MTH 427 - Spring 2023
Exam 3 - Spring 2023 - MTH 427

Conceptual problems

(Make sure to clearly show your work in order to earn full credit)

1 Exercise 1 (18 points)

For each of the following ARMA models, find the roots of the AR and MA polynomials, identify the values of p and q for which they are ARMA(p, q) (watch out for parameter redundancy), determine whether they are causal, and determine whether they are invertible.

(a) $x_t = -0.3x_{t-1} + w_t - 0.4w_{t-1} - 0.21w_{t-2}$

$$(1 + 0.3B)x_t = (1 - 0.4B - 0.21B^2)w_t$$
$$\cancel{(1 + 0.3z)} = \cancel{(1 + 0.3z)}(1 - 0.7z)$$
$$\underbrace{1}_{\phi(z)} = \underbrace{1 - 0.7z}_{\theta(z)}$$
$$x_t = w_t - 0.7w_{t-1} \quad \Longrightarrow \quad \text{ARMA}(0, 1)$$

Since $\phi(z) = 1$ has no roots, the model is causal.
Since $\theta(z)$ has one root, $-\frac{10}{3}$, and $\left|-\frac{10}{3}\right| > 1$, the model is invertible.

(b) $x_t = 3x_{t-1} + w_t + w_{t-1} - 2w_{t-2}$

$$(1 - 3B)x_t = (1 + B - 2B^2)w_t$$
$$(1 - 3z) = (1 - z)(1 + 2z)$$
$$\underbrace{1 - 3z}_{\phi(z)} = \underbrace{(1 - z)(1 + 2z)}_{\theta(z)}$$
$$x_t = 3x_{t-1} + w_t + w_{t-1} - 2w_{t-2} \quad \Longrightarrow \quad \text{ARMA}(1, 2)$$

Since $\phi(z)$ has root $\frac{1}{3} \leq 1$, the model is not causal.
Since $\theta(z)$ has roots 1 and $-\frac{1}{2}$, and the $|1| \leq 1$, the model is not invertible.

(c) $x_t = 0.5x_{t-1} + 0.5x_{t-2} + w_t - w_{t-1}$

$$(1 - 0.5B - 0.5B^2)x_t = (1 - B)w_t$$
$$(1 - 0.5z - 0.5z^2) = (1 - z)$$
$$(1 + 0.5z)\cancel{(1 - z)} = \cancel{(1 - z)}$$
$$\underbrace{1 + 0.5z}_{\phi(z)} = \underbrace{1}_{\theta(z)}$$
$$x_t + 0.5x_{t-1} = w_t \quad \Longrightarrow \quad \text{ARMA}(1, 0)$$

Since $\phi(z)$ only has root 2 and $2 > 1$, the model is causal.
Since $\theta(z)$ has no roots, the model is invertible.

2 Exercise 2 (8 points)

Suppose $x_t = w_t - w_{t-4}$, where w_t is a Gaussian white noise $W \sim N(0, \sigma_w^2)$. Find the mean and the autocovariance function (ACVF) of this series. Show that its autocorrelation function is nonzero only for lag $h = 4.(h \geq 1)$.

Solution:

The mean is zero, $\mathrm{E}[x_t] = \mathrm{E}[w_t - w_{t-4}] = \mathrm{E}[w_t] - \mathrm{E}[w_{t-4}] = 0 - 0 = 0$.

The ACVF is

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t+h}) &= \mathrm{Cov}(w_t - w_{t-4},\ w_{t+h} - w_{t+h-4}) \\ &= \underbrace{\mathrm{Cov}(w_t, w_{t+h})}_{\substack{\sigma_w^2 \text{ when } h=0 \\ 0 \text{ else}}} - \underbrace{\mathrm{Cov}(w_t, w_{t+h-4})}_{\substack{\sigma_w^2 \text{ when } h=4 \\ 0 \text{ else}}} - \underbrace{\mathrm{Cov}(w_{t-4}, w_{t+h})}_{\text{always } 0} + \underbrace{\mathrm{Cov}(w_{t-4}, w_{t+h-4})}_{\substack{\sigma_w^2 \text{ when } h=0 \\ 0 \text{ else}}}\end{aligned}$$

Since $\mathrm{Cov}(w_a, w_b)$ is zero for $a \neq b$ and σ_w^2 for $a = b$.

Restricting to $h \geq 1$ means that

$$\mathrm{ACVF} = \begin{cases} -\sigma_w^2 & h = 4 \\ 0 & h \neq 4 \end{cases} \qquad h \geq 1$$

Therefore the autocorrelation function is only nonzero at lag $h = 4$.

3 Exercise 3 (11 points)

For those models of "Exercise 1" that are causal and invertible, compute the first four coefficients $\pi_0, \pi_1, \pi_2, \pi_3$ in the invertible linear process representation $\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$
(Hint: set $\sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}$)

(a) $x_t = w_t - 0.7w_{t-1}$ $\qquad \phi(z) = 1 \qquad \theta(z) = 1 - 0.7z$

$$\begin{aligned}&\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots = \frac{\phi(z)}{\theta(z)} = \frac{1}{1 - 0.7z} \\ \iff\ &(1 - 0.7z)(\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots) = 1 \\ \iff\ &\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots - 0.7\pi_0 z - 0.7\pi_1 z^2 - 0.7\pi_2 z^3 - 0.7\pi_3 z^4 + \cdots = 1\end{aligned}$$

$$\begin{cases} \pi_0 = 1 & \implies \pi_0 = 1 \\ \pi_1 z - 0.7(1)z = 0 & \implies \pi_1 = 0.7 \\ \pi_2 z^2 - 0.7(0.7)z^2 = 0 & \implies \pi_2 = 0.49 \\ \pi_3 z^3 - 0.7(0.49)z^3 = 0 & \implies \pi_3 = 0.343 \end{cases}$$

(c) $x_t + 0.5x_{t-1} = w_t$ $\qquad \phi(z) = 1 + 0.5z \qquad \theta(z) = 1$

$$\begin{aligned} \pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots &= \frac{\phi(z)}{\theta(z)} = 1 + 0.5z \\ \iff \pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots &= 1 + 0.5z \end{aligned}$$

$$\begin{cases} \pi_0 = 1 \\ \pi_1 = 0.5 \\ \pi_2 = 0 \\ \pi_3 = 0 \end{cases}$$

4 Exercise 4 (14 points)

Exhibit an equation of the following models. (hint: use you may first use backshift operators, then expand both sides to obtain an equation in difference form.)

(a) ARIMA$(2, 1, 1)$.

$p = 2,\ d = 1,\ q = 1.$

$$\begin{aligned} & (1 - \phi_1 B - \phi_2 B^2)(1 - B)^1 x_t = (1 + \theta_1 B) w_t \\ \iff & \left(1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + \phi_2 B^3\right) x_t = (1 + \theta_1 B) w_t \\ \iff & x_t - (1 + \phi_1)x_{t-1} + (\phi_1 - \phi_2)x_{t-2} + \phi_2 x_{t-3} = w_t + \theta_1 w_{t-1} \end{aligned}$$

(b) ARIMA$(1, 1, 1) \times (0, 1, 1)_{12}$ or SARIMA$(1, 1, 1) \times (0, 1, 1)_{12}$ (notation from other text books).

