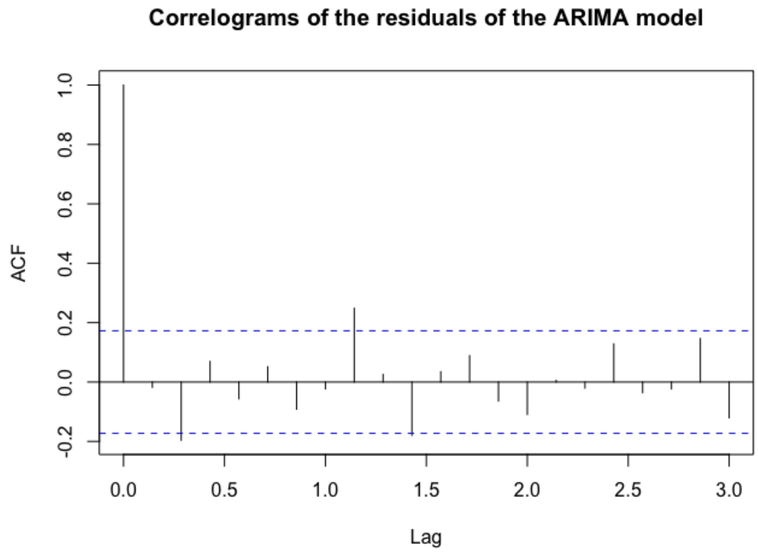
4. Plot the residuals after fitting the ARIMA(1,1,0) model.

We find the residuals of the ARIMA(1,1,0) model.

Plot of residuals after fitting ARIMA(1,1,0) model



Plot of acf of the residuals after fitting ARIMA(1,1,0) model



5. Do the residuals look like white noise? Explain your opinion.

Yes the residuals look like white noise. We can see that the correlogram of the residuals doesn't show any significant autocorrelations. Only the autocorrelation at lag 0 is high, rest all are more close to 0(within the significant bands) . Hence we can say that the residuals can be white noise i.e, that is, the series is independent and identically distributed with mean zero and constant variance.

```
> Box.test(resid.euro, type = "Ljung-Box")

	Box-Ljung test

data:  resid.euro
X-squared = 0.041716, df = 1, p-value = 0.8382
```

We know that , The Ljung-Box test uses the hypothesis,

H0 -> residuals are independently distributed (white noise), HA -> residuals are not independently distributed (serial correlation).

We see that the p-value is greater than 0.05, which means that the residuals are independent (white noise).

We see the same observation for Box-Pierce test also

```
> Box.test(resid.euro, type = "Box-Pierce")

        Box-Pierce test

data:  resid.euro
X-squared = 0.040761, df = 1, p-value = 0.84
```

Hence from the Box tests on the residuals also we can say that the residuals look like white noise.

