University of Glasgow

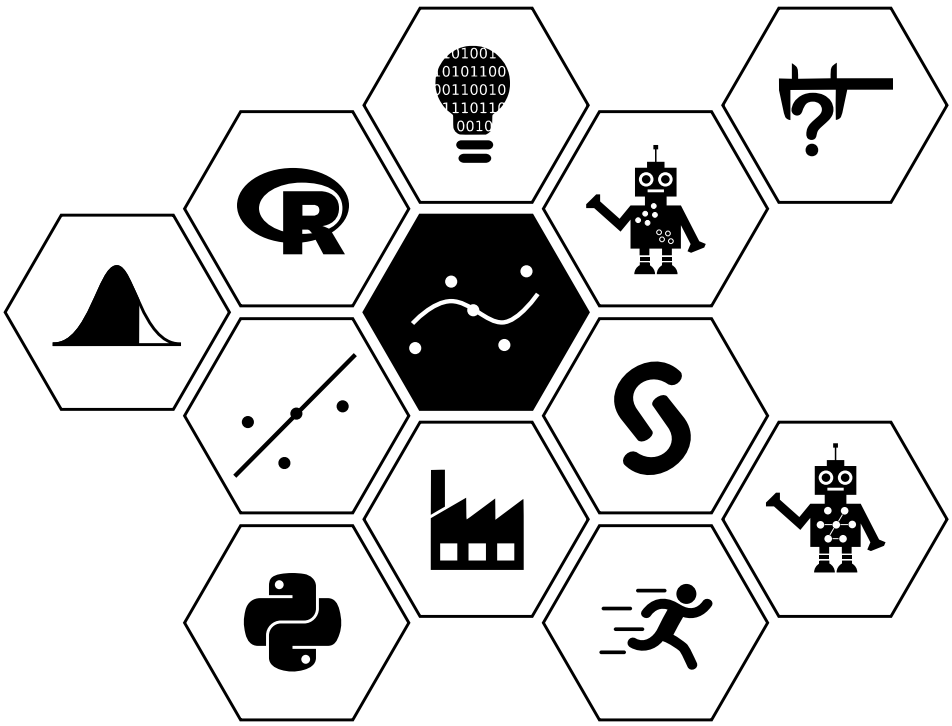
Advanced Predictive Models

Tereza Neocleous

Academic Year 2020-21

Week 7:

Autoregressive processes, moving average processes and the ARIMA approach



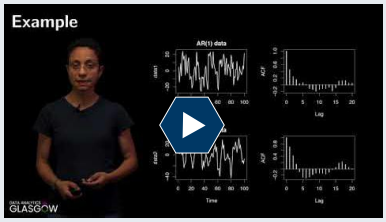
DATA ANALYTICS GLASGOW

Last week we described how to remove trend and seasonal variation from time series data, leaving a stationary residual series with short-term correlation. This week we will introduce the two main classes of time series process for stationary time series data, which differ in their short-term correlation structures.

We begin with defining an autoregressive (AR) process, and introducing the partial autocorrelation function (PACF). We then proceed to define a moving average (MA) process and describe its properties and how to identify it.

Autoregressive processes

In Week 6, we described how to remove trend and seasonal variation from time series data, leaving a stationary residual series with short-term correlation. Now we introduce an autoregressive process, and for this we assume that the time series being modelled is stationary, which can be achieved by removing the trend and working with the residual series.



Autoregressive processes

https://youtu.be/bS1y4GUvBSY

Duration: 11m55s

Definition 1 (Autoregressive process of order p).

*Let Z_t be a purely random process (i.e. each Z_t is independent) with mean 0 and variance σ_z^2. An **autoregressive process of order** p, denoted AR(p), is given by*

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t$$

where Z_t is a purely random process with mean zero and variance σ_z^2 and we assume that $X_0 = X_{-1} = \cdots = X_{-p} = 0$.

In this model correlation is introduced between the random variables by regressing X_t on past values $X_{t-1}, \ldots, X_{t-p}$. The parameters $\alpha_1, \ldots, \alpha_p$ are the coefficients of the autoregressive process, where α_1 is called the lag 1 coefficient, α_2 is called the lag 2 coefficient and so on.

This model essentially regresses X_t on its own past values, which is the reason for the prefix *auto*. We will describe the properties of autoregressive processes, discuss how to choose the order p when modelling real data, and how to estimate the coefficients $\alpha_1, \ldots, \alpha_p, \sigma_z^2$. For simplicity, we begin by describing the properties of a first order process.

AR(1) process

An AR(1) process is given by

$$X_t = \alpha X_{t-1} + Z_t$$

where Z_t is a purely random process with mean zero and variance σ_z^2 and $X_0 = 0$.

We now derive its mean, variance and correlation structure, which will tell us the conditions under which the process is stationary.

Mean of AR(1) process

To calculate the mean of the process we need to re-write it as an infinite sum of a purely random process, using back substitution as follows.

$$
\begin{aligned}
X_t &= \alpha X_{t-1} + Z_t \\
&= \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t \\
&= \alpha^2 X_{t-2} + \alpha Z_{t-1} + Z_t \\
&= \alpha^2(\alpha X_{t-3} + Z_{t-2}) + \alpha Z_{t-1} + Z_t \\
&= \alpha^3 X_{t-3} + \alpha^2 Z_{t-2} + \alpha Z_{t-1} + Z_t \\
&= \vdots \\
&= \sum_{j=0}^{\infty} \alpha^j Z_{t-j}
\end{aligned}
$$

Then the mean is given by

$$
\begin{aligned}
E(X_t) &= E(\sum_{j=0}^{\infty} \alpha^j Z_{t-j}) \\
&= \sum_{j=0}^{\infty} \alpha^j E(Z_{t-j}) \\
&= \sum_{j=0}^{\infty} \alpha^j \times 0 \\
&= 0
\end{aligned}
$$

So regardless of the value of the lag one coefficient α, an AR(1) process has mean zero.

Variance of AR(1) process

The variance can also be calculated by writing X_t as an infinite sum of a purely random process.

$$
\begin{aligned}
\mathrm{Var}(X_t) &= \mathrm{Var}(\sum_{j=0}^{\infty} \alpha^j Z_{t-j}) \\
&= \sum_{j=0}^{\infty} \mathrm{Var}(\alpha^j Z_{t-j}) \\
&= \sum_{j=0}^{\infty} \alpha^{2j} \mathrm{Var}(Z_{t-j}) \\
&= \sum_{j=0}^{\infty} \alpha^{2j} \times \sigma_z^2 \\
&= \sigma_z^2 \sum_{j=0}^{\infty} \alpha^{2j}
\end{aligned}
$$

There are no covariance terms in the above expression because each Z_t is independent as it is a purely random process. The variance is obtained from an infinite sum, so its value depends on α. Two cases are possible.

1. If $|\alpha| \geq 1$ then $\mathrm{Var}(X_t) = \infty$.
2. If $|\alpha| < 1$ then $\alpha^2 < 1$, and we know that for a geometric progression

$$
\sum_{j=0}^{\infty} \alpha^{2j} = 1 + \alpha^2 + \alpha^4 + \alpha^6 + \ldots = \frac{1}{1-\alpha^2}
$$

so we have

$$\text{Var}(X_t) = \sigma_z^2 \sum_{j=0}^{\infty} \alpha^{2j} = \frac{\sigma_z^2}{1-\alpha^2}$$

Notes

1. As a weakly stationary process must have a finite constant variance, an AR(1) process is not stationary if $|\alpha| \geq 1$.
2. To prove that an AR(1) process is stationary for $|\alpha| < 1$, we need to check the autocorrelation function does not depend on t.

Autocorrelation function of AR(1) process

The autocovariance function at lag τ is calculated as follows.

$$\begin{aligned}
\text{Cov}(X_t, X_{t+\tau}) &= E(X_t X_{t+\tau}) - E(X_t)E(X_{t+\tau}) \\
&= E(X_t X_{t+\tau}) \\
&= E\left[\sum_{j=0}^{\infty} (\alpha^j Z_{t-j}) \times \sum_{k=0}^{\infty} \left(\alpha^k Z_{t-k+\tau}\right)\right] \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \alpha^j \alpha^k E\left(Z_{t-j} Z_{t-k+\tau}\right)
\end{aligned}$$

Multiplying out the above sum will give pairs $E(Z_r Z_s)$ (excluding constants).

- For any $r \neq s$ we have $E(Z_r Z_s) = E(Z_r)E(Z_s) = 0$ because the Z_t are independent.
- For $r = s$ we have $E(Z_r^2) = \text{Var}(Z_r) + [E(Z_r)]^2 = \sigma_z^2$.

Therefore the only pairs of terms that remain are when $j = k + \tau$, simplifying the sum to

$$\begin{aligned}
\text{Cov}(X_t, X_{t+\tau}) &= \sigma_z^2 \sum_{j=0}^{\infty} \alpha^j \alpha^{j+\tau} \\
&= \sigma_z^2 \alpha^\tau \sum_{j=0}^{\infty} \alpha^{2j}
\end{aligned}$$

Therefore the covariance is also infinite for $|\alpha| \geq 1$. For $|\alpha| < 1$ the covariance is given by

$$\gamma_\tau = \text{Cov}(X_t, X_{t+\tau}) = \frac{\sigma_z^2 \alpha^\tau}{1-\alpha^2}$$

The correlation function is therefore given by

$$\rho_\tau = \text{Corr}(X_t, X_{t+\tau}) = \frac{\gamma_\tau}{\gamma_0} = \frac{\sigma_z^2 \alpha^\tau/(1-\alpha^2)}{\sigma_z^2/(1-\alpha^2)} = \alpha^{|\tau|}$$

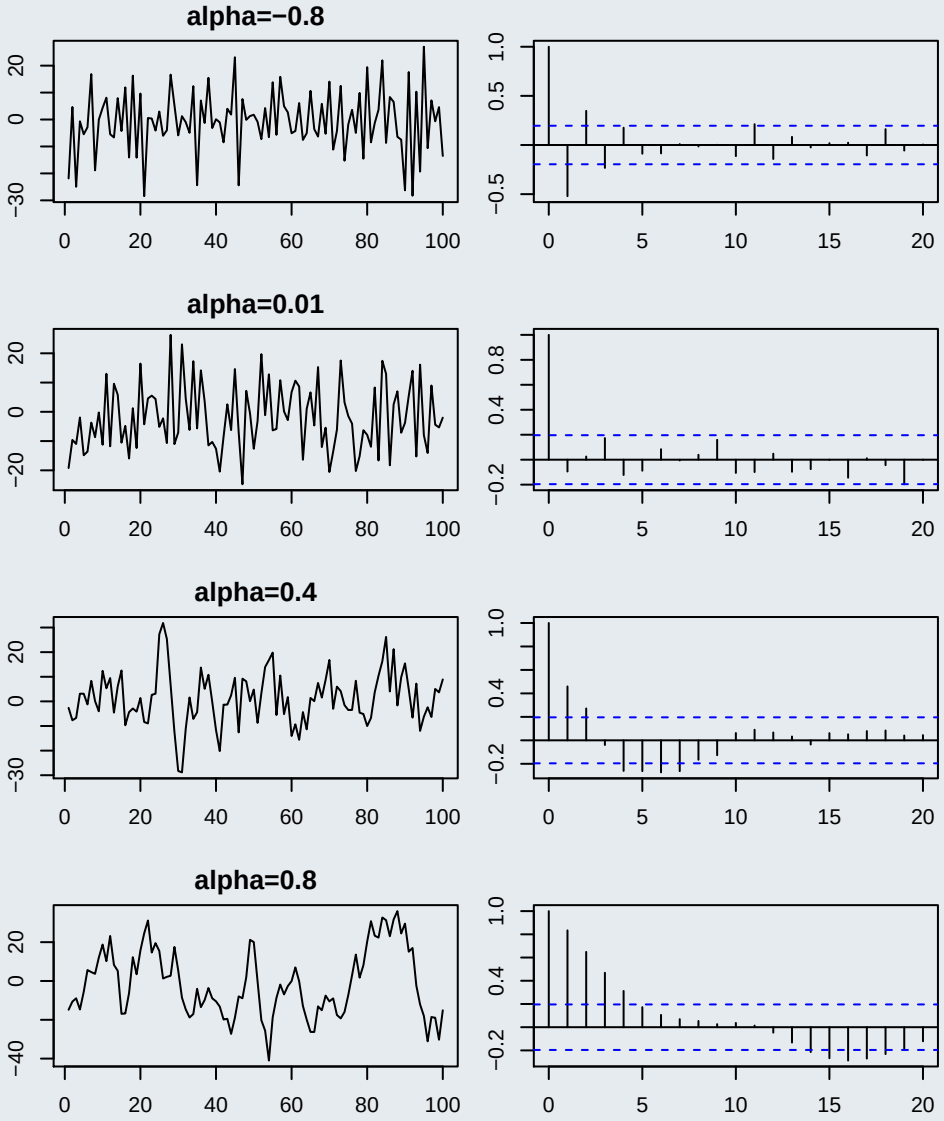
where the modulus operator is required because $\gamma_\tau = \gamma_{-\tau}$. Therefore if $|\alpha| < 1$ the mean and variance are constant and finite, and the autocorrelation function only depends on τ. Hence the process is weakly stationary.

Example 1.

Here are some time plots and correlograms of AR(1) processes with different values of α. The processess are simulated and graphed using the following R code.

```
data1 <- arima.sim(model=list(ar=-0.8), n=100, sd=10)
data2 <- arima.sim(model=list(ar=0.01), n=100, sd=10)
data3 <- arima.sim(model=list(ar=0.4), n=100, sd=10)
```

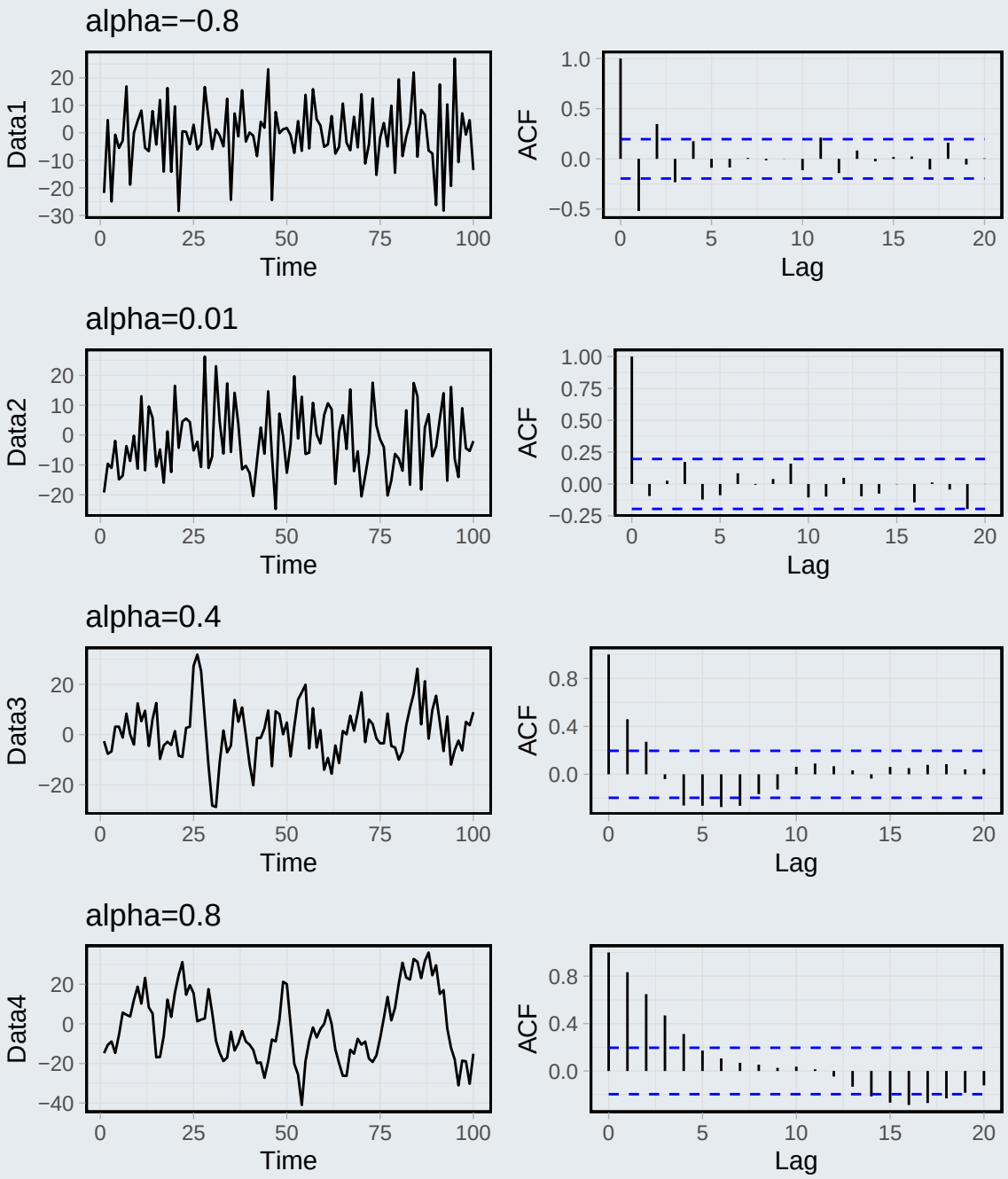
```
data4 <- arima.sim(model=list(ar=0.8), n=100, sd=10)

par(mfrow=c(4,2), mar=c(3,2,2,1))
plot(data1, main="alpha=-0.8")
acf(data1, main =" ")
plot(data2, main="alpha=0.01")
acf(data2, main =" ")
plot(data3, main="alpha=0.4")
acf(data3, main =" ")
plot(data4, main="alpha=0.8")
acf(data4, main =" ")
```



