Time Series Analysis

Homework Assignment 3

Problem 1

a) Prove causality.

$$
\begin{aligned}
X_t &= -0.4X_{t-1} + 0.45X_{t-2} + Z_t \\
Z_t &= X_t + 0.4X_{t-1} - 0.45X_{t-1} \\
\Phi(x) &= 1 + 0.4x - 0.45x^2 \\
0 &= 1 + 0.4x - 0.45x^2
\end{aligned}
$$

Can calculate the value of *x* using the quadratic equation;

$$
\begin{aligned}
x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-0.4 \pm \sqrt{0.4^2 - (4 \cdot -0.45 \cdot 1)}}{2 \cdot -0.45} \\
x &= \frac{-0.5 \pm 1.4}{-0.9} \\
x &= -\frac{10}{9} \quad or \quad x = 2
\end{aligned}
$$

$$
\begin{aligned}
x_1 &= -\frac{10}{9} \\
x_2 &= 2
\end{aligned}
$$

Therefore we can conclude that the time series X_t is causal since $x_1 \, and \, x_2$ are not in a complex unit circle in that $|x_1|, |x_2| > 0$.

b) If causal, calculate the first four coefficients ψ0, ψ1, ψ2, ψ3.

From the script we know ; $\Phi(x) \;=\; 1 - \phi_1 x \;-\; \phi_2 x^2 \;-\; \ldots \;-\; \phi_p x^p$

$\Phi(x) \;=\; 1 + 0.4x \;-\; 0.45x^2$
$\phi_1 \;=\; -\,0.4$, $\phi_2 \;=\; 0.45$

$\psi_0 \;= 0$

$$\psi_1 \;=\; \sum_{i=1}^{p} \phi_i \psi_{1-i} \;=\; \sum_{i=1}^{2} \phi_i \psi_{1-i} \;=\; \phi_1 \psi_0 \;+\; \phi_2 \psi_{-1}$$
$$=\; (-0.4 \cdot 1) \;+ (0.45 \cdot 0) \;=\; -\,0.4$$

$$\psi_2 \;=\; \sum_{i=1}^{p} \phi_i \psi_{1-i} \;=\; \sum_{i=1}^{2} \phi_i \psi_{2-i} \;=\; \phi_1 \psi_1 \;+\; \phi_2 \psi_0$$
$$=\; (-0.4 \cdot -0.4) \;+\; (0.45 \cdot 1) \;=\; 0.61$$

$$\psi_3 \;=\; \sum_{i=1}^{p} \phi_i \psi_{1-i} \;=\; \sum_{i=1}^{2} \phi_i \psi_{3-i} \;=\; \phi_1 \psi_2 \;+\; \phi_2 \psi_1$$
$$=\; (-0.4 \cdot 0.61) \;+\; (0.45 \cdot -0.4) \;=\; -\,0.424$$

c)

d) The simulated time series is shown in Figure 1;

```
> library(randtests)
> # EXERCISE 1
>
> ## Part D
> set.seed(7)
> X <- rnorm(500)  #initialize X_t
> Z <- rnorm(500, mean = 0, sd = sqrt(1)) # initialize Z_t
>
> #simulate X_t
> for (t in 3:500) X[t] <- (-0.4 *X[t-1]) + (0.45 * X[t-2]) +Z[t]
>
> Xts <- ts(X) #change to time series
>
> plot(Xts,ylab = "x_t", main = "Time Series x_t") #plot data
>
```

