Week 3: Autoregressive Moving Average(ARMA) Models

- Shumway, R.H. & Stoffer, D.S. (2016). Time series analysis and its applications with R examples. springer.
  - Chapter 3: ARIMA Models
- Brockwell, P.J., & Davis, R.A. (2009). Time series: theory and methods. Springer.
  - Chapter 3: Stationary ARMA Processes

Chapter 3

Autoregressive Moving Average(ARMA) Models

Classical regression is often insufficient for explaining all of the interesting dynamics of a time series. Instead, the introduction of correlation that may be generated through lagged linear relations leads to proposing the autoregressive (AR) and autoregressive moving average (ARMA) models.

3.1 What are ARMA Models?

The classical regression model was developed for the static case, namely, we only allow the dependent variable to be influenced by current values of the independent variables. In the time series case, it is desirable to allow the dependent variable to be influenced by the past values of the independent variables and possibly by its own past values. If the present can be plausibly modeled in terms of only the past values of the independent inputs, we have the enticing prospect that forecasting will be possible.

3.1.1 Introduction to Autoregressive Models

Autoregressive models are based on the idea that the current value of the series, x_t, can be explained as a function of p past values, $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$, where p determines the number of steps into the past needed to forecast the current value.

Definition 3.1 (AR(p)) *An autoregressive model of order p, abbreviated AR(p), is of the form*

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t, \tag{3.1.1}$$

where x_t is stationary, $w_t \sim wn(0, \sigma_w^2)$, and $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). The mean of x_t in (3.1.1) *is zero. If the mean, μ, of x_t is not zero, replace x_t by $x_t - \mu$:*

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + (\phi_2 x_{t-2} - \mu) + \ldots + (\phi_p x_{t-p} - \mu) + w_t,$$

or write

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t, \tag{3.1.2}$$

where $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$.

We note that (3.1.2) is similar to the regression model, and hence the term auto (or self) regression. Some technical difficulties, however, develop from applying that model because the regressors, $x_{t-1}, \cdots, x_{t-p}$, are random components, whereas the predictors z_t was assumed to be fixed. A useful form follows by using the backshift operator to write (3.1.1) as

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots, \phi_p B^p)x_t = w_t, \tag{3.1.3}$$

or even more concisely as

$$\phi(B)x_t = w_t, \tag{3.1.4}$$

The properties of $\phi(B)$ are important in solving (3.1.4) for x_t. This leads to the following definition.

Definition 3.2 (Autoregressive operator) *The autoregressive operator is defined to be*

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots, \phi_p B^p). \tag{3.1.5}$$

Example 3.1 (The AR(1) Model) *The AR model of order 1 is given by* $x_t = \phi x_{t-1} + w_t$. *Iterating backwards k times, we get*

$$\begin{aligned} x_t &= \phi x_{t-1} + w_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t \\ &= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\ &\vdots \\ &= \phi^k x_{t-k} + \sum_{j=0}^{k} \phi^j w_{t-j} \end{aligned}$$

This method suggests that, by continuing to iterate backward, and provided that $|\phi| < 1$ *and* $\sup_t Var(x_t) < \infty$, *we can represent an AR(1) model as a linear process given by*

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} \tag{3.1.6}$$

Representation (3.1) *is called the* **stationary solution of the model**. *In fact, by simple substitution,*

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} = \phi\left(\sum_{k=0}^{\infty} \phi^k w_{t-1-k}\right) + w_t = \phi x_{t-1} + w_t.$$

The AR(1) process defined by is stationary with mean

$$E(x_t) = \sum_{j=0}^{\infty} \phi^j E(w_{t-j}) = 0,$$

and autocovariance function,

$$\begin{aligned}\gamma(h) &= Cov(x_{t+h}, x_t) \\ &= \sigma_w^2 \sum_{j=0}^{\infty} \phi^{h+j}\phi^j \\ &= \phi^h \sigma_w^2 \sum_{j=0}^{\infty} \phi^{2j} \\ &= \frac{\phi^h \sigma_w^2}{1-\phi^2}, \qquad h \geq 0. \end{aligned} \tag{3.1.7}$$

Recall that $\gamma(h) = \gamma(-h)$, so we will only exhibit the autocovariance function for $h \geq 0$. From (3.1.7)*, the ACF of an AR(1) is*

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \qquad h > 0, \tag{3.1.8}$$

and $\rho(h)$ satisfies the recursion

$$\rho(h) = \phi\rho(h-1), \qquad h = 1, 2, \ldots. \tag{3.1.9}$$

We will discuss the ACF of a general AR(p) model later.

Example 3.2 (The Sample Path of an AR(1) Process) *Figure 3.1 shows a time plot of two AR(1) processes, one with ϕ = .9 and one with ϕ = -.9; in both cases, $\sigma_w^2 = 1$. In the first case, $\rho(h) = .9^h$, for $h \geq 0$, so observations close together in time are positively correlated with each other. This result means that observations at contiguous time points will tend to be close in value to each other. This fact shows up in the top of Figure 3.1 as a very smooth sample path for x_t. Now, contrast this with the case in which $\phi = -.9$, so that $\rho(h) = (-.9)^h$, for $h \geq 0$. This result means that observations at contiguous time points are negatively correlated but observations two time points apart are positively correlated. This fact shows up in the bottom of Figure 3.1. In this case, the sample path is very choppy.*

