2 Univariate Stationary Processes

As mentioned in the introduction, the publication of the textbook by GEORGE E.P. BOX and GWILYM M. JENKINS in 1970 opened a new road to the analysis of economic time series. This chapter presents the Box-Jenkins Approach, its different models and their basic properties in a rather elementary and heuristic way. These models have become an indispensable tool for short-run forecasts. We first present the most important approaches for statistical modelling of time series. These are autoregressive (AR) processes (*Section 2.1*) and moving average (MA) processes (*Section 2.2*), as well as a combination of both types, the so-called ARMA processes (*Section 2.3*). In *Section 2.4* we show how this class of models can be used for predicting the future development of a time series in an optimal way. Finally, we conclude this chapter with some remarks on the relation between the univariate time series models described in this chapter and the simultaneous equations systems of traditional econometrics (*Section 2.5*).

2.1 Autoregressive Processes

We know autoregressive processes from traditional econometrics: Already in 1949, DONALD COCHRANE and GUY H. ORCUTT used the first order autoregressive process for modelling the residuals of a regression equation. We will start with this process, then treat the second order autoregressive process and finally show some properties of autoregressive processes of an arbitrary but finite order.

2.1.1 First Order Autoregressive Processes

Derivation of Wold's Representation

A *first order autoregressive process*, an AR(1) process, can be written as an inhomogeneous stochastic first order difference equation,

$$x_t = \delta + \alpha x_{t-1} + u_t, \tag{2.1}$$

G. Kirchgässner et al., *Introduction to Modern Time Series Analysis*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-33436-8_2, © Springer-Verlag Berlin Heidelberg 2013

where the inhomogeneous part $\delta + u_t$ consists of a constant term δ and a pure random process u_t. Let us assume that for $t = t_0$ the initial value x_{t_0} is given. By successive substitution in (2.1) we get

$$\begin{aligned}
x_{t_0+1} &= \delta + \alpha x_{t_0} + u_{t_0+1} \\
x_{t_0+2} &= \delta + \alpha x_{t_0+1} + u_{t_0+2} \\
&= \delta + \alpha(\delta + \alpha x_{t_0} + u_{t_0+1}) + u_{t_0+2} \\
&= \delta + \alpha\delta + \alpha^2 x_{t_0} + \alpha u_{t_0+1} + u_{t_0+2} \\
x_{t_0+3} &= \delta + \alpha x_{t_0+2} + u_{t_0+3} \\
x_{t_0+3} &= \delta + \alpha\delta + \alpha^2\delta + \alpha^3 x_{t_0} + \alpha^2 u_{t_0+1} + \alpha u_{t_0+2} + u_{t_0+3} \\
&\vdots \\
x_{t_0+\tau} &= (1 + \alpha + \alpha^2 + \ldots + \alpha^{\tau-1})\delta + \alpha^\tau x_{t_0} \\
&\quad + \alpha^{\tau-1} u_{t_0+1} + \alpha^{\tau-2} u_{t_0+2} + \ldots + \alpha u_{t_0+\tau-1} + u_{t_0+\tau},
\end{aligned}$$

or

$$x_{t_0+\tau} = \alpha^\tau x_{t_0} + \frac{1-\alpha^\tau}{1-\alpha}\delta + \sum_{j=0}^{\tau-1} \alpha^j u_{t_0+\tau-j} .$$

For $t = t_0 + \tau$, we get

$$x_t = \alpha^{t-t_0} x_{t_0} + \frac{1-\alpha^{t-t_0}}{1-\alpha}\delta + \sum_{j=0}^{t-t_0-1} \alpha^j u_{t-j} . \tag{2.2}$$

The development and thus the properties of this process are mainly determined by the assumptions on the initial condition x_{t_0}.

The case of a *fixed (deterministic) initial condition* is given if x_0 is assumed to be a fixed (real) number, for example for $t_0 = 0$, i.e. no random variable. Then we can write:

$$x_t = \alpha^t x_0 + \frac{1-\alpha^t}{1-\alpha}\delta + \sum_{j=0}^{t-1} \alpha^j u_{t-j} .$$

This process consists of time dependent deterministic and stochastic parts. Thus, it can never be weakly stationary, since first and second order mo-

ments are time dependent. It is, however, *asymptotically stationary* because the time dependence vanishes for $t_0 \to -\infty$.

We can imagine the case of *stochastic initial conditions* as (2.1) being generated along the whole time axis, i.e. $-\infty < t < \infty$. If we observe the process only for positive values of t, the initial value x_0 is a random variable which is generated by this process. Formally, the process with stochastic initial conditions results from (2.2) if the solution of the homogeneous difference equation has disappeared. This is only possible if $|\alpha| < 1$. Therefore, in the following, we restrict α to the interval $-1 < \alpha < 1$. If $\lim_{t_0 \to -\infty} x_{t_0}$ is bounded, (2.2) for $t_0 \to -\infty$ converges to

$$x_t = \frac{\delta}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j} . \tag{2.3}$$

The time dependence has disappeared. According to *Section 1.5*, the AR(1) process (2.1) has the Wold representation (2.3) with $\psi_j = \alpha^j$ and $|\alpha| < 1$. This results in the convergence of

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \alpha^{2j} = \frac{1}{1-\alpha^2} .$$

Thus, assuming stochastic initial conditions, the process (2.1) is weakly stationary.

The Lag Operator

Equation (2.3) can also be derived from relation (2.1) by using the lag operator defined in *Section 1.3*:

$$(1-\alpha L)x_t = \delta + u_t . \tag{2.1'}$$

If we solve for x_t we get

$$x_t = \frac{\delta}{1-\alpha L} + \frac{1}{1-\alpha L} u_t . \tag{2.4}$$

The expression $1/(1-\alpha L)$ can formally be expanded to a geometric series,

$$\frac{1}{1-\alpha L} = 1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots .$$

Thus, we get

$$\begin{aligned} x_t &= (1 + \alpha L + \alpha^2 L + \dots)\delta + (1 + \alpha L + \alpha^2 L + \dots)u_t \\ &= (1 + \alpha + \alpha^2 + \dots)\delta + u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + \dots , \end{aligned}$$

and because of $|\alpha| < 1$

$$x_t = \frac{\delta}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j} .$$

The first term could have been derived immediately if we substituted the value '1' for L in the first term of (2.4). (See also relation (1.8) on p. 11).

Calculation of Moments

Due to representation (2.3), the first and second order moments can be calculated. As $E[u_t] = 0$ holds for all t, we get for the mean

$$E[x_t] = E\left[\frac{\delta}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j u_{t-j}\right]$$

$$E[x_t] = \frac{\delta}{1-\alpha} + \sum_{j=0}^{\infty} \alpha^j E\left[u_{t-j}\right] = \frac{\delta}{1-\alpha} = \mu$$

i.e. the mean is constant. It is different from zero if and only if $\delta \neq 0$. Because of $1 - \alpha > 0$, the sign of the mean is determined by the sign of δ. For the variance we get

$$V[x_t] = E\left[\left(x_t - \frac{\delta}{1-\alpha}\right)^2\right] = E\left[\left(\sum_{j=0}^{\infty} \alpha^j u_{t-j}\right)^2\right]$$

$$= E[(u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + \ldots)^2]$$

$$= E[u_t^2 + \alpha^2 u_{t-1}^2 + \alpha^4 u_{t-2}^2 + \ldots + 2\alpha u_t u_{t-1} + 2\alpha^2 u_t u_{t-2} + \ldots]$$

$$= \sigma^2(1 + \alpha^2 + \alpha^4 + \ldots),$$

because $E[u_t u_s] = 0$ for $t \neq s$ and $E[u_t u_s] = \sigma^2$ for $t = s$. Applying the summation formula for the geometric series, and because of $|\alpha| < 1$, we get the constant variance

$$V[x_t] = \frac{\sigma^2}{1-\alpha^2} .$$

The covariances can be calculated as follows:

$$\text{Cov}\,[x_t, x_{t-\tau}] = E\left[\left(x_t - \frac{\delta}{1-\alpha}\right)\left(x_{t-\tau} - \frac{\delta}{1-\alpha}\right)\right]$$

$$
\begin{aligned}
&= E[(u_t + \alpha u_{t-1} + ... + \alpha^{\tau} u_{t-\tau} + ...) \\
&\quad \cdot (u_{t-\tau} + \alpha u_{t-\tau-1} + \alpha^2 u_{t-\tau-2} + ...)] \\
&= E[(u_t + \alpha u_{t-1} + ... + \alpha^{\tau-1} u_{t-\tau+1} \\
&\quad + \alpha^{\tau}(u_{t-\tau} + \alpha u_{t-\tau-1} + \alpha^2 u_{t-\tau-2} + ...)) \\
&\quad \cdot (u_{t-\tau} + \alpha u_{t-\tau-1} + \alpha^2 u_{t-\tau-2} + ...)] \\
&= \alpha^{\tau} E[(u_{t-\tau} + \alpha u_{t-\tau-1} + \alpha^2 u_{t-\tau-2} + ...)^2] .
\end{aligned}
$$

Thus, we get

$$
\text{Cov}[x_t, x_{t-\tau}] = \alpha^{\tau} V[x_{t-\tau}] = \alpha^{\tau} \frac{\sigma^2}{1-\alpha^2} .
$$

The autocovariances are only a function of the time difference τ and not of time t, and we can write:

$$
\gamma(\tau) = \alpha^{\tau} \frac{\sigma^2}{1-\alpha^2}, \quad \tau = 0, 1, 2, ... \tag{2.5}
$$

Therefore, the AR(1) process with $|\alpha| < 1$ and stochastic initial conditions is weakly stationary.

An Alternative Method for the Calculation of Moments

Under the condition of weak stationarity, i.e. for $|\alpha| < 1$ and stochastic initial conditions, the mean of x_t is constant. If we apply the expectation operator on equation (2.1), we get:

$$
E[x_t] = E[\delta + \alpha x_{t-1} + u_t] = \delta + \alpha E[x_{t-1}] + E[u_t] .
$$

Because of $E[u_t] = 0$ and $E[x_t] = E[x_{t-1}] = \mu$ for all t we can write

$$
E[x_t] = \mu = \frac{\delta}{1-\alpha} .
$$

If we consider the deviations from the mean,

$$
\tilde{x}_t = x_t - \mu
$$

and substitute this in relation (2.1), we get:

$$
\tilde{x}_t + \mu = \delta + \alpha \tilde{x}_{t-1} + \alpha \mu + u_t .
$$

From this it follows that

$$\tilde{x}_t = \delta + \mu(\alpha - 1) + \alpha\,\tilde{x}_{t-1} + u_t$$

$$= \delta + \frac{\delta}{1-\alpha}(\alpha - 1) + \alpha\,\tilde{x}_{t-1} + u_t$$

$$\tilde{x}_t = \alpha\,\tilde{x}_{t-1} + u_t \,. \tag{2.6}$$

This is the AR(1) process belonging to (2.1) with $E[\tilde{x}_t] = 0$.

If we multiply equation (2.6) with $\tilde{x}_{t-\tau}$ for $\tau \geq 0$ and take expectations we can write:

$$E[\tilde{x}_{t-\tau}\,\tilde{x}_t] = \alpha\,E[\tilde{x}_{t-\tau}\,\tilde{x}_{t-1}] + E[\tilde{x}_{t-\tau}\,u_t] \,. \tag{2.7}$$

Because of (2.3) we get

$$\tilde{x}_{t-\tau} = u_{t-\tau} + \alpha\,u_{t-\tau-1} + \alpha^2\,u_{t-\tau-2} + \ldots \,.$$

This leads to

$$E[\tilde{x}_{t-\tau}\,u_t] = \begin{cases} \sigma^2 & \text{for} \quad \tau = 0 \\ 0 & \text{for} \quad \tau > 0 \end{cases} \,. \tag{2.8}$$

Because of the stationarity assumption and because of the (even) symmetry of the autocovariances, $\gamma(\tau) = \gamma(-\tau)$, equation (2.7) results in

$$\tau = 0: \quad E[\tilde{x}_t^2] = \alpha\,E[\tilde{x}_t\,\tilde{x}_{t-1}] + \sigma^2,$$

or

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

$$\tau = 1: \quad E[\tilde{x}_t\,\tilde{x}_{t-1}] = \alpha\,E[\tilde{x}_{t-1}^2],$$

or

$$\gamma(1) = \alpha\,\gamma(0) \,.$$

This leads to the variance of the AR(1) process

$$\gamma(0) = \frac{\sigma^2}{1-\alpha^2} \,.$$

For $\tau \geq 1$ (2.7) implies

$$\gamma(1) = \alpha\,\gamma(0)$$

$$\gamma(2) = \alpha\,\gamma(1) = \alpha^2\,\gamma(0)$$

$$\gamma(3) = \alpha\,\gamma(2) = \alpha^3\,\gamma(0)$$

$$\vdots$$

$$\gamma(\tau) = \alpha\,\gamma(\tau\text{-}1) = \alpha^\tau\,\gamma(0)\ .$$

Thus, the covariances can be calculated from the linear homogeneous first order difference equation

$$\gamma(\tau) - \alpha\,\gamma(\tau\text{-}1) = 0$$

with the initial value $\gamma(0) = \sigma^2/(1 - \alpha^2)$.

The Autocorrelogram

Because of $\rho(\tau) = \gamma(\tau)/\gamma(0)$, the autocorrelation function (the autocorrelogram) of the AR(1) process is

$$\rho(\tau) = \alpha^\tau\,,\quad \tau = 1, 2, \ldots\ . \tag{2.9}$$

This function converges geometrically to zero for $\tau \to \infty$, and its infinite sum equals $1/(1 - \alpha)$ since $|\alpha| < 1$. This convergence is monotone for positive and oscillating for negative values of α.

Example 2.1

For $\delta = 0$ and $\alpha \in \{0.9, 0.5, -0.9\}$, *Figures 2.1 to 2.3* each present one realisation of the corresponding AR(1) process with T = 240 observations. To generate these series, we used realisations of normally distributed pure random processes with mean zero and variance one. We always dropped the first 60 observations to eliminate the dependence of the initial values.

The realisation for $\alpha = 0.9$, presented in *Figure 2.1*, is relatively smooth. This is to be expected given the theoretical autocorrelation function because random variables with a considerable distance between each other still have high positive correlations.

The development of the realisation in *Figure 2.2* with $\alpha = 0.5$ is much less systematic. The geometric decrease of the theoretical autocorrelation function is rather fast. The fourth order autocorrelation coefficient is only 0.0625.

Contrary to this, the realisation of the AR(1) process with $\alpha = -0.9$, presented in *Figure 2.3*, follows a well pronounced zigzag course with, however, alternating positive and negative amplitudes. This is consistent with the theoretical autocorrelation function indicating that all random variables with even-numbered distance are positively correlated and those with odd-numbered distance negatively correlated.