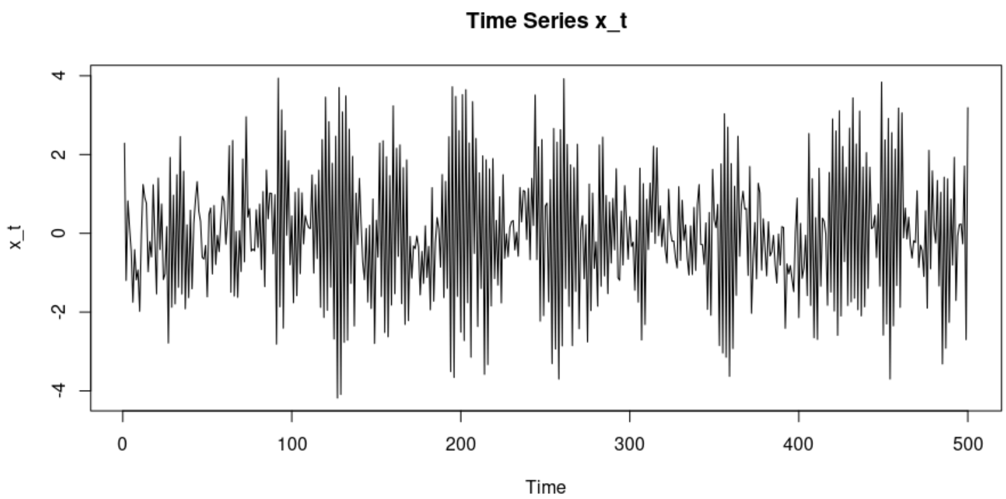

Figure 1

e) Figure 2 contains the corresponding autocorrelation function and Figure 3 shows the partial autocorrelation function.

```
## Part E
acf(Xts, main = "ACF for Time Series (x_t)")

pacf(Xts,  main = "PACF for Time Series (x_t)")
```

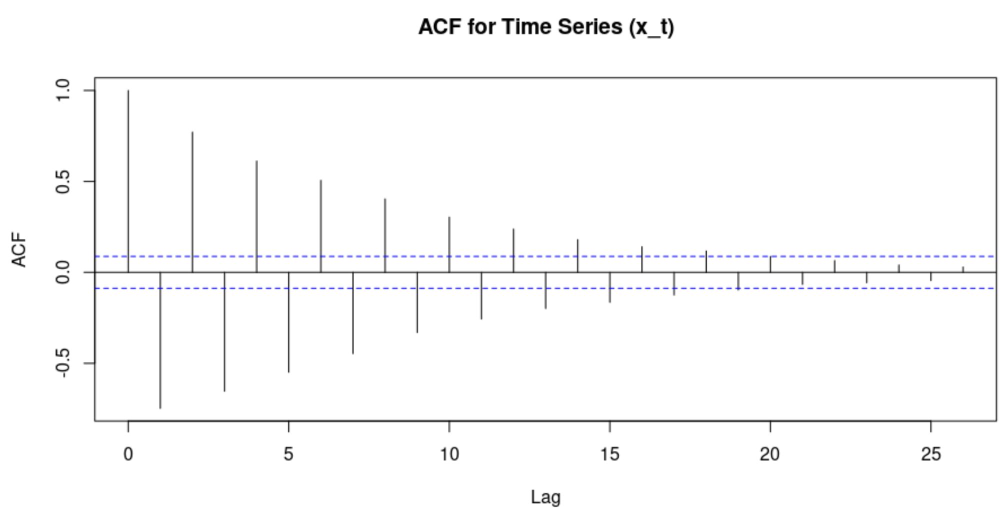

Figure 2

We can observe in the autocorrelation function in Figure 2, that there is a significant correlation when lag < 18, meaning that the correlation is diminishing with the increase in lag.