Figure 3.1 can be reproduced in R as follows.

```
par(mfrow=c(2,1))
# in the expressions below, ~ is a space and == is equal
tsplot(sarima.sim(ar= .9, n=100), col=4, ylab="", main=(expression(AR
    (1)~~~phi==+.9)))
tsplot(sarima.sim(ar=-.9, n=100), col=4, ylab="", main=(expression(AR
    (1)~~~phi==-.9)))
```

Listing 3.1: The code to reproduce Figure 3.1

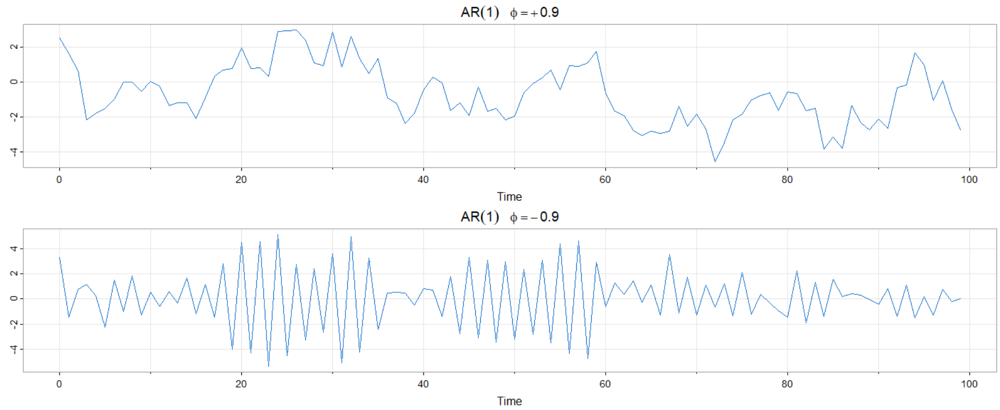

Figure 3.1: Simulated AR(1) models: ϕ = .9 (top); ϕ = -.9 (bottom)

The technique of iterating backward to get an idea of the stationary solution of AR models works well when $p = 1$, but not for larger orders. A general technique is that of **matching coefficients**. Consider the AR(1) model in operator form $\phi(B)x_t = w_t$, (3.1.4), where $\phi(B) = 1 - \phi B$ and $|\phi| < 1$. Also, write the model in equation (3.1) using operator form as

$$\begin{aligned} x_t &= \sum_{j=0}^{\infty} \phi^j w_{t-j} \\ &= \sum_{j=0}^{\infty} \phi^j B^j w_t \\ &= \psi(B) w_t \end{aligned} \tag{3.1.10}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\psi_j = \phi^j$.

Suppose we did not know that $\psi_j = \phi^j$. We could substitute $\psi(B)w_t$ from (3.1.10) for x_t to obtain

$$\phi(B)x_t = w_t \qquad \Rightarrow \qquad \phi(B)\psi(B)w_t = w_t \tag{3.1.11}$$

The coefficients of B on the left-hand side of (3.1.11) must be equal to those on right-hand side of (3.1.11), which means

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \ldots + \psi_j B^j + \ldots) = 1 \tag{3.1.12}$$

Reorganizing the coefficients in (3.1.12),

$$1 + (\psi_1 - \phi)B + (\psi_2 - \psi_1\phi)B^2 + \ldots + (\psi_j - \psi_{j-1}\phi)B^j + \ldots = 1.$$

We see that for each $j = 1, 2, \ldots$, the coefficient of B^j on the left must be zero because it is zero on the right. Therefore,

$$\begin{aligned} &\psi_1 - \phi = 0 && \Rightarrow && \psi_1 = \phi \\ &\psi_2 - \psi_1\phi = 0 && \Rightarrow && \psi_2 = \phi^2 \\ &\vdots \\ &\psi_j - \psi_{j-1}\phi = 0 && \Rightarrow && \psi_j = \psi_{j-1}\phi = \phi^j. \end{aligned}$$

Note that $\psi_0 = 1$.

Another way to think about the operations we just performed is to consider the AR(1) model in operator form, $\phi(B)x_t = w_t$. Now multiply both sides by $\phi^{-1}(B)$ (assuming the inverse operator exists) to get:

$$\phi^{-1}(B)\phi(B)x_t = \phi^{-1}(B)w_t \qquad \Rightarrow \qquad x_t = \phi^{-1}(B)w_t$$

We know already that

$$\phi^{-1}(B) = 1 + \phi B + \phi^2 B^2 + \ldots + \phi^j B^j + \ldots,$$

that is, $\phi^{-1}(B) = \psi(B)$. Thus, we notice that working with operators is like working with polynomials. That is, consider the polynomial $\phi(z) = 1 - \phi z$, where z is a complex number and $|\phi| < 1$. Then,

$$\phi^{-1}(z) = 1 + \phi z + \phi^2 z^2 + \ldots + \phi^j z^j + \ldots, \qquad |z| \leq 1.$$

and the coefficients of B^j in $\phi^{-1}(B)$ are the same as the coefficients of z_j in $\phi^{-1}(z)$. These results will be generalized in our discussion of ARMA models. We will find the polynomials corresponding to the operators useful in exploring the general properties of ARMA models.

