Autoregressive Moving Average Processes (ARMA) 3

3.1 Summary

This chapter is concerned with the modelling of serial correlation (or autocorrelation) that is characteristic of many time series dynamics. To that end we cover a class of stochastic processes widely used in practice. They are discrete-time processes: $\{x_t\}_{t \in \mathbb{T}}$ with $\mathbb{T} \subseteq \mathbb{Z}$. Throughout this chapter, the innovations or shocks $\{\varepsilon_t\}$ behind $\{x_t\}$ are assumed to form a white noise sequence as defined in Example 2.6. The next section treats the rather simple moving average structure. The third section addresses the inversion of lag polynomials at a general level. The fourth section breaks down the technical aspects to the application with ARMA processes.

3.2 Moving Average Processes

We define the **moving average process** of order q (MA(q)) as

$$x_t = \mu + b_0\, \varepsilon_t + b_1\, \varepsilon_{t-1} + \cdots + b_q\, \varepsilon_{t-q}\,, \quad b_0 = 1\,,\ t \in \mathbb{T}\,. \tag{3.1}$$

In the following, we assume a white noise process for the so-called innovation $\{\varepsilon_t\}$ governing the processes considered. In general, we assume $b_q \neq 0$. Let us consider a special case before we enter the general discussion of the model.

MA(1)

We set μ to zero and q to one in (3.1) and thereby obtain

$$x_t = \varepsilon_t + b\, \varepsilon_{t-1}.$$

U. Hassler, *Stochastic Processes and Calculus*, Springer Texts in Business and Economics, DOI 10.1007/978-3-319-23428-1_3

As the innovations are WN$(0, \sigma^2)$, the moments are independent of time: Firstly, it holds that $\mathrm{E}(x_t) = \mu = 0$. Secondly, one obtains from

$$\gamma(h) = \mathrm{Cov}(x_t, x_{t+h}) = \mathrm{E}\left((\varepsilon_t + b\,\varepsilon_{t-1})\,(\varepsilon_{t+h} + b\,\varepsilon_{t+h-1})\right)$$

immediately

$$\gamma(0) = \sigma^2(1 + b^2)\,, \quad \gamma(1) = \sigma^2 b\,,$$

and

$$\gamma(h) = 0 \quad \text{for} \quad h > 1\,.$$

For the MA(1) process considered the autocorrelation function $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$ therefore is:

$$\rho(1) = \frac{b}{1 + b^2}, \quad \rho(h) = 0,\ h > 1.$$

This is why the process is always (weakly) stationary without any assumptions with respect to b or to the index set $\mathbb{T}$. Elementary curve sketching shows that $\rho(1)$ considered with respect to b,

$$\rho(1; b) = \frac{b}{1 + b^2},$$

becomes extremal for $b = \pm 1$. For $b = -1$, $\rho(1)$ takes the minimum $-\frac{1}{2}$, and for $b = 1$ it takes the maximum $\frac{1}{2}$ (see Problem 3.1).

In Fig. 3.1 realizations were simulated by means of so-called pseudo random number with each 50 observations of moving average processes. All three graphs were generated by the same realizations of $\{\varepsilon_t\}$. The first graph with $b_1 = b = 0$ shows the simulation of a pure random process $\{\varepsilon_t\}$: The times series oscillates arbitrarily around the expected value zero. The third graph with $b_1 = b = 0.9$ depicts the case of (strong) positive autocorrelation of first order: Positive values are followed by positive values whereas negative values tend to entail negative values, i.e. the zero line is less often crossed than in the first case. Finally, the graph in the middle lies in between both extreme cases as weak positive autocorrelation ($b_1 = b = 0.3$) is present.

MA(q)

The results obtained for the MA(1) process can be generalized. Every MA process is – for all parameter values independent of starting value conditions – always stationary. For the MA(q) process, it holds that its autocorrelation sequence vanishes from the order q on (i.e. it becomes zero). The proof is elementary and is therefore omitted.

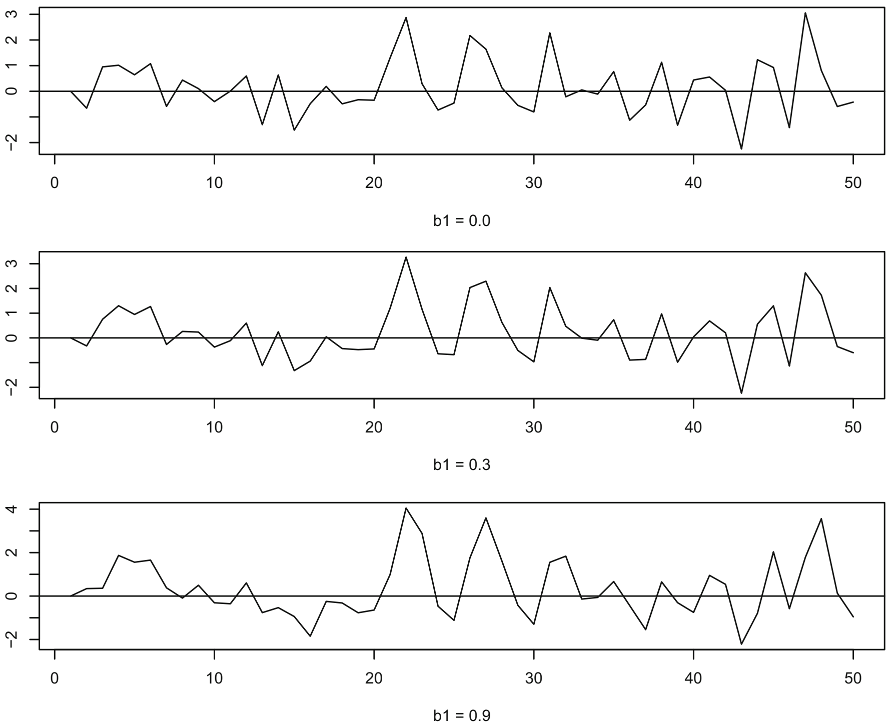

Fig. 3.1 Simulated MA(1) processes with $\sigma^2 = 1$ and $\mu = 0$

Proposition 3.1 (MA(q)) *Assume an MA(q) process from (3.1).*

(a) The process is stationary with expectation μ.
(b) For the autocovariances it holds that,

$$\gamma(h) = \sigma^2(b_h + b_{h+1}b_1 + \ldots + b_q b_{q-h}), \quad h = 0, 1, \ldots q,$$

and $\gamma(h) = 0$ for $h > q$.

Example 3.1 (Seasonal MA) Let S denote the seasonality, e.g. $S = 4$ or $S = 12$ for quarterly or monthly data, or $S = 5$ for (work) daily observations. Hence, we define as a special MA process[1]

$$x_t = \varepsilon_t + b\,\varepsilon_{t-S}.$$

[1]For $S = 1$ we obtain the MA(1) case, however, without a seasonal interpretation.

In this context, $q = S$ and $b_q = b$ hold and

$$b_1 = b_2 = \ldots = b_{q-1} = 0\,.$$

Proposition 3.1 (b) yields:

$$\gamma(0) = \sigma^2(b_0 + b_1^2 + \cdots + b_q^2) = \sigma^2(1 + b^2)\,,$$
$$\gamma(1) = \sigma^2(b_1 + b_2\,b_1 + \cdots + b_q\,b_{q-1}) = 0\,,$$

and also

$$\gamma(h) = 0 \quad \text{for} \quad h = 1, 2, \ldots, S - 1\,.$$

The case $h = S$, however, yields

$$\gamma(S) = \sigma^2 b_S b_0 = \sigma^2 b\,.$$

For $h > S$, according to Proposition 3.1 it holds that $\gamma(h) = 0$. Hence, the process at hand is exclusively autocorrelated at lag S, which is why it is also called a seasonal MA process. ■

MA(∞) Processes

We now let q go off to infinity. For reasons that become obvious in the fourth section, the MA coefficients, however, are not denoted by b_j anymore. Instead, consider the infinite real sequence $\{c_j\}, j \in \mathbb{N}$, to define:

$$x_t = \mu + \sum_{j=0}^{\infty} c_j\,\varepsilon_{t-j}\,, \quad \sum_{j=0}^{\infty} |c_j| < \infty\,,\ c_0 = 1\,. \tag{3.2}$$

Sometimes the process from (3.2) is called "causal" as there are only past or contemporaneous random variables ε_{t-j}, $j \geq 0$, entering the process at time t. The condition on the coefficients $\{c_j\}$ of being absolutely summable guarantees that $\sum_{j=0}^{\infty} c_j\,\varepsilon_{t-j}$ in (3.2) is a well-defined random variable, which then can be called x_t, see e.g. Fuller (1996, Theorem 2.2.1). Absolute summability naturally implies square summability,

$$\sum_{j=0}^{\infty} c_j^2 < \infty\,,$$

and, indeed, sometimes we define the MA(∞) process upon this weaker assumption. Square summability of $\{c_j\}$ is sufficient for stationarity with $\mathrm{E}(x_t) = \mu$ and

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} c_j \, c_{j+h} \, ,$$

see Fuller (1996, Theorem 2.2.3), which is the first result of the follow proposition. The second one is established in Problem 3.3.

