Chapter 3

ARMA Processes

3.1 Introduction

In this chapter *autoregressive moving average* processes are discussed. They play a crucial role in specifying time series models for applications. As the solutions of *stochastic difference equations* with constant coefficients and these processes possess a linear structure.

Definition 3.1.1 (ARMA processes)**.** *(a) A weakly stationary process* $(X_t\colon t \in \mathbb{Z})$ *is called an autoregressive moving average time series of order* (p, q)*, abbreviated by ARMA*(p, q)*, if it satisfies the difference equations*

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \qquad t \in \mathbb{Z}, \tag{3.1.1}$$

where $\phi_1, \ldots, \phi_p$ *and* $\theta_1, \ldots, \theta_q$ *are real constants,* $\phi_p \neq 0 \neq \theta_q$*, and* $(Z_t\colon t \in \mathbb{Z}) \sim \mathrm{WN}(0, \sigma^2)$*.*

(b) A weakly stationary stochastic process $(X_t\colon t \in \mathbb{Z})$ *is called an ARMA*(p, q) *time series with mean* μ *if the process* $(X_t - \mu\colon t \in \mathbb{Z})$ *satisfies the equation system* (3.1.1)*.*

A more concise representation of (3.1.1) can be obtained with the use of the backshift operator B. To this end, define the *autoregressive polynomial* and the *moving average polynomial* by

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p, \qquad z \in \mathbb{C},$$

and

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q, \qquad z \in \mathbb{C},$$

respectively, where $\mathbb{C}$ denotes the set of complex numbers. Inserting the backshift operator into these polynomials, the equations in (3.1.1) become

$$\phi(B)X_t = \theta(B)Z_t, \qquad t \in \mathbb{Z}. \tag{3.1.2}$$

Example 3.1.1. Figure 3.1 displays realizations of three different autoregressive moving average time series based on independent, standard normally distributed $(Z_t\colon t \in \mathbb{Z})$. The left panel is an ARMA(2,2) process with parameter specifications $\phi_1 = .2$, $\phi_2 = -.3$, $\theta_1 = -.5$ and $\theta_2 = .3$. The middle plot is

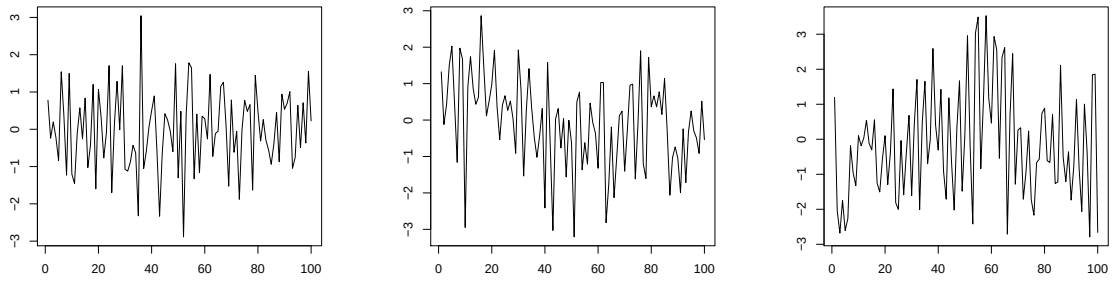

Figure 3.1: Realizations of three autoregressive moving average processes.

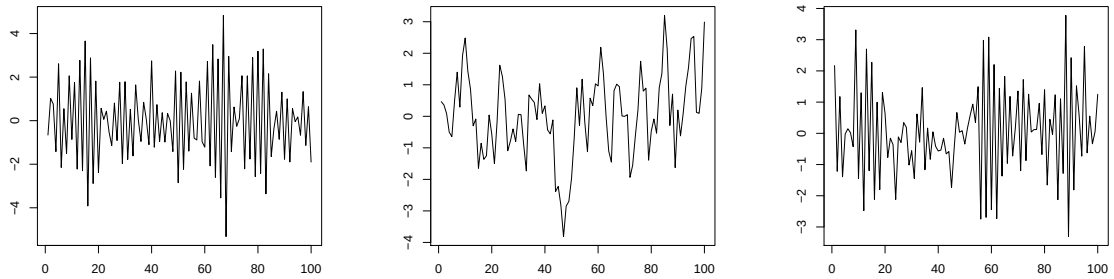

Figure 3.2: Realizations of three autoregressive processes.

obtained from an ARMA(1,4) process with parameters $\phi_1 = .3$, $\theta_1 = -.2$, $\theta_2 = -.3$, $\theta_3 = .5$, and $\theta_4 = .2$, while the right plot is from an ARMA(4,1) with parameters $\phi_1 = -.2$, $\phi_2 = -.3$, $\phi_3 = .5$ and $\phi_4 = .2$ and $\theta_1 = .6$. The plots indicate that ARMA models can provide a flexible tool for modeling diverse residual sequences. It will turn out in the next section that all three realizations here come from (strictly) stationary processes. Similar time series plots can be produced in R using the commands

```
> arima22 =
   arima.sim(list(order=c(2,0,2), ar=c(.2,-.3), ma=c(-.5,.3)), n=100)
> arima14 =
   arima.sim(list(order=c(1,0,4), ar=.3, ma=c(-.2,-.3,.5,.2)), n=100)
> arima41 =
   arima.sim(list(order=c(4,0,1), ar=c(-.2,-.3,.5,.2), ma=.6), n=100)
```

Some special cases covered in the following two examples have particular relevance in time series analysis.

Example 3.1.2 (AR processes)**.** If the moving average polynomial in (3.1.2) is equal to one, that is, if $\theta(z) \equiv 1$, then the resulting $(X_t \colon t \in \mathbb{Z})$ is referred to as *autoregressive process of order p, AR(p)*. These time series interpret the value of the current variable X_t as a linear combination of p previous variables $X_{t-1}, \ldots, X_{t-p}$ plus an additional distortion by the white noise Z_t. Figure 3.2 displays two AR(1) processes with respective parameters $\phi_1 = -.9$ (left) and $\phi_1 = .8$ (middle) as well as an AR(2) process with parameters $\phi_1 = -.5$ and $\phi_2 = .3$. The corresponding R commands are

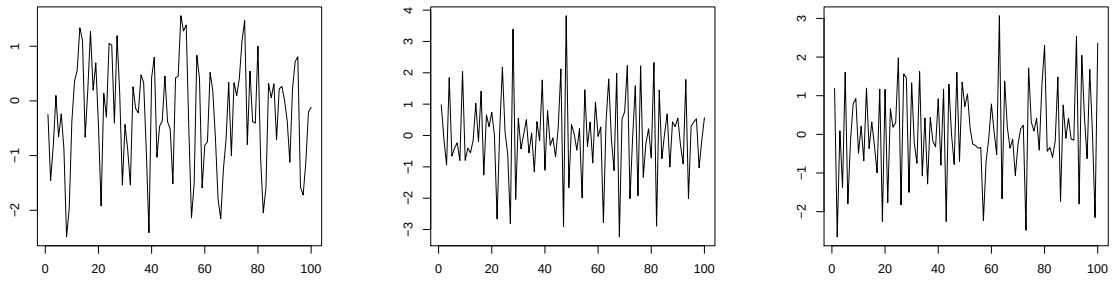

Figure 3.3: Realizations of three moving average processes.

```
> ar1neg = arima.sim(list(order=c(1,0,0), ar=-.9), n=100)
> ar1pos = arima.sim(list(order=c(1,0,0), ar=.8), n=100)
> ar2 = arima.sim(list(order=c(2,0,0), ar=c(-.5,.3)), n=100)
```

Example 3.1.3 (MA processes)**.** If the autoregressive polynomial in (3.1.2) is equal to one, that is, if $\phi(z) \equiv 1$, then the resulting $(X_t \colon t \in \mathbb{Z})$ is referred to as *moving average process of order q, MA(q)*. Here the present variable X_t is obtained as superposition of q white noise terms $Z_t, \ldots, Z_{t-q}$. Figure 3.3 shows two MA(1) processes with respective parameters $\theta_1 = .5$ (left) and $\theta_1 = -.8$ (middle). The right plot is observed from an MA(2) process with parameters $\theta_1 = -.5$ and $\theta_2 = .3$. In R one may use

```
> ma1pos = arima.sim(list(order=c(0,0,1), ma=.5), n=100)
> ma1neg = arima.sim(list(order=c(0,0,1), ma=-.8), n=100)
> ma2 = arima.sim(list(order=c(0,0,2), ma=c(-.5,.3)), n=100)
```

For the analysis upcoming in the next chapters, we now introduce moving average processes of infinite order ($q = \infty$). They are an important tool for determining stationary solutions to the difference equations (3.1.1).

Definition 3.1.2 (Linear processes)**.** *A stochastic process $(X_t \colon t \in \mathbb{Z})$ is called linear process or MA(∞) time series if there is a sequence $(\psi_j \colon j \in \mathbb{N}_0)$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z}, \tag{3.1.3}$$

where $(Z_t \colon t \in \mathbb{Z}) \sim \mathrm{WN}(0, \sigma^2)$.

Moving average time series of any order q are special cases of linear processes. Just pick $\psi_j = \theta_j$ for $j = 1, \ldots, q$ and set $\psi_j = 0$ if $j > q$. It is common to introduce the power series

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \qquad z \in \mathbb{C},$$

to express a linear process in terms of the backshift operator. Display (3.1.3) can now be rewritten in the compact form

$$X_t = \psi(B) Z_t, \qquad t \in \mathbb{Z}.$$

With the definitions of this section at hand, properties of ARMA processes, such as stationarity and invertibility, are investigated in the next section. The current section is closed giving meaning to the notation $X_t = \psi(B)Z_t$. Note that one is possibly dealing with an *infinite* sum of random variables. For completeness and later use, in the following example the mean and ACVF of a linear process are derived.