Note 3.1

- *Consider the random walk $x_t = x_{t-1} + w_t$. As mentioned in chapter 1, this process is not stationary. In general. we might wonder if there is an stationary AR(1) process with $|\phi| \geq 1$. Such processes are called* **explosive** *because the values of the time series quickly become large in magnitude.*
- *An explosive process $x_t = \phi x_{t-1} + w_t$ with $|\phi| > 1$ can be rewritten as*

$$x_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}.$$

- *Unfortunately, this model is useless because it requires us to know the future to be able to predict the future. When a process does not depend on the future, such as the AR(1) when $|\phi| < 1$, we will say the process is causal. In the explosive case, the process is stationary, but it is also future dependent, and not causal.*

3.1.2 Introduction to Moving Average Models

As an alternative to the autoregressive representation in which the x_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as MA(q), assumes the white noise w_t on the right-hand side of the defining equation are combined linearly to form the observed data.

Definition 3.3 (MA(q)) *The moving average model of order q is defined to be*

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}, \tag{3.1.13}$$

where $w_t \sim wn(0, \sigma_w^2)$, and $\theta_1, \theta_2, \ldots, \theta_q$ ($\theta_q \neq 0$) are parameters.

The system is the same as the infinite moving average defined as the linear process (3.1.10), where $\psi_0 = 1$, $\psi_j = \theta_j$, for $j = 1, \ldots, q$, and $\psi_j = 0$ for other values. We may also write the MA(q) process in the equivalent form

$$x_t = \theta(B)w_t, \tag{3.1.14}$$

using the following definition.

Definition 3.4 (Moving average operator) *The moving average operator is*

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q, \tag{3.1.15}$$

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters $\theta_1, \ldots, \theta_q$.

Example 3.3 (The MA(1) Process) *Consider the MA(1) model $x_t = w_t + \theta w_{t-1}$. Then, $E(x_t) = 0$,*

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2 & h = 0, \\ \theta\sigma_w^2 & h = 0, \\ 0 & h > 1 \end{cases}$$

and the ACF is

$$\rho(h) = \begin{cases} 1 & h = 0, \\ \frac{\theta}{(1+\theta^2)} & h = 0, \\ 0 & h > 1 \end{cases}$$

Note $|\rho(1)| \leq 1/2$ for all values of θ. Also, x_t is correlated with x_{t-1}, but not with $x_{t-2}, x_{t-3}, \ldots$. Contrast this with the case of the AR(1) model in which the correlation between x_t and x_{t-k} is never zero. When $\theta = .9$, for example, x_t and x_{t-1} are positively correlated, and $\rho(1) = .497$. When $\theta = -.9$, x_t and x_{t-1} are negatively correlated, $\rho(1) = -.497$. Figure 3.2 shows a time plot of these two processes with σ_w^2= 1. The series for which θ = .9 is smoother than the series for which $\theta = .9$.

Figure 3.2 can be reproduced in R as follows.

```
par(mfrow=c(2,1))
tsplot(sarima.sim(ma= .9, n=100), col=4, ylab="", main=(expression(MA
    (1)~~~theta==+.9)))
tsplot(sarima.sim(ma=-.9, n=100), col=4, ylab="", main=(expression(MA
    (1)~~~theta==-.9)))
```

Listing 3.2: The code to reproduce Figure 3.1

Note 3.2

- *In general, there is no restriction on the value of θ in MA(1) process. But, it can be shown that without any restriction different MA(1) processes will end up with the same ACF. (Check the previous example with $\theta = 5$ and $\theta = 1/5$.)*

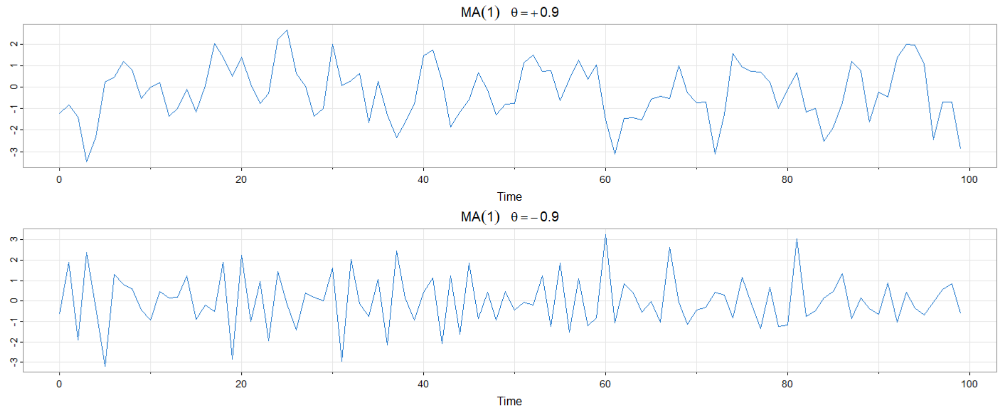

Figure 3.2: Simulated MA(1) models: θ = .9 (top); θ = -.9 (bottom)

- *To get a unique representation for the process, by mimicking the criterion of causality for AR models, we will choose the model with an infinite AR representation. Such a process is called an* **invertible** *process. In general, a time series is invertible if errors can be inverted into a representation of past observations.*

- *To discover which model is the invertible model, we can reverse the roles of x_t and w_t (because we are mimicking the AR case) and write the MA(1) model as $w_t = -\theta w_{t-1} + x_t$. It can be shown that, if $|\theta| < 1$, then $w_t = -\sum_{j=0}^{\infty}(-\theta)^j x_{t-j}$, which is the desired infinite AR representation of the model.*

As in the AR case, the polynomial, $\theta(z)$, corresponding to the moving average operators, $\theta(B)$, will be useful in exploring general properties of MA processes. For example, following the steps of equations (3.1.10)-(3.1.12), we can write the MA(1) model as $x_t = \theta(B)w_t$, where $\theta(B) = 1 + \theta B$. If $|\theta| < 1$, then we can write the model as $\pi(B)x_t = w_t$, where $\pi(B) = \theta^{-1}(B)$. Let $\theta(z) = 1 + \theta z$, for $|z| \leq 1$, then

$$\pi(z) = \theta^{-1}(z) = \frac{1}{1+\theta z} = \sum_{j=0}^{\infty}(-\theta)^j z^j,$$

3.1.3 Autoregressive Moving Average Models

We now proceed with the general development of autoregressive, moving average, and mixed autoregressive moving average (ARMA), models for stationary time series.