Proposition 3.2 (Infinite MA) *Assume an MA(∞) process,*

$$x_t = \mu + \sum_{j=0}^{\infty} c_j \, \varepsilon_{t-j} \, , \quad \{\varepsilon_t\} \sim WN(0, \sigma^2) \, , \; c_0 = 1 \, ,$$

with $\sum_{j=0}^{\infty} c_j^2 < \infty$.

(a) The process is stationary with expected value μ*, and for the autocovariances it holds that*

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} c_j \, c_{j+h} \, , \quad h = 0, 1, \ldots .$$

(b) Under absolute summability, $\sum_{j=0}^{\infty} |c_j| < \infty$*, the sequence of autocovariances is absolutely summable:*

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty \, .$$

The fact that the sequence of autocovariances is absolutely summable under (3.2), see (b) of the proposition, has an immediate implication: the autocovariances tend to zero with growing lag:

$$\gamma(h) \to 0 \quad \text{as } h \to \infty \, .$$

This means that the correlation between x_t and x_{t-h} tends to decrease with growing lag h.

Note that x_t from (3.2) or Proposition 3.2 is defined as a linear combination of ε_{t-j}, $j \geq 0$. Therefore, one sometimes speaks of a **linear process**. Other authors reserve this label for the more restricted case of iid innovations, where all temporal dependence of $\{x_t\}$ arises exclusively from the MA coefficients $\{c_j\}$. The results of Proposition 3.2, however, hold under the weaker assumption that $\{\varepsilon_t\}$ is white noise.

Impulse Responses

The MA(∞) coefficients are often called **impulse responses**, since they measure the effect of a shock j periods ago on x_t:

$$\frac{\partial x_t}{\partial \varepsilon_{t-j}} = c_j \,.$$

Square summability implies that shocks are transient in that $c_j \to 0$ as $j \to \infty$. The speed at which the impulse responses converge to zero characterizes the dynamics. In particular, Campbell and Mankiw (1987) popularized the cumulated impulse responses as measure of **persistence**, where the cumulated effect is defined as

$$CIR(J) = \sum_{j=0}^{J} c_j \,.$$

This measure quantifies the total effect up to J periods back, if there occurred a unit shock in each past period, including the present period at time t. Asymptotically, one obtains the so-called long-run effect, often called total multiplier in economics. This measure is defined as

$$CIR := \lim_{J\to\infty} CIR(J) \,, \tag{3.3}$$

provided that this quantity exists. Clearly, under absolute summability of the impulse responses, CIR is well defined. Andrews and Chen (1994) advocated CIR as being superior to alternative measures of persistence. We will return to the measurement and interpretation of persistence in the next chapter.

The MA(∞) process is of considerable generality. In fact, Wold (1938) showed that every stationary process with expected value zero can be decomposed into a square summable, purely non-deterministic MA(∞) component and an uncorrelated component, say $\{\delta_t\}$, which is deterministic in the sense that it is perfectly predictable from its past:[2]

$$\sum_{j=0}^{\infty} c_j\, \varepsilon_{t-j} + \delta_t \,, \quad \sum_{j=0}^{\infty} c_j^2 < \infty \,, \; \mathrm{E}(\varepsilon_{t-j}\delta_t) = 0 \,, \tag{3.4}$$

[2] Typically, one assumes that δ_t is identically equal to zero for all t, since "perfectly predictable" only allows for trivial processes like e.g. $\delta_t = (-1)^t A$ or $\delta_t = A$, where A is some random variable, such that $\delta_t = -\delta_{t-1}$ or $\delta_t = \delta_{t-1}$, respectively. Of course, this does not rule out the case of a constant mean μ different from zero as assumed in (3.2).

where $\mathrm{E}(\varepsilon_t) = \mathrm{E}(\delta_t) = 0$. For details on this **Wold decomposition** see Fuller (1996, Theorem 2.10.2) or Brockwell and Davis (1991, Theorem 5.7.1).

In practice it is not reasonable to construct a model with infinitely many parameters c_j as they cannot be estimated from a finite number of observations without further restrictions. In the fourth section of this chapter, however, we will learn that very simple, so-called autoregressive processes with a finite number of parameters possess an MA(∞) representation. In order to discuss autoregressive processes rigorously, we need to concern ourselves with polynomials in the lag operator and their invertibility.

3.3 Lag Polynomials and Invertibility

Frequently, time series models are written by means of the **lag operator**[3] L. The lag operator shifts the process $\{x_t\}$ by one unit in time: $Lx_t = x_{t-1}$. By the inverse operator L^{-1} the shift is just reversed, $L^{-1}Lx_t = x_t$, or $L^{-1}x_t = x_{t+1}$. Successive use of the operator is denoted by its power, $L^j x_t = x_{t-j}$, $j \in \mathbb{Z}$. The identity is described with L^0, $L^0 x_t = x_t$. Applied to a constant c, the operator leaves the value unchanged, $Lc = c$.

Causal Linear Filters

Let us consider an input process $\{x_t\}$, $t \in \mathbb{T}$, which is transferred into an output process $\{y_t\}$ by linear filtering,

$$y_t = \sum_{j=0}^{p} w_j\, x_{t-j}\,, \tag{3.5}$$

with the real filter coefficients $\{w_j\}$ (which need not to add up to one). Therefore, y_t is a linear combination of values of x_{t-j} where we assume constant filters, i.e. the weights do not depend on t. In particular, the **filter** is called causal as y_t is defined by past and contemporaneous values of $\{x_t\}$ only.

The general linear causal filter from (3.5) can be formulated as **polynomial in the lag operator**:

$$F(L) = \sum_{j=0}^{p} w_j\, L^j \quad \text{with} \quad y_t = F(L)\, x_t\,.$$

[3]Many authors also speak of the backshift operator and write B.

Occasionally, two filters are put in a row:

$$y_t = F_1(L)\,x_t\,,\quad z_t = F_2(L)\,y_t = F_2(L)\,F_1(L)\,x_t\,,$$

$$F_1(L) = \sum_{j=0}^{p_1} w_{1,j}\,L^j\,,\quad F_2(L) = \sum_{j=0}^{p_2} w_{2,j}\,L^j\,.$$

Then, the filter $F(L)$ transforming $\{x_t\}$ into $\{z_t\}$ is defined by multiplying the filters, $F(L) = F_2(L)F_1(L)$, which is called a "convolution":

$$F_2(L)\,F_1(L) = \sum_{k=0}^{p_1+p_2} v_k L^k\,,\quad v_k = \sum_{j=0}^{k} w_{2,j} w_{1,k-j} = \sum_{j=0}^{k} w_{2,k-j} w_{1,j}\,.$$

This convolution is commutative:

$$F_1(L)\,F_2(L) = F_2(L)\,F_1(L)\,.$$

By expressing filters by means of the lag operator, we can manipulate them just as ordinary (complex-valued) polynomials. As an example we consider so-called difference filters.

