

# MSAN 604 - Homework 2

November 17, 2016

## 1 Textbook Problems

**3.1** Determine which of the following ARMA processes are causal and which of them are invertible. (In each case  $\{Z_t\}$  denotes white noise.)

a.  $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$

We have

$$X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t \Leftrightarrow \phi^2(B)X_t = Z_t$$

with  $\phi^2(z) = 1 + 0.2z - 0.48z^2$ .

The roots of  $\phi(z)$  are  $z = \frac{-0.2 \pm \sqrt{(0.2)^2 + 4(0.48)^2}}{2(-0.48)} = \frac{0.2 \pm 1.4}{0.96}$ . We get  $z_1 = -1.25$  and  $z_2 = 5/3$ .

We have  $|z_1| > 1$  and  $|z_2| > 1$ , so the process is **stationary**.

In addition, we have  $\theta(z) = 1$ . So the process is **invertible**.

b.  $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$

We have

$$X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2} \Leftrightarrow \phi^2(B)X_t = \theta^2(B)Z_t$$

with  $\phi^2(z) = 1 + 1.9z + 0.88z^2$  and  $\theta^2(z) = 1 + 0.2z + 0.7z$ .

The roots of  $\phi^2(z)$  are  $z = \frac{-1.9 \pm \sqrt{(1.9)^2 - 4(0.88)}}{2(0.88)} = \frac{-1.9 \pm \sqrt{0.09}}{1.76}$ . We get  $z_1 = -10/11$  and  $z_2 = -5/4$ .

We have  $|z_1| < 1$ , so the process is **not stationary**.

The roots of  $\theta^2(z)$  are  $z = \frac{-0.2 \pm \sqrt{(0.2)^2 - 4(0.7)}}{2(0.7)} = \frac{-0.2}{1.4} \pm i \frac{\sqrt{2.76}}{1.4}$ .

We have  $|z| = \sqrt{(0.2/1.4)^2 + (\sqrt{2.76}/1.4)^2} \simeq 1.2 > 1$ . So the process is **invertible**.

c.  $X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1}$

We have

$$X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1} \Leftrightarrow \phi(B)X_t = \theta(B)Z_t$$

with  $\phi(z) = 1 + 0.6z$  and  $\theta(z) = 1 + 1.2z$ .

The root of  $\phi(z)$  is  $z = \frac{-1}{0.6}$ .

We have  $|z| > 1$ , so the process is **stationary**.

The root of  $\theta(z)$  is  $z = \frac{-1}{1.2}$ .

We have  $|z| < 1$ , so the process is **not invertible**.

d.  $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$

We have

$$X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t \Leftrightarrow \phi^2(B)X_t = Z_t$$

with  $\phi^2(z) = 1 + 1.8z + 0.81z^2$ .

The root of  $\phi^2(z)$  is  $z = \frac{-1.8 \pm \sqrt{(1.8)^2 - 4(0.81)}}{2(0.81)} = \frac{-1.8 \pm \sqrt{3.24 - 3.24}}{1.62} = -10/9$ .

We have  $|z| > 1$ , so the process is **causal**.

In addition, we have  $\theta(z) = 1$ . So the process is **invertible**.

e.  $X_t + 1.6X_{t-1} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-2}$

We have

$$X_t + 1.6X_{t-1} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-2} \Leftrightarrow \phi(B)X_t = \theta^2(B)Z_t$$

with  $\phi(z) = 1 + 1.6z$  and  $\theta^2(z) = 1 - 0.4z + 0.04z^2$ .

The root of  $\phi(z)$  is  $z = -1/1.6 < 1$ , so the process is **not stationary**.

The root of  $\theta(z)$  is  $z = \frac{0.4 \pm \sqrt{(0.4)^2 - 4(0.04)}}{2(0.04)} = \frac{0.4 \pm \sqrt{0.16 - 0.16}}{0.08} = 5 > 1$ . So the process is **invertible**.

**1.10** If  $m_t = \sum_{k=0}^p c_k t^k$ ,  $t = 0, \pm 1, \dots$ , show that  $\nabla m_t$  is a polynomial of degree  $p - 1$  in  $t$  and hence that  $\nabla^{p+1} m_t = 0$ .