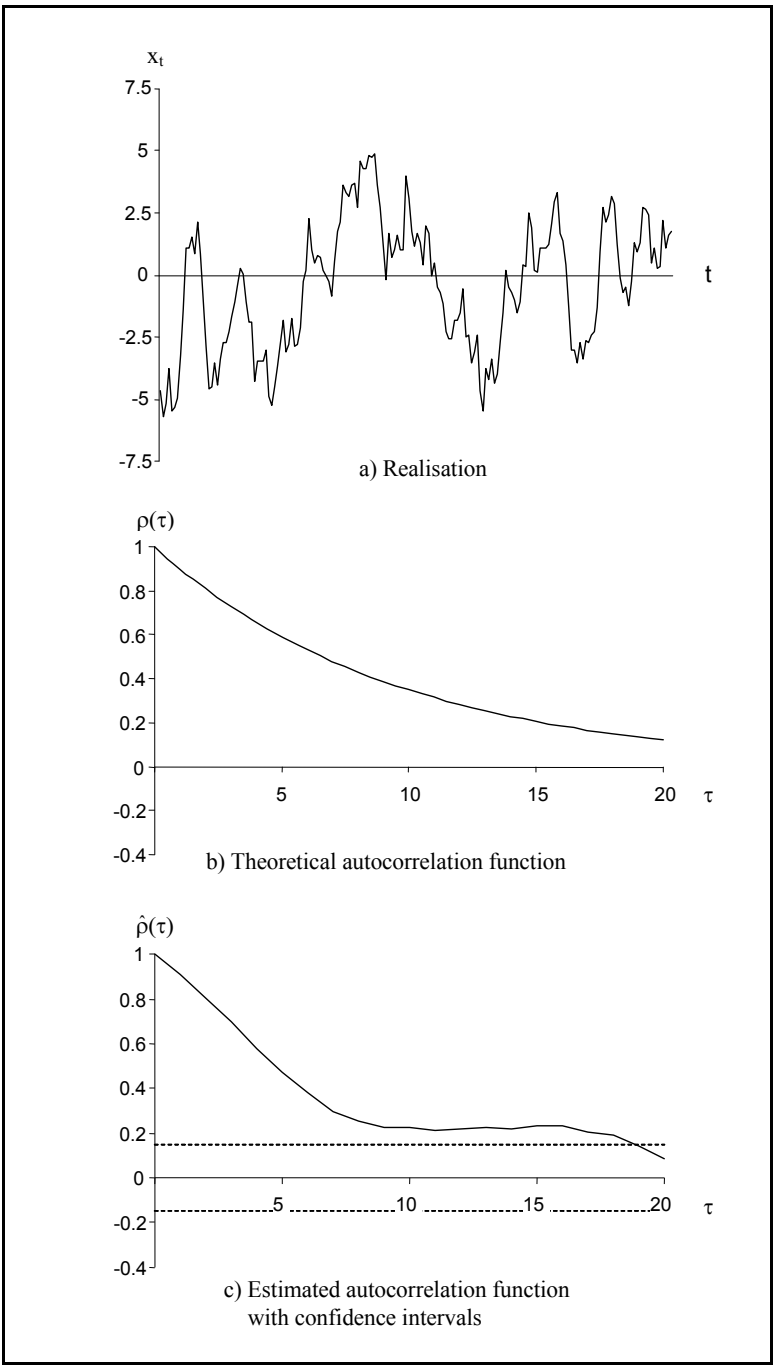

Figure 2.1: AR(1) process with $\alpha = 0.9$

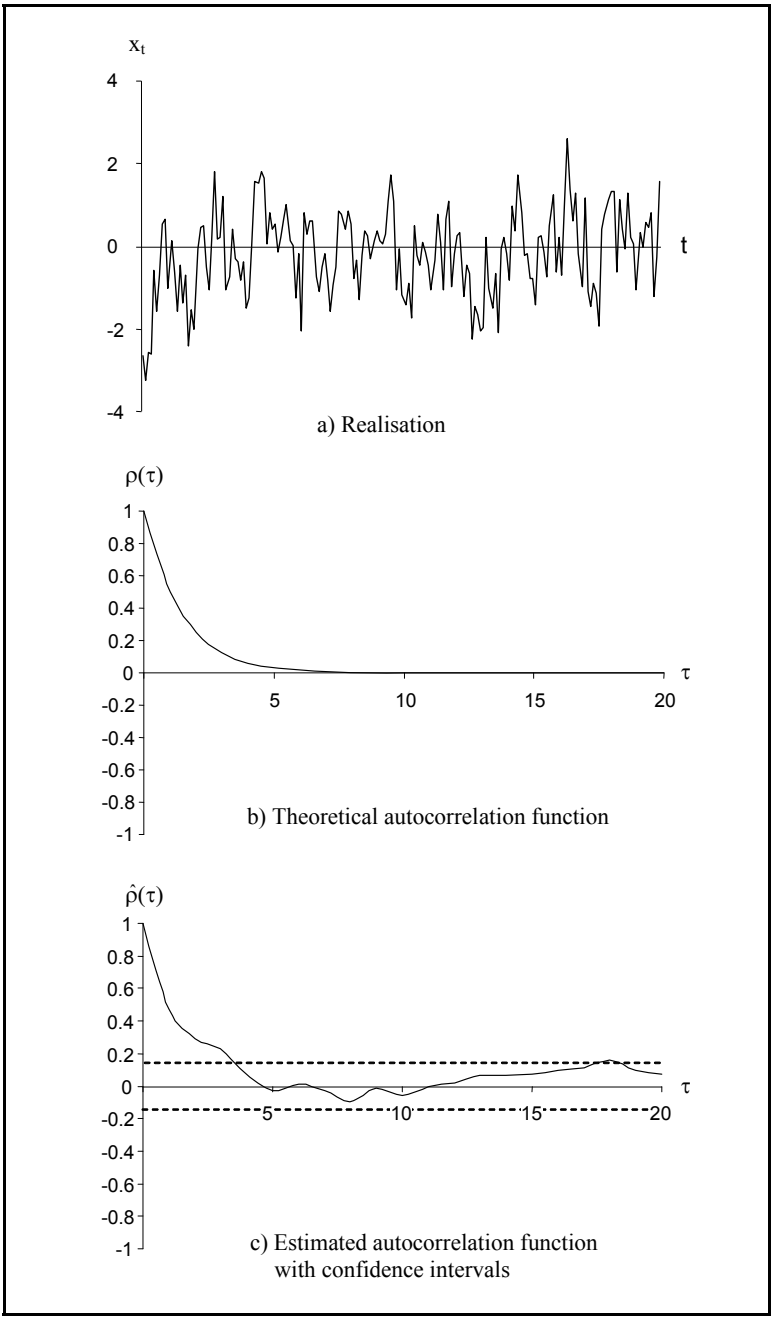

Figure 2.2: AR(1) process with α= 0.5

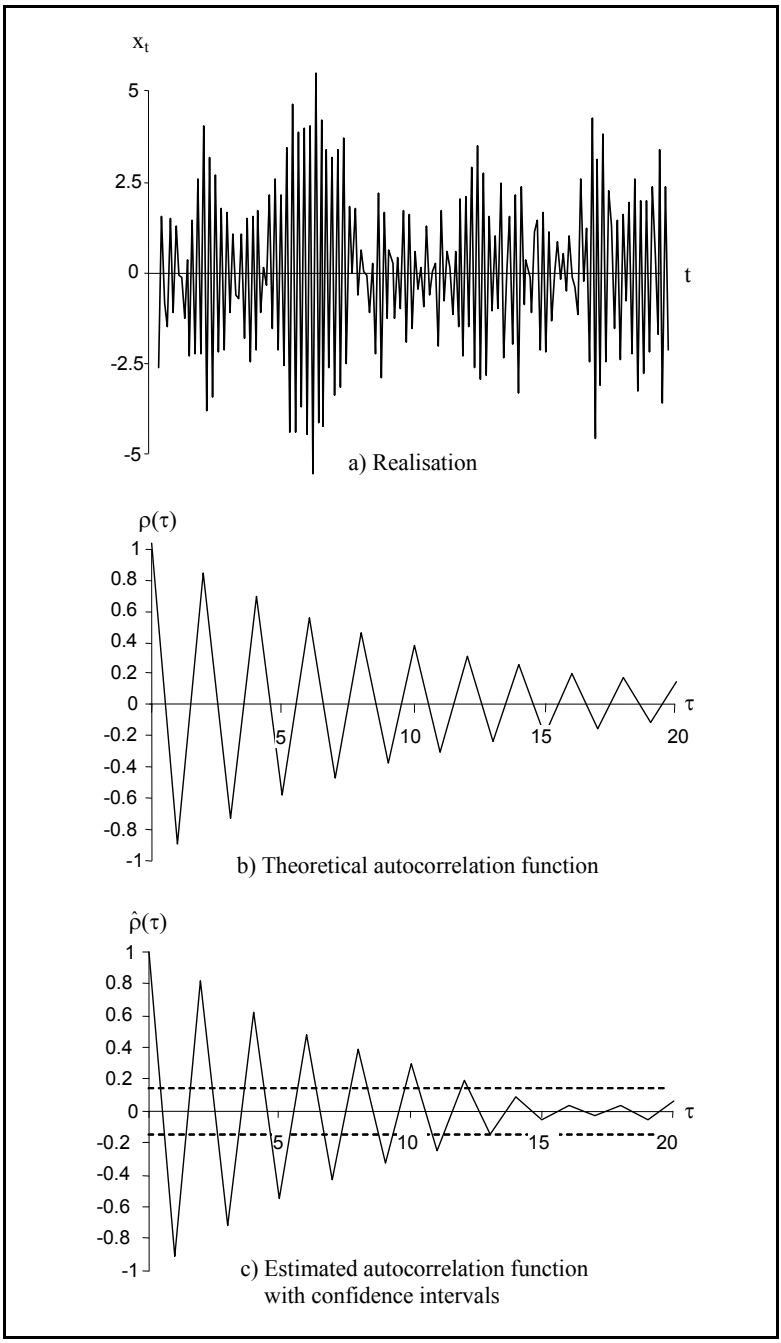

Figure 2.3: AR(1) process with α = -0.9

It generally holds that the closer the parameter α is to +1, the smoother the realisations will be. For negative values of α we get zigzag developments which are the more pronounced the closer α is to -1. For $\alpha = 0$ we get a pure random process. The autocorrelation functions estimated by means of relation (1.10) with the given realisations are also presented in *Figures 2.1 to 2.3*. The dotted parallel lines show approximate 95 percent confidence intervals for the null hypothesis assuming that the true process is a pure random process. In all three cases, the estimated functions reflect quite well the typical development of the theoretical autocorrelations.

Example 2.2

In a paper on the effect of economic development on the electoral chances of the German political parties during the period of the social-liberal coalition from 1969 to 1982, GEBHARD KIRCHGÄSSNER (1985) investigated (besides other issues) the time series properties of the popularity series of the parties constructed by monthly surveys of the Institute of Demoscopy in Allensbach (Germany). For the period from January 1971 to April 1982, the popularity series of the Christian Democratic Union (CDU), i.e. the share of voters who answered that they would vote for this party (or its Bavarian sister party, the CSU) if there were a general election by the following Sunday, is given in *Figure 2.4*. The autocorrelation and the partial autocorrelation function (which is discussed in *Section 2.1.4*) are also presented in this figure. While the autocorrelation function goes slowly towards zero, the partial autocorrelation function breaks off after $\tau = 1$. This argues for an AR(1) process.

The model has been estimated with Ordinary Least Squares (OLS), the method proposed in *Section 2.1.5* for the estimation of autoregressive models. Thus, we get:

$$\begin{array}{lcl} CDU_t & = & \underset{(3.43)}{8.053} + \underset{(17.10)}{0.834}\, CDU_{t-1} + \hat{u}_t, \\ \bar{R}^2 = 0.683, & \text{SE} = 1.586, & Q(11) = 12.516 \ (p = 0.326). \end{array}$$

The estimated t values are given in parentheses, SE denotes the standard error of the residuals. The autocorrelogram, which is also given in *Figure 2.4*, does not indicate any higher-order process. Moreover, given the high p-value, the Ljung-Box Q statistic with 12 correlation coefficients (i.e. with 11 degrees of freedom) gives no reason to reject this model. The mean is calculated as

$$\hat{\mu} = \frac{8.053}{1 - 0.834} = 48.512 .$$

It shows that about 48.5 percent of the voters voted on average for the CDU during this period.

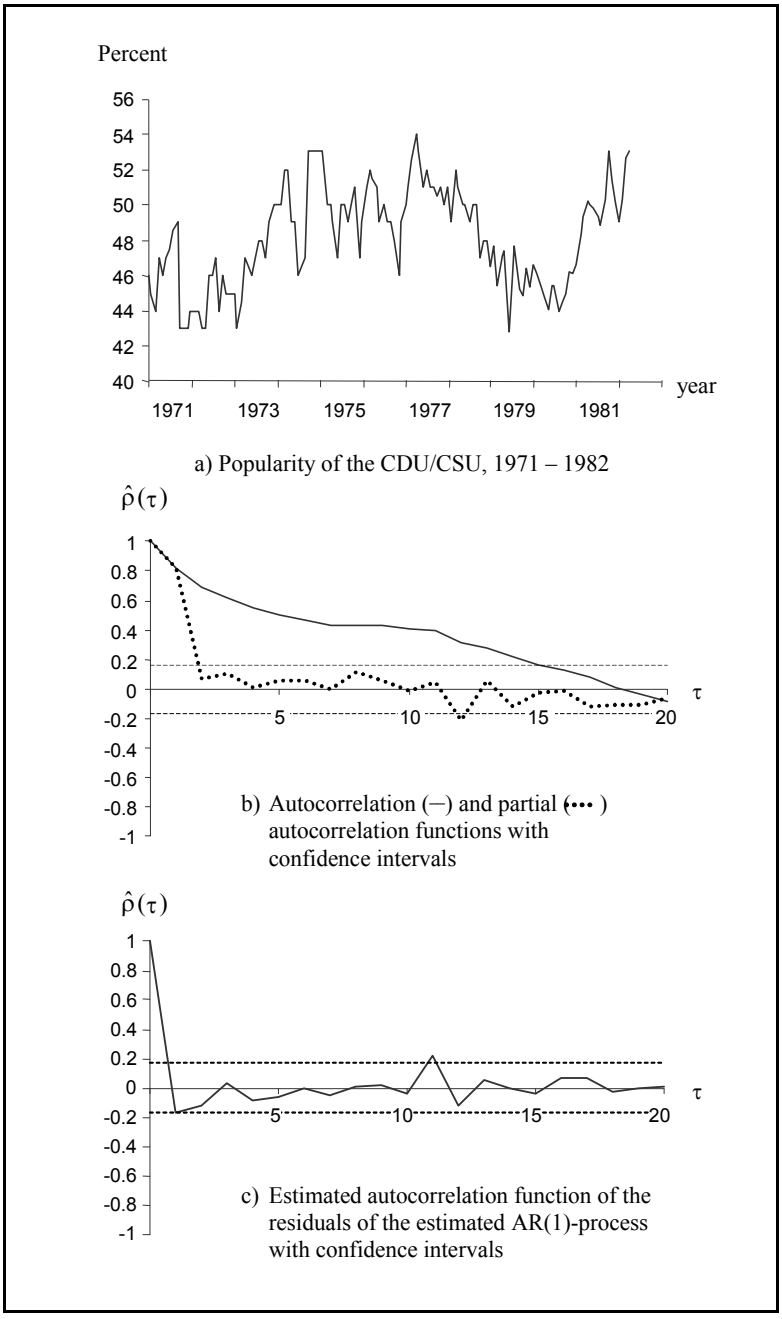

Figure 2.4: Popularity of the CDU/CSU, 1971 – 1982

Stability Conditions

Along with the stochastic initial value, the condition $|\alpha| < 1$, the so-called stability condition, is crucial for the stationarity of the AR(1) process. We can also derive the stability condition from the linear homogeneous difference equation, which is given for the process itself by

$$x_t - \alpha x_{t-1} = 0,$$

for its autocovariances by

$$\gamma(\tau) - \alpha \gamma(\tau-1) = 0$$

and for the autocorrelations by

$$\rho(\tau) - \alpha \rho(\tau-1) = 0.$$

These difference equations have stable solutions, i.e. $\lim_{\tau\to\infty}\rho(\tau) = 0$, if and only if their characteristic equation

$$\lambda - \alpha = 0 \tag{2.10}$$

has a solution (root) with an absolute value smaller than one, i.e. if $|\alpha| < 1$ holds. We get an equivalent condition if we do not consider the characteristic equation but the lag polynomial of the corresponding difference equations,

$$1 - \alpha L = 0. \tag{2.11}$$

This implies that the solution has to be larger than one in absolute value. (Strictly speaking, L, which denotes an operator, has to be substituted by a variable, which is often denoted by 'z'. To keep the notation simple, we use L in both meanings.)

Example 2.3

Let us consider the stochastic process

$$y_t = x_t + v_t . \tag{E2.1}$$

In this equation, x_t is a stationary AR(1) process, $x_t = \alpha x_{t-1} + u_t$, with $|\alpha| < 1$; v_t is a pure random process with mean zero and constant variance σ_v^2 which is uncorrelated with the other pure random process u_t with mean zero and constant variance σ_u^2.

We can interpret the stochastic process y_t as an additive decomposition of two stationary components. Then y_t itself is stationary. In the sense of MILTON FRIEDMAN (1957) we can interpret x_t as the permanent (systematic) and v_t as the transitory component.

What does the correlogram of y_t look like? As both x_t and v_t have zero mean, $E[y_t] = 0$. Multiplying (E2.1) with $y_{t-\tau}$ and taking expectations results in

$$E[y_{t-\tau}\, y_t] \;=\; E[y_{t-\tau}\, x_t] \;+\; E[y_{t-\tau}\, v_t]\,.$$

Due to $y_{t-\tau} \;=\; x_{t-\tau} + v_{t-\tau}$, we get

$$E[y_{t-\tau}\, y_t] \;=\; E[x_{t-\tau}\, x_t] \;+\; E[v_{t-\tau}\, x_t] \;+\; E[x_{t-\tau}\, v_t] + \; E[v_{t-\tau}\, v_t].$$

As u_t and v_t are uncorrelated, it holds that $E[v_{t-\tau}\, x_t] = E[x_{t-\tau}\, v_t] = 0$, and because of the stationarity of the two processes, we can write

(E2.2) $$\gamma_y(\tau) \;=\; \gamma_x(\tau) + \gamma_v(\tau)\,.$$

For $\tau = 0$ we get the variance of y_t as

$$\gamma_y(0) \;=\; \gamma_x(0) + \sigma_v^2 \;=\; \frac{\sigma_u^2}{1-\alpha^2} \;+\; \sigma_v^2\,.$$

For $\tau > 0$, because of $\gamma_v(\tau) = 0$ for $\tau \neq 0$, we get from (E2.2)

$$\gamma_y(\tau) \;=\; \gamma_x(\tau) \;=\; \alpha^{\tau}\,\frac{\sigma_u^2}{1-\alpha^2}\;.$$

Thus, we finally get

$$\rho_y(\tau) \;=\; \frac{\alpha^{\tau}}{1+(1-\alpha^2)\sigma_v^2/\sigma_u^2}\;,\quad \tau \;=\; 1, 2, ...,$$

for the correlogram of y_t. The overlay of the systematic component by the transitory component reduces the autocorrelation generated by the systematic component. The larger the variance of the transitory component, the stronger is this effect.

2.1.2 Second Order Autoregressive Processes

Generalising (2.1), the *second order autoregressive process* (AR(2)) can be written as

(2.12) $$x_t \;=\; \delta + \alpha_1\, x_{t-1} + \alpha_2\, x_{t-2} + u_t,$$

with u_t denoting a pure random process with variance σ^2 and $\alpha_2 \neq 0$. With the lag operator L we get

(2.13) $$(1 - \alpha_1\, L - \alpha_2\, L^2)\, x_t \;=\; \delta + u_t.$$

With $\alpha(L) = 1 - \alpha_1\, L - \alpha_2\, L^2$ we can write

(2.14) $$\alpha(L)\, x_t \;=\; \delta + u_t.$$

As for the AR(1) process, we get the Wold representation from (2.14) if we invert $\alpha(L)$; i.e. under the assumption that $\alpha^{-1}(L)$ exists and has the property

$$\alpha(L)\,\alpha^{-1}(L) = 1 \tag{2.15}$$

we can 'solve' for x_t in (2.14):

$$x_t = \alpha^{-1}(L)\,\delta + \alpha^{-1}(L)\,u_t\,. \tag{2.16}$$

If we use the series expansion with undetermined coefficients for

$$\alpha^{-1}(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \ldots$$

it has to hold that

$$1 = (1 - \alpha_1 L - \alpha_2 L^2)(\psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \ldots)$$

because of (2.15). This relation is an identity only if the coefficients of L^j, $j = 0, 1, 2, \ldots$, are equal on both the right and the left hand side. We get

$$\begin{array}{lccccccccc} 1 = & \psi_0 & + & \psi_1 L & + & \psi_2 L^2 & + & \psi_3 L^3 & + & \ldots \\ & & - & \alpha_1\psi_0 L & - & \alpha_1\psi_1 L^2 & - & \alpha_1\psi_2 L^3 & - & \ldots \\ & & & & - & \alpha_2\psi_0 L^2 & - & \alpha_2\psi_1 L^3 & - & \ldots \end{array}\,.$$

Comparing the coefficients of the lag polynomials on the right- and left-hand side finally leads to

$$\begin{array}{lll} L^0: & & \psi_0 = 1 \\ L^1: & \psi_1 - \alpha_1\psi_0 = 0 & \Rightarrow \quad \psi_1 = \alpha_1. \\ L^2: & \psi_2 - \alpha_1\psi_1 - \alpha_2\psi_0 = 0 & \Rightarrow \quad \psi_2 = \alpha_1^2 + \alpha_2\,. \\ L^3: & \psi_3 - \alpha_1\psi_2 - \alpha_2\psi_1 = 0 & \Rightarrow \quad \psi_3 = \alpha_1^3 + 2\alpha_1\alpha_2\,. \end{array}$$

By applying this so-called method of undetermined coefficients, we get the values ψ_j, $j = 2, 3, \ldots$, from the linear homogeneous difference equation

$$\psi_j - \alpha_1\,\psi_{j-1} - \alpha_2\,\psi_{j-2} = 0$$

with the initial conditions $\psi_0 = 1$ and $\psi_1 = \alpha_1$.

The stability condition for the AR(2) process requires that, for $j \to \infty$, the ψ_j converge to zero, i.e. that the characteristic equation of (2.12),

$$\lambda^2 - \alpha_1\,\lambda - \alpha_2 = 0, \tag{2.17}$$

has only roots with absolute values smaller than one, or that all solutions of the lag polynomial in (2.13),

(2.18) $$1 - \alpha_1 L - \alpha_2 L^2 = 0$$

are larger than one in modulus. Together with stochastic initial conditions, this guarantees the stationarity of the process. The stability conditions are fulfilled if the following parameter restrictions hold jointly for (2.17) and (2.18):

$$1 + (-\alpha_1) + (-\alpha_2) > 0,$$
$$1 - (-\alpha_1) + (-\alpha_2) > 0,$$
$$1 - (-\alpha_2) > 0.$$

As a constant is not changed by the application of the lag operator, the number '1' can substitute the lag operator in the corresponding terms. Thus, due to (2.16), the Wold representation of the AR(2) process is given by

(2.19) $$x_t = \frac{\delta}{1-\alpha_1-\alpha_2} + \sum_{j=0}^{\infty} \psi_j u_{t-j} \quad , \psi_0 = 1.$$

Under the assumption of stationarity, the expected value of the stochastic process can be calculated directly from (2.12) since $E[x_t] = E[x_{t-1}] = E[x_{t-2}] = \mu$. We get

$$\mu = \delta + \alpha_1 \mu + \alpha_2 \mu$$

or

(2.20) $$E[x_t] = \mu = \frac{\delta}{1-\alpha_1-\alpha_2} .$$

As the stability conditions are fulfilled, $1 - \alpha_1 - \alpha_2 > 0$ holds, i.e. the sign of δ also determines the sign of μ.