Example 3.2 (Difference Filters) By means of the lag operator, filters can be constructed, for example the difference filter $\Delta = 1 - L$ or the difference of the previous year for quarterly data $\Delta_4 = 1 - L^4$:

$$\Delta x_t = x_t - x_{t-1}\,,\quad \Delta_4 x_t = x_t - x_{t-4}\,.$$

The seasonal difference filter for monthly observations ($S = 12$) as well as for daily observations ($S = 5$) are defined analogously:

$$\Delta_S = 1 - L^S\,.$$

Instead of extensively calculating the double difference,

$$\begin{aligned}\Delta(\Delta x_t) &= \Delta(x_t - x_{t-1}) = (x_t - x_{t-1}) - L(x_t - x_{t-1})\\ &= (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2\,x_{t-1} + x_{t-2}\,,\end{aligned}$$

we write in short by expanding $(1 - L)^2$:

$$\Delta^2 x_t = (1 - L)^2 x_t = (1 - 2\,L + L^2)x_t = x_t - 2\,x_{t-1} + x_{t-2}\,.$$

While the ordinary difference operator Δ eliminates a linear time trend from a time series, the second differences naturally remove a quadratic time trend:

$$\Delta^2(a + bt + ct^2) = c(t^2 - 2(t-1)^2 + (t-2)^2) = 2c.$$

The fact that the order of filtering is exchangeable is well demonstrated by means of the example of seasonal and ordinary differences:

$$\begin{aligned}\Delta\,\Delta_S &= (1-L)\,(1-L^S) = 1 - L - L^S + L^{S+1}\\ &= (1-L^S)\,(1-L) = \Delta_S\,\Delta\,. \quad \blacksquare\end{aligned}$$

Invertibility of Lag Polynomials

We define as a polynomial of degree p (also of order p) in the lag operator

$$P(L) = 1 + b_1 L + \cdots + b_p L^p\,,\ b_p \neq 0\,, \tag{3.6}$$

with the real coefficients b_1 to b_p. For brevity, $P(L)$ is also called lag polynomial.

Consider a first degree polynomial as a special case of (3.6),

$$A_1(L) = 1 - a\,L\,,$$

where the reason for the negative sign will immediately be obvious. When and how can this polynomial be inverted? A **comparison of coefficients** ("method of undetermined coefficients") results in the following series expansion (see Problem 3.4):

$$(1-aL)^{-1} = \frac{1}{A_1(L)} = \sum_{j=0}^{\infty} a^j L^j\,.$$

As is well known, it holds that (infinite geometric series, see Problem 3.2)

$$\sum_{j=0}^{\infty} a^j = \frac{1}{1-a} < \infty \quad \Longleftrightarrow \quad |a| < 1\,. \tag{3.7}$$

Hence, it holds that

$$(1-aL)^{-1} = \frac{1}{A_1(L)} = \sum_{j=0}^{\infty} a^j L^j \quad \text{with} \sum_{j=0}^{\infty} |a^j| = \frac{1}{1-|a|} < \infty$$

if and only if $|a| < 1$. This condition of invertibility is frequently reformulated. In order to do this, we determine the so-called z-transform of the lag polynomial with z being an element of the complex numbers ($z \in \mathbb{C}$): $A_1(z) = 1 - az$. Now, $|a| < 1$

implies that the z-transform $A_1(z) = 1 - az$ exhibits only roots outside the unit circle, i.e. roots being greater than one in absolute value:

$$|a| < 1 \quad \Longleftrightarrow \quad \left[A_1(z) = 0 \Rightarrow |z| = \frac{1}{|a|} > 1\right]. \tag{3.8}$$

Causally Invertible Polynomials

This condition of invertibility for $A_1(L)$ from (3.8) is easily conveyed to a polynomial $P(L)$ of the order p. We say $P(L)$ is **causally invertible** if there exists a power series expansion with non-negative powers and absolutely summable coefficients:

$$(P(L))^{-1} = \frac{1}{P(L)} = \sum_{j=0}^{\infty} \alpha_j L^j \quad \text{with} \sum_{j=0}^{\infty} |\alpha_j| < \infty.$$

The invertibility depends on the z-transform

$$P(z) = 1 + b_1 z + \cdots + b_p z^p, \quad z \in \mathbb{C}, \tag{3.9}$$

or rather on the absolute value of its roots. The following condition of invertibility is adopted from Brockwell and Davis (1991, Thm. 3.1.1), and it is discussed as an exercise (Problem 3.5).

Proposition 3.3

(a) The polynomial $1 - aL$ *is causally invertible,*

$$\frac{1}{1 - aL} = \sum_{j=0}^{\infty} a^j L^j \quad \textit{with} \sum_{j=0}^{\infty} |a^j| < \infty,$$

if and only if $|a| < 1$.

(b) The polynomial $P(L)$ *from (3.6) is causally invertible, i.e. for* $(P(L))^{-1}$ *there exits the absolutely summable series expansion,*

$$(P(L))^{-1} = \frac{1}{P(L)} = \sum_{j=0}^{\infty} \alpha_j L^j \quad \textit{with} \sum_{j=0}^{\infty} |\alpha_j| < \infty,$$

if and only if it holds for all roots of $P(z)$ *that they are greater than one in absolute value:*

$$P(z) = 0 \quad \Rightarrow \quad |z| > 1. \tag{3.10}$$

Example 3.3 (Invertible MA Processes) The MA(1) process, $x_t = \varepsilon_t + b\varepsilon_{t-1}$, can now be formulated alternatively by applying the MA(1) polynomial $B(L) = 1 + b\,L$. Although not required for stationarity, it is usually assumed that $|b| < 1$. What for? This implies for the MA polynomial that:

$$B(z) = 0 \quad \Rightarrow \quad |z| = \frac{1}{|b|} > 1.$$

According to Proposition 3.3 the condition of invertibility is fulfilled and there exists

$$\frac{1}{B(L)} = \frac{1}{1 + b\,L} = \sum_{j=0}^{\infty} \alpha_j\,L^j,$$

where the coefficients $\{\alpha_j\}$ are absolutely summable. The $\{\alpha_j\}$ are obtained explicitly by comparison of coefficients in

$$\begin{aligned} 1 &= (1 + b\,L) \sum_{j=0}^{\infty} \alpha_j\,L^j \\ &= \alpha_0 + \alpha_1\,L + \alpha_2\,L^2 + \alpha_3\,L^3 + \cdots \\ &\quad + b\,(\alpha_0\,L + \alpha_1\,L^2 + \alpha_2\,L^3 + \cdots), \end{aligned}$$

yielding

$$1 = \alpha_0, \quad 0 = \alpha_1 + b\alpha_0, \quad 0 = \alpha_2 + b\alpha_1, \quad \text{etc.},$$

or

$$\alpha_0 = 1,\ \alpha_1 = -b,\ \alpha_2 = b^2,\ \text{and}\ \alpha_j = (-1)^j\,b^j,\ j \geq 0.$$

Hence, the MA(1) process $x_t = B(L)\,\varepsilon_t$ can be reformulated as follows:

$$\varepsilon_t = \frac{x_t}{1 + b\,L} = x_t - b\,x_{t-1} + b^2\,x_{t-2} - b^3\,x_{t-3} \pm \ldots,$$

or

$$\begin{aligned} x_t &= b\,x_{t-1} - b^2\,x_{t-2} + b^3\,x_{t-3} \pm \ldots + \varepsilon_t \\ &= \sum_{j=1}^{\infty} (-1)^{j-1}\,b^j\,x_{t-j} + \varepsilon_t\,, \quad \sum_{j=0}^{\infty} |b^j| < \infty. \end{aligned}$$

In other words: The invertible MA(1) process (for $|b| < 1$) can be expressed (in an absolutely summable manner) as a process depending on its own infinitely many lags. Such a process is called autoregressive (of infinite order). ■

Frequently, one is interested in the special case of a quadratic polynomial, $p = 2$. In this case, the so-called **Schur criterion** provides an equivalent reformulation of the condition of invertibility, rephrasing (3.10) in terms of the polynomial coefficients b_i directly. These have to fulfill three conditions simultaneously. We take the corresponding statement from e.g. Sydsæter, Strøm, and Berck (1999, p. 58).

Corollary 3.1 *For $p = 2$ the polynomial from (3.6),*

$$P(L) = 1 + b_1 L + b_2 L^2 \,,$$

is causally invertible with absolutely summable series expansion $(P(L))^{-1}$ if and only if:

$$\begin{aligned} & (i) && 1 - b_2 && > 0\,, \\ \textit{and } & (ii) && 1 - b_1 + b_2 && > 0\,, \\ \textit{and } & (iii) && 1 + b_1 + b_2 && > 0\,. \end{aligned}$$

Instead of checking $|z_{1,2}| > 1$ for

$$z_{1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2}}{2b_2} \,,$$

it may in practice be simpler to check (i) through (iii) from Corollary 3.1.

3.4 Autoregressive and Mixed Processes

Let $\{x_t\}$ be given by the following stochastic difference equation,

$$x_t = \nu + a_1 x_{t-1} + \cdots + a_p x_{t-p} + \varepsilon_t \,, \quad a_p \neq 0 \,, \quad t \in \mathbb{T} \,,$$

defining an **autoregressive process** of the order p, AR(p). The properties of the general AR process can be illustrated well at the example $p = 1$.

