Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 2

Moving Average and Autoregressive Processes

In any study, one of the first undertakings is to describe the item under investigation. Typically there is no unique description and, as the discipline develops, alternatives appear that prove more useful in some applications.

The joint distribution function (1.2.1) gives a complete description of the distributional properties of a time series. The covariance function provides a less complete description, but one from which useful conclusions can be drawn. As we have noted, the correlation function has functional properties analogous to those of the characteristic function of a statistical distribution function. Therefore, the distribution function associated with the correlation function provides an alternative representation for a time series. This representation will be introduced in Chapter 3 and discussed in Chapter 4.

Another method of describing a time series is to express it as a function of more elementary time series. A sequence of independent identically distributed random variables is a very simple type of time series. In applications it is often possible to express the observed time series as a relatively simple function of this elementary time series. In particular, many time series have been represented as linear combinations of uncorrelated $(0, \sigma^2)$ or of independent $(0, \sigma^2)$ random variables. We now consider such representations.

2.1. MOVING AVERAGE PROCESSES

We shall call the time series $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$, defined by

$$X_t = \sum_{j=-M}^{M} \alpha_j e_{t-j}, \tag{2.1.1}$$

where M is a nonnegative integer, α_j are real numbers, $\alpha_{-M} \neq 0$, and the e_t are uncorrelated $(0, \sigma^2)$ random variables, a *finite moving average time series* or a *finite moving average process.*

If there exists no finite M such that $\alpha_j = 0$ for all j with absolute value greater than M, then $\{X_t\}$ is an *infinite moving average process* and is given the representation

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j} . \tag{2.1.2}$$

The time series defined by

$$X_t = \sum_{j=0}^{M} \alpha_j e_{t-j} , \tag{2.1.3}$$

where $\alpha_0 \neq 0$ and $\alpha_M \neq 0$, is called a *one-sided moving average* of *order M*. Note that the number of terms in the sum on the right-hand side of (2.1.3) is $M + 1$. In particular, (2.1.3) is a *left-moving average*, since nonzero weights are applied only to e_t and to e's with indexes less than t. We shall most frequently use the one-sided representation of moving average time series. Note that there would be no loss of generality in considering only the one-sided representation. If in the representation (2.1.1) we define the random variable ϵ_t by $\epsilon_t = e_{t+M}$, we have, for (2.1.1),

$$\begin{aligned} X_t &= \sum_{j=-M}^{M} \alpha_j e_{t-j} = \sum_{s=0}^{2M} \alpha_{s-M} \epsilon_{t-s} \\ &= \sum_{s=0}^{2M} \beta_s \epsilon_{t-s} , \end{aligned} \tag{2.1.4}$$

where $\beta_s = \alpha_{s-M}$.

Likewise we can, without loss of generality, take β_0 in (2.1.4) to be one. If β_0 is not one, we simply define

$$\begin{aligned} u_{t-s} &= \beta_0 \epsilon_{t-s} , \\ \delta_s &= \beta_s \beta_0^{-1} , \qquad s = 0, 1, 2, \ldots, 2M , \end{aligned}$$

and obtain the representation

$$X_t = \sum_{s=0}^{2M} \delta_s u_{t-s} , \tag{2.1.5}$$

where $\delta_0 = 1$ and the u_t are uncorrelated $(0, \beta_0^2 \sigma^2)$ random variables.

The covariance function of an Mth order moving average is distinctive in that it is zero for all h satisfying $|h| > M$. For the time series (2.1.3) we have

$$\begin{aligned} &E\{X_t\} = 0 , \\ &\gamma_X(h) = E\{X_t X_{t+h}\} = \begin{cases} \displaystyle\sum_{t=0}^{M-|h|} \alpha_t \alpha_{t+|h|} \sigma^2 , & 0 \leq |h| \leq M , \\ 0 , & |h| > M . \end{cases} \end{aligned} \tag{2.1.6}$$