Alternatively, we can use `library(ggfortify)` to plot the time series using `ggplot` graphics:

```
library(ggfortify)
library(gridExtra)
p1 <- autoplot(data1, xlab = "Time", ylab = "Data1", main = "alpha=-0.8")
p1acf <- autoplot(acf(data1, plot = FALSE), main = "")

p2 <- autoplot(data2, xlab = "Time", ylab = "Data2", main = "alpha=0.01")
p2acf <- autoplot(acf(data2, plot = FALSE), main = "")

p3 <- autoplot(data3, xlab = "Time", ylab = "Data3", main = "alpha=0.4")
p3acf <- autoplot(acf(data3, plot = FALSE), main = "")
```

```
p4 <- autoplot(data4, xlab = "Time", ylab = "Data4", main = "alpha=0.8")
p4acf <- autoplot(acf(data4, plot = FALSE), main = "")

grid.arrange(p1, p1acf, p2, p2acf, p3, p3acf, p4, p4acf, nrow=4)
```



Notes

1. A random walk process occurs when $\alpha = 1$ and is hence not stationary.
2. A purely random process occurs when $\alpha = 0$ and is hence stationary.

Supplementary material:
Yule-Walker equations

Conditional on the mean and variance of the AR(1) process calculated above, we can derive the covariance in a shorter way assuming that the process is stationary.

$$\begin{aligned} X_t &= \alpha X_{t-1} + Z_t \\ X_{t-k}X_t &= \alpha X_{t-k}X_{t-1} + X_{t-k}Z_t \\ E(X_{t-k}X_t) &= \alpha E(X_{t-k}X_{t-1}) + E(X_{t-k}Z_t) \\ E(X_{t-k}X_t) &= \alpha E(X_{t-k}X_{t-1}) \end{aligned}$$

Writing $X_{t-k} = \sum_{j=0}^{\infty} \alpha^j Z_{t-k-j}$, the last line is true because

$$\begin{aligned} E(X_{t-k}Z_t) &= E(\sum_{j=0}^{\infty} \alpha^j Z_{t-k-j}Z_t) \\ &= \sum_{j=0}^{\infty} \alpha^j E(Z_{t-k-j}Z_t) \\ &= 0 \end{aligned}$$

Furthermore, as $E(X_t) = 0$ then

$$\gamma_{-k} = E(X_{t-k}X_t) - E(X_t)E(X_{t-k}) = E(X_{t-k}X_t)$$

so that the above equation becomes

$$\gamma_{-k} = \alpha\gamma_{-k+1} \qquad \text{and} \qquad \rho_{-k} = \alpha\rho_{-k+1}$$

Finally, as both the autocovariance and autocorrelation functions are even, i.e. $\gamma_{-k} = \gamma_k$ we have

$$\gamma_k = \alpha\gamma_{k-1} \qquad \text{and} \qquad \rho_k = \alpha\rho_{k-1}$$

This recursive relationship is a special case of the **Yule-Walker equations**, named after G.U. Yule and G. Walker. The equations are one method used to calculate the autocorrelation function for a general AR(p) model. The equations can be used to determine the autocorrelations by noting that $\rho_0 = 1$, which leads to $\rho_1 = \alpha$ and so on.

AR(p) processes

An AR(p) process is given by

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t$$

where Z_t is a purely random process with mean zero and variance σ_z^2. It can be written in the following alternative form:

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t \\ X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} &= Z_t \\ (1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p)X_t &= Z_t \\ \phi(B)X_t &= Z_t \end{aligned}$$

where B is the backshift operator. The function $\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$ is called the **characteristic polynomial** and is important for determining whether the process is stationary.

Mean of AR(p) process

We can use the **invertibility theorem** to find the mean of an AR(p) process.

Theorem 1 (Invertibility).

Any AR(p) process can be written as an infinite sum of a purely random process

$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j}$.

Note that the relationship between the α's and β's can be hard (algebraically messy) to find.

We can use the invertibility theorem to show that the mean of an autoregressive process is zero:

$$\begin{aligned} E(X_t) &= E(\sum_{j=0}^{\infty} \beta_j Z_{t-j}) \\ &= \sum_{j=0}^{\infty} \beta_j E(Z_{t-j}) \\ &= 0 \end{aligned}$$

Stationarity for AR(p) processes

The variance and autocorrelation function can be calculated by determining the β_j's, although these are hard to calculate. Instead, we calculate the variance and autocorrelation function conditional on the process being stationary. The following theorem tells us when an AR(p) process is stationary.

Theorem 2 (Stationarity).

Write an AR(p) process as

$$\phi(B) X_t = Z_t$$

*where $\phi(B)$ is the characteristic polynomial $\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$. Then the process is stationary if the roots of the **characteristic equation***

$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p = 0$$

have modulus greater than 1, i.e. they lie outside the unit circle. Here we consider B as the variable of the polynomial equation.

Example 2.

To work out for which values of α the AR(1) process $X_t = \alpha X_{t-1} + Z_t$ is stationary, we need to look at the roots of the characteristic polynomial. The characteristic polynomial is given by $\phi(B) = 1 - \alpha B$, which has a single root $1/\alpha$. Hence the process is stationary when $1/\alpha$ is greater than 1 in absolute value, i.e. when $|\alpha| < 1$.

Example 3.

The AR(2) process $X_t = X_{t-1} - 0.5 X_{t-2} + Z_t$ is stationary. The characteristic equation is given by $1 - B + 0.5B^2 = 0$. Using the quadratic formula gives

$$\begin{aligned}
\text{roots} &= \frac{1 \pm \sqrt{(-1)^2 - 4 \times 0.5 \times 1}}{2 \times 0.5} \\
\text{roots} &= \frac{1 \pm \sqrt{-1}}{1} \\
\text{roots} &= 1 + i \text{ and } 1 - i
\end{aligned}$$

where i is the complex part of the number, i.e. $i = \sqrt{-1}$. Therefore as both roots have modulus greater than one the process is stationary.

Task 1.

Is the AR(2) process $X_t = 0.75X_{t-1} - 0.125X_{t-2} + Z_t$ stationary?

Variance of AR(p) process

The variance of an AR(p) process is only finite if the process is stationary. For a stationary AR(p) process the variance is given by

$$\text{Var}[X_t] = \sigma_z^2 + \sum_{\tau=1}^{p} \alpha_\tau \gamma_\tau.$$

Note that the variance depends on the autocovariance functions at lags 1 to p.

Supplementary material:
Calculation of the variance of an AR(p) process

Assuming the process is stationary, we calculate its variance as follows.

$$\begin{aligned}
X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t \\
X_t^2 &= \alpha_1 X_t X_{t-1} + \ldots + \alpha_p X_t X_{t-p} + X_t Z_t \\
E(X_t^2) &= \alpha_1 E(X_t X_{t-1}) + \ldots + \alpha_p E(X_t X_{t-p}) + E(X_t Z_t) \\
E(X_t^2) &= \alpha_1 E(X_t X_{t-1}) + \ldots + \alpha_p E(X_t X_{t-p}) + \sigma_z^2
\end{aligned}$$

The last line holds because

$$\begin{aligned}
E(X_t Z_t) &= E[(\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t) Z_t] \\
&= E[(\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}) Z_t] + E(Z_t^2) \\
&= E(\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}) E(Z_t) + \text{Var}(Z_t) + [E(Z_t)]^2 \\
&= \sigma_z^2
\end{aligned}$$

Therefore as $E(X_t) = 0$ we have that

$\text{Var}(X_t) = E(X_t^2) - E(X_t)^2 = E(X_t^2)$

and

$\gamma_\tau = \text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(X_t, X_{t-\tau}) = E(X_t X_{t-\tau}) - E(X_t)E(X_{t-\tau}) = E(X_t X_{t-\tau}).$

so that the variance simplifies to

$$\text{Var}[X_t] = \sigma_z^2 + \sum_{\tau=1}^{p} \alpha_\tau \gamma_\tau.$$

Autocorrelation function of AR(p) process

The autocovariance function is given by the recursive equation

$$\gamma_\tau = \alpha_1\gamma_{\tau-1} + \ldots + \alpha_p\gamma_{\tau-p}.$$

Dividing by the variance gives the autocorrelation function as

$$\rho_\tau = \alpha_1\rho_{\tau-1} + \ldots + \alpha_p\rho_{\tau-p}$$

which is called the **Yule-Walker** equation.

Supplementary material:
Calculation of the autocovariance and autocorrelation function of an AR(p) process

The autocovariance and autocorrelation functions can be calculated in the same way as the variance.

$$\begin{aligned}
X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t \\
X_t X_{t-\tau} &= \alpha_1 X_{t-\tau} X_{t-1} + \ldots + \alpha_p X_{t-\tau} X_{t-p} + X_{t-\tau} Z_t \\
E(X_t X_{t-\tau}) &= \alpha_1 E(X_{t-\tau} X_{t-1}) + \ldots + \alpha_p E(X_{t-\tau} X_{t-p}) + E(X_{t-\tau} Z_t) \\
E(X_t X_{t-\tau}) &= \alpha_1 E(X_{t-\tau} X_{t-1}) + \ldots + \alpha_p E(X_{t-\tau} X_{t-p})
\end{aligned}$$

As $\gamma_{-\tau} = \gamma_\tau$, the autocovariance function is given by the recursive equation

$$\gamma_\tau = \alpha_1\gamma_{\tau-1} + \ldots + \alpha_p\gamma_{\tau-p}.$$

Dividing by the variance gives the autocorrelation function as

$$\rho_\tau = \alpha_1\rho_{\tau-1} + \ldots + \alpha_p\rho_{\tau-p}$$

which is called the **Yule-Walker** equation.

These equations can be written in matrix form for $\tau = 1, 2, \ldots, p$ as

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

Note: The Yule-Walker equations can be used repeatedly to calculate the autocorrelation function for any AR(p) process. However, as p increases the equations get more complex.

Model identification

Given a stationary time series how do you determine:

(a) if an autoregressive process is an appropriate model, and

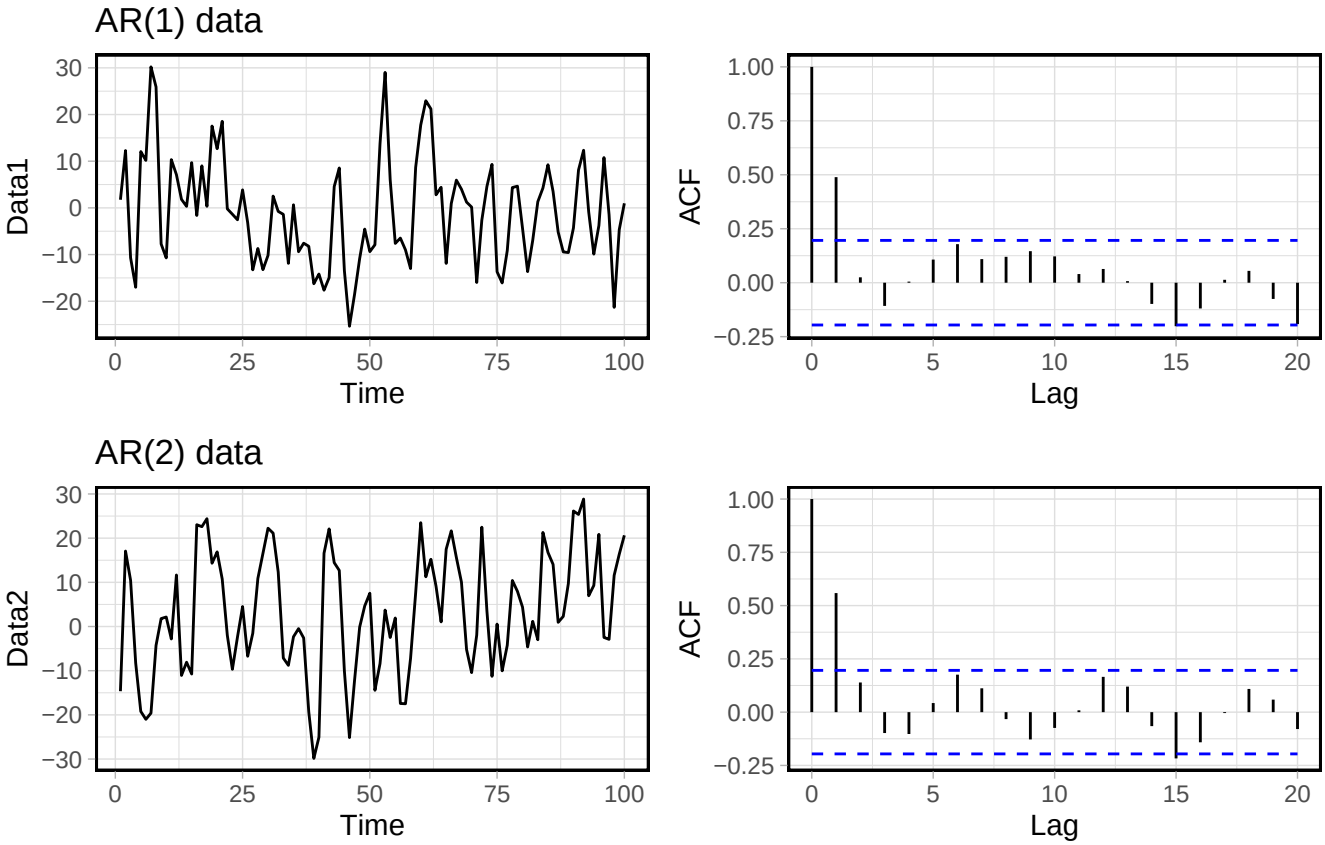
(b) if it is, which order process should you use (i.e. which value of p)?

The obvious solution would be to base your decision on the correlogram of the time series, but this does not help. Let us illustrate this with an example, by simulating AR(1) and AR(2) data.

Task 2.