$p = 1,\ d = 1,\ q = 1,\ P = 0,\ D = 1,\ Q = 1,\ S = 12.$

Using the model

$$\Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d x_t = \Theta_Q(B^s)\theta_q(B) w_t$$

with

$$\Phi_P(B^s) = 1, \quad \phi_p(B) = (1 - \phi_1 B), \quad \Theta_Q(B^s) = (1 + \Theta_1 B^{12}), \quad \theta_q(B) = (1 + \theta_1 B)$$

gives

$$1(1 - \phi_1 B)(1 - B^{12})^1(1 - B)^1 x_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B) w_t.$$

Distributing yields

$$(1 - B - B^{12} + B^{13} - \phi_1 B + \phi_1 B^2 + \phi_1 B^{13} - \phi_1 B^{14}) x_t = (1 + B\theta_1 + \Theta_1 B^{12} + \theta_1 \Theta_1 B^{13}) w_t.$$

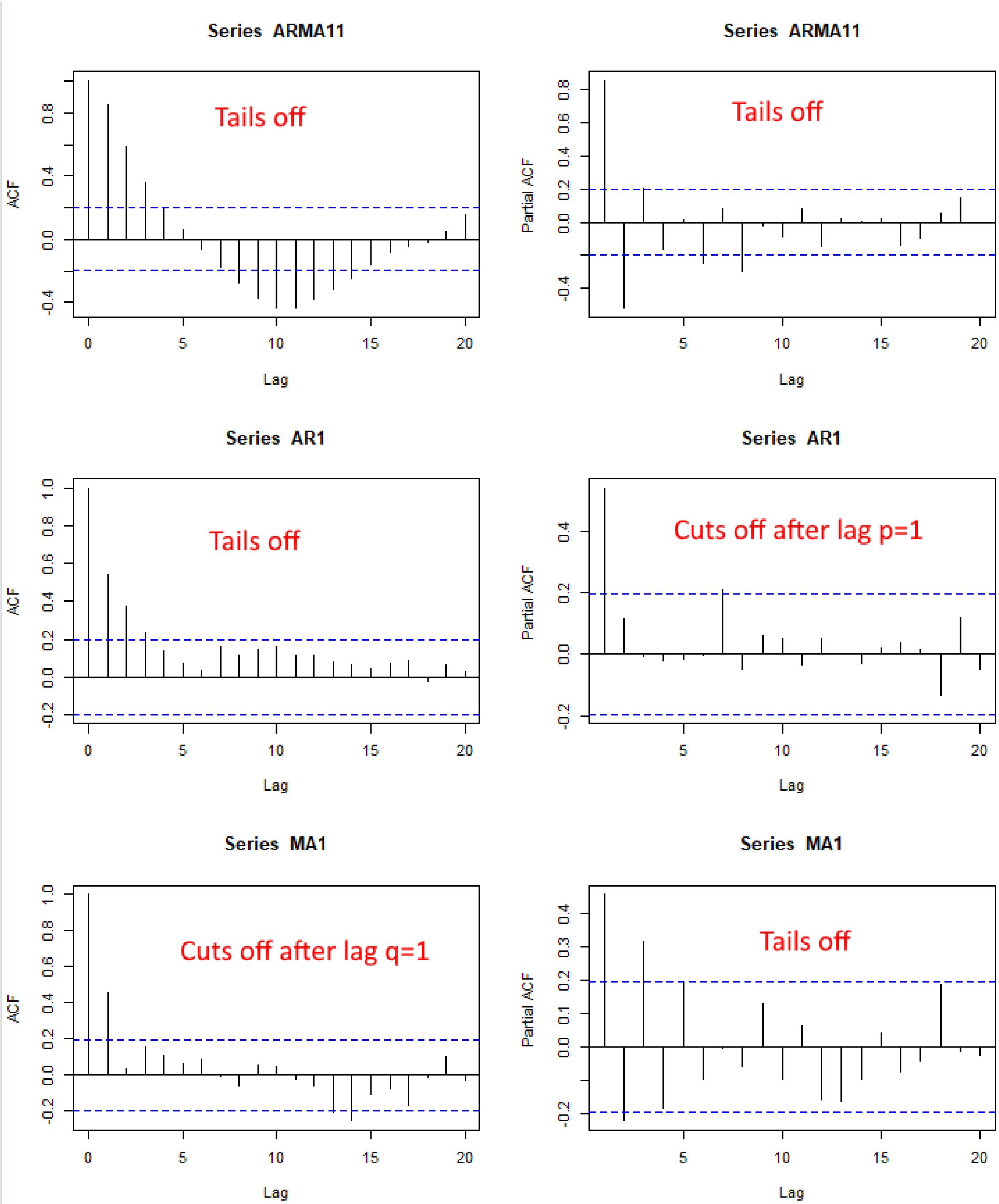
Writing as a difference equation it becomes

$$x_t - (1 + \phi_1)x_{t-1} + \phi_1 x_{t-2} - x_{t-12} + (1 + \phi_1)x_{t-13} - \phi_1 x_{t-14} = w_t + \theta_1 w_{t-1} + \Theta_1 w_{t-12} + \theta_1 \Theta_1 w_{t-13}.$$

R Project

5 Exercise 5 (8 points)

Generate $n = 100$ observations from ARMA$(1, 1)$, AR(1) and MA(1) processes respectively and plot the sample ACF and PACF of each series for $\phi = 0.6$, $\theta = 0.9$. Compare the sample ACFs and PACFs of the generated series with the results given in Table 3.1. (Are these series consistent with the table?)



From the following R code:

