Andre Guimaraes Duarte
20406263

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 $\theta(z) = 1 \neq 0 \ \forall z$. 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 $m_t = \sum_{k=0}^p c_k t^k$, $t = 0, \pm 1, \ldots$, show that ∇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

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

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

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

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, main="Height of Lake Huron Through the Years",
     xlab = "Year", ylab = "Height (ft)")
acf(LakeHuron)
adf.test(LakeHuron)
```
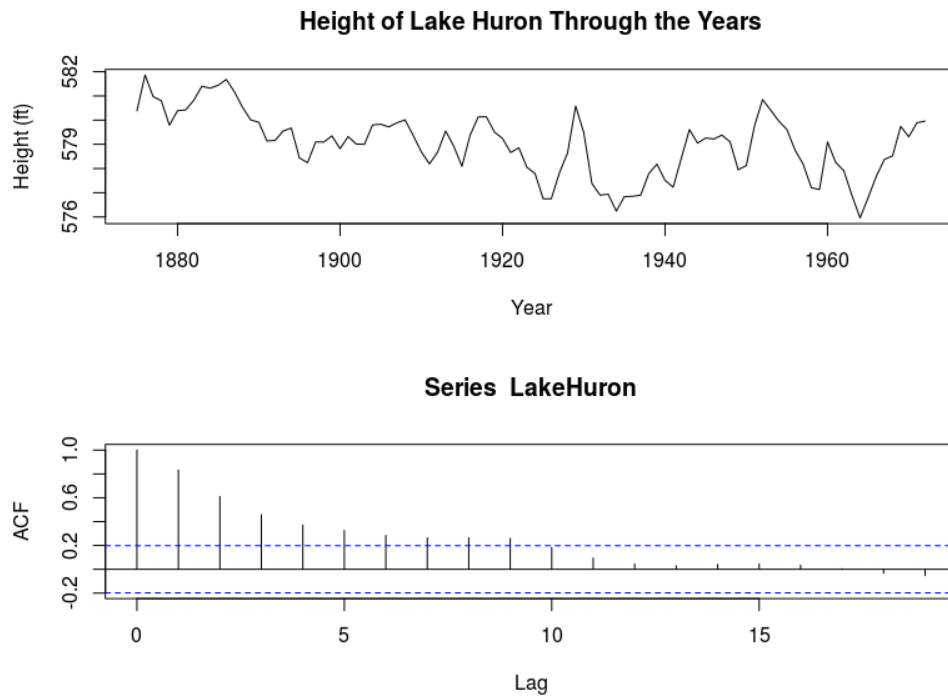

Figure 1: Plot of the data and ACF plot.

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, main="Differenced Height of Lake Huron Through the Years",
     xlab = "Year", ylab = "Height (ft)")
acf(LakeHuron1)
adf.test(LakeHuron1)
```

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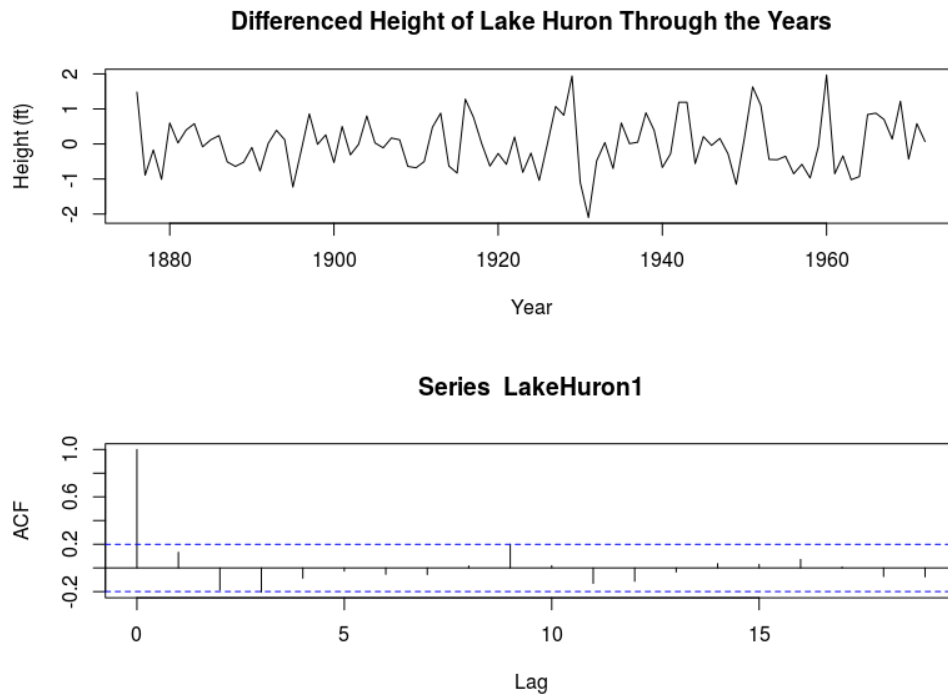
Figure 2: Plot of the 1-differenced data and ACF plot.

(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))
summary(m1)
```

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(2) model to the 1-differenced data:

```
m2 <- arima(LakeHuron1, order = c(2,0,0))
summary(m2)
```

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*(m1$loglik - m2$loglik)
pval <- 1-pchisq(D,1)
print(c("Test Statistic:",round(D,4),"P-value:",round(pval,4)))
```

5

We get a p-value of 0.0299, so we reject the null hypothesis that the two models fit equally well, and accept the alternative that the AR(2) model fits better.

(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 AR(2) model `m2`. By comparing AIC and $\hat{\sigma}^2$, we see that they are lower for `m2`, confirming that this model fits better.