Example 3.1.4 (Mean and ACVF of a linear process)**.** Let $(X_t : t \in \mathbb{Z})$ be a linear process according to Definition 3.1.2. Then, it holds that

$$E[X_t] = E\left[\sum_{j=0}^{\infty} \psi_j Z_{t-j}\right] = \sum_{j=0}^{\infty} \psi_j E[Z_{t-j}] = 0, \qquad t \in \mathbb{Z}.$$

Next observe also that

$$\begin{aligned}\gamma(h) &= \mathrm{Cov}(X_{t+h}, X_t) \\ &= E\left[\sum_{j=0}^{\infty} \psi_j Z_{t+h-j} \sum_{k=0}^{\infty} \psi_k Z_{t-k}\right] \\ &= \sigma^2 \sum_{k=0}^{\infty} \psi_{k+h}\psi_k < \infty\end{aligned}$$

by assumption on the sequence $(\psi_j : j \in \mathbb{N}_0)$.

3.2 Causality and Invertibility

While a moving average process of order q will always be stationary without conditions on the coefficients $\theta_1, \ldots, \theta_q$, some deeper thoughts are required in the case of AR(p) and ARMA(p, q) processes. For simplicity, we start by investigating the autoregressive process of order one, which is given by the equations $X_t = \phi X_{t-1} + Z_t$ (writing $\phi = \phi_1$). Repeated iterations yield that

$$X_t = \phi X_{t-1} + Z_t = \phi^2 X_{t-2} + Z_t + \phi Z_{t-1} = \ldots = \phi^N X_{t-N} + \sum_{j=0}^{N-1} \phi^j Z_{t-j}.$$

Letting $N \to \infty$, it could now be shown that, with probability one,

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

is the weakly stationary solution to the AR(1) equations, provided that $|\phi| < 1$. These calculations would indicate moreover, that an autoregressive process of order one can be represented as linear process with coefficients $\psi_j = \phi^j$.

Example 3.2.1 (Mean and ACVF of an AR(1) process)**.** Since an autoregressive process of order one has been identified as an example of a linear process, one can easily determine its expected value as

$$E[X_t] = \sum_{j=0}^{\infty} \phi^j E[Z_{t-j}] = 0, \qquad t \in \mathbb{Z}.$$

For the ACVF, it is obtained that

$$\begin{aligned}\gamma(h) &= \operatorname{Cov}(X_{t+h}, X_t) \\ &= E\left[\sum_{j=0}^{\infty} \phi^j Z_{t+h-j} \sum_{k=0}^{\infty} \phi^k Z_{t-k}\right] \\ &= \sigma^2 \sum_{k=0}^{\infty} \phi^{k+h}\phi^k = \sigma^2\phi^h \sum_{k=0}^{\infty} \phi^{2k} = \frac{\sigma^2\phi^h}{1-\phi^2},\end{aligned}$$

where $h \geq 0$. This determines the ACVF for all h using that $\gamma(-h) = \gamma(h)$. It is also immediate that the ACF satisfies $\rho(h) = \phi^h$. See also Example 3.1.1 for comparison.

Example 3.2.2 (Nonstationary AR(1) processes)**.** In Example 1.2.3 we have introduced the random walk as a nonstationary time series. It can also be viewed as a nonstationary AR(1) process with parameter $\phi = 1$. In general, autoregressive processes of order one with coefficients $|\phi| > 1$ are called *explosive* for they do not admit a weakly stationary solution that could be expressed as a linear process. However, one may proceed as follows. Rewrite the defining equations of an AR(1) process as

$$X_t = -\phi^{-1} Z_{t+1} + \phi^{-1} X_{t+1}, \qquad t \in \mathbb{Z}.$$

Apply now the same iterations as before to arrive at

$$X_t = \phi^{-N} X_{t+N} - \sum_{j=1}^{N} \phi^{-j} Z_{t+j}, \qquad t \in \mathbb{Z}.$$

Note that in the weakly stationary case, the present observation has been described in terms of past innovations. The representation in the last equation however contains only future observations with time lags larger than the present time t. From a statistical point of view this does not make much sense, even though by identical arguments as above we may obtain

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}, \qquad t \in \mathbb{Z},$$

as the weakly stationary solution in the explosive case.

The result of the previous example leads to the notion of causality which means that the process $(X_t \colon t \in \mathbb{Z})$ has a representation in terms of the white noise $(Z_s \colon s \leq t)$ and that is hence uncorrelated with the future as given by $(Z_s \colon s > t)$. We give the definition for the general ARMA case.

Definition 3.2.1 (Causality)**.** *An ARMA(p, q) process given by (3.1.1) is causal if there is a sequence $(\psi_j \colon j \in \mathbb{N}_0)$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z}.$$

Causality means that an ARMA time series can be represented as a linear process. It was seen earlier in this section how an AR(1) process whose coefficient satisfies the condition $|\phi| < 1$ can be converted into a linear process. It was also shown that this is impossible if $|\phi| > 1$. The conditions on the autoregressive parameter ϕ can be restated in terms of the corresponding autoregressive polynomial $\phi(z) = 1 - \phi z$ as follows. It holds that

$$|\phi| < 1 \quad \text{if and only if} \quad \phi(z) \neq 0 \ \text{ for all } |z| \leq 1,$$

$$|\phi| > 1 \quad \text{if and only if} \quad \phi(z) \neq 0 \ \text{ for all } |z| \geq 1.$$

It turns out that the characterization in terms of the zeroes of the autoregressive polynomials carries over from the AR(1) case to the general ARMA(p, q) case. Moreover, the ψ-weights of the resulting linear process have an easy representation in terms of the polynomials $\phi(z)$ and $\theta(z)$. The result is summarized in the next theorem.

Theorem 3.2.1. *Let $(X_t \colon t \in \mathbb{Z})$ be an ARMA(p, q) process such that the polynomials $\phi(z)$ and $\theta(z)$ have no common zeroes. Then $(X_t \colon t \in \mathbb{Z})$ is causal if and only if $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. The coefficients $(\psi_j \colon j \in \mathbb{N}_0)$ are determined by the power series expansion*

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

A concept closely related to causality is *invertibility*. This notion is motivated with the following example that studies properties of a moving average time series of order 1.

Example 3.2.3. Let $(X_t \colon t \in \mathbb{N})$ be an MA(1) process with parameter $\theta = \theta_1$. It is an easy exercise to compute the ACVF and the ACF as

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2, & h = 0, \\ \theta\sigma^2, & h = 1 \\ 0 & h > 1, \end{cases} \qquad \rho(h) = \begin{cases} 1 & h = 0. \\ \theta(1+\theta^2)^{-1}, & h = 1. \\ 0 & h > 1. \end{cases}$$

These results lead to the conclusion that $\rho(h)$ does not change if the parameter θ is replaced with θ^{-1}. Moreover, there exist pairs (θ, σ^2) that lead to the same ACVF, for example $(5, 1)$ and $(1/5, 25)$. Consequently, we arrive at the fact that the two MA(1) models

$$X_t = Z_t + \frac{1}{5} Z_{t-1}, \qquad t \in \mathbb{Z}, \qquad (Z_t \colon t \in \mathbb{Z}) \sim \text{iid } \mathcal{N}(0, 25),$$

and

$$X_t = \tilde{Z}_t + 5\tilde{Z}_{t-1}, \qquad t \in \mathbb{Z}, \qquad (\tilde{Z} \colon t \in \mathbb{Z}) \sim \text{iid } \mathcal{N}(0, 1),$$

are indistinguishable because we only observe X_t but not the noise variables Z_t and $\tilde{Z}_t$.

For convenience, the statistician will pick the model which satisfies the invertibility criterion which is to be defined next. It specifies that the noise sequence can be represented as a linear process in the observations.

Definition 3.2.2 (Invertibility)**.** *An ARMA(p, q) process given by (3.1.1) is invertible if there is a sequence* $(\pi_j : j \in \mathbb{N}_0)$ *such that* $\sum_{j=0}^{\infty} |\pi_j| < \infty$ *and*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z}.$$

Theorem 3.2.2. *Let* $(X_t : t \in \mathbb{Z})$ *be an ARMA(p, q) process such that the polynomials* $\phi(z)$ *and* $\theta(z)$ *have no common zeroes. Then* $(X_t : t \in \mathbb{Z})$ *is invertible if and only if* $\theta(z) \neq 0$ *for all* $z \in \mathbb{C}$ *with* $|z| \leq 1$*. The coefficients* $(\pi_j)_{j \in \mathbb{N}_0}$ *are determined by the power series expansion*

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

From now on it is assumed that all ARMA sequences specified in the sequel are causal and invertible unless explicitly stated otherwise. The final example of this section highlights the usefulness of the established theory. It deals with parameter redundancy and the calculation of the causality and invertibility sequences $(\psi_j : j \in \mathbb{N}_0)$ and $(\pi_j : j \in \mathbb{N}_0)$.

