

# Autoregressive moving average models

by Kevin Kotzé

When introducing the prominent features of most economic and financial time series we noted that most of the data that we will encounter contains evidence of autocorrelation. For example, if economic growth in period 1 is high, then one would often note that the value for economic growth in period 2 is also high. Therefore, when asked to describe such data, we could possibly use a model that may be specified as,

$$y_t = \phi y_{t-1} + \varepsilon_t$$

If we do not incorporate the persistence in the data in the explanatory portion of the model, (i.e.  $\phi y_{t-1}$ ), then it will be captured in the unexplained error term,  $\varepsilon_t$ . This would be of concern, as in those cases where there is a degree of persistence in the error, the condition that  $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_{\varepsilon_t}^2)$  would be violated. The result of this would be that the parameter estimates would be inefficient and their standard errors (which measure the certainty with which we are able to estimate these parameters) would be incorrect.

In this chapter we consider one of the dominant approaches to modelling this form of persistence in time series data, with the aid of the method that was first articulated in Box and Jenkins (1979). A more recent exposition of this classical text is contained in Box, Jenkins, and Reinsel (2008), where they make extensive use of autocorrelation and partial autocorrelation functions to identify the structure of an ARMA( $p, q$ ) model.

## 1 Moving average (MA) models

The first models that we will consider in detail pertain to those that seek to describe a moving average (MA) process, which is a linear combination of white noise errors (i.e.  $\varepsilon_t$ ). The simplest variant of this model is the MA(1) that may be expressed as,

$$y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \quad (1.1)$$

where  $\mu$  is a constant,  $y_t$  is the weighted sum of a constant,  $\mu$ , and the two most recent values of  $\varepsilon_t$  and  $\varepsilon_{t-1}$ , which are assumed to be independent and identically distributed white noise,  $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$ .

To ascertain whether the MA(1) process is stationary, we need to calculate the different moments of the underlying process. To calculate the mean (i.e. the first moment) of the process,

$$\begin{aligned} \mathbb{E}[y_t] &= \mathbb{E}[\mu + \varepsilon_t + \theta \varepsilon_{t-1}] \\ &= \mu + \mathbb{E}[\varepsilon_t] + \theta \mathbb{E}[\varepsilon_{t-1}] \\ &= \mu \end{aligned} \quad (1.2)$$

Since the error terms are i.i.d. and their expected mean value is zero, the mean for this process is  $\mu$ . We are then able to conclude that the mean of the process is constant and does not depend on time. To

calculate the second moment of the process we compute the variance as,

$$\begin{aligned}
 \text{var}[y_t] &= \mathbb{E}[y_t - \mathbb{E}[y_t]]^2 \\
 &= \mathbb{E}[(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) - \mu]^2 \\
 &= \mathbb{E}[\varepsilon_t]^2 + 2\theta\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] + \mathbb{E}[\theta\varepsilon_{t-1}]^2 \\
 &= \sigma^2 + 0 + \theta\sigma^2 \\
 &= (1 + \theta^2)\sigma^2
 \end{aligned} \tag{1.3}$$

where  $\mathbb{E}[\varepsilon_t] = 0$  and  $\mathbb{E}[(\varepsilon_t)^2] = \sigma^2$ . Once again, this value is constant and does not depend on time. The first covariance may then be calculated as,

$$\begin{aligned}
 \text{cov}[y_t, y_{t-1}] &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-1} - \mathbb{E}[y_{t-1}])] \\
 &= \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\
 &= \mathbb{E}[\varepsilon_t\varepsilon_{t-1}] + \theta\mathbb{E}[\varepsilon_{t-1}^2] + \mathbb{E}[\theta\varepsilon_t\varepsilon_{t-2}] + \mathbb{E}[\theta^2\varepsilon_{t-1}\varepsilon_{t-2}] \\
 &= 0 + \theta\sigma^2 + 0 + 0 \\
 &= \theta\sigma^2
 \end{aligned} \tag{1.4}$$

And in the more general case where we have a lag order of  $j$ ,

$$\begin{aligned}
 \text{cov}[y_t, y_{t-j}] &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-j} - \mathbb{E}[y_{t-j}])] \\
 &= \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1})] \\
 &= 0 \quad \text{for } j > 1
 \end{aligned} \tag{1.5}$$

Since the mean in (1.2), the variance in (1.3), and the covariances in (1.4) and (1.5) depend on time, the MA(1) process is covariance stationary. Note that such a MA(1) process is stationary regardless of the value  $\theta$ .

The autocorrelation function (ACF) for a MA(1) process may then be derived from the expression,

$$\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}.$$

In this case  $\gamma(0)$  would refer to the variance and  $\gamma(j)$  would refer to the covariance for lag  $j$ . This would imply that,

$$\begin{aligned}
 \rho(1) &= \frac{\theta\sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{(1 + \theta^2)} \\
 \rho(j) &= 0 \quad \text{for } j > 1
 \end{aligned}$$

We see that for lag orders  $j > 1$ , the autocorrelations are zero. Hence, the autocorrelation function for a MA(1) process will go very quickly to zero as  $j$  becomes large.



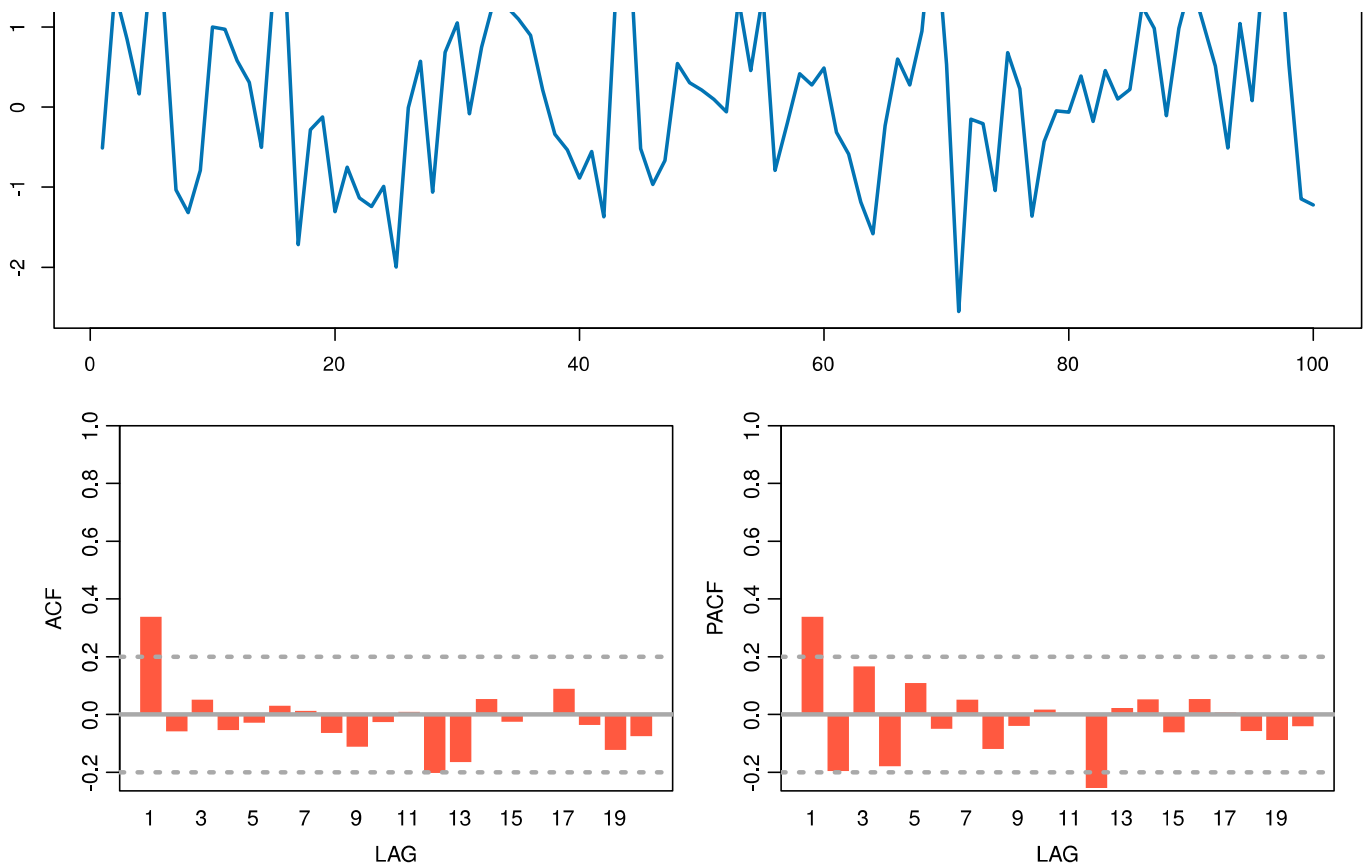
Figure 1: **MA(1)** with  $\theta = 0.5$ 

Figure 1 plots an MA(1) processes, with  $\theta = 0.5$ . The corresponding ACF, which is a plot of  $\rho(j)$  against  $j$ , is included in this diagram. Positive values of  $\theta$  induce positive autocorrelation in the series. As such, an unusually large value of  $y_t$  is likely to be followed by a larger-than-average value for  $y_{t+1}$ , and vice versa. In contrast, negative values of  $\theta$  would imply negative autocorrelation. Hence, a large positive value for  $y_t$  is possibly going to be followed by a large negative value for  $y_{t+1}$ .

Note that for the MA(1) processes, the ACF function returns very quickly to zero after one lag. Hence, we note that for our simulated MA(1) process that includes a small amount of white noise, for lags orders  $j > 1$ , the autocorrelations are largely equivalent to zero. Note also that in this case the PACF represents a slowly decaying function.