(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 <- m2$residuals # residuals
r <- e/sqrt(m2$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 0.9888, 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, as seen in Figure 4. 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
```
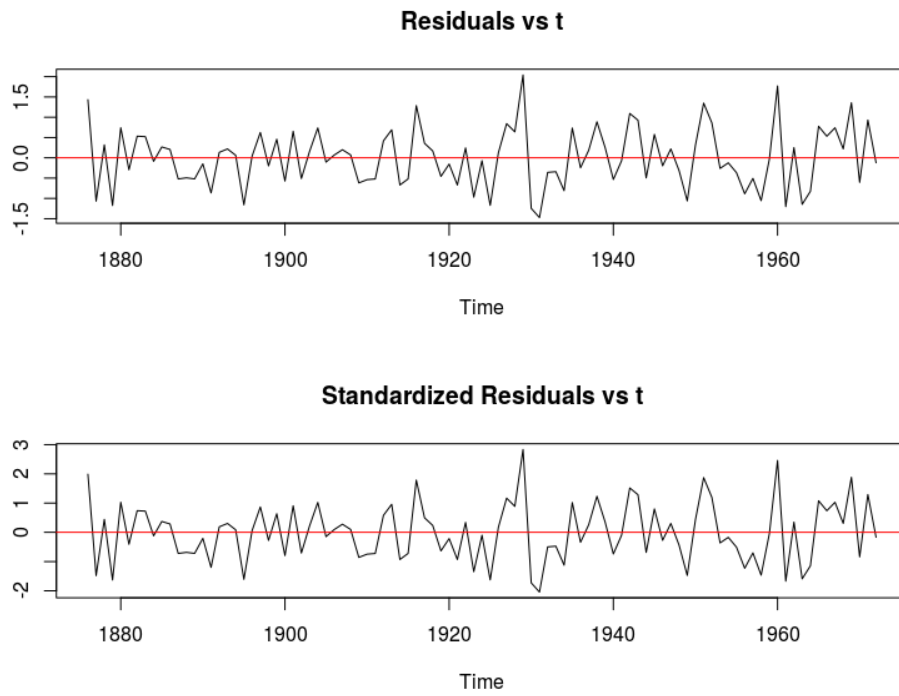
Residuals vs t

Standardized Residuals vs t

Figure 3: Sequential plot of residuals and standardized residuals.
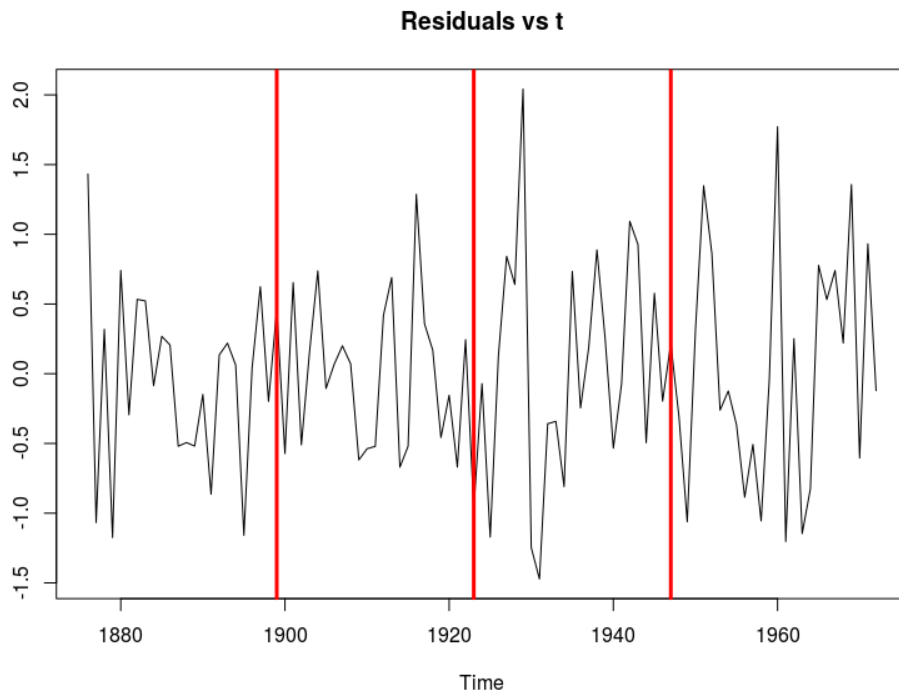
Residuals vs t

Figure 4: Data splitting for homoskedasticity tests.

Bartlett's test produces a p-value of 0.1629, and Levene's test 0.2029. Therefore neither null hypothesis is

7

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)
```
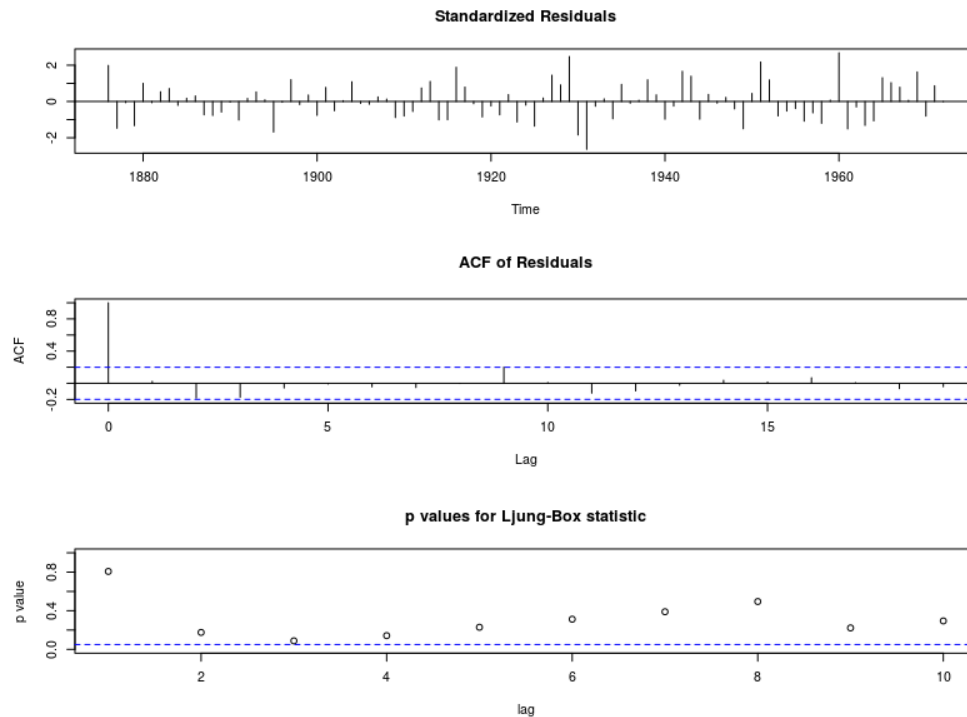

Figure 5: Plots for the Ljung-Box test.