In order to calculate the second order moments, we can assume – without loss of generality – that $\mu = 0$, which is equivalent to $\delta = 0$. Multiplying (2.12) with $x_{t-\tau}$, $\tau \geq 0$, and taking expectations leads to

(2.21) $$E[x_{t-\tau} x_t] = \alpha_1 E[x_{t-\tau} x_{t-1}] + \alpha_2 E[x_{t-\tau} x_{t-2}] + E[x_{t-\tau} u_t] .$$

Because of representation (2.19), relation (2.8) holds here as well. This leads to the following equations

(2.22) $$\begin{aligned} \tau = 0 &: \gamma(0) = \alpha_1\gamma(1) + \alpha_2\gamma(2) + \sigma^2 \\ \tau = 1 &: \gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(1) \\ \tau = 2 &: \gamma(2) = \alpha_1\gamma(1) + \alpha_2\gamma(0) \end{aligned} ,$$

and, more generally, the following difference equation holds for the autocovariances $\gamma(\tau)$, $\tau \geq 2$,

$$\gamma(\tau) - \alpha_1 \gamma(\tau\text{-}1) - \alpha_2 \gamma(\tau\text{-}2) = 0. \tag{2.23}$$

As the stability conditions hold, the autocovariances which can be recursively calculated with (2.23) are converging to zero for $\tau \to \infty$.

The relations (2.22) result in

$$V[x_t] = \gamma(0) = \frac{1-\alpha_2}{(1+\alpha_2)\ [(1-\alpha_2)^2 - \alpha_1^2]}\sigma^2 \tag{2.24}$$

for the variance of the AR(2) process, and in

$$\gamma(1) = \frac{\alpha_1}{(1+\alpha_2)\ [(1-\alpha_2)^2 - \alpha_1^2]}\sigma^2,$$

and

$$\gamma(2) = \frac{\alpha_1^2 + \alpha_2 - \alpha_2^2}{(1+\alpha_2)\ [(1-\alpha_2)^2 - \alpha_1^2]}\sigma^2,$$

for the autocovariances of order one and two.

The autocorrelations can be calculated accordingly. If we divide (2.23) by the variance $\gamma(0)$ we get the linear homogeneous second order difference equation,

$$\rho(\tau) - \alpha_1 \rho(\tau\text{-}1) - \alpha_2 \rho(\tau\text{-}2) = 0 \tag{2.25}$$

with the initial conditions $\rho(0) = 1$ and $\rho(1) = \alpha_1/(1 - \alpha_2)$ for the autocorrelation function. Depending on the values of α_1 and α_2, AR(2) processes can generate quite different developments, and, therefore, these processes can show considerably different characteristics.

Example 2.4

Let us consider the AR(2) process

$$x_t = 1 + 1.5\, x_{t\text{-}1} - 0.56\, x_{t\text{-}2} + u_t \tag{E2.3}$$

with a variance of u_t of 1. Because the characteristic equation

$$\lambda^2 - 1.5\,\lambda + 0.56 = 0$$

has the two roots $\lambda_1 = 0.8$ and $\lambda_2 = 0.7$, (E2.3) is stationary, given that we have stochastic initial conditions. The expected value of this process is

$$\mu = \frac{1}{1 - 1.5 + 0.56} = 16.\overline{6}\,.$$

The variance of (E2.3) can be calculated from (2.24) as $\gamma(0) = 19.31$. A realisation of this process (with 180 observations) is given in *Figure 2.5* in which the (estimated) mean was subtracted. Thus, the realisations fluctuate around zero, and the process always tends to go back to the mean. This *mean-reverting behaviour* is a typical property of stationary processes.

Due to (2.25) we get

$$\rho(\tau) - 1.5\,\rho(\tau\text{-}1) + 0.56\,\rho(\tau\text{-}2) = 0, \quad \tau = 2, 3, \ldots,$$
$$\text{with } \rho(0) = 1, \quad \rho(1) = 0.96$$

for the autocorrelation function. The general solution of this homogeneous difference equation is

$$\rho(\tau) = C_1\,(0.8)^{\tau} + C_2\,(0.7)^{\tau}\,,$$

where C_1 and C_2 are two arbitrary constants. Taking into account the two initial conditions we get

$$\rho(\tau) = 2.6\,(0.8)^{\tau} - 1.6\,(0.7)^{\tau}$$

for the autocorrelation coefficients. This development is also expressed in *Figure 2.5*. The coefficients are always positive but strictly monotonically decreasing. Initially, the estimated autocorrelogram using the given realisation is also monotonically decreasing, but, contrary to the theoretical development, the values begin to fluctuate from the tenth lag onwards. However, except for the coefficient for $\tau = 16$, the estimates are not significantly different from zero; they are all inside the approximate 95 percent confidence interval indicated by the dotted lines.

The characteristic equations of stable autoregressive processes of second or higher order can result in conjugate complex roots. In this case, the time series exhibit dampened oscillations, which are shocked again and again by the pure random process. The solution of the homogeneous part of (2.12) for conjugate complex roots can be represented by

$$x_t = d^t\,(C_1 \cos(f\,t) + C_2 \sin(f\,t))$$

with C_1 and C_2 again being arbitrary constants that can be determined by using the initial conditions. The dampening factor

$$d = \sqrt{-\alpha_2}$$

corresponds to the modulus of the two roots, and

$$f = \arccos\left(\frac{\alpha_1}{2\sqrt{-\alpha_2}}\right)$$

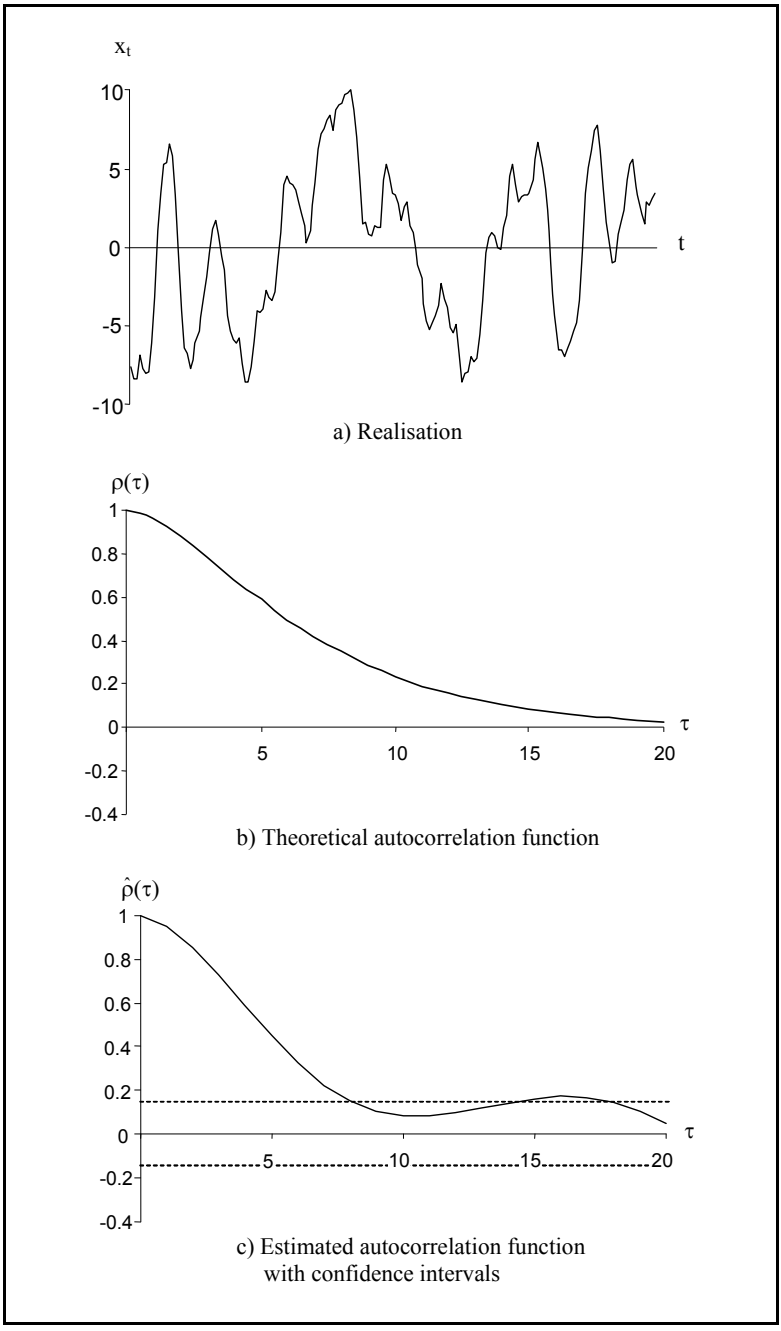

Figure 2.5: AR(2) process with $\alpha_1 = 1.5$, $\alpha_2 = -0.56$

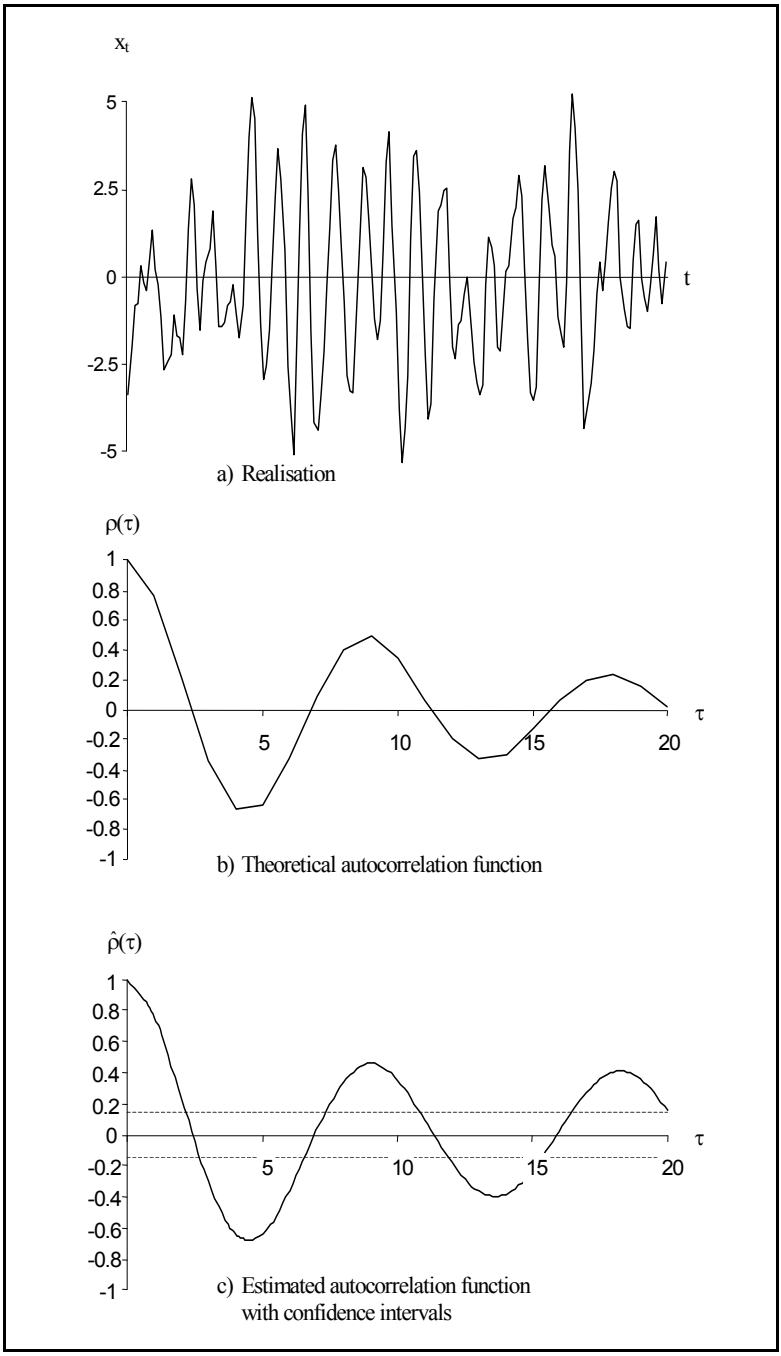

Figure 2.6: AR(2) process with $\alpha_1 = 1.4$ and $\alpha_2 = -0.85$

is the frequency of the oscillation. The period of the cycles is $P = 2\pi/f$. Processes with conjugate complex roots are well-suited to describe business cycle fluctuations.

Example 2.5

Consider the AR(2) process

$$x_t = 1.4\, x_{t-1} - 0.85\, x_{t-2} + u_t, \tag{E2.4}$$

with a variance of u_t of 1. The characteristic equation

$$\lambda^2 - 1.4\,\lambda + 0.85 = 0$$

has the two solutions $\lambda_1 = 0.7 + 0.6i$ and $\lambda_2 = 0.7 - 0.6i$. ('i' stands for the imaginary unit: $i^2 = -1$.) The modulus (dampening factor) is $d = 0.922$. Thus, (E2.4) with stochastic initial conditions and a mean of zero is stationary. According to (2.24) the variance is given by $\gamma(0) = 8.433$.

A realisation of this process with 180 observations is given in *Figure 2.6*. Its development is cyclical around its zero mean. For the autocorrelation function we get

$$\rho(\tau) - 1.4\,\rho(\tau\text{-}1) + 0.85\,\rho(\tau\text{-}2) = 0, \quad \tau = 2, 3, \ldots,$$
$$\rho(0) = 1, \quad \rho(1) = 0.76,$$

because of (2.25).

The general solution is

$$\rho(\tau) = 0.922^{\tau}\,(C_1 \cos(0.709\,\tau) + C_2 \sin(0.709\,\tau)) .$$

Taking into account the two initial conditions, we get for the autocorrelation coefficients

$$\rho(\tau) = 0.922^{\tau}\,(\cos(0.709\,\tau) + 0.1 \sin(0.709\,\tau)) ,$$

with a frequency of $f = 0.709$.

In case of quarterly data, this corresponds to a period length of about 9 quarters. Both the theoretical and the estimated autocorrelations in *Figure 2.6* show this kind of dampened periodical behaviour.

Example 2.6

Figure 2.7 shows the development of the three month money market rate in Frankfurt (GSR) from the first quarter of 1970 to the last quarter of 1998 as well as the autocorrelation and the partial autocorrelation functions explained in *Section 2.1.4*. Whereas the autocorrelation function tends only slowly towards zero, the partial autocorrelation function breaks off after two lags. As will be shown below, this indicates an AR(2) process. For the period from 1970 to 1998, estimation with OLS results in the following:

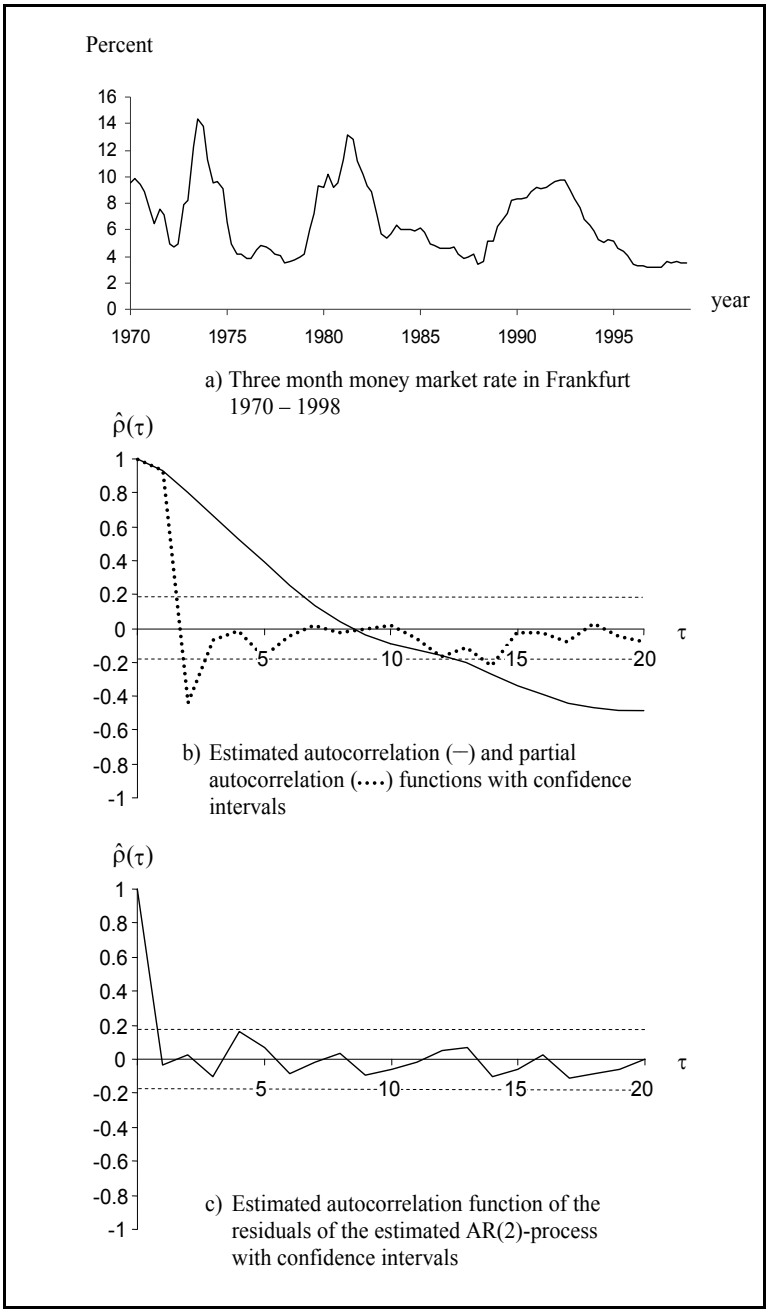

Figure 2.7: Three month money market rate in Frankfurt, 1970 – 1998

$$GSR_t = \underset{(2.82)}{0.575} + \underset{(17.50)}{1.407}\, GSR_{t-1} - \underset{(-6.16)}{0.498}\, GSR_{t-2} + \hat{u}_t,$$

$$\bar{R}^2 = 0.910, \quad SE = 0.812, \quad Q(6) = 6.475 \ (p = 0.372),$$

with t values being again given in parentheses. On the 0.1 percent level, both estimated coefficients of the lagged interest rates are significantly different from zero. The autocorrelogram of the estimated residuals (given in *Figure 2.7c*) as well as the Ljung-Box Q statistic which is calculated with 8 correlation coefficients (and 6 degrees of freedom) does not indicate any higher order process.
The two roots of the process are 0.70 ± 0.06i, i.e. they indicate dampened cycles. The modulus (dampening factor) is d = 0.706; the frequency f = 0.079 corresponds to a period of 79.7 quarters and therefore of nearly 20 years. Correspondingly, this oscillation cannot be detected in the estimated autocorrelogram presented in *Figure 2.7b*.

2.1.3 Higher Order Autoregressive Processes

An AR(p) process can be described by the following stochastic difference equation,

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + u_t, \tag{2.26}$$

with $\alpha_p \neq 0$, where u_t is again a pure random process with zero mean and variance σ^2. Using the lag operator we can also write:

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p)\, x_t = \delta + u_t. \tag{2.26'}$$

If we assume stochastic initial conditions, the AR(p) process in (2.26) is stationary if the stability conditions are satisfied, i.e. if the characteristic equation

$$\lambda^p - \alpha_1 \lambda^{p-1} - \alpha_2 \lambda^{p-2} - \ldots - \alpha_p = 0 \tag{2.27}$$

only has roots with absolute values smaller than one, or if the solutions of the lag polynomial

$$1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p = 0 \tag{2.28}$$

only have roots with absolute values larger than one.

If the stability conditions are satisfied, we get the Wold representation of the AR(p) process by the series expansion of the inverse lag polynomial,

$$\frac{1}{1 - \alpha_1 L - \ldots - \alpha_p L^p} = 1 + \psi_1 L + \psi_2 L^2 + \ldots$$

as

$$x_t = \frac{\delta}{1 - \alpha_1 - \ldots - \alpha_p} + \sum_{j=0}^{\infty} \psi_j u_{t-j} \quad . \tag{2.29}$$

Generalising the approach that was used to calculate the coefficients of the AR(2) process, the series expansion can again be calculated by the method of undetermined coefficients.