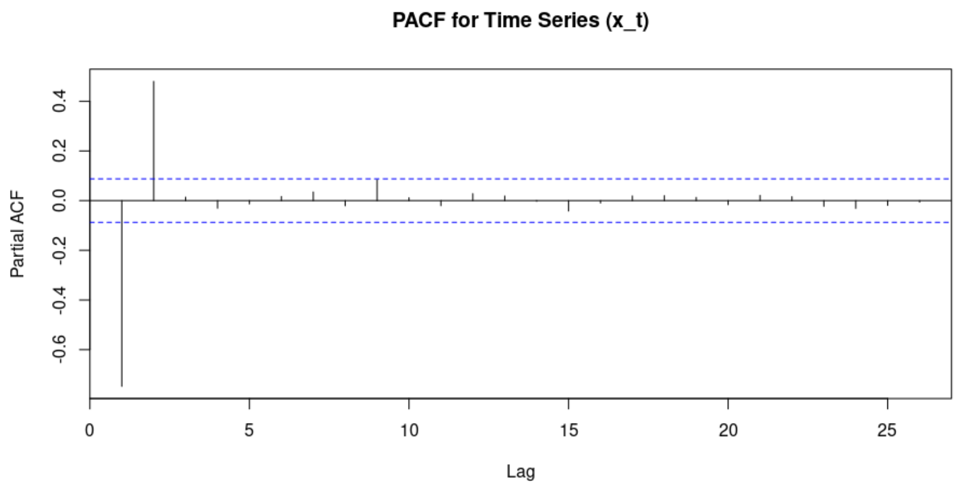

Figure 3

In the partial autocorrelation function, we can observe that we have a significant correlation for lag = 2, and the correlation diminishes with the increase in lag.

f) From the PACF graph above, the autoregressive process is of order 2. This is because, after lag = 2, the correlation diminishes and they are all within the confidence bounds.

Problem 2

a) The residual series X_t , ACF and PACF are shown in Figure 4, Figure 5 and Figure 6 respectively.

```
## Part A
lakelevelRes <- read.table(file = "lakelevelRes.dat", header = TRUE)

#View(lakelevelRes)
attach(lakelevelRes)

residualsTs <- ts(lakelevelRes, start = c(1875), end = c(1972),frequency = 1)

length(residualsTs)

#plot the time series
plot(residualsTs, ylab = "Residuals" , main = "Residuals Time Series for Lake Huron Water
     Level (1875-1972)")

#acf
acf(residualsTs, na.action = na.omit, main = "ACF for Lake Huron Water Level Residuals")

#pacf
pacf(residualsTs, na.action = na.omit, main = "PACF Lake Huron Water Level Residuals")
```