Definition 3.5 *[ARMA(p,q)] A time series $\{x_t;\ t = 0, \pm1, \pm2, \ldots\}$ is ARMA(p,q) if it is stationary and*

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \tag{3.1.16}$$

with $\phi_p \neq 0$ $\theta_q \neq 0$ and $\sigma_w^2 > 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively. If x_t has a nonzero mean μ, we set $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$ and write the model as

$$x_t = \alpha + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}, \tag{3.1.17}$$

where $w_t \sim wn(0, \sigma_w^2)$

As previously noted, when $q = 0$, the model is called an autoregressive model of order p, AR(p), and when $p = 0$, the model is called a moving average model of order q, MA(q). To aid in the investigation of ARMA models, it will be useful to write them using the AR and the MA operators. In particular, the ARMA(p, q) model in (3.1.16) can then be written in concise form as

$$\phi(B)x_t = \theta(B)w_t. \tag{3.1.18}$$

There are some problems with the general definition of ARAM(p, q) process, such as,

(i) parameter redundant models (refer to Example 3.7 in Shumway and Stoffer),

(ii) stationary AR models that depend on the future, and

(iii) MA models that are not unique.

To overcome these problems, we will require some additional restrictions on the model parameters. First, we make the following definitions.

Definition 3.6 *The AR and MA polynomials are defined as*

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p, \qquad \phi_p \neq 0 \tag{3.1.19}$$

and

$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z_q, \qquad \theta_q \neq 0, \tag{3.1.20}$$

respectively, where z *is a complex number.*

To address the first problem, we will henceforth refer to an ARMA(p, q) model to mean that it is in its simplest form. That is, in addition to the original definition given in equation (3.1.16), we will also require that $\phi(z)$ and $\theta(z)$ have no common factors.

To address the problem of future-dependent models, we formally introduce the concept of **causality**.

Definition 3.7 (Causality) *An ARMA(p, q) model is said to be causal, if the time series* $\{x_t;\ t = 0, \pm1, \pm2, \ldots\}$ *can be written as a one-sided linear process:*

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t \tag{3.1.21}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ *and* $\sum_{j=0}^{\infty} |\psi_j| > \infty.$*; we set* $\psi_0 = 1$.

Property 3.1 *[Causality of an ARMA(p, q) Process] An ARMA(p, q) model is causal if and only if* $\phi(z) \neq 0$ *for* $|z| \leq 1$*. The coefficients of the linear process given in 3.1.21 can be determined by solving*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Another way to phrase Property 3.1 is that an ARMA process is causal only when the roots of $\phi(z)$ lie outside the unit circle; that is, $\phi(z) = 0$ only when $|z| > 1$. Finally, to address the problem of uniqueness problem, we choose the model that allows an infinite autoregressive representation.

Definition 3.8 (Invertibility) *An ARMA(p, q) model is said to be invertible, if the time series* $\{x_t; t = 0, \pm 1, \pm 2, \ldots\}$ *can be written as*

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_i, \tag{3.1.22}$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ *and* $\sum_{j=0}^{\infty} |\pi_j| < \infty$*; we set* $\pi_0 = 1$.

Analogous to Property 3.1, we have the following property.

Property 3.2 *[Invertibility of an ARMA(p, q) Process] An ARMA(p, q) model is invertible if and only if* $\theta(z) \neq 0$ *for* $|z| \leq 1$. *The coefficients* π_j *of* $\pi(B)$ *given in* (3.1.22) *can be determined by solving*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

Another way to phrase Property 3.2 is that an ARMA process is invertible only when the roots of $\theta(z)$ lie outside the unit circle; that is, $\theta(z) = 0$ only then $|z| > 1$.

Example 3.4 (Parameter Redundancy, Causality, Invertibility) *Consider the process*

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

or, in operator form,

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t.$$

At first, x_t *appears to be an ARMA(2, 2) process. But notice that*

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

and

$$\theta(B) = (1 + B + .25B2) = (1 + .5B)^2$$

have a common factor that can be canceled. After cancellation, the operators are $\phi(B) = (1 - .9B)$ *and* $\theta(B) = (1 + .5B)$*, so the model is an ARMA(1, 1) model,* $(1 - .9B)x_t = (1 + .5B)w_t$*, or*

$$x_t = .9x_{t-1} + .5w_{t-1} + wt. \tag{3.1.23}$$

The model is causal because $\phi(z) = (1 - .9z) = 0$ *when* $z = 10/9$*, which is outside the unit circle. The model is also invertible because the root of* $\theta(z) = (1 + .5z)$ *is* $z = -2$, *which is outside the unit circle. To write the model as a linear process, we can obtain the ψ-weights using Property 3.1, $\phi(z)\psi(z) = \theta(z)$, or*

$$(1 - .9z)(1 + \psi_1 z + \psi_2 z^2 + \ldots + \psi_j z^j + \ldots) = 1 + .5z.$$