Example 3.2.4 (Parameter redundancy)**.** Consider the ARMA equations

$$X_t = .4X_{t-1} + .21X_{t-2} + Z_t + .6Z_{t-1} + .09Z_{t-2},$$

which seem to generate an ARMA(2,2) sequence. However, the autoregressive and moving average polynomials have a common zero:

$$\tilde{\phi}(z) = 1 - .4z - .21z^2 = (1 - .7z)(1 + .3z),$$

$$\tilde{\theta}(z) = 1 + .6z + .09z^2 = (1 + .3z)^2.$$

Therefore, one can reset the ARMA equations to a sequence of order (1,1) and obtain

$$X_t = .7X_{t-1} + Z_t + .3Z_{t-1}.$$

Now, the corresponding polynomials have no common roots. Note that the roots of $\phi(z) = 1 - .7z$ and $\theta(z) = 1 + .3z$ are $10/7 > 1$ and $-10/3 < -1$, respectively. Thus Theorems 3.2.1 and 3.2.2 imply that causal and invertible solutions exist. In the following, the corresponding coefficients in the expansions

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \qquad \text{and} \qquad Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad t \in \mathbb{Z},$$

are calculated. Starting with the causality sequence $(\psi_j : j \in \mathbb{N}_0)$. Writing, for $|z| \leq 1$,

$$\sum_{j=0}^{\infty} \psi_j z^j = \psi(z) = \frac{\theta(z)}{\phi(z)} = \frac{1 + .3z}{1 - .7z} = (1 + .3z) \sum_{j=0}^{\infty} (.7z)^j,$$

it can be obtained from a comparison of coefficients that

$$\psi_0 = 1 \qquad \text{and} \qquad \psi_j = (.7 + .3)(.7)^{j-1} = (.7)^{j-1}, \qquad j \in \mathbb{N}.$$

Similarly one computes the invertibility coefficients $(\pi_j \colon j \in \mathbb{N}_0)$ from the equation

$$\sum_{j=0}^{\infty} \pi_j z^j = \pi(z) = \frac{\phi(z)}{\theta(z)} = \frac{1 - .7z}{1 + .3z} = (1 - .7z) \sum_{j=0}^{\infty} (-.3z)^j$$

($|z| \leq 1$) as

$$\pi_0 = 1 \qquad \text{and} \qquad \pi_j = (-1)^j (.3 + .7)(.3)^{j-1} = (-1)^j (.3)^{j-1}.$$

Together, the previous calculations yield to the explicit representations

$$X_t = Z_t + \sum_{j=1}^{\infty} (.7)^{j-1} Z_{t-j} \qquad \text{and} \qquad Z_t = X_t + \sum_{j=1}^{\infty} (-1)^j (.3)^{j-1} X_{t-j}.$$

In the remainder of this section, a general way is provided to determine the weights $(\psi_j \colon j \geq 1)$ for a causal ARMA(p, q) process given by $\phi(B)X_t = \theta(B)Z_t$, where $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Since $\psi(z) = \theta(z)/\phi(z)$ for these z, the weight ψ_j can be computed by matching the corresponding coefficients in the equation $\psi(z)\phi(z) = \theta(z)$, that is,

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots)(1 - \phi_1 z - \ldots - \phi_p z^p) = 1 + \theta_1 z + \ldots + \theta_q z^q.$$

Recursively solving for $\psi_0, \psi_1, \psi_2, \ldots$ gives

$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 - \phi_1 \psi_0 &= \theta_1, \\ \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 &= \theta_2, \end{aligned}$$

and so on as long as $j < \max\{p, q+1\}$. The general solution can be stated as

$$\psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = \theta_j, \qquad 0 \leq j < \max\{p, q+1\}, \tag{3.2.1}$$

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0, \qquad j \geq \max\{p, q+1\}, \tag{3.2.2}$$

if we define $\phi_j = 0$ if $j > p$ and $\theta_j = 0$ if $j > q$. To obtain the coefficients ψ_j one therefore has to solve the homogeneous linear difference equation (3.2.2) subject to the initial conditions specified by (3.2.1). For more on this subject, see Section 3.6 of Brockwell and Davis (1991) and Section 3.3 of Shumway and Stoffer (2006).

In R, these computations can be performed using the command `ARMAtoMA`. For example, one can use the commands

```
> ARMAtoMA(ar=.7,ma=.3,25)
> plot(ARMAtoMA(ar=.7,ma=.3,25))
```

which will produce the output displayed in Figure 3.4. The plot shows nicely the exponential decay of the ψ-weights which is typical for ARMA processes. The table shows row-wise the weights $\psi_0, \ldots, \psi_{24}$. This is enabled by the choice of `25` in the argument of the function `ARMAtoMA`.

1.0000000000	0.7000000000	0.4900000000	0.3430000000	0.2401000000
0.1680700000	0.1176490000	0.0823543000	0.0576480100	0.0403536070
0.0282475249	0.0197732674	0.0138412872	0.0096889010	0.0067822307
0.0047475615	0.0033232931	0.0023263051	0.0016284136	0.0011398895
0.0007979227	0.0005585459	0.0003909821	0.0002736875	0.0001915812

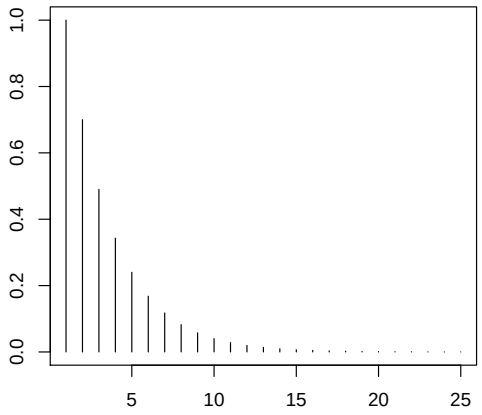

Figure 3.4: The R output for the ARMA(1,1) process of Example 3.2.4.

3.3 The PACF of a causal ARMA Process

In this section, the *partial autocorrelation function (PACF)* is introduced to further assess the dependence structure of stationary processes in general and causal ARMA processes in particular. To start with, let us compute the ACVF of a moving average process of order q.

Example 3.3.1 (The ACVF of an MA(q) process)**.** Let $(X_t\colon t \in \mathbb{Z})$ be an MA(q) process specified by the polynomial $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$. Then, letting $\theta_0 = 1$, it holds that

$$E[X_t] = \sum_{j=0}^{q} \theta_j E[Z_{t-j}] = 0.$$

To compute the ACVF, suppose that $h \geq 0$ and write

$$\begin{aligned}
\gamma(h) &= \text{Cov}(X_{t+h}, X_t) = E[X_{t+h}X_t] \\
&= E\left[\left(\sum_{j=0}^{q} \theta_j Z_{t+h-j}\right)\left(\sum_{k=0}^{q} \theta_k Z_{t-k}\right)\right] \\
&= \sum_{j=0}^{q}\sum_{k=0}^{q} \theta_j\theta_k E[Z_{t+h-j}Z_{t-k}] \\
&= \begin{cases} \sigma^2 \displaystyle\sum_{k=0}^{q-h} \theta_{k+h}\theta_k, & 0 \leq h \leq q. \\ 0, & h > q. \end{cases}
\end{aligned}$$

The result here is a generalization of the MA(1) case, which was treated in Example 3.2.3. It is also a special case of the linear process in Example 3.1.4. The structure of the ACVF for MA processes indicates a possible strategy to determine in practice the unknown order q: plot the the sample ACF and select as order q the largest lag such that $\rho(h)$ is significantly different from zero.

While the sample ACF can potentially reveal the true order of an MA process, the same is not true anymore in the case of AR processes. Even for the AR(1) time series it has been shown in Example 3.2.1 that its ACF $\rho(h) = \phi^{|h|}$ is nonzero for all lags. As further motivation, however, we discuss the following example.

Example 3.3.2. Let $(X_t\colon t \in \mathbb{Z})$ be a causal AR(1) process with parameter $|\phi| < 1$. It holds that

$$\gamma(2) = \text{Cov}(X_2, X_0) = \text{Cov}(\phi^2 X_0 + \phi Z_1 + Z_2, X_0) = \phi^2\gamma(0) \neq 0.$$

To break the linear dependence between X_0 and X_2, subtract ϕX_1 from both variables. Calcualting the resulting covariance yields

$$\text{Cov}(X_2 - \phi X_1, X_0 - \phi X_1) = \text{Cov}(Z_2, X_0 - \phi X_1) = 0,$$

since, due to the causality of this AR(1) process, $X_0 - \phi X_1$ is a function of $Z_1, Z_0, Z_{-1}, \ldots$ and therefore uncorrelated with $X_2 - \phi X_1 = Z_2$.

The previous example motivates the following general definition.

Definition 3.3.1 (Partial autocorrelation function)**.** *Let $(X_t\colon t \in \mathbb{Z})$ be a weakly stationary stochastic process with zero mean. Then, the sequence $(\phi_{hh}\colon h \in \mathbb{N})$ given by*

$$\begin{aligned}\phi_{11} &= \rho(1) = \text{Corr}(X_1, X_0),\\ \phi_{hh} &= \text{Corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}), \qquad h \geq 2,\end{aligned}$$

is called the partial autocorrelation function (PACF) of $(X_t\colon t \in \mathbb{Z})$. Therein,

$$\begin{aligned}X_h^{h-1} &= \textit{regression of } X_h \textit{ on } (X_{h-1}, \ldots, X_1)\\ &= \beta_1 X_{h-1} + \beta_2 X_{h-2} + \ldots + \beta_{h-1} X_1\\ X_0^{h-1} &= \textit{regression of } X_0 \textit{ on } (X_1, \ldots, X_{h-1})\\ &= \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{h-1} X_{h-1}.\end{aligned}$$

Notice that there is no intercept coefficient β_0 in the regression parameters, since it is assumed that $E[X_t] = 0$. The following example demonstrates how to calculate the regression parameters in the case of an AR(1) process.