Can you work out the equations of the AR models simulated using the following code?

```
data1 <- arima.sim(model=list(ar=0.6), n=100, sd=10)
data2 <- arima.sim(model=list(ar=c(0.75, -0.125)), n=100, sd=10)

p1 <- autoplot(data1, xlab = "Time", ylab = "Data1", main = "AR(1) data")
p1acf <- autoplot(acf(data1, plot = FALSE), main = "")
p2 <- autoplot(data2, xlab = "Time", ylab = "Data2", main = "AR(2) data")
p2acf <- autoplot(acf(data2, plot = FALSE), main = "")

grid.arrange(p1, p1acf, p2, p2acf, nrow=2)
```



The correlograms for the AR(1) and AR(2) process look very similar, so the correlogram is not an appropriate tool for choosing p.

A second approach to choosing p is to start by fitting an AR(1) process because it is the simplest, and look at the correlogram of the residuals. If the residuals resemble a purely random process, then the chosen model is adequate. If there is still residual correlation you could increase p by one (to an AR(2) model) and re-fit the model, again checking for the presence of residual correlation. However, this approach assumes that an autoregressive process of some order is an appropriate model. What happens if it isn't? Therefore we need an alternative approach.

Definition 2 (Partial autocorrelation function).

*The **partial autocorrelation function (PACF)** at lag τ is equal to the estimated lag τ coefficient $\hat{\alpha}_\tau$, obtained when fitting an AR(τ) model to a data set. It is denoted by π_τ, and represents the excess correlation in the time series that has not been accounted for by the $\tau - 1$ smaller lags.*

Notes

The partial autocorrelation function at lag τ is calculated by fitting an AR(τ) process to the data. For example:

- Fit the AR(1) model $X_t = \alpha_1 X_{t-1} + Z_t$ to the time series; then $\hat{\pi}_1 = \hat{\alpha}_1$.
- Fit the AR(2) model $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$ to the time series; then $\hat{\pi}_2 = \hat{\alpha}_2$.
- Fit the AR(3) model $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + Z_t$ to the time series, then $\hat{\pi}_3 = \hat{\alpha}_3$, and so on.

The partial autocorrelation coefficients $\hat{\alpha}_\tau$ from fitting an AR(τ) model can be estimated using a number of methods.

- The partial autocorrelation function is not defined for lag zero, unlike the autocorrelation function.

The partial autocorrelation function allows us to determine an appropriate AR(p) process for a given data set. This is because if the data do come from an AR(p) process then

$$\pi_\tau = \begin{cases} \text{non-zero} & \tau \leq p \\ 0 & \tau > p \end{cases}$$

The justification for this is as follows.

1. Consider an AR(1) process, $X_t = \alpha_1 X_{t-1} + Z_t$. Then
   - $\hat{\pi}_1$ is estimated from fitting an AR(1) model so $\hat{\pi}_1 = \hat{\alpha}_1 \approx \alpha_1$, which is not zero as this is an AR(1) process.
   - for any $\tau > 1$, $\hat{\pi}_\tau$ is estimated from fitting an AR(τ) model so $\hat{\pi}_\tau = \hat{\alpha}_\tau \approx \alpha_\tau = 0$, as it is an AR(1) process.
2. Consider an AR(2) process, $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$. Then
   - $\hat{\pi}_1$ is estimated from fitting an AR(1) model so $\hat{\pi}_1 = \hat{\alpha}_1 \approx \alpha_1$ (non-zero), as it is an AR(2) process.
   - $\hat{\pi}_2$ is estimated from fitting an AR(2) model so $\hat{\pi}_2 = \hat{\alpha}_2 \approx \alpha_2$ (non-zero), as it is an AR(2) process.
   - for any $\tau > 2$, $\hat{\pi}_\tau$ is estimated from fitting an AR(τ) model so $\hat{\pi}_\tau = \hat{\alpha}_\tau \approx \alpha_\tau = 0$, as it is an AR(2) process.

The same arguments hold for AR(p) processes for $p > 2$.

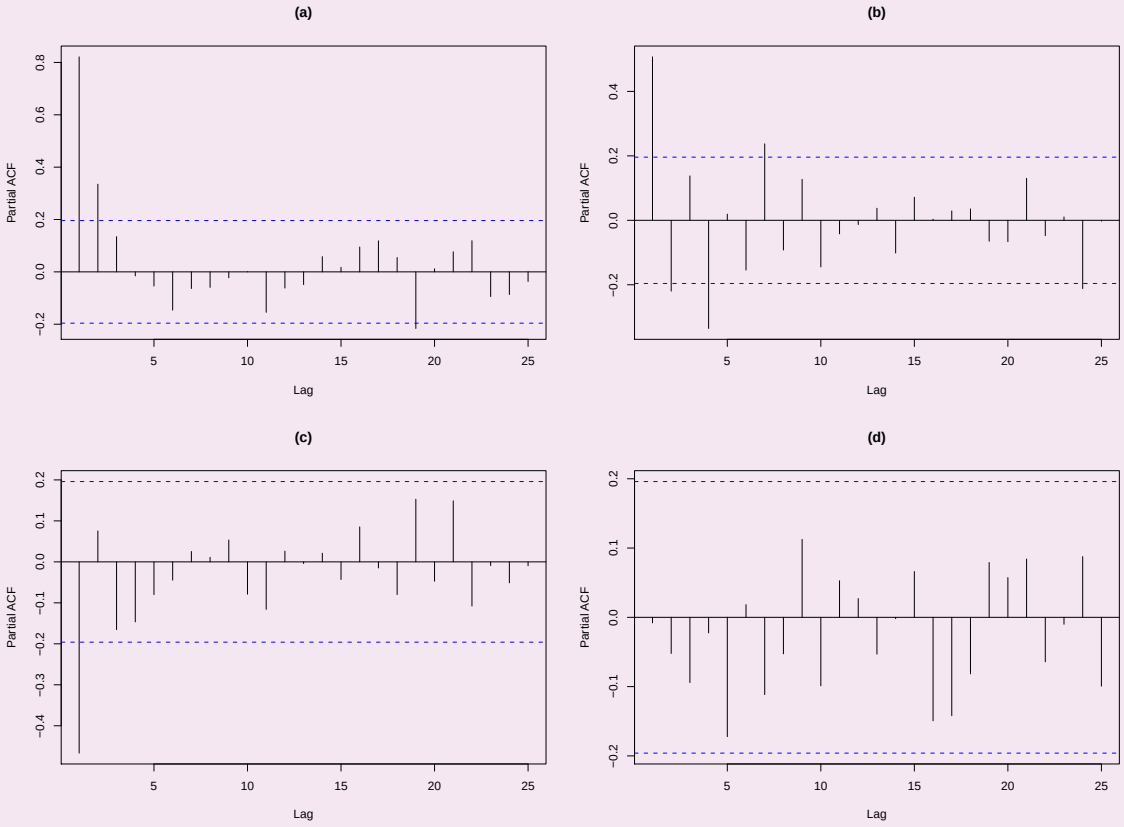
It can also be shown that for an AR(p) process, for any lag $\tau > p$

$$\pi_\tau \sim \text{N}\left(0, \frac{1}{n}\right)$$

which gives an approximate 95% confidence interval for π_τ of $\pm 1.96/\sqrt{n}$, when the true value $\pi_\tau = 0$. Therefore to choose the order of an AR(p) process, plot the partial autocorrelation function and choose p as the smallest value that is significantly different from zero.

Task 3.

From the PACFs below, determine whether an AR(p) process would be an appropriate, and if so what value of p you would use.



Supplementary material:
Parameter estimation

Parameter estimation is not generally done by hand so there is no need to learn how to carry out parameter estimation in this course. However, it is useful to be aware of what the methods are. The main two methods are:

1. **The autocorrelation method (Yule-Walker estimation)** uses the Yule-Walker recursive equations and replaces the theoretical autocorrelation function by its sample estimates. For the variance, we use the form of the variance for an AR(p) process, again, replacing the theoretical quantities by their sample estimates.
2. **Ordinary least squares estimation** estimates the α's by minimising the ordinary least squares objective function and the residual variance is estimated from the residuals.

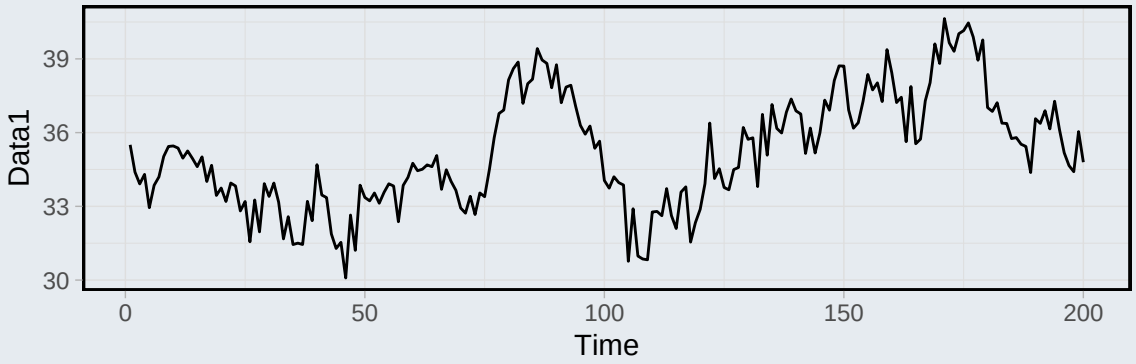
Notes:

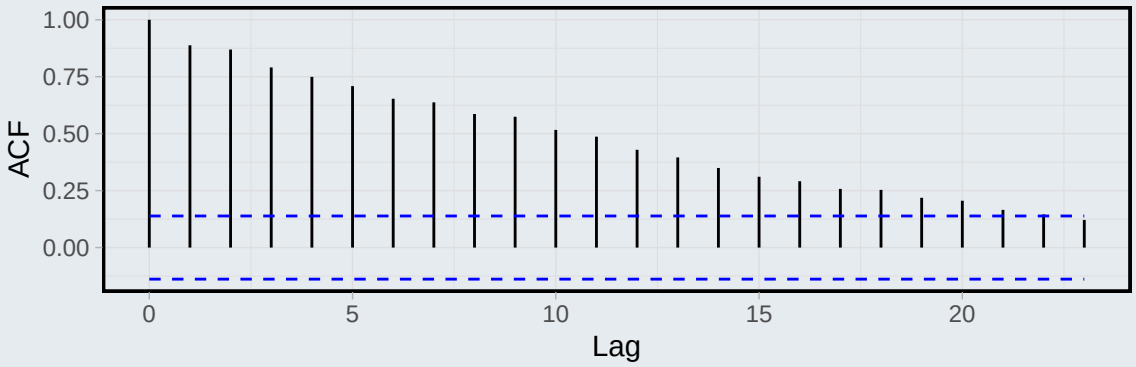
- Ordinary least squares (which assumes independence) is used because the correlation in the time series is modelled by the AR(p) process, hence the errors Z_t are independent.
- When the data are normally distributed the ordinary least squares estimates are also the maximum likelihood estimates.
- The Yule-Walker method does not produce accurate estimates when *n* is small (less than 50) and the model is close to being non-stationary.
- There are many other methods of estimation for time series. R uses either maximum likelihood or least squares.

Example 4.

Consider the data set generated using the following R code.

```
#### Generate AR(2) data with a linear trend
time <- 1:200
data.ar <- arima.sim(model=list(ar=c(0.6, 0.3)), n=200, sd=1)
data <- data.ar + 30 + 0.05 * time

## Plot the correlogram and time plot
p1 <- autoplot(data, xlab = "Time", ylab = "Data1", main = "")
p1acf <- autoplot(acf(data, plot = FALSE), main = "")

grid.arrange(p1, p1acf, nrow=2)
```





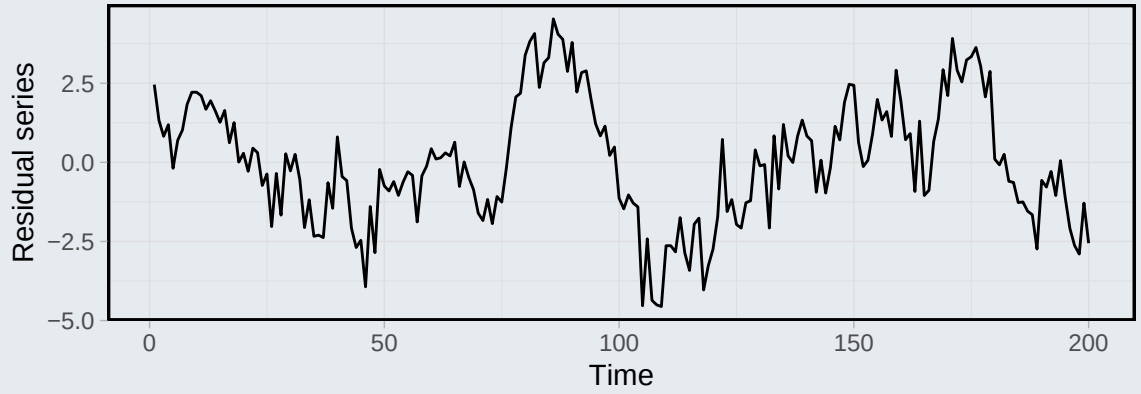
From looking at the time plot and correlogram the data appear to have a linear trend, which we remove before trying to model the correlation. We remove the trend using regression methods (see Week 6 for details), i.e. by fitting $m_t = \beta_0 + \beta_1 t$ as shown below, and obtain the following residual series.

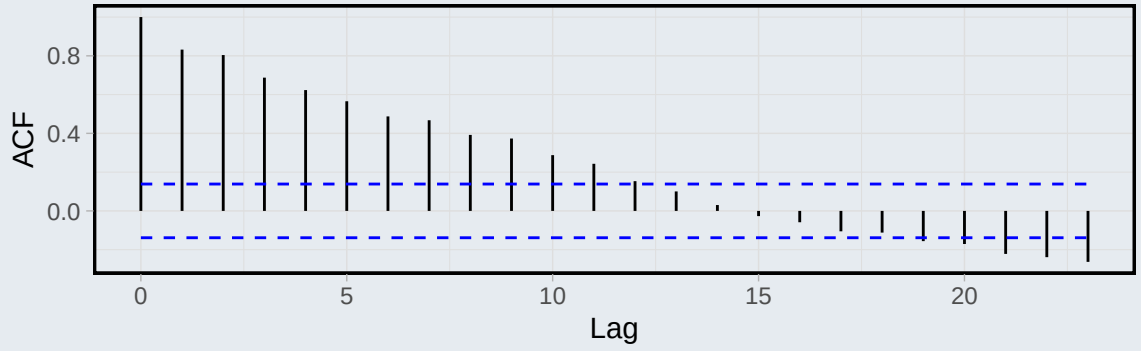
```
## Remove the trend
linear.model <- lm(data~time)
summary(linear.model)
```

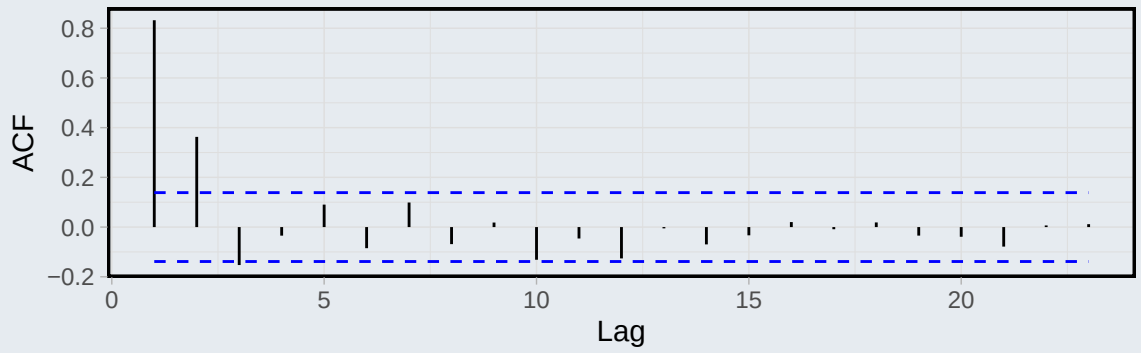
```
Call:
lm(formula = data ~ time)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5568 -1.2720 -0.0947  1.2576  4.5332

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.025323   0.273792 120.622   <2e-16 ***
time         0.021630   0.002362   9.156   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.929 on 198 degrees of freedom
```

```
Multiple R-squared:  0.2975,    Adjusted R-squared:  0.2939
F-statistic: 83.84 on 1 and 198 DF,  p-value: < 2.2e-16
```

```
residual.series <- data - linear.model$fitted.values

## Plotting
p1 <- autoplot(residual.series, xlab = "Time", ylab = "Residual series")
p1acf <- autoplot(acf(residual.series, plot = FALSE),
                  main = "")
p1pacf<- autoplot(pacf(residual.series, plot = FALSE),
                  main = "")
grid.arrange(p1, p1acf, p1pacf, nrow=3)
```