We have  $\nabla m_t = (1 - B)m_t = m_t - m_{t-1} = \sum_{k=0}^p c_k t^k - \sum_{k=0}^p c_k (t-1)^k$ .

It can easily be seen that the terms of order  $p$  cancel each other out ( $(c_p - c_p)t^p = 0$ ).

In addition, the term of order  $p - 1$  is  $pc_{p-1} \neq 0$ . So the resulting polynomial is of degree  $p - 1$ .

Therefore, every time the difference operator is applied to  $m_t$ , we get a polynomial of one less degree than  $m_t$ . As a result, we can have that  $\nabla^p m_t$  is a polynomial of degree 0 (i.e. a constant). Therefore, we conclude that  $\nabla^{p+1} m_t = 0$ .

**1.15** Let  $\{Y_t\}$  be a stationary process with mean zero and let  $a$  and  $b$  be constants.

a. If  $X_t = a + bt + s_t + Y_t$ , where  $s_t$  is a seasonal component with period 12, show that  $\nabla \nabla_{12} X_t = (1 - B)(1 - B^{12})X_t$  is stationary and express its autocovariance function in terms of that of  $\{Y_t\}$ .

Let's call  $W_t = \nabla \nabla_{12} X_t$ .

We have

$$\begin{aligned} W_t &= \nabla \nabla_{12} X_t \\ &= (1 - B)(1 - B^{12})X_t \\ &= (1 - B)(X_t - X_{t-12}) \\ &= X_t - X_{t-12} - X_{t-1} + X_{t-13} \\ &= a + bt + s_t + Y_t - a - b(t-12) - s_{t-12} - Y_{t-12} - a - b(t-1) - s_{t-1} - Y_{t-1} + a + b(t-13) + s_{t-13} + Y_{t-13} \\ &= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} \end{aligned}$$

So we have  $\mu_W(t) = E[W_t] = E[Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}] = 0$  does not depend on  $t$ .

We also have

$$\begin{aligned}
\gamma_W(t, t+h) &= \text{Cov}(W_t, W_{t+h}) \\
&= \text{Cov}(Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}, Y_{t+h} - Y_{t+h-1} - Y_{t+h-12} + Y_{t+h-13}) \\
&= \text{Cov}(Y_t, Y_{t+h}) - \text{Cov}(Y_t, Y_{t+h-1}) - \text{Cov}(Y_t, Y_{t+h-12}) + \text{Cov}(Y_t, Y_{t+h-13}) \\
&\quad - \text{Cov}(Y_{t-1}, Y_{t+h}) + \text{Cov}(Y_{t-1}, Y_{t+h-1}) + \text{Cov}(Y_{t-1}, Y_{t+h-12}) - \text{Cov}(Y_{t-1}, Y_{t+h-13}) \\
&\quad - \text{Cov}(Y_{t-12}, Y_{t+h}) + \text{Cov}(Y_{t-12}, Y_{t+h-1}) + \text{Cov}(Y_{t-12}, Y_{t+h-12}) - \text{Cov}(Y_{t-12}, Y_{t+h-13}) \\
&\quad + \text{Cov}(Y_{t-13}, Y_{t+h}) - \text{Cov}(Y_{t-13}, Y_{t+h-1}) - \text{Cov}(Y_{t-13}, Y_{t+h-12}) + \text{Cov}(Y_{t-13}, Y_{t+h-13}) \\
&= \gamma_Y(h) - \gamma_Y(h-1) - \gamma_Y(h-12) + \gamma_Y(h-13) \\
&\quad - \gamma_Y(h+1) + \gamma_Y(h) + \gamma_Y(h-11) - \gamma_Y(h-12) \\
&\quad - \gamma_Y(h+12) + \gamma_Y(h+11) + \gamma_Y(h) - \gamma_Y(h-1) \\
&\quad + \gamma_Y(h+13) - \gamma_Y(h+12) - \gamma_Y(h+1) + \gamma_Y(h) \\
&= 4\gamma_Y(h) - 2\gamma_Y(h-1) - 2\gamma_Y(h-12) + \gamma_Y(h-13) \\
&\quad - 2\gamma_Y(h+1) + \gamma_Y(h-11) - 2\gamma_Y(h+12) + \gamma_Y(h+11) + \gamma_Y(h+13)
\end{aligned}$$

So we get that  $\gamma_W(t, t+h)$  does not depend on  $t$ .