Example 3.3.3 (PACF of an AR(1) process)**.** If $(X_t\colon t \in \mathbb{Z})$ is a causal AR(1) process, then $\phi_{11} = \rho(1) = \phi$. To calculate ϕ_{22}, calculate first $X_2^1 = \beta X_1$, that is β. This coefficient is determined by minimizing the mean-squared error between X_2 and βX_1:

$$E[X_2 - \beta X_1]^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0)$$

which is minimized by $\beta = \rho(1) = \phi$. (This follows easily by taking the derivative and setting it to zero.) Therefore $X_2^1 = \phi X_1$. Similarly, one computes $X_0^1 = \phi X_1$ and it follows from Example 3.3.2 that $\phi_{22} = 0$. Indeed all lags $h \geq 2$ of the PACF are zero.

More generally, consider briefly a causal AR(p) process given by $\phi(B)X_t = Z_t$ with $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$. Then, for $h > p$,

$$X_h^{h-1} = \sum_{j=1}^{p} \phi_j X_{h-j}$$

and consequently

$$\phi_{hh} = \text{Corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}) = \text{Corr}(Z_h, X_0 - X_0^{h-1}) = 0$$

if $h > p$ by causality (the same argument used in Example 3.3.2 applies here as well). Observe, however, that ϕ_{hh} is not necessarily zero if $h \leq p$. The forgoing suggests that the sample version of the PACF can be utilized to identify the order of an autoregressive process from data: use as p the largest lag h such that ϕ_{hh} is significantly different from zero.

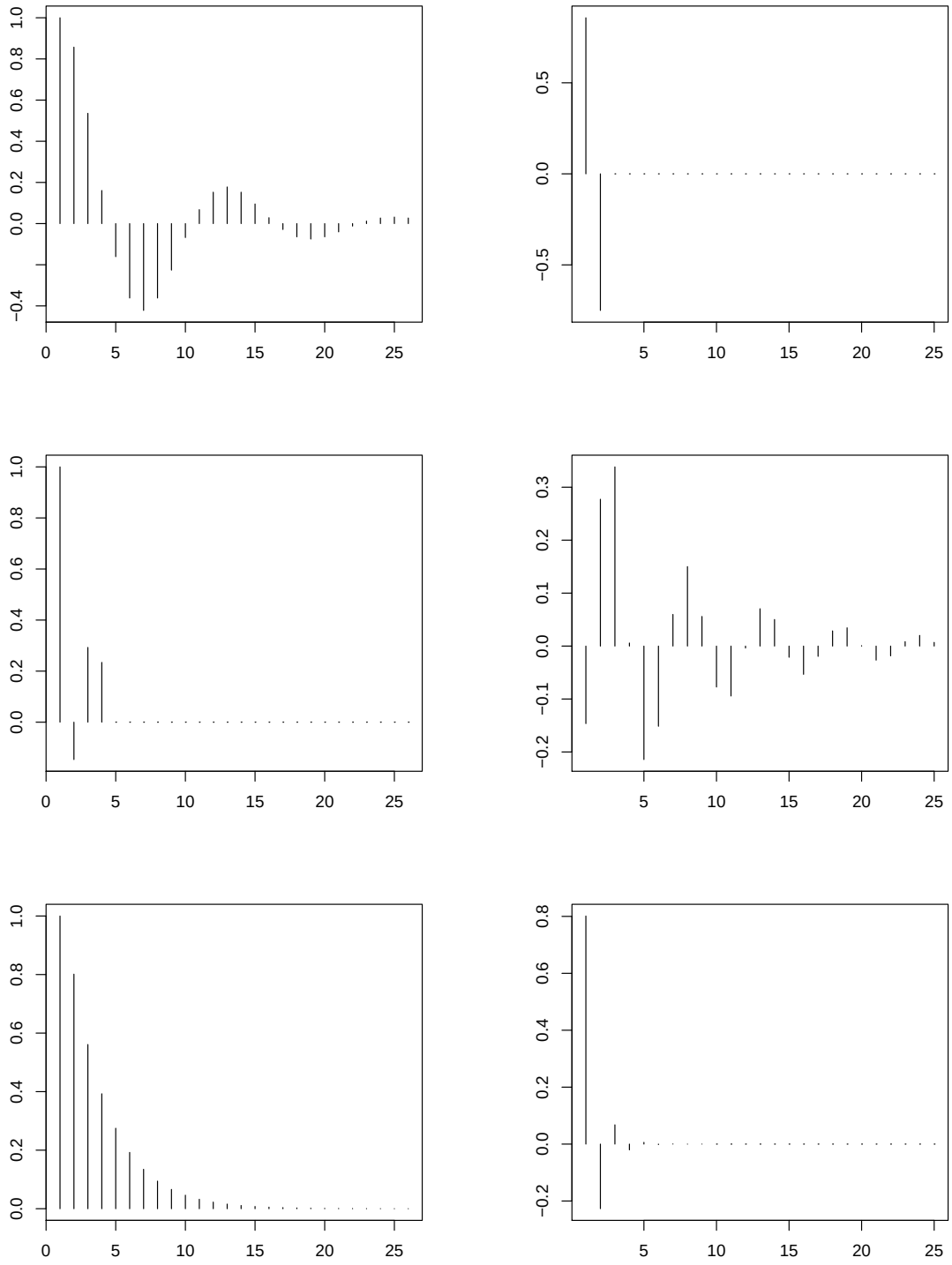

Figure 3.5: The ACFs and PACFs of an AR(2) process (upper panel), and MA(3) process (middle panel) and and ARMA(1,1) process (lower panel).

On the other hand, for an invertible MA(q) process, one can write $Z_t = \pi(B)X_t$ or, equivalently,

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

which shows that the PACF of an MA(q) process will be nonzero for all lags, since for a "perfect" regression one would have to use all past variables $(X_s\colon s < t)$ instead of only the quantity X_t^{t-1} given in Definition 3.3.1.

In summary, the PACF reverses the behavior of the ACVF for autoregressive and moving average processes. While the latter have an ACVF that vanishes after lag q and a PACF that is nonzero (though decaying) for all lags, AR processes have an ACVF that is nonzero (though decaying) for all lags but a PACF that vanishes after lag p.

ACVF (ACF) and PACF hence provide useful tools in assessing the dependence of given ARMA processes. If the estimated ACVF (the estimated PACF) is essentially zero after some time lag, then the underlying time series can be conveniently modeled with an MA (AR) process—and no general ARMA sequence has to be fitted. These conclusions are summarized in Table 3.1.

Example 3.3.4. Figure 3.5 collects the ACFs and PACFs of three ARMA processes. The upper panel is taken from the AR(2) process with parameters $\phi_1 = 1.5$ and $\phi_2 = -.75$. It can be seen that the ACF tails off and displays cyclical behavior (note that the corresponding autoregressive polynomial has complex roots). The PACF, however, cuts off after lag 2. Thus, inspecting ACF and PACF, we would correctly specify the order of the AR process.

The middle panel shows the ACF and PACF of the MA(3) process given by the parameters $\theta_1 = 1.5$, $\theta_2 = -.75$ and $\theta_3 = 3$. The plots confirm that $q = 3$ because the ACF cuts off after lag 3 and the PACF tails off.

Finally, the lower panel displays the ACF and PACF of the ARMA(1,1) process of Example 3.2.4. Here, the assessment is much harder. While the ACF tails off as predicted (see Table 3.1), the PACF basically cuts off after lag 4 or 5. This could lead to the wrong conclusion that the underlying process is actually an AR process of order 4 or 5. (The reason for this behavior lies in the fact that the dependence in this particular ARMA(1,1) process can be well approximated by that of an AR(4) or AR(5) time series.)

To reproduce the graphs in R, you can use the commands

```
> ar2.acf = ARMAacf(ar=c(1.5,-.75), ma=0, 25)
> ar2.pacf = ARMAacf(ar=c(1.5,-.75), ma=0, 25, pacf=T)
```

for the AR(2) process. The other two cases follow from straightforward adaptations of this code.

Example 3.3.5 (Recruitment Series)**.** The data considered in this example consists of 453 months of observed recruitment (number of new fish) in a certain part of the Pacific Ocean collected over the years 1950–1987. The corresponding time series plot is given in the left panel of Figure 3.6. The corresponding ACF and PACF displayed in the middle and right panel of the same figure recommend fitting an AR process of order $p = 2$ to the recruitment data. Assuming that the data is in `rec`, the R code to reproduce Figure 3.6 is

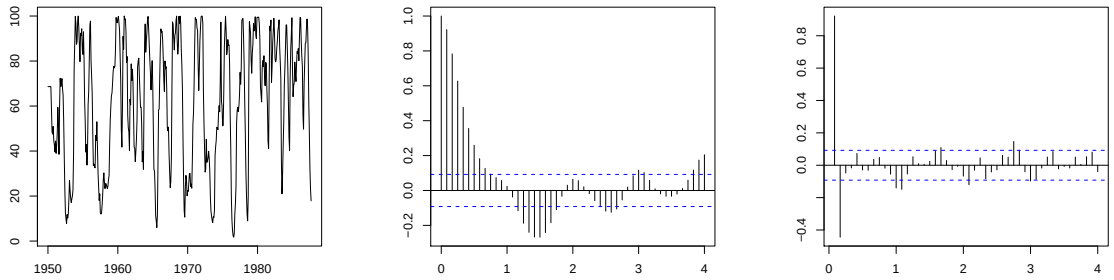

Figure 3.6: The recruitment series of Example 3.3.5 (left), its sample ACF (middle) and sample PACF (right).