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 0.4233, 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(2) model to model this process.
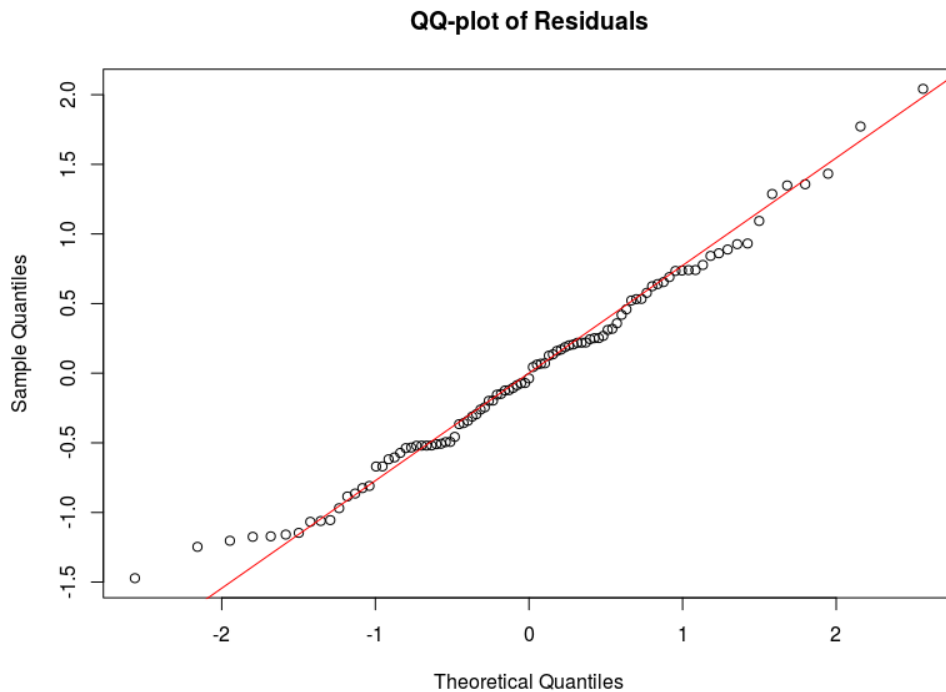
QQ-plot of Residuals

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 and plotting the series:

```
# Load data
beer <- read.csv("beer.csv", header=T)
beer <- ts(beer, start = c(1956,1), frequency = 12)
plot(beer, main="Beer Production in Australia",
     xlab="Year", ylab="Production (x1e6 L)")
```

The term `frequency=12` is used to indicate that the data is separated by months. We get the plot shown in Figure 7. We can see that there seems to be a polynomial trend as well as seasonality in the process. In addition, the amplitude of the variations seem to increase with time, so we may think about doing a log transformation on the data. This is what we do, and we check the resulting plot:

```
# Take log-transform of the data
lbeer <- log(beer)
plot(lbeer, main="Beer Production in Australia",
     xlab="Year", ylab="Log of Production (x1e6 L)")
```
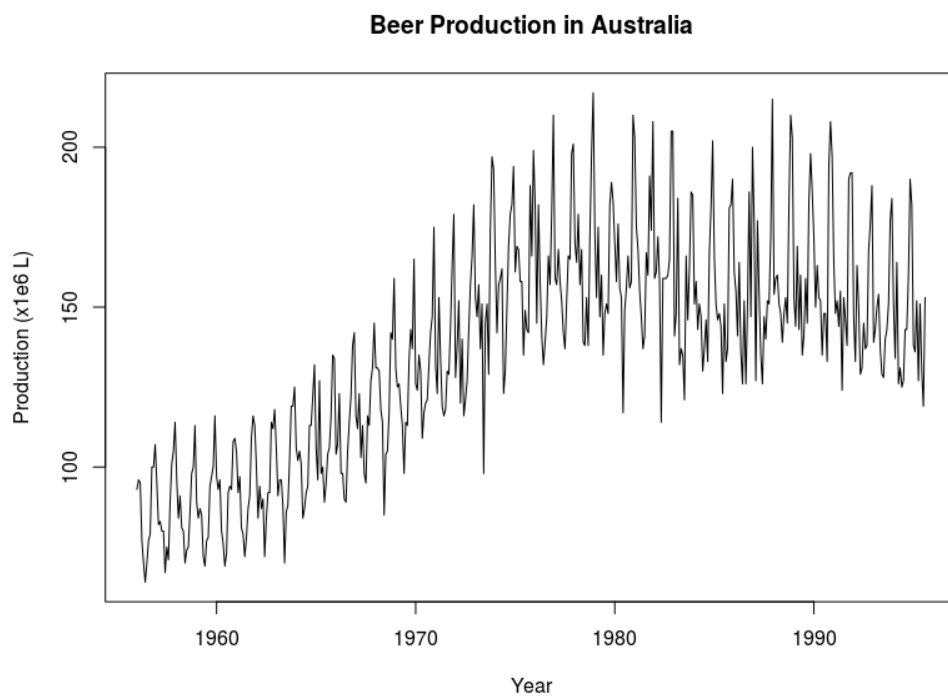
9

Beer Production in Australia

Figure 7: Plot of beer data set.
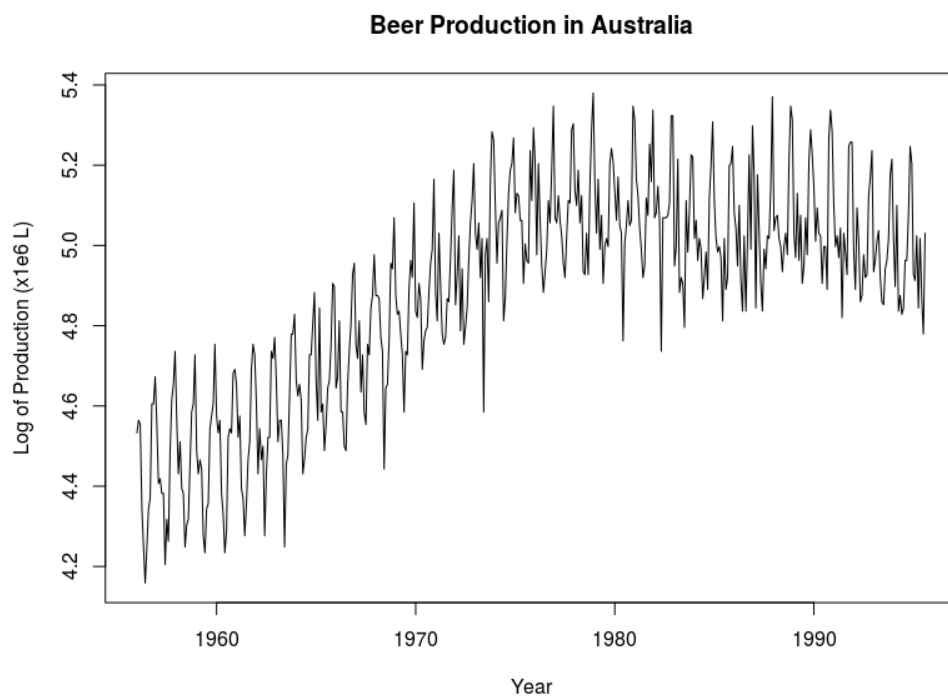
Beer Production in Australia

Figure 8: Plot of log-transformed beer data set.

As seen in Figure 8, the amplitude of the oscillations now seems more even throughout time.