## 1.1 Higher order MA process

A finite MA( $q$ ) process will have the following form:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1.6)$$

And the infinite-order moving average process, MA( $\infty$ ), may be expressed as,

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} = \mu + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \quad (1.7)$$

with  $\theta_0 = 1$ . It can be shown that results for a first order MA(1) carry over to an infinite order MA( $\infty$ ),

provided the following relationship is satisfied,

$$\sum_{j=0}^{\infty} |\theta_j| < \infty \quad (1.8)$$

which implies that the coefficients are absolute summable.<sup>1</sup> An  $MA(\infty)$  process with absolute summable coefficients has absolute summable covariances. Moreover, the process is covariance-stationary when,

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty \quad (1.9)$$

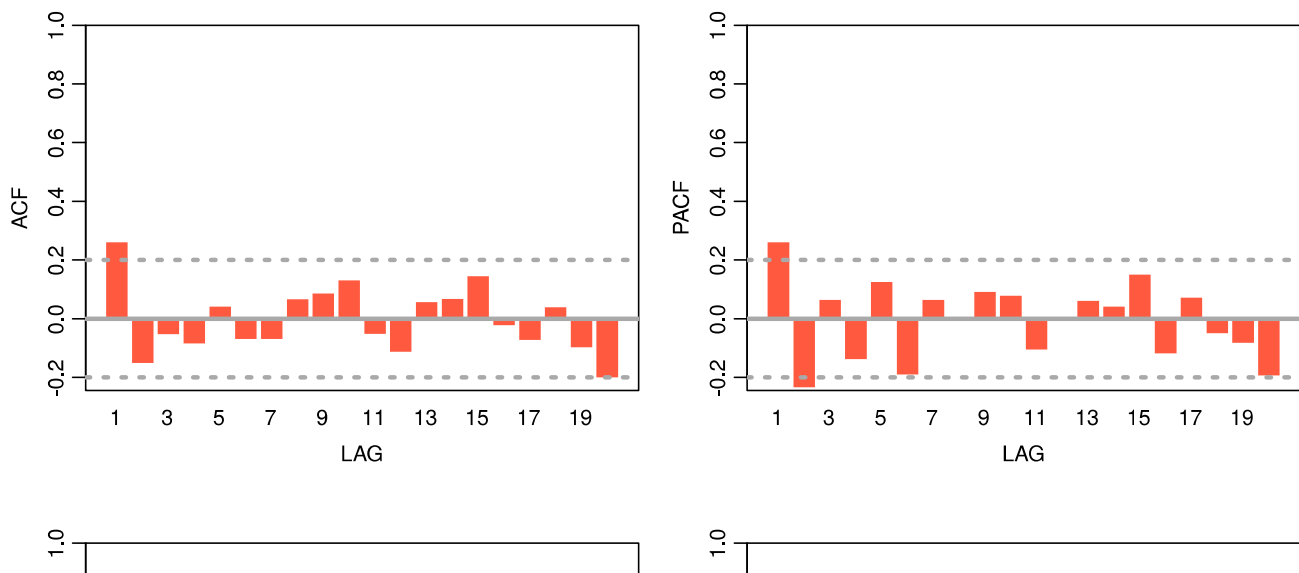
The importance of these conditions will become apparent after we show how the moving average process is related to the autoregressive representation. It will also allow us to describe the conditions that give rise to these extreme cases.

## 1.2 The identification of the order of MA processes

As noted previously, the ACF may be used to display the effect of a shock on a particular process. When dealing with a  $MA(1)$  process that takes the form  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ , the effect of the shock in the previous period,  $\varepsilon_{t-1}$ , would affect the value of  $y_t$ . Hence, the value for the first autocorrelation,  $\rho(1)$ , in the ACF should be significantly different from zero, while the remaining autocorrelation values,  $\{\rho(2), \rho(3), \dots\}$ , should not be significantly different from zero.

Similarly, in the case of a  $MA(2)$  process that takes the form  $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$ , the values for the first two autocorrelation coefficients,  $\rho(1)$  and  $\rho(2)$ , should be significantly different from zero in the ACF, while the remaining autocorrelation values should take on a value that is close to zero.

Therefore, if we simulate a number of MA processes, we should be able to identify the order of the process from the number of autocorrelations that are different from zero. An example of such an exercise is provided in Figure 2, where we have simulated MA processes with different coefficient values for three different MA processes.



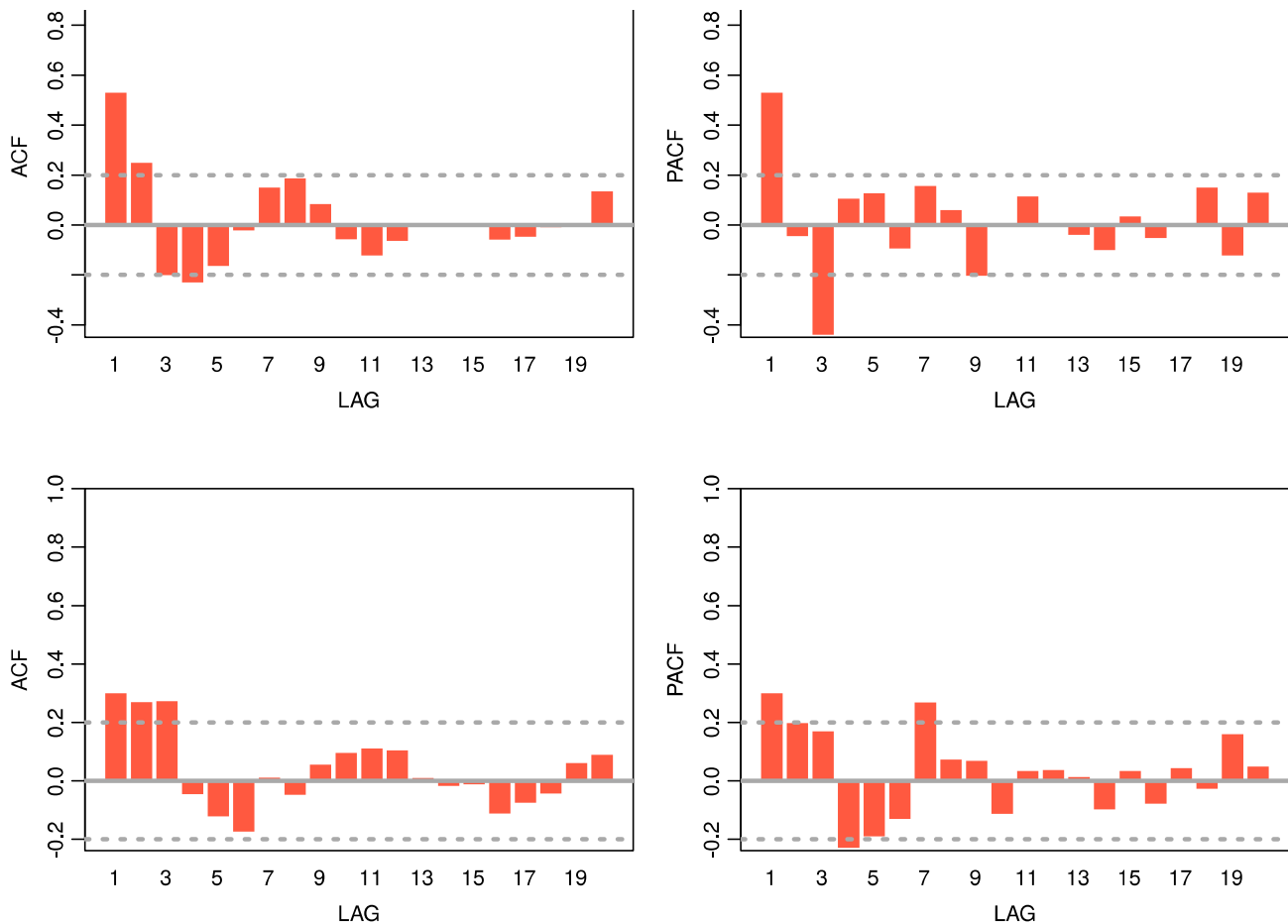


Figure 2: **Simulated MA(1) with  $\theta_1 = 0.5$ , MA(2) with  $\theta_1 = 0.6$  and  $\theta_2 = 0.7$ , MA(3) with  $\theta_1 = 0.3$ ,  $\theta_2 = 0.4$ , and  $\theta_3 = 0.6$**

## 2 Autoregressive (AR) process

Another important time series process may be described with the aid of an autoregressive model. For example where we believe that the underlying process exhibits a certain degree of persistence, we may wish to describe the process as,

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (2.1)$$

The AR(1) relates the value of a variable  $y$  at time  $t$ , to its previous value at time  $(t - 1)$  and a random disturbances  $\varepsilon$ , also at time  $t$ . Once again, we assume that  $\varepsilon_t$  is independent and identically distributed white noise,  $\varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, \sigma^2)$ .

In the previous chapter we showed that if  $|\phi| < 1$ , the AR(1) is covariance-stationary, with finite variance,

$$\begin{aligned} \mathbb{E}[y_t] &= 0 \\ \text{var}[y_t] &= \frac{\sigma^2}{1 - \phi^2} \\ \text{cov}[y_t, y_{t-j}] &= \phi^j \text{var}[y_t] \end{aligned} \quad (2.2)$$

To prove these conditions, we could make use recursive substitution, the method of undetermined

coefficients or lag operators. The use of recursive substitution and lag operators is consider below, as it is highly intuitive.<sup>2</sup>

## 2.1 Solving an AR(1) with recursive substitution

We can solve  $y_t = \phi y_{t-1} + \varepsilon_t$  with recursive substitution, where we start at some period of time and continually substitute in expressions on the right-hand side.

$$\begin{aligned}
 y_t &= \phi y_{t-1} + \varepsilon_t \\
 &= \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
 &= \phi^2(\phi y_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\
 &= \vdots \\
 &= \phi^{j+1} y_{t-(j+1)} + \phi^j \varepsilon_{t-j} + \dots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t
 \end{aligned} \tag{2.3}$$

where we have made use of the expression,  $y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1}$ , in the second line. After we substitute in repeated terms for the right-hand side, we see the emergence of a pattern that is captured in the final expression.