From (2.29) we get the constant (unconditional) expectation as

$$E[x_t] = \frac{\delta}{1 - \alpha_1 - \ldots - \alpha_p} = \mu \, .$$

Again, similarly to the AR(1) and AR(2) cases, a necessary condition for stability is

$$1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p > 0.$$

Without loss of generality we can set $\delta = 0$, i.e. $\mu = 0$, in order to calculate the autocovariances. Because of $\gamma(\tau) = E[x_{t-\tau} \, x_t]$, we get according to (2.26)

$$\gamma(\tau) = E[x_{t-\tau} \, (\alpha_1 \, x_{t-1} + \alpha_2 \, x_{t-2} + \ldots + \alpha_p \, x_{t-p} + u_t)] \, . \tag{2.30}$$

For $\tau = 0, 1, \ldots, p$, it holds that

$$\begin{aligned}
\gamma(0) &= \alpha_1\gamma(1) + \alpha_2\gamma(2) + \cdots + \alpha_p\gamma(p) + \sigma^2 \\
\gamma(1) &= \alpha_1\gamma(0) + \alpha_2\gamma(1) + \cdots + \alpha_p\gamma(p-1) \\
&\vdots \\
\gamma(p) &= \alpha_1\gamma(p-1) + \alpha_2\gamma(p-2) + \cdots + \alpha_p\gamma(0)
\end{aligned} \tag{2.31}$$

because of the symmetry of the autocovariances and because of $E[x_{t-\tau} \, u_t] = \sigma^2$ for $\tau = 0$ and zero for $\tau > 0$.

This is a linear inhomogeneous equation system for given α_i and σ^2 to derive the $p + 1$ unknowns $\gamma(0), \gamma(1), \ldots, \gamma(p)$. For $\tau > p$ we get the linear homogeneous difference equation to calculate the autocovariances of order $\tau > p$:

$$\gamma(\tau) - \alpha_1 \, \gamma(\tau\text{-}1) - \ldots - \alpha_p \, \gamma(\tau\text{-}p) = 0. \tag{2.32}$$

If we divide (2.32) by $\gamma(0)$, we get the corresponding difference equation to calculate the autocorrelations:

$$\rho(\tau) - \alpha_1 \, \rho(\tau\text{-}1) - \ldots - \alpha_p \, \rho(\tau\text{-}p) = 0. \tag{2.33}$$

The initial conditions $\rho(1), \rho(2), \ldots, \rho(p)$ can be derived from the so-called *Yule-Walker equations*. We get those if we successively insert $\tau = 1, 2, \ldots, p$ in (2.33), or, if the last p equations in (2.31) are divided by $\gamma(0)$,

$$
\begin{aligned}
\rho(1) &= \alpha_1 + \alpha_2\,\rho(1) + \alpha_3\,\rho(2) + \ldots + \alpha_p\,\rho(p\text{-}1)\\
\rho(2) &= \alpha_1\,\rho(1) + \alpha_2 + \alpha_3\,\rho(1) + \ldots + \alpha_p\,\rho(p\text{-}2)\\
&\vdots\\
\rho(p) &= \alpha_1\,\rho(p\text{-}1) + \alpha_2\,\rho(p\text{-}2) + \alpha_3\,\rho(p\text{-}3) + \ldots + \alpha_p
\end{aligned}
\tag{2.34}
$$

If we define $\boldsymbol{\rho}' = (\rho(1), \rho(2), \ldots, \rho(p))$, $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ and

$$
\underset{p\times p}{R} = \begin{bmatrix}
1 & \rho(1) & \rho(2) & \cdots & \rho(p-1)\\
\rho(1) & 1 & \rho(1) & \cdots & \rho(p-2)\\
\vdots & & & & \\
\rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1
\end{bmatrix}
$$

we can write the Yule-Walker equations (2.34) in matrix form,

$$
\boldsymbol{\rho} = R\,\boldsymbol{\alpha}\,. \tag{2.35}
$$

If the first p autocorrelation coefficients are given, the coefficients of the AR(p) process can be calculated according to (2.35) as

$$
\boldsymbol{\alpha} = R^{-1}\,\boldsymbol{\rho}\,. \tag{2.36}
$$

Equations (2.35) and (2.36) show that there is a one-to-one mapping between the p coefficients $\boldsymbol{\alpha}$ and the first p autocorrelation coefficients $\boldsymbol{\rho}$ of an AR(p) process. If there is a generating pure random process, it is sufficient to know either $\boldsymbol{\alpha}$ or $\boldsymbol{\rho}$ to identify the AR(p) process. Thus, there are two possibilities to describe the structure of an autoregressive process of order p: the parametric representation that uses the parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$, and the non-parametric representation with the first p autocorrelation coefficients $\rho(1), \rho(2), \ldots, \rho(p)$. Both representations contain exactly the same information. Which representation is used depends on the specific situation. We usually use the parametric representation to describe finite order autoregressive processes (with known order).

Example 2.7

Let the fourth order autoregressive process

$$
x_t = \alpha_4\,x_{t-4} + u_t,\quad 0 < \alpha_4 < 1,
$$

be given, where u_t is again white noise with zero mean and variance σ^2. Applying (2.31) we get:

$$
\begin{aligned}
\gamma(0) &= \alpha_4\,\gamma(4) + \sigma^2,\\
\gamma(1) &= \alpha_4\,\gamma(3),\\
\gamma(2) &= \alpha_4\,\gamma(2),
\end{aligned}
$$

$$\gamma(3) = \alpha_4 \gamma(1),$$
$$\gamma(4) = \alpha_4 \gamma(0).$$

From these relations we get

$$\gamma(0) = \frac{\sigma^2}{1 - \alpha_4^2},$$

$$\gamma(1) = \gamma(2) = \gamma(3) = 0,$$

$$\gamma(4) = \alpha_4 \frac{\sigma^2}{1 - \alpha_4^2}.$$

As can easily be seen, only the autocovariances with lag $\tau = 4j$, $j = 1, 2, \ldots$ are different from zero, while all other autocovariances are zero. Thus, for $\tau > 0$ we get the autocorrelation function

$$\rho(\tau) = \begin{cases} \alpha_4^j & \text{for } \tau = 4j,\ j = 1, 2, \ldots \\ 0 & \text{elsewhere.} \end{cases}$$

Only every fourth autocorrelation coefficient is different from zero; the sequence of these autocorrelation coefficients decreases monotonically like a geometric series. Employing such a model for quarterly data, this AR(4) process captures the correlation between random variables that are distant from each other by a multiplicity of four periods, i.e. the structure of the correlations of all variables which belong to the i-th quarter of a year, $i = 1, 2, 3, 4$, follows an AR(1) process while the correlations between variables that belong to different quarters are always zero. Such an AR(4) process provides a simple possibility of modelling seasonal effects which typically influence the same quarters of different years. For empirical applications, it is advisable to first eliminate the deterministic component of a seasonal variation by employing seasonal dummies and then to model the remaining seasonal effects by such an AR(4) process.

2.1.4 The Partial Autocorrelation Function

Due to the stability conditions, autocorrelation functions of stationary finite order autoregressive processes are always sequences that converge to zero but do not break off. This makes it difficult to distinguish between processes of different orders when using the autocorrelation function. To cope with this problem, we introduce a new concept, the *partial autocorrelation function*. The partial correlation between two random variables is the correlation that remains if the possible impact of all other random variables has been eliminated. To define the partial autocorrelation coefficient, we use the new notation,

$$x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \ldots + \phi_{kk}x_{t-k} + u_t,$$

where ϕ_{ki} is the coefficient of the variable with lag i if the process has order k. (According to the former notation it holds that $\alpha_i = \phi_{ki}$ i = 1,2,…,k.) The coefficients ϕ_{kk} are the partial autocorrelation coefficients (of order k), k = 1,2,… . The partial autocorrelation measures the correlation between x_t and x_{t-k} which remains when the influences of x_{t-1}, x_{t-2}, ..., x_{t-k+1} on x_t and x_{t-k} have been eliminated.

Due to the Yule-Walker equations (2.35), we can derive the partial autocorrelation coefficients ϕ_{kk} from the autocorrelation coefficients if we calculate the coefficients ϕ_{kk}, which belong to x_{t-k}, for k = 1, 2, ... from the corresponding linear equation systems

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \rho(2) & \cdots & \rho(k-2) \\ \vdots & & & & \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix}, \quad k = 1, 2, \ldots .$$

With Cramer's rule we get

$$(2.37) \qquad \phi_{kk} = \frac{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(1) \\ \rho(1) & 1 & \cdots & \rho(2) \\ \vdots & \vdots & & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \cdots & \rho(k-2) \\ \vdots & \vdots & & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & 1 \end{vmatrix}}, \quad k = 1, 2, \ldots .$$

Thus, if the data generating process (DGP) is an AR(1) process, we get for the partial autocorrelation function:

$$\phi_{11} = \rho(1)$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = 0,$$

because of $\rho(2) = \rho(1)^2$. Generally, the partial autocorrelation coefficients $\phi_{kk} = 0$ for $k > 1$ in an AR(1) process.

If the DGP is an AR(2) process, we get

$$\phi_{11} = \rho(1), \quad \phi_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}, \phi_{kk} = 0 \quad \text{for} \quad k > 2 .$$

The same is true for an AR(p) process: all partial autocorrelation coefficients of order higher than p are zero. Thus, for finite order autoregressive processes, the partial autocorrelation function provides the possibility of identifying the order of the process by the order of the last non-zero partial autocorrelation coefficient. We can estimate the partial autocorrelation coefficients consistently by substituting the theoretical values in (2.37) by their consistent estimates (1.10). For the partial autocorrelation coefficients which have a theoretical value of zero, i.e. the order of which is larger than the order of the process, we get asymptotically that they are normally distributed with $E[\hat{\phi}_{kk}] = 0$ and $V[\hat{\phi}_{kk}] = 1/T$ for $k > p$.

Example 2.8

The AR(1) process of *Example 2.1* has the following theoretical partial autocorrelation function: $\phi_{11} = \rho(1) = \alpha$ and zero elsewhere. In this example, α takes on the values 0.9, 0.5 and -0.9. The estimates of the partial autocorrelation functions for the realisations in *Figures 2.1* and *2.3* are presented in *Figure 2.8*. It is obvious for both processes that these are AR(1) processes. The estimated value for the process with $\alpha = 0.9$ is $\hat{\phi}_{11} = 0.91$, while all other partial autocorrelation coefficients are not significantly different from zero. We get $\hat{\phi}_{11} = -0.91$ for the process with $\alpha = -0.9$, while all estimated higher order partial autocorrelation coefficients do not deviate significantly from zero.

The AR(2) process of *Example 2.4* has the following theoretical partial autocorrelation function: $\phi_{11} = 0.96$, $\phi_{22} = -0.56$ and zero elsewhere. The realisation of this process, which is given in *Figure 2.5*, leads to the empirical partial autocorrelation function in *Figure 2.8*. It corresponds quite closely to the theoretical function; we get $\hat{\phi}_{11} = 0.95$ and $\hat{\phi}_{22} = -0.60$ and all higher order partial autocorrelation coefficients are not significantly different from zero. The same holds for the AR(2) process with the theoretical non-zero partial autocorrelations $\phi_{11} = 0.76$ and $\phi_{22} = -0.85$ given in *Example 2.5*. We get the estimates $\hat{\phi}_{11} = 0.76$ and $\hat{\phi}_{22} = -0.78$, whereas all higher order partial correlation coefficients are not significantly different from zero.

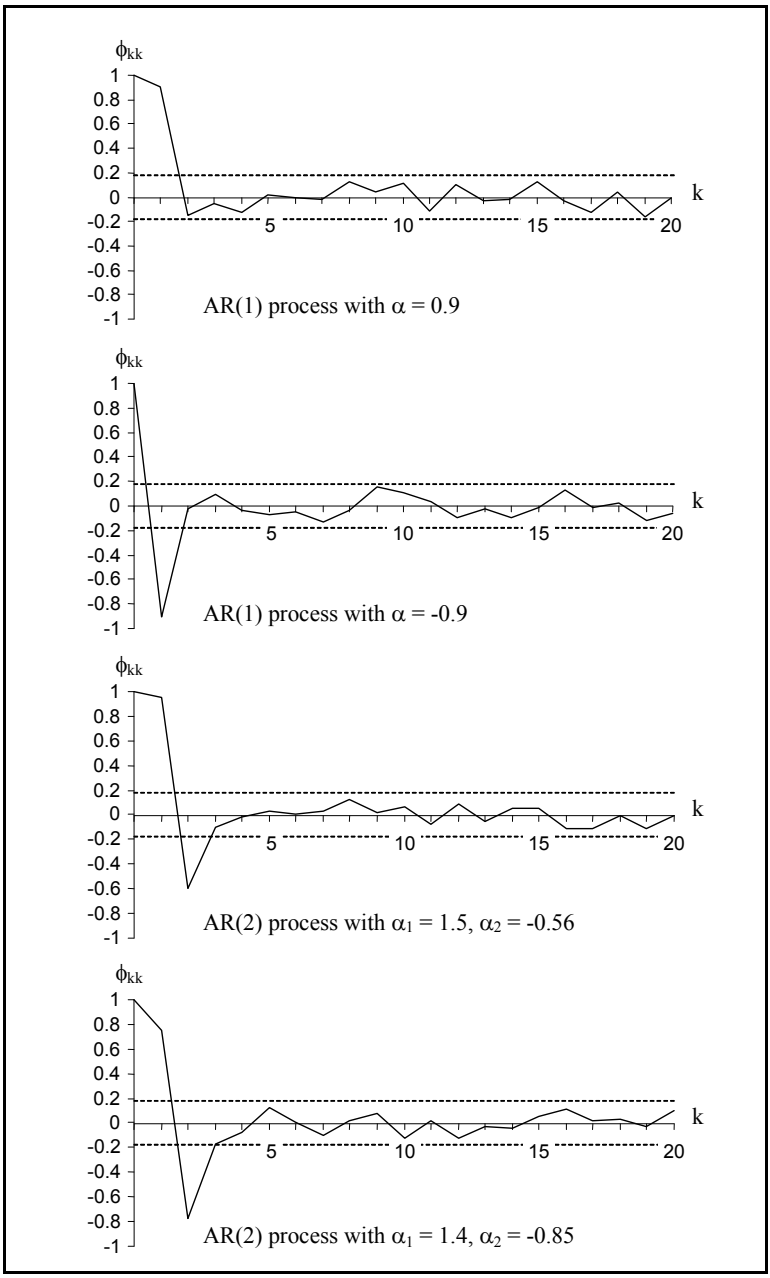

Figure 2.8: Estimated partial autocorrelation functions

2.1.5 Estimating Autoregressive Processes

Under the assumption of a known order p we have different possibilities to estimate the parameters:

(i) If we know the distribution of the white noise process that generates the AR(p) process, the parameters can be estimated by using maximum likelihood (ML) methods.

(ii) The parameters can also be estimated with the method of moments by using the Yule-Walker equations.

(iii) A further possibility is to treat

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + u_t, \tag{2.26}$$

as a regression equation and apply the ordinary least squares (OLS) method for estimation. OLS provides consistent estimates. Moreover, if (2.26) fulfils the stability conditions, $\sqrt{T}(\hat{\delta} - \delta)$ as well as $\sqrt{T}(\hat{\alpha}_i - \alpha_i)$, i = 1, 2, ..., p, are asymptotically normally distributed.

If the order of the AR process is unknown, it can be estimated with the help of information criteria. For this purpose, AR processes with successively increasing orders p = 1, 2, ..., p^{max} are estimated. Finally, the order p* is chosen which minimises the respective criterion. The following criteria are often used:

(i) The final prediction error which goes back to HIROTUGU AKAIKE (1969)

$$FPE = \frac{T+m}{T-m} \cdot \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t^{(p)})^2 .$$

(ii) Closely related to this is the Akaike information criterion (HIROTUGU AKAIKE (1974))

$$AIC = \ln \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t^{(p)})^2 + m \frac{2}{T} .$$

(iii) Alternatives are the Bayesian criterion of GIDEON SCHWARZ (1978)

$$SC = \ln \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t^{(p)})^2 + m \frac{\ln T}{T}$$

(iv) as well as the criterion developed by EDWARD J. HANNAN and BARRY G. QUINN (1979)

$$HQ = \ln \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t^{(p)})^2 + m \frac{2 \ln(\ln T)}{T} .$$

$\hat{u}_t^{(p)}$ are the estimated residuals of the AR(p) process, while m is the number of estimated parameters. If the constant term is estimated, too, $m = p + 1$ for an AR(p) process. These criteria are always based on the same principle: They consist of one part, the sum of squared residuals (or its logarithm), which decreases when the number of estimated parameters increases, and of a 'penalty term', which increases when the number of estimated parameters increases. Whereas the first two criteria overestimate the true (finite) order asymptotically, the two other criteria estimate the true order of the process consistently. For $T \geq 16$, the penalty term of SC is larger than the one of HQ which itself is larger than the one of AIC. This leads to the following ordering of the estimated AR orders:

$$\text{SC order} \leq \text{HQ order} \leq \text{AIC order}.$$

Please note that choosing such an order does not always imply that we have white noise residuals. This has to be checked independently. Many computer programmes like, for example, EViews, do not exactly report the criteria given in (ii) through (iv). Relying on the log-likelihood function instead of on the sum of squared residuals directly, they add $1 + \ln(2\pi) \approx 2.8379$, which does, of course, neither affect the order nor which value of p minimises the information criteria.

Example 2.9

As in *Example 2.6*, we take a look at the development of the three month money market interest rate in Frankfurt am Main. If, for this series, we estimate AR processes up to the order $p = 4$, we get the following results (for $T = 116$):

$p = 0$: AIC = 4.8334, HQ = 4.8430, SC = 4.8571;

$p = 1$: AIC = 2.7180, HQ = 2.7373, SC = 2.7655;

$p = 2$: AIC = 2.4457, HQ = 2.4746, SC = 2.5169;

$p = 3$: AIC = 2.4609, HQ = 2.4995, SC = 2.5559;

$p = 4$: AIC = 2.4778, HQ = 2.5260, SC = 2.5965.

With all three criteria we get the minimum for $p = 2$. Thus, the optimal number of lags is $p^* = 2$, as used in *Example 2.6*.

2.2 Moving Average Processes

Moving average processes of an infinite order have already occurred when we presented the Wold decomposition theorem. They are, above all, of theoretical importance as, in practice, only a finite number of (different) parameters can be estimated. In the following, we consider finite order moving average processes. We start with the first order moving average process and then discuss general properties of finite order moving average processes.

2.2.1 First Order Moving Average Processes

The first order moving average process (MA(1)) is given by the following equation:

$$x_t = \mu + u_t - \beta u_{t-1}, \tag{2.38}$$

or

$$x_t - \mu = (1 - \beta L)u_t, \tag{2.38'}$$

with u_t again being a pure random process. The Wold representation of an MA(1) process (as of any finite order MA process) has a finite number of terms. In this special case, the Wold coefficients are $\psi_0 = 1$, $\psi_1 = -\beta$ and $\psi_j = 0$ for $j \geq 2$. Thus, $\sum_j \psi_j^2$ is finite for all finite values of β, i.e. an MA(1) process is always stationary.

Taking expectations of (2.38) leads to

$$E[x_t] = \mu + E[u_t] - \beta E[u_{t-1}] = \mu .$$

The variance can also be calculated directly,

$$\begin{aligned} V[x_t] &= E[(x_t - \mu)^2] \\ &= E[(u_t - \beta u_{t-1})^2] \\ &= E[(u_t^2 - 2\beta u_t u_{t-1} + \beta^2 u_{t-1}^2)] \\ &= (1 + \beta^2)\sigma^2 = \gamma(0) . \end{aligned}$$

Therefore, the variance is constant at any point of time.