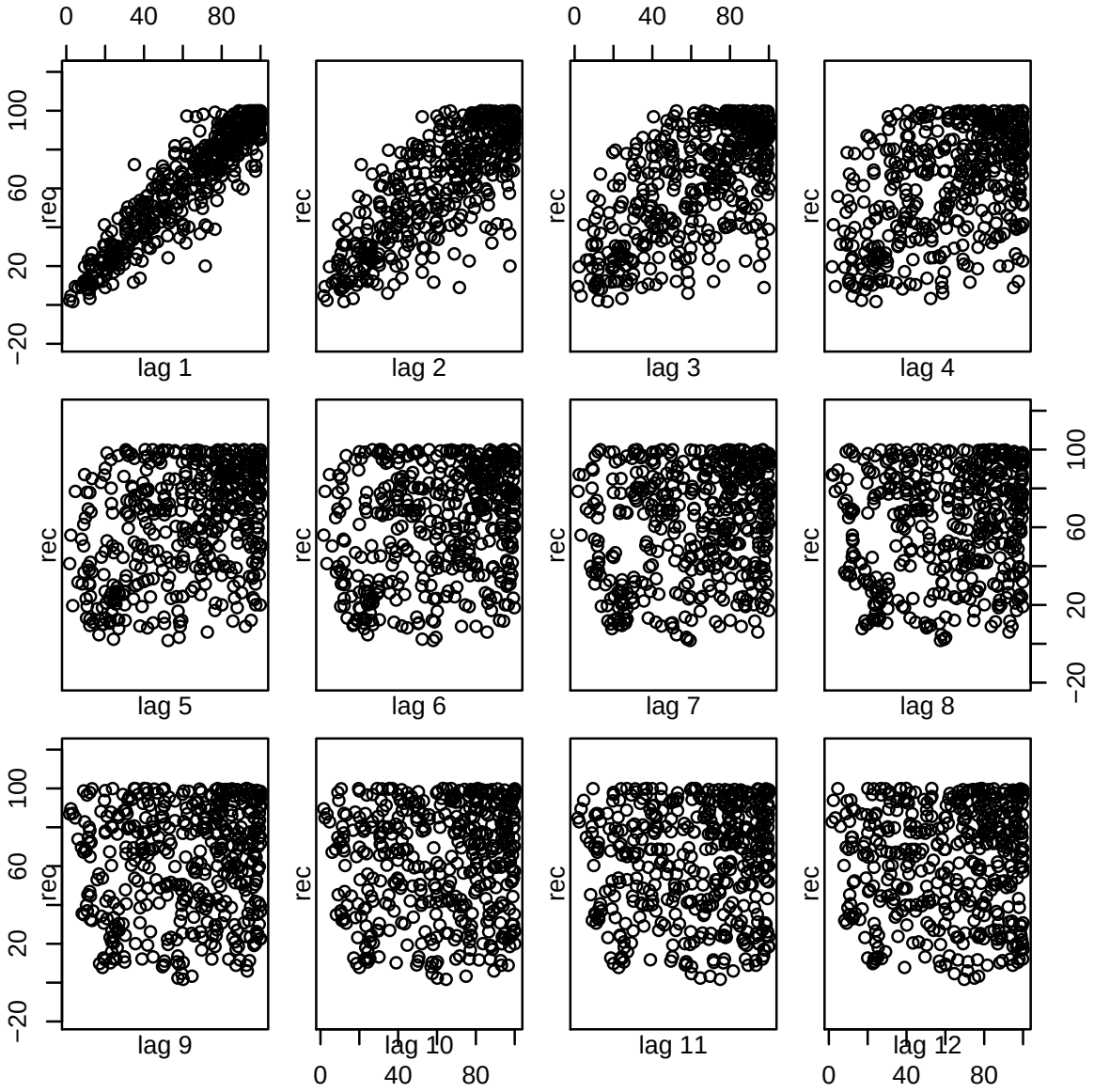

Figure 3.7: Scatterplot matrix relating current recruitment to past recruitment for the lags $h = 1, \ldots, 12$.

	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: The behavior of ACF and PACF for AR, MA, and ARMA processes.

```
> rec = ts(rec, start=1950, frequency=12)
> plot(rec, xlab="", ylab="")
> acf(rec, lag=48)
> pacf(rec, lag=48)
```

This assertion is also consistent with the scatterplots that relate current recruitment to past recruitment at several time lags, namely $h = 1, \ldots, 12$. For lag 1 and 2, there seems to be a strong linear relationship, while this is not the case anymore for $h \geq 3$. The corresponding R commands are

```
> lag.plot(rec, lags=12, layout=c(3,4), diag=F)
```

Denote by X_t the recruitment at time t. To estimate the AR(2) parameters, run a regression on the observed data triplets included in the set $\{(x_t, x_{t-1}, x_{t-2})\colon j = 3, \ldots, 453\}$ to fit a model of the form

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \qquad t = 3, \ldots, 453,$$

where $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$. This task can be performed in R as follows.

```
> fit.rec = ar.ols(rec, aic=F, order.max=2, demean=F, intercept=T)
```

These estimates can be assessed with the command `fit.rec` and the corresponding standard errors with `fit.rec$asy.se`. Here the parameter estimates $\hat{\phi}_0 = 6.737(1.111)$, $\hat{\phi}_1 = 1.3541(.042)$, $\hat{\phi}_2 = -.4632(.0412)$ and $\hat{\sigma}^2 = 89.72$ are obtained. The standard errors are given in brackets.

3.4 Forecasting

Suppose that the variables $X_1, \ldots, X_n$ of a weakly stationary time series $(X_t\colon t \in \mathbb{Z})$ have been observed with the goal to predict or forecast the future values of $X_{n+1}, X_{n+2}, \ldots$. The focus is here on so-called *one-step best linear predictors (BLP)*. These are, by definition, linear combinations

$$\hat{X}_{n+1} = \phi_{n0} + \phi_{n1} X_n + \ldots + \phi_{nn} X_1 \tag{3.4.1}$$

of the observed variables $X_1, \ldots, X_n$ that minimize the mean-squared error

$$E\left[\{X_{n+1} - g(X_1, \ldots, X_n)\}^2\right]$$

for functions g of $X_1, \ldots, X_n$. Straightforward generalizations yield definitions for the m-step best linear predictors $\hat{X}_{n+m}$ of X_{n+m} for arbitrary $m \in \mathbb{N}$ in the same fashion. Using Hilbert space theory, one can prove the following theorem which will be the starting point for our considerations.

Theorem 3.4.1 (Best linear prediction)**.** *Let $(X_t\colon t \in \mathbb{Z})$ be a weakly stationary stochastic process of which $X_1, \dots, X_n$ are observed. Then, the one-step BLP $\hat{X}_{n+1}$ of X_{n+1} is determined by the equations*

$$E\left[(X_{n+1} - \hat{X}_{n+1})X_{n+1-j}\right] = 0$$

for all $j = 1, \dots, n+1$, where $X_0 = 1$.

The equations specified in Theorem 3.4.1 can be used to calculate the coefficients $\phi_{n0}, \dots, \phi_{nn}$ in (3.4.1). It suffices to focus on mean zero processes $(X_t\colon t \in \mathbb{Z})$ and thus to set $\phi_{n0} = 0$ as the following calculations show. Assume that $E[X_t] = \mu$ for all $t \in \mathbb{Z}$. Then, Theorem 3.4.1 gives that $E[\hat{X}_{n+1}] = E[X_{n+1}] = \mu$ (using the equation with $j = n+1$). Consequently, it holds that

$$\mu = E[\hat{X}_{n+1}] = E\left[\phi_{n0} + \sum_{\ell=1}^{n} \phi_{n\ell} X_{n+1-\ell}\right] = \phi_{n0} + \sum_{\ell=1}^{n} \phi_{n\ell}\mu.$$

Using now that $\phi_{n0} = \mu(1 - \phi_{n1} - \ldots - \phi_{nn})$, equation (3.4.1) can be rewritten as

$$\hat{Y}_{n+1} = \phi_{n1}Y_n + \ldots + \phi_{nn}Y_1,$$

where $\hat{Y}_{n+1} = \hat{X}_{n+1} - \mu$ has mean zero.

With the ACVF γ of $(X_t\colon t \in \mathbb{Z})$, the equations in Theorem 3.4.1 can be expressed as

$$\sum_{\ell=1}^{n} \phi_{n\ell}\gamma(j-\ell) = \gamma(j), \qquad j = 1, \dots, n. \tag{3.4.2}$$

Note that due to the convention $\phi_{n0} = 0$, the last equation in Theorem 3.4.1 (for which $j = n+1$) is omitted. More conveniently, this is restated in matrix notation. To this end, let $\boldsymbol{\Gamma}_n = (\gamma(j-\ell))_{j,\ell=1,\dots,n}$, $\boldsymbol{\phi}_n = (\phi_{n1}, \dots, \phi_{nn})^T$ and $\boldsymbol{\gamma}_n = (\gamma(1), \dots, \gamma(n))^T$, where ^{T} denotes the transpose. With these notations, (3.4.2) becomes

$$\boldsymbol{\Gamma}_n\boldsymbol{\phi}_n = \boldsymbol{\gamma}_n \qquad \Longleftrightarrow \qquad \boldsymbol{\phi}_n = \boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n, \tag{3.4.3}$$

provided that $\boldsymbol{\Gamma}_n$ is nonsingular. The determination of the coefficients $\phi_{n\ell}$ has thus been reduced to solving a linear equation system and depends only on second-order properties of $(X_t\colon t \in \mathbb{Z})$ which are given by the ACVF γ.