The final line in (2.3) explains  $y_t$  as a linear function of an initial value  $y_{t-(j+1)}$  and the historical values of  $\varepsilon_t$ . If  $|\phi| < 1$  and  $j$  is large,  $\phi^{j+1} y_{t-(j+1)} \rightarrow 0$ . Thus, the AR(1) can be expressed as an MA( $\infty$ ) when  $|\phi| < 1$ . This is an extremely important concept in the analysis of time series and we will return to it on a number of occasions.

Note that if  $|\phi| > 1$  then  $\phi^j \rightarrow \infty$  as  $j \rightarrow \infty$ . Hence, this would suggest that when an AR(1) model is represented by a random walk process, the sum of all the coefficients in an equivalent moving average representation would equate to infinity (i.e. it is not summable).

## 2.2 Solving an AR(1) with lag operators

The application of lag operators is particularly useful when dealing with more complex model structures, such as those for vector autoregressive models. In this relatively simple case, the lag operator the autoregressive model,  $y_t = \phi y_{t-1} + \varepsilon_t$ , can be written as,

$$(1 - \phi L) y_t = \varepsilon_t \tag{2.4}$$

A sequence  $\{y_t\}_{t=-\infty}^{\infty}$  is said to be bounded if there exists a finite number  $k$ , such that  $|y_t| < k$  for all  $t$ . Provided  $|\phi| < 1$ , and we restrict ourselves to bounded sequences, the expression in (2.4) can be multiplied by  $(1 - \phi L)^{-1}$  on both sides of the equality sign to obtain,

$$\begin{aligned}
 (1 - \phi L)^{-1} (1 - \phi L) y_t &= (1 - \phi L)^{-1} \varepsilon_t \\
 y_t &= (1 - \phi L)^{-1} \varepsilon_t
 \end{aligned} \tag{2.5}$$

This expression could then be expanded as follows,

$$y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \dots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \tag{2.6}$$

To consider the mechanism behind this, firstly note that if  $t \rightarrow \infty$ , we can under the assumption that  $|\phi| < 1$ , apply the following approximation (also known as the geometric rule),

$$(1 - \phi L)^{-1} = \lim_{j \rightarrow \infty} (1 + \phi L + (\phi L)^2 + \dots + (\phi L)^j) \quad (2.7)$$

This is based on the result,  $(1 - z)^{-1} = 1 + z + z^2 + z^3 + \dots$ , which holds if  $|z| < 1$ . Hence, as  $(1 - \phi L)^{-1} (1 - \phi L) = 1$ , the left-hand side of equation (2.6) follows naturally from equation (2.5), while the right-hand side of equation (2.6) would also follow from equation (2.5) after employing the result stated in (2.7).

Importantly, we can only perform this inversion (multiplication by  $(1 - \phi L)^{-1}$  if  $|\phi| < 1$ ). When it holds, we say that the process is invertible.

As an alternative, equation (2.6) could be written with the aid of the notation for the coefficients that we used for the MA( $\infty$ ) processes,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \dots = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad (2.8)$$

Therefore, when  $|\phi| < 1$ , the following equality is satisfied,

$$\sum_{j=0}^{\infty} |\theta_j| = \sum_{j=0}^{\infty} |\phi^j| \quad (2.9)$$

In this case, the finite sum of  $\phi^j$  equals  $\frac{1}{1-\phi}$ . This ensures that when an AR( $p$ ) process is stationary, an MA( $\infty$ ) representation exists.

## 2.3 Mean, variance and autocovariance of an AR(1)

To compute the unconditional first and second order moments of a stable AR(1) process, we can use the MA( $\infty$ ) representation of the process, derived in equation (2.6), and the properties of Gaussian white noise that were defined in the previous chapter. The mean could then be calculated as,

$$\mathbb{E}[y_t] = \mathbb{E}[\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \dots] = 0 \quad (2.10)$$

The variance is then,

$$\begin{aligned} \gamma[0] &= \text{var}[y_t] = \mathbb{E}[y_t - \mathbb{E}[y_t]]^2 \\ &= \mathbb{E}[\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \dots]^2 \\ &= \text{var}[\varepsilon_t] + \phi^2 \text{var}[\varepsilon_{t-1}] + \phi^4 \text{var}[\varepsilon_{t-2}] + \phi^6 \text{var}[\varepsilon_{t-3}] + \dots \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \dots) \sigma^2 \\ &= \frac{1}{1 - \phi^2} \sigma^2 \end{aligned} \quad (2.11)$$

The first order covariance is,

$$\begin{aligned}
\gamma(1) &= \mathbb{E} \left[ (y_t - \mathbb{E}[y_t]) (y_{t-1} - \mathbb{E}[y_{t-1}]) \right] \\
&= \mathbb{E} \left[ (\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \dots) (\varepsilon_{t-1} + \phi \varepsilon_{t-2} + \phi^2 \varepsilon_{t-3} + \phi^3 \varepsilon_{t-4} + \dots) \right] \\
&= (\phi + \phi^3 + \phi^5 + \dots) \sigma^2 = \phi (1 + \phi^2 + \phi^4 + \phi^6 + \dots) \sigma^2 \\
&= \phi \frac{1}{1 - \phi^2} \sigma^2 \\
&= \phi \text{var}[y_t]
\end{aligned} \tag{2.12}$$

While for  $j > 1$  we have,

$$\gamma(j) = \mathbb{E} \left[ (y_t - \mathbb{E}[y_t]) (y_{t-j} - \mathbb{E}[y_{t-j}]) \right] = \phi^j \text{var}[y_t] \tag{2.13}$$

Thus, we have finally derived a proof for the results that were stated in (2.2).

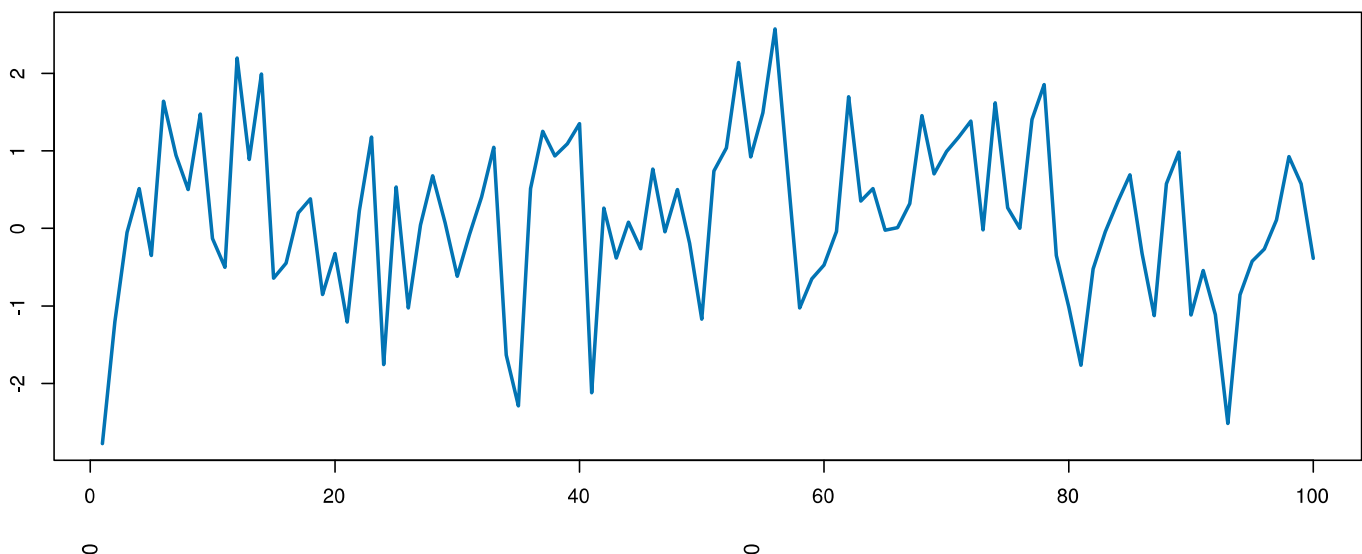
It is worth noting that the autocorrelation function for an AR(1) process will always coincide with the impulse response function,  $\partial y_{t+j} / \partial \varepsilon_t$ . In this case the ACF for an AR(1) for  $j = \{1, \dots, J\}$ , is given by

$$\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1, \quad \rho(1) = \frac{\gamma(1)}{\gamma(0)} = \phi, \quad \dots, \quad \rho(j) = \frac{\gamma(j)}{\gamma(0)} = \phi^j \tag{2.14}$$