Therefore, we can conclude that the process  $\{W_t\}$  is stationary.

b. If  $X_t = (a + bt)s_t + Y_t$ , where  $s_t$  is a seasonal component with period 12, show that  $\nabla_{12}^2 X_t = (1 - B^{12})^2 X_t$  is stationary and express its autocovariance function in terms of that of  $\{Y_t\}$ .

Let's call  $Z_t = \nabla_{12}^2 X_t$ .

We have

$$\begin{aligned}
Z_t &= \nabla_{12}^2 X_t \\
&= (1 - B^{12})^2 X_t \\
&= (1 - B^{12})(1 - B^{12})X_t \\
&= (1 - B^{12})(X_t - X_{t-12}) \\
&= X_t - X_{t-12} - X_{t-12} + X_{t-24} \\
&= X_t - 2X_{t-12} + X_{t-24} \\
&= (a + bt)s_t + Y_t - 2((a + b(t-12))s_{t-12} + Y_{t-12}) + (a + b(t-24))s_{t-24} + Y_{t-24} \\
&= a(s_t - 2s_{t-12} + s_{t-24}) + b(ts_t - 2(t-12)s_{t-12} + (t-24)s_{t-24}) + Y_t - 2Y_{t-12} + Y_{t-24} \\
&= Y_t - 2Y_{t-12} + Y_{t-24}
\end{aligned}$$

So we have  $\mu_Z(t) = E[Z_t] = E[Y_t - 2Y_{t-12} + Y_{t-24}] = 0$  does not depend on  $t$ .

We also have

$$\begin{aligned}
\gamma_Z(t, t+h) &= \text{Cov}(Z_t, Z_{t+h}) \\
&= \text{Cov}(Y_t - 2Y_{t-12} + Y_{t-24}, Y_{t+h} - 2Y_{t+h-12} + Y_{t+h-24}) \\
&= \text{Cov}(Y_t, Y_{t+h}) - 2\text{Cov}(Y_t, Y_{t+h-12}) + \text{Cov}(Y_t, Y_{t+h-24}) \\
&\quad - 2\text{Cov}(Y_{t-12}, Y_{t+h}) + 4\text{Cov}(Y_{t-12}, Y_{t+h-12}) - 2\text{Cov}(Y_{t-12}, Y_{t+h-24}) \\
&\quad + \text{Cov}(Y_{t-24}, Y_{t+h}) - 2\text{Cov}(Y_{t-24}, Y_{t+h-12}) + \text{Cov}(Y_{t-24}, Y_{t+h-24}) \\
&= \gamma_Y(h) - 2\gamma_Y(h-12) + \gamma_Y(h-24) \\
&\quad - 2\gamma_Y(h+12) + 4\gamma_Y(h) - 2\gamma_Y(h-12) \\
&\quad + \gamma_Y(h+24) - 2\gamma_Y(h+12) + \gamma_Y(h) \\
&= 6\gamma_Y(h) - 4\gamma_Y(h-12) + \gamma_Y(h-24) - 4\gamma_Y(h+12) + \gamma_Y(h+24)
\end{aligned}$$

So we get that  $\gamma_Z(t, t+h)$  does not depend on  $t$ .

Therefore, we can conclude that the process  $\{Z_t\}$  is stationary.

## 2 Additional Problems

### 2.1 ARIMA fitting with the LakeHuron dataset.

The LakeHuron dataset contains annual measurements of the level, in feet, of Lake Huron between 1875 and 1972.

(a) Take ordinary differences of the data until the resulting time series is suitably stationary, using the Augmented Dickey-Fuller Test to aid in your choice of  $d$ .