Let $\boldsymbol{X}_n = (X_n, X_{n-1}, \dots, X_1)^T$. Then, $\hat{X}_{n+1} = \boldsymbol{\phi}_n^T\boldsymbol{X}_n$. To assess the quality of the prediction, one computes the mean-squared error with the help of (3.4.3) as follows:

$$\begin{aligned}
P_{n+1} &= E\left[(X_{n+1} - \hat{X}_{n+1})^2\right] \\
&= E\left[(X_{n+1} - \boldsymbol{\phi}_n^T\boldsymbol{X}_n)^2\right] \\
&= E\left[(X_{n+1} - \boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X}_n)^2\right] \\
&= E\left[X_{n+1}^2 - 2\boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X}_n X_{n+1} + \boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{X}_n\boldsymbol{X}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n\right] \\
&= \gamma(0) - 2\boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\Gamma}_n\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n \\
&= \gamma(0) - \boldsymbol{\gamma}_n^T\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n.
\end{aligned} \tag{3.4.4}$$

As an initial example, we explain the prediction procedure for an autoregressive process of order 2.

Example 3.4.1 (Prediction of an AR(2) Process)**.** Let $(X_t\colon t \in \mathbb{Z})$ be the causal AR(2) process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$. Suppose that only an observation of X_1 is available to forecast the value of X_2. In this simplified case, the single prediction equation (3.4.2) is

$$\phi_{11}\gamma(0) = \gamma(1),$$

so that $\phi_{11} = \rho(1)$ and $\hat{X}_{1+1} = \rho(1)X_1$. In the next step, assume that observed values of X_1 and X_2 are at hand to forecast the value of X_3. Then, one similarly obtains from (3.4.2) that the predictor can be computed from

$$\begin{aligned}\hat{X}_{2+1} &= \phi_{21}X_2 + \phi_{22}X_1 = \boldsymbol{\phi}_2^T \boldsymbol{X}_2 = (\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\gamma}_2)^T \boldsymbol{X}_2 \\ &= (\gamma(1), \gamma(2)) \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} X_2 \\ X_1 \end{pmatrix}.\end{aligned}$$

However, applying the arguments leading to the definition of the PACF in Section 3.3, one finds that

$$E\left[\{X_3 - (\phi_1 X_2 + \phi_2 X_1)\}X_1\right] = E[Z_3 X_1] = 0,$$

$$E\left[\{X_3 - (\phi_1 X_2 + \phi_2 X_1)\}X_2\right] = E[Z_3 X_2] = 0.$$

Hence, $\hat{X}_{2+1} = \phi_1 X_2 + \phi_2 X_1$ and even $\hat{X}_{n+1} = \phi_1 X_n + \phi_2 X_{n-1}$ for all $n \geq 2$, exploiting the particular autoregressive structure. Since similar results can be proved for general causal AR(p) processes, the one-step predictors have the form

$$\hat{X}_{n+1} = \phi_1 X_n + \ldots + \phi_p X_{n-p+1}$$

whenever the number of observed variables n is at least p.

The major drawback of this approach is immediately apparent from the previous example: For larger sample sizes n, the prediction procedure requires the calculation of the inverse matrix $\boldsymbol{\Gamma}_n^{-1}$ which is computationally expensive. In the remainder of this section, two recursive prediction methods are introduced that bypass the inversion altogether. They are known as *Durbin-Levinson algorithm* and *innovations algorithm.* Finally, predictors based on the *infinite past* are introduced which are often easily applicable for the class of causal and invertible ARMA processes.

Method 1 (The Durbin-Levinson algorithm) If $(X_t\colon t \in \mathbb{Z})$ is a zero mean weakly stationary process with ACVF γ such that $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$, then the coefficients $\phi_{n\ell}$ in (3.4.2) and the mean squared errors P_n in (3.4.4) satisfy the recursions

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)}, \qquad P_0 = \gamma(0),$$

and, for $n \geq 1$,

$$\phi_{nn} = \frac{1}{P_{n-1}}\left(\gamma(n) - \sum_{\ell=1}^{n-1} \phi_{n-1,\ell}\gamma(n-\ell)\right),$$

$$\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

and

$$P_n = P_{n-1}(1 - \phi_{nn}^2).$$

It can be shown that under the assumptions made on the process $(X_t\colon t \in \mathbb{Z})$, it holds indeed that ϕ_{nn} is equal to the value of the PACF of $(X_t\colon t \in \mathbb{Z})$ at lag n. The result is formulated as Corollary 5.2.1 in Brockwell and Davis (1991). This fact is highlighted in an example.

Example 3.4.2 (The PACF of an AR(2) process)**.** Let $(X_t\colon t \in \mathbb{Z})$ be a causal AR(2) process. Then, $\rho(1) = \phi_1/(1-\phi_2)$ and all other values can be computed recursively from

$$\rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) = 0, \qquad h \geq 2.$$

Note that the ACVF γ satisfies a difference equation with the same coefficients, which is seen by multiplying the latter equation with $\gamma(0)$. Applying the Durbin-Levinson algorithm gives first that

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)} = \rho(1) \qquad \text{and} \qquad P_1 = P_0(1-\phi_{11}^2) = \gamma(0)(1-\rho(1)^2).$$

Ignoring the recursion for the error terms P_n in the following, the next $\phi_{n\ell}$ values are obtained as

$$\begin{aligned} \phi_{22} &= \frac{1}{P_1}\left[\gamma(2) - \phi_{11}\gamma(1)\right] = \frac{1}{1-\rho(1)^2}\left[\rho(2) - \rho(1)^2\right] \\ &= \frac{\phi_1^2(1-\phi_2)^{-1} + \phi_2 - [\phi_1(1-\phi_2)^{-1}]^2}{1 - [\phi_1(1-\phi_2)^{-1}]^2} = \phi_2, \end{aligned}$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \rho(1)(1-\phi_2) = \phi_1,$$

$$\phi_{33} = \frac{1}{P_2}\left[\gamma(3) - \phi_{21}\gamma(2) - \phi_{22}\gamma(1)\right] = \frac{1}{P_2}\left[\gamma(3) - \phi_1\gamma(2) - \phi_2\gamma(2)\right] = 0.$$

Now, referring to the remarks after Example 3.3.3, no further computations are necessary to determine the PACF because $\phi_{nn} = 0$ for all $n > p = 2$.

Method 2 (The innovations algorithm) In contrast to the Durbin-Levinson algorithm, this method can also be applied to nonstationary processes. It should thus, in general, be preferred over Method 1. The innovations algorithm gets its name from the fact that one directly uses the form of the prediction equations in Theorem 3.4.1 which are stated in terms of the *innovations* $(X_{t+1} - \hat{X}_{t+1})_{t\in\mathbb{Z}}$. Observe that the sequence consists of uncorrelated random variables.

The one-step predictors $\hat{X}_{n+1}$ can be calculated from the recursions

$$\hat{X}_{0+1} = 0, \qquad P_1 = \gamma(0)$$

and, for $n \geq 1$,

$$\hat{X}_{n+1} = \sum_{\ell=1}^{n} \theta_{n\ell}(X_{n+1-\ell} - \hat{X}_{n+1-\ell})$$

$$P_{n+1} = \gamma(0) - \sum_{\ell=0}^{n-1} \theta_{n,n-\ell}^2 P_{\ell+1},$$

where the coefficients are obtained from the equations

$$\theta_{n,n-\ell} = \frac{1}{P_{\ell+1}} \left[\gamma(n-\ell) - \sum_{i=0}^{\ell-1} \theta_{\ell,\ell-i}\theta_{n,n-i}P_{i+1} \right], \qquad \ell = 0, 1, \ldots, n-1.$$

As example we show how the innovations algorithm is applied to a moving average time series of order 1.

Example 3.4.3 (Prediction of an MA(1) Process)**.** Let $(X_t \colon t \in \mathbb{Z})$ be the MA(1) process $X_t = Z_t + \theta Z_{t-1}$. Note that

$$\gamma(0) = (1+\theta^2)\sigma^2, \qquad \gamma(1) = \theta\sigma^2 \qquad \text{and} \qquad \gamma(h) = 0 \quad (h \geq 2).$$

Using the innovations algorithm, one can compute the one-step predictor from the values

$$\theta_{n1} = \frac{\theta\sigma^2}{P_n}, \qquad \theta_{n\ell} = 0 \quad (\ell = 2, \ldots, n-1),$$

and

$$P_1 = (1+\theta^2)\sigma^2,$$

$$P_{n+1} = (1 + \theta^2 - \theta\theta_{n1})\sigma^2$$

as

$$\hat{X}_{n+1} = \frac{\theta\sigma^2}{P_n}(X_n - \hat{X}_n).$$

Method 3 (Prediction based on the infinite past) Suppose that a causal and invertible ARMA(p, q) process is analyzed. Assume further that (unrealistically) the complete history of the process can be stored and that thus all past variables $(X_t \colon t \leq n)$ can be accessed. Define then

$$\tilde{X}_{n+m} = E[X_{n+m}|X_n, X_{n-1}, \ldots],$$

as the m-step ahead predictor based on the infinite past. It can be shown that, for large sample sizes n, the difference between the values of $\hat{X}_{n+m}$ and $\tilde{X}_{n+m}$ vanishes at an exponential rate. Exploiting causality and invertibility of the ARMA process, one can transform the predictor $\tilde{X}_{n+m}$ so that it is in a computationally more feasible form. To do so, note that by causality

$$\tilde{X}_{n+m} = E[X_{n+m}|X_n, X_{n-1}, \ldots]$$

$$= E\left[\sum_{j=0}^{\infty} \psi_j Z_{n+m-j} \Big| X_n, X_{n-1}, \ldots\right]$$

$$= \sum_{j=m}^{\infty} \psi_j Z_{n+m-j} \tag{3.4.5}$$

because $E[Z_t|X_n, X_{n-1}, \ldots]$ equals zero if $t > n$ and equals Z_t if $t \leq n$ (due to invertibility!). The representation in (3.4.5) can be used to compute the mean squared prediction error $\tilde{P}_{n+m}$. It follows from causality that

$$\tilde{P}_{n+m} = E[(X_{n+m} - \tilde{X}_{n+m})^2] = E\left[\left(\sum_{j=0}^{m-1} \psi_j Z_{n+m-j}\right)^2\right] = \sigma^2 \sum_{j=0}^{m-1} \psi_j^2. \tag{3.4.6}$$