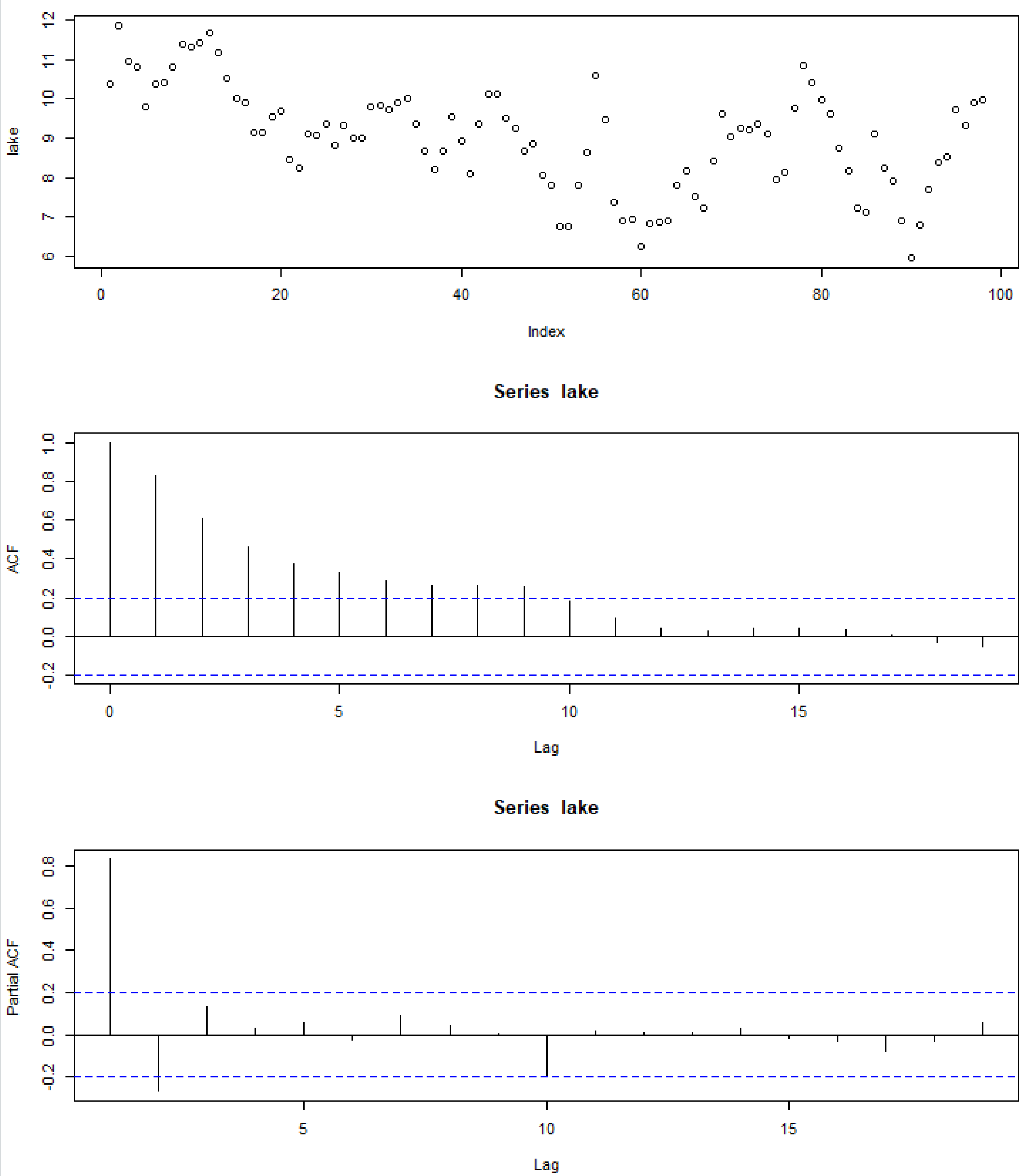
```
par(mfrow = c(3,2))

ARMA11=arima.sim(list(order=c(1,0,1), ar=.6, ma=0.9), n=100)
acf(ARMA11)
pacf(ARMA11)

AR1=arima.sim(list(order=c(1,0,0), ar=.6), n=100)
acf(AR1)
pacf(AR1)

MA1=arima.sim(list(order=c(0,0,1), ma=0.9), n=100)
acf(MA1)
pacf(MA1)
```

Compared to the table (red text), these results are consistent. When the graph says it cuts off, it *cuts* off. The AR(1) model looks a little misleading because R isn't labeling the start lag of 1, but tracking back from lag-5 the results hold true. Then all the ones that tail off slowly go towards zero, or oscillate around it.

6 Exercise 6 (13 points)

Consider the Lake Huron dataset ("LakeHuron") in R or in the package "itsmr" as "lake". It consists of $n = 98$ observations of annual levels from 1875-1972. Let's investigate if there is evidence of decline in the level of Lake Huron.

(a) Plot this series. Plot a sample **acf** and the sample **pacf** for this series.



Using the following R code:

```
library(itsmr)
par(mfrow = c(3,1))
plot(lake)
acf(lake)
pacf(lake)
```

(b) Based on the **acf** and **pacf** in part (a), fit an appropriate ARMA model using `sarima(p,d,q)` function in R, performing all necessary diagnostics. Comment.

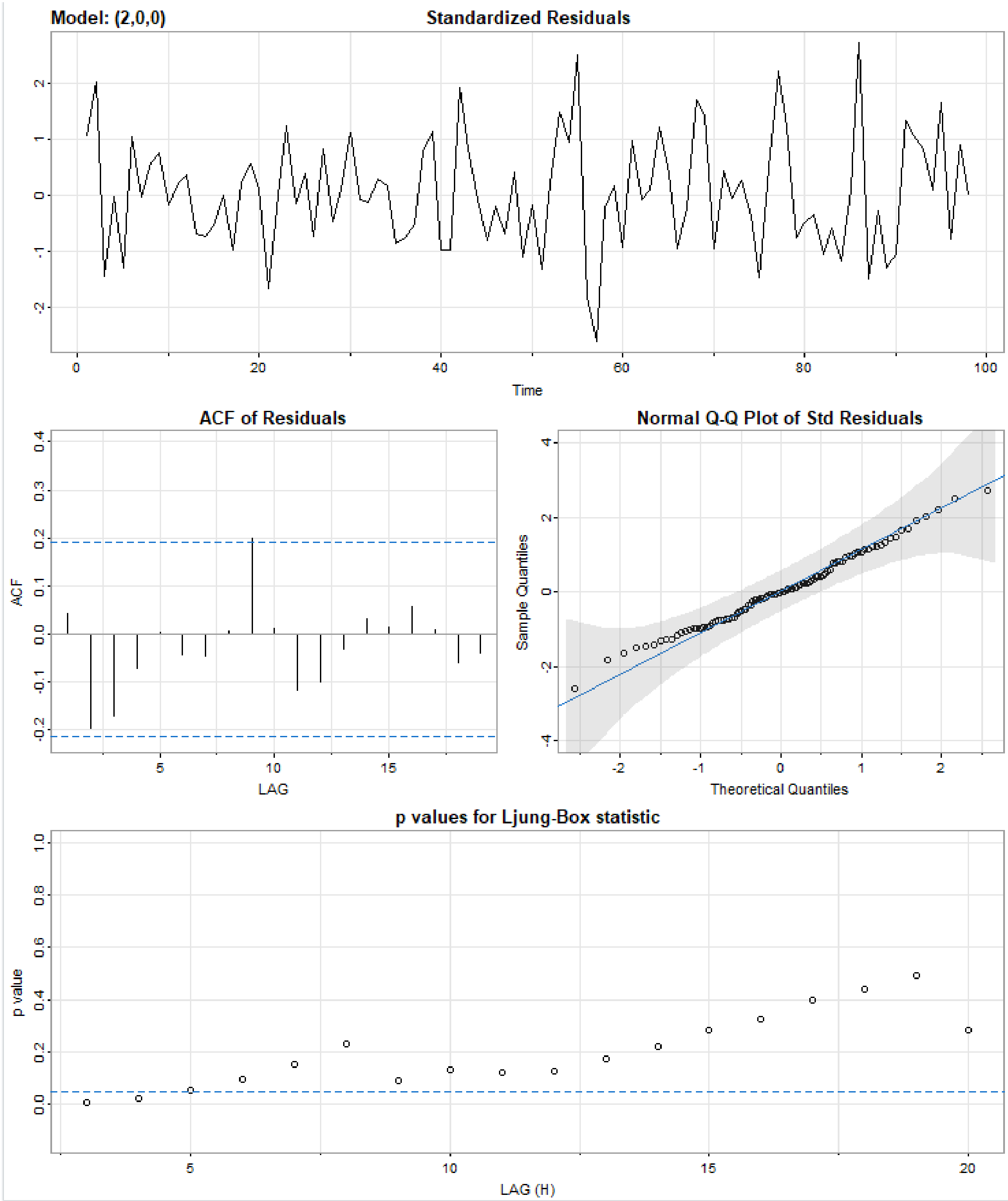
Since the ACF tails off, the model shouldn't be an MA(q). Since the PACF cuts off after lag 2, the model could be an AR(2). This can be fitted with