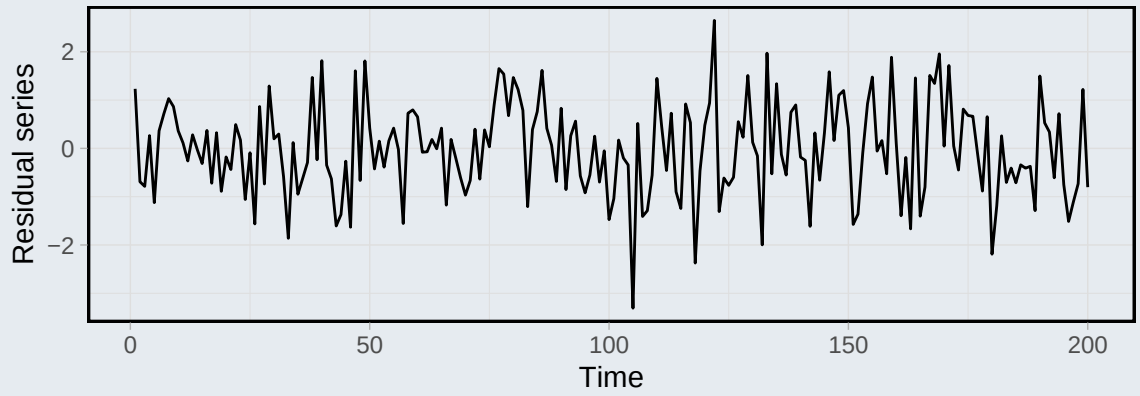
The time plot suggests the residual series is stationary, while the partial autocorrelation function suggests that an AR(2) process is appropriate. Note that the autocorrelation function does not really tell us a great deal here. Finally we fit an AR(2) model to the data using the R function `arima()`, and plot the residuals to ensure they are independent. The R code for this is as follows:

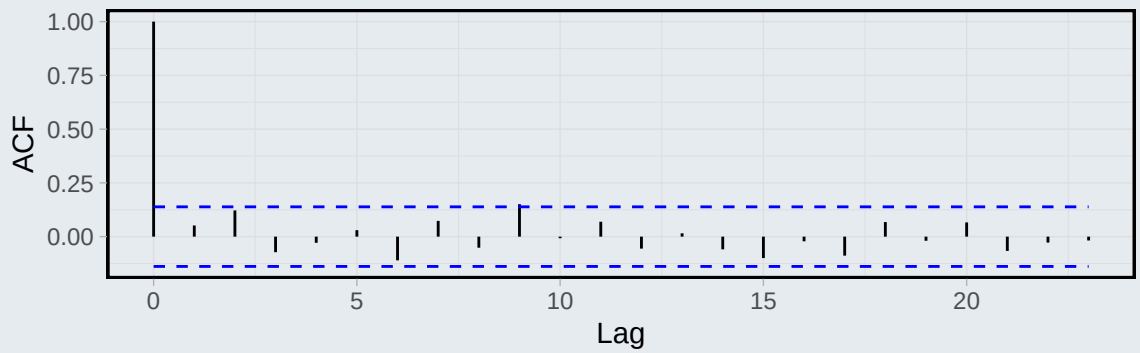
```
## Fit an AR(2) model to the data
model.ar <- arima(residual.series, order=c(2,0,0))
model.ar
```

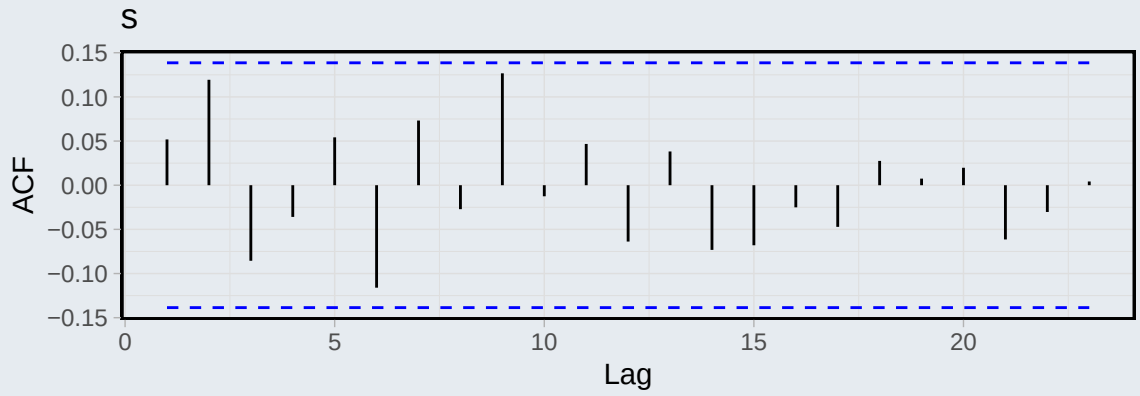
```
Call:
```

```
arima(x = residual.series, order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.5332  0.3671    -0.0030
s.e.  0.0660  0.0661     0.6481

sigma^2 estimated as 0.9423:  log likelihood = -278.61,  aic = 565.22
```

```
## Plotting
p2 <- autoplot(model.ar$residuals, xlab = "Time",  ylab = "Residual series")
p2acf <- autoplot(acf(model.ar$residuals, plot = FALSE),
                  main = "")
p2pacf<- autoplot(pacf(model.ar$residuals, plot = FALSE),
                  main = "s")
grid.arrange(p2, p2acf, p2pacf, nrow=3)
```





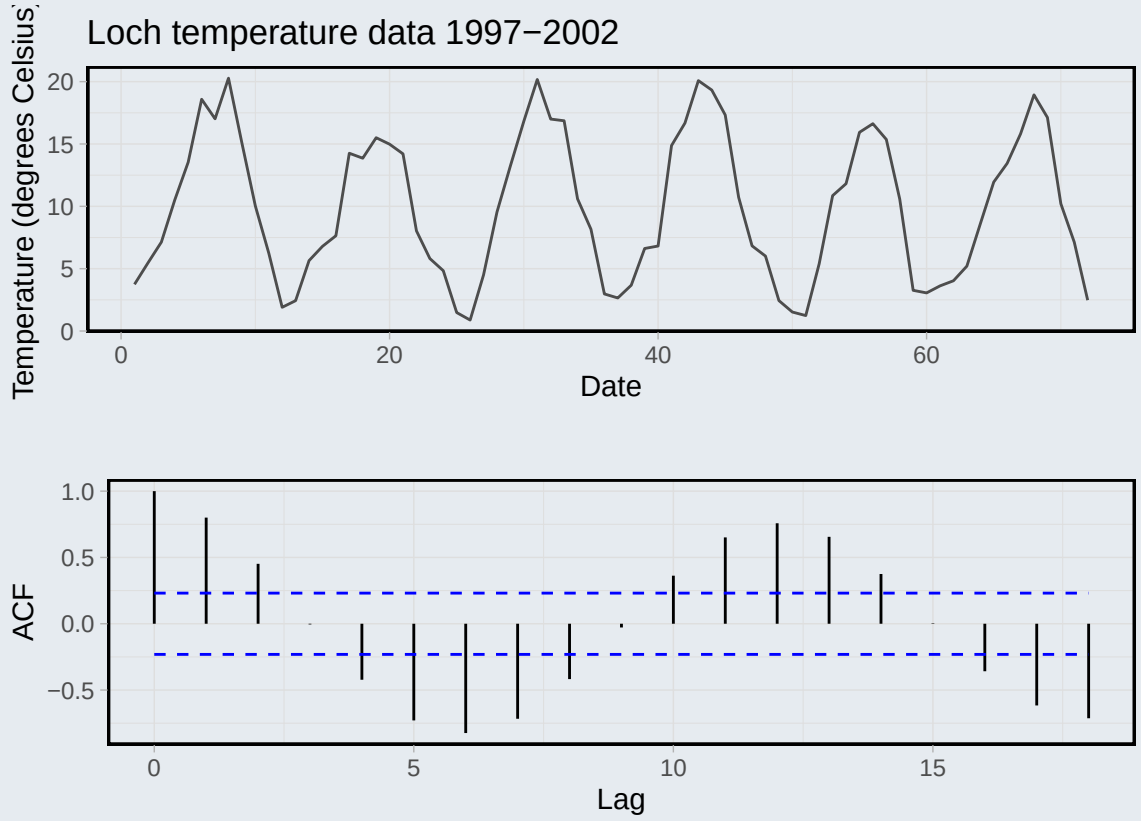


The residuals look independent, so the model we fitted is appropriate.

Example 5 (Temperature data in a Scottish loch).

The Centre for Ecology and Hydrology (CEH) in Edinburgh routinely measure the temperature in a number of Scottish lochs (lakes). The data below are monthly temperature levels in a particular loch between 1997 and 2002. The plot below shows the temperature levels exhibit little trend but prominent seasonal variation, with higher values in the summer than in the winter.

```
data <- read.csv(url("http://www.stats.gla.ac.uk/~tereza/rp/lochdata.csv"))
n <- nrow(data)
time <- 1:n
x <- data[,3]

## Plotting time series and ACF
p1 <- ggplot(data, aes(time, temperature))+ geom_line(color = "#4d4d4d")+
    xlab("Date") + ylab("Temperature (degrees Celsius)") +
    ggtitle("Loch temperature data 1997-2002")

p1acf <- autoplot(acf(data$temperature, plot = FALSE),main = "")

grid.arrange(p1, p1acf, nrow=2)
```



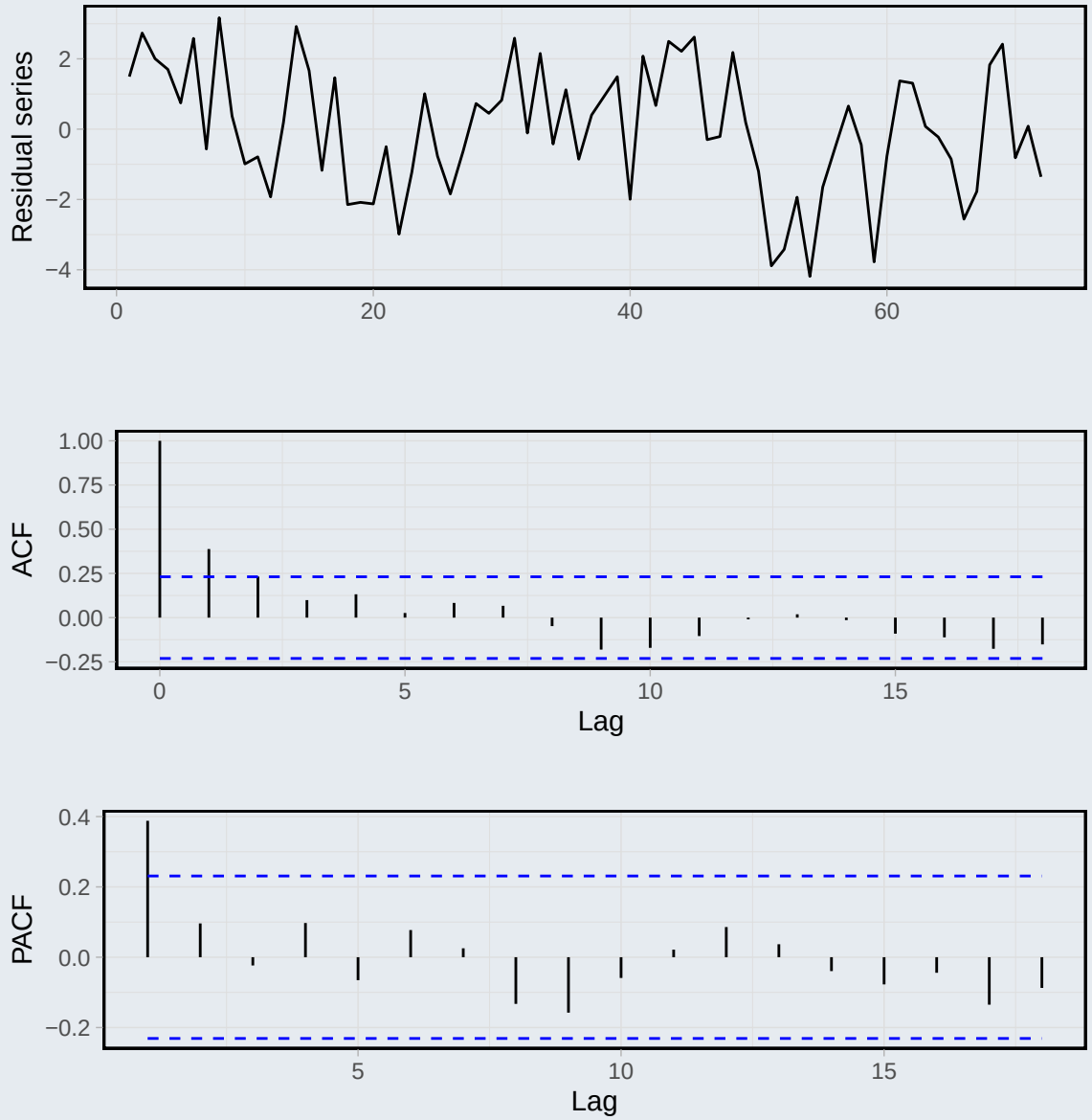
Given that the seasonal pattern seems regular (the same each year), an appropriate model seems to be

$$s_t = \beta_0 + \beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12)$$

where we divide by 12 because there are 12 periods (i.e. months) before the seasonal pattern repeats itself. We fit the seasonal model as shown below and obtain the following residual series.