We can then look at the ACF plot of the series to check for ordinary and/or seasonal differencing that may be necessary:

```
# check whether ordinary and/or seasonal differencing is necessary
acf(lbeer, lag.max = 96)
```
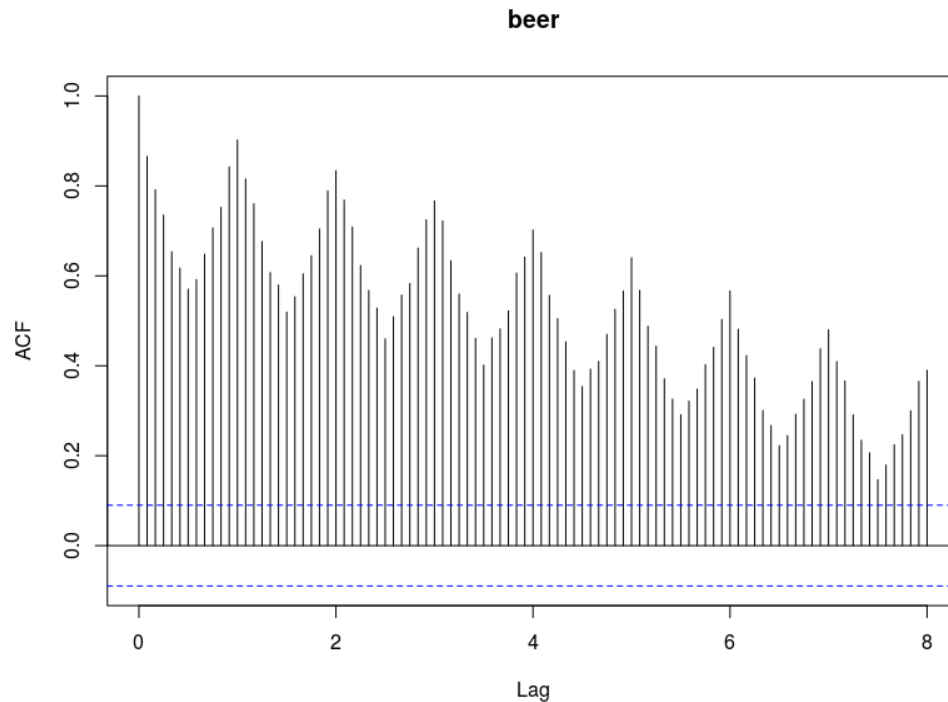

Figure 9: ACF of log-transformed beer data set.

Figure 9 shows slow decay as well as a sinusoidal pattern. This indicates that both trend and seasonality are present in the data and need to be accounted for. We start with ordinary differencing:

```
# Both forms of differencing seem necessary. Let's do ordinary first:
par(mfrow=c(2,1))
dlbeer <- diff(lbeer)
plot(dlbeer, main="Trend corrected Beer Production in Australia",
     xlab="Year", ylab="Log of Production (x1e6 L)")
acf(dlbeer, lag.max=48)
adf.test(dlbeer)
```

The resulting plots shown in Figure 10 seem ok. Also, the ADF test produces a p-value inferior to 0.01, so we reject the null hypothesis that the process is not stationary: we do not need to difference again. We have d=1.

Now, we perform seasonal differencing with lag 12 (this value was determined by counting how many "peaks" are in one period in the ACF plot):
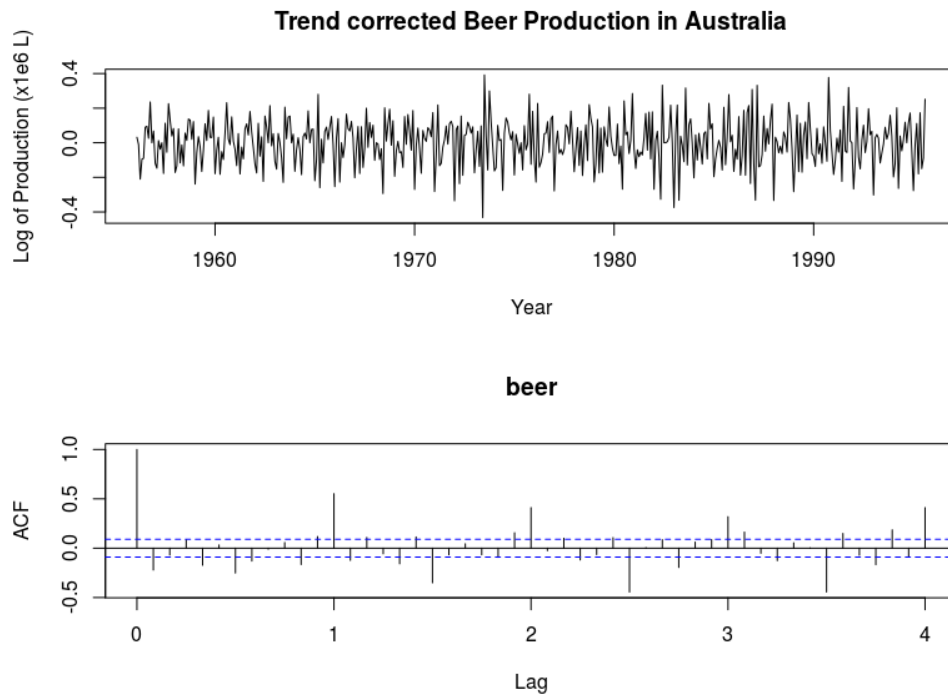
11

Figure 10: Trend-corrected log-transformed beer data set.

```
# Still need seasonal differencing:
dlbeer.12 <- diff(dlbeer, lag=12)
plot(dlbeer.12, main="Trend and Seasonality Corrected Beer Production in Australia",
     xlab="Year", ylab="Log of Production (x1e6 L)")
acf(dlbeer.12, lag.max=48)
adf.test(dlbeer.12, k=12)
```

The resulting plots shown in Figure 11 seem ok. Also, the ADF test produces a p-value inferior to 0.01, so we reject the null hypothesis that the process is not stationary: we do not need to seasonally difference again. We have D=1 and s=12.