AR(1)

Particularly, let $p = 1$:

$$x_t = \nu + a\, x_{t-1} + \varepsilon_t \,, \quad t \in \mathbb{T} \,. \tag{3.11}$$

When replacing x_{t-1} by this defining equation, one obtains:

$$\begin{aligned} x_t &= \nu + a(\nu + a\,x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \nu + a\,\nu + a^2\,x_{t-2} + a\,\varepsilon_{t-1} + \varepsilon_t . \end{aligned}$$

Hence, x_t and x_{t-2} prove to be correlated. Continued substitution yields

$$x_t = \nu + a\,\nu + a^2\,\nu + a^3\,x_{t-3} + a^2\,\varepsilon_{t-2} + a\,\varepsilon_{t-1} + \varepsilon_t\,,$$

or for any $h \geq 0$:

$$x_t = (1 + a + \ldots + a^{h-1})\,\nu + a^h\,x_{t-h} + a^{h-1}\,\varepsilon_{t-h+1} + \ldots + a\,\varepsilon_{t-1} + \varepsilon_t .$$

Now, let us suppose that the index set $\mathbb{T}$ does not have a lower bound at zero but includes an infinite past, then h can be arbitrarily large and the substitution can be repeated ad infinitum. If it furthermore holds that

$$|a| < 1,$$

then the geometric series yields ($h \to \infty$)

$$1 + a + \ldots + a^{h-1} = \frac{1-a^h}{1-a} \to \frac{1}{1-a}\,,$$

and $a^h\,x_{t-h} \to 0$ in a sense that can be made rigoros, see Brockwell and Davis (1991, p. 71) or Fuller (1996, p. 39). In this manner, it follows for $h \to \infty$ under the aforementioned conditions that:

$$x_t = \frac{\nu}{1-a} + 0 + \sum_{j=0}^{\infty} a^j\,\varepsilon_{t-j} .$$

This way one obtains an infinite MA representation with geometrically decaying coefficients, $c_j = a^j$ in (3.2). In fact, this representation can formally be obtained by inverting $1 - aL$, see Proposition 3.3 (a). The process is therefore stationary with (see Proposition 3.2)

$$\begin{aligned} \mathrm{E}(x_t) &= \mu = \frac{\nu}{1-a}, \\ \mathrm{Var}(x_t) &= \sigma^2 \sum_{j=0}^{\infty} a^{2j} = \frac{\sigma^2}{1-a^2} = \gamma(0) \end{aligned}$$

and

$$\begin{aligned}\text{Cov}(x_t\,, x_{t+h}) &= \sigma^2 \sum_{j=0}^{\infty} a^j a^{j+h} \\ &= a^h\,\gamma(0).\end{aligned}$$

It follows

$$\rho(h) = a^h.$$

Note that these results are obtained for $|a| < 1$. Furthermore, stationarity also depends on the index set. That is to say, if it holds that $\mathbb{T} = \{0, 1, 2, \ldots\}$ then the above-mentioned repeated substitution cannot be performed infinitely often. From

$$x_t = (1 + a + \ldots + a^{t-1})\,\nu + a^t\,x_0 + a^{t-1}\,\varepsilon_1 + \cdots + a\,\varepsilon_{t-1} + \varepsilon_t$$

the expected value follows as time-dependent:

$$\text{E}(x_t) = (1 + a + \ldots + a^{t-1})\,\nu + a^t\text{E}(x_0), \quad t \in \{0, 1, \ldots\}.$$

In particular, this example shows that the stationarity behavior of a process can depend on the index set $\mathbb{T}$. Therefore, in general a stochastic process is not completely characterized without specifying $\mathbb{T}$.

Example 3.4 (AR(1)) Figure 3.2 displays 50 realizations each of AR(1) processes obtained by simulation. However, in the first case $a_1 = a = 0$ such that the graph depicts the realizations of $\{\varepsilon_t\}$. On the right the theoretical autocorrelogram, $\rho(h) = 0$ for $h > 0$, is shown. In the second panel, the case of a positive autocorrelation ($a_1 = a = 0.75$) is illustrated: Positive values tend to be followed by positive values (and vice versa for negative values), such that phases of positive realizations tend to alternate with phases of negative values. The corresponding autocorrelogram shows the geometrically decaying positive autocorrelations up to the order 10. In the last case, there is a negative autocorrelation ($a_1 = a = -0.75$); consistently, negative values tend to be followed by positive ones and vice versa positive values tend to be followed by negative ones. Therefore, the zero line is more often crossed than in the first case of no serial correlation. The corresponding autocorrelogram is alternating: $\rho(h) = |a|^h(-1)^h$ for $a = -0.75$. Qualitatively different patterns of autocorrelation cannot be generated by the simple AR(1) model. Note that the impulse responses or MA(∞) coefficients of the AR(1) model are $c_j = a^j$. Consequently, the cumulated effect defined in (3.3) becomes for $|a| < 1$:

$$CIR = \sum_{j=0}^{\infty} a^j = \frac{1}{1-a}\,.$$

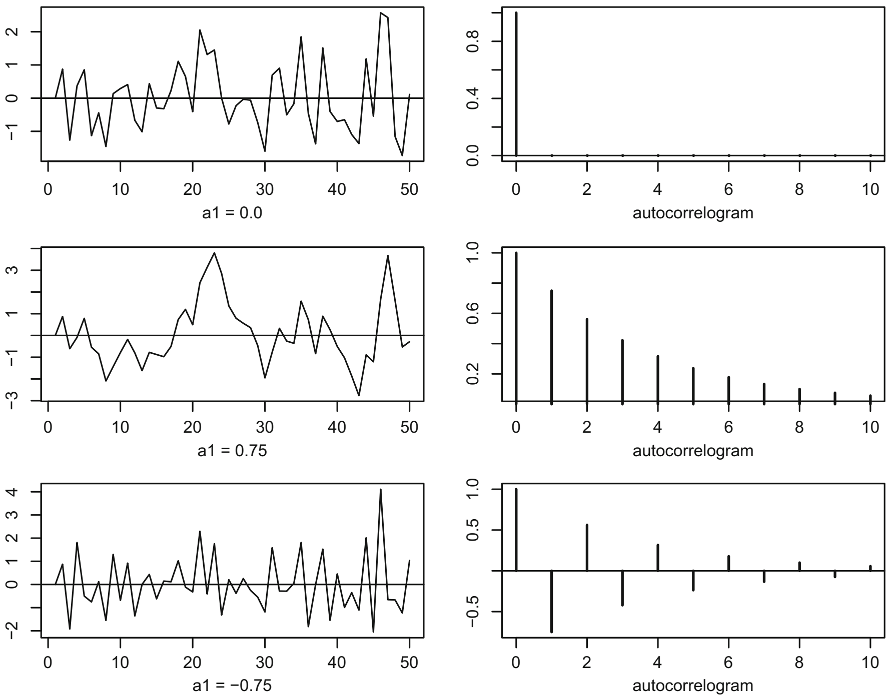

Fig. 3.2 Simulated AR(1) processes with $\sigma^2 = 1$ and $\nu = 0$

The larger a, the larger is *CIR*, which hence quantifies the persistence described above for positive a. ■

We have seen that the stationarity of an AR(1) process depends essentially on the absolute value of a. For the Markov property, however, this value is irrelevant. We talk about a Markov process if the entire past information $\mathcal{I}_t$ up to time t is concentrated in the last observation x_t:

$$\mathrm{P}\,(x_{t+s} \leq x | \mathcal{I}_t) = \mathrm{P}\,(x_{t+s} \leq x | x_t)$$

for all $s > 0$ and $x \in \mathbb{R}$. Assuming a normal distribution, we show in Problem 3.7 that every AR(1) process is a Markov process.