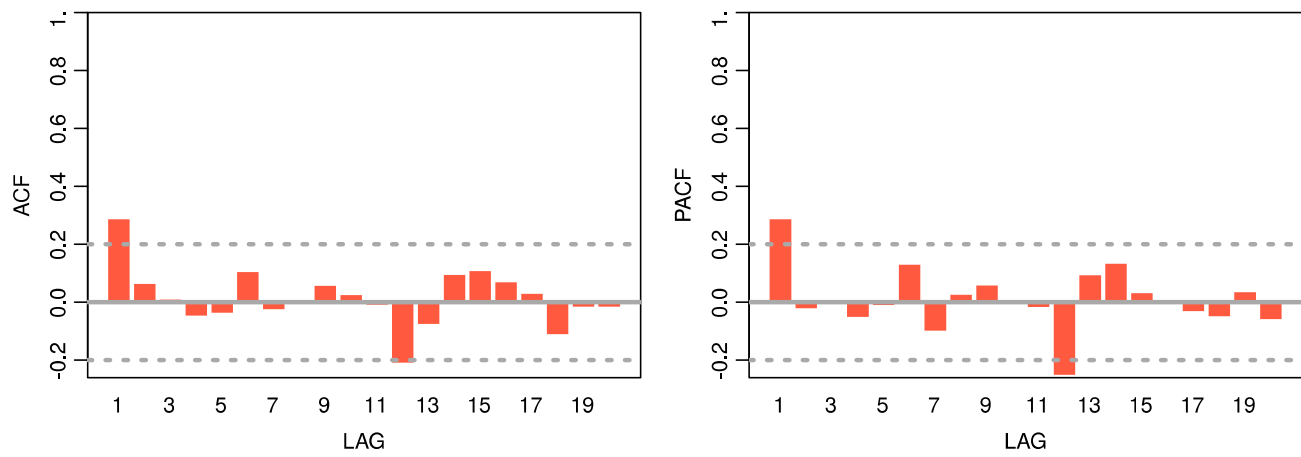
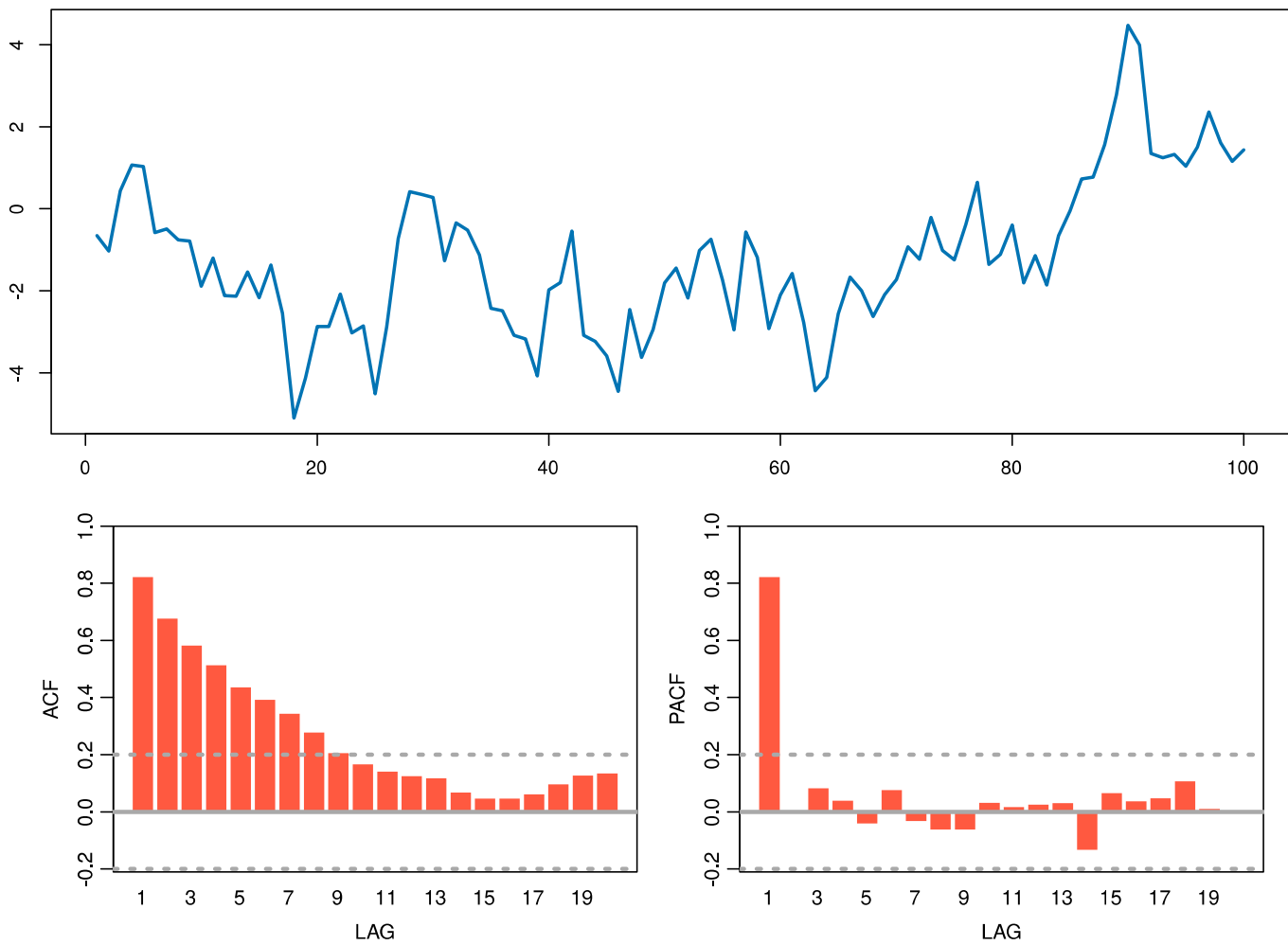
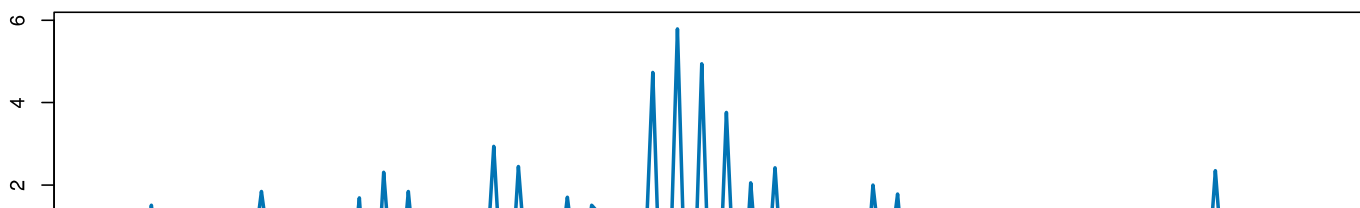
It is also worth noting that the coefficients relating to the ACF in (2.14) are equivalent to those that are contained in the moving average representation of the autoregressive process in (2.6) and (2.10). In addition, the expression in (2.10) would facilitate a convenient way for calculating the dynamic multipliers that are summarised by the impulse response function since,

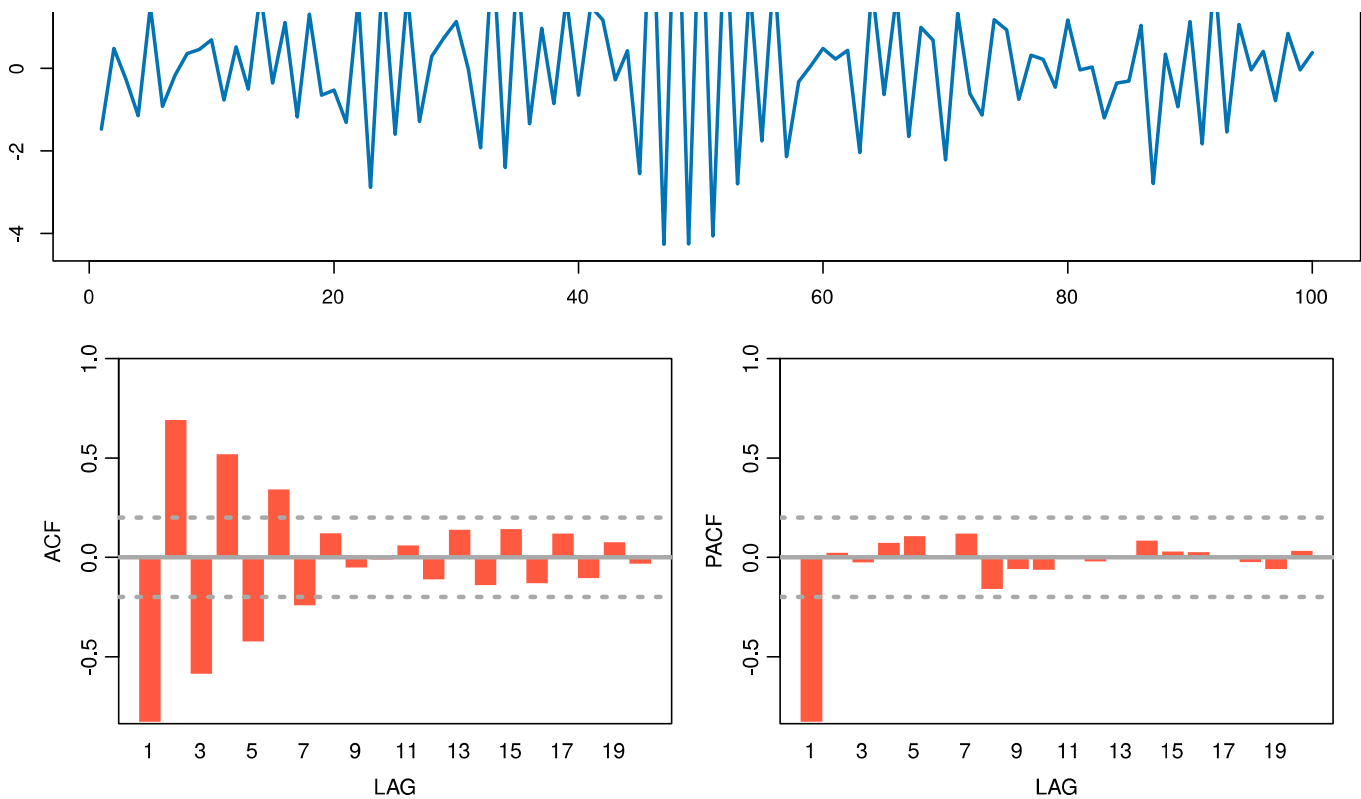
$$\frac{\partial y_t}{\partial \varepsilon_t} = 1, \quad \frac{\partial y_t}{\partial \varepsilon_{t-1}} = \phi, \quad \frac{\partial y_t}{\partial \varepsilon_{t-2}} = \phi^2, \quad \dots, \quad \frac{\partial y_t}{\partial \varepsilon_{t-j}} = \phi^j \tag{2.15}$$

Figure 3 to 4 plot three different AR(1) processes, with different coefficient values. The first of these, in Figure 3, shows the result of a simulated AR(1) process with  $\phi = 0.4$ , while Figure 5 shows the results of a simulated AR(1) process with  $\phi = 0.9$ . Then finally, the results of simulated AR(1) process with  $\phi = -0.8$  is shown in Figure 4. The respective ACF and PACF functions for these processes have been included with each of the figures.