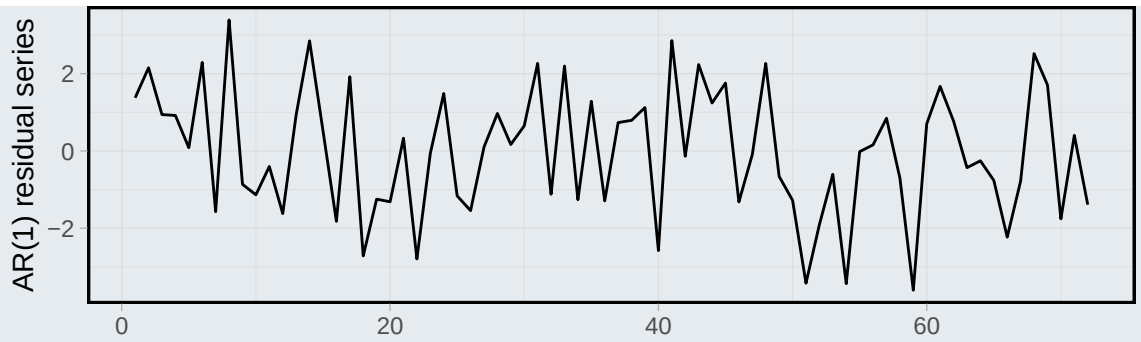
```
model.season <- lm(x~sin(2*pi*time/12) + cos(2*pi*time/12))
residual.series <- x - model.season$fitted.values

# Plotting
p2 <- autoplot(ts(residual.series), ylab = "Residual series")
```

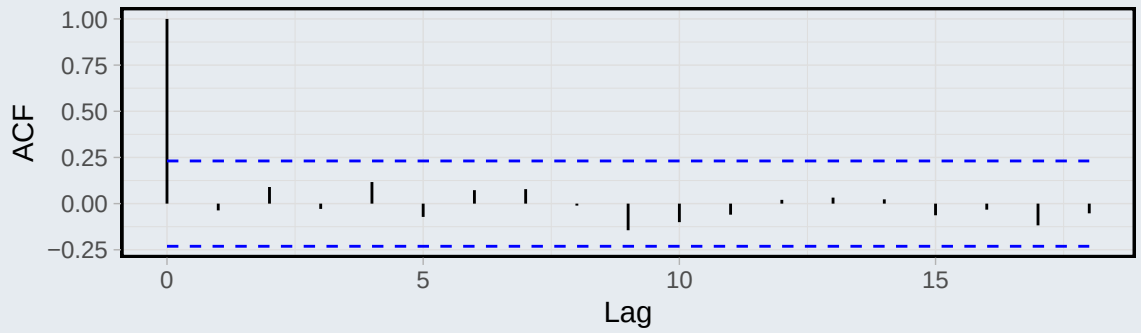
```
p2acf <- autoplot(acf(residual.series, plot = FALSE),
                  main = "")
p2pacf <- autoplot(pacf(residual.series, plot = FALSE),
                  main = "", ylab="PACF")

grid.arrange(p2, p2acf, p2pacf, nrow=3)
```

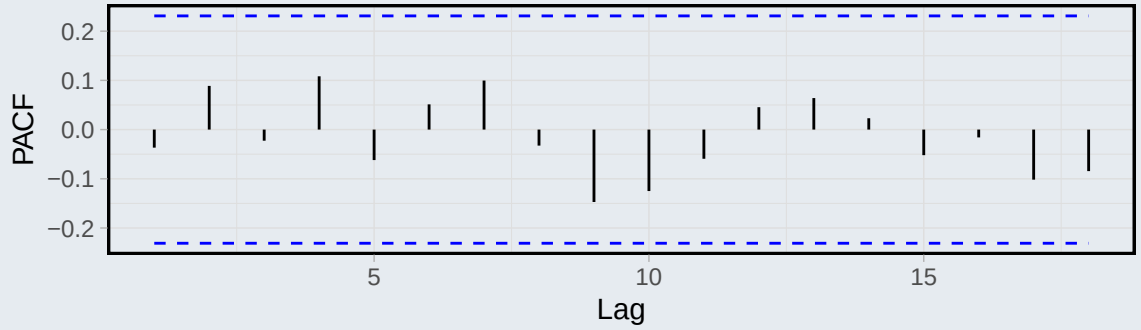


The time plot suggests that the residual series is stationary, while the partial autocorrelation function suggests that an AR(1) process is appropriate. We fit the AR(1) model as shown below and obtain the following residuals.

```
model.ar <- arima(residual.series, order=c(1,0,0))

# Plotting
p3 <- autoplot(model.ar$residuals, ylab = "AR(1) residual series")
p3acf <- autoplot(acf(model.ar$residuals, plot = FALSE),
                  main = "")
p3pacf <- autoplot(pacf(model.ar$residuals, plot = FALSE),
                  main = "", ylab="PACF")

grid.arrange(p3, p3acf, p3pacf, nrow=3)
```







The residuals look independent, so the model we fitted is appropriate.

Task 4.

The dataset `glasgowtemp.csv` available from this link contains the daily minimum temperature levels in Glasgow between 2005 and 2007.

Produce descriptive plots of the data, remove any trend or seasonal variation by fitting an appropriate model and check the residual series for the presence of short-term correlation.

Then proceed to identify an appropriate process to model short-term correlation. Fit the chosen model to the data using R, and analyse the residuals to ensure the chosen model is appropriate.

Moving average processes

The second most common type of stationary time series model is called a moving average process. Here we assume that the time series being modelled is weakly stationary, which can be obtained by removing any trend or seasonal variation using the methods mentioned in last week's learning material.

Moving average processes

https://youtu.be/hXWa9cF42HM

Duration: 7m33s

Definition 3 (Moving average process).

*Let Z_t be a purely random process (i.e. each Z_t is independent) with mean 0 and variance σ_z^2. A **moving average process of order** q, denoted MA(q), is given by*

$$X_t = \lambda_0 Z_t + \lambda_1 Z_{t-1} + \ldots + \lambda_q Z_{t-q}$$

where traditionally Z_t is given a coefficient of 1, i.e. $\lambda_0 = 1$. The model can be re-written using the backshift operator, $BZ_t = Z_{t-1}$, as

$$\begin{aligned} X_t &= Z_t + \lambda_1 Z_{t-1} + \ldots + \lambda_q Z_{t-q} \\ &= (1 + \lambda_1 B + \ldots + \lambda_q B^q) Z_t \\ &= \theta(B) Z_t \end{aligned}$$

*where $\theta(B) = 1 + \lambda_1 B + \ldots + \lambda_q B^q$ is called the **characteristic polynomial**.*

Correlation is introduced between X_t and X_{t-1} because their values both depend on the elements of the purely random process $Z_{t-1}, \ldots, Z_{t-q}$. In common with an autoregressive process, λ_1 is the lag 1 coefficient, λ_2 is the lag 2 coefficient and so on.

Mean and variance of MA(q) process

The mean and variance of an MA(q) process are straightforward to calculate. Writing the process as

$$X_t = \sum_{j=0}^{q} \lambda_j Z_{t-j}$$

where again $\lambda_0 = 1$, the mean can be calculated as

$$\begin{aligned} E(X_t) &= E(\sum_{j=0}^{q} \lambda_j Z_{t-j}) \\ &= \sum_{j=0}^{q} \lambda_j E(Z_{t-j}) \\ &= 0 \end{aligned}$$

So for any parameters $\lambda_1, \ldots, \lambda_q$ an MA(q) process has mean zero. The variance is given by

$$\begin{aligned}\mathrm{Var}(X_t) &= \mathrm{Var}(\sum_{j=0}^{q}\lambda_j Z_{t-j}) \\ &= \sum_{j=0}^{q}\lambda_j^2\mathrm{Var}(Z_{t-j}) \\ &= \sigma_z^2\sum_{j=0}^{q}\lambda_j^2\end{aligned}$$

where again $\lambda_0 = 1$ so

$$\mathrm{Var}(X_t) = \sigma_z^2[1+\sum_{j=1}^{q}\lambda_j^2]$$

There are no covariance terms in the calculation because $Z_t, Z_{t-1}, \ldots, Z_{t-q}$ are a purely random process and are therefore independent ($\mathrm{Cov}(Z_t, Z_{t-k}) = 0$ for $k \neq 0$).

Autocorrelation function of MA(q) process

The autocovariance function is given by

$$\gamma_\tau = \mathrm{Cov}(X_t, X_{t+\tau}) = \begin{cases} \sigma_z^2\sum_{j=0}^{q-\tau}\lambda_j\lambda_{j+\tau} & \text{if } \tau = 0,1,\ldots,q \\ 0 & \tau > q \end{cases}$$

and the autocorrelation function is given by

$$\rho_\tau = \mathrm{Corr}(X_t, X_{t+\tau}) = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{\sum_{j=0}^{q-\tau}\lambda_j\lambda_{j+\tau}}{\sum_{j=0}^{q}\lambda_j^2} & \text{if } \tau = 1,\ldots,q \\ 0 & \tau > q \end{cases}$$

where $\lambda_0 = 1$.

Supplementary material:
Calculation of the autocovariance function

The autocovariance function is calculated as:

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t+\tau}) &= E(X_t X_{t+\tau}) - E(X_t)E(X_{t+\tau}) \\ &= E(X_t X_{t+\tau}) \\ &= E\left\{\sum_{j=0}^{q}\lambda_j Z_{t-j}\right\}\times\left\{\sum_{k=0}^{q}\lambda_k Z_{t-k+\tau}\right\} \\ &= \sum_{j=0}^{q}\sum_{k=0}^{q}\lambda_j\lambda_k E(Z_{t-j}Z_{t-k+\tau})\end{aligned}$$

The expectations $E(Z_r Z_s)$ can be split into two cases:

1. If $r \neq s$ we have $E(Z_r Z_s) = E(Z_r)E(Z_s) = 0$ because the Z_t are independent.
2. If $r = s$ we have $E(Z_r^2) = \mathrm{Var}(Z_r) + [E(Z_r)]^2 = \sigma_z^2$.

Therefore, only the Z_r included in both X_t and $X_{t+\tau}$ contribute to the covariance function.

- When $\tau = 0$, $Z_t, \ldots, Z_{t-q}$ are in both X_t and X_t.
- When $\tau = 1$, $Z_t, \ldots, Z_{t-q+1}$ are in both X_t and X_{t+1}.
- When $\tau = q-1$, Z_t, Z_{t-1} are in both X_t and X_{t+q-1}.

- When $\tau = q$, Z_t is in both X_t and X_{t+q}.
- When $\tau = q + 1$ there are no elements in common.

Therefore the autocovariance function is given by

$$\gamma_\tau = \text{Cov}(X_t, X_{t+\tau}) = \begin{cases} \sigma_z^2 \sum_{j=0}^{q-\tau} \lambda_j \lambda_{j+\tau} & \text{if } \tau = 0, 1, \ldots, q \\ 0 & \tau > q \end{cases}$$

where $\lambda_0 = 1$.

Notes

1. The mean and variance of any MA(q) process are finite and constant, while the autocorrelation function is finite and does not depend on t. Therefore any MA(q) is weakly stationary.
2. The autocorrelation function of an MA(q) process is positive at lags $1, \ldots, q$ and zero for any lag greater than q. This gives us a method for detecting whether an MA(q) process is an appropriate model for a given data set.

Example 6.

Consider the MA(1) process $X_t = Z_t + \lambda Z_{t-1}$. Its variance is given by

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(Z_t + \lambda Z_{t-1}) \\ &= \sigma_z^2(1 + \lambda^2) \end{aligned}$$

Its lag one autocovariance is given by

$$\begin{aligned} \text{Cov}(X_t, X_{t+1}) &= \text{Cov}(Z_t + \lambda Z_{t-1},\ Z_{t+1} + \lambda Z_t) \\ &= \text{Cov}(Z_t,\ Z_{t+1}) + \lambda \text{Cov}(Z_t,\ Z_t) + \lambda \text{Cov}(Z_{t-1},\ Z_{t+1}) + \lambda^2 \text{Cov}(Z_{t-1},\ Z_t) \\ &= \lambda \sigma_z^2 \end{aligned}$$

while its lag 2 autocovariance is given by

$$\begin{aligned} \text{Cov}(X_t, X_{t+2}) &= \text{Cov}(Z_t + \lambda Z_{t-1},\ Z_{t+2} + \lambda Z_{t+1}) \\ &= \text{Cov}(Z_t,\ Z_{t+2}) + \lambda \text{Cov}(Z_t,\ Z_{t+1}) + \lambda \text{Cov}(Z_{t-1},\ Z_{t+2}) + \lambda^2 \text{Cov}(Z_{t-1},\ Z_{t+1}) \\ &= 0 \end{aligned}$$

The same argument holds true for $\tau > 2$. Therefore the autocorrelation function is given by

$$\rho_\tau = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{\lambda}{1+\lambda^2} & \text{if } \tau = 1 \\ 0 & \tau > 1 \end{cases}$$

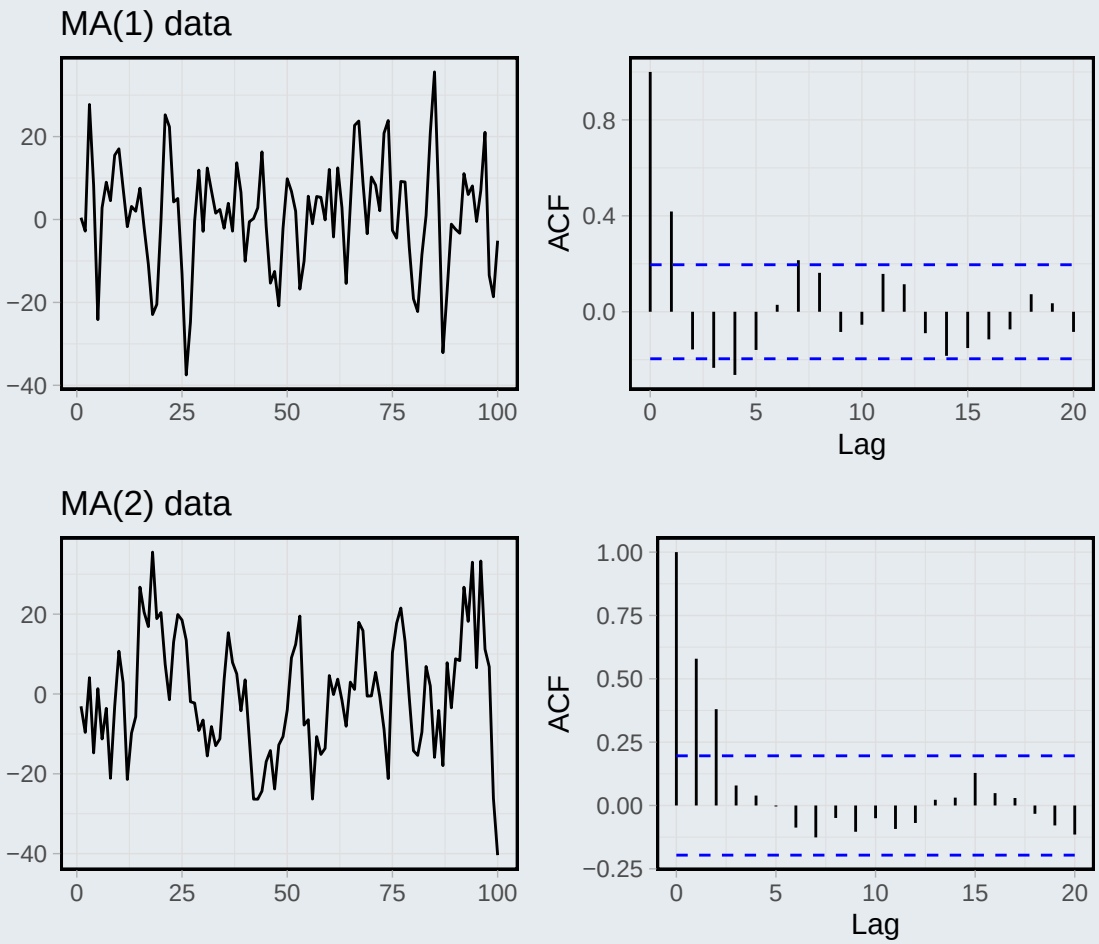
Task 5.

For the MA(1) processes $X_t = Z_t + \lambda Z_{t-1}$ and $X_t = Z_t + \frac{1}{\lambda} Z_{t-1}$, show that they have the same autocorrelation function and find the autocorrelation function when $\lambda = 1$ and $\lambda = -1$. Can you find a λ so that $|\rho_1| > 1/2$?

Example 7.

The graphs below show time plots and correlograms of simulated data with MA(1) and MA(2) correlation structures. As suggested by theory, the sample autocorrelation functions are only significantly different from zero for lags less than or equal to q. The code to simulate the data is given below.

```
# Simulated MA(1) and MA(2) data
data1 <- arima.sim(model=list(ma=0.8), n=100, sd=10)
data2 <- arima.sim(model=list(ma=c(0.8, 0.8)), n=100, sd=10)

# Plotting
p1 <- autoplot(data1, main = "MA(1) data")
p1acf <- autoplot(acf(data1, plot = FALSE),
                  main = "", ylab="ACF")
p2 <- autoplot(data2, main = "MA(2) data")
p2acf <- autoplot(acf(data2, plot = FALSE),
                  main = "", ylab="ACF")

grid.arrange(p1, p1acf, p2, p2acf, nrow=2)
```



Invertibility

If our data gives a sample autocorrelation function similar to an MA(1) process, we cannot estimate λ from the data, because λ and $1/\lambda$ will both fit the data equally well, as we've seen in Task 5.