Rearranging, we get

$$1 + (\psi_1 - .9)z + (\psi_2 - .9\psi_1)z^2 + \ldots + (\psi_j - .9\psi_{j-1})z_j + \ldots = 1 + .5z.$$

Matching the coefficients of z on the left and right sides we get $\psi_1 - .9 = .5$ and $\psi_j - .9\psi_{j-1} = 0$ for $j > 1$. Thus, $\psi_j = 1.4(.9)^{j-1}$ for $j \geq 1$ and (3.1.23) can be written as

$$x_t = w_t + 1.4 \sum_{j=1}^{\infty} .9^{j-1} w_{t-j}.$$

The invertible representation using Property 3.2 is obtained by matching coefficients in $\theta(z)\pi(z) = \phi(z)$,

$$(1 + .5z)(1 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \ldots) = 1 - .9z.$$

In this case, the π-weights are given by $\pi_j = (-1)^j 1.4(.5)^{j-1}$, for $j \geq 1$, and hence, because $w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$, we can also write (3.1.23) as

$$x_t = 1.4 \sum_{j=1}^{\infty} (-.5)^{j-1} x_{t-j} + w_j$$

```
ARMAtoMA(ar = .9, ma = .5, 10)   # first 10 psi-weights
# [1] 1.4000000 1.2600000 1.1340000 1.0206000 0.9185400 0.8266860
    0.7440174 0.6696157
# [9] 0.6026541 0.5423887

ARMAtoAR(ar = .9, ma = .5, 10)   # first 10 pi-weights
# [1] -1.400000000  0.700000000 -0.350000000  0.175000000 -0.087500000
     0.043750000
# [7] -0.021875000  0.010937500 -0.005468750  0.002734375
```

Listing 3.3: The code to calculte ψ_j in (3.1.21)

3.2 Autocorrelation and Partial Autocorrelation

We begin by exhibiting the ACF of an MA(q) process, $x_t = \theta(B)w_t$, where $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$. Because x_t is a finite linear combination of white noise terms, the process is stationary with mean

$$E(x_t) = \sum_{j=0}^{q} \theta_j E(w_{t-j}) = 0,$$

where we have written $\theta_0 = 1$, and with autocovariance function

$$\begin{aligned} \gamma(h) = Cov(x_{t+h}, x_t) &= Cov\left(\sum_{j=0}^{q} \theta_j w_{t+h-j}, \sum_{k=0}^{q} \theta_k w_{t-k}\right) \\ &= \begin{cases} \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, & 0 \leq h \leq q \\ 0 & h > q. \end{cases} \end{aligned} \tag{3.2.1}$$

Recall that $\gamma(h) = \gamma(-h)$, so we will only display the values for $h \geq 0$. Note that $\gamma(q)$ cannot be zero because $\theta_q \neq 0$. The cutting off of $\gamma(h)$ after q lags is the signature of the MA(q) model. Dividing (3.2.1) by $\gamma(0)$ yields the ACF of an MA(q):

$$\rho(h) = = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1+\theta_1^2+\ldots+\theta_q^2}, & 0 \leq h \leq q \\ 0 & h > q. \end{cases} \tag{3.2.2}$$

For a causal ARMA(p, q) model, $\phi(B)x_t = \theta(B)w_t$, where the zeros of $\phi(z)$ are outside the unit circle, write

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}. \tag{3.2.3}$$

It follows immediately that $E(x_t) = 0$ and the autocovariance function of x_t is

$$\gamma(h) = Cov(x_{t+h}, x_t) = \sigma_w^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \quad h \geq 0.(3.45) \tag{3.2.4}$$

Besides, it is also possible to obtain a homogeneous difference equation directly in terms of $\gamma(h)$. First, we write

$$\begin{aligned} \gamma(h) = Cov(x_{t+h}, x_t) &= Cov\left(\sum_{j=1}^{p} \phi_j x_{t+h-j} + \sum_{j=0}^{q} \theta_j w_{t+h-j}, x_t\right) \\ &= \sum_{j=1}^{p} \phi_j \gamma(h-j) + \sigma_w^2 \sum_{j=0}^{q} \theta_j \psi_{j-h}, \qquad h \geq 0, \end{aligned} \tag{3.2.5}$$

where we have used the fact that, for $h \geq 0$,

$$Cov(w_{t+h-j}, x_t) = Cov\left(w_{t+h-j}, \sum_{k=0}^{\infty} \psi_k w_{t-k}\right) = \psi_{j-h}\sigma_w^2.$$

From (3.2.5), we can write a general homogeneous equation for the ACF of a causal ARMA process:

$$\gamma(h) - \phi_1\gamma(h-1) - \ldots - \phi_p\gamma(h-p) = 0, \qquad h \geq \max(p, q+1), \tag{3.2.6}$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = \sigma_w^2 \sum_{j=0}^{q} \theta_j \psi_{j-h}, \qquad 0 \leq h \leq \max(p, q+1), \tag{3.2.7}$$

Dividing (3.2.6) and (3.2.7) through by $\gamma(0)$ will allow us to solve for the ACF, $\rho(h) = \gamma(h)/\gamma(0)$.