`sarima(lake,2,0,0, no.constant=TRUE).`

The results of this model are

```
sigma^2 estimated as 0.5443:  log likelihood = -111.85,  aic = 229.7

$degrees_of_freedom
[1] 96

$ttable
    Estimate     SE t.value p.value
ar1   1.1360 0.1020 11.1397  0.0000
ar2  -0.1393 0.1024 -1.3615  0.1766
```

whereby the p-value of one of the coefficients is 0.1766, which is relatively high, and we want low values. Looking further at the graphs for the model, the ACF of the residuals peaks outside and the Ljung-Box indicates non iid residuals since $\alpha < 0.05$ on some datapoints. The AIC was 229.7 with $\sigma^2 = 0.5443$, which is fine.

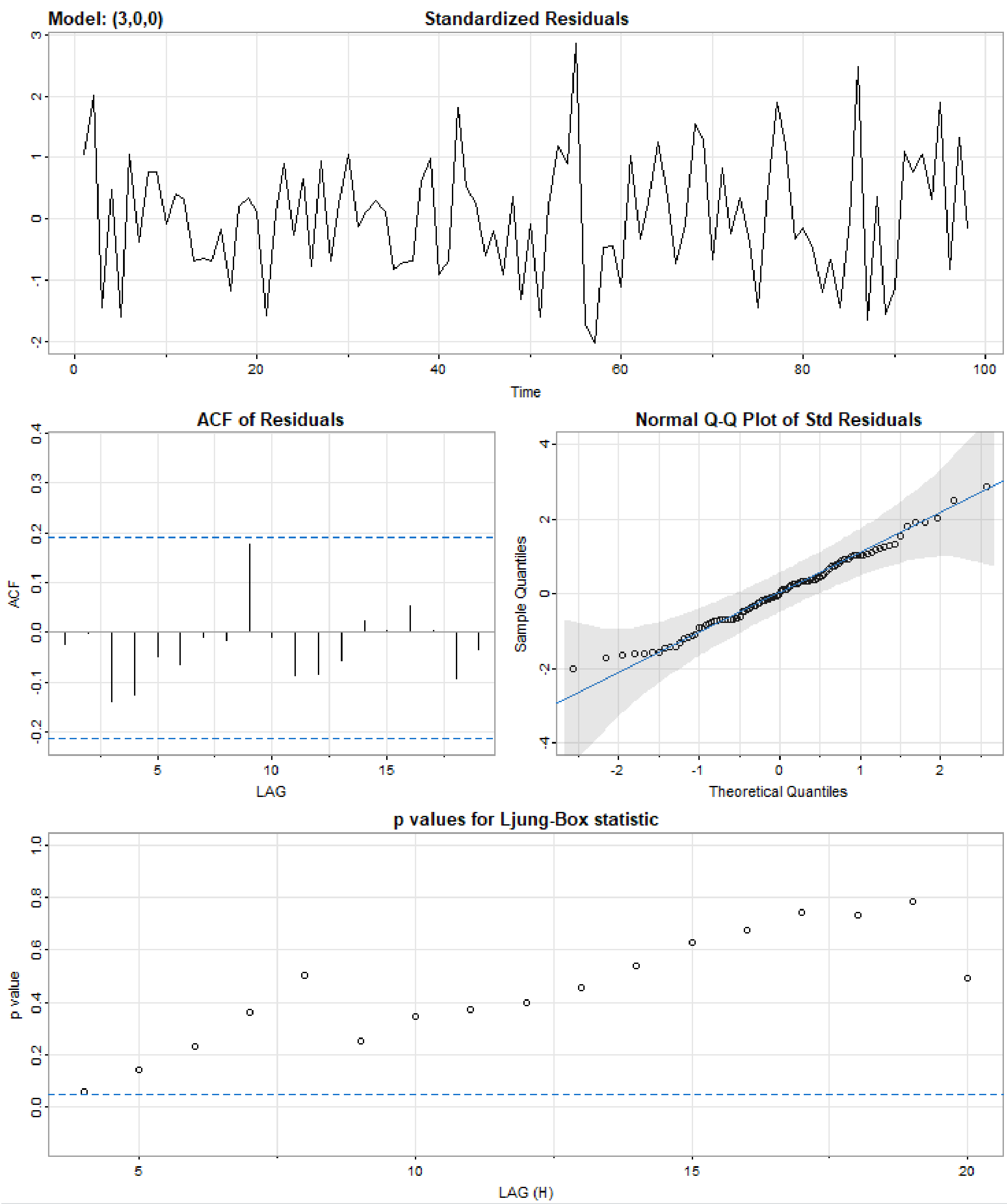


Moving on to a different model to try and get IID residuals, we can see that the PACF drops even sharper after lag 3, and has almost a large enough drop off after lag 1.

Lag 1 gives good p-values for the coeffient ($p = 0$), and has a similar $\sigma^2 = 0.5547$ and AIC of 229.54.

```
sigma^2 estimated as 0.5547:  log likelihood = -112.77,  aic = 229.54

$degrees_of_freedom
[1] 97

$ttable
    Estimate     SE  t.value p.value
ar1   0.9972 0.0036 277.2471       0
```