Now that we have appropriately differenced, we know that we are fitting a SARIMA model, and therefore need to choose p, P, q, and Q. In order to do so, we look at the ACF and PACF plots of the differences data:

```
# This seems fine now. Since we seasonally differenced, we are fitting a SARIMA model
# and need to choose p, P, q, Q.
# Let's look at the ACF/PACF plots for this
par(mfrow=c(2,1))
acf(dlbeer.12, lag.max=48)
pacf(dlbeer.12, lag.max=48)
```

From the two plots shown in Figure 12, we can infer the following hypotheses: $p \leq 5, q \leq 4, P = 0, Q = 1$. Therefore, we fit the SARIMA $(5, 1, 4) \times (0, 1, 1)_{12}$ model (more details on the selection of this particular model in the Annex):
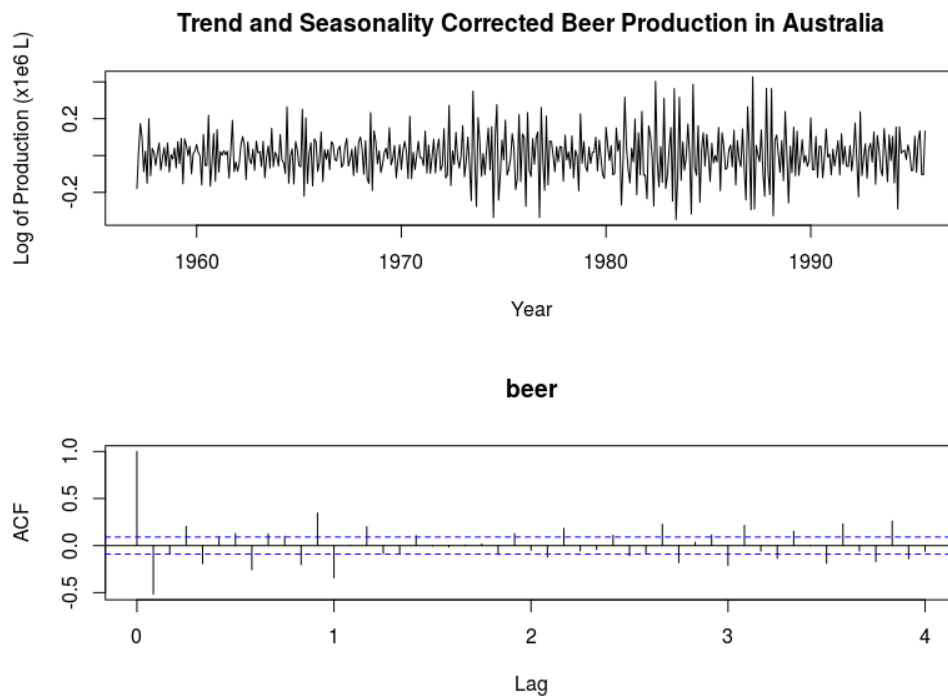
12

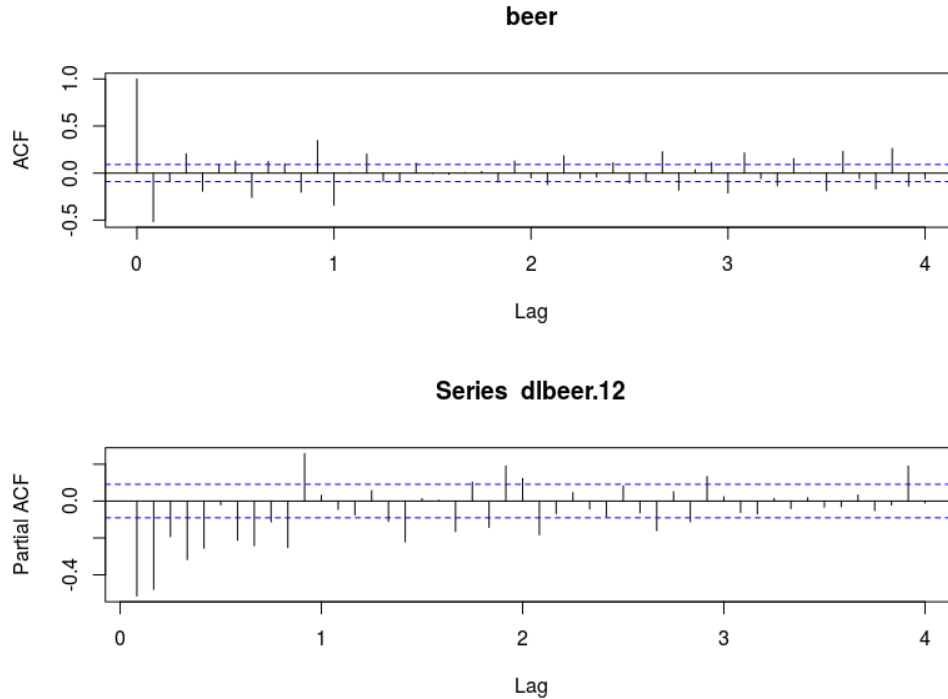Figure 11: Seasonal and trend-corrected log-transformed beer data set.

Figure 12: ACF and PACF plots of log-transformed beer data set.

```
# Fit model
```

```
m <- arima(lbeer, c(3,1,4), list(order=c(0,1,1), period=12))
summary(m)
```

This model is fit using Maximum Likelihood Estimation (after finding starting values with Conditional Sum of Squares). For this model, we get the following coefficients and standard errors:

Coefficients	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3	θ_4	Θ_1
Values	-0.3895	-0.1163	0.7648	-0.6452	-0.2173	-0.8401	0.8296	-0.8385
S.E.	0.0627	0.0724	0.0627	0.0489	0.0156	0.0172	0.0459	0.0297

In addition, we have $\hat{\sigma}^2 = 0.003702$, log-likelihood = 626.99, AIC = -1235.99.

We can now visualize how this model fits our data:

```
# Let's visualize how well this model fits
f <- lbeer - m$residuals # fitted values
par(mfrow=c(1,1))
plot(beer, main="Beer Production in Australia",
     xlab="Year", ylab="Production (x1e6 L)")
lines(exp(f), col="red")
legend("bottomright", legend = c("Observed", "Predicted"),
       lty = 1, col = c("black", "red"))
```
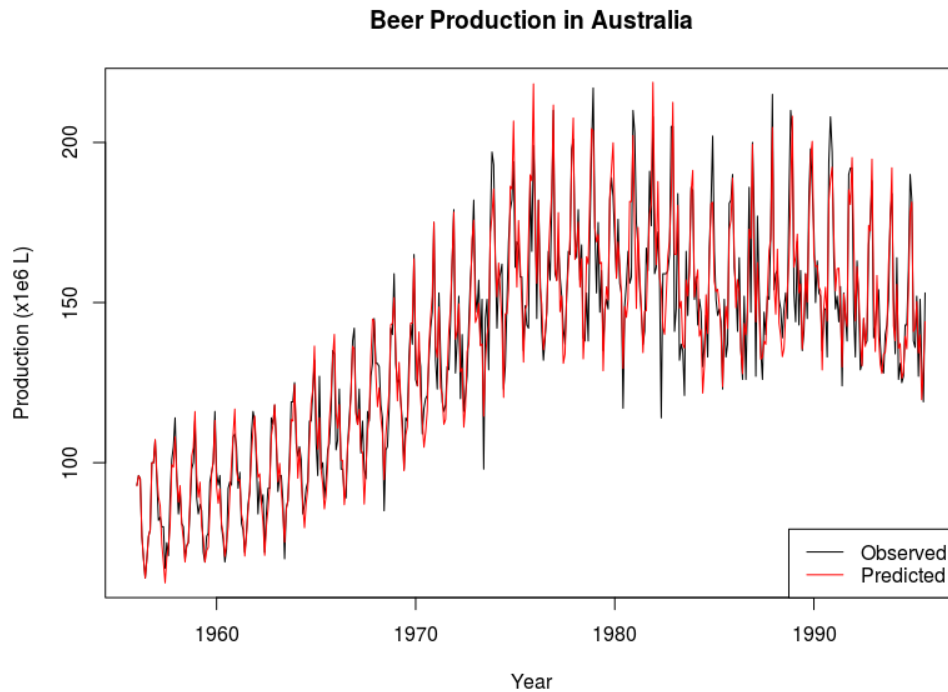

Figure 13: Plot of original beer data set and model predictions.