For the covariances of the process we get

$$\begin{aligned} E[(x_t - \mu)(x_{t+\tau} - \mu)] &= E[(u_t - \beta u_{t-1})(u_{t+\tau} - \beta u_{t+\tau-1})] \\ &= E[(u_t u_{t+\tau} - \beta u_t u_{t+\tau-1} - \beta u_{t-1} u_{t+\tau} + \beta^2 u_{t-1} u_{t+\tau-1})] . \end{aligned}$$

The covariances are different from zero only for $\tau = \pm 1$, i.e. for adjoining random variables. In this case

$$\gamma(1) = -\beta\sigma^2 .$$

Thus, for an MA(1) process, all autocovariances and therefore all autocorrelations with an order higher than one disappear, i.e. $\gamma(\tau) = \rho(\tau) = 0$ for $\tau \geq 2$.

The correlogram of an MA(1) process is

$$\rho(0) = 1, \quad \rho(1) = \frac{-\beta}{1+\beta^2}, \quad \rho(\tau) = 0 \text{ for } \tau \geq 2.$$

If we consider $\rho(1)$ as a function of β, $\rho(1) = f(\beta)$, it holds that $f(0) = 0$ and $f(\beta) = -f(-\beta)$, i.e. that $f(\beta)$ is point symmetric to the origin, and that $|f(\beta)| \leq 0.5$. $f(\beta)$ has its maximum at $\beta = -1$ and its minimum at $\beta = 1$. Thus, an MA(1) process cannot have a first order autocorrelation above 0.5 or below -0.5.

If we know the autocorrelation coefficient $\rho(1) = \rho_1$, for example, by estimation, we can derive (estimate) the corresponding parameter β by using the equation for the first order autocorrelation coefficient,

$$(1+\beta^2)\rho_1 + \beta = 0 .$$

The quadratic equation can also be written as

$$\beta^2 + \frac{1}{\rho_1}\beta + 1 = 0, \tag{2.39}$$

and it has the two solutions

$$\beta_{1,2} = -\frac{1}{2\rho_1}\left(1 \pm \sqrt{1-4\rho_1^2}\right) .$$

Thus, the parameters of the MA(1) process can be estimated non-linearly with the method of moments: the theoretical moments are substituted by their consistent estimates and the resulting equation is used for estimating the parameters consistently.

Because of $|\rho_1| \leq 0.5$, the quadratic equation always results in real roots. They also have the property that $\beta_1\beta_2 = 1$. This gives us the possibility to model the same autocorrelation structure with two different parameters, where one is the inverse of the other.

In order to get a unique parameterisation, we require a further property of the MA(1) process. We ask under which conditions the MA(1) process (2.38) can have an autoregressive representation. By using the lag operator representation (2.38') we get

$$u_t = -\frac{\mu}{1-\beta} + \frac{1}{1-\beta L} x_t .$$

An expansion of the series $1/(1 - \beta L)$ is only possible for $|\beta| < 1$ and results in the following AR(∞) process

$$u_t = -\frac{\mu}{1-\beta} + x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \ldots$$

or

$$x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \ldots = \frac{\mu}{1-\beta} + u_t .$$

This representation requires the condition of *invertibility* ($|\beta| < 1$). In this case, we get a unique parameterisation of the MA(1) process. Applying the lag polynomial in (2.38'), we can formulate the *invertibility condition* in the following way: An MA(1) process is invertible if and only if the root of the lag polynomial

$$1 - \beta L = 0$$

is larger than one in modulus.

Example 2.10

The following MA(1) process is given:

(E2.5) $$x_t = \varepsilon_t - \beta \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, 2^2),$$

with $\beta = -0.5$. For this process we get

$$E[x_t] = 0,$$

$$V[x_t] = (1 + 0.5^2) \cdot 4 = 5,$$

$$\rho(1) = \frac{0.5}{1+0.5^2} = 0.4,$$

$$\rho(\tau) = 0 \quad \text{for} \quad \tau \geq 2.$$

Solving the corresponding quadratic equation (2.39) for this value of $\rho(1)$ leads to the two roots $\beta_1 = -2.0$ and $\beta_2 = -0.5$. If we now consider the process

(E2.5a) $$y_t = \eta_t + 2 \eta_{t-1}, \quad \eta_t \sim N(0, 1),$$

we obtain the following results:

$$E[y_t] = 0,$$

$$V[y_t] = (1 + 2.0^2) \cdot 1 = 5,$$

$$\rho(1) = \frac{2.0}{1+2.0^2} = 0.4,$$

$$\rho(\tau) = 0 \quad \text{for} \quad \tau \geq 2,$$

i.e. the variances and the autocorrelogram of the two processes (E2.5) and (E2.5a) are identical. The only difference between them is that (E2.5) is invertible, because the invertibility condition $|\beta| < 1$ holds, whereas (E2.5a) is not invertible. Thus, given the structure of the correlations, we can choose the one of the two processes that fulfils the invertibility condition without imposing any restrictions on the structure of the process.

With equation (2.37), the partial autocorrelation function of the MA(1) process can be calculated in the following way:

$$\phi_{11} = \rho(1),$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = -\frac{\rho(1)^2}{1-\rho(1)^2} < 0,$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & 0 \\ 0 & \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & 0 \\ \rho(1) & 1 & \rho(1) \\ 0 & \rho(1) & 1 \end{vmatrix}} = \frac{\rho(1)^3}{1-2\rho(1)^2} \lessgtr 0 \quad \text{for} \quad \beta \lessgtr 0,$$

$$\phi_{44} = \frac{\begin{vmatrix} 1 & \rho(1) & 0 & \rho(1) \\ \rho(1) & 1 & \rho(1) & 0 \\ 0 & \rho(1) & 1 & 0 \\ 0 & 0 & \rho(1) & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & 0 & 0 \\ \rho(1) & 1 & \rho(1) & 0 \\ 0 & \rho(1) & 1 & \rho(1) \\ 0 & 0 & \rho(1) & 1 \end{vmatrix}} = \frac{-\rho(1)^4}{(1-\rho(1)^2)^2-\rho(1)^2} < 0,$$

etc.

If β is positive, $\rho(1)$ is negative and vice versa. This leads to the two possible patterns of partial autocorrelation functions, exemplified by $\beta = \pm 0.8$:

$$\beta = 0.8, \quad \phi_{ii} \in \{-0.49,-0.31,-0.22, -0.17, \ldots\},$$

$$\beta = -0.8, \quad \phi_{ii} \in \{0.49,-0.31, 0.22, -0.17, \ldots\}.$$

Thus, contrary to the AR(1) process, the autocorrelation function of the MA(1) process breaks off, while the partial autocorrelation function does not. These properties hold generally, since invertible finite order MA processes are equivalent to infinite order AR processes.

2.2.2 MA(1) and Temporal Aggregation

The time series which are discussed in this book are measured in discrete time, with intervals of equal length. Exchange rates, for example, are normally quoted at the end of each trading day. For econometric analyses, however, monthly, quarterly, or even annual data are used, rather than these daily values. Usually, averages or end-of-period data are used for temporal aggregation.

Thus, two aggregation schemes have to be distinguished. The first one is skip sampling (or: systematic sampling) where only every m^{th} data point is recorded. If x_t is the basic series at $t = 1, 2, 3, \ldots$, the skip sampled series y_s with new time scale s is end-of-period data,

$$y_1 = x_m, \quad y_2 = x_{2m}, \quad y_3 = x_{3m}, \ldots, y_s = x_{sm}.$$

Such an aggregation is typical for stock variables. However, the second scheme of averaging over m non-overlapping periods is also widely used, in particular for rates or indices:

$$\begin{aligned}
\tilde{y}_1 &= \frac{1}{m}\left(x_m + x_{m-1} + \ldots + x_1\right) \\
\tilde{y}_2 &= \frac{1}{m}\left(x_{2m} + x_{2m-1} + \ldots + x_{m+1}\right) \\
&\vdots \\
\tilde{y}_s &= \frac{1}{m}\left(x_{sm} + x_{sm-1} + \ldots + x_{(s-1)m+1}\right).
\end{aligned}$$

In the following, we do not present a general theory of temporal aggregation but just discuss a special case of particular applied interest, the random walk, with

$$x_t = x_{t-1} + u_t,$$

where an artificial MA(1) structure arises due to aggregation by averaging. It is straightforward to see that systematic sampling does not affect the random walk property, since in this case we can write

$$y_s = x_0 + \sum_{t=1}^{sm} u_t .$$

From this representation we get

$$y_s = y_{s-1} + \eta_s,$$

with η_s being white noise:

$$\eta_s = u_{sm} + u_{sm-1} + \ldots + u_{(s-1)m+1},$$

with $E[\eta_s] = 0$ and

$$E(\eta_s \cdot \eta_{s-\tau}) = \begin{cases} m\sigma_u^2 & \text{for} \quad \tau = 0 \\ 0 & \text{elsewhere} \end{cases} .$$

Hence, the random walk property is inherited by y_s, only the variance of the differences $y_s - y_{s-1}$ is inflated in the obvious way. In case of averaging, $\tilde{y}_s$, matters get more complicated. It can, however, be shown that the differences

$$\tilde{y}_s - \tilde{y}_{s-1} = \tilde{\eta}_s$$

follow no longer a white noise process but an MA(1) scheme hidden behind

$$\tilde{\eta}_s = \frac{1}{m}\left(u_{sm} + 2u_{sm-1} + \ldots + mu_{(s-1)m+1} + \ldots + 2u_{(s-2)m+3} + u_{(s-2)m+2}\right).$$

We omit details but refer to HOLBROOK WORKING (1960) who showed that with increasing aggregation level, $m \to \infty$, one obtains the autocorrelation function

$$\rho(\tau) = \frac{E\left[\tilde{\eta}_s \tilde{\eta}_{s-\tau}\right]}{V\left[\tilde{\eta}_s\right]} \rightarrow \begin{cases} 1, & \tau = 0 \\ \frac{1}{4}, & |\tau| = 1 \\ 0, & \text{elsewhere} \end{cases} .$$

Note that the above autocorrelation function corresponds to the following MA(1)-process

$$\tilde{\eta}_s = \tilde{u}_s - \tilde{\beta}\tilde{u}_{s-1}$$

where $\tilde{u}_s$ is white noise, and the limiting value (for $m \rightarrow \infty$) of the MA parameter is

$$\tilde{\beta} = \sqrt{3} - 2 \approx -0.268.$$

GEORGE C. TIAO (1972) generalised this result the following way: If $x_t - x_{t-1}$ is not generated by white noise but by an invertible MA(1) process, then $\tilde{y}_s - \tilde{y}_{s-1}$ behaves with growing *m* like the MA(1) process $\tilde{u}_s - \tilde{\beta}\tilde{u}_{s-1}$, where $\tilde{\beta}$ is independent of the underlying MA(1) structure of $x_t - x_{t-1}$. This result even continues to hold when the assumption that $x_t - x_{t-1}$ is MA(1) is replaced by a more general moving average process of higher order as introduced in subsection 2.2.3.

Example 2.11

Consider averaging over m = 2 periods,

$$\tilde{y}_s = \frac{1}{2}\left(x_{2s} + x_{2s-1}\right).$$

For the random walk $x_t = x_{t-1} + u_t$, it holds that

$$\begin{aligned} \tilde{\eta}_s &= \tilde{y}_s - \tilde{y}_{s-1} \\ &= \frac{1}{2}\left(x_{2s} + x_{2s-1} - x_{2s-2} - x_{2s-3}\right) \\ &= \frac{1}{2}\left(u_{2s} + 2\,u_{2s-1} + u_{2s-2}\right). \end{aligned}$$

This process can be described as

$$\tilde{\eta}_s = \tilde{u}_s - \beta\tilde{u}_{s-1}$$

with $\beta = 2\sqrt{2} - 3 \approx -0.172$, and

$$E(\tilde{\eta}_s \cdot \tilde{\eta}_{s-\tau}) = \begin{cases} \frac{3}{2}\sigma_u^2 & \text{for} \quad \tau = 0 \\ \frac{1}{4}\sigma_u^2 & \text{for} \quad |\tau| = 1 \\ 0 & \text{elsewhere} \end{cases},$$

such that for m = 2 the autocorrelation coefficient at lag one becomes $\rho(1) = 1/6$.

Example 2.12

Example 1.3 as well as *Figure 1.8* present the end-of-month exchange rate between the Swiss Franc and the U.S. Dollar over the period from January 1974 to December 2011. The autocorrelogram of the first differences of the logarithms of this time series indicates that they follow a pure random process. The tests we applied did not reject this null hypothesis.

If we use monthly averages instead of end-of-month data, the following MA(1) process can be estimated for the first difference of the logarithms of this exchange rate:

$$\begin{aligned} \Delta\ln(e_t) &= \underset{(-1.53)}{-0.003} + \hat{u}_t + \underset{(6.91)}{0.308}\,\hat{u}_{t-1}, \\ \bar{R}^2 &= 0.082, \quad SE = 0.028, \quad Q(11) = 8.216 \ (p = 0.694), \\ JB &= 21.194 \ (p = 0.000), \end{aligned}$$

with the t values again given in parentheses. $\ln(\cdot)$ denotes the natural logarithm. The estimated coefficient of the MA(1) term is highly significantly different from zero. The Ljung-Box Q-statistic indicates that there is no longer any significant autocorrelation in the residuals. As $m \approx 20$ is relatively large (in this context), the estimated values of the MA(1) term should not be too different from the theoretical value given by GEORGE C. TIAO (1972). The theoretical value -0.268 lies in the two-sigma confidence interval of the estimated parameter -0.308.

2.2.3 Higher Order Moving Average Processes

In general, the *moving average process of order q* (MA(q)) can be written as

$$x_t = \mu + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \ldots - \beta_q u_{t-q} \tag{2.40}$$

with $\beta_q \neq 0$ and u_t as a pure random process. Using the lag operator we get

$$\begin{aligned} x_t - \mu &= (1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_q L^q) u_t \\ &= \beta(L) u_t \,. \end{aligned} \tag{2.40'}$$

From (2.40) we see that we already have a finite order Wold representation with $\psi_k = 0$ for $k > q$. Thus, there are no problems of convergence, and every finite MA(q) process is stationary, no matter what values are used for β_j, $j = 1, 2, ..., q$.

For the expectation of (2.40) we immediately get $E[x_t] = \mu$. Thus, the variance can be calculated as:

$$\begin{aligned} V[x_t] &= E[(x_t - \mu)^2] \\ &= E[(u_t - \beta_1 u_{t-1} - \ldots - \beta_q u_{t-q})^2] \\ &= E[(u_t^2 + \beta_1^2 u_{t-1}^2 + \ldots + \beta_q^2 u_{t-q}^2 - 2\beta_1 u_t u_{t-1} - \ldots \\ &\quad - 2\beta_{q-1}\beta_q u_{t-q+1} u_{t-q})] . \end{aligned}$$

From this we obtain

$$V[x_t] = (1 + \beta_1^2 + \beta_2^2 + \ldots + \beta_q^2)\,\sigma^2 .$$

For the covariances of order τ we can write

$$\begin{aligned} Cov[x_t, x_{t+\tau}] &= E[(x_t - \mu)(x_{t+\tau} - \mu)] \\ &= E[(u_t - \beta_1 u_{t-1} - \ldots - \beta_q u_{t-q}) \\ &\quad \cdot (u_{t+\tau} - \beta_1 u_{t+\tau-1} - \ldots - \beta_q u_{t+\tau-q})] \\ &= E[u_t(u_{t+\tau} - \beta_1 u_{t+\tau-1} - \ldots - \beta_q u_{t+\tau-q}) \\ &\quad - \beta_1 u_{t-1}(u_{t+\tau} - \beta_1 u_{t+\tau-1} - \ldots - \beta_q u_{t+\tau-q}) \\ &\quad \vdots \\ &\quad - \beta_q u_{t-q}(u_{t+\tau} - \beta_1 u_{t+\tau-1} - \ldots - \beta_q u_{t+\tau-q})] . \end{aligned}$$

Thus, for $\tau = 1, 2, ..., q$ we get

$$\begin{aligned} &\tau = 1: \quad \gamma(1) = (-\beta_1 + \beta_1\beta_2 + \ldots + \beta_{q-1}\beta_q)\sigma^2, \\ &\tau = 2: \quad \gamma(2) = (-\beta_2 + \beta_1\beta_3 + \ldots + \beta_{q-2}\beta_q)\sigma^2, \\ &\qquad \vdots \\ &\tau = q: \quad \gamma(q) = -\beta_q\sigma^2, \end{aligned} \tag{2.41}$$

while we have $\gamma(\tau) = 0$ for $\tau > q$.

Consequently, all autocovariances and autocorrelations with orders higher than the order of the process are zero. It is – at least theoretically – possible to identify the order of an MA(q) process by using the autocorrelogram.

It can be seen from (2.41) that there exists a system of non-linear equations for given (or estimated) second order moments that determines (makes it possible to estimate) the parameters $\beta_1, ..., \beta_q$. As we have al-

ready seen in the case of the MA(1) process, such non-linear equation systems have multiple solutions, i.e. there exist different values for β_1, β_2, ... and β_q that all lead to the same autocorrelation structure. To get a unique parameterisation, the invertibility condition is again required, i.e. it must be possible to represent the MA(q) process as a stationary AR(∞) process. Starting from (2.40'), this implies that the inverse operator $\beta^{-1}(L)$ can be represented as an infinite series in the lag operator, where the sum of the coefficients has to be bounded. Thus, the representation we get is an AR(∞) process

$$u_t = -\frac{\mu}{\beta(1)} + \beta^{-1}(L)\, x_t$$

$$= -\frac{\mu}{\beta(1)} + \sum_{j=0}^{\infty} c_j x_{t-j} \ ,$$

where

$$1 = (1 - \beta_1 L - ... - \beta_q L^q)(\, 1 + c_1 L + c_2 L^2 + ... \,),$$

and the parameters c_i, i = 1, 2, ... are calculated by using again the method of undetermined coefficients. Such a representation exists if all roots of

$$1 - \beta_1 L - ... - \beta_q L^q = 0$$

are larger than one in absolute value.