We start by plotting the data and the ACF plot:

```
# Plot original data
par(mfrow=c(2,1))
plot(LakeHuron)
acf(LakeHuron)
adf.test(LakeHuron)
```

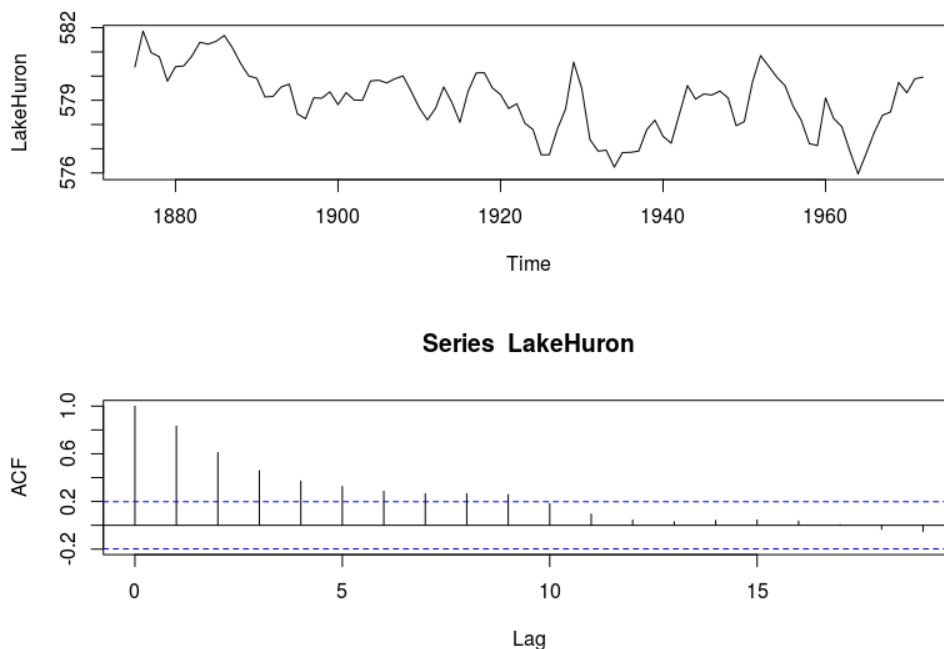


Figure 1: Plot of the data and ACF plot.

We get the plots shown in Figure 1, where we can see that the process is clearly not stationary. Indeed, the Augmented Dickey-Fuller test produces a p-value of 0.254. So we do not reject the null hypothesis that the time series is not stationary. Therefore, we try differencing once and look at the result:

```
# Difference once
LakeHuron1 <- diff(LakeHuron)
plot(LakeHuron1)
acf(LakeHuron1)
adf.test(LakeHuron1)
```

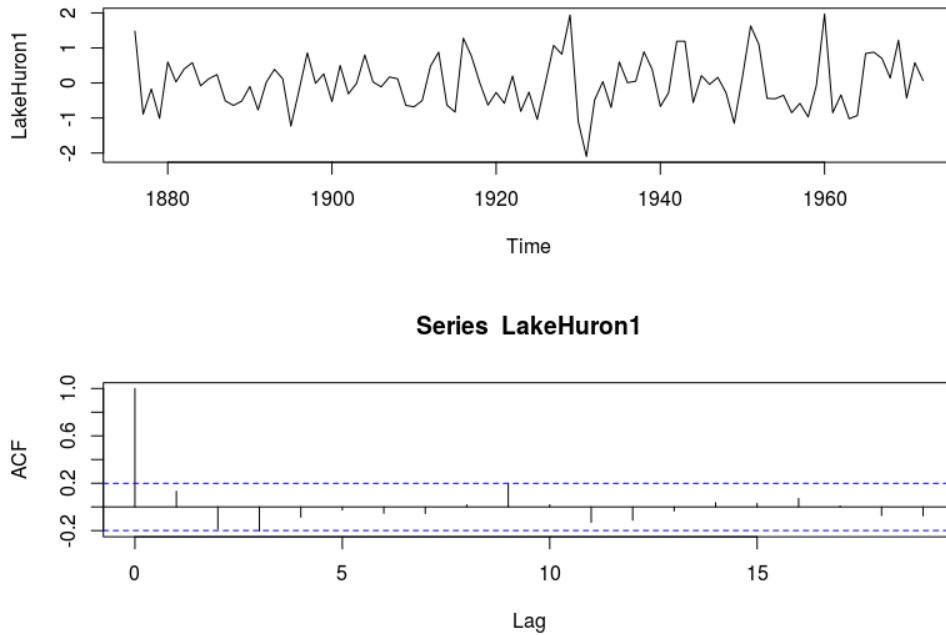


Figure 2: Plot of the 1-differenced data and ACF plot.