From Figure 13, we can see that this model fits the data well. The predicted values closely follow the original data.

(b) Using appropriate formal and informal residual diagnostics, investigate whether the "optimal" model you've found in (a) satisfies the following assumptions

14

- Zero-Mean

- Homoscedasticity

- Zero-Correlation

- Normality

We start by extracting the residuals from the model:

```
# Residual diagnostics
e <- m$residuals # residuals
r <- e/sqrt(m$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)
```
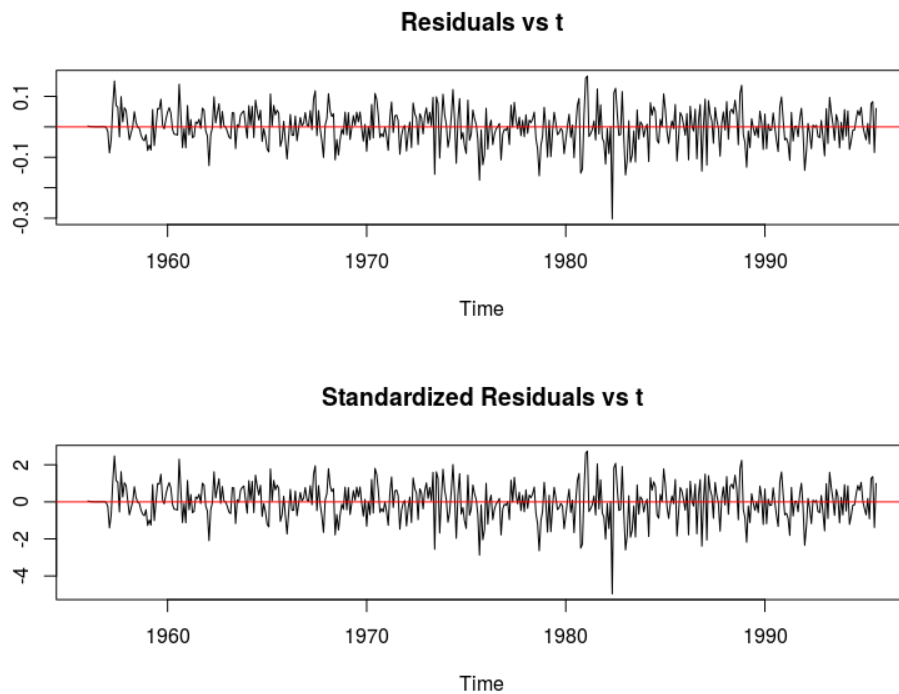

Figure 14: Sequential plot of residuals and standardized residuals.

15

The plots in Figure 14 show that the process seems to have zero-mean, with maybe one issue around 1982. We can't discern any correlation nor outliers, except again for one observation around 1982. The result from the t-test produces a p-value of 0.7775, 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, as shown in Figure 15. 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(1966,1976,1986), lwd=3, col="red")
group <- c(rep(1,119),rep(2,119),rep(3,119),rep(4,119))
levene.test(e,group) #Levene
bartlett.test(e,group) #Bartlett
```
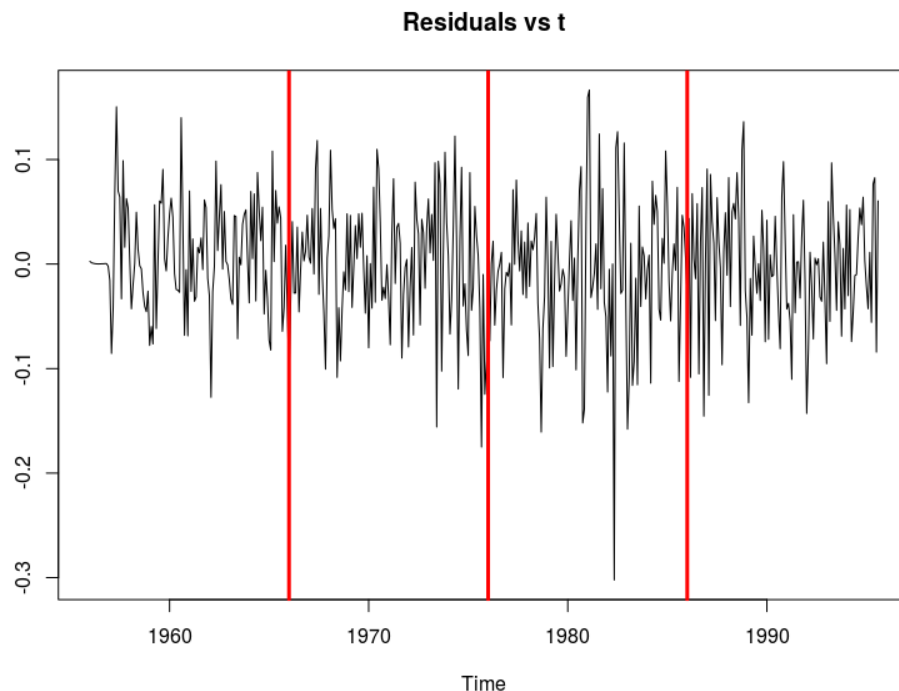

Figure 15: Sequential plot of residuals and standardized residuals.

Bartlett's test produces a p-value of 0.001243, and Levene's test 0.04535. Both p-values lead us to reject the null hypothesis that the variance is the same between each group. This model fails the homoskedasticity assumption. This may be due to the observation in 1982 that increases the variance in its group (and may also cause issues with normality later on). Informally, by only looking at the sequential plot of the residuals, it does not seem like the data is very heteroskedastic.

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

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

The plots in Figure 16 indicate that the data is random/uncorrelated.
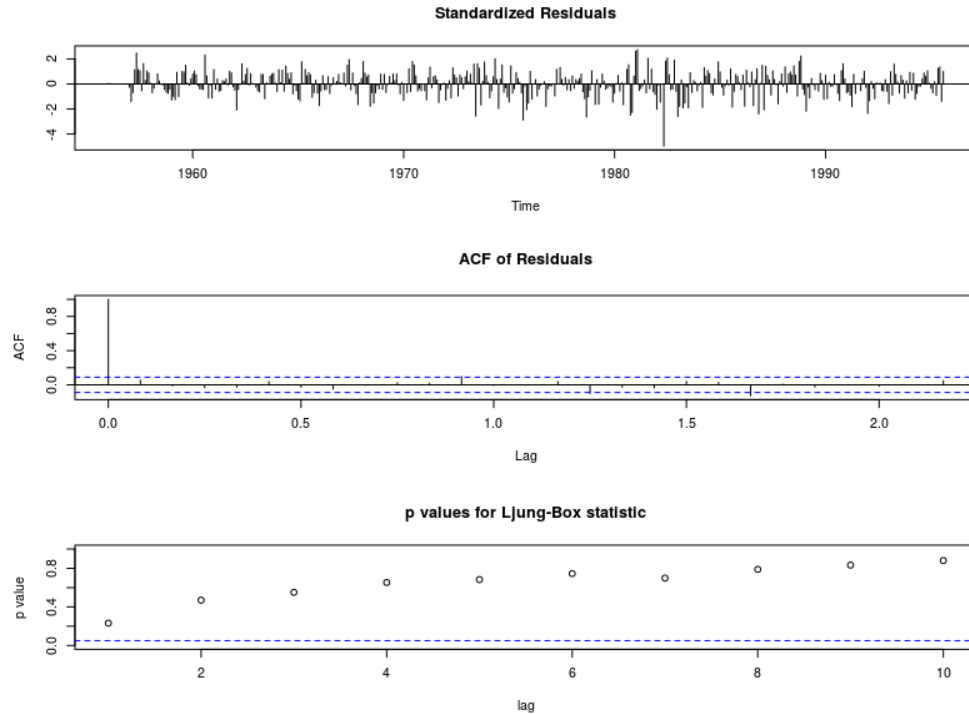
16

Figure 16: Plots for the Ljung-Box test.

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
# Histogram
hist(e, freq = F, xlim = c(-.4,.4), main = "Histogram of residuals")
curve(dnorm(x, mean(e), sd(e)), col="darkblue", lwd=1, add=TRUE, yaxt="n")
legend("topright", legend=c("N(-0.001, 0.004)"), lty = c(1), lwd=c(1),
       col=c("darkblue"))
```