For a given MA(q) autocorrelation function there are two sets of possible parameters $(\lambda_1, \ldots, \lambda_q)$ that could have produced it. However, only one of these has the desirable property of being **invertible**. **Invertibility** roughly means that we can invert the formula for the MA process and express X_t in terms of past values X_{t-1}, X_{t-2}, etc. The other solution in general has the property that X_t can only be expressed in terms of future values X_{t+1}, X_{t+2}, etc. Since the latter is less intuitive than the former, typically one focuses in estimation on the invertible solution.

Definition 4 (Invertibility).

An MA(q) process is ***invertible*** *if it can be written as an autoregressive process of infinite order, i.e. as*

$$X_t = \sum_{j=1}^{\infty} b_j X_{t-j} + Z_t$$

where the coefficients b_j form a convergent sum, i.e. $\sum_{j=1}^{\infty} b_j < \infty$.

Example 8.

Consider the MA(1) process $X_t = Z_t + \lambda Z_{t-1}$. We can invert it as follows.

$$\begin{aligned} Z_t &= X_t - \lambda Z_{t-1} \\ &= X_t - \lambda(X_{t-1} - \lambda Z_{t-2}) \\ &= X_t - \lambda X_{t-1} + \lambda^2(X_{t_2} - \lambda Z_{t-3}) \\ &= \vdots \\ &= \sum_{j=0}^{\infty} (-\lambda)^j X_{t-j} \end{aligned}$$

Then using this description we can write the MA(1) process as

$$X_t = Z_t + \lambda Z_{t-1} = Z_t + \lambda \sum_{j=0}^{\infty} (-\lambda)^j X_{t-j-1}$$

which is an AR(∞) process. This process is invertible if the coefficients form a convergent sum, i.e. if

$$\sum_{j=0}^{\infty} (-\lambda)^j = 1 + (-\lambda) + (-\lambda)^2 + (-\lambda)^3 + \ldots < \infty$$

We know that the sum of a geometric progression is only convergent if $|\lambda| < 1$. Therefore, for the MA(1) processes

$$\begin{aligned} \textbf{A} \qquad X_t &= Z_t + \lambda Z_{t-1} \text{ and} \\ \textbf{B} \qquad X_t &= Z_t + \frac{1}{\lambda} Z_{t-1}, \end{aligned}$$

either $|\lambda| < 1$ or $|1/\lambda| < 1$, so only one of them is invertible.

Let us now specify the condition that determines whether a process is invertible.

Theorem 3 (Invertibility).

The MA(q) process

$$\begin{aligned} X_t &= Z_t + \lambda_1 Z_{t-1} + \ldots + \lambda_q Z_{t-q} \\ &= (1 + \lambda_1 B + \ldots + \lambda_q B^q) Z_t \\ &= \theta(B) Z_t \end{aligned}$$

is invertible if and only if the roots of the characteristic polynomial $\theta(B)$ have modulus greater than one and hence lie outside the unit circle.

Note: The condition for invertibility of an MA(q) process is identical to the condition for stationarity of an AR(p) process.

Task 6.

Is the MA(2) process $X_t = Z_t + 4.25Z_{t-1} + Z_{t-2}$ invertible?

Model identification

Given a stationary time series, how do we determine if:

(a) a moving average process is appropriate, and

(b) if it is, which order process should we use (i.e. which value of q)?

We can answer both of these questions using the autocorrelation function (correlogram). Recall that for an MA(q) process the ACF is given by

$$\rho_\tau = \mathrm{Corr}[X_t, X_{t+\tau}] = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{\sum_{j=0}^{q-\tau} \lambda_j \lambda_{j+\tau}}{\sum_{j=0}^{q} \lambda_j^2} & \text{if } \tau = 1, \ldots, q \\ 0 & \tau > q \end{cases}$$

This function is non-zero for $\tau \leq q$ and zero for $\tau > q$. It can also be shown that for $\tau > q$

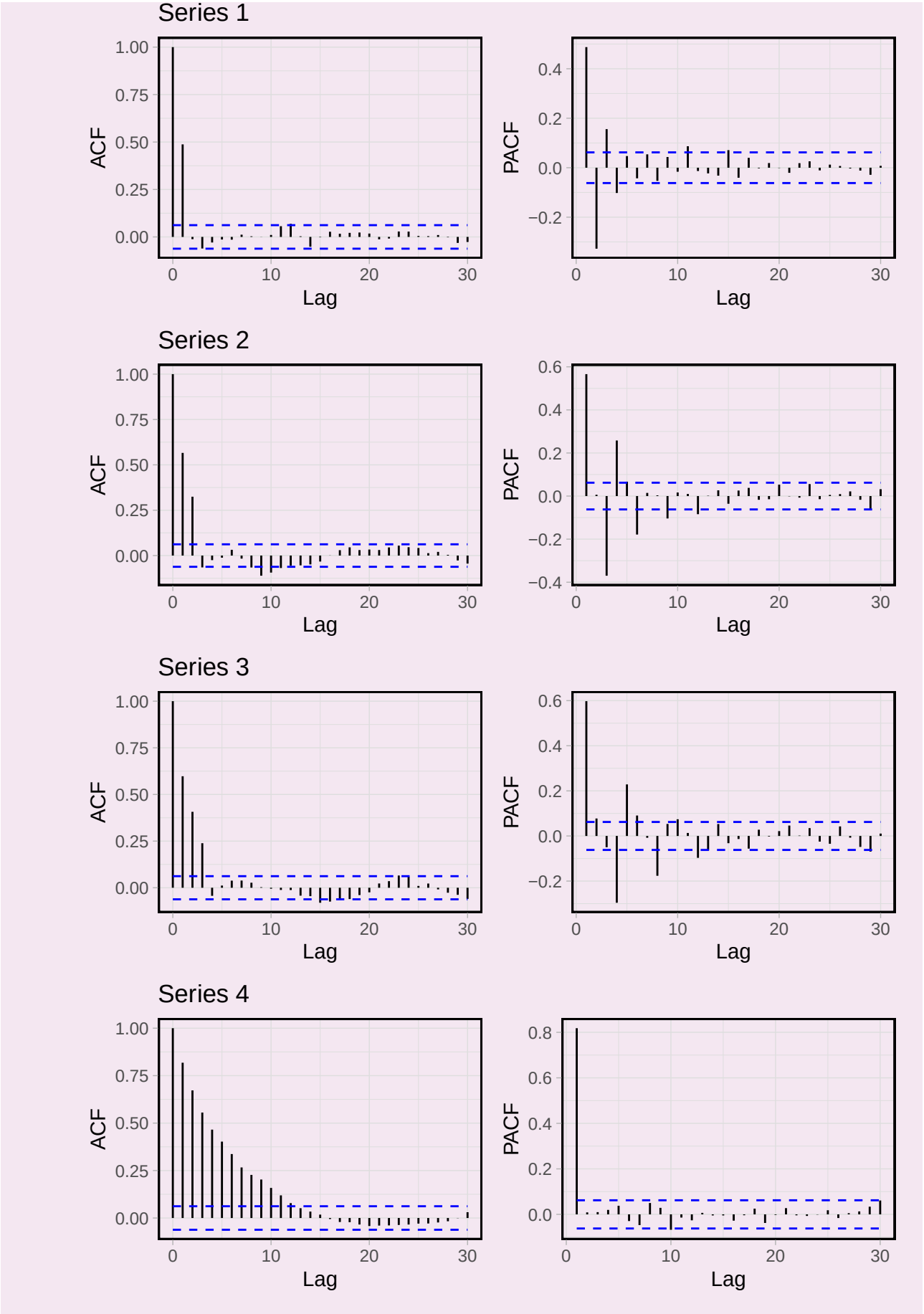
$$\rho_\tau \sim \mathrm{N}\left(0, \frac{1}{n}\right)$$

which results in approximate 95% confidence intervals for ρ_τ of $\pm 1.96/\sqrt{n}$. Therefore it is straightforward to identify whether an MA(q) process is appropriate by plotting the sample autocorrelation function (correlogram). If an MA(q) process is appropriate the correlogram will be significantly different from zero at lags $\tau = 1, \ldots, q$, and within the 95% confidence intervals for $\tau > q$.

Note: The autocorrelation function (ACF) tells us whether an MA(q) process is appropriate while the partial autocorrelation function (PACF) suggests whether an AR(p) process is appropriate. Therefore for a given time series, both should be plotted to show which process would be a good model.

Task 7.

Sample autocorrelation and partial autocorrelation functions are shown below for four data sets, one data set on each row. What type of MA(q) process is an appropriate model in each case?


Series 1
ACF
PACF
Lag
Series 2
ACF
PACF
Lag
Series 3
ACF
PACF
Lag
Series 4
ACF
PACF
Lag

Supplementary material:
Parameter estimation

Parameter estimation is harder than for an AR(p) process because in the AR(p) case, ordinary least squares can be used. For an MA(q) process an ordinary least squares procedure would require minimisation of random variables of which values are not known. Instead, a modification of ordinary least squares is used, called **conditional least squares**.

Consider the MA(1) model $X_t = \mu + Z_t + \lambda Z_{t-1}$, where $\mathrm{Var}(Z_t) = \sigma_z^2$. The conditional least squares algorithm works as follows.

1. Select starting values for μ and λ,

$$\tilde{\mu} = \frac{1}{n}\sum_{t=1}^{n} x_t \qquad \text{and by solving} \qquad \hat{\rho}_1 = \frac{\tilde{\lambda}}{1+\tilde{\lambda}^2}$$

2. Calculate the conditional residual sum of squares

$$S(\tilde{\mu}, \tilde{\lambda}) = \sum_{t=1}^{n} [x_t - \tilde{\mu} - \tilde{\lambda} Z_{t-1}]^2$$

where $Z_0 = 0$ and Z_t is calculated recursively using

$$Z_t = x_t - \tilde{\mu} - \tilde{\lambda} Z_{t-1}$$

3. Repeat Step 2 for a range of values of (μ, λ) that are close to the initial estimates in Step 1. Then determine the estimates $(\hat{\mu}, \hat{\lambda})$ as the values that minimise $S(\tilde{\mu}, \tilde{\lambda})$ over all those values considered.
4. Using the fact that the variance of an MA(1) process is $\mathrm{Var}(X_t) = \sigma_z^2(1+\lambda^2)$, we obtain that

$$\hat{\sigma}_z^2 = \frac{\hat{\sigma}^2}{(1+\hat{\lambda}^2)}$$

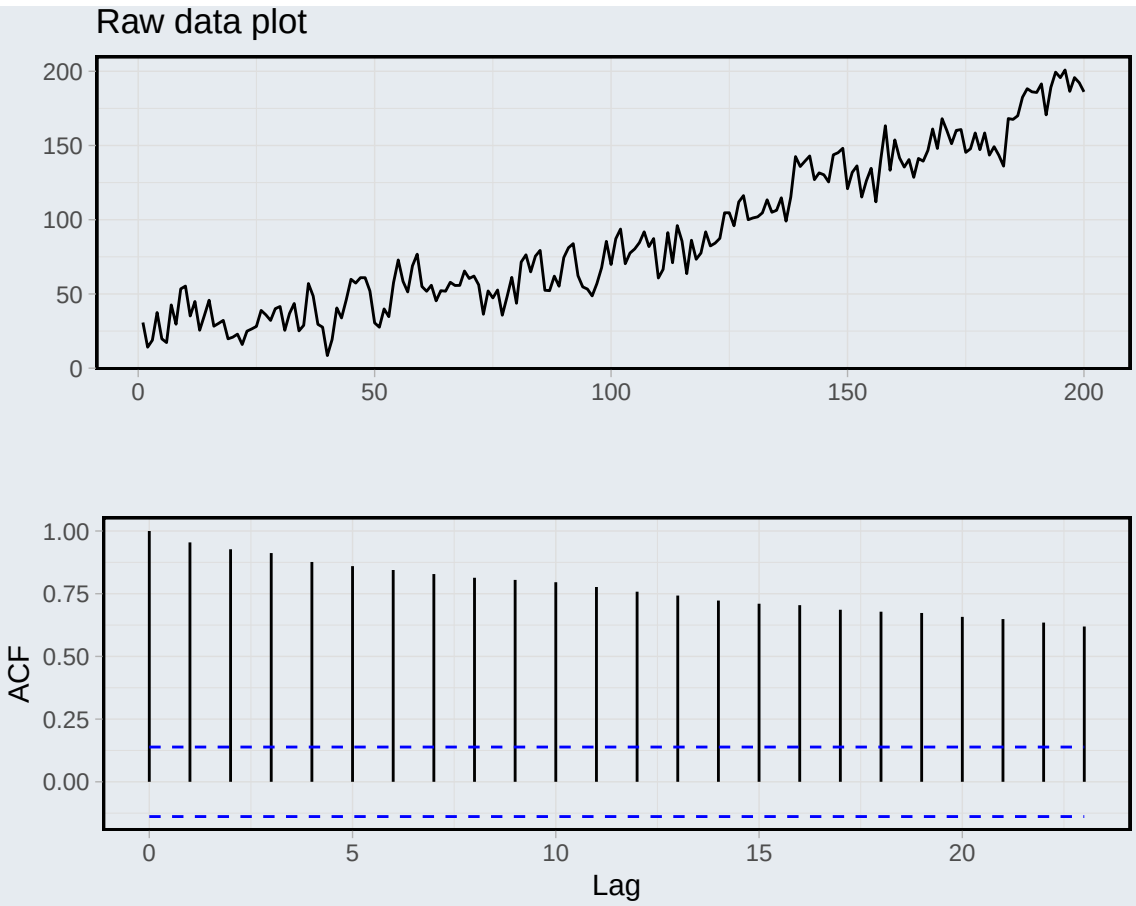
where $\hat{\sigma}^2$ is the overall variance of the process and is estimated as we saw earlier.

Parameter estimation for more general MA(q) processes can be implemented in a similar way. The conditional least squares method is used by R.

Example 9.