Figure 3: **AR(1)** with  $\phi = 0.4$ Figure 4: **AR(1)** with  $\phi = 0.9$ 

Figure 5: **AR(1)** with  $\phi = -0.8$ 

Since the AR(1) processes are stationary by construction (i.e.  $|\phi| < 1$ ), all the simulated processes fluctuate around a constant mean. However, the fluctuations in the  $\phi = 0.9$  case are much more persistent than those of the  $\phi = 0.4$  case (as can be seen from the ACF plots). When there is a high degree of persistence, shocks that occur far back in time would continue to affect  $y_t$ , but by a relative smaller amount than shocks from the immediate past. For the AR(1) process with  $\phi = 0.4$ , the effects of the shocks have more-or-less faded out after 7 to 8 periods. For the AR(1) process with  $\phi = 0.9$ , the effect of the shocks are considered to be more persistent, as noted by the ACF in Figure 5. Of course, even in the case of  $\phi = 0.9$ , the effect of the shocks are certainly *not* permanent.

When the autoregressive coefficient is negative, as in the  $\phi = -0.8$  case, large positive values of  $y_t$  are likely to be followed by negative values for  $y_{t+1}$ . Hence, the ACF oscillates between positive and negative values, depending on whether  $j$  is an even or odd number. However, as for the  $\phi = 0.9$  case, the effect of a shock will eventually dissipate with time.

When comparing the visual representation of GDP growth in the previous chapter, with the AR(1) simulated data in Figure 3, the degree of persistence is relatively similar. Of course, the exact timing of oscillations in GDP growth and the simulated data do not coincide, but the general pattern and frequency of movements around a constant mean seems to be similar. Thus, knowledge about the theoretical properties of AR processes is of relevance when seeking to describe real world data. During the tutorial we will seek to fit an autoregressive model to South African GDP growth to see how this particular model could be used to describe the underlying data generating process.

### 2.3.1 Adding a constant

Let us now add a constant to the AR(1) model, where,

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \quad (2.16)$$

To ascertain how the results in equations (2.10) to (2.13) change, we can define  $v_t = \mu + \varepsilon_t$ , such that,

$$\begin{aligned} y_t &= \phi y_{t-1} + v_t \\ y_t &= (1 - \phi L)^{-1} v_t \\ &= \left( \frac{1}{1 - \phi} \right) \mu + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots \end{aligned} \quad (2.17)$$

Therefore, the unconditional first moment is,

$$\mathbb{E}[y_t] = \left( \frac{1}{1 - \phi} \right) \mu \quad (2.18)$$

Although the expected mean is no longer zero it also does not dependent on time. After deriving the variance and covariance for this process, one would be able to show that it will continue to be stationary, as long as  $|\phi| < 1$ .

## 2.4 Higher order AR processes

As was shown earlier, it is possible to generalise a MA(1) process to an infinite order MA( $\infty$ ) process, where the same rules apply. However, for higher-order AR processes, things become a bit more complicated. For example, consider the AR(2) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (2.19)$$

In this case we are no longer able to consider the value of  $\phi_1$  alone, to determine whether or not it is stationary, as we now also have to consider the effect of  $\phi_2$ . As it turns out, there are different ways of doing this. The way we will do it here, which is consistent with the description that will be provided when discussing vector autoregressive models, makes use of the premise that we have to rewrite the AR(2) expression in equation (2.19) as a first-order difference equation, by using matrix notation. For example, consider the vector  $Z_t$ , which is of dimension  $(2 \times 1)$ ,

$$Z_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$

To apply this notation to the error term, we make use of a  $(2 \times 1)$  vector for  $v_t$ , which takes the form,

$$v_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

The coefficients will then be contained in a  $(2 \times 2)$  matrix,  $\Gamma$ , where

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \quad (2.22)$$

The following first-order difference equation could then take the form,

$$Z_t = \Gamma Z_{t-1} + v_t \quad (2.23)$$

which would be equivalent to,

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \quad (2.24)$$

The matrix  $\Gamma$  in (2.22) is the so-called *companion form* matrix of the AR(2) process. As described in the chapter on vector autoregressive models, any higher order autoregressive process can be represented as a first-order autoregressive process using this formulation and generalizations of the companion form matrix.

With  $\Gamma$  at hand we may check for stationarity by computing the eigenvalues of this matrix. Moreover, the eigenvalues of  $\Gamma$  are two solutions of the  $x$  polynomial that satisfy the characteristic equation:

$$x^2 - \phi_1 x - \phi_2 = 0 \quad (2.25)$$

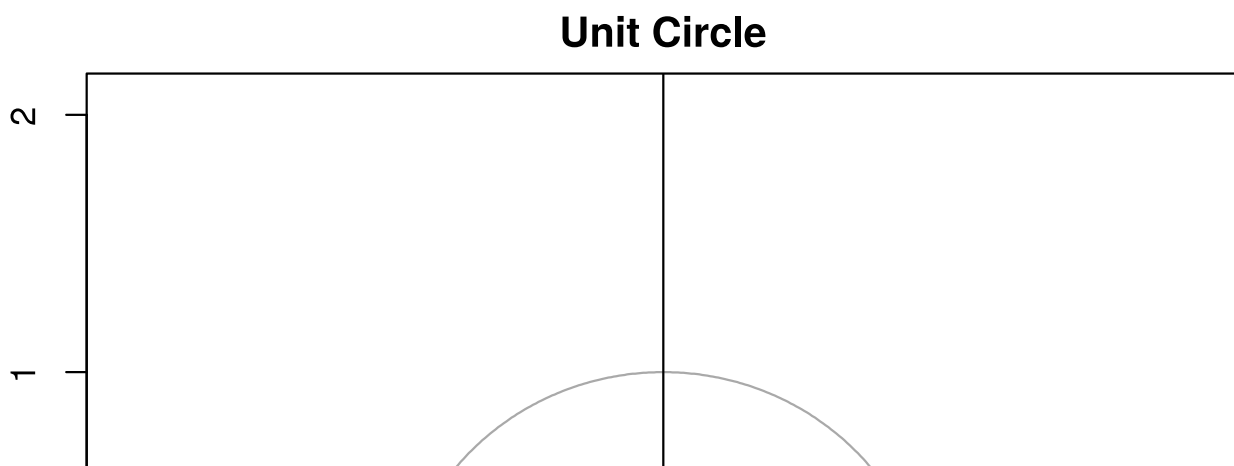
These eigenvalues ( $m_1$  and  $m_2$ ) must then satisfy  $(x - m_1)(x - m_2)$ , and can be found from the solution formula:

$$m_1, m_2 = \frac{\left( \phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2} \right)}{2} \quad (2.26)$$

Stationarity requires that the eigenvalues are less than one in absolute value. In the AR(2) case, one can show that this will be the case if,<sup>3</sup>

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ -\phi_1 + \phi_2 &< 1 \\ \phi_2 &> -1 \end{aligned} \quad (2.27)$$

In the above expression, it will be the case that if the eigenvalues are less than unity then the system will be stable and stationary. As noted in Hamilton (1994), some authors express the above characteristic equation as  $1 - x^2 - \phi_1 x - \phi_2 = 0$ , where the roots would need to be outside of the unit circle if the system is to be stationary. In an attempt to avoid this potentially confusing situation, we will try to make use the eigenvalues (and reference thereto) wherever possible, as they cannot be interpreted as anything other than what is provided above (i.e. the values  $m_1$  and  $m_2$  for the case of an AR(2) model).