Example 3.5 (The ACF of an AR(p)) *Consider AR(p) process. To obtain its ACF, it follows immediately from* (3.2.6) *that*

$$\rho(h) - \phi_1\rho(h-1) - \ldots - \phi_p\rho(h-p) = 0, \qquad h \geq p. \tag{3.2.8}$$

Let $z_1, \ldots, z_r$ denote the roots of $\phi(z)$, each with multiplicity $m_1, \ldots, m_r$, respectively, where $m_1 + \ldots + m_r = p$. Then, it can be shown that the general solution is

$$\rho(h) = z_1^{-h} P_1(h) + z_2^{-h} P_2(h) + \ldots + z_r^{-h} P_r(h), \qquad h \geq p, \tag{3.2.9}$$

where $P_j(h)$ is a polynomial in h of degree $m_j - 1$.

Recall that for a causal model, all of the roots are outside the unit circle, $|z_i| > 1$, for $i = 1, \ldots, r$. If all the roots are real, then $\rho(h)$ dampens exponentially fast to zero as $h \to \infty$. If some of the roots are complex, then they will be in conjugate pairs and $\rho(h)$ will dampen, in a sinusoidal fashion, exponentially fast to zero as $h \to \infty$. In the case of complex roots, the time series will appear to be cyclic in nature. This, of course, is also true for ARMA models in which the AR part has complex roots.

Example 3.6 (The ACF of an ARMA(1, 1)) *Consider the ARMA(1, 1) process $x_t = \phi x_{t-1} + \theta w_{t-1} + w_t$, where $|\phi| < 1$. Based on* (3.2.6)*, the autocovariance function satisfies*

$$\gamma(h) - \phi\gamma(h-1) = 0, \qquad h = 2, 3, \ldots,$$

and it can be shown that the general solution is

$$\gamma(h) = c\phi^h, \qquad h = 1, 2, \ldots \tag{3.2.10}$$

To obtain the initial conditions, we use (3.2.7)*:*

$$\gamma(0) = \phi\gamma(1) + \sigma_w^2[1 + \theta\phi + \theta^2]$$

and

$$\gamma(1) = \phi\gamma(0) + \sigma_w^2\theta.$$

Solving for $\gamma(0)$ and $\gamma(1)$, we obtain:

$$\gamma(0) = \sigma_w^2 \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2}$$

and

$$\gamma(1) = \sigma_w^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}.$$

To solve for c, note that from (3.2.10)*, $\gamma(1) = c\phi$ or $c = \gamma(1)/\phi$. Hence, the specific solution for $h \geq 1$ is*

$$\gamma(h) = \frac{\gamma(1)}{\phi}\phi^h = \sigma_w^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}\phi^{h-1}$$

Finally, dividing through by $\gamma(0)$ yields the ACF

$$\rho(h) = \frac{(1 + \theta\phi)(\phi + \theta)}{1 - 2\phi\theta + \theta^2}\phi^{h-1}, \qquad h \geq 1 \tag{3.2.11}$$

Notice that the general pattern of $\rho(h)$ versus h in (3.2.11) is not different from that of an AR(1) given in (3.1.8). Hence, it is unlikely that we will be able to tell the difference between an ARMA(1,1) and an AR(1) based solely on an ACF estimated from a sample. This consideration will lead us to the partial autocorrelation function.

3.2.1 The Partial Autocorrelation Function (PACF)

We have seen in (3.2.2), for MA(q) models, the ACF will be zero for lags greater than q. Moreover, because $\theta_q \neq 0$, the ACF will not be zero at lag q. Thus, the ACF provides a considerable amount of information about the order of the dependence when the process is a moving average process. If the process, however, is ARMA or AR, the ACF alone tells us little about the orders of dependence. Hence, it is worthwhile pursuing a function that will behave like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF). Recall that if X, Y, and Z are random variables, then the partial correlation between X and Y given Z is obtained by regressing X on Z to obtain $\hat{X}$, regressing Y on Z to obtain $\hat{Y}$, and then calculating

$$\rho_{XY|Z} = Corr\{X - \hat{X}, Y - \hat{Y}\}.$$

The idea is that $\rho_{XY|Z}$ measures the correlation between X and Y with the linear effect of Z removed (or partialled out). If the variables are multivariate normal, then this definition coincides with $\rho_{XY|Z} = Corr(X, Y|Z)$.

To motivate the idea for time series, consider a causal AR(1) model, $x_t = \phi x_{t-1} + w_t$. Then,

$$\begin{aligned}\gamma_x(2) = Cov(x_t, x_{t-2}) &= Cov(\phi x_{t-1} + w_t, x_{t-2})\\ &= Cov(\phi^2 x_{t-2} + \phi w_{t-1} + w_t, x_{t-2}) = \phi^2\gamma(0).\end{aligned}$$

This result follows from causality because x_{t-2} involves $\{w_{t-2}, w_{t-3}, \ldots\}$, which are all uncorrelated with w_t and w_{t-1}. The correlation between x_t and x_{t-2} is not zero, as it would be for an MA(1), because x_t is dependent on x_{t-2} through x_{t-1}. Suppose we break this chain of dependence by removing (or partial out) the effect x_{t-1}. That is, we consider the correlation between $x_t - \phi x_{t-1}$ and $x_{t-2} - \phi x_{t-1}$, because it is the correlation between x_t and x_{t-2} with the linear dependence of each on x_{t-1} removed.

In this way, we have broken the dependence chain between x_t and x_{t-2}. In fact,

$$Cov(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = Cov(w_t, x_{t-2} - \phi x_{t-1}) = 0.$$

Hence, the tool we need is partial autocorrelation, which is the correlation between x_s and x_t with the linear effect of everything "in the middle" removed.