AR(p)

In general, the AR(p) process can be formulated equivalently by means of a lag polynomial:

$$A(L)\, x_t = \nu + \varepsilon_t\,, \quad A(L) = 1 - a_1\, L - \cdots - a_p\, L^p\,, \quad t \in \mathbb{T}\,. \tag{3.12}$$

Under the condition

$$A(z) = 0 \quad \Rightarrow \quad |z| > 1$$

the autoregressive polynomial $A(L)$ is causally invertible with absolutely summable coefficients according to Proposition 3.3:

$$\frac{1}{A(L)} = \sum_{j=0}^{\infty} \alpha_j L^j , \qquad \sum_{j=0}^{\infty} |\alpha_j| < \infty .$$

Therefore, under the condition of invertibility it holds that

$$x_t = \frac{\nu + \varepsilon_t}{A(L)} = \frac{\nu}{A(1)} + \sum_{j=0}^{\infty} \alpha_j \, \varepsilon_{t-j} ,$$

and $\{x_t\}$ is a stationary MA(∞) process, see Proposition 3.2. Hence, the autocovariance sequence is absolutely summable. Furthermore, in this case it holds for $h > 0$ (w.l.o.g.[4] we set $\nu = 0$ for simplification),

$$\begin{aligned}
\gamma(h) &= \mathrm{Cov}(x_t, x_{t+h}) \\
&= \mathrm{E}(x_t \, x_{t+h}) \\
&= \mathrm{E}\left(x_t(a_1 \, x_{t+h-1} + \ldots + a_p \, x_{t+h-p} + \varepsilon_{t+h})\right) \\
&= a_1\gamma(h-1) + \ldots + a_p\gamma(h-p) + 0 .
\end{aligned}$$

Dividing by $\gamma(0)$ yields the recursive relation from the subsequent proposition: The autocorrelations are given by a deterministic difference equation of order p. The still missing *necessary* condition of stationarity from Proposition 3.4 ($A(1) > 0$) will be derived in Problem 3.6.

Proposition 3.4 (AR(p)) *Let $\{x_t\}$ be an AR(p) process from (3.12) with index set*[5] $\mathbb{T} = \{-\infty, \ldots, T\}$.

[4] The abbreviation stands for "without loss of generality". It is frequently used for assumptions that are substantially not necessary and that are only made to simplify the argument or the notation. In the example at hand, generally it would have to be written $(x_{t-j} - \mu)$ for all j; just as well, one can set $\mu = \nu/A(1)$ equal to zero and simply write x_{t-j}.

[5] Also in the following, the notation $\{-\infty, \ldots, T\}$ is always to denote the set of all integers without $\{T+1, T+2, \ldots\}$.

(a) *The process has an absolutely summable MA(∞) representation according to (3.2) if and only if it holds that*

$$A(z) = 0 \quad \Rightarrow \quad |z| > 1\,.$$

Then the process is stationary with expectation $\mu = \nu/A(1)$. The condition $A(1) = 1 - \sum_{j=1}^{p} a_j > 0$ is necessary for this.

(b) *For stationary processes the autocorrelation sequence is absolutely summable where it holds that $\rho(h) \neq 0$ for all integer numbers h and*

$$\rho(h) = a_1\,\rho(h-1) + \ldots + a_p\,\rho(h-p)\,, \quad h > 0\,.$$

Again note that the absolute summability of $\rho(h)$ implies: $\rho(h) \to 0$ for $h \to \infty$. The farther x_t and x_{t+h} are apart from each other the weaker tends to be their correlation.

Certain properties of the AR(1) process are lost for $p > 1$. In Problem 3.8 we show for a special case ($p = 2$) that the AR(p) process, $p > 1$, is not a Markov process in general.

AR(2)

Let $p = 2$,

$$x_t = \nu + a_1\,x_{t-1} + a_2\,x_{t-2} + \varepsilon_t\,,$$

with the autoregressive polynomial

$$A(L) = 1 - a_1 L - a_2 L^2\,.$$

From Corollary 3.1, we know the conditions under which $(A(L))^{-1}$ can be expanded as an absolutely summable filter:

$$\begin{array}{rll} (i) & 1 + a_2 & > 0\,, \\ \text{and } (ii) & 1 + a_1 - a_2 & > 0\,, \\ \text{and } (iii) & 1 - a_1 - a_2 & > 0\,. \end{array}$$

Consequently, under these three parameter restrictions the AR(2) process is stationary. The restrictions become even more obvious if they are solved for a_2 and depicted in a coordinate system:

$$\begin{array}{rll} (i) & a_2 > & -1\,, \\ \text{and } (ii) & a_2 < & 1 + a_1\,, \\ \text{and } (iii) & a_2 < & 1 - a_1\,. \end{array}$$

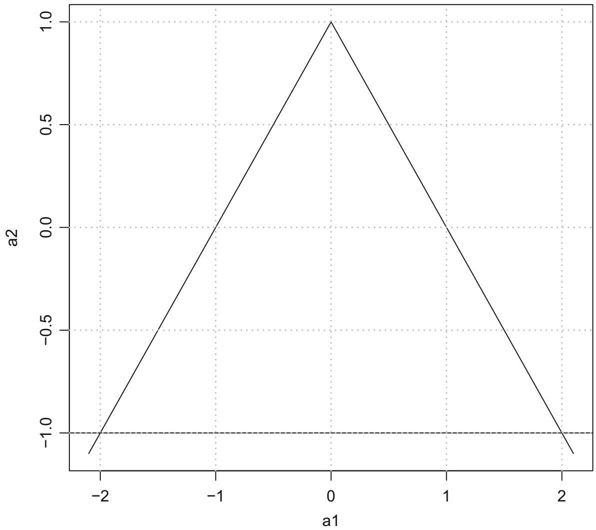

Fig. 3.3 Stationarity triangle for AR(2) processes

So, three lines are given, above or below which a_2 has to be, respectively. This is why one also talks about the stability or stationarity triangle, see Fig. 3.3. Within the triangle lies the stationarity region.

The autocorrelation series of the AR(2) case is determined from Proposition 3.4. For $h = 0$, it naturally holds that $\rho(0) = 1$. For $h = 1$, one obtains

$$\rho(1) = a_1\, \rho(0) + a_2\, \rho(-1).$$

Because of the symmetry, $\rho(-h) = \rho(h)$, it follows that

$$\rho(1) = \frac{a_1}{1 - a_2}.$$

Similarly, it follows that

$$\begin{aligned} \rho(2) &= a_1\, \rho(1) + a_2\, \rho(0) \\ &= \frac{a_1^2}{1 - a_2} + a_2. \end{aligned}$$

By repeated insertion into the second order difference equation,

$$\rho(h) = a_1\, \rho(h-1) + a_2\, \rho(h-2)\,, \quad h \geq 2\,,$$

the entire autocorrelation sequence is determined. Next, four numerical examples will be considered.

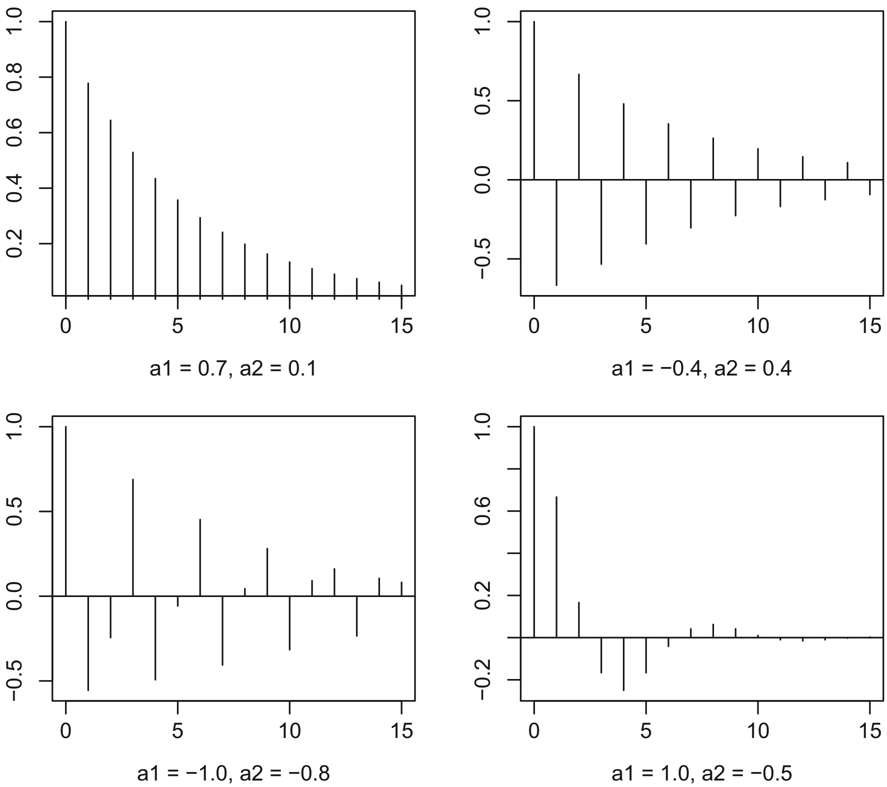

Fig. 3.4 Autocorrelograms for AR(2) processes

Example 3.5 (AR(2)) We now consider four numerical examples for AR(2) processes in order to characterize typical patterns of autocorrelation. In all cases, the stability conditions can be proven to be fulfilled. The corresponding autocorrelograms can be found in Fig. 3.4.