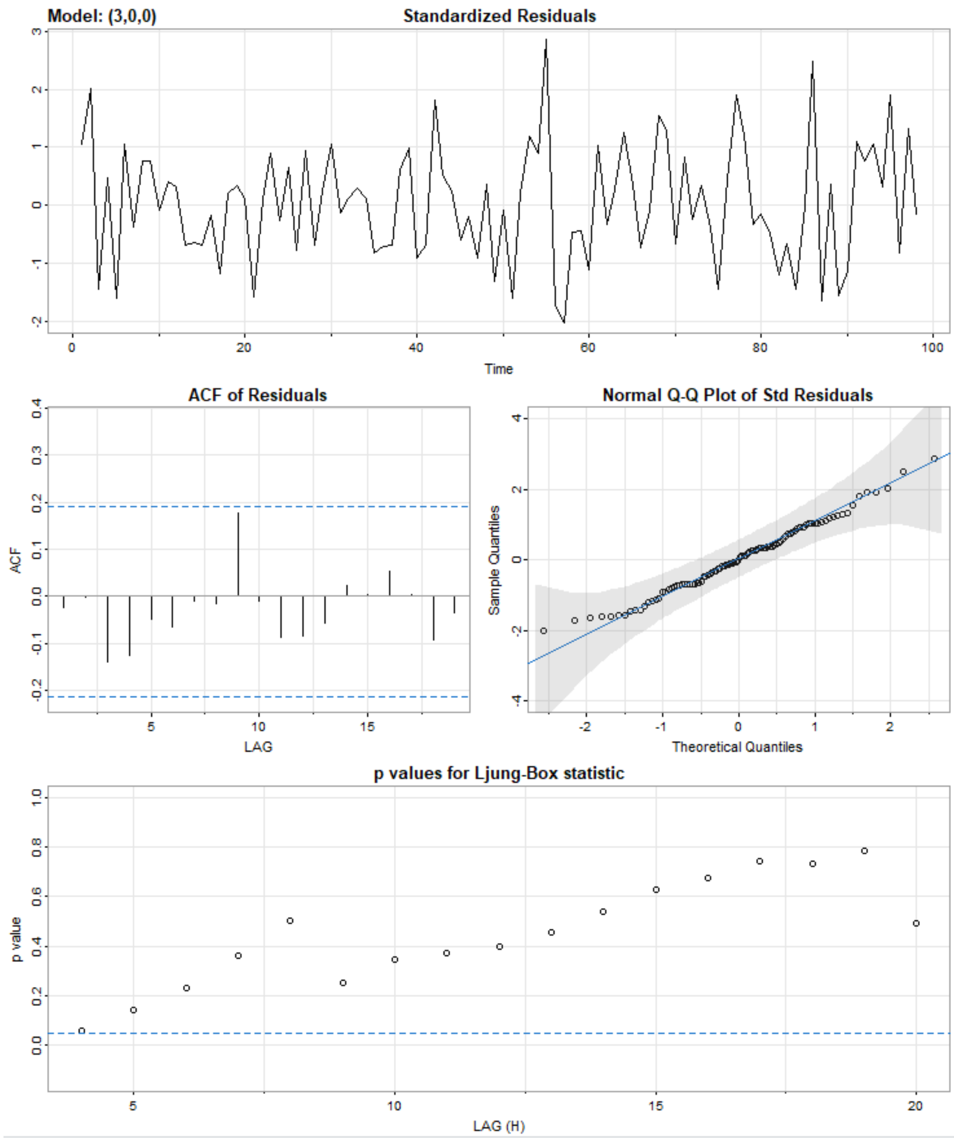
The same problems appear with the diagnostic graphs, though. The Ljung-Box statistic indicates the residuals are not iid and the ACF still spikes.



Finally we settle on Lag 3 because the coefficients are all relatively small with 0.03 being the max. σ^2 = 0.5183 is the lowest so far, and same with the AIC of 227.06.

```
sigma^2 estimated as 0.5183:  log likelihood = -109.53,  aic = 227.06

$degrees_of_freedom
[1] 95

$ttable
    Estimate     SE t.value p.value
ar1   1.1723 0.1011 11.5933  0.0000
ar2  -0.3965 0.1545 -2.5657  0.0119
ar3   0.2216 0.1015  2.1833  0.0315
```

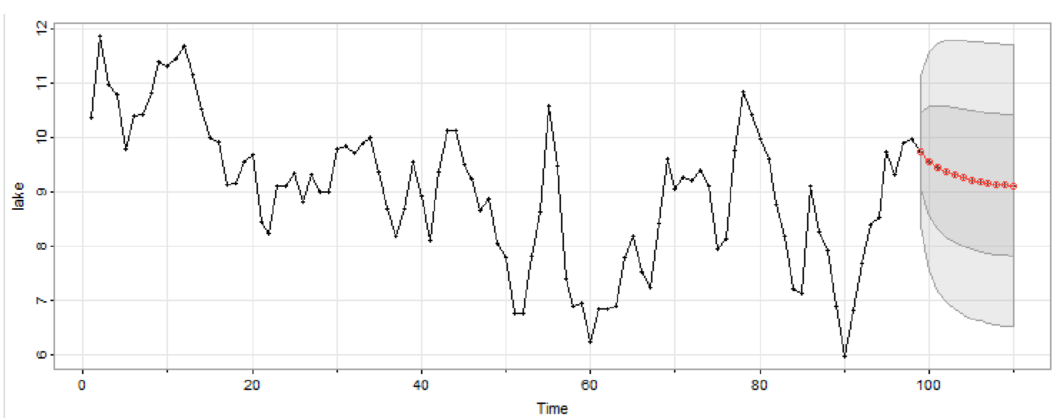
Additionally, the diagnostic graphs are promising too. The Ljung-Box statistic indicates the residuals are iid and the ACF of the residuals has the least spikes of all the tested models. The QQ plot is also a little more normal. For this reason we can conclude that AR(3) is a good fit for the model. It should be noted that while AR(4) has a better σ^2, AIC, and Ljung Box statistic, the p-values of the coefficients are bad, with one of them being 0.80. First differencing also suffered from bad p-values on the coefficients on AR(2) and AR(3), but had lower AICs around 217. Their predictions were flat lines however, so I will ignore those models



(c) After deciding on an appropriate model, forecast the data into the future 12 time periods ahead. Make sure to display the plot.

To make the prediction on the selected model, use the following code:

```
sarima.for(lake, 12, 3, 0, 0)
```