On the other hand, (3.4.5) does not allow to directly calculate the forecasts because $\tilde{X}_{n+m}$ is given in terms of the noise variables Z_{n+m-j}. Instead invertibility will be utilized. Observe first that

$$E[X_{n+m-j}|X_n, X_{n-1}, \ldots] = \begin{cases} \tilde{X}_{n+m-j}, & j < m. \\ X_{n+m-j}, & j \geq m. \end{cases}$$

By invertibility (the "0 =" part follows again from causality),

$$0 = E[Z_{n+m}|X_n, X_{n-1}, \ldots]$$

$$= E\left[\sum_{j=0}^{\infty} \pi_j X_{n+m-j} \Big| X_n, X_{n-1}, \ldots\right]$$

$$= \sum_{j=0}^{\infty} \pi_j E[X_{n+m-j}|X_n, X_{n-1}, \ldots].$$

Combining the previous two statements, yields

$$\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}. \tag{3.4.7}$$

The equations can now be solved recursively for $m = 1, 2, \ldots$ Note, however, that for any $m \geq 1$ the sequence $(X_{n+m+t} - \tilde{X}_{n+m+t} \colon t \in \mathbb{Z})$ does not consist of uncorrelated random variables. In fact, if $h \in \mathbb{N}_0$, it holds that

$$E[(X_{n+m} - \tilde{X}_{n+m})(X_{n+m+h} - \tilde{X}_{n+m+h})]$$

$$= E\left[\sum_{j=0}^{m-1} \psi_j Z_{n+m-j} \sum_{i=0}^{m+h-1} \psi_i Z_{n+m+h-i}\right]$$

$$= \sigma^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+h}.$$

Finally, for practical purposes the given forecast needs to be truncated. This is accomplished by setting

$$\sum_{j=n+m}^{\infty} \pi_j X_{n+m-j} = 0.$$

The resulting equations (see (3.4.7) for comparison) yield recursively the truncated m-step predictors X^*_{n+m}:

$$X^*_{n+m} = -\sum_{j=1}^{m-1} \pi_j X^*_{n+m-j} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}. \tag{3.4.8}$$

3.5 Parameter Estimation

Let $(X_t\colon t \in \mathbb{Z})$ be a causal and invertible ARMA(p,q) process with known orders p and q, possibly with mean μ. This section is concerned with estimation procedures for the unknown parameter vector

$$\boldsymbol{\beta} = (\mu, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2)^T. \tag{3.5.1}$$

To simplify the estimation procedure, it is assumed that the data has already been adjusted by subtraction of the mean and the discussion is therefore restricted to zero mean ARMA models.

In the following, three estimation methods are introduced. The method of moments works best in case of pure AR processes, while it does not lead to optimal estimation procedures for general ARMA processes. For the latter, more efficient estimators are provided by the maximum likelihood and least squares methods which will be discussed subsequently.

Method 1 (Method of Moments) Since this method is only efficient in their case, the presentation here is restricted to AR(p) processes

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t, \qquad t \in \mathbb{Z},$$

where $(Z_t\colon t \in \mathbb{Z}) \sim \text{WN}(0, \sigma^2)$. The parameter vector $\boldsymbol{\beta}$ consequently reduces to $(\boldsymbol{\phi}, \sigma^2)^T$ with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$ and can be estimated using the *Yule-Walker equations*

$$\boldsymbol{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p \qquad \text{and} \qquad \sigma^2 = \gamma(0) - \boldsymbol{\phi}^T \boldsymbol{\gamma}_p,$$

where $\boldsymbol{\Gamma}_p = (\boldsymbol{\gamma}(k-j))_{k,j=1,\ldots,p}$ and $\boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))^T$. Observe that the equations are obtained by the same arguments applied to derive the Durbin-Levinson algorithm in the previous section. The method of moments suggests to replace every quantity in the Yule-Walker equations with their estimated counterparts, which yields the *Yule-Walker estimators*

$$\widehat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p = \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \tag{3.5.2}$$

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_p^T \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p = \hat{\gamma}(0) \left[1 - \hat{\boldsymbol{\rho}}_p^T \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p\right]. \tag{3.5.3}$$

Therein, $\hat{\boldsymbol{R}}_p = \hat{\gamma}(0)^{-1} \hat{\boldsymbol{\Gamma}}_p$ and $\hat{\boldsymbol{\rho}}_p = \hat{\gamma}(0)^{-1} \hat{\boldsymbol{\gamma}}_p$ with $\hat{\gamma}(h)$ defined as in (1.2.1). Using $\hat{\gamma}(h)$ as estimator for the ACVF at lag h, a dependence on the sample size n is obtained in an implicit way. This dependence is suppressed in the notation used here. The following theorem contains the limit behavior of the Yule-Walker estimators as n tends to infinity.

Theorem 3.5.1. *If $(X_t\colon t \in \mathbb{Z})$ is a causal AR(p) process, then*

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \sigma^2 \boldsymbol{\Gamma}_p^{-1}) \qquad \text{and} \qquad \hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

as $n \to \infty$, where $\to^P$ indicates convergence in probability.

A proof of this result is given in Section 8.10 of Brockwell and Davis (1991). Since equations (3.5.2) and (3.5.3) have the same structure as the corresponding equations (3.4.3) and (3.4.4), the Durbin-Levinson algorithm can be used to solve recursively for the estimators $\widehat{\boldsymbol{\phi}}_h = (\widehat{\phi}_{h1}, \ldots, \widehat{\phi}_{hh})$. Moreover, since ϕ_{hh} is equal to the value of the PACF of $(X_t\colon t \in \mathbb{Z})$ at lag h, the estimator $\widehat{\phi}_{hh}$ can be used as its proxy. Since it is already known that, in the case of AR(p) processes, $\phi_{hh} = 0$ if $h > p$, Theorem 3.5.1 implies immediately the following corollary.

Corollary 3.5.1. *If $(X_t\colon t \in \mathbb{Z})$ is a causal AR(p) process, then*

$$\sqrt{n}\widehat{\phi}_{hh} \xrightarrow{\mathcal{D}} Z \qquad (n \to \infty)$$

for all $h > p$, where Z stands for a standard normal random variable.

Example 3.5.1 (Yule-Walker estimates for AR(2) processes)**.** Suppose that $n = 144$ values of the autoregressive process $X_t = 1.5X_{t-1} - .75X_{t-2} + Z_t$ have been observed, where $(Z_t\colon t \in \mathbb{Z})$ is a sequence of independent standard normal variates. Assume further that $\hat{\gamma}(0) = 8.434$, $\hat{\rho}(1) = 0.834$ and $\hat{\rho}(2) = 0.476$ have been calculated from the data. The Yule-Walker estimators for the parameters are then given by

$$\widehat{\boldsymbol{\phi}} = \begin{pmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{pmatrix} = \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} \begin{pmatrix} 0.834 \\ 0.476 \end{pmatrix} = \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix}$$

and

$$\hat{\sigma}^2 = 8.434 \left[1 - (0.834, 0.476) \begin{pmatrix} 1.439 \\ -0.725 \end{pmatrix} \right] = 1.215.$$

To construct asymptotic confidence intervals using Theorem 3.5.1, the unknown limiting covariance matrix $\sigma^2 \boldsymbol{\Gamma}_p^{-1}$ needs to be estimated. This can be done using the estimator

$$\frac{\hat{\sigma}^2 \hat{\boldsymbol{\Gamma}}_p^{-1}}{n} = \frac{1}{144} \frac{1.215}{8.434} \begin{pmatrix} 1.000 & 0.834 \\ 0.834 & 1.000 \end{pmatrix}^{-1} = \begin{pmatrix} 0.057^2 & -0.003 \\ -0.003 & 0.057^2 \end{pmatrix}.$$

Then, the $1 - \alpha$ level confidence interval for the parameters ϕ_1 and ϕ_2 are computed as

$$1.439 \pm 0.057 z_{1-\alpha/2} \qquad \text{and} \qquad -0.725 \pm 0.057 z_{1-\alpha/2},$$

respectively, where $z_{1-\alpha/2}$ is the corresponding normal quantile.

Example 3.5.2 (Recruitment Series)**.** Let us reconsider the recruitment series of Example 3.3.5. There, an AR(2) model was first established as appropriate for the data and the model parameters were then estimated using an ordinary least squares approach. Here, the coefficients will instead be estimated with

the Yule-Walker procedure. The R command is

```
> rec.yw = ar.yw(rec, order=2)
```

The mean estimate can be obtained from `rec.yw$x.mean` as $\hat{\mu} = 62.26$, while the autoregressive parameter estimates and their standard errors are accessed with the commands `rec.yw$ar` and `sqrt(rec.yw$asy.var.coef)` as $\hat{\phi}_1 = 1.3316(.0422)$ and $\hat{\phi}_2 = -.4445(.0422)$. Finally, the variance estimate is obtained from `rec.yw$var.pred` as $\hat{\sigma}^2 = 94.7991$. All values are close to their counterparts in Example 3.3.5.