We get the plots shown in Figure 2, where the process seems stationary. Indeed, the Augmented Dickey-Fuller test now produces a p-value  $< 0.01$ , so we reject the null hypothesis and conclude that the time series is stationary. We do not need to difference again. We conclude that  $d = 1$ .

(b) Fit an  $AR(1)$  model to the d-differenced **LakeHuron** data, using maximum likelihood estimation.

We fit an  $AR(1)$  model to the 1-differenced data:

```
m1 <- arima(LakeHuron1, order = c(1,0,0))
```

For this model, we get  $\hat{\sigma}^2 = 0.5452$ , log likelihood = -108.23, and aic = 222.45.

(c) Fit an  $AR(2)$  model to the d-differenced **LakeHuron** data, using maximum likelihood estimation.

We fit an  $AR(1)$  model to the 1-differenced data:

```
m2 <- arima(LakeHuron1, order = c(1,0,0))
```

For this model, we get  $\hat{\sigma}^2 = 0.5188$ , log likelihood = -105.87, and aic = 219.74.

(d) Perform a likelihood ratio test comparing the  $AR(1)$  and  $AR(2)$  models and state whether or not the null hypothesis is rejected.

We perform a Likelihood Ratio Test to compare the two models. Here, the null model is **m1**, and the alternative model is **m2**:

```
# Compare the two models using LRT
D <- -2*(m2$loglik - m1$loglik)
```

```
pval <- 1-pchisq(D,1)
print(c("Test Statistic:",round(D,4),"P-value:",round(pval,4)))
```

We get a p-value of 1, so we do not reject the null hypothesis that the two models fit equally well.

(e) Using your results from (d) and an assessment of  $AIC$  and  $\hat{\sigma}^2$  for each model, state which model you deem to be “optimal” and briefly explain this choice.

From the result in (d), we are inclined to choose the simpler model  $m_1$ . By comparing  $AIC$  and  $\hat{\sigma}^2$ , we see that they are slightly higher for  $m_1$ . However, the difference does not seem enough to warrant using the AR(2) model in this case.

(f) Using appropriate formal and informal residual diagnostics, investigate whether the “optimal” model you’ve chosen in (e) satisfies the following assumptions

- Zero-Mean
- Homoscedasticity
- Zero-Correlation
- Normality

We start by extracting the residuals from the model:

```
# Residual diagnostics
e <- m1$residuals # residuals
r <- e/sqrt(m1$sigma2) # standardized residuals
```

Then, we plot the residuals against time. The horizontal line at 0 makes it easier to see whether the data seems to have zero-mean.

```
# Plot residuals vs t
par(mfrow=c(2,1))
plot(e, main="Residuals vs t", ylab="")
abline(h=0, col="red")
plot(r, main="Standardized Residuals vs t", ylab="")
abline(h=0, col="red")

# test whether residuals have zero mean
t.test(e)
```

The plots in Figure 3 show that the process seems to have zero-mean. We can also say that it looks homoskedastic, and we can’t discern any correlation nor outliers. The result from the t-test produces a p-value of 0.9765, so we do not reject the null hypothesis that the mean is zero.

In order to test Homoskedasticity, we split our data into four groups of similar length. Then we use Bartlett’s and Levene’s tests to check whether the variance in each group is statistically equal.

```
# test for heteroscedasticity
par(mfrow=c(1,1))
plot(e, main="Residuals vs t", ylab="")
abline(v=c(1899,1923,1947), lwd=3, col="red")
group <- c(rep(1,24),rep(2,24),rep(3,24),rep(4,25))
levene.test(e,group) #Levene
bartlett.test(e,group) #Bartlett
```