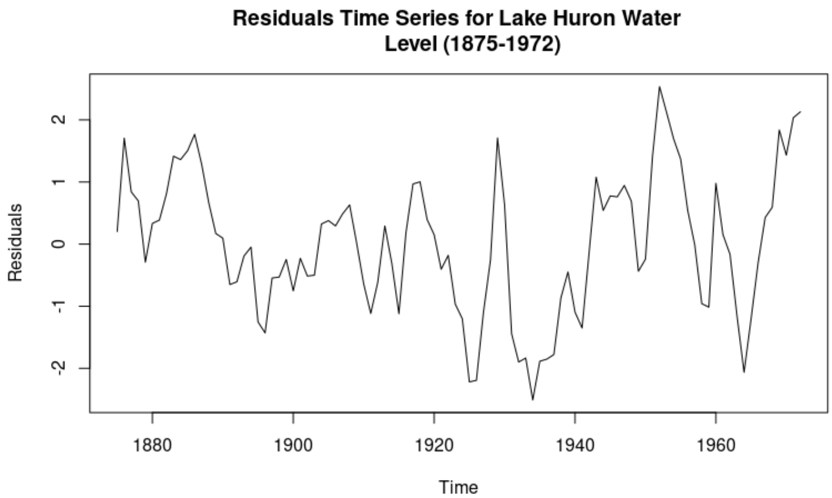

Figure 4

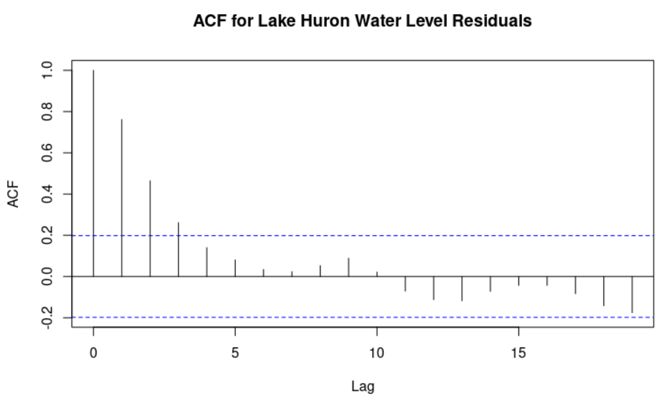

Figure 5

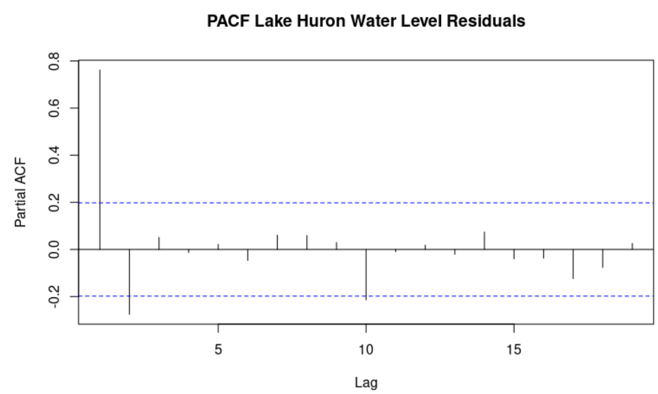

Figure 6

We can observe that in the ACF, there is a significant correlation for lag < 4, in that the correlation lies outside the confidence bounds. In the PACF figure, we can see that we can see a significant correlation when the lag is 1 and 2.

b) From the results, I will choose an autoregressive model of order 2. We can determine from the PACF results that lag 2 has a significant correlation and all the lags after fall within the confidence bounds.

c) Order and coefficients of the model;

```
> ## Part C
> arModel <- ar(residualsTs)
>
> arModel$order  # order
[1] 2
>
> arModel$ar #coefficients
[1]  0.9713674 -0.2754360
>
```

The model results to order 2 and the coefficients 0.9714 and -0.2754.

d) In order to investigate if the resulting residual (rt) comes from an i.i.d. , we will need to first extract the residuals from the autoregressive model as shown below;

```
> arResiduals <- ts(ar(residualsTs)$resid)
>
>
> length(arResiduals)
[1] 98
>
```

We will then use three tests; Ljung-Box test, turning point test and difference sign test. We form our hypothesis for the different tests as follows;

Ljung-Box test

Setting; Given an observation $x_1, \ldots, x_N$ for a given time series X_N where ACF ρ is true but unknown and N< 100, where L is the lag and N is the length of the time series .

Test Problem;

$$H_0 : \rho(1) = \ldots = p(L) = 0$$
$$H_1 : \rho(1), \ldots, p(L) \neq 0$$

Turning Point Test:

Given observation $r_1, ..., r_N$;

Test problem ;

H_0 : $r_{1,} ..., r_n$ are iid

H_1 : $r_{1,} ..., r_n$ are not iid

Difference Sign Test:

Given observation $r_1, ..., r_N$;

Test problem ;

H_0 : $r_{1,} ..., r_n$ are iid

H_1 : $r_{1,} ..., r_n$ are not iid

The results of the result above are shown in the code snippet below;

```
> #Box-Pierce Test
> Box.test(na.omit(arResiduals), lag = min(20, length(arResiduals)), type = "Ljung-Box")

	Box-Ljung test

data:  na.omit(arResiduals)
X-squared = 9.1111, df = 20, p-value = 0.9816

>
> turning.point.test(na.omit(arResiduals))

	Turning Point Test

data:  na.omit(arResiduals)
statistic = 0.57022, n = 96, p-value = 0.5685
alternative hypothesis: non randomness

>
> #difference sign test
> difference.sign.test(na.omit(arResiduals))

	Difference Sign Test

data:  na.omit(arResiduals)
statistic = -0.52759, n = 96, p-value = 0.5978
alternative hypothesis: nonrandomness

>
```