6. Fit an GARCH(1,1) model to the residuals of the ARIMA(1,1,0) model.

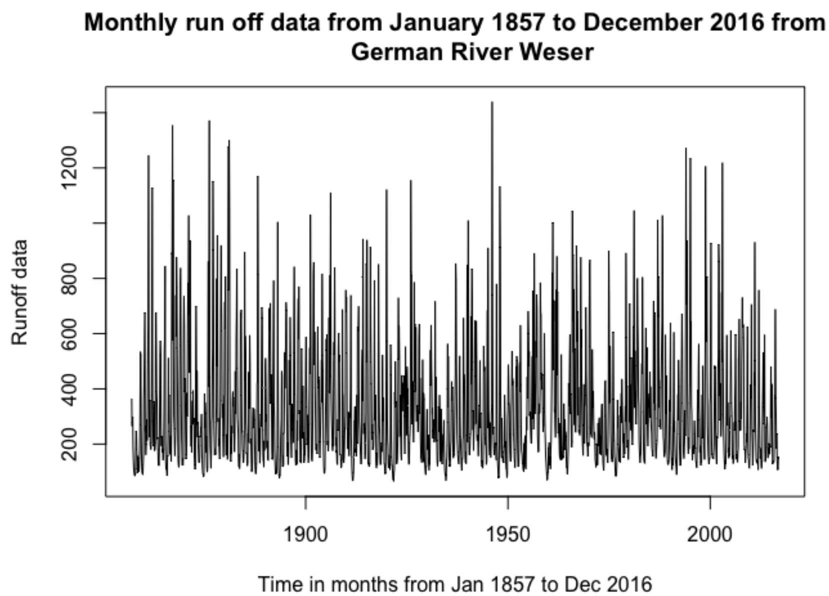
GARCH(1,1) is fitted to the residuals of the ARIMA(1,1,0) model, then summary of GARCH(1,1) model and AIC of the GARCH(1,1) model is found

```
> euro.garch

Call:
garch(x = resid.euro, order = c(1, 1), trace = F)

Coefficient(s):
       a0         a1         b1  
8.582e-05  1.036e-07  2.702e-02  

> summary(euro.garch)

Call:
garch(x = resid.euro, order = c(1, 1), trace = F)

Model:
GARCH(1,1)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.6440 -0.5731  0.1333  0.7014  2.4580 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 8.582e-05          NA       NA       NA
a1 1.036e-07          NA       NA       NA
b1 2.702e-02          NA       NA       NA

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 4.1143, df = 2, p-value = 0.1278


        Box-Ljung test

data:  Squared.Residuals
X-squared = 1.5655, df = 1, p-value = 0.2109

> AIC(euro.garch)
[1] -825.7375
```

Problem 4

The statistic of the runoff data from the German river Weser, measured at Intschede, is given in the data file named "runoff.txt". This gives the monthly data from January 1857 to December 2016. Runoff = amount of water passing by the station. Source: "The Global Runoff Data Centre, 56068 Koblenz, Germany".

1. Load the data into R, store the column named "Calculated" as a time series named "Weser" with the correct frequency. Plot the data.



2. Choose a best-fitting seasonal ARIMA model from the following: ARIMA(1, 1, 0)(1, 1, 0)$_{12}$, ARIMA(1, 1, 0)(0, 1, 1)$_{12}$, ARIMA(0, 1, 1)(1, 1, 0)$_{12}$, ARIMA(1, 0, 0)(1, 0, 1)$_{12}$, ARIMA(1, 1, 1)(1, 1, 0)$_{12}$, ARIMA(1, 0, 1)(0, 1, 0)$_{12}$. Base your choice on the AIC

To choose the best-fitting seasonal ARIMA model, we need to calculate the Akaike Information Criterion (AIC) for all the given ARIMA models. The model with the least AIC value is taken as the best fitting model.

```
> AIC(arima(Weser, order = c(1,1,0), seas = list(order = c(1,1,0), 12)))
[1] 25490.32
> AIC(arima(Weser, order = c(1,1,0), seas = list(order = c(0,1,1), 12)))
[1] 24855.32
> AIC(arima(Weser, order = c(0,1,1), seas = list(order = c(1,1,0), 12)))
[1] 25287.3
> AIC(arima(Weser, order = c(1,0,0), seas = list(order = c(1,0,1), 12)))
[1] 24673.06
> AIC(arima(Weser, order = c(1,1,1), seas = list(order = c(1,1,0), 12)))
[1] 25170.83
> AIC(arima(Weser, order = c(1,0,1), seas = list(order = c(0,1,0), 12)))
[1] 25576.06
>
```

The model with the minimized AIC is the best fitting model, and from the results above, we can see that the model $ARIMA(1,0,1)(1,0,1)_{12}$ has the lowest AIC value, which is 24673.06

3. Write down the best-fitting model.

```
arima(x = Weser, order = c(1, 0, 0), seasonal = list(order = c(1, 0, 1), 12))

Coefficients:
         ar1    sar1     sma1  intercept
      0.4728  0.9997  -0.9853   320.5402
s.e.  0.0202  0.0004   0.0084    47.7995

sigma^2 estimated as 21843:  log likelihood = -12331.53,  aic = 24673.06
>
```

4. Predict the value for January 2017 using R.

The predicted value for January 2017 is 380.4432 using the predict().

```
> pred = predict(arima.best, n.ahead = 1)
> pred
$pred
          Jan
2017 380.4432

$se
          Jan
2017 147.8111
```