Example 3.5.3. Consider the invertible MA(1) process $X_t = Z_t + \theta Z_{t-1}$, where $|\theta| < 1$. Using invertibility, each X_t has an infinite autoregressive representation

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

that is nonlinear in the unknown parameter θ to be estimated. The method of moments is here based on solving

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1+\hat{\theta}^2}.$$

for $\hat{\theta}$. The foregoing quadratic equation has the two solutions

$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)},$$

of which we pick the invertible one. Note moreover, that $|\hat{\rho}(1)|$ is not necessarily less or equal to $1/2$ which is required for the existence of real solutions. (The theoretical value $|\rho(1)|$, however, is always less than 1/2 for any MA(1) process, as an easy computation shows). Hence, θ can not always be estimated from given data samples.

Method 2 (Maximum Likelihood Estimation) The innovations algorithm of the previous section applied to a causal ARMA(p, q) process $(X_t \colon t \in \mathbb{Z})$ gives

$$\hat{X}_{i+1} = \sum_{j=1}^{i} \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \qquad 1 \le i < \max\{p, q\},$$

$$\hat{X}_{i+1} = \sum_{j=1}^{p} \phi_j X_{i+1-j} + \sum_{j=1}^{q} \theta_{ij}(X_{i+1-j} - \hat{X}_{i+1-j}), \qquad i \ge \max\{p, q\},$$

with prediction error

$$P_{i+1} = \sigma^2 R_{i+1}.$$

In the last expression, σ^2 has been factored out due to reasons that will become apparent from the form of the likelihood function to be discussed below. Recall that the sequence $(X_{i+1} - \hat{X}_{i+1} \colon i \in \mathbb{Z})$ consists of uncorrelated random variables if the parameters are known. Assuming normality for the errors, we moreover obtain even independence. This can be exploited to define the *Gaussian maximum*

likelihood estimation (MLE) procedure. Throughout, it is assumed that $(X_t\colon t \in \mathbb{Z})$ has zero mean ($\mu = 0$). The parameters of interest are collected in the vectors $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)^T$ and $\boldsymbol{\beta}' = (\boldsymbol{\phi}, \boldsymbol{\theta})^T$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$. Assume finally that we have observed the variables $X_1, \ldots, X_n$. Then, the Gaussian likelihood function for the innovations is

$$L(\boldsymbol{\beta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left(\prod_{i=1}^{n} R_i^{1/2} \right) \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{R_j} \right). \tag{3.5.4}$$

Taking the partial derivative of $\ln L(\boldsymbol{\beta})$ with respect to the variable σ^2 reveals that the MLE for σ^2 can be calculated from

$$\hat{\sigma}^2 = \frac{S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})}{n}, \qquad S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{R_j}.$$

Therein, $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ denote the MLEs of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ obtained from minimizing the *profile likelihood* or *reduced likelihood*

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln\left(\frac{S(\boldsymbol{\phi}, \boldsymbol{\theta})}{n} \right) + \frac{1}{n} \sum_{j=1}^{n} \ln(R_j).$$

Observe that the profile likelihood $\ell(\boldsymbol{\phi}, \boldsymbol{\theta})$ can be computed using the innovations algorithm. The speed of these computations depends heavily on the quality of initial estimates. These are often provided by the non-optimal Yule-Walker procedure. For numerical methods, such as the *Newton-Raphson* and *scoring algorithms*, see Section 3.6 in Shumway and Stoffer (2006).

The limit distribution of the MLE procedure is given as the following theorem. Its proof can be found in Section 8.8 of Brockwell and Davis (1991).

Theorem 3.5.2. *Let $(X_t\colon t \in \mathbb{Z})$ be a causal and invertible ARMA(p, q) process defined with an iid sequence $(Z_t\colon t \in \mathbb{Z})$ satisfying $E[Z_t] = 0$ and $E[Z_t^2] = \sigma^2$. Consider the MLE $\hat{\boldsymbol{\beta}}'$ of $\boldsymbol{\beta}'$ that is initialized with the moment estimators of Method 1. Then,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}' - \boldsymbol{\beta}') \xrightarrow{\mathcal{D}} N(0, \sigma^2 \boldsymbol{\Gamma}_{p,q}^{-1}) \qquad (n \to \infty).$$

The result is optimal. The covariance matrix $\boldsymbol{\Gamma}_{p,q}$ is in block form and can be evaluated in terms of covariances of various autoregressive processes.

Example 3.5.4 (Recruitment Series)**.** The MLE estimation procedure for the recruitment series can be applied in R as follows:

```
> rec.mle = ar.mle(rec, order=2)
```

The mean estimate can be obtained from `rec.mle$x.mean` as $\hat{\mu} = 62.26$, while the autoregressive parameter estimates and their standard errors are accessed with the commands `rec.mle$ar` and `sqrt(rec.mle$asy.var.coef)` as $\hat{\phi}_1 = 1.3513(.0410)$ and $\hat{\phi}_2 = -.4099(.0410)$. Finally, the variance estimate is obtained from `rec.yw$var.pred` as $\hat{\sigma}^2 = 89.3360$. All values are very close to their counterparts in Example 3.3.5.

Method 3 (Least Squares Estimation) An alternative to the method of moments and the MLE is provided by the least squares estimation (LSE). For causal and invertible ARMA(p, q) processes, it is based on minimizing the weighted sum of squares

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{R_j} \tag{3.5.5}$$

with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, respectively. Assuming that $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\theta}}$ denote these LSEs, the LSE for σ^2 is computed as

$$\tilde{\sigma}^2 = \frac{S(\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}})}{n - p - q}.$$

The least squares procedure has the same asymptotics as the MLE.

Theorem 3.5.3. *The result of Theorem 3.5.2 holds also if $\hat{\boldsymbol{\beta}}'$ is replaced with $\tilde{\boldsymbol{\beta}}'$.*

Example 3.5.5 (Recruitment Series)**.** The least squares estimation has already been discussed in Example 3.3.5, including the R commands.

3.6 Model Selection

In this section, a rough guide for going about the data analysis will be provided. It consists of several parts, most of which have been discussed previously. The main focus is on the selection of p and q in the likely case that these parameters are unknown.

Step 1. Plot the data and check whether or not the variability remains reasonably stable throughout the observation period. If that is not the case, use preliminary transformations to stabilize the variance. One popular class is given by the *Box-Cox transformations* (Box and Cox, 1964)

$$f_\lambda(U_t) = \begin{cases} \lambda^{-1}(U_t^\lambda - 1), & U_t \geq 0,\ \lambda > 0. \\ \ln U_t & U_t > 0,\ \lambda = 0. \end{cases}$$

In practice f_0 or $f_{1/2}$ are often adequate choices. (Recall, for instance, the Australian wine sales data of Example 1.4.1.)

Step 2. Remove, if present, trend and seasonal components from the data. Chapter 1 introduced a number of tools to do so, based on the classical decomposition of a time series

$$Y_t = m_t + s_t + X_t$$

into a trend, a seasonality and a residual component. Note that differencing works also without the specific representation in the last display. If the data appears stationary, move on to the next step. Else apply, for example, another set of difference operations.

Step 3. Suppose now that Steps 1 and 2 have provided us with observations that are well described by a stationary sequence $(X_t \colon t \in \mathbb{Z})$. The goal is then to find the most appropriate ARMA(p, q) model

to describe the process. In the unlikely case that p and q can be assumed known, utilize the estimation procedures of Section 3.5 directly. Otherwise, choose them according to one of the following criteria.

(a) The standard criterion that is typically implemented in software packages is a modification of *Akaike's information criterion*, see Akaike (1969), which was given by Hurvich and Tsai (1989). In this paper, it is suggested that the ARMA model parameters be chosen to minimize the objective function

$$\mathrm{AIC}_C(\boldsymbol{\phi}, \boldsymbol{\theta}, p, q) = -2 \ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, S(\boldsymbol{\phi}, \boldsymbol{\theta})/n) + \frac{2(p+q+1)n}{n-p-q-2}. \tag{3.6.1}$$

Here, $L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ denotes the Gaussian likelihood defined in (3.5.4) and $S(\boldsymbol{\phi}, \boldsymbol{\theta})$ is the weighted sum of squares in (3.5.5). It can be seen from the definition that the AIC_C does not attempt to minimize the log-likelihood function directly. The introduction of the penalty term on the right-hand side of (3.6.1) reduces the risk of overfitting.

(b) For pure autoregressive processes, Akaike (1969) introduced a criterion that is based on a minimization of the *final prediction error*. Here, the order p is chosen as the minimizer of the objective function

$$\mathrm{FPE} = \hat{\sigma}^2 \frac{n+p}{n-p},$$

where $\hat{\sigma}^2$ denotes the MLE of the unknown noise variance σ^2. For more on this topic and other procedures that help fit a model, we refer here to Section 9.3 of Brockwell and Davis (1991).

Step 4. The last step in the analysis is concerned with *diagnostic checking* by applying the goodness of fit tests of Section 1.5.

3.7 Summary

The class of autoregressive moving average processes has been introduced to model stationary stochastic processes. Theoretical properties such as causality and invertibility have been examined, which depend on the zeroes of the autoregressive and moving average polynomials, respectively.

It has been shown how the causal representation of an ARMA process can be utilized to compute its covariance function which contains all information about the dependence structure.

Assuming known parameter values, several forecasting procedures have been discussed. The Durbin-Levinson algorithm works well for pure AR processes, while the innovations algorithm is particularly useful for pure MA processes. Predictions using an infinite past work well for causal and invertible ARMA processes. For practical purposes, however, a truncated version is more relevant.

Since the exact parameter values are in general unknown, several estimation procedures were introduced. The Yule-Walker procedure is only optimal in the AR case but provides useful initial estimates that can be used for the numerical derivation of maximum likelihood or least squares estimates.

Finally, a framework has been provided that may potentially be useful when facing the problem of analyzing a data set in practice.