(i) $a_1 = 0.7$, $a_2 = 0.1$: In this case, all the autocorrelations are positive and they converge to zero with h; their behavior is similar to the autocorrelogram of a AR(1) process with $a_1 > 0$, see Fig. 3.2.
(ii) $a_1 = -0.4$, $a_2 = 0.4$: Starting with $\rho(1) < 0$, the autocorrelations alternate similarly to an AR(1) process with $a_1 < 0$, cf. Fig. 3.2.
(iii) $a_1 = -1.0$, $a_2 = -0.8$: In this case, we find a dynamic which cannot be generated by an AR(1) model; two negative autocorrelations of the first and second order are followed by a seemingly irregularly alternating pattern.
(iv) $a_1 = 1.0$, $a_2 = -0.5$: In the last case, the autocorrelations swing from the positive area to the negative area, then to the positive one and again to the negative one whereas the last-mentioned can hardly be perceived because of the small absolute values.

Therefore, the cases (iii) and (iv) show that the AR(2) process allows for richer dynamics and dependence structures than the simpler AR(1) process. ■

Autoregressive Moving Average Processes

Now, we consider a combination of AR and MA processes. Again, $\{x_t\}$ is given by a stochastic difference equation of order p, only that ε_t from (3.12) is replaced by an MA(q) process:

$$x_t = \nu + a_1 x_{t-1} + \cdots + a_p x_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}\,, \quad t \in \mathbb{T}\,.$$

Abbreviating, we talk about **ARMA(p, q) processes** assuming $a_p \neq 0$ and $b_q \neq 0$. Again, a more compact representation follows by using lag polynomials,

$$A(L)\, x_t = \nu + B(L)\, \varepsilon_t\,, \quad t \in \mathbb{T}\,, \tag{3.13}$$

where it is assumed that both polynomials

$$A(L) = 1 - a_1 L - \cdots - a_p L^p \;\text{ and }\; B(L) = 1 + b_1 L + \cdots + b_q L^q$$

do not have common roots.

A stationary MA(∞) representation hinges on the autoregressive polynomial such that the stationarity condition can be adopted from the pure AR(p) case in Proposition 3.4. It amounts to an absolutely summable expansion of $(A(L))^{-1}$ such that the process possesses an absolutely summable representation as MA(∞) process, see (3.2). If stationarity is given, the absolutely summable autocorrelation sequence can again be determined from a stable difference equation (see e.g. Brockwell & Davis, 1991, p. 93)

Proposition 3.5 (ARMA(p, q)) *Let $\{x_t\}$ be an ARMA(p, q) process from (3.13) with $\mathbb{T} = \{-\infty, \ldots, T\}$.*

(a) The process has an absolutely summable MA(∞) representation according to (3.2) if and only if it holds that

$$A(z) = 0 \quad \Rightarrow \quad |z| > 1\,.$$

Then the process is stationary with expectation $\mu = \nu/A(1)$. The condition $A(1) = 1 - \sum_{j=1}^{p} a_j > 0$ is necessary for this.

(b) For stationary processes the autocorrelation sequence is absolutely summable where it holds that $\rho(h) \neq 0$ for all integer numbers h and

$$\rho(h) = a_1\, \rho(h-1) + \ldots + a_p\, \rho(h-p)\,, \quad h \geq \max(p, q+1)\,.$$

For $h \geq \max(p, q + 1)$ the autocorrelations satisfy the same difference equation as in the pure AR(p) case. It is known that the solution to such a difference equation is bounded exponentially. Consequently, we have the following result, see e.g. Brockwell and Davis (1991, Prob. 3.11): For a stationary ARMA process there exist positive constants c and g with $0 < g < 1$, such that

$$|\rho(h)| \leq c\, g^h \tag{3.14}$$

for all $h = 1, 2, \ldots$. Hence, the decay rate of the autocorrelations is bounded exponentially. This shows again that stationary ARMA process are characterized be absolutely summable autocorrelations, see Problem 3.2. Proposition 3.5 will be discussed below for $p = q = 1$.

Before turning to the ARMA(1,1) case, we note that an analogous result to Proposition 3.5 (a) is available for an AR(∞) representation according to Proposition 3.3: The ARMA process has an absolutely summable AR(∞) representation (see Example 3.3) if and only if

$$B(z) = 0 \quad \Rightarrow \quad |z| > 1\,.$$

In this case the ARMA process is called invertible.

ARMA(1,1)

We now wish to obtain the autocorrelation structure of the ARMA(1,1) process,

$$x_t = a\, x_{t-1} + \varepsilon_t + b\, \varepsilon_{t-1}, \quad |a| < 1,\ |b| < 1\,,$$

where $|a| < 1$ for stationarity, and $|b| < 1$ to ensure invertibility (see Example 3.3). The condition of no common roots of $1 - aL$ and $1 + bL$ is given for $a \neq -b$. In the case of common roots, the lag polynomials could be reduced and one would obtain

$$x_t = \frac{1 + bL}{1 + bL}\, \varepsilon_t = \varepsilon_t \quad \text{for} \quad a = -b\,.$$

Due to the invertibility of $1 - aL$, the process can be formulated as an infinite MA process,

$$\begin{aligned} x_t &= \frac{\varepsilon_t + b\, \varepsilon_{t-1}}{1 - aL} \\ &= \sum_{j=0}^{\infty} a^j\, \varepsilon_{t-j} + b \sum_{j=0}^{\infty} a^j\, \varepsilon_{t-1-j} \\ &= \varepsilon_t + \sum_{j=1}^{\infty} \left(a^j + b\, a^{j-1}\right)\, \varepsilon_{t-j}\,, \end{aligned}$$

where a shift of subscripts was carried out. In the notation of (3.2) the MA(∞) coefficients result as

$$c_j = a^{j-1}(a+b)\,, \quad j \geq 1\,.$$

In this way Proposition 3.2 yields for the variance:

$$\gamma(0) = \sigma^2 \left(1 + \sum_{j=1}^{\infty} a^{2j-2}\,(a+b)^2\right) = \ldots = \sigma^2\,\frac{(1+b^2+2ab)}{1-a^2}\,.$$

The autocovariance at lag one follows in the same way,

$$\gamma(1) = \sigma^2\frac{(a+b)(1+ab)}{1-a^2},$$

such that it holds:

$$\rho(1) = \frac{(a+b)(1+ab)}{1+b^2+2ab}.$$

Furthermore we learn from Proposition 3.5:

$$\rho(h) = a\,\rho(h-1), \quad h \geq 2.$$

Hence, the MA(1) component (that is b) influences directly only $\rho(1)$ having only an indirect effect beyond the autocorrelation of the first order: For $h \geq 2$ a recursive relation between $\rho(h)$ and $\rho(h-1)$ holds true just as it applies to the pure AR(1) process. Thus, this yields four typical patterns. In order to identify these, it suffices entirely to concentrate on the numerator of $\rho(1)$ as well as on the sign of a as the denominator of $\rho(1)$ is always positive because it is a multiple of the variance. As $1+ab$ is positive due to the stationarity and the invertibility of the MA polynomial, the behavior of the autocorrelogram depends on the signs of $a+b$ and a only. The exponential bound for the autocorrelations is easily verified:

$$\rho(h) = a\,\rho(h-1) = \cdots = a^{h-1}\rho(1)\,. \quad h \geq 2.$$

Therefore, (3.14) applies with $g = |a|$ and $c = |\rho(1)|/|a|$.

Example 3.6 (ARMA(1,1)) The four possible patterns of the autocorrelogram of a stationary and invertible ARMA(1,1) model will be discussed and illustrated by numerical examples, cf. Fig. 3.5.