Example 2.13

Let the following MA(2) process

$$x_t = u_t + 0.6\, u_{t-1} - 0.1\, u_{t-2}$$

be given, with a variance of 1 given for the pure random process u. For the variance of x we get

$$V[x_t] = (1 + 0.36 + 0.01) \cdot 1 = 1.37 \ .$$

Corresponding to (2.41) the covariances are

$$\gamma(1) = +0.6 - 0.06 = 0.54$$

$$\gamma(2) = -0.1 \qquad .$$

$$\gamma(\tau) = 0 \quad \text{for } \tau > 2$$

This leads to the autocorrelation coefficients $\rho(1) = 0.39$ and $\rho(2) = -0.07$. To check whether the process is invertible, the quadratic equation

$$1 + 0.6\, L - 0.1\, L^2 = 0$$

has to be solved. As the two roots -1.36 and 7.36 are larger than 1 in absolute value, the invertibility condition is fulfilled, i.e. the MA(2) process can be written as an AR(∞) process

$$\begin{aligned} x_t &= (1 + 0.6\,L - 0.1\,L^2)\,u_t\,, \\ u_t &= \frac{1}{1+0.6L-0.1L^2}\,x_t \\ &= (1 + c_1\,L + c_2\,L^2 + c_3\,L^3 + \ldots)\,x_t\,. \end{aligned}$$

The unknowns c_i, i = 1, 2, ..., can be determined by comparing the coefficients of the polynomials in the following way:

$$\begin{aligned} 1 &= (1 + 0.6\,L - 0.1\,L^2)(1 + c_1\,L + c_2\,L^2 + c_3\,L^3 + \ldots) \\ 1 &= 1 + c_1\,L + c_2\,L^2 + c_3\,L^3 + \ldots \\ &\quad + 0.6\,L + 0.6\,c_1\,L^2 + 0.6\,c_2\,L^3 + \ldots \\ &\quad - 0.1\,L^2 - 0.1\,c_1\,L^3 - \ldots \end{aligned}$$

It holds that

$$\begin{aligned} c_1 + 0.6 = 0 &\Rightarrow c_1 = -0.60, \\ c_2 + 0.6\,c_1 - 0.1 = 0 &\Rightarrow c_2 = 0.46, \\ c_3 + 0.6\,c_2 - 0.1\,c_1 = 0 &\Rightarrow c_3 = -0.34, \\ c_4 + 0.6\,c_3 - 0.1\,c_2 = 0 &\Rightarrow c_4 = 0.25, \\ &\vdots \qquad\quad . \end{aligned}$$

Thus, we get the following AR(∞) representation

$$x_t - 0.6\,x_{t-1} + 0.46\,x_{t-2} - 0.34\,x_{t-3} + 0.25\,x_{t-4} - \ldots = u_t\,.$$

Similarly to the MA(1) process, the partial autocorrelation function of the MA(q) process does not break off. As long as the order q is finite, the MA(q) process is stationary whatever its parameters are. If the order tends towards infinity, however, for the process to be stationary the series of the coefficients has to converge just like in the Wold representation.

2.3 Mixed Processes

If we take a look at the two different functions that can be used to identify autoregressive and moving average processes, we see from *Table 2.1* that the situation in which neither of them breaks off can only arise if there is an MA(∞) process that can be inverted to an AR(∞) process, i.e. if the Wold representation of an AR(∞) process corresponds to an MA(∞) process. However, as pure AR or MA representations, these processes cannot

be used for empirical modelling because they can only be characterised by means of infinitely many parameters. After all, according to the *principle of parsimony*, the number of estimated parameters should be as small as possible when applying time series methods.

In the following, we introduce processes which contain both an autoregressive (AR) term of finite order p and a moving average (MA) term of finite order q. Hence, these mixed processes are denoted as ARMA(p,q) processes. They enable us to describe processes in which neither the autocorrelation nor the partial autocorrelation function breaks off after a finite number of lags. Again, we start with the simplest case, the ARMA(1,1) process, and consider the general case afterwards.

Table 2.1: Characteristics of the Autocorrelation and the Partial Autocorrelation Functions of AR and MA Processes

	Autocorrelation Function	Partial Autocorrelation Function
MA(q)	breaks off with q	does not break off
AR(p)	does not break off	breaks off with p

2.3.1 ARMA(1,1) Processes

An ARMA(1,1) process can be written as follows,

$$x_t = \delta + \alpha x_{t-1} + u_t - \beta u_{t-1}, \tag{2.42}$$

or, by using the lag operator

$$(1 - \alpha L)\, x_t = \delta + (1 - \beta L)\, u_t, \tag{2.42'}$$

where u_t is a pure random process. To get the Wold representation of an ARMA(1,1) process, we solve (2.42') for x_t,

$$x_t = \frac{\delta}{1-\alpha} + \frac{1-\beta L}{1-\alpha L}\, u_t .$$

It is obvious that $\alpha \neq \beta$ must hold, because otherwise x_t would be a pure random process fluctuating around the mean $\mu = \delta/(1 - \alpha)$. The ψ_j, $j = 0, 1, 2, \ldots$, can be determined as follows:

$$\frac{1-\beta L}{1-\alpha L} = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots$$

$$1-\beta L = (1-\alpha L)(\psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots)$$

$$1-\beta L = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots - \alpha\psi_0 L - \alpha\psi_1 L^2 - \alpha\psi_2 L^3 - \dots .$$

Comparing the coefficients of the two lag polynomials we get

$$\begin{aligned}
L^0&: \quad \psi_0 = 1 \\
L^1&: \quad \psi_1 - \alpha\psi_0 = -\beta \Rightarrow \psi_1 = \alpha - \beta \\
L^2&: \quad \psi_2 - \alpha\psi_1 = 0 \Rightarrow \psi_2 = \alpha(\alpha - \beta) \\
L^3&: \quad \psi_3 - \alpha\psi_2 = 0 \Rightarrow \psi_3 = \alpha^2(\alpha - \beta) \\
&\vdots \\
L^j&: \quad \psi_j - \alpha\psi_{j-1} = 0 \Rightarrow \psi_j = \alpha^{j-1}(\alpha - \beta) .
\end{aligned}$$

The ψ_j, $j \geq 2$ can be determined from the linear homogeneous difference equation

$$\psi_j - \alpha\psi_{j-1} = 0$$

with $\psi_1 = \alpha - \beta$ as initial condition. The ψ_j converge towards zero if and only if $|\alpha| < 1$. This corresponds to the stability condition of the AR term. Thus, the ARMA(1,1) process is stationary if, with stochastic initial conditions, it has a stable AR(1) term. The Wold representation is

$$(2.43) \quad x_t = \frac{\delta}{1-\alpha} + u_t + (\alpha - \beta)\, u_{t-1} + \alpha(\alpha - \beta)\, u_{t-2} + \alpha^2(\alpha - \beta)\, u_{t-3} + \dots .$$

Thus, the ARMA(1,1) process can be written as an MA(∞) process.

To invert the MA(1) part, $|\beta| < 1$ must hold. Starting from (2.42') leads to

$$u_t = \frac{-\delta}{1-\beta} + \frac{1-\alpha L}{1-\beta L}\, x_t .$$

If $1/(1-\beta L)$ is developed into a geometric series we get

$$\begin{aligned}
u_t &= \frac{-\delta}{1-\beta} + (1-\alpha L)(1 + \beta L + \beta^2 L^2 + \dots)\, x_t \\
&= \frac{-\delta}{1-\beta} + x_t + (\beta - \alpha)\, x_{t-1} + \beta(\beta - \alpha)\, x_{t-2} + \beta^2(\beta - \alpha)\, x_{t-3} + \dots .
\end{aligned}$$

This proves to be an AR(∞) representation. It shows that the combination of an AR(1) and an MA(1) term leads to a process with both MA(∞) and AR(∞) representation if the AR term is stable and the MA term invertible.

We obtain the first and second order moments of the stationary process in (2.42) as follows:

$$\begin{aligned} E[x_t] &= E[\delta + \alpha x_{t-1} + u_t - \beta u_{t-1}] \\ &= \delta + \alpha E[x_{t-1}] . \end{aligned}$$

Due to $E[x_t] = E[x_{t-1}] = \mu$, we get

$$\mu = \frac{\delta}{1-\alpha} ,$$

i.e. the expectation is the same as in an AR(1) process.

If we set $\delta = 0$ without loss of generality, the expectation is zero. The autocovariance of order $\tau \geq 0$ can then be written as

$$E[x_{t-\tau}x_t] = E[x_{t-\tau}(\alpha x_{t-1} + u_t - \beta u_{t-1})], \tag{2.44}$$

which leads to

$$\gamma(0) = \alpha \gamma(1) + E[x_t u_t] - \beta E[x_t u_{t-1}]$$

for $\tau = 0$. Due to (2.43), $E[x_t u_t] = \sigma^2$ and $E[x_t u_{t-1}] = (\alpha - \beta) \sigma^2$. Thus, we can write

$$\gamma(0) = \alpha \gamma(1) + (1 - \beta(\alpha - \beta)) \sigma^2. \tag{2.45}$$

(2.44) leads to

$$\gamma(1) = \alpha \gamma(0) + E[x_{t-1} u_t] - \beta E[x_{t-1} u_{t-1}]$$

for $\tau = 1$. Because of (2.43) this can be written as

$$\gamma(1) = \alpha \gamma(0) - \beta \sigma^2 . \tag{2.46}$$

If we insert (2.46) in (2.45) and solve for $\gamma(0)$, the resulting variance of the ARMA(1,1) process is

$$\gamma(0) = \frac{1+\beta^2-2\alpha\beta}{1-\alpha^2} \sigma^2. \tag{2.47}$$

Inserting this into (2.46), we get

$$\gamma(1) = \frac{(\alpha-\beta)(1-\alpha\beta)}{1-\alpha^2} \sigma^2 \tag{2.48}$$

for the first order autocovariance. For $\tau \geq 2$, (2.44) results in the autocovariances

$$\gamma(\tau) = \alpha\,\gamma(\tau\text{-}1) \tag{2.49}$$

and the autocorrelations

$$\rho(\tau) = \alpha\,\rho(\tau\text{-}1)\,. \tag{2.50}$$

This results in the same difference equation as in an AR(1) process but, however, with the different initial condition

$$\rho(1) = \frac{(\alpha-\beta)(1-\alpha\beta)}{1+\beta^2-2\alpha\beta}\,.$$

The first order autocorrelation coefficient is influenced by the MA term, while the higher order autocorrelation coefficients develop in the same way as in an AR(1) process.

If the process is stable and invertible, i.e. for $|\alpha| < 1$ and $|\beta| < 1$, the sign of $\rho(1)$ is determined by the sign of $(\alpha - \beta)$ because of $(1 + \beta^2 - 2\alpha\beta) > 0$ and $(1 - \alpha\beta) > 0$. Moreover, it follows from (2.49) that the autocorrelation function – as in the AR(1) process – is monotonic for $\alpha > 0$ and oscillating for $\alpha < 0$. Due to $|\alpha| < 1$ with τ increasing, the autocorrelation function also decreases in absolute value.

Thus, the following typical autocorrelation structures are possible:

(i) $\alpha > 0$ and $\alpha > \beta$: The autocorrelation function is always positive.

(ii) $\alpha < 0$ and $\alpha < \beta$: The autocorrelation function oscillates; the initial condition $\rho(1)$ is negative.

(iii) $\alpha > 0$ and $\alpha < \beta$: The autocorrelation function is negative from $\rho(1)$ onwards.

(iv) $\alpha < 0$ and $\alpha > \beta$: The autocorrelation function oscillates; the initial condition $\rho(1)$ is positive.

Figure 2.9 shows the development of the corresponding autocorrelation functions up to $\tau = 20$ for the parameter values $\alpha, \beta \in \{0.8, 0.5, -0.5, -0.8\}$ in which, of course, $\alpha \neq \beta$ must always hold, as otherwise the ARMA(1,1) process degenerates to a pure random process.

For the partial autocorrelation function we get

$$\phi_{11} = \rho(1) = \frac{(\alpha-\beta)(1-\alpha\beta)}{1+\beta^2-2\alpha\beta}\,,$$

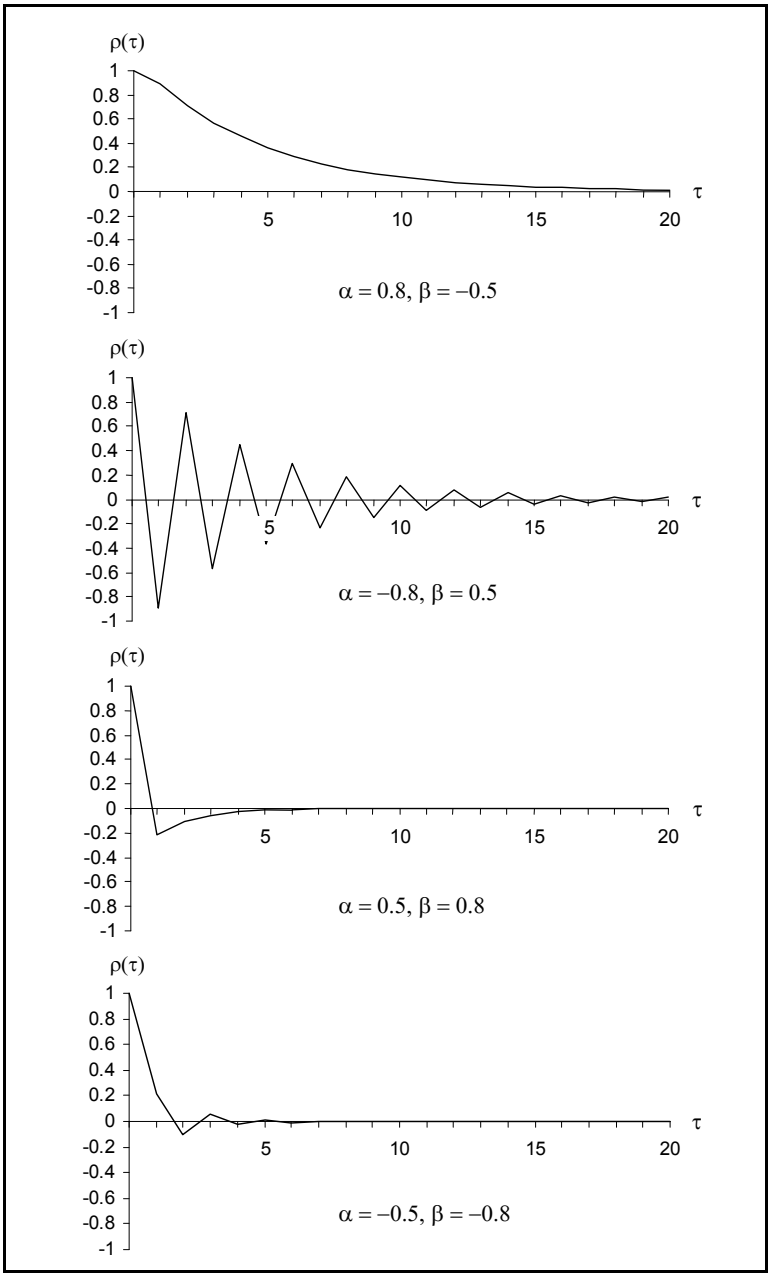

Figure 2.9: Theoretical autocorrelation functions of ARMA(1,1) processes

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} = \frac{\rho(1)(\alpha-\rho(1))}{1-\rho(1)^2},$$

because of $\rho(2) = \alpha\,\rho(1)$,

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \alpha\rho(1) \\ \alpha\rho(1) & \rho(1) & \alpha^2\rho(1) \end{vmatrix}}{1+2\alpha\rho(1)^3-\rho(1)^2(2+\alpha^2)}$$

$$= \frac{\rho(1)(\alpha-\rho(1))^2}{1+2\alpha\rho(1)^3-\rho(1)^2(2+\alpha^2)}, \text{ etc.}$$

Thus, the ARMA(1,1) process is a stationary stochastic process where neither the autocorrelation nor the partial autocorrelation function breaks off.

The following example shows how, due to measurement error, an AR(1)-process becomes an ARMA(1,1) process.

Example 2.14

The 'true' variable $\tilde{x}_t$ is generated by a stationary AR(1) process,

(E2.8) $$\tilde{x}_t = \alpha\,\tilde{x}_{t-1} + u_t,$$

but it can only be measured with an error v_t, i.e. for the observed variable x_t it holds that

(E2.9) $$x_t = \tilde{x}_t + v_t,$$

where v_t is a pure random process uncorrelated with the random process u_t. (The same model was used in *Example 2.3* but with a different interpretation.) If we transform (E2.8) to

$$\tilde{x}_t = \frac{u_t}{1-\alpha L}$$

and insert it into (E2.9) we get

$$(1-\alpha L)\,x_t = u_t + v_t - \alpha\,v_{t-1}.$$

For the combined error term $\zeta_t = u_t + v_t - \alpha\, v_{t-1}$ we get

$$\gamma_\zeta(0) = \sigma_u^2 + (1+\alpha^2)\,\sigma_v^2$$

$$\gamma_\zeta(1) = -\alpha\,\sigma_v^2$$

$$\gamma_\zeta(\tau) = 0 \quad \text{for} \quad \tau \geq 2,$$

or

$$\rho_\zeta(1) = \frac{-\alpha\sigma_v^2}{\sigma_u^2 + (1+\alpha^2)\,\sigma_v^2}, \quad \rho_\zeta(\tau) = 0 \quad \text{for} \quad \tau \geq 2.$$

Thus, the observable variable x_t follows an ARMA(1,1) process,

$$(1 - \alpha\, L)\, x_t = (1 - \beta\, L)\, \eta_t ,$$

where β can be calculated by means of $\rho_\zeta(1)$ and η_t is a pure random process. (See also the corresponding results in *Section 2.2.1*.)

2.3.2 ARMA(p,q) Processes

The general *autoregressive moving average process* with AR order p and MA order q can be written as

$$(2.51)\quad x_t = \delta + \alpha_1\, x_{t-1} + \ldots + \alpha_p\, x_{t-p} + u_t - \beta_1\, u_{t-1} - \ldots - \beta_q\, u_{t-q} ,$$

with u_t being a pure random process and $\alpha_p \neq 0$ and $\beta_q \neq 0$ having to hold. Using the lag operator, we can write

$$(2.51')\quad (1 - \alpha_1 L - \ldots - \alpha_p L^p)\, x_t = \delta + (1 - \beta_1 L - \ldots - \beta_q L^q)\, u_t ,$$

or

$$(2.51'')\quad \alpha(L)\, x_t = \delta + \beta(L)\, u_t .$$

As factors that are common in both polynomials can be reduced, $\alpha(L)$ and $\beta(L)$ cannot have identical roots. The process is stationary if – with stochastic initial conditions – the stability conditions of the AR term are fulfilled, i.e. if $\alpha(L)$ only has roots that are larger than 1 in absolute value. Then we can derive the Wold representation for which

$$\beta(L) = \alpha(L)(1 + \psi_1 L + \psi_2 L^2 + \ldots)$$

must hold. Again, the ψ_j, $j = 1, 2, \ldots$, can be calculated by comparing the coefficients. If, likewise, all roots of $\beta(L)$ are larger than 1 in absolute value, the ARMA(p,q) process is also invertible.

A stationary and invertible ARMA(p,q) process may either be represented as an AR(∞) or as an MA(∞) process. Thus, neither its autocorrela-

tion nor its partial autocorrelation function breaks off. In short, it is possible to generate stationary stochastic processes with infinite AR and MA orders by using only a finite number of parameters.

Under the assumption of stationarity, (2.51) directly results in the constant mean

$$E[x_t] = \mu = \frac{\delta}{1-\alpha_1-\ldots-\alpha_p} .$$

If, without loss of generality, we set $\delta = 0$ and thus also $\mu = 0$, we get the following relation for the autocovariances:

$$\begin{aligned}\gamma(\tau) &= E[x_{t-\tau}x_t] \\ &= E[x_{t-\tau}(\alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + u_t - \beta_1 u_{t-1} - \ldots - \beta_q u_{t-q})] .\end{aligned}$$

This relation can also be written as

$$\begin{aligned}\gamma(\tau) = {} & \alpha_1\gamma(\tau\text{-}1) + \alpha_2\gamma(\tau\text{-}2) + \ldots + \alpha_p\gamma(\tau\text{-}p) \\ & + E[x_{t-\tau}u_t] - \beta_1 E[x_{t-\tau}u_{t-1}] - \ldots - \beta_q E[x_{t-\tau}u_{t-q}] .\end{aligned}$$

Due to the Wold representation, the covariances between $x_{t-\tau}$ and u_{t-i}, $i = 0, \ldots, q$, are zero for $\tau > q$, i.e. the autocovariances for $\tau > q$ and $\tau > p$ are generated by the difference equation of an AR(p) process,

$$\gamma(\tau) - \alpha_1\gamma(\tau\text{-}1) - \alpha_2\gamma(\tau\text{-}2) - \ldots - \alpha_p\gamma(\tau\text{-}p) = 0 \quad \text{for } \tau > q \wedge \tau > p$$

whereas the first q autocovariances are also influenced by the MA part. Normalisation with $\gamma(0)$ leads to exactly the same results for the autocorrelations.