The QQ-plot in Figure 17 seems OK, except on the lower tail, where some points start to drift away from the line. Shapiro-Wilk's test produces a p-value of 0.0001, so we are inclined to reject the null hypothesis that the residuals are normally distributed. The histogram on Figure 18 however suggests that the process does not deviate too much from normality. The low p-value observed from Shapiro-Wilk's test may be due to some outlier observations. On this basis, we can choose to accept normality for this model.

Therefore, we have seen that aside from possible heteroskedasticity, the SARIMA $(5, 1, 4) \times (0, 1, 1)_{12}$ model satisfies the residual diagnostics (at least informally). This seems like an adequate choice to model this process.
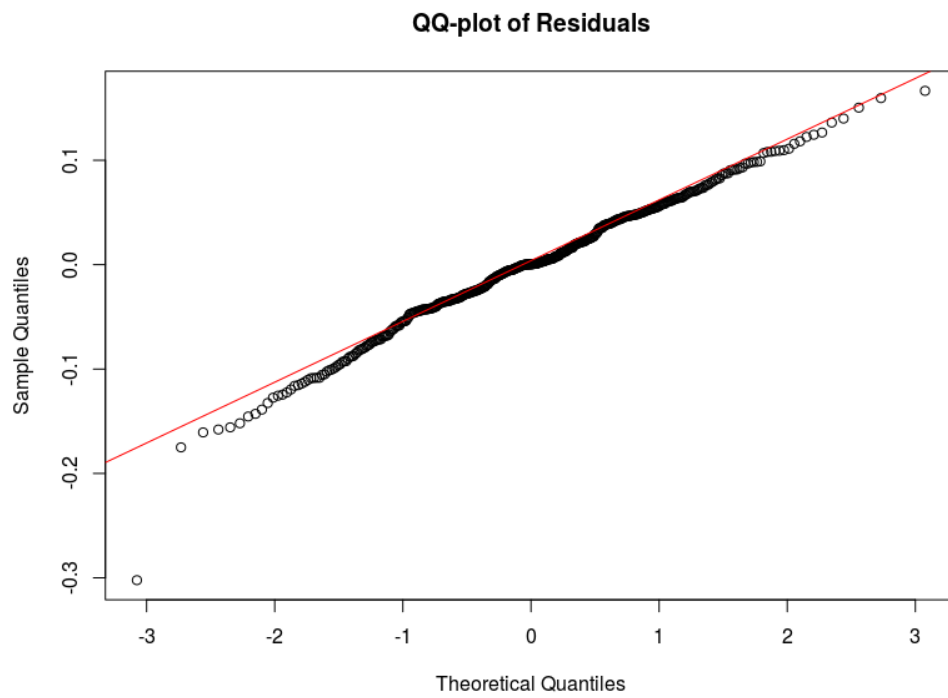
17

Figure 17: QQ-plot of the residuals.
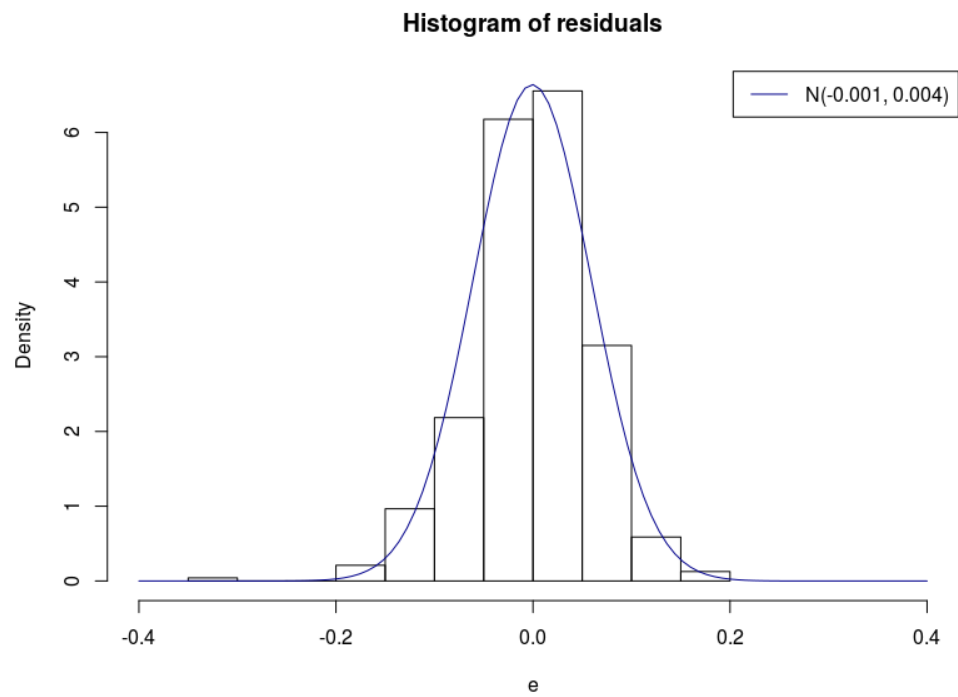

Figure 18: Histogram of residuals.