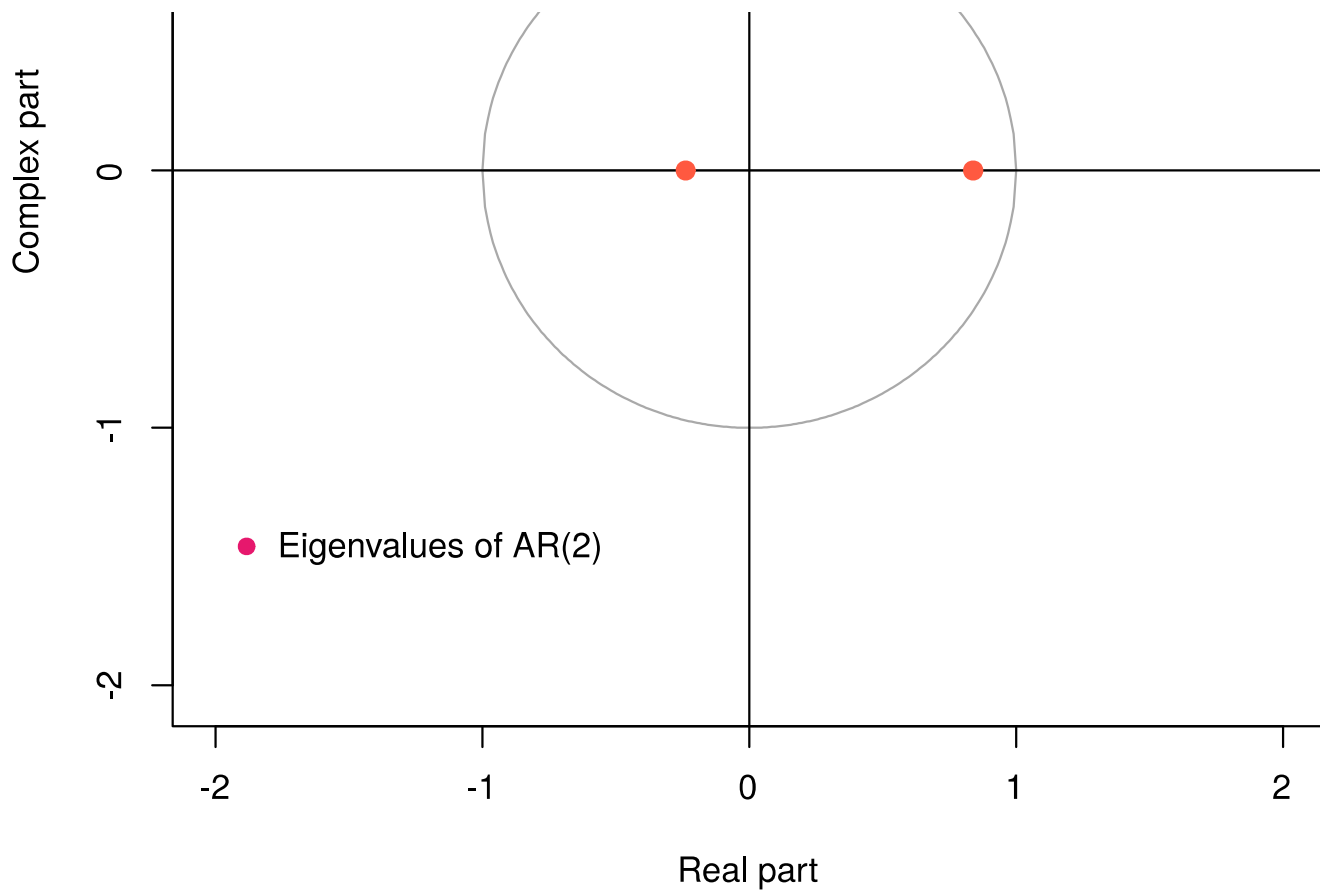


Figure 6: **Eigenvalues for difference equation**  $x^2 - 0.6x - 0.2 = 0$

Higher-order autoregressive processes, denoted  $AR(p)$ , can be written as,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.28)$$

Checking for stationarity for such processes involves the same calculations as above. In this case, the  $\Gamma$  matrix will take the form,

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (2.29)$$

Provided the roots of the characteristic equation satisfy,

$$x^p - \phi_1 x^{p-1} - \dots - \phi_p = 0 \quad (2.30)$$

it will be the case that all the eigenvalues are less than one in absolute value (i.e. they lie within the unit circle). In such an instance, the  $p^{\text{th}}$  order autoregression will be stable and stationary.

Although it might seem as if it would be arduous to investigate these conditions, they are relatively simple to compute in practice. We only need to construct the companion matrix, before we determine whether or not all the eigenvalues of this matrix are less than 1 in absolute value. Most good

econometric or statistical software packages will be able to do this for you.

## 2.5 The identification of the order of AR processes

As in the case with the identification of the order of the MA( $q$ ) process, we could make use of the ACF coefficients to determine the order of the simulated process. However, in this case the intervening lags of the AR( $p$ ) process would pass on the persistence to successive coefficients in the ACF. Thankfully, the PACF removes the effects of the persistence that is passed on from intervening lags of the AR( $p$ ) process and as a result, it may be used to identify the order of the AR( $p$ ) process.

## 3 Autoregressive moving average (ARMA) models

By combining the two models that were described above, namely the MA and AR models, we get what is called an autoregressive moving average (ARMA) model. In the simplest case, we can specify an ARMA(1,1) process as,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (3.1)$$

Using the lag polynomial described in the previous chapter a more general formulation of an ARMA model is:

$$\phi(L) y_t = \theta(L) \varepsilon_t \quad (3.2)$$

where  $\phi$  and  $\theta$  represent the (autoregressive) and (moving average) lag polynomials,

$$\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i \quad \text{and} \quad \theta(L) = 1 + \sum_{i=1}^q \theta_i L^i \quad (3.3)$$

This specification ensures that the number of lags, ( $p$ ) and ( $q$ ), can differ. For instance, an ARMA(2,1) combines an AR(2) with a MA(1) to provide:

$$\begin{aligned} (1 - \phi_1 L - \phi_2 L^2) y_t &= (1 + \theta_1 L) \varepsilon_t \\ y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \end{aligned} \quad (3.4)$$

To determine whether or not an ARMA process is stationary could focus our attention on its autoregressive past. By way of example, if we assume that we are dealing with an ARMA(1,1) process then we can test for stationarity by converting it into an equivalent moving average representation. Using the lag operator the ARMA(1,1) process can be written as,

$$(1 - \phi L) y_t = (1 + \theta L) \varepsilon_t \quad (3.5)$$

Multiplying by  $(1 - \phi L)^{-1}$  on both sides of the equality sign provides,

$$\begin{aligned} y_t &= \frac{(1 + \theta L)}{(1 - \phi L)} \varepsilon_t \\ &= (1 - \phi L)^{-1} \varepsilon_t + (1 - \phi L)^{-1} \theta_1 \varepsilon_{t-1} \end{aligned} \quad (3.6)$$

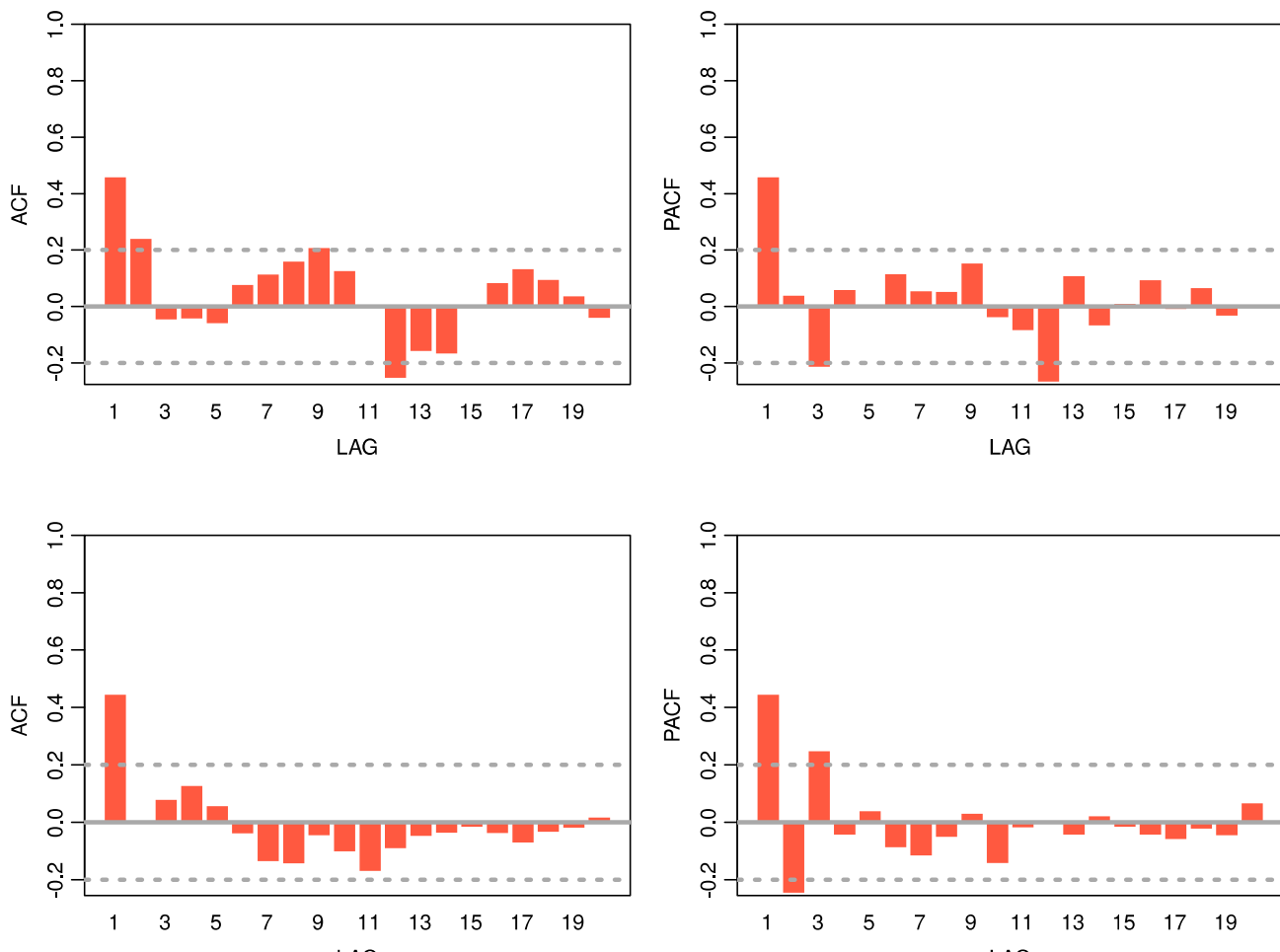
When  $|\phi| < 1$ , this can be written as the geometric process,

$$\begin{aligned}
 y_t &= \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_t + \theta \sum_{j=0}^{\infty} (\phi L)^j \varepsilon_{t-1} \\
 &= \varepsilon_t + \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j} + \theta \sum_{j=1}^{\infty} \phi^{j-1} \varepsilon_{t-j} \\
 &= \varepsilon_t + \sum_{j=1}^{\infty} (\phi^j + \theta \phi^{j-1}) \varepsilon_{t-j}
 \end{aligned} \tag{3.7}$$

Where we have used the result in equation (2.7) to expand the terms in  $(1 - \phi L)^{-1}$ . With  $|\phi| < 1$ , the weights will need to decline sufficiently for the process to have finite variance and autocovariances.