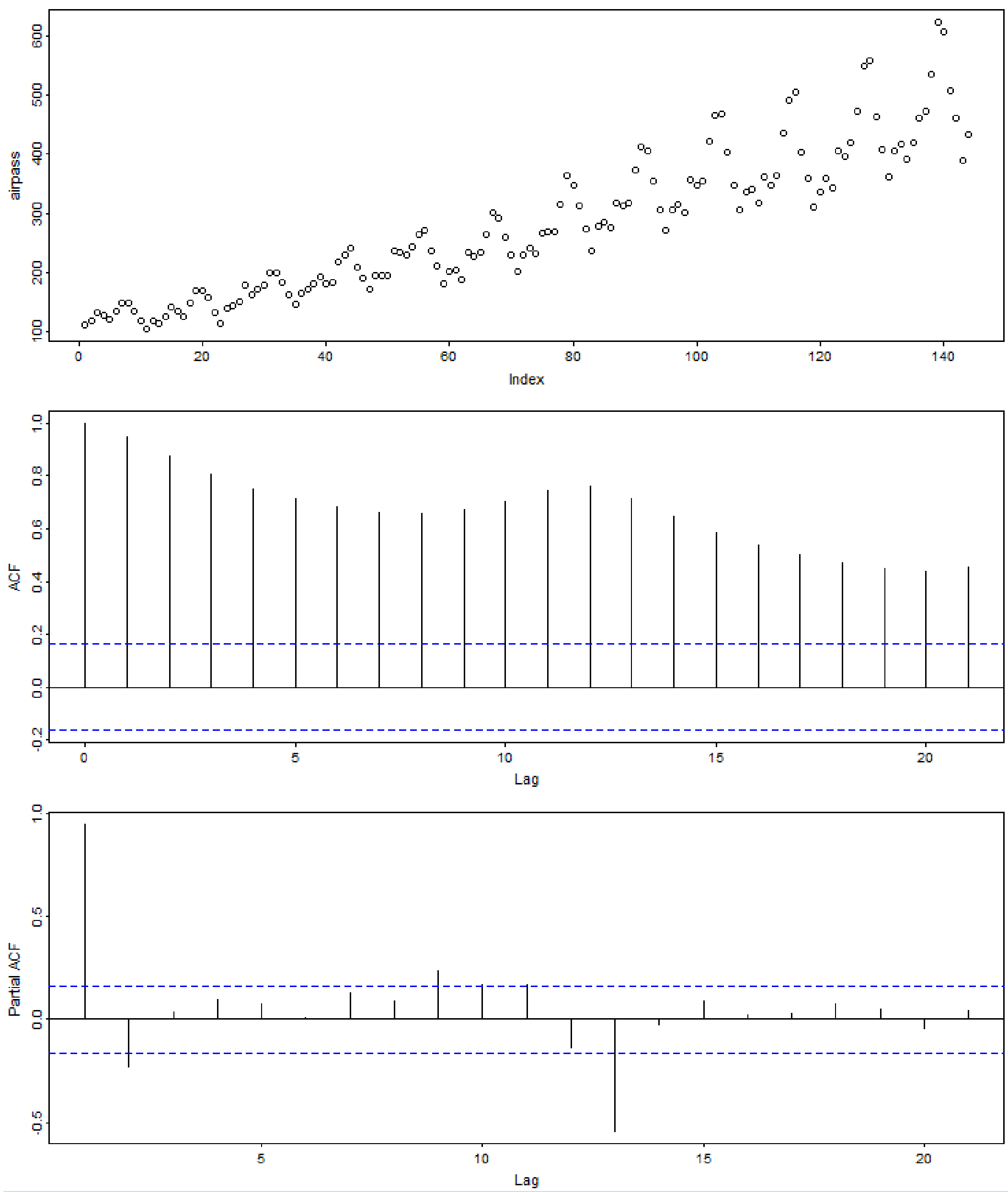


7 Exercise 7 (28 points)

Consider the Monthly Airlines Passengers numbers 1949-1960 (airpass) from the package "itsmr", or (AirPassengers) from package "astsa" in R.

(a) Display the plot of this series and its **acf** and **pacf**. Comment on the behavior of this series.

```
par(mfrow = c(3,1))
plot(airpass)
acf(airpass)
pacf(airpass)
```



It appears to have a linear increasing trend as well as increasing variance with time. Based on the ACF and PACF plots, an AR(2) seems like a good fit and MA(q) would be really bad.

(b) Is any transformation necessary? Why?

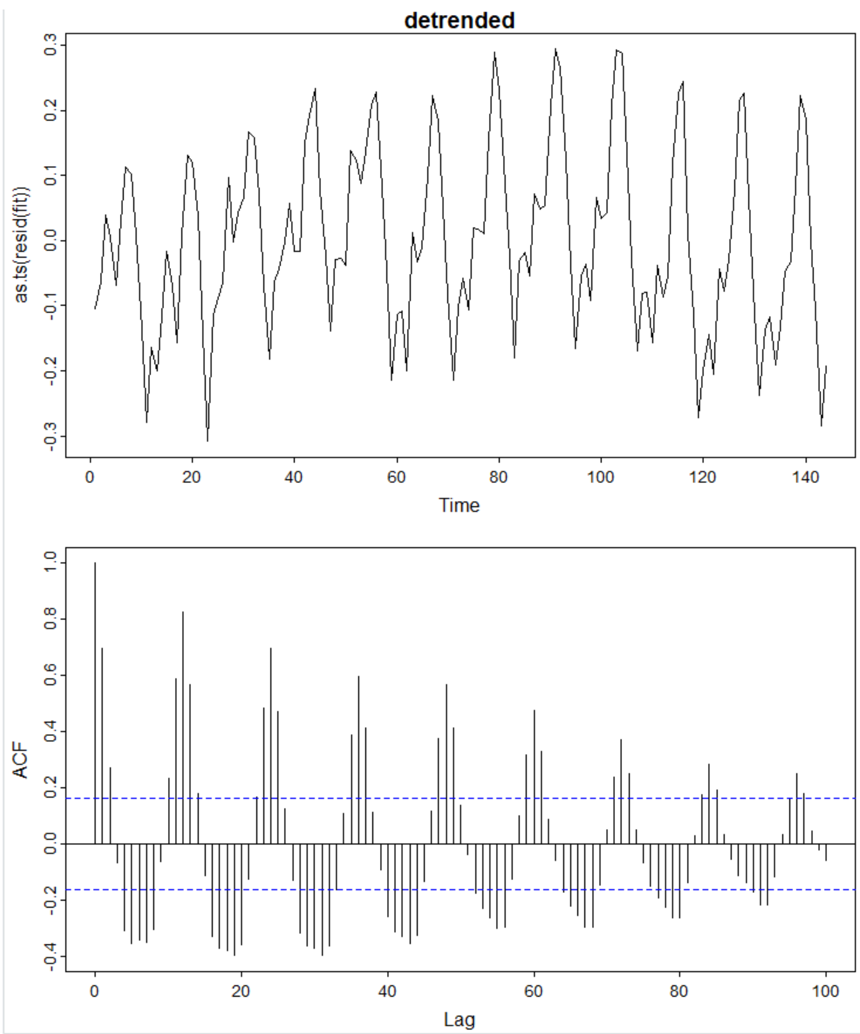
A log transformation is necessary in order to stabilize the variance. Also detrending the series should be done to remove the linear growth.

(c) Detrend the series by fitting a linear regression of the log-transformed of the series on time t. Based on the R^2 of this regression fit, can we say that there is a significant trend? Why? (Make sure to display your code and results).

Utilizing the following code:

```
stableAirpass <- log(airpass)
fit=lm(stableAirpass~time(stableAirpass))
par(mfrow=c(2,1))
plot(as.ts(resid(fit)), main="detrended")
summary(fit)
acf(resid(fit),100, main="detrended")
```

```
Residual standard error: 0.139 on 142 degrees of freedom
Multiple R-squared:  0.9015,	Adjusted R-squared:  0.9008
F-statistic:  1300 on 1 and 142 DF,  p-value: < 2.2e-16
```



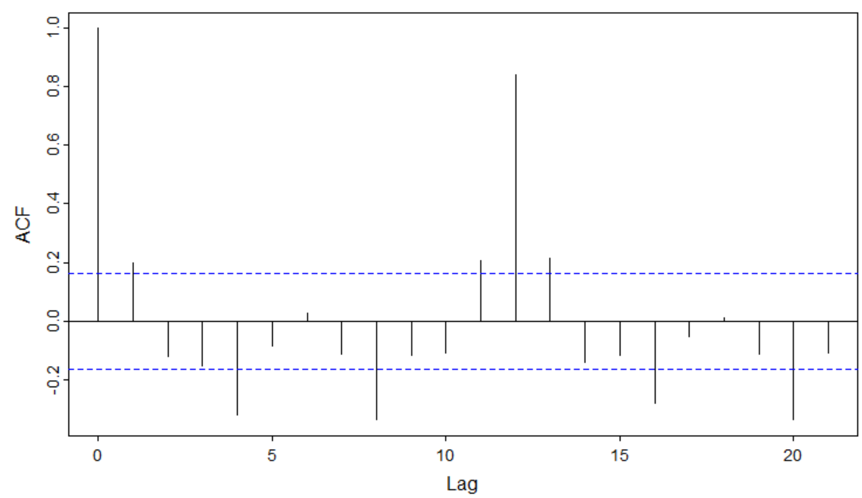
An $R^2 = 0.9$ is a strong linear trend because 90% of the variance is explained by the model.