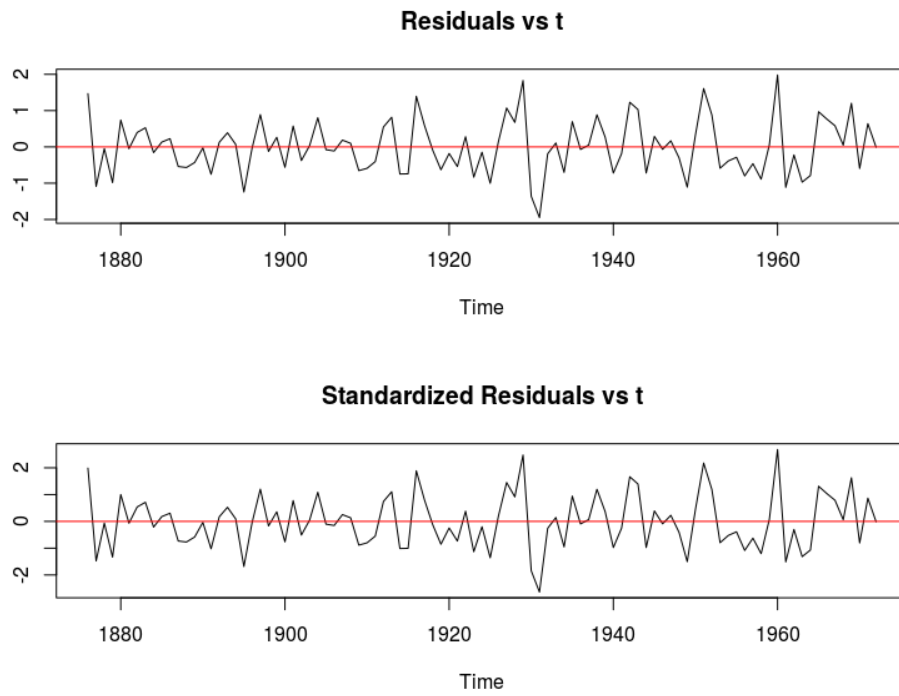


Figure 3: Sequential plot of residuals and standardized residuals.

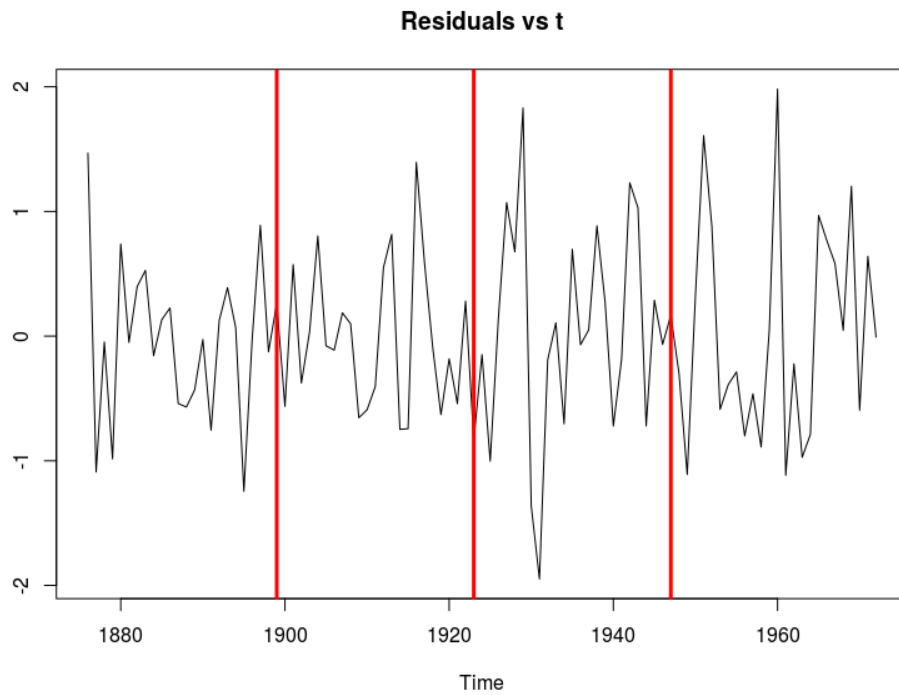


Figure 4: Data splitting for homoskedasticity tests.

Bartlett's test produces a p-value of 0.2522, and Levene's test 0.1644. Therefore neither null hypothesis is rejected, and we conclude that the process is homoskedastic.