In the discussion relating to moving average processes, we noted that the ACF would give an indication of the order of the process, while the PACF would slowly decline. In contrast, when discussing the autoregressive models, we noted that the PACF would give an indication of the order of the process, while the ACF would slowly decline.

When considering the combined ARMA process we would usually note that both of these functions should decay slightly and as such it may become difficult to decipher the order of the combined ARMA model. An example of the ACF and PACF functions for an ARMA(1,1) model are provided in Figure 7, where we note that there is a trace of persistence in both functions. Note that despite the fact that we know that we are dealing with an ARMA(1,1) model, the autocorrelations in the ACF and PACF both appear to differ from zero at  $j = 2$ .



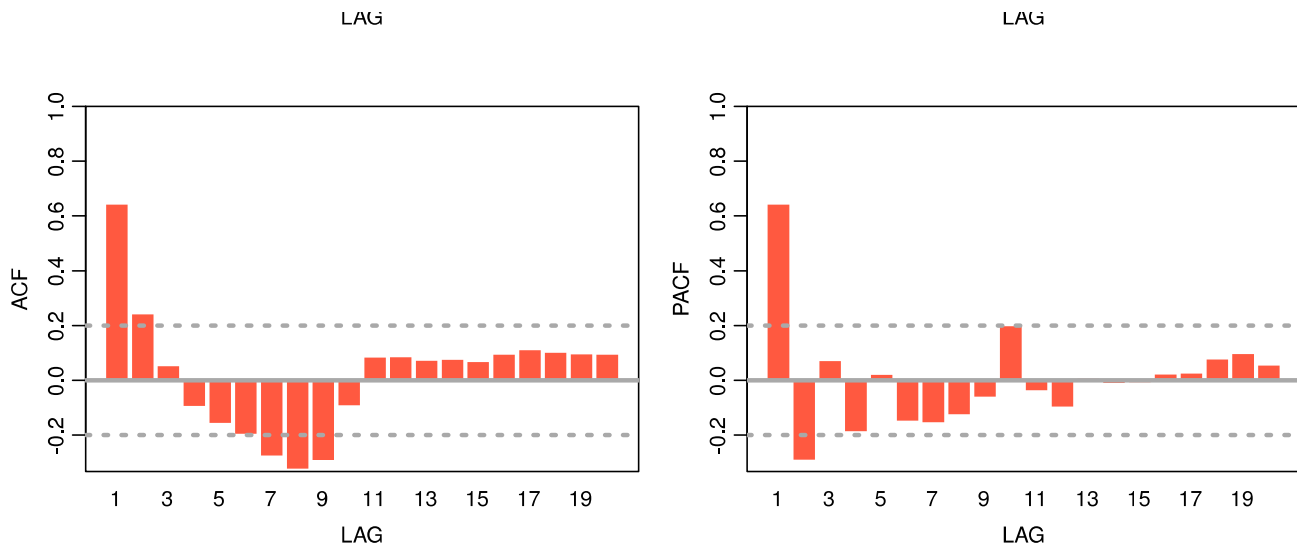


Figure 7: **Autocorrelation functions for AR(1) with  $\phi = 0.5$ , MA(1) with  $\theta = 0.6$ , and ARMA(1,1) with  $\phi = 0.5$  and  $\theta = 0.6$**

## 4 Seasonal ARMA Models

The framework of the ARMA model may be extended to model the seasonal behaviour of the underlying processes. In several cases, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag  $s$ . For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of  $s = 12$ , because of the strong connections of activities during particular months in a successive calendar years. Similarly, data that has a quarterly frequency will exhibit repetitive behaviour after  $s = 4$  quarters, where  $s$  would obviously represent the frequency of the seasonal.

As a result of this phenomena, it would be appropriate to introduce autoregressive and moving average terms that are able to identify with the seasonal lags in such a process. The resulting pure seasonal autoregressive moving average representation, may be expressed as an  $\text{ARMA}(p, q)_s$  model.<sup>4</sup> Hence, a first-order seasonal autoregressive moving average series that seeks to incorporate a monthly seasonal could be written as,

$$y_t = \phi y_{t-12} + \varepsilon_t + \theta \varepsilon_{t-12}$$

This model seeks to describe the variable  $y_t$  in terms of past lags at the multiple of the seasonal, where  $s = 12$  months. The estimation of the parameters in this model and its use in forecasting exercises would involve relatively straightforward modifications of the unit lag case that has already been considered. The stationarity conditions that were discussed previously, would also continue to hold.

When looking to identify such a process from the sample ACF and PACF functions, we would note that the values for the MA(1) with a seasonal ( $s = 12$ ), which could be written as,  $y_t = \varepsilon_t + \theta \varepsilon_{t-12}$ , it is easy to verify that

$$\begin{aligned}\gamma(0) &= (1 + \theta^2)\sigma^2 \\ \gamma(12) &= \theta\sigma^2 \\ \gamma(j) &= 0, \text{ for values where } j \neq 12\end{aligned}$$



Thus, the only non-zero autocorrelation, aside from lag zero is,  $\rho(12) = \theta/(1 + \theta^2)$ . Similarly, for the AR(1) model with seasonal ( $s = 12$ ), we could calculate,

$$\begin{aligned}\gamma(0) &= \sigma^2/(1 - \phi^2) \\ \gamma(12) &= \sigma^2\phi^k/(1 - \phi^2) \text{ for } k = 1, 2, \dots \\ \gamma(j) &= 0, \text{ for values where } j \neq 12\end{aligned}$$

These results suggest that the PACF would have the analogous extensions from non-seasonal to seasonal models.

Of course, one could allow for mixed seasonal models in an ARMA framework, where the AR element takes on a seasonal setup and the MA element takes on a nonseasonal setup. Such a model may take the form

$$y_t = \phi y_{t-12} + \varepsilon_t + \theta \varepsilon_{t-1}$$

While it would be relatively straightforward to estimate the parameters in such a model, it would be relatively difficult to identify the structural form of such a model from the ACF and PACF coefficients in practical applications.

## 5 Model specification and parameter estimation

In a real world application, the parameters and functional form of the underlying data generating process are unknown. In addition, the respective parameters in these models would then need to be estimated, before one would be able to assess the model fit. This essential procedure is encapsulated in the Box and Jenkins (1979) methodology for time series modelling, which consists of the stages for the *identification* of an appropriate functional form for the model; *estimation* of the model parameters; and *diagnostic checking* to assess the fit of the model.

### 5.1 Identification

As a part of the initial identification stage, one would plot the data that is to be modelled to inspect for any obvious outliers, missing values, or pronounced structural breaks. Thereafter, one may wish to consider whether there is a possibility that the time series is nonstationary, by considering whether there is a pronounced trend in the data.

If the data does not appear to contain any of the above features, then an ACF could be applied to the variable. If the results suggest that the autocorrelation lags return relatively quickly to zero, then the variable will obviously be stationary. However, if there is any doubt about whether or not the series is stationary, one would need to perform the tests for stationarity to determine the functional form of nonstationarity before the recommended transformations are performed to induce stationarity.

As a part of the identification stage for a stationary variable, a more detailed inspection of the autocorrelation and partial autocorrelation functions would provide an indication of the number of autoregressive and/or moving average terms that should be included in the selection of possible model specifications.

If the real world data that we are looking to model provided results that were similar to ACF and PACF

in the lowest panel of Figure 7, then we may wish to consider the use of an ARMA(2,2), an ARMA(1,2), an ARMA(2,1) or an ARMA(1,1). One could also think about using a MA(2) or an AR(2), but you would not consider a MA(1) or an AR(1).

In addition to the use of the ACF and PACF, one may also wish to calculate  $Q$ -statistic to test whether a group of autocorrelations is different from zero. This would be useful when the degree of persistence is very low and you would like to investigate whether or not the variable could be regarded as white noise. This statistic was original development by Box and Pierce (1970), while better small-sample performance was achieved following the Ljung and Box (1978) improvements. It is usually represented as,

$$Q = T(T+2) \sum_{k=1}^s \rho_k^2 / (T-k)$$

In this case, if the sample value of  $Q$  exceeds the critical value that is provided by the  $\chi^2$  distribution with  $s$  degrees of freedom, then at least one value of  $\rho_k$  is statistically different from zero at the specified significance level (i.e. the data does not represent white noise).

## 5.2 Estimation

The next stage involves the estimation of the parameters in the various candidate models. When completing this procedure it is worth bearing in mind that additional coefficients may improve the in-sample fit at the expense of a reduction in the degrees of freedom. In addition, models with less parameters often produce better out-of-sample fit. Hence, we often prefer more parsimonious to less parsimonious models.

In addition, it is also important to emphasize that the data must be stationary if the model is invertible. The distribution theory that underlies the use of sample ACF and PACF as approximations for the true data generating process assume that  $y_t$  is stationary, while the  $t$ -statistics and  $Q$ -statistics also rely on this principle. Hence, one should be extremely careful when the estimated value of  $|\phi_1|$  or  $\rho(1)$  is close to unity.

When assessing the measures for goodness-of-fit, it would obviously be the case that the most appropriate model would fit the data better than other models. This may be represented by the coefficient of determination (or  $R^2$ ) and average of the residual sum of squares. In addition, as a part of the estimation stage one should calculate the information criteria and/or apply various likelihood-ratio tests to obtain additional information that relates to the appropriate lag-order that is to be used in the model.