One can also remove the trend by differencing

(d) Compute the first difference of the log-transformed of the original series. Examine its sample **acf** plot and compare it with the **acf** plot of the residuals in part (c). Make sure to display your plots.

Using the following code:

```
diffAirpass <- diff(stableAirpass)
acf(diffAirpass)
```



The ACF of the first difference has less data points spiking above the blue dashed line than the linear regression did. This indicates that the residuals of the first difference are less correlated. The first difference stabilizes near zero much faster than the linear regression model, of which slowly decreases in residual size over time. Both graphs still follow a cyclic pattern of period 12.

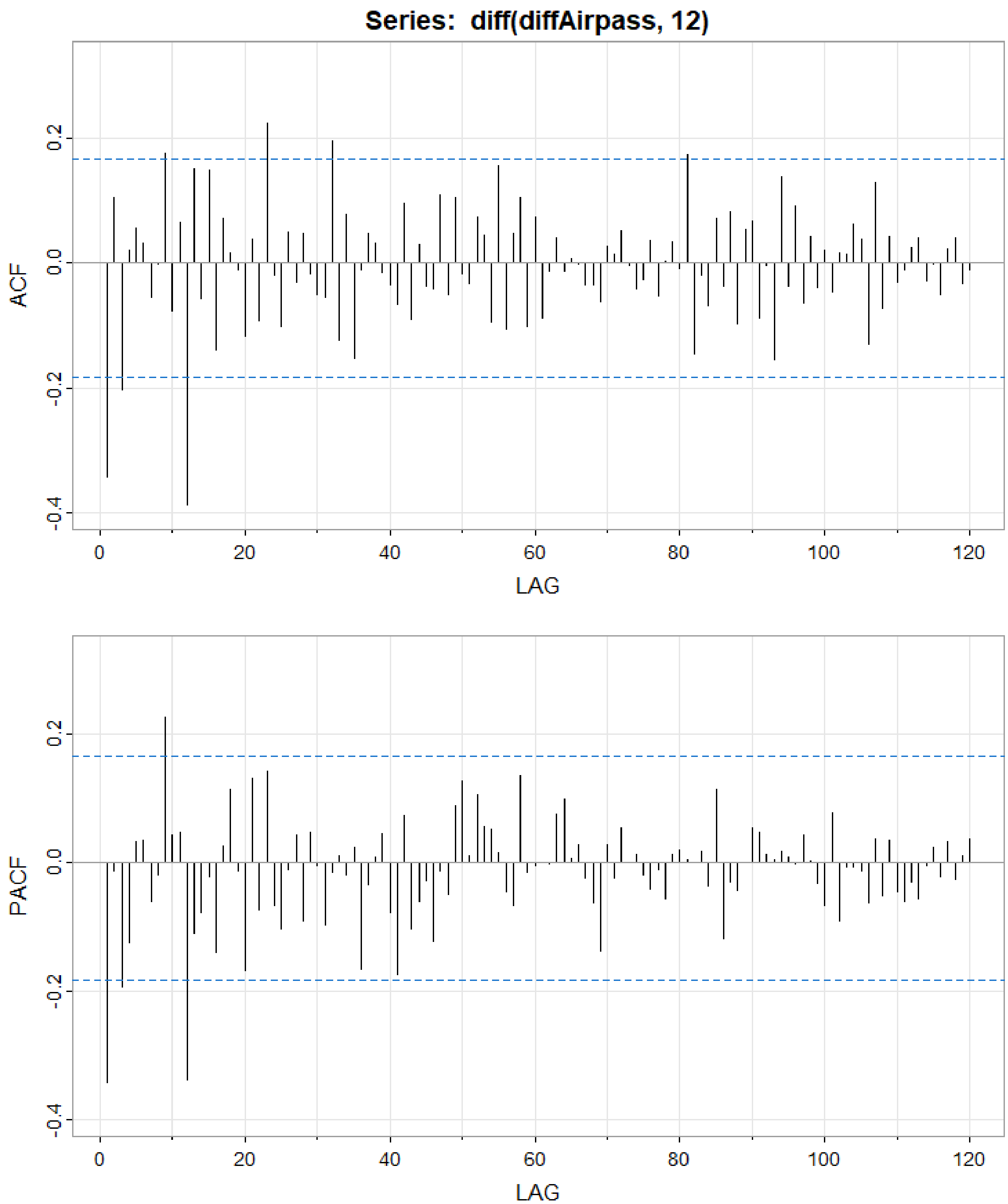
(e) Is there a strong seasonality in the series? Explain.

There is strong seasonality in the series because of the ACF plots having periodic spikes in the residuals. These occur every 12 datapoints, also known as every year (season). This can also be observed in the detrended data from part (c) whereby it follows a fairly consistent wave pattern with evenly spaced peaks.

(f) Display a sample **acf** and **pacf** of the seasonal difference of the differenced series in part (d). *hint: this line of code is given in the text file "Rcode time series-hw8..." posted in Canvas for "**unemp**" series.*

The following code was added, using the series assigned in part (d):

```
acf2(diff(diffAirpass, 12), 120)
```



(g) Based on the **acf** and **pacf** in (f), fit an appropriate seasonal ARIMA model to the log transformed Airlines Pasengers series. Make sure to explain why you select that model (model diagnostics).

For the nonseasonal component, both the ACF and PACF cut off after lag 1. This would indicate either AR(1), MA(1), or ARMA(1, 1) to be a good starting model.

For the seasonal component, the ACF and PACF both appear to cut off after lag 1, though its not a perfectly clear seasonal drop, which would also be SAR(1), SMA(1), or ARMA(1, 1).