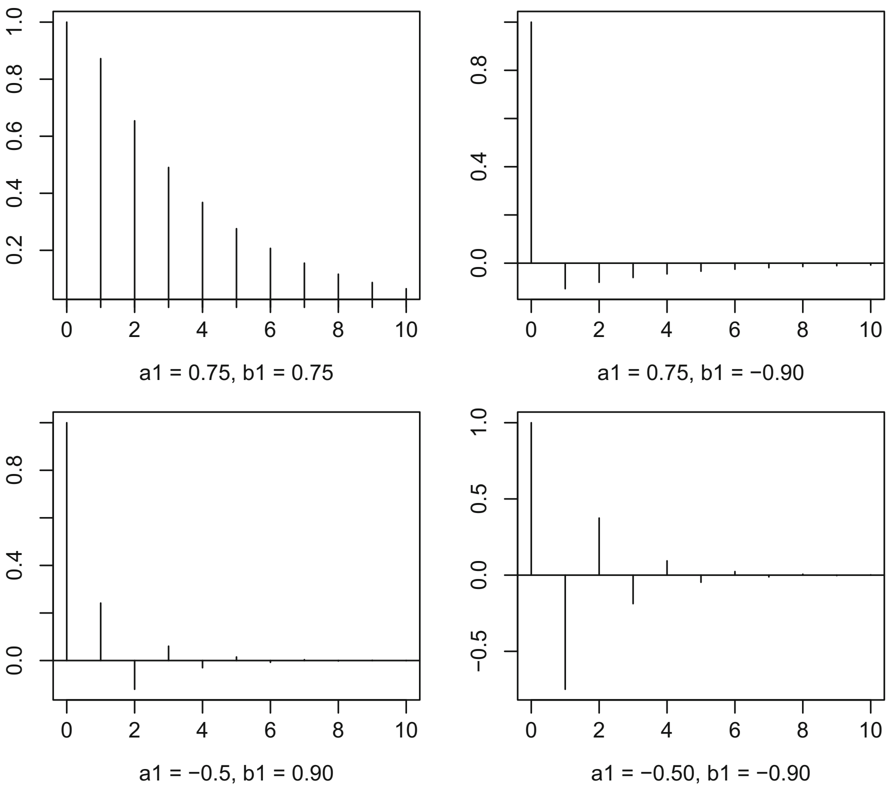

Fig. 3.5 Autocorrelograms for ARMA(1,1) processes

Case 1: Obviously, for $a + b > 0$ it holds that $\rho(1) > 0$. If, furthermore, $a > 0$ then the entire autocorrelogram proceeds above the zero line.

Case 2: For $a + b < 0$ and $a > 0$ one obtains exclusively negative autocorrelations.

Case 3: An alternating pattern, starting with $\rho(1) > 0$, is obtained for $a < 0$ and $a > -b$.

Case 4: For $a < 0$ and $a < -b$ the autocorrelation series is as well alternating but starts with a negative value.

Note that cases 1 and 4 can be generated qualitatively by a pure AR(1) process as well. For cases 2 and 3, however, there occur patterns which cannot be produced by an AR(1) process. Therefore, the ARMA(1,1) process allows for richer dynamic modeling than the AR(1) model does. Also when comparing with the AR(2) case, we find that the ARMA(1,1) model allows for additional dynamics. ■

3.5 Problems and Solutions

Problems

3.1 Where does the first order autocorrelation of an MA(1) process ($\rho(1) = \frac{b}{1+b^2}$) have its maximum and minimum?

3.2 Show for $g \in \mathbb{R} \setminus \{1\}$ (geometric series):

$$\sum_{i=0}^{n} g^i = \frac{1-g^{n+1}}{1-g}.$$

Conclusion: For $|g| < 1$ and $n \to \infty$ it holds that:

$$\sum_{i=0}^{\infty} g^i = \frac{1}{1-g}.$$

3.3 Prove part (b) from Proposition 3.2.

3.4 Derive the series expansion

$$(1 - aL)^{-1} = \sum_{j=0}^{\infty} a^j L^j$$

for real a with $|a| < 1$.

3.5 Prove Proposition 3.3 (b).

3.6 Prove the necessary condition of causal invertibility from Proposition 3.4, that is:

$$\{A(z) = 0 \Rightarrow |z| > 1\} \implies \{A(1) > 0\} \,,$$

where $A(z) = 1 - a_1 z - \ldots - a_p z^p$.

3.7 Let $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$ be a Gaussian process. Show that $\{x_t\}$ with $x_t = a_1 x_{t-1} + \varepsilon_t$ is a Markov process.

3.8 Let $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$. Show that $\{x_t\}$ with $x_t = a_2 x_{t-2} + \varepsilon_t$ is not a Markov process ($a_2 \neq 0$).

Solutions

3.1 The traditional way to solve the problem is curve sketching. We consider the first order autocorrelation to be a function of b:

$$f(b) = \rho(1) = \frac{b}{1+b^2}.$$

Then the quotient rule for the first order derivative yields

$$f'(b) = \frac{1+b^2-2b^2}{(1+b^2)^2} = \frac{(1-b)(1+b)}{(1+b^2)^2}.$$

The roots of the derivative are given by $|b| = 1$. In $b = -1$ there is a change of sign of $f'(b)$, namely from a negative to a positive slope. Hence, in $b = -1$ there is a relative (and also an absolute) minimum. Because of $f(b)$ being an odd function (symmetric about the origin), there is a maximum in $b = 1$. Therefore, the maximum possible correlation in absolute value is

$$|f(-1)| = f(1) = \frac{1}{2}.$$

One may also tackle the problem by more elementary means. Note that for $b \neq 0$

$$\frac{1}{|b|} - 2 + |b| = \left(\frac{1}{\sqrt{|b|}} - \sqrt{|b|}\right)^2 \geq 0\,,$$

which is equivalent to

$$\frac{1}{2} \geq \frac{1}{\frac{1}{|b|} + |b|} = \frac{|b|}{1+b^2} = |f(b)|\ ,$$

with $f(b) = \rho(1)$ defined above. Since $|f(-1)| = f(1) = \frac{1}{2}$, this solves the problem.

3.2 By S_n we denote the following sum for finite n:

$$S_n = \sum_{i=0}^{n} g^i = 1 + g + \ldots + g^{n-1} + g^n.$$

Multiplication by g yields

$$g\,S_n = g + g^2 + \ldots + g^n + g^{n+1}.$$

Therefore, it holds that

$$S_n - g\,S_n = 1 - g^{n+1}.$$

By ordinary factorization the formula

$$S_n = \frac{1 - g^{n+1}}{1 - g}$$

and therefore the claim is verified.

3.3 The absolute summability of $\gamma(h)$ follows from the absolute summability of the linear coefficients $\{c_j\}$ allowing for a change of the order of summation. In order to do so, we first apply the triangle inequality:

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{h=0}^{\infty}|\gamma(h)| &= \sum_{h=0}^{\infty}\left|\sum_{j=0}^{\infty} c_j\, c_{j+h}\right| \\
&\leq \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\left|c_j\, c_{j+h}\right| = \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\left|c_j\right|\,\left|c_{j+h}\right| \\
&= \sum_{j=0}^{\infty}\left|c_j\right|\left(\sum_{h=0}^{\infty}\left|c_{j+h}\right|\right),
\end{aligned}$$

where at the end round brackets were placed for reasons of clarity. The final term is further bounded by enlarging the expression in brackets:

$$\sum_{j=0}^{\infty}\left|c_j\right|\left(\sum_{h=0}^{\infty}\left|c_{j+h}\right|\right) \leq \sum_{j=0}^{\infty}\left|c_j\right|\left(\sum_{h=0}^{\infty}\left|c_h\right|\right).$$

Therefore, the claim follows indeed from the absolute summability of $\{c_j\}$.

3.4 For the proof we denote $(1 - aL)^{-1}$ as $\sum_{j=0}^{\infty}\alpha_j L^j$,

$$\frac{1}{1 - aL} = \sum_{j=0}^{\infty}\alpha_j L^j\,,$$

and determine the coefficients α_j. By multiplying this equation with $1 - aL$, we obtain

$$\begin{aligned}1 &= (1 - aL) \sum_{j=0}^{\infty} \alpha_j L^j \\ &= \alpha_0 + \alpha_1 L^1 + \alpha_2 L^2 + \dots \\ &\qquad - a\,\alpha_0 L^1 - a\,\alpha_1 L^2 - a\,\alpha_2 L^3 - \dots .\end{aligned}$$

Now, we compare the coefficients associated with L^j on the left- and on the right-hand side:

$$\begin{aligned}1 &= \alpha_0\,, \\ 0 &= \alpha_1 - a\,\alpha_0\,, \\ 0 &= \alpha_2 - a\,\alpha_1\,, \\ &\vdots \\ 0 &= \alpha_j - a\,\alpha_{j-1}\,, \quad j \geq 1\,.\end{aligned}$$

As claimed, the solution of the difference equation obtained in this way, ($\alpha_j = a\,\alpha_{j-1}$), is obviously $\alpha_j = a^j$.