It is usually suggested that the application of the Akaike and Bayesian Information criteria would provide useful measures for the goodness-of-fit for a particular model, by balancing the fitted errors against the number of parameters in the model.<sup>5</sup> To calculate these measures we would firstly need to find the residual sum of squares,  $SSR_k$  under the model with  $k$  regression coefficients. Thereafter, we could find the variance for the estimator as,

$$\hat{\sigma}_k^2 = \frac{SSR_k}{T}$$

These may then be used to calculate the information criteria where,

$$AIC = \log \hat{\sigma}_k^2 + \frac{T + 2k}{T}$$

$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log T}{T}$$

In this case the value of  $k$  that yields the minimum  $AIC$  is the one that relates to the best model. The construction of these statistics suggest that minimizing  $\hat{\sigma}_k^2$  would be a reasonable objective, except that it decreases monotonically as  $k$  increases. Therefore, we ought to penalize the error variance by a term proportional to the number of parameters. The choice for the penalty term given by these expressions are not the only forms of information criteria, but they are possible the most widely used.

It is worth noting that the  $BIC$  criteria would infer that the lag order should be more parsimonious than what is suggested by the  $AIC$ . In this way these two criteria could be used to describe the bounds of the lags that should be included in the model (moving from a more parsimonious to a less parsimonious setting).

Note that for both the  $AIC$  and  $BIC$  tests, smaller values for the calculated test statistic are preferred (or where the statistic is negative, you should select the model with the test statistics that is closest to  $-\infty$ ).

## 5.3 Diagnostic checking

In the last stage of this procedure, we need to perform a battery of diagnostic checks on the most appropriate model to ensure that it is suitable. In this case we would need to focus our attention on the residuals to ensure they represent white noise (possibly with the aid of ACF, PACF and Box-Ljung test).<sup>6</sup> If there is evidence of remaining serial correlation in the error term then the model would need to be re-estimated. In addition, it is usually a good idea to test the normality of the residuals, whilst other tests for either over-fitting or under-fitting could also be considered.

As noted previously, parsimonious models are usually preferred, so it is worth checking the standard errors of the respective coefficients to ensure that the specification does not include terms that are not able to make a significant contribution in the explanation of the left-hand side variable. If you find that the standard errors for a particular autoregressive or moving average coefficient suggest that the parameter is insignificant then it should be removed and the model should be re-estimated.

## 6 Structural Breaks

The problem of detecting structural changes in linear regression relationships has been an important topic in statistical and econometric research over time. Given recent economic events, this area has received increased attention. In what follows we review some of the most commonly used procedures that have been applied to economic data. Perron (2006) and Perron (2010) provide useful surveys of the literature.

### 6.1 Chow's Breakpoint Test

The Chow (1960) breakpoint test seeks to fit the same regression model to separate sub-samples of the

data, to see whether there are significant differences in the parameter estimates. Formally, this test investigates whether the null hypothesis of “no structural change” holds after constructing a  $F$ -test statistic for the parameters in the two models. A significant difference indicates a structural change in the relationship.

To gain the intuition behind this statistic, consider the simple linear regression model,

$$y_t = x_t^\top \beta_t + \varepsilon_t$$

where  $\varepsilon_t$  is a serially uncorrelated error term. In this case  $\beta_t$  is a time varying coefficient that may take on two possible values. To test whether or not the coefficient estimate changes at date  $\tau$  we could consider the values for  $\beta$ , where

$$\beta_t = \begin{cases} \beta & t \leq \tau \\ \beta + \delta & t > \tau \end{cases}$$

If the break date  $t$  is known, then the problem of testing the null hypothesis of no break (that is,  $\delta = 0$ ) against the alternative of a nonzero break ( $\delta \neq 0$ ) is equivalent to testing the hypothesis that the coefficient  $\delta$  is zero in the augmented version of

$$y_t = x_t^\top \beta_t + \delta Z_t(\tau) + \varepsilon_t$$

where  $Z_t(\tau) = X_t$  if  $t > \tau$  and  $Z_t(\tau) = 0$  otherwise.

This test can be computed using a conventional  $t$ -statistic when estimating the regression with ordinary least squares. The hypothesis of no break is rejected at the 5% significance level if the absolute value of this  $t$ -statistic is greater than 1.96. Or alternatively, we could test for a break in all the model parameters at this point in time with the aid of an  $F$ -test.

The major drawback of this procedure is that the change point must be known *a priori* as it is required to split the sample into two sub-samples. In addition, to perform this test you would need to ensure that each sub-sample has at least as many observations as the number of estimated parameters.

## 6.2 Quandt Likelihood Ratio Test

The Quandt (1960) likelihood ratio (QLR) test is a natural extension of the Chow test where a  $F$ -test statistic is calculated for all potential breakpoints within an interval  $[\underline{i}, \bar{i}]$ . One would then reject the null hypothesis of no structural change if the absolute value of any of the test statistics are too large. It has been shown that the QLR statistic has controlled size under the null of no break and good power against the alternative of a breakpoint.

Unfortunately the asymptotic properties of this statistic are non-standard so we would need to make use of the critical values of either Andrews (1993), Andrews and Ploberger (1994), Hansen (1997) or Stock and Watson (2010). It is usually performed as a  $\sup F$ -test although other variants exist. The associated  $p$ -values for the statistic would then provide an indication of the probability that a structural break may have occurred at a particular point in time.

## 6.3 Testing for multiple structural breaks

Bai and Perron (1998) and Bai and Perron (2003) extend this approach to use  $F$ -tests for 0 vs.  $\ell$  breaks and  $\ell$  vs.  $\ell + 1$  breaks respectively with arbitrary but fixed  $\ell$ . Hence, where it is assumed that there are  $m$  possible breakpoints we would like to consider the least squares estimates for the different  $\beta_j$  coefficients that would minimise the resulting residual sum of squares. To implement this procedure Bai and Perron (2003) calculate the residual sum of squares from a regression that includes a single breakpoint, for which the date of which is not observed. The value for this statistic is then compared to the residual sum of squares for the regression that does not include a breakpoint. Thereafter, a second breakpoint is included and the residual sum of squares is compared to that of a single breakpoint. Therefore, the problem for identifying all the dates for the structural changes is to find the  $m$  breakpoints that minimize the objective function:

$$\min_{1 \rightarrow m} RSS$$

over all  $m$  partitions. Information criteria are often used for model selection, which would identify the selection of  $m$  breakpoints in this case. Bai and Perron (2003) suggest that the AIC usually overestimates the number of breaks and as such the BIC is usually preferred.

## 6.4 CUSUM Test

The CUSUM test is based on the cumulative sum of the recursive residuals that utilises a generalized fluctuation test framework. To investigate for a structural break, one would plot the cumulative sum together with the 5% critical boundaries. The test finds parameter instability if the cumulative sum breaks either of the two boundaries. It was originally proposed by Brown, Durbin, and Evans (1975) and is particularly useful when the process is initially relatively stable, and then goes through a particularly turbulent period.

## 7 Conclusion

The use of relatively simple univariate autoregressive moving average regression models are usually appropriate for the modelling of most stationary time series. As the models may be estimated with relative ease they should be applied before considering a more complex model structure. In addition, the use of the Box and Jenkins (1979) technique may be applied to assist with the identification of the model structure that could explain the underlying data generating process.

Where the data may be subject to an obvious structural break, one would apply a breakpoint test such as the QLR statistic or a CUSUM test. It is relatively easy to implement either of these intuitive test statistics.

## 8 References

- Akaike, H. 1981. "Likelihood of a Model and Information Criteria." *Journal of Econometrics* 16: 3–14.
- Andrews, D. W. K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61(4): 821–56.
- Andrews, D. W. K., and W. Ploberger. 1994. "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative." *Econometrica* 62: 1383–1414.

- Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66 (1): 47–78.
- . 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18 (1): 1–22.
- Box, George, and Gwilym Jenkins. 1979. *Time Series Analysis: Forecasting and Control*. New York: Wiley.
- Box, George, Gwilym Jenkins, and Glen Reinsel. 2008. *Time Series Analysis: Forecasting and Control*. 4th ed. New York: John Wiley & Sons.
- Box, George, and D. A. Pierce. 1970. "Distribution of Residual Correlations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association* 65: 1509–26.
- Brown, R. L., J. Durbin, and J. M. Evans. 1975. "Techniques for Testing the Constancy of Regression Relationships over Time." *Journal of the Royal Statistical Society* 37: 149–63.
- Chow, G. C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28: 591–605.
- Enders, Walter. 2010. *Applied Econometric Time Series*. 3rd ed. New York: John Wiley & Sons.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, B.E. 1997. "Approximating Asymptotic  $p$  Values for Structural-Change Tests." *Journal of Business and Economic Statistics* 15 (1): 60–67.
- Harvey, Andrew C. 1993. *Time Series Models*. Cambridge, Mass: MIT Press.
- Ljung, G. M., and George Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65: 553–64.
- Perron, Pierre. 2006. "Palgrave Handbook of Econometrics, Volume 1." In, 278–352. Pgrave MacMillan.
- . 2010. "Macroeconometrics and Time Series Analysis." In, edited by S.N. Durlauf and L.E. Blume. The New Palgrave Economics Collection. London: Palgrave Macmillan.
- Quandt, R.E. 1960. "Tests of the Hypothesis That a Linear Regression Obeys Two Separate Regimes." *Journal of the American Statistical Association* 55: 324–30.
- Schwarz, H. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6: 461–64.
- Stock, James H., and Mark W. Watson. 2010. *Introduction to Econometrics*. 3rd ed. New York: Addison-Wesley.

- 
1. See, Hamilton (1994) for a proof.↩
  2. Those who would like to consider the use of the method of undetermined coefficients, which involves a guess and verify routine, should consult Hamilton (1994) or Enders

(2010).↵

3. See, Harvey (1993) for a proof.↵
4. In a large number of multivariate investigations one would remove the seasonal component, however, in cases where we are interested in the degree of seasonal variation in a univariate process, this procedure would be inappropriate.↵
5. See, Akaike (1981) and Schwarz (1978).↵
6. Note that when applying the Box-Ljung test to residuals, one would need to correct for the degrees of freedom.↵