To formally check for uncorrelatedness, we use the Ljung-Box test:

```
# test for uncorrelatedness / randomness
tsdiag(m1)
```

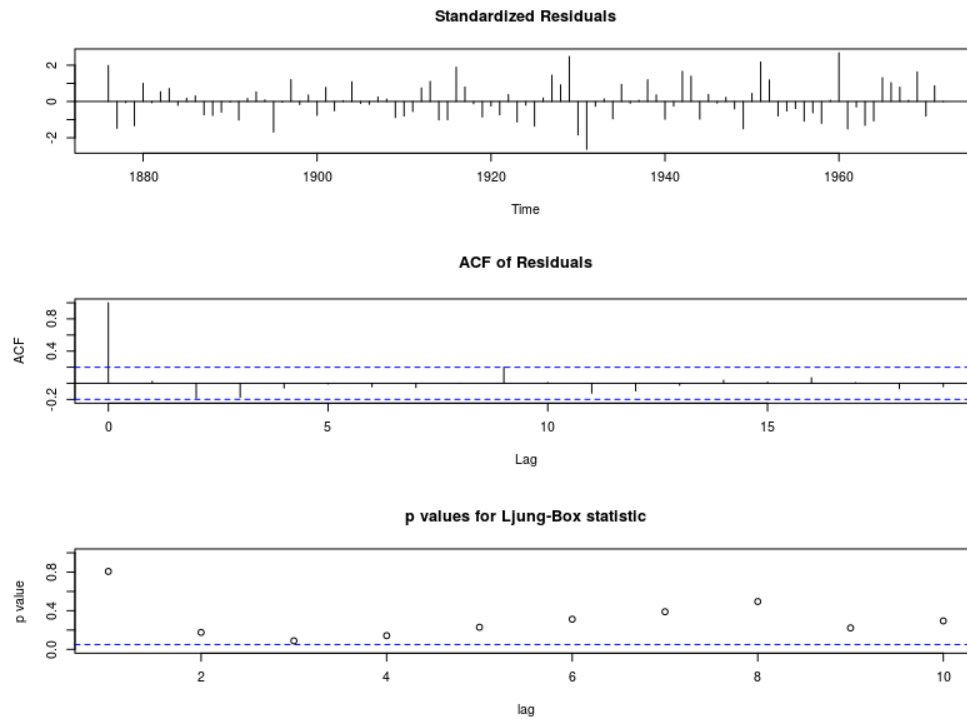


Figure 5: Plots for the Ljung-Box test.

The plots in Figure 5 indicate that the data is random/uncorrelated.

Finally, we need to verify if the residuals are normally distributed, since we used Maximum Likely Estimators:

```
# test for normality
par(mfrow=c(1,1))
qqnorm(e, main="QQ-plot of Residuals")
qqline(e, col = "red")
shapiro.test(e) #SW test
```

The QQ-plot in Figure 6 seems OK, and Shapiro-Wilk's test produces a p-value of 0.5982, so we do not reject the null hypothesis that the residuals are normally distributed.

All the residual diagnostics satisfy the assumptions, so we can accept the AR(1) model to model this process.



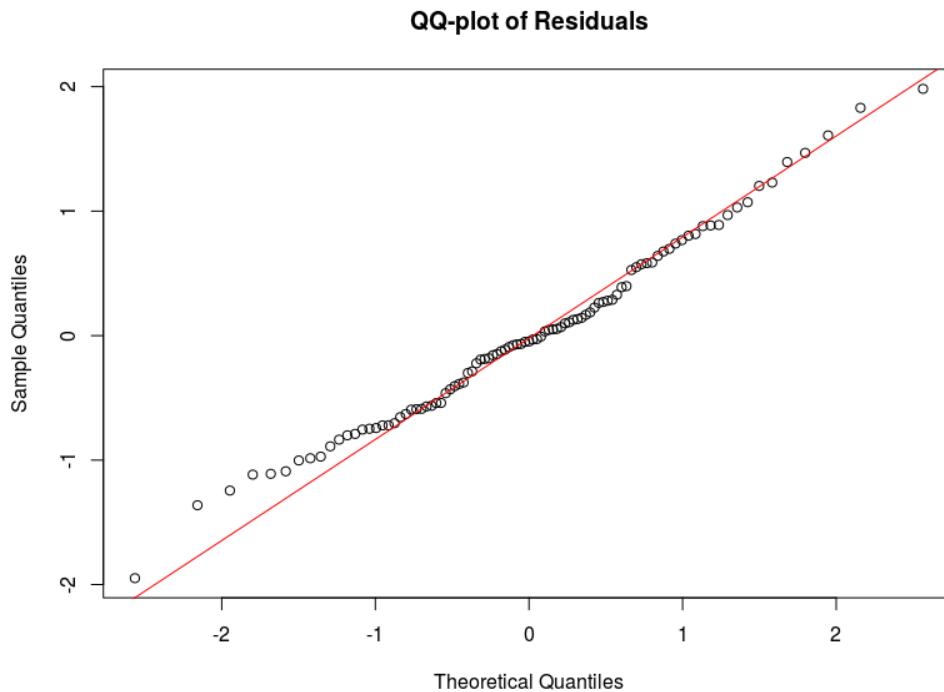


Figure 6: QQ-plot of the residuals.