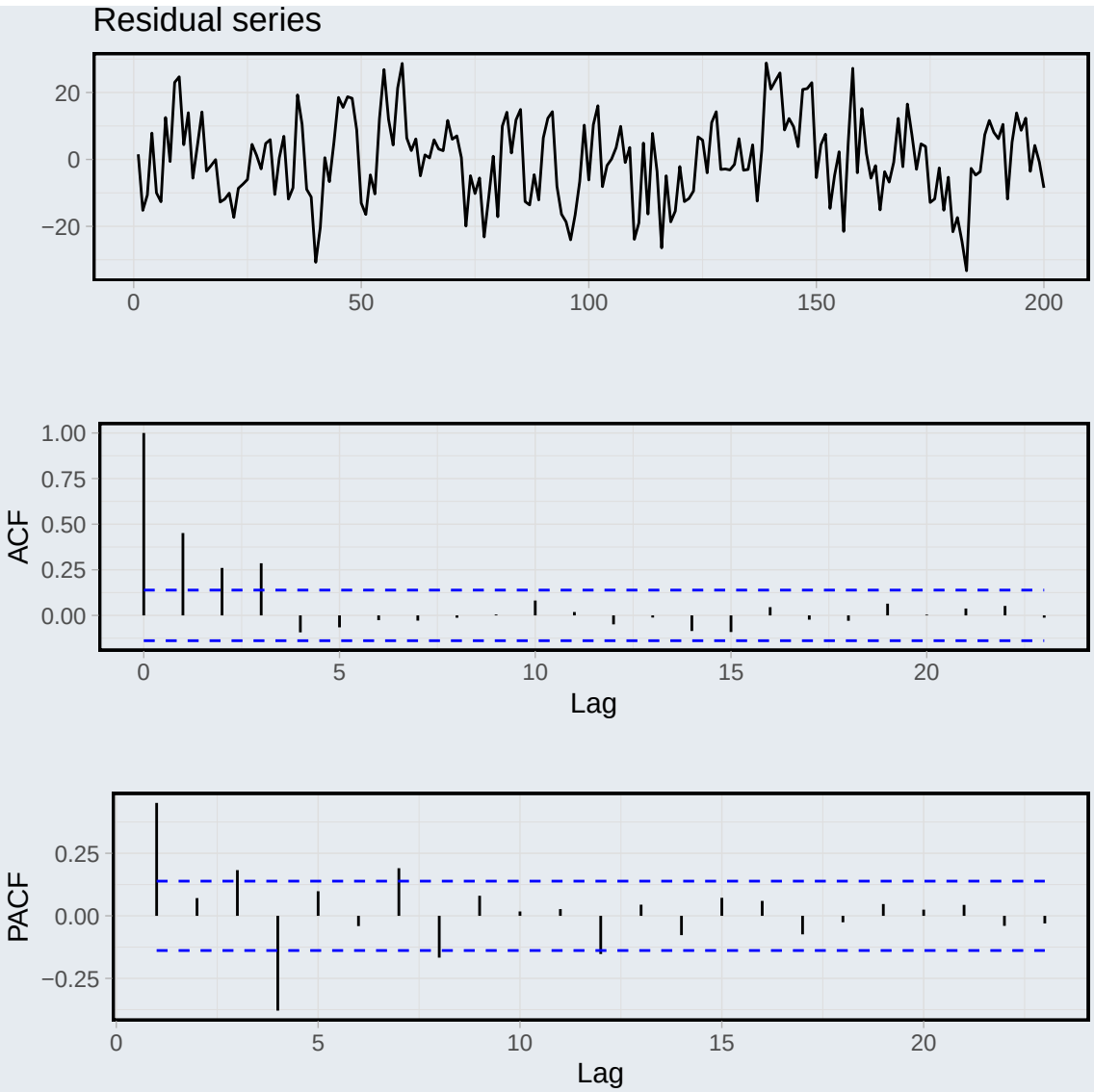
Consider the simulated data set shown below, which was generated using the following R code.

```
time <- 1:200
data.ar <- arima.sim(model=list(ma=c(0.6, 0.3, 0.6)), n=200, sd=10)
data <- data.ar + 30 + 0.05 * time + 0.004*time^2

# Plotting

p1 <- autoplot(data, main="Raw data plot")
p1acf <- autoplot(acf(data, plot = FALSE),
                  main = "")

grid.arrange(p1, p1acf, nrow=2)
```



From looking at the time plot and correlogram the data appear to have a quadratic trend, which we remove before trying to model the correlation. Removing the trend using regression methods (see Week 6 for details), i.e. fitting $m_t = \beta_0 + \beta_1 t + \beta_2 t^2$, gives the following residual series.

```
## Remove the trend
time2 <- time^2
linear.model <- lm(data~time+time2)
residual.series <- data - linear.model$fitted.values

# Plotting
    p <- autoplot(residual.series, main="Residual series")
 pacf <- autoplot(acf(residual.series, plot = FALSE),
                  ylab="ACF", main="")
ppacf <- autoplot(pacf(residual.series, plot = FALSE),
                  ylab="PACF", main="")

grid.arrange(p, pacf, ppacf, nrow=3)
```



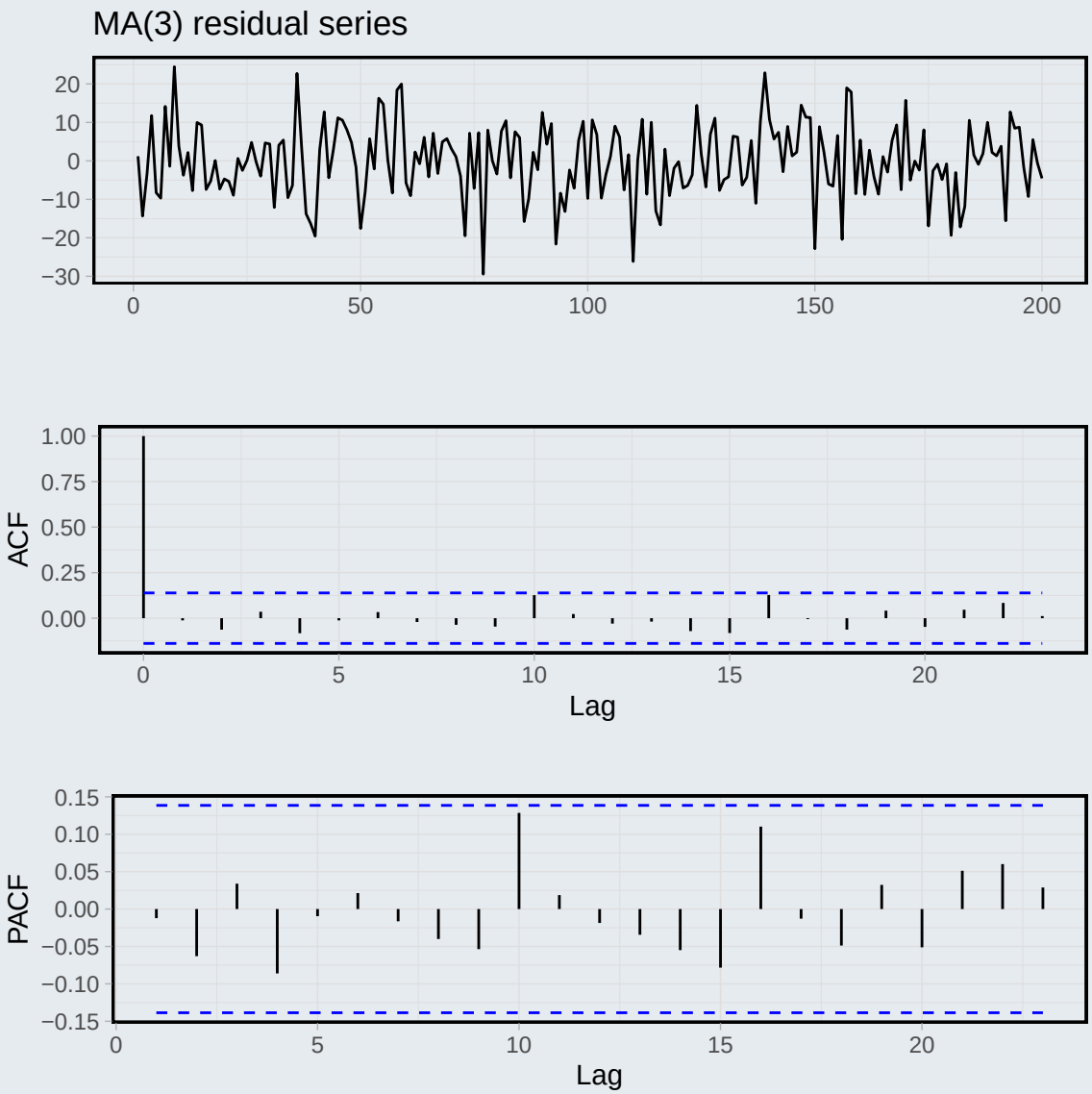
The time plot suggests the residual series is stationary, while the autocorrelation function (correlogram) suggests an MA(3) process is appropriate. Note that the partial autocorrelation function does not really tell us a great deal here. Finally we fit an MA(3) model to the data using the R function `arima()`, and plot the residuals to ensure they are independent.

```
## Fit an MA(3) model to the data
model.ma <- arima(residual.series, order=c(0,0,3))
model.ma
```

```
Call:
arima(x = residual.series, order = c(0, 0, 3))

Coefficients:
         ma1     ma2     ma3  intercept
      0.5921  0.2630  0.5375    -0.0468
s.e.  0.0566  0.0838  0.0664     1.6203

sigma^2 estimated as 92.77:  log likelihood = -737.63,  aic = 1485.26
```

```
# Plotting
p <- autoplot(model.ma$residuals, main="MA(3) residual series")
pacf <- autoplot(acf(model.ma$residuals, plot = FALSE),
                 main="", ylab="ACF")
ppacf <- autoplot(pacf(model.ma$residuals, plot = FALSE),
                 main="", ylab="PACF")
```

```
grid.arrange(p, pacf, ppacf, nrow=3)
```



The residuals look independent, so the model below is appropriate.

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \lambda_3 Z_{t-3} + Z_t$$

Additional resources on AR and MA processes

Time Series Analysis with Applications in R by Cryer and Chan:

- **Chapter 4**: Models for stationary time series

Time Series Analysis and Its Applications: With R Examples by Shumway and Stoffer:

- **Chapter 3**: ARIMA Models

Week 7 learning outcomes

By the end of this week, you should be able to:

- identify when an autoregressive process is appropriate
- fit autoregressive processes in R to model short-term correlation
- identify when a moving average process is appropriate
- fit moving average processes in R to model short-term correlation the partial autocorrelation function

Answers to tasks

Answer to Task 1. The characteristic polynomial is given by

$$\begin{aligned} X_t &= 0.75X_{t-1} - 0.125X_{t-2} + Z_t \\ X_t - 0.75X_{t-1} + 0.125X_{t-2} &= Z_t \\ (1 - 0.75B + 0.125B^2)X_t &= Z_t \end{aligned}$$

So solving $1 - 0.75B + 0.125B^2 = 0$ using the quadratic formula gives

$$\begin{aligned} \text{roots} &= \frac{0.75 \pm \sqrt{(-0.75)^2 - 4 \times 0.125 \times 1}}{2 \times 0.125} \\ \text{roots} &= \frac{0.75 \pm 0.25}{0.25} \\ \text{roots} &= 2 \text{ and } 4 \end{aligned}$$

Therefore as both roots have modulus greater than one the process is stationary.

Answer to Task 2. For`data1` the `model` argument of `arima.sim()` takes the value `list(ar=0.6)`, suggesting that there is only one AR coefficient and it's equal to 0.6. This gives the model equation for `data1` as

$$X_t = 0.6X_{t-1} + Z_t$$

Similarly for `data2` we have two AR coefficients, and the model equation is given by

$$X_t = 0.75X_{t-1} - 0.125X_{t-2} + Z_t$$

Answer to Task 3. For panel (a) an AR(2) process looks appropriate.

For panel (b), if anything an AR(2) process, but the significant correlations at lags 4 and 7 are slightly concerning. Actually this data were an MA(1) process.

For panel (c), an AR(1) process looks appropriate.

In panel (d) there is no evidence of short-term correlation, so a purely random process is appropriate.

Answer to Task 4. First let us read in the data:

```
temp <- read.csv(url("http://www.stats.gla.ac.uk/~tereza/rp/glasgowtemp.csv"))
dim(temp)
```

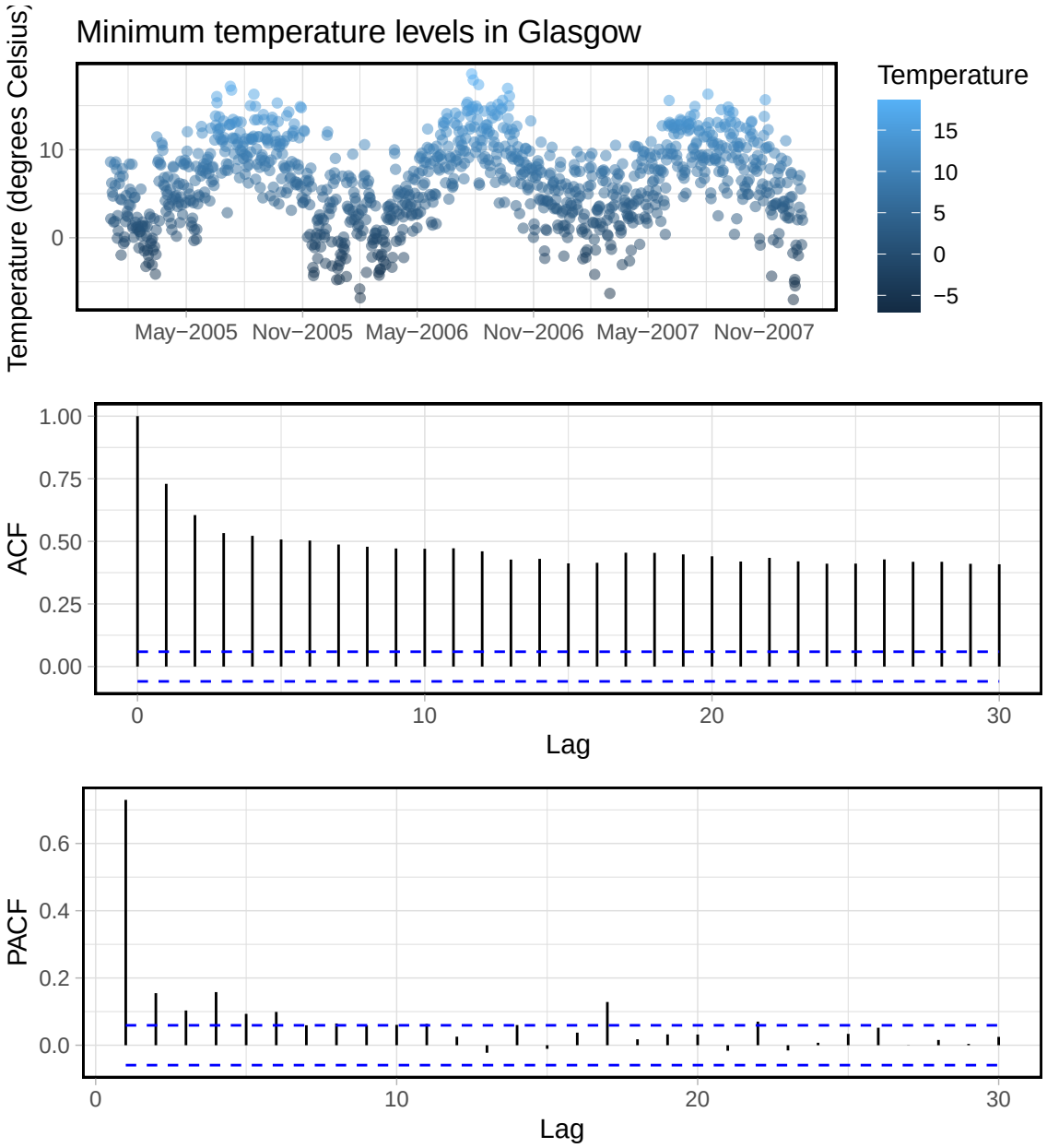
```
[1] 1095    3
```

```
head(temp)
```

```
        Date Day Temperature
1 01/01/2005   1    8.619557
2 02/01/2005   2    2.152232
3 03/01/2005   3    6.223116
4 04/01/2005   4    5.153802
5 05/01/2005   5    4.734165
6 06/01/2005   6    1.787761
```

```
# Define date variable and plot data
temp$Date2 <- as.Date(temp$Date, format="%d/%m/%Y" )

p1 <- ggplot(temp, aes(Date2, Temperature)) +
    geom_jitter(aes(colour = Temperature), alpha = 0.5) +
    xlab(" ") + scale_x_date(date_labels = "%b-%Y", date_breaks = "6 month") +
    ylab("Temperature (degrees Celsius)") + ggtitle("Minimum temperature levels in Glasgow")

p1acf <- autoplot(acf(temp$Temperature, plot=FALSE))
```

```
p1pacf<- autoplot(pacf(temp$Temperature, plot=FALSE), ylab="PACF")

grid.arrange(p1, p1acf, p1pacf, nrow=3)
```