When trying a SARIMA$(1,1,1)\times(1,1,1)_{12}$ in R, we can look at the coefficents' p-values.

```
sarima(stableAirpass, 1,1,1, 1,1,1, 12,no.constant=TRUE)
```

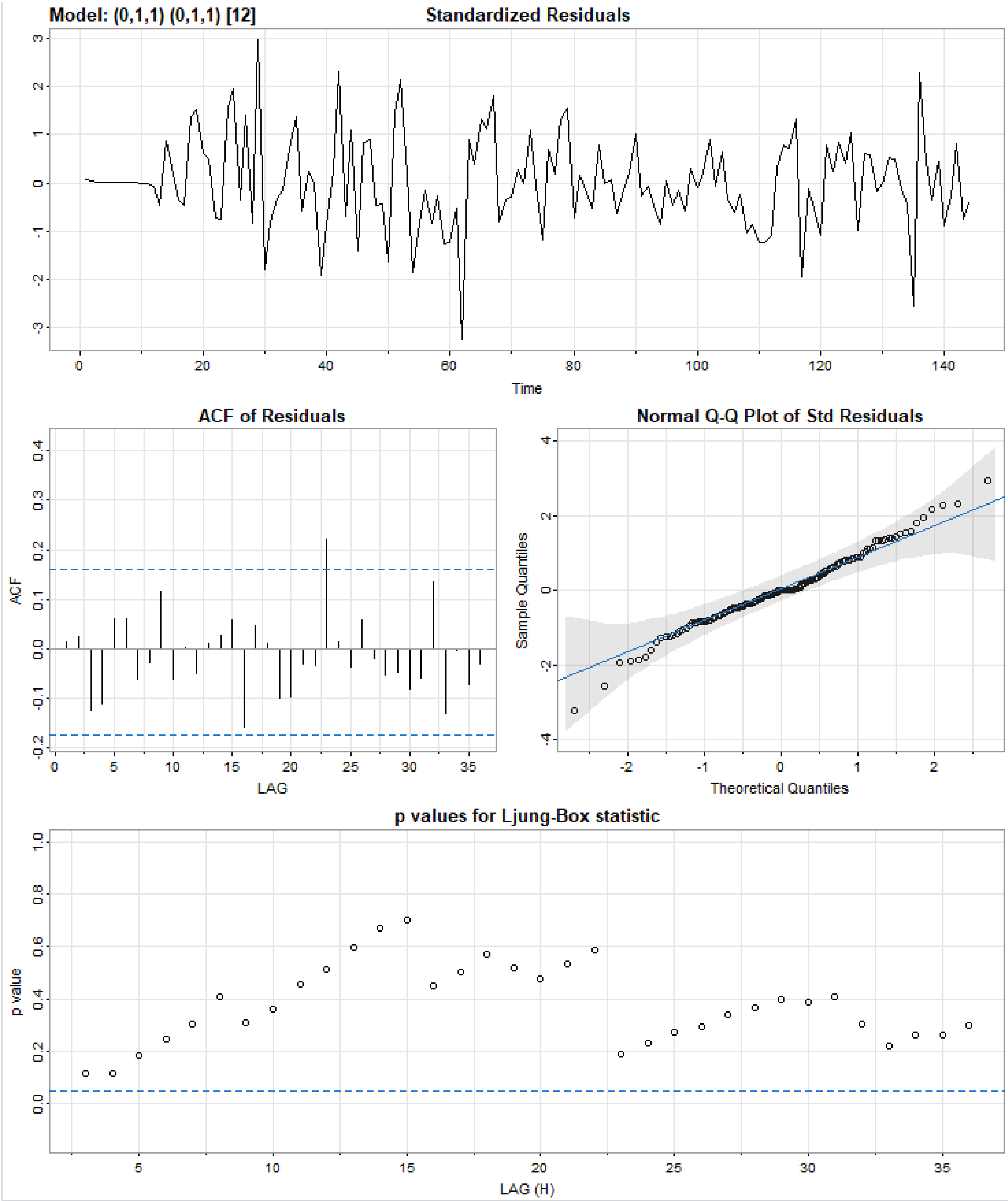
	Estimate	SE	t.value	p.value
ar1	0.1666	0.2459	0.6777	0.4992
ma1	-0.5615	0.2116	-2.6542	0.0090
sar1	-0.0990	0.1540	-0.6430	0.5214
sma1	-0.4973	0.1360	-3.6576	0.0004

Dropping the ar1 and sar1 terms due to their large p-values yields a new model

```
sarima(stableAirpass, 0,1,1, 0,1,1, 12,no.constant=TRUE)
```

	Estimate	SE	t.value	p.value
ma1	-0.4018	0.0896	-4.4825	0
sma1	-0.5569	0.0731	-7.6190	0

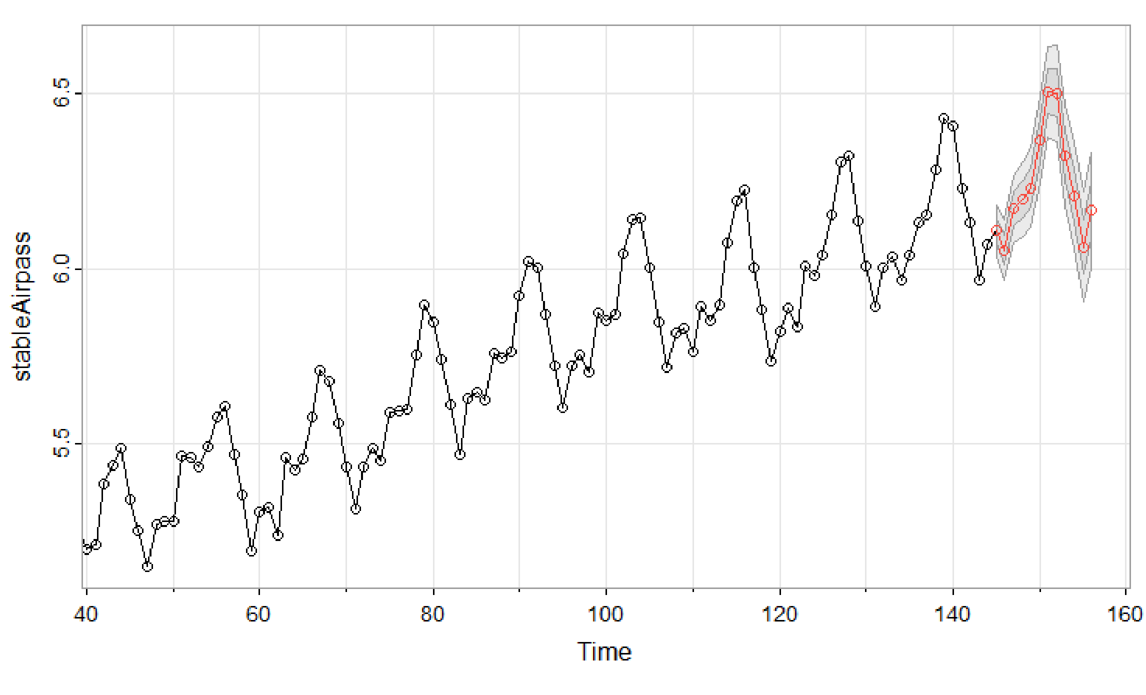
With the following diagnostic plots:



The Ljung-Box statistic being greater than $\alpha = 0.05$ indicates iid residuals and the QQ plot indicates normally distributed residuals. The R console results also show an estimated $\sigma^2 = 0.0013$ which is very low, along with an AIC of -483.4. This seems to be a very good model based on these additional numbers.

(h) Use the estimated model to forecast the next 12 values.

```
sarima.for(stableAirpass, n.ahead=12, 0, 1, 1, 0, 1, 1, 12)
```



With a console log for `$pred` indicating the following exact values:

```
6.110186
6.053775
6.171715
6.199300
6.232556
6.368779
6.507294
6.502906
6.324698
6.209008
6.063487
6.168025
```