To formally define the PACF for mean-zero stationary time series, let x_{t+h}, for $h \geq 2$, denote the regression of x_{t+h} on $\{x_{t+h-1}, x_{t+h-2}, \ldots, x_{t+1}\}$, which we write as

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \ldots + \beta_{h-1} x_{t+1}. \tag{3.2.12}$$

No intercept term is needed in (3.2.12) because the mean of x_t is zero (otherwise, replace x_t by $x_t - \mu_x$ in this discussion). In addition, let $\hat{x}_t$ denote the regression of x_t on $\{x_{t+1}, x_{t+2}, \ldots, x_{t+h-1}\}$, then

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \ldots + \beta_{h-1} x_{t+h-1}. \tag{3.2.13}$$

Because of stationarity, the coefficients, $\beta_1, \ldots, \beta_{h-1}$ are the same in (3.2.12) and (3.2.13); we will explain this result in the next section, but it will be evident from the examples.

Definition 3.9 (The partial autocorrelation function (PACF)) *The partial autocorrelation function of a stationary process, x_t, denoted by ϕ_{hh}, for $h = 1, 2, \ldots,$ is*

$$\phi_{11} = Corr(x_{t+1}, x_t) = \rho(1, \tag{3.2.14}$$

and

$$\phi_{hh} = Corr(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), \qquad h \geq 2. \tag{3.2.15}$$

Note 3.3
- *The PACF, ϕ_{hh}, is the correlation between x_{t+h} and x_t with the linear dependence of $\{x_{t+1}, \ldots, x_{t+h-1}\}$ on each, removed.*
- *If the process x_t is Gaussian, then $\phi_{hh} = Corr(x_{t+h}, x_t | x_{t+1}, \ldots, x_{t+h-1})$; that is, ϕ_{hh} is the correlation coefficient between x_{t+h} and x_t in the bivariate distribution of (x_{t+h}, x_t) conditional on $\{x_{t+1}, \ldots, x_{t+h-1}\}$.*

Example 3.7 (The PACF of an AR(1)) *Consider the PACF of the AR(1) process given by $x_t = \phi x_{t-1} + w_t$, with $|\phi| < 1$. By definition, $\phi_{11} = \rho(1) = \phi$. To calculate ϕ_{22}, consider the regression of x_{t+2} on x_{t+1}, say, $\hat{x}_{t+2} = \beta x_{t+1}$. We choose β to minimize*

$$E(x_{t+2} - \hat{x}_{t+2})^2 = E(x_{t+2} - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0).$$

Taking derivatives with respect to β and setting the result equal to zero, we have $\beta = \gamma(1)/\gamma(0) = \rho(1) = \phi$. Next, consider the regression of x_t on $x_{t+1,}$ say $\hat{x}_t = \beta x_{t+1}$. We choose β to minimize

$$E(x_t - \hat{x}_t)^2 = E(x_t - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0).$$

This is the same equation as before, so $\beta = \phi$. Hence,

$$\begin{aligned}\phi_{22} &= Corr(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = Corr(x_{t+2} - \phi x_{t+1}, x_t - \phi x_{t+1}) \\ &= Corr(w_{t+2}, x_t - \phi x_{t+1}) = 0\end{aligned}$$

by causality. Thus, $\phi_{22} = 0$. In the next example, we will see that in this case, $\phi_{hh} = 0$ for all $h > 1$.

Example 3.8 (The PACF of an AR(p)) *The model implies $x_{t+h} = \sum_{j=1}^{p} \phi_j x_{t+h-j} + w_{t+h}$, where the roots of $\phi(z)$ are outside the unit circle. When $h > p$, the regression of x_{t+h} on $\{x_{t+1}, \ldots, x_{t+h-1}\}$, is*

$$\hat{x}_{t+h} = \sum_{j=1}^{p} \phi_j x_{t+h-j}$$

We have not proved this obvious result yet, but we will prove it later. Thus, when $h > p$,

$$\phi_{hh} = Corr(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) = Corr(w_{t+h}, x_t - \hat{x}_t) = 0,$$

because, by causality, $x_t - \hat{x}_t$ depends only on $\{w_{t+h-1}, w_{t+h-2}, \ldots\}$; recall equation (3.2.13). Whenh $\leq p$, ϕ_{pp} is not zero, and $\phi_{11}, \ldots, \phi_{p-1,p-1}$ are not necessarily zero. We will see later that, in fact, $\phi_{pp} = \phi_p$. Figure 3.3 shows the ACF and the PACF of the AR(2) with $\phi_1 = 1.5$ and $\phi_2 = -.75$. To reproduce Figure 3.3 in R, use the following commands:

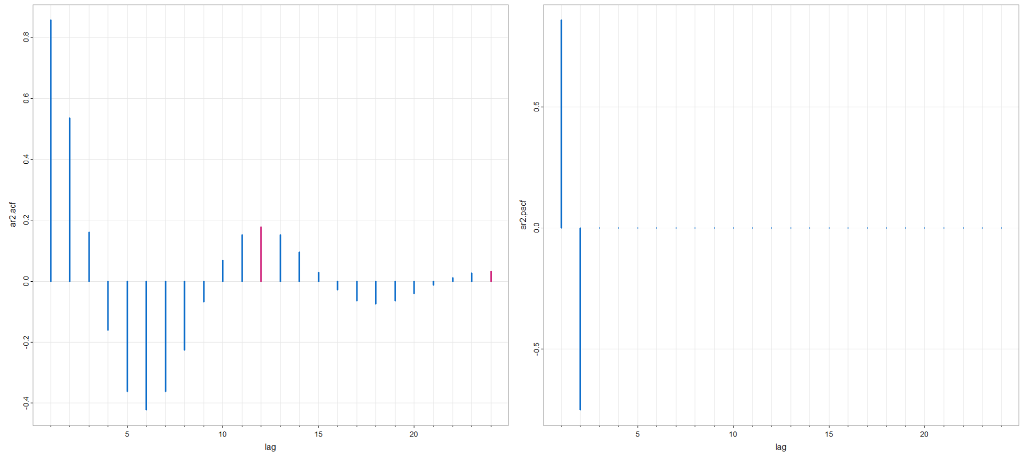

Figure 3.3: The ACF and PACF of an AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -.75$

```
ar2.acf  = ARMAacf(ar=c(1.5,-.75), ma=0, 24)[-1]
ar2.pacf = ARMAacf(ar=c(1.5,-.75), ma=0, 24, pacf=TRUE)
par(mfrow=1:2)
tsplot(ar2.acf,  type="h", xlab="lag", lwd=3, nxm=5, col=c(rep(4,11),
   6))
tsplot(ar2.pacf, type="h", xlab="lag", lwd=3, nxm=5, col=4)
```

Listing 3.4: The code to reproduce Figure 3.3

Example 3.9 (The PACF of an Invertible MA(q)) *For an invertible MA(q), we can write $x_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} + w_t$ .Moreover, no finite representation exists. From this result, it should be apparent that the PACF will never cut off, as in the case of an AR(p). For an MA(1), $x_t = w_t + \theta w_{t-1}$, with $|\theta| < 1$, it can be shown that $\phi_{22} = -\theta^2/(1 + \theta^2 + \theta^4)$. For the MA(1) in general, we can show that*

$$\phi_{hh} = -\frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2(h+1)}}, \qquad h \geq 1.$$

In the next section, we will discuss methods of calculating the PACF. The PACF for MA models behaves much like the ACF for AR models. Also, the PACF for AR models behaves much like the ACF for MA models. Because an invertible ARMA model has an infinite AR representation, the PACF will not cut off. We may summarize these results in Table 3.1.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Table 3.1: Behavior of the ACF and PACF for ARMA models

Example 3.10 (Preliminary Analysis of the Recruitment Series) *We consider the problem of modeling the Recruitment series, discussed in previous chapters. There are 453* *months of observed recruitment ranging over the years 1950-1987. The ACF and the PACF given in Figure 3.4 are consistent with the behavior of an AR(2). The ACF has cycles corresponding roughly to a 12-month period, and the PACF has large values for* $h = 1, 2$ *and then is essentially zero for higher order lags. Based on Table 3.1, these results suggest that a second-order* ($p = 2$) *autoregressive model might provide a good fit.*

We ran a regression using the data triplets $\{(x; z_1, z_2) : (x_3; x_2, x_1), (x_4; x_3, x_2), \ldots, (x_{453}; x_{452}, x_{451})\}$ *to fit a model of the form* $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$ *for* $t = 3, 4, \ldots, 453$. *The estimates and standard errors (in parentheses) are*$\hat{\phi}_0 = 6.74(1.11)$, $\hat{\phi}_1 = 1.35(.04)$, $\hat{\phi}_2 = -.46(.04)$, *and* $\sigma_w^2 = 89.72$. *The following R code can be used for this analysis.We use acf2 from astsa to print and plot the ACF and PACF.*

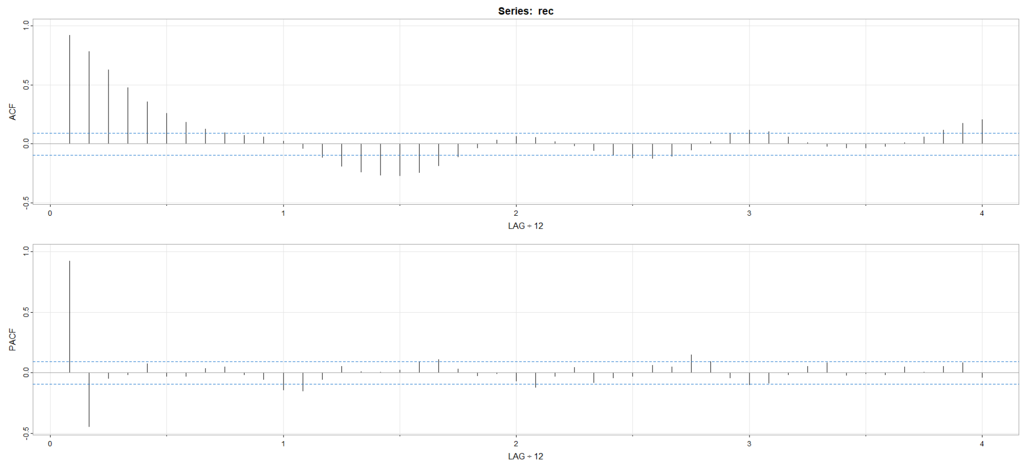

Figure 3.4: ACF and PACF of the Recruitment series. Note that the lag axes are in terms of season (12 months in this case)

```
acf2(rec, 48)      # will produce values and a graphic
(regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE))   # regression
regr$asy.se.coef  # standard errors
```

Listing 3.5: The code to reproduce Figure 3.4 and estimate the values