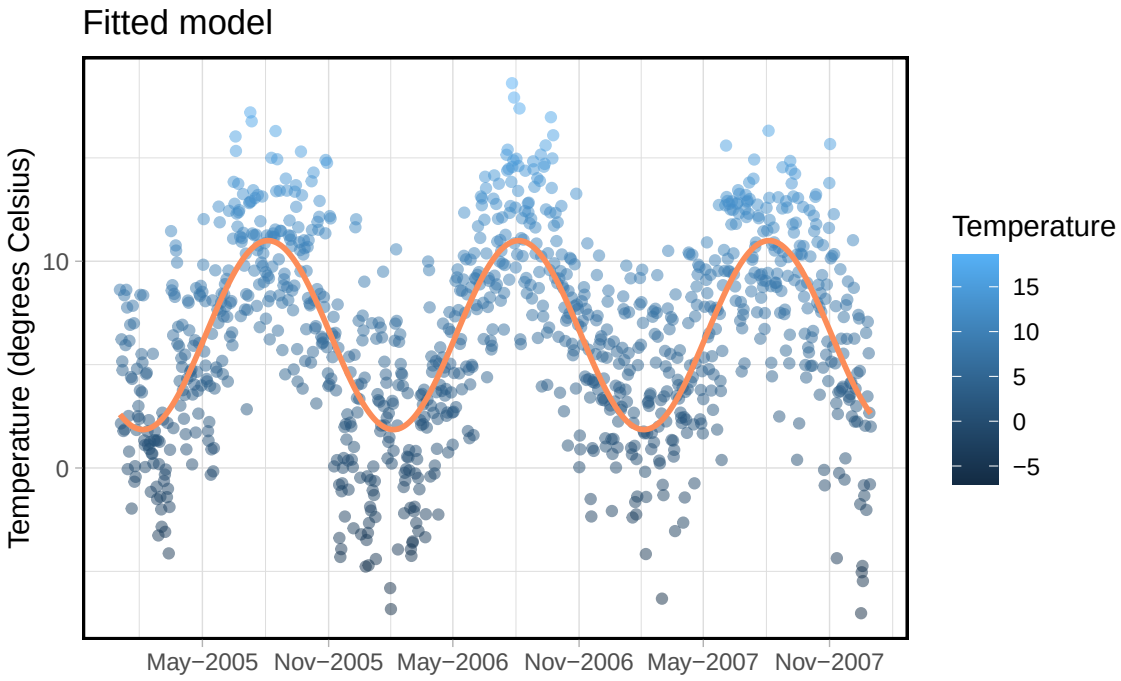
Next model the trend using a sinusoidal curve and obtain the residual series:

```
## Model the trend using a sine and cosine model
## that repeats itself after 365 days
model <- lm(temp$Temperature~sin(2*pi*temp$Day/365)+cos(2*pi*temp$Day/365))
summary(model)
```

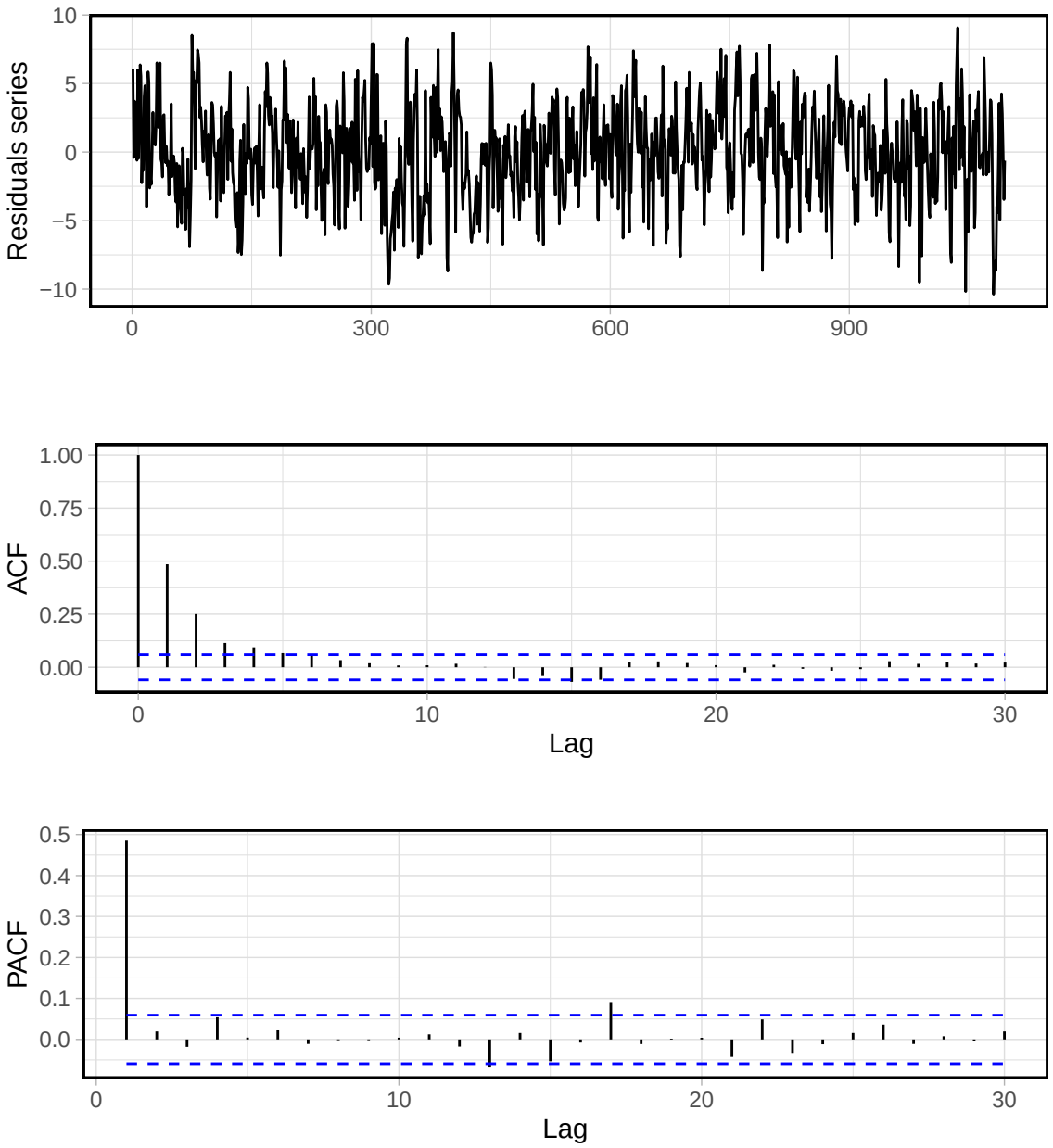
```
Call:
lm(formula = temp$Temperature ~ sin(2 * pi * temp$Day/365) +
    cos(2 * pi * temp$Day/365))

Residuals:
     Min       1Q   Median       3Q      Max
-10.3755  -2.1802  -0.0533   2.4349   9.0749

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  6.4263     0.1029   62.47   <2e-16 ***
sin(2 * pi * temp$Day/365)  -2.5477     0.1455  -17.51   <2e-16 ***
```

```
cos(2 * pi * temp$Day/365)  -3.7933     0.1455  -26.07   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.404 on 1092 degrees of freedom
Multiple R-squared:  0.4746,    Adjusted R-squared:  0.4736
F-statistic: 493.2 on 2 and 1092 DF,  p-value: < 2.2e-16
```

```r
# Fitted model
temp$fits <- model$fitted.values
p2 <- p1 + geom_line(aes(y=model$fitted.values), color = "#fc8d59", size= 1)
p2  + ggtitle("Fitted model")
```

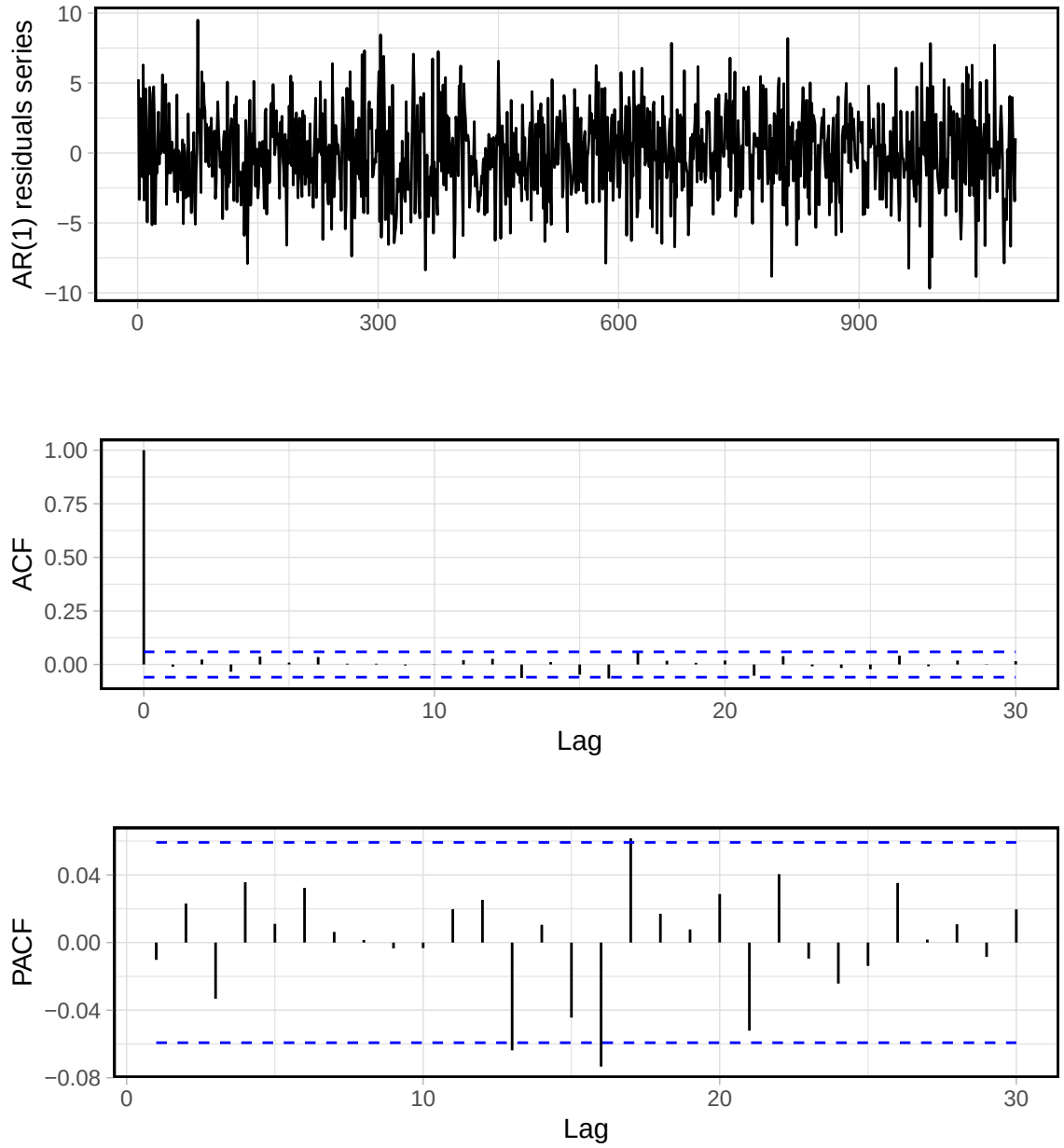



```r
# Residuals
residuals <- temp$Temperature - temp$fits

res <- autoplot(as.ts(residuals), ylab = "Residuals series")
racf <- autoplot(acf(residuals, plot=FALSE), main = "")
rpacf<- autoplot(pacf(residuals, plot=FALSE), main = "", ylab="PACF")

grid.arrange(res, racf, rpacf, nrow=3)
```



The time plot suggests that the residual series is stationary, while the partial autocorrelation function suggests that an AR(1) process is appropriate. We fit the AR(1) model as shown below and obtain the following residuals.

```
model.ar <- arima(residuals, order=c(1,0,0))

# Plotting
p3 <- autoplot(model.ar$residuals, ylab = "AR(1) residuals series")
p3acf <- autoplot(acf(model.ar$residuals, plot = FALSE),
                  main = "")
p3pacf <- autoplot(pacf(model.ar$residuals, plot = FALSE),
                  main = "", ylab="PACF")

grid.arrange(p3, p3acf, p3pacf, nrow=3)
```



The residuals look independent, so the model we fitted is appropriate.

Answer to Task 5. As shown in the lecture notes, the ACF for an MA(1) process is given by

$$\rho_\tau = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{\lambda}{1+\lambda^2} & \text{if } \tau = 1 \\ 0 & \tau > 1 \end{cases}$$

Then replacing λ with $1/\lambda$ gives

$$\rho_1 = \frac{\frac{1}{\lambda}}{1+\frac{1}{\lambda^2}} = \frac{\lambda}{\lambda^2+1}$$

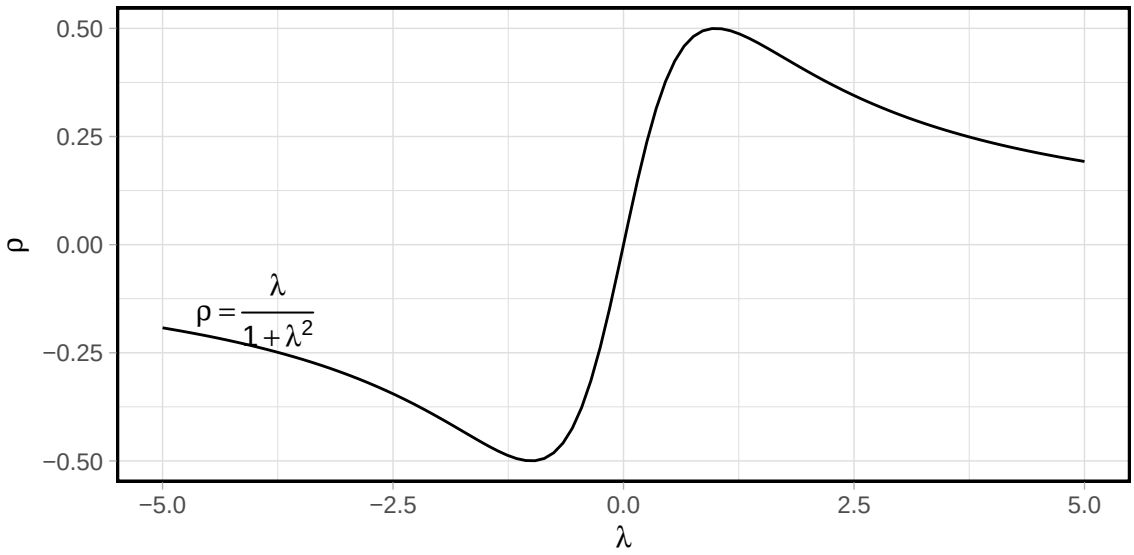
hence the autocorrelation function remains unchanged.

Then replacing λ with $1/\lambda$ gives

$$\rho_1 = \frac{\frac{1}{\lambda}}{1+\frac{1}{\lambda^2}} = \frac{\lambda}{\lambda^2+1}$$

hence the autocorrelation function remains unchanged.

When $\lambda = 1$ $\rho_1 = 0.5$ and when $\lambda = -1$ $\rho_1 = -0.5$, so no, we cannot find a λ so that $|\rho| > 1/2$. To see this simply plot ρ_1 against λ, which gives the following plot.



Answer to Task 6. The characteristic polynomial is given by

$$\theta(B) \;=\; 1 + 4.25B + B^2.$$

Solving $1 + 4.25B + B^2 = 0$ using the quadratic formula gives

$$\begin{aligned} \text{roots} &= \frac{-4.25 \pm \sqrt{(4.25)^2 - 4 \times 1 \times 1}}{2 \times 1} \\ \text{roots} &= \frac{-4.25 \pm 3.75}{2} \\ \text{roots} &= -0.25 \text{ and } -4 \end{aligned}$$

Therefore as one of the roots has modulus less than one the process is not invertible.

Answer to Task 7. Series 1 is an MA(1), Series 2 is an MA(2), Series 3 is an MA(3) and Series 4 is an AR(1). Notice that when an MA(q) process is not appropriate, the sample ACF does not become zero at low lag.