If the orders p and q are given and the distribution of the white noise process u_t is known, the parameters of an ARMA(p,q) process can be estimated consistently by using maximum likelihood methods. These estimates are also asymptotically efficient. If there is no such programme available, it is possible to estimate the parameters consistently with least squares. As every invertible ARMA(p,q) process is equivalent to an AR(∞) process, first of all an AR(k) process is estimated with k sufficiently larger than p. From this, one can get estimates of the non-observable residuals $\hat{u}_t$. By employing these residuals, the ARMA(p,q) process can be estimated with the least squares method,

$$x_t = \delta + \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} - \beta_1 \hat{u}_{t-1} - \ldots - \beta_q \hat{u}_{t-q} + v_t .$$

This approach can also be used if p and q are unknown. These orders can, for example, be determined by using the information criteria shown in *Section 2.1.5*.

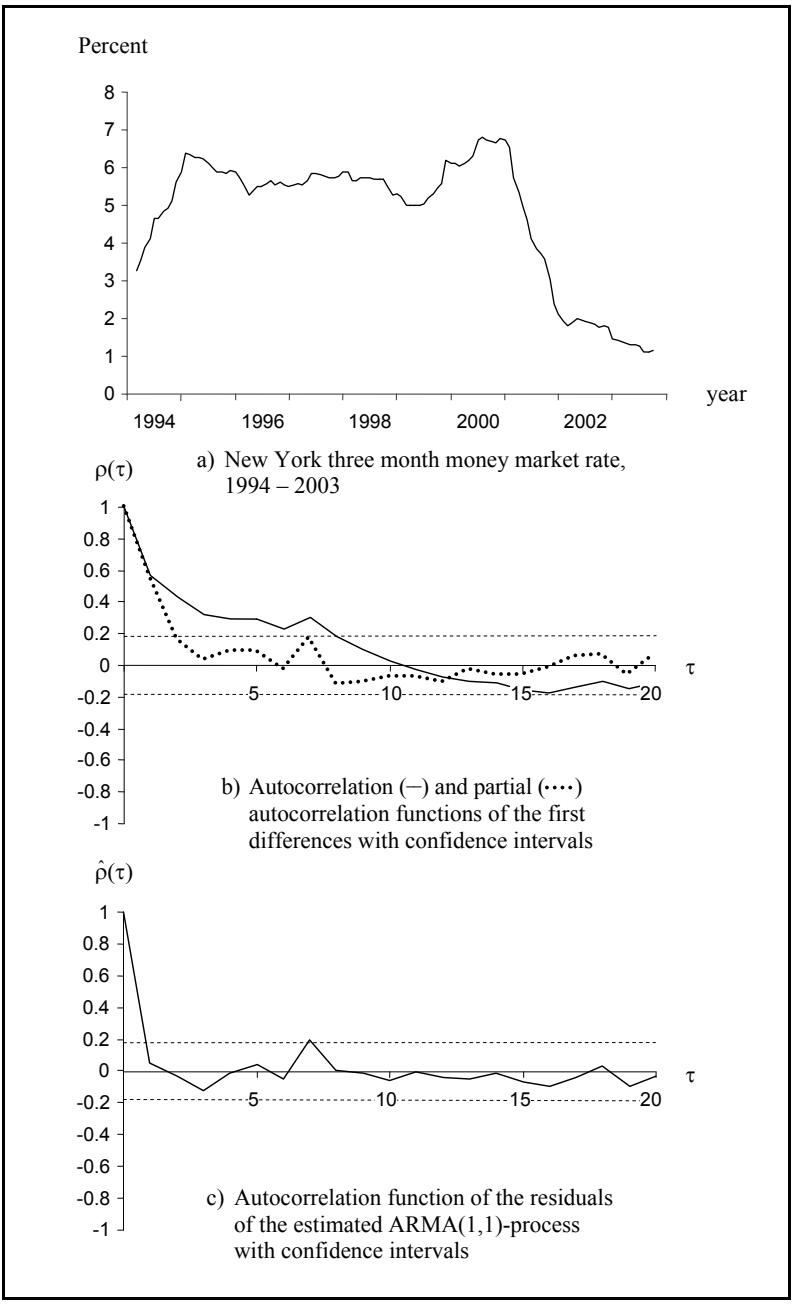

Figure 2.10: Three month money market rate in New York, 1994 – 2003

Example 2.15

Figure 2.10 shows the development of the US three month money market rate (USR) as well as the estimated autocorrelation and partial autocorrelation function of the first differences of this time series for the period from March 1994 to August 2003 (114 observations). Both functions do not show a clear break-off behaviour. Therefore, the following ARMA(1,1) model has been estimated for this time series:

$$\begin{array}{rcccccc} \Delta USR_t & = & -0.006 & + & 0.831\,\Delta USR_{t-1} & + & \hat{u}_t - 0.457\,\hat{u}_{t-1},. \\ & & (-0.73) & & (10.91) & & (-3.57) \end{array}$$

$$\bar{R}^2 = 0.351, \quad SE = 0.166, \quad Q(10) = 7.897 \ (p = 0.639).$$

The AR(1) as well as the MA(1) terms are different from zero and from one at any usual significance level. The autocorrelogram of the estimated residuals, which is also given in *Figure 2.10*, as well as the Ljung-Box Q statistic, which is calculated for this model with 12 autocorrelation coefficients (i.e. with 10 degrees of freedom), do not provide any evidence of a higher order process.

2.4 Forecasting

As mentioned in the introduction, in the 1970's, one of the reasons for the broad acceptance of time series analysis using the Box-Jenkins approach was the fact that forecasts with this comparably simple method often outperformed forecasts generated by large econometric models. In the following, we show how ARMA models can be used for making forecasts about the future development of time series. In doing so, we assume that all observations of the time series up to time t are known.

2.4.1 Forecasts with Minimal Mean Squared Errors

We want to solve the problem of making a τ-step ahead forecast for x_t with a *linear prediction function*, given a stationary and/or invertible data generating process.

Let $\hat{x}_t(\tau)$ be such a prediction function for $x_{t+\tau}$. Thus, $\hat{x}_t(\tau)$ is a random variable for given t and τ. As all stationary ARMA processes have a Wold representation, we assume the existence of such a representation without loss of generality. Thus,

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

where u_t is a pure random process with the usual properties $E[u_t] = 0$,

$$E[u_t u_s] = \begin{cases} \sigma^2 & \text{for} \quad t = s \\ 0 & \text{for} \quad t \neq s \end{cases}.$$

Therefore, it also holds that

$$x_{t+\tau} = \mu + \sum_{j=0}^{\infty} \psi_j u_{t+\tau-j}, \quad \tau = 1, 2, \ldots . \tag{2.52}$$

For a linear prediction function with the information given up to time t, we assume the following representation

$$\hat{x}_t(\tau) = \mu + \sum_{k=0}^{\infty} \theta_k^{\tau} u_{t-k}, \quad \tau = 1, 2, \ldots , \tag{2.53}$$

where the θ_k^{τ}, k = 0, 1, 2, ..., τ = 1, 2, ..., are unknown. The forecast error of a τ-step forecast is $f_t(\tau) = x_{t+\tau} - \hat{x}_t(\tau)$, τ = 1, 2, ..., . In order to make a good forecast, these errors should be small. The expected quadratic forecast error $E[(x_{t+\tau} - \hat{x}_t(\tau))^2]$, which should be minimised, is used as the criterion to determine the unknowns θ_k^{τ}. Taking into account (2.52) and (2.53) we can write

$$E[f_t^2(\tau)] = E\left[\left(\sum_{j=0}^{\infty} \psi_j u_{t+\tau-j} - \sum_{k=0}^{\infty} \theta_k^{\tau} u_{t-k}\right)^2\right]$$

$$= E\left[\left(u_{t+\tau} + \psi_1 u_{t+\tau-1} + \ldots + \psi_{\tau-1} u_{t+1} + \sum_{k=0}^{\infty} (\psi_{\tau+k} - \theta_k^{\tau}) u_{t-k}\right)^2\right].$$

From this it follows that

$$E[f_t^2(\tau)] = \left(1 + \psi_1^2 + \ldots + \psi_{\tau-1}^2\right)\sigma^2 + \sigma^2 \sum_{k=0}^{\infty} \left(\psi_{\tau+k} - \theta_k^{\tau}\right)^2 . \tag{2.54}$$

The variance of the forecast error reaches its minimum if we set $\theta_k^{\tau} = \psi_{\tau+k}$ for k = 0, 1, 2, ..., . Thus, we get the optimal linear prediction function for a τ-step ahead forecast from (2.53), as

$$\hat{x}_t(\tau) = \mu + \sum_{k=0}^{\infty} \psi_{\tau+k} u_{t-k}, \quad \tau = 1, 2, \ldots . \tag{2.55}$$

For the conditional expectation of u_{t+s}, given $u_t, u_{t-1}, \ldots$, it holds that

$$E[u_{t+s}|u_t, u_{t-1}, \ldots] = \begin{cases} u_{t+s} & \text{for } s \leq 0 \\ 0 & \text{for } s > 0 \end{cases}.$$

Thus, we get the conditional expectation of $x_{t+\tau}$, because of (2.52), as

$$E[x_{t+\tau}|u_t, u_{t-1}, \ldots] = \mu + \sum_{k=0}^{\infty} \psi_{\tau+k} u_{t-k}.$$

Due to (2.55), the conditional expectation of $x_{t+\tau}$, with all information available at time t given, is identical to the optimal prediction function. This leads to the following result: The conditional expectation of $x_{t+\tau}$, with all information up to time t given, provides the τ-step forecast with minimal mean squared prediction error.

With (2.52) and (2.55) the τ-step forecast error can be written as

$$(2.56) \quad f_t(\tau) = x_{t+\tau} - \hat{x}_t(\tau) = u_{t+\tau} + \psi_1 u_{t+\tau-1} + \psi_2 u_{t+\tau-2} + \ldots + \psi_{\tau-1} u_{t+1}$$

with

$$E[f_t(\tau)|u_t, u_{t-1}, \ldots] = E[f_t(\tau)] = 0.$$

From these results we can immediately draw some *conclusions*:

1. Best linear unbiased predictions (BLUP) of stationary ARMA processes are given by the conditional expectation for $x_{t+\tau}$, $\tau = 1,2, \ldots$

$$\hat{x}_t(\tau) = E[x_{t+\tau}|x_t, x_{t-1}, \ldots] = E_t[x_{t+\tau}].$$

2. For the one-step forecast errors ($\tau = 1$), $f_t(1) = u_{t+1}$, we get

$$E[f_t(1)] = E[u_{t+1}] = 0, \text{ and}$$

$$E[f_t(1)f_s(1)] = E[u_{t+1}u_{s+1}] = \begin{cases} \sigma^2 & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases}.$$

The one-step forecast errors are a pure random process; they are identical with the residuals of the data generating process. If the one-step prediction errors were correlated, the prediction could be improved by using the information contained in the prediction errors. In such a case, however, $\hat{x}_t(1)$ would not be an optimal forecast.

3. For the τ-step forecast errors ($\tau > 1$) we get

$$f_t(\tau) = u_{t+\tau} + \psi_1 u_{t+\tau-1} + \psi_2 u_{t+\tau-2} + \ldots + \psi_{\tau-1} u_{t+1},$$

i.e. they follow a MA(τ-1) process with $E[f_t(\tau)] = 0$ and the variance

$$V[f_t(\tau)] = \left(1 + \psi_1^2 + \ldots + \psi_{\tau-1}^2\right) \sigma^2 \quad . \tag{2.57}$$

This variance can be used for constructing confidence intervals for τ-step forecasts. However, these intervals are too narrow for practical applications because they do not take into account the uncertainty in the estimation of the parameters ψ_i, i = 1, 2, ..., τ-1.

4. It follows from (2.57) that the forecast error variance increases monotonically with increasing forecast horizon τ:

$$V[f_t(\tau)] \geq V[f_t(\tau\text{-}1)] \, .$$

5. Due to (2.57) we get for the limit

$$\lim_{\tau\to\infty} V[f_t(\tau)] = \lim_{\tau\to\infty}\left(1 + \psi_1^2 + \ldots + \psi_{\tau-1}^2\right)\sigma^2 = \sigma^2 \sum_{j=0}^{\infty}\psi_j^2 = V[x_t] \, ,$$

i.e. the variance of the τ-step forecast error is not larger than the variance of the underlying process.

6. The following variance decomposition follows from (2.55) and (2.56):

$$V[x_{t+\tau}] = V[\hat{x}_t(\tau)] + V[f_t(\tau)] \, . \tag{2.58}$$

7. Furthermore,

$$\lim_{\tau\to\infty} \hat{x}_t(\tau) = \lim_{\tau\to\infty}\left(\mu + \sum_{k=0}^{\infty}\psi_{\tau+k}u_{t-k}\right) = \mu = E[x_t] \, ,$$

i.e. for increasing forecast horizons, the forecasts converge to the (unconditional) mean of the series.

The concept of 'weak' rational expectations whose information set is restricted to the current and past values of a variable exactly corresponds to the optimal prediction approach used here.

2.4.2 Forecasts of ARMA(p,q) Processes

The Wold decomposition employed in the previous section has advantages when it comes to the derivation of theoretical results, but it is not practically useful for forecasting. Thus, in the following, we will discuss forecasts directly using AR, MA, or ARMA representations.

Forecasts with a Stationary AR(1) Process

For this process, it holds that

$$x_t = \delta + \alpha\, x_{t-1} + u_t\,,$$

with $|\alpha| < 1$. The optimal τ-step forecast is the conditional mean of $x_{t+\tau}$, i.e.

$$E_t[x_{t+\tau}] = E_t[\delta + \alpha\, x_{t+\tau-1} + u_{t+\tau}] = \delta + \alpha\, E_t[x_{t+\tau-1}]\,.$$

Due to the *first conclusion*, we get the following first order difference equation for the prediction function

$$\hat{x}_t(\tau) = \delta + \alpha\, \hat{x}_t\,(\tau\text{-}1)\,,$$

which can be solved recursively:

$$\tau = 1: \quad \hat{x}_t(1) = \delta + \alpha\, \hat{x}_t(0) = \delta + \alpha\, x_t$$

$$\tau = 2: \quad \hat{x}_t(2) = \delta + \alpha\, \hat{x}_t(1) = \delta + \alpha\,\delta + \alpha^2 x_t$$

$$\vdots$$

$$\hat{x}_t(\tau) = \delta(1 + \alpha + \ldots + \alpha^{\tau-1}) + \alpha^\tau x_t$$

$$\hat{x}_t(\tau) = \frac{1-\alpha^\tau}{1-\alpha}\,\delta + \alpha^\tau x_t = \frac{\delta}{1-\alpha} + \alpha^\tau\left(x_t - \frac{\delta}{1-\alpha}\right).$$

As $\mu = \delta/(1-\alpha)$ is the mean of a stationary AR(1) process,

$$\hat{x}_t(\tau) = \mu + \alpha^\tau (x_t - \mu) \quad \text{with} \quad \lim_{\tau\to\infty} \hat{x}_t(\tau) = \mu\,,$$

i.e., with increasing forecast horizon τ, the predicted values of an AR(1) process converge geometrically to the unconditional mean μ of the process. The convergence is monotonic if α is positive, and oscillating if α is negative.

To calculate the τ-step prediction error, the Wold representation, i.e. the MA(∞) representation of the AR(1) process, can be used,

$$x_t = \mu + u_t + \alpha\, u_{t-1} + \alpha^2 u_{t-2} + \alpha^3 u_{t-3} + \ldots\,.$$

Due to (2.56) and (2.57) we get the MA(τ-1) process

$$f_t(\tau) = u_{t+\tau} + \alpha\, u_{t+\tau-1} + \alpha^2 u_{t+\tau-2} + \ldots + \alpha^{\tau-1} u_{t+1}$$

for the forecast error with the variance

$$V[f_t(\tau)] = \left(1 + \alpha^2 + \ldots + \alpha^{2(\tau-1)}\right)\sigma^2 = \frac{1-\alpha^{2\tau}}{1-\alpha^2}\,\sigma^2.$$

With increasing forecast horizons, it follows that

$$\lim_{\tau\to\infty} V[f_t(\tau)] \;=\; \frac{\sigma^2}{1-\alpha^2} \;=\; V[x_t]\,,$$

i.e. the prediction error variance converges to the variance of the AR(1) process.

Forecasts with Stationary AR(p) Processes

Starting with the representation

$$x_t \;=\; \delta \;+\; \alpha_1\, x_{t-1} \;+\; \alpha_2\, x_{t-2} \;+\; \ldots \;+\; \alpha_p\, x_{t-p} \;+\; u_t\,,$$

the conditional mean of $x_{t+\tau}$ is given by

$$E_t[x_{t+\tau}] \;=\; \delta \;+\; \alpha_1\, E_t[x_{t+\tau-1}] \;+\; \ldots \;+\; \alpha_p\, E_t[x_{t+\tau-p}]\,.$$

Here,

$$E_t[x_{t+s}] \;=\; \begin{cases} \hat{x}_t(s) & \text{for} \quad s > 0 \\ x_{t+s} & \text{for} \quad s \le 0 \end{cases}\,.$$

Thus, the above difference equation can be solved recursively:

$$\tau = 1\text{:}\quad \hat{x}_t(1) \;=\; \delta \;+\; \alpha_1\, x_t \;+\; \alpha_2\, x_{t-1} \;+\; \ldots + \;\alpha_p\, x_{t+1-p}$$

$$\tau = 2\text{:}\quad \hat{x}_t(2) \;=\; \delta \;+\; \alpha_1\, \hat{x}_t(1) \;+\; \alpha_2\, x_t \;+\; \ldots \;+\; \alpha_p\, x_{t+2-p}\,,\ \text{etc.}$$

Forecasts with an Invertible MA(1) Process

For this process, it holds that

$$x_t \;=\; \mu \;+\; u_t \;-\; \beta\, u_{t-1}$$

with $|\beta| < 1$. The conditional mean of $x_{t+\tau}$ is

$$E_t[x_{t+\tau}] \;=\; \mu \;+\; E_t[u_{t+\tau}] \;-\; \beta\, E_t[u_{t+\tau-1}]\,.$$

For $\tau = 1$, this leads to

$$\hat{x}_t(1) \;=\; \mu \;-\; \beta\, u_t\,, \tag{2.59}$$

and for $\tau \ge 2$, we get

$$\hat{x}_t(\tau) \;=\; \mu\,,$$

i.e. the unconditional mean is the optimal forecast of $x_{t+\tau}$, $\tau = 2, 3, \ldots,$. For the τ-step prediction errors and their variances we get:

$$
\begin{aligned}
f_t(1) &= u_{t+1}, & V[f_t(1)] &= \sigma^2 \\
f_t(2) &= u_{t+2} - \beta\, u_{t+1}, & V[f_t(2)] &= (1+\beta^2)\,\sigma^2 \\
&\vdots & &\vdots \\
f_t(\tau) &= u_{t+\tau} - \beta\, u_{t+\tau-1}, & V[f_t(\tau)] &= (1+\beta^2)\,\sigma^2 .
\end{aligned}
$$

To be able to perform the one-step forecasts (2.59), the unobservable variable u has to be expressed as a function of the observable variable x. To do this, it must be taken into account that for $s \leq t$, the one-step forecast errors can be written as

$$u_s = x_s - \hat{x}_{s-1}(1). \tag{2.60}$$

For t = 0, we get from (2.59)

$$\hat{x}_0(1) = \mu - \beta\, u_0$$

with the non-observable but fixed u_0. Taking (2.60) into account, we get for t = 1

$$
\begin{aligned}
\hat{x}_1(1) = \mu - \beta\, u_1 &= \mu - \beta\,(x_1 - \hat{x}_0\,(1)) \\
&= \mu - \beta\, x_1 + \beta\,(\mu - \beta\, u_0) \\
&= \mu(1+\beta) - \beta\, x_1 - \beta^2\, u_0 .
\end{aligned}
$$

Correspondingly, we get for t = 2

$$
\begin{aligned}
\hat{x}_2(1) = \mu - \beta\, u_2 &= \mu - \beta\,(x_2 - \hat{x}_1\,(1)) \\
&= \mu - \beta\, x_2 + \beta(\mu(1+\beta) - \beta\, x_1 - \beta^2\, u_0) \\
&= \mu(1+\beta+\beta^2) - \beta\, x_2 - \beta^2\, x_1 - \beta^3\, u_0 .
\end{aligned}
$$

If we continue this procedure, the so-called *backcasting*, we finally arrive at a representation of the one-step prediction which – except for u_0 – consists only of observable terms,

$$\hat{x}_t(1) = \mu(1+\beta+\ldots+\beta^t) - \beta\, x_t - \beta^2\, x_{t-1} - \ldots - \beta^t\, x_1 - \beta^{t+1}\, u_0 .$$

Due to the invertibility of the MA(1) process, i.e. for $|\beta| < 1$, the impact of the unknown initial value u_0 finally disappears.