## 2.2 SARIMA fitting with the `beers.csv` dataset.

The `beers.csv` dataset contains information on monthly Australian beer production, in millions of litres, from January 1956 to August 1995.

(a) Using techniques discussed in class, fit (using maximum likelihood estimation) an “optimal”  $SARIMA(p, d, q) \times (P, D, Q)$  model to the `beers.csv` time series, and justify your choice of  $p, d, q, P, D, Q$  and  $s$ .

We start by importing the data:

```
# Load data
beer <- read.csv("beer.csv", header=T)
beer <- ts(beer, start = 1956, end = 1995, frequency = 12)
```

The term `frequency=12` is used to indicate that the data is separated by months.

Then, we can plot the process, as well as the ACF plot to determine stationarity and seasonality:

```
# Plot the data
par(mfrow=c(2,1))
plot(beer)
acf(beer, lag.max = 144)
adf.test(beer)
```

The resulting plots shown in Figure 7 show that the time series is not stationary. However, the Augmented Dickey-Fuller test produces a significant p-value, which would suggest that we reject the null hypothesis that the process is not stationary.

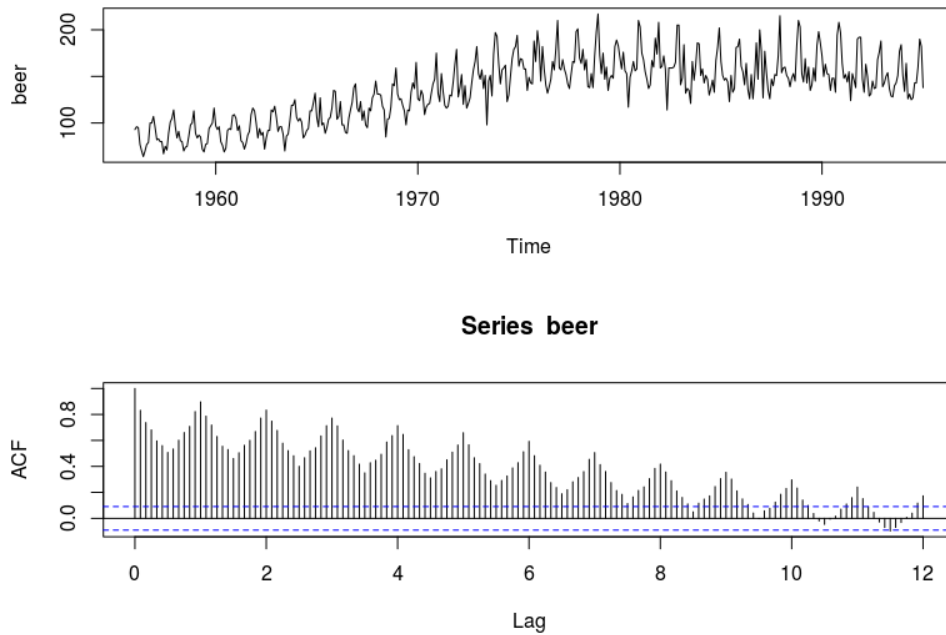


Figure 7: Plot of beer data set and ACF plot.

Now, we difference again in order to remove the seasonality pattern.

```
# Difference for seasonality (period=12)
beer1.12 <- diff(beer1, lag = 12)
plot(beer1.12)
acf(beer1.12, lag.max = 144)
```

(b) Using appropriate formal and informal residual diagnostics, investigate whether the “optimal” model you’ve found in (a) satisfies the following assumptions “optimal” model you’ve chosen in (e) satisfies the following assumptions

- Zero-Mean
- Homoscedasticity
- Zero-Correlation
- Normality