We can make the following conclusion for the tests while taking α = 0.05 is as below;

- Ljung-Box; p-value = 0.9816, given that the p-value > α, we do not have enough evidence to reject the hypothesis.
- Turning Point test; p-value = 0.5685, given that the p-value > α, we do not have enough evidence to reject the hypothesis.
- Difference sign test; p-value = 0.5978, given that the p-value > α, we do not have enough evidence to reject the hypothesis.

In general, since all the three tests show that there is not enough evidence to reject the hypothesis, we can conclude that the time series r_t is iid.

e) To test for Gaussianity, we can use the Gaussian qq-plot for visualization and the Shapiro-Wilk test. Figure 7 shows the qq-plot.

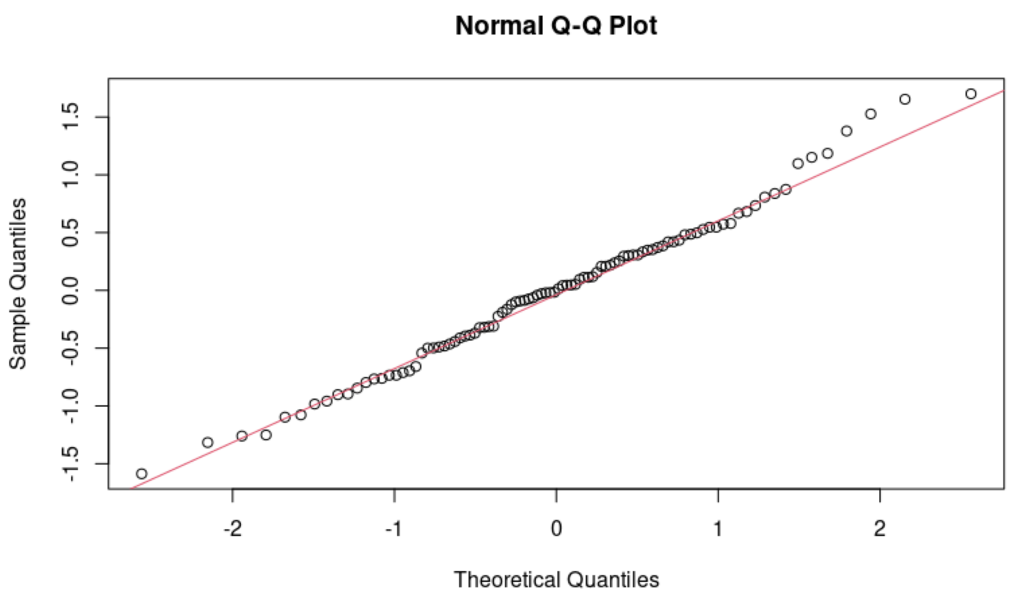

Figure 7

From the qq-plot above, we can observe that the fitted line fits our time series data, and can conclude that it has a Gaussian distribution. We can use the Shapiro-Wilk test to further confirm stated as;

Setting : Given an iid random variables $R_1, \ldots R_N$ given that continuous distribution is unknown and n < 5000;

Test Problems;

H_0 : R_1 has a gaussian distribution

H_1 : R_1 does not have a gaussian distribution

The result of the Shapiro -Wilk test is shown in the snippet below;

```
> shapiro.test(arResiduals)

        Shapiro-Wilk normality test

data:  arResiduals
W = 0.98974, p-value = 0.6711

> |
```

We can observe that the p-value = 0.6711 and that the p-value > α, and therefore we do not have enough evidence to reject the hypothesis. We can then conclude that the time series has a Gaussian distribution.