Similarly, one can show that, after q forecast steps, the optimal forecasts of invertible MA(q) processes, $q > 1$ are equal to the unconditional mean of the process and that the variance of the forecast errors is equal to the variance of the underlying process. The forecasts in observable terms are represented similarly to those of the MA(1) process.

Forecasts with ARMA(p,q) Processes

Forecasts for these processes result from combining the approaches of pure AR and MA processes. Thus, the one-step ahead forecast for a stationary and invertible ARMA(1,1) process is given by

$$\hat{x}_t(1) = \delta + \alpha x_t - \beta u_t.$$

Starting with t = 0 and taking (2.60) into account, forecasts are successively generated by backcasting. We first get

$$\hat{x}_0(1) = \delta + \alpha x_0 - \beta u_0,$$

where x_0 and u_0 are assumed to be any fixed numbers. For t = 1 we get

$$\hat{x}_1(1) = \delta + \alpha x_1 - \beta u_1 = \delta + \alpha x_1 - \beta(x_1 - \hat{x}_0(1))$$

$$= \delta(1 + \beta) + (\alpha - \beta) x_1 + \beta \alpha x_0 - \beta^2 u_0 ,$$

which finally leads to

$$\hat{x}_t(1) = \delta(1 + \beta + ... + \beta^t) + (\alpha - \beta) x_t + \beta(\alpha - \beta) x_{t-1} + ... \tag{2.61}$$

$$+ \beta^{t-1}(\alpha - \beta) x_1 + \beta^t \alpha x_0 - \beta^{t+1} u_0 .$$

Due to the invertibility condition, i.e. for $|\beta| < 1$, the one-step forecast for large values of t does no longer depend on the unknown initial values x_0 and u_0.

For the τ-step forecast, $\tau = 2, 3, ...$, we get

$$\hat{x}_t(2) = \delta + \alpha \hat{x}_t(1)$$

$$\hat{x}_t(3) = \delta + \alpha \hat{x}_t(2)$$

$$\vdots$$

Using (2.61), these forecasts can be calculated recursively.

2.4.3 Evaluation of Forecasts

Forecasts can be evaluated ex post, i.e. when the realised values are available. There are many kinds of measures to do this. Quite often, only graphs and/or scatter diagrams of the predicted values and the corresponding observed values of a time series are plotted. Intuitively, a forecast is 'good' if the predicted values describe the development of the series in the graphs relatively well or if the points in the scatter diagram are concentrated around the angle bisecting line in the first and/or third quadrant. Such intu-

itive arguments are, however, not founded on the above-mentioned considerations on optimal predictions. For example, as (2.59) shows, the optimal one-step forecast of a MA(1) process is a pure random process. This implies that the graphs compare two quite different processes. *Conclusion 6* given above states that the following decomposition holds for the variances of the data generating processes, the forecasts and the forecast errors,

$$V[x_{t+\tau}] \;=\; V[\,\hat{x}_t(\tau)] \;+\; V[f_t(\tau)]\,.$$

Thus, it is obvious that predicted and realised values are generally generated by different processes.

As a result, a measure for the *predictability* of stationary processes can be developed. It is defined as follows,

$$P(\tau)^2 \;=\; \frac{V[\hat{x}_t(\tau)]}{V[x_{t+\tau}]} \;=\; 1 \;-\; \frac{V[f_t(\tau)]}{V[x_{t+\tau}]}\,, \tag{2.62}$$

with $0 \le P(\tau)^2 \le 1$. At the same time, $P(\tau)^2$ is the correlation coefficient between the predicted and the realised values of x. The optimal forecast of a pure random process with mean zero is $\hat{x}_t(\tau) = 0$, i.e. $P(\tau)^2 = 0$. Such a process cannot be predicted. On the other hand, for the one-step forecast of a MA(1) process, we can write

$$P(1)^2 \;=\; \frac{\beta^2\;\sigma^2}{(1 \;+\; \beta^2)\sigma^2} \;=\; \frac{\beta^2}{1 \;+\; \beta^2} \;>\; 0.$$

However, the decomposition (2.58), theoretically valid for optimal forecasts, does not hold for actual (empirical) forecasts, even if they are generated by using (estimated) ARMA processes. This is due to the fact that forecast errors are hardly ever totally uncorrelated with the forecasts. Therefore, the value of $P(\tau)^2$ might even become negative for 'bad' forecasts.

JACOB MINCER and VICTOR ZARNOWITZ (1969) made the following suggestion to check the consistency of forecasts. By using OLS the following regression equation is estimated

$$x_{t+\tau} \;=\; a_0 \;+\; a_1\;\hat{x}_t(\tau) \;+\; \varepsilon_{t+\tau}. \tag{2.63}$$

It is tested either individually with t tests or commonly with a F test whether $a_0 = 0$ and $a_1 = 1$. If this is fulfilled, the forecasts are said to be consistent. However, such a regression produces consistent estimates of the parameters if and only if $\hat{x}_t(\tau)$ and $\varepsilon_{t+\tau}$ are asymptotically uncorrelated.

Moreover, to get consistent estimates of the variances, which is necessary for the validity of the test results, the residuals have to be pure random processes. Even under the null hypothesis of optimal forecasts, this only holds for one-step predictions. Thus, the usual F and t tests can only be used for $\tau = 1$. For $\tau > 1$, the MA(τ-1) process of the forecast errors has to be taken into account when the variances are estimated. A procedure for such situations combines Ordinary Least Squares for the estimation of the parameters and Generalised Least Squares for the estimation of the variances, as proposed by BRYAN W. BROWN and SHLOMO MAITAL (1981).

JINOOK JEONG and GANGADHARRAO S. MADDALA (1991) have pointed out another problem which is related to these tests. Even rational forecasts are usually not without errors; they contain measurement errors. This implies, however, that (2.63) cannot be estimated consistently with OLS; an instrumental variables estimator must be used. An alternative to the estimation of (2.63) is therefore to estimate a univariate MA(τ-1) model for the forecast errors of a τ-step prediction,

$$\hat{f}_t(\tau) = a_0 + u_t + a_1 u_{t-1} + a_2 u_{t-2} + \ldots + a_{\tau-1} u_{t-\tau+1} ,$$

and to check the null hypothesis H_0: $a_0 = 0$ and whether the estimated residuals $\hat{u}_t$ are white noise.

On the other hand, simple descriptive measures, which are often employed to evaluate the performance of forecasts, are based on the average values of the forecast errors over the forecast horizon. The simple arithmetic mean indicates whether the values of the variable are – on average – over- or underestimated. However, the disadvantage of this measure is that large over- and underestimates cancel each other out. The *mean absolute error* is often used to avoid this effect. Starting the forecasts from a fixed point of time, t_0, and assuming that realisations are available up to t_0+m, we get

$$\mathrm{MAE}(\tau) = \frac{1}{m + 1 - \tau} \sum_{j=0}^{m-\tau} \left| f_{t_0+j}(\tau) \right| , \quad \tau = 1, 2, \ldots .$$

Every forecast error gets the same weight in this measure. The *root mean square error* is often used to give particularly large errors a stronger weight:

$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{m+1-\tau} \sum_{j=0}^{m-\tau} f^2_{t_0+j}(\tau)} , \quad \tau = 1, 2, \ldots .$$

These measures are not normalised, i.e. their size depends on the scale of the data.

The inequality measure proposed by HENRY THEIL (1961) avoids this problem by comparing the actual forecasts with so-called naïve forecasts, i.e. the realised values of the last available observation,

$$U(\tau) = \sqrt{\frac{\sum_{j=0}^{m-\tau} f_{t_0+j}^2(\tau)}{\sum_{j=0}^{m-\tau} (x_{t_0+\tau+j} - x_{t_0+j})^2}}, \quad \tau = 1, 2, \ldots .$$

If $U(\tau) = 1$, the forecast is as good as the naïve forecast, $\hat{x}_t(\tau) = x_t$. For $U(\tau) < 1$ the forecasts perform better than the naïve one. MAE, RMSE and Theil's U all become zero if predicted and realised values are identical over the whole forecast horizon.

Example 2.16

All these measures can also be applied to forecasts which are not generated by ARMA models, as, for example, the forecasts of the Council of Economic Experts or the Association of German Economic Research Institutes. Since the end of the 1960's, both institutions have published forecasts of the German economic development for the following year, the institutes usually in October and the Council at the end of November. HANNS MARTIN HAGEN and GEBHARD KIRCHGÄSSNER (1996) investigated the annual forecasts of the growth rates of GNP for the period from 1970 to 1995 as well as for the sub-periods from 1970 to 1982 and from 1983 to 1995. These periods correspond to the social-liberal government of SPD and FDP and the conservative-liberal government of CDU/CSU and FDP.

The results are given in *Table 2.2*. Besides the criteria given above, the table also indicates the square of the correlation coefficient between realised and predicted values (R^2), the estimated regression coefficient $\hat{a}_1$ of the test equation (2.63) as well as the mean error (ME). According to almost all criteria, the forecasts of the Council outperform those of the institutes. This was to be expected, as the Council's forecasts are produced slightly later, at a time when more information is available. It holds for the forecasts of both institutions that the mean absolute error, the root mean squared error as well as Theil's U are smaller in the second period compared to the first one. This is some evidence that the forecasts might have improved over time. On the other hand, the correlation coefficient between predicted and realised values has also become smaller. This indicates a deterioration of the forecasts. It has to be taken into account that the variance of the variable to be predicted was considerably smaller in the second period as compared to the first one. Thus, the smaller errors do not necessarily indicate improvements of the forecasts. It is also interesting to note that on average the forecast errors of both institutions were negative in the first and positive in the second sub-period. They tended to overestimate the development in the period of the social-liberal coalition and to underestimate it in the period of the conservative-liberal coalition.

Table 2.2: Forecasts of the Council of Economic Experts and of the Economic Research Institutes

	Period	R^2	RMSE	MAE	ME	$\hat{a}_1$	U
Institutes	1970 – 1995	0.369	1.838	1.346	-0.250*	1.005*	0.572
	1970 – 1982	0.429	2.291	1.654	-0.731	1.193*	0.625
	1983 – 1995	0.399	1.229	1.038	0.231	1.081	0.457
Council of Economic Experts	1970 – 1995	0.502*	1.647*	1.171*	-0.256	1.114	0.512*
	1970 – 1982	0.599*	2.025*	1.477*	-0.723*	1.354	0.552*
	1983 – 1995	0.472*	1.150*	0.865*	0.212*	1.036*	0.428*

'*' denotes the 'better' of the two forecasts.

2.5 The Relation between Econometric Models and ARMA Processes

The ARMA model-based forecasts discussed in the previous section are *unconditional forecasts*. The only information that is used to generate these forecasts is the information contained in the current and past values of the time series. There is demand for such forecasts, and – as mentioned above – one of the reasons for the development and the popularity of the Box-Jenkins methodology presented in this chapter is that by applying the above-mentioned approaches, these predictions perform – at least partly – much better than forecasts generated by large scale econometric models. Thus, the Box-Jenkins methodology seems to be a (possibly much better) alternative to the traditional econometric methodology.

However, this perspective is rather restricted. On the one hand, conditional rather than unconditional forecasts are required in many cases, for example, in order to evaluate the effect of a tax reform on economic growth. Such forecasts cannot be generated by using (only) univariate models. On the other hand, and more importantly, the separation of the two approaches is much less strict than it seems to be at first glance. As ARNOLD ZELLNER and FRANZ C. PALM (1974) showed, linear dynamic simultaneous equation systems as used in traditional econometrics can be transformed into ARMA models. (Inversely, multivariate time series models as discussed in the next chapters can be transformed into traditional econometric models.) The univariate ARMA models correspond to the *fi-*

nal equations of econometric models in the terminology of JAN TINBERGEN (1940).

Let us consider a very simple model. An exogenous, weakly stationary variable x, as defined in (2.64b), has a current and lagged impact on the dependent variable y, while the error term might be autocorrelated. Thus, we get the model

$$y_t = \eta_1(L)\, x_t + \eta_2(L)\, u_{1,t}\,, \tag{2.64a}$$

$$\alpha(L)\, x_t = \beta(L)\, u_{2,t}\,, \tag{2.64b}$$

where $\eta_1(L)$ and $\eta_2(L)$ are lag polynomials of finite order. If we insert (2.64b) in (2.64a), we get for y the univariate model

$$\alpha(L)\, y_t = \zeta(L)\, v_t \tag{2.64a'}$$

with

$$\zeta(L)\, v_t := \eta_1(L)\, \beta(L)\, u_{2,t} + \eta_2(L)\, \alpha(L)\, u_{1,t}\,.$$

As $\zeta(L)v_t$ is an MA process of finite order, we get a finite order ARMA representation for y. It must be pointed out that the univariate representations of the two variables have the same finite order AR term.

References

Since the time when HERMAN WOLD developed the class of ARMA processes in his dissertation and GEORGE E.P. BOX and GWILYM M. JENKINS (1970) popularised and further developed this model class in the textbook mentioned above, there have been quite a lot of **textbooks** dealing with these models at different technical levels. An introduction focusing on empirical applications is, for example, to be found in

ROBERT S. PINDYCK and DANIEL L. RUBINFELD, *Econometric Models and Economic Forecasts*, McGraw-Hill, Boston et al., 4th edition 1998, Chapter 17f. pp. 521 – 578,

PETER J. BROCKWELL and RICHARD A. DAVIS, *Introduction to Time Series and Forecasting*, Springer, New York et al. 1996, as well as

TERENCE C. MILLS, *Time Series Techniques for Economists*, Cambridge University Press, Cambridge (England) 1990. Contrary to this,

PETER J. BROCKWELL and RICHARD A. DAVIS, *Time Series: Theory and Methods*, Springer, New York et al. 1987,

give a rigorous presentation in probability theory. Along with the respective proofs of the theorems, this textbook shows, however, many empirical examples.

Autoregressive processes for the residuals of an estimated regression equation were used for the first time in econometrics by

DONALD COCHRANE and GUY H. ORCUTT, Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Association* 44 (1949), pp. 32 – 61.

The different **information criteria** to detect the order of an autoregressive process are presented in

HIROTUGU AKAIKE, Fitting Autoregressive Models for Prediction, *Annals of the Institute of Statistical Mathematics* AC-19 (1974), pp. 364 – 385,

HIROTUGU AKAIKE, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control* 21 (1969), pp. 234 – 237,

GIDEON SCHWARZ, Estimating the Dimensions of a Model, *Annals of Statistics* 6 (1978), pp. 461 – 464, as well as in

EDWARD J. HANNAN and BARRY G. QUINN, The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society* B 41 (1979), pp. 190 – 195.

The effect of **temporal aggregation** on the first differences of temporal averages have first been investigated by

HOLBROOK WORKING, Note on the Correlation of First Differences of Averages in a Random Chain, *Econometrica* 28 (1960), pp. 916 – 918

and later on, in more detail, by

GEORGE C. TIAO, Asymptotic Behaviour of Temporal Aggregates of Time Series, *Biometrika* 59 (1972), pp. 525 – 531.

The approach to check the **consistency of predictions** was developed by

JACOB MINCER and VICTOR ZARNOWITZ, The Evaluation of Economic Forecasts, in: J. MINCER (ed.), *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York 1969.

The use of MA processes of the forecast errors to estimate the variances of the estimated parameters was presented by

BRYAN W. BROWN and SHLOMO MAITAL, What Do Economists Know? An Empirical Study of Experts' Expectations, *Econometrica* 49 (1981), pp. 491 – 504.

The fact that measurement errors also play a role in rational forecasts and that, therefore, instrumental variable estimators should be used, was indicated by

JINOOK JEONG and GANGADHARRAO S. MADDALA, Measurement Errors and Tests for Rationality, *Journal of Business and Economic Statistics* 9 (1991), pp. 431 – 439.

These procedures have been applied to the common forecasts of the German economic research institutes by

GEBHARD KIRCHGÄSSNER, Testing Weak Rationality of Forecasts with Different Time Horizons, *Journal of Forecasting* 12 (1993), pp. 541 – 558.

Moreover, the forecasts of the German Council of Economic Experts as well as those of the German Economic Research Institutes were investigated in

HANNS MARTIN HAGEN and GEBHARD KIRCHGÄSSNER, Interest Rate Based Forecasts of German Economic Growth: A Note, *Weltwirtschaftliches Archiv* 132 (1996), pp. 763 – 773.

The **measure of inequality** (Theil's U) was proposed by

HENRY THEIL, *Economic Forecasts and Policy*, North-Holland, Amsterdam 1961.

An alternative measure is given in

HENRY THEIL, *Applied Economic Forecasting*, North-Holland, Amsterdam 1966.

Today, both measures are used in computer programmes. Quite generally, **forecasts for time series data** are discussed in

CLIVE W.J. GRANGER, *Forecasting in Business and Economics*, Academic Press, 2nd edition 1989.

On the **evaluation of the predictive accuracy** of forecasts see

FRANCIS X. DIEBOLD and ROBERTO S. MARIANO, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13 (1995), pp. 253 – 263.

The **relationship between time series models and econometric equation systems** is analysed in

ARNOLD ZELLNER and FRANZ C. PALM, Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics* 2 (1974), pp. 17 – 54.

See for this also

FRANZ C. PALM, Structural Econometric Modeling and Time Series Analysis: An Integrated Approach, in: A. ZELLNER (ed.), Applied Time Series Analysis of Economic Data, U.S. Department of Commerce, Economic Research Report ER-S, Washington 1983, pp. 199 – 230.

The term **final equation** originates from

JAN TINBERGEN, Econometric Business Cycle Research, *Review of Economic Studies* 7 (1940), pp. 73 – 90.

An introduction into the solution of **difference equations** is given in

WALTER ENDERS, *Applied Econometric Time Series*, 3rd edition, Wiley, Hoboken, N.J. 2010, Chapter 1.

The **permanent income hypothesis** as a determinant of consumption expenditure was developed by

MILTON FRIEDMAN, *A Theory of the Consumption Function*, Princeton University Press, Princeton N.J. 1957.

The example of the **estimated popularity function** is given in

GEBHARD KIRCHGÄSSNER, Causality Testing of the Popularity Function: An Empirical Investigation for the Federal Republic of Germany, 1971 – 1982, *Public Choice* 45 (1985), pp. 155 – 173.

http://www.springer.com/978-3-642-33435-1

Introduction to Modern Time Series Analysis
Kirchgässner, G.; Wolters, J.; Hassler, U.
2013, XII, 320 p., Hardcover
ISBN: 978-3-642-33435-1