3   Annex

In order to select the right p, q, we fit all possible models such that $p \leq 5$ and $q \leq 5$, then compare $\hat{\sigma}^2$, log-likelihood, and AIC in order to choose the "best" one:

```
# Compare different models
m0<-arima(lbeer, c(1,1,1), list(order=c(0,1,1), period=12))
m1<-arima(lbeer, c(1,1,2), list(order=c(0,1,1), period=12))
m2<-arima(lbeer, c(1,1,3), list(order=c(0,1,1), period=12))
m3<-arima(lbeer, c(1,1,4), list(order=c(0,1,1), period=12))
m4<-arima(lbeer, c(2,1,1), list(order=c(0,1,1), period=12))
m5<-arima(lbeer, c(2,1,2), list(order=c(0,1,1), period=12))
m6<-arima(lbeer, c(2,1,3), list(order=c(0,1,1), period=12))
m7<-arima(lbeer, c(2,1,4), list(order=c(0,1,1), period=12))
m8<-arima(lbeer, c(3,1,1), list(order=c(0,1,1), period=12))
m9<-arima(lbeer, c(3,1,2), list(order=c(0,1,1), period=12))
m10<-arima(lbeer, c(3,1,3), list(order=c(0,1,1), period=12))
m11<-arima(lbeer, c(3,1,4), list(order=c(0,1,1), period=12))
m12<-arima(lbeer, c(4,1,1), list(order=c(0,1,1), period=12))
m13<-arima(lbeer, c(4,1,2), list(order=c(0,1,1), period=12))
m14<-arima(lbeer, c(4,1,3), list(order=c(0,1,1), period=12))
m15<-arima(lbeer, c(4,1,4), list(order=c(0,1,1), period=12))
m16<-arima(lbeer, c(5,1,1), list(order=c(0,1,1), period=12))
m17<-arima(lbeer, c(5,1,2), list(order=c(0,1,1), period=12))
m18<-arima(lbeer, c(5,1,3), list(order=c(0,1,1), period=12))
m19<-arima(lbeer, c(5,1,4), list(order=c(0,1,1), period=12))

sigma2<-c(m0$sigma2,m1$sigma2,m2$sigma2,m3$sigma2,m4$sigma2,m5$sigma2,m6$sigma2,m7$sigma2,
          m8$sigma2,m9$sigma2,m10$sigma2,m11$sigma2,m12$sigma2,m13$sigma2,m14$sigma2,
          m15$sigma2,m16$sigma2,m17$sigma2,m18$sigma2,m19$sigma2)
loglik<-c(m0$loglik,m1$loglik,m2$loglik,m3$loglik,m4$loglik,m5$loglik,m6$loglik,m7$loglik,
          m8$loglik,m9$loglik,m10$loglik,m11$loglik,m12$loglik,m13$loglik,m14$loglik,
          m15$loglik,m16$loglik,m17$loglik,m18$loglik,m19$loglik)
AIC<-c(m0$aic,m1$aic,m2$aic,m3$aic,m4$aic,m5$aic,m6$aic,m7$aic,m8$aic,m9$aic,
       m10$aic,m11$aic,m12$aic,m13$aic,m14$aic,m15$aic,m16$aic,m17$aic,m18$aic,m19$aic)
d <- data.frame(pdq = c("(1,1,1)","(1,1,2)","(1,1,3)","(1,1,4)",
                        "(2,1,1)","(2,1,2)","(2,1,3)","(2,1,4)",
                        "(3,1,1)","(3,1,2)","(3,1,3)","(3,1,4)",
                        "(4,1,1)","(4,1,2)","(4,1,3)","(4,1,4)",
                        "(5,1,1)","(5,1,2)","(5,1,3)","(5,1,4)"),sigma2,loglik,AIC)

# Order this by sigma2
d[order(d$sigma2),]

# Order this by loglik
d[order(-d$loglik),]

# Order this by AIC
d[order(d$AIC),]
```

By looking at the different orderings of the table `d`, by ascending $\hat{\sigma}^2$, descending log-likelihood, and ascending AIC, we obtain four models that consitently give the best results: $(5,1,4) \times (0,1,1)_{12}$, $(4,1,4) \times (0,1,1)_{12}$,

$(3,1,4) \times (0,1,1)_{12}$, and $(4,1,3) \times (0,1,1)_{12}$. We use Likelihood Ratio Tests in order to determine which model is the "best".

First, we compare $(4,1,4) \times (0,1,1)_{12}$ with $(5,1,4) \times (0,1,1)_{12}$. The p-value obtained is 0.7521, so we do not reject the null hypothesis that the models fit equally well. We keep the simpler one. Then, we compare $(3,1,4) \times (0,1,1)_{12}$ with $(4,1,4) \times (0,1,1)_{12}$. The p-value is 0.2111, so we do not reject the null hypothesis that the models fit equally well. We keep the simpler one. We cannot compare $(3,1,4) \times (0,1,1)_{12}$ with $(4,1,3) \times (0,1,1)_{12}$ since they have the same number of parameters. Finally, we compare $(3,1,4) \times (0,1,1)_{12}$ with $(2,1,4) \times (0,1,1)_{12}$. The p-value is less than 0.01, so we reject the null hypothesis that the models fit equally well, and keep the one with more parameters, i.e. $(3,1,4) \times (0,1,1)_{12}$.

```
# Likelihood Ratio Tests
D <- -2*(m15$loglik - m19$loglik)
pval <- 1-pchisq(D,2)
print(c("Test Statistic:",D,"P-value:",pval)) #choose m15

D <- -2*(m11$loglik - m15$loglik)
pval <- 1-pchisq(D,2)
print(c("Test Statistic:",D,"P-value:",pval)) #choose m11

# Cannot compare m11 to m14 because they have the same number of parameters

D <- -2*(m7$loglik - m11$loglik)
pval <- 1-pchisq(D,2)
print(c("Test Statistic:",D,"P-value:",pval)) #choose m11
```