3.5 We factorize $P(z) = 1 + b_1 z + \ldots + b_p z^p$ with roots $z_1, \ldots, z_p$ of this polynomial (fundamental theorem of algebra):

$$P(z) = b_p\,(z - z_1) \ldots (z - z_p)\,.$$

From each bracket we factorize $-z_j$ out such that

$$P(z) = b_p\,(-1)^p\,z_1\,\ldots z_p \left(1 - \frac{z}{z_1}\right) \ldots \left(1 - \frac{z}{z_p}\right)\,.$$

Because of $P(0) = 1$, we obtain $b_p\,(-1)^p\,z_1\,\ldots z_p = 1$. Therefore the factorization simplifies to

$$\begin{aligned}P(z) &= \left(1 - \frac{z}{z_1}\right) \ldots \left(1 - \frac{z}{z_p}\right) \\ &= P_1(z) \cdots P_p(z)\,,\end{aligned}$$

with

$$P_k(z) = 1 - \frac{z}{z_k} = 1 - \zeta_k z\,, \quad k = 1, \ldots, p\,,$$

where $\zeta_k = 1/z_k$. From part a) we know that

$$\frac{1}{P_k(L)} = \sum_{j=0}^{\infty} \zeta_k^j L^j \quad \text{with} \sum_{j=0}^{\infty} |\zeta_k^j| < \infty$$

if and only if

$$|z_k| = \frac{1}{|\zeta_k|} > 1 .$$

Now, consider the convolution (sometimes called Cauchy product) for $k \neq \ell$:

$$\frac{1}{P_k(L)} \frac{1}{P_\ell(L)} = \sum_{j=0}^{\infty} c_j L^j$$

with

$$c_j := \sum_{i=0}^{j} \zeta_k^i \zeta_\ell^{j-i} .$$

We have $\sum_{j=0}^{\infty} |c_j| < \infty$ if and only if both $P_k^{-1}(L)$ and $P_\ell^{-1}(L)$ are absolutely summable, which holds true if and only if

$$|z_k| > 1 \quad \text{and} \quad |z_\ell| > 1 .$$

Repeating this argument we obtain that

$$\frac{1}{P(L)} = \frac{1}{P_1(L)} \cdots \frac{1}{P_p(L)} = \sum_{j=0}^{\infty} c_j L^j \quad \text{with} \sum_{j=0}^{\infty} |c_j| < \infty$$

if and only if (3.10) holds. Quod erat demonstrandum.

3.6 At first we reformulate the autoregressive polynomial $A(z) = 1 - a_1 z - \ldots - a_p z^p$ in its factorized form with roots $z_1, \ldots, z_p$ (again by the fundamental theorem of algebra):

$$A(z) = -a_p (z - z_1) \ldots (z - z_p) .$$

For $z = 1$ this amounts to

$$A(1) = -a_p (1 - z_1) \ldots (1 - z_p) . \tag{3.15}$$

Because of $A(0) = 1$ we obtain as well:

$$1 = -a_p(-1)^p z_1 \, \ldots \, z_p \,. \tag{3.16}$$

Now we proceed in two steps, treating the cases of complex and real roots separately.

(A) Complex roots: Note that for a root $z_1 \in \mathbb{C}$ it holds that the complex conjugate, $z_2 = \overline{z_1}$, is a root as well. Then calculating with complex numbers yields for the product

$$\begin{aligned}(1 - z_1)(1 - z_2) &= (1 - z_1)(1 - \overline{z_1}) \\ &= (1 - z_1)(\overline{1 - z_1}) \\ &= |1 - z_1|^2 > 0\,.\end{aligned}$$

Hence, for $p > 2$, complex roots contribute positively to $A(1)$ in (3.15). If $p = 2$, the roots are only complex if $a_2 < 0$, since the discriminant is $a_1^2 + 4a_2$; hence, $A(1) > 0$ by (3.15).

(B) Since the effect of complex roots is positive, we now concentrate on real roots z_i, for which it holds that $|z_i| > 1$ by assumption. So, we assume without loss of generality that the polynomial has no complex roots, or that all complex roots have been factored out. Two sub-cases have to be distinguished. (1) Even degree: For an even p we again distinguish between two cases. Case 1, $a_p > 0$: Because of (3.16) there has to be an odd number of negative roots and therefore there has to be an odd number of positive roots as well. For the latter it holds that $(1 - z_i) < 0$ while the first naturally fulfill $(1 - z_i) > 0$. Hence, as claimed, it follows from (3.15) that $A(1)$ is positive. Case 2, $a_p < 0$: In this case one argues quite analogously. Because of (3.16) there is an even number of positive and negative roots such that the requested claim follows from (3.15) as well. (2) Odd degree: For an odd p one obtains the requested result as well by distinction of the two cases for the sign of a_p. We omit details.

Hence, the proof is complete.

3.7 The normality of $\{\varepsilon_t\}$ implies a multivariate Gaussian distribution of

$$\begin{pmatrix} \varepsilon_{t+1} \\ \vdots \\ \varepsilon_{t+s} \end{pmatrix} \sim \text{ii}\,\mathcal{N}_s\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 I_s\right)$$

with the identity matrix I_s of dimension s. The s-fold substitution yields

$$\begin{aligned}x_{t+s} &= a_1^s x_t + a_1^{s-1}\varepsilon_{t+1} + \ldots + a_1\varepsilon_{t+s-1} + \varepsilon_{t+s} \\ &= a_1^s x_t + \sum_{i=0}^{s-1} a_1^i \varepsilon_{t+s-i}.\end{aligned}$$

The sum over the white noise process has the moments

$$\mathrm{E}\left(\sum_{i=0}^{s-1} a_1^i \varepsilon_{t+s-i}\right) = 0, \quad \mathrm{Var}\left(\sum_{i=0}^{s-1} a_1^i \varepsilon_{t+s-i}\right) = \sigma^2 \sum_{i=0}^{s-1} a_1^{2i},$$

and, furthermore, it is normally distributed:

$$\sum_{i=0}^{s-1} a_1^i \varepsilon_{t+s-i} \sim \mathcal{N}\left(0, \, \sigma^2 \sum_{i=0}^{s-1} a_1^{2i}\right).$$

Hence, x_{t+s} given x_t follows a Gaussian distribution with the corresponding moments:

$$x_{t+s} | \, x_t \sim \mathcal{N}\left(a_1^s x_t, \, \sigma^2 \sum_{i=0}^{s-1} a_1^{2i}\right).$$

As x_{t+s} can be expressed as a function of x_t and $\varepsilon_{t+1}, \ldots, \varepsilon_{t+s}$ alone, the further past of the process does not matter for the conditional distribution of x_{t+s}. Therefore, for the entire information $\mathcal{I}_t$ up to time t it holds that:

$$x_{t+s} | \, \mathcal{I}_t \sim \mathcal{N}\left(a_1^s x_t, \, \sigma^2 \sum_{i=0}^{s-1} a_1^{2i}\right).$$

Hence, the Markov property (2.9) has been shown. It holds independently of the concrete value of a_1.

3.8 For

$$x_t = a_2 x_{t-2} + \varepsilon_t,$$

we obtain for $s = 1$ the conditional expectations $\mathrm{E}(x_{t+1} | \, \mathcal{I}_t) = a_2 \, x_{t-1}$, and

$$\mathrm{E}(x_{t+1} | \, x_t) = \mathrm{E}(a_2 x_{t-1} + \varepsilon_{t+1} | \, x_t) = a_2 \, \mathrm{E}(x_{t-1} | \, x_t),$$

with $\mathcal{I}_t = \sigma(x_t, x_{t-1}, \ldots, x_1)$. As the conditional expectations are not equivalent, the conditional distributions are not the same. Hence, it generally holds that

$$\mathrm{P}(x_{t+1} \leq x | \, x_t) \neq \mathrm{P}(x_{t+1} \leq x | \, \mathcal{I}_t),$$

which proves that $\{x_t\}$ is not a Markov process.

References

Andrews, D. W. K., & Chen, H.-Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics, 12*, 187–204.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). New York: Springer.

Campbell, J. Y., & Mankiw, N. G. (1987). Are output fluctuations transitory? *Quarterly Journal of Economics, 102*, 857–880.

Fuller, W. A. (1996). *Introduction to statistical time series* (2nd ed.). New York: Wiley.

Sydsæter, K., Strøm, A., & Berck, P. (1999). *Economists' mathematical manual* (3rd ed.). Berlin/New York: Springer.

Wold, H. O. A. (1938). *A study in the analysis of stationary time series*. Stockholm: Almquist & Wiksell.