As an example, consider the simple time series

$$X_t = \sum_{i=0}^{4} e_{t-i} .$$

Clearly,

$$E\{X_t\} = \sum_{i=0}^{4} E\{e_{t-i}\} = 0 \qquad \text{for all } t$$

and

$$E\{X_t X_{t+h}\} = E\left\{\left(\sum_{i=0}^{4} e_{t-i}\right)\left(\sum_{i=0}^{4} e_{t+h-i}\right)\right\}$$
$$= \begin{cases} (5 - |h|)\sigma^2 , & |h| \leq 4 , \\ 0 & \text{otherwise} . \end{cases}$$

By (2.1.6) only the first autocorrelation of a first order moving average is nonzero, while all autocorrelations of order greater than two for a second order moving average are zero. For a qth order moving average, $\rho(q) \neq 0$, and all higher order autocorrelations are zero. There are further restrictions on the correlation function of moving average time series. Consider the first order moving average

$$X_t = e_t + \alpha e_{t-1} . \tag{2.1.7}$$

Then

$$\rho(1) = \frac{\alpha}{1 + \alpha^2} . \tag{2.1.8}$$

Using elementary calculus, it is easy to prove that $\rho(1)$ achieves a maximum of 0.5 for $\alpha = 1$ and a minimum of -0.5 for $\alpha = -1$.

Thus, the first autocorrelation of a first order moving average process will always fall in the interval $[-0.5, 0.5]$. For any ρ in $(0, 0.5)$ [or in $(-0.5, 0)$] there are two α-values yielding such a ρ, one in the interval $(0, 1)$ [or in $(-1, 0)$] and one in the interval $(1, \infty)$ [or $(-\infty, -1)$].

Example 2.1.1. In Figure 2.1.1 we display 100 observations from a realization of the time series

$$X_t = e_t + 0.7e_{t-1} ,$$

where the e_t are normal independent $(0, 1)$ random variables. The positive correlation between adjacent observations is clear to the eye and is emphasized by the plot of X_t against X_{t-1} of Figure 2.1.2. The correlation between X_t and X_{t-1} is $(1.49)^{-1}(0.7) = 0.4698$, and this is the slope of the line plotted on the figure. On the other hand, since the e_t are independent, X_t is independent of X_{t-2}, and this is clear from the plot of X_t against X_{t-2} in Figure 2.1.3.

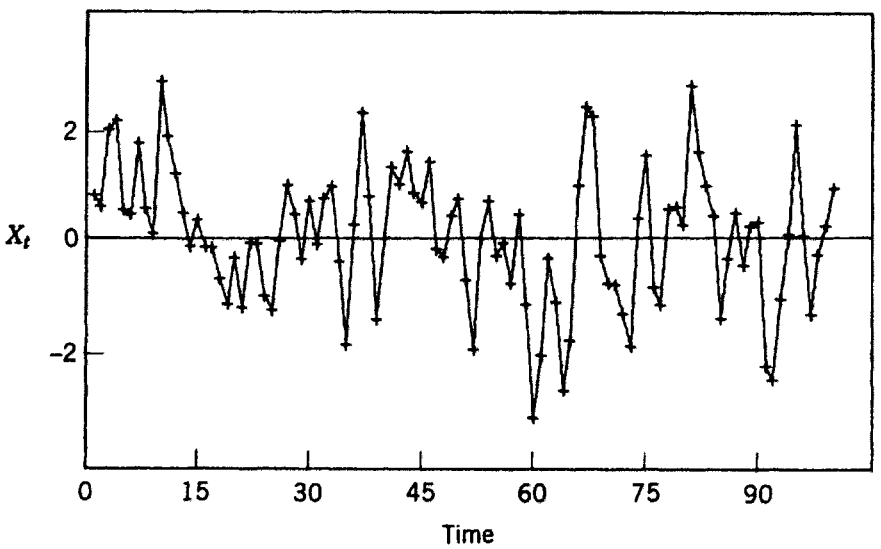

Figure 2.1.1. One hundred observations from the time series $X_t = e_t + 0.7e_{t-1}$.

While the simple correlation between X_t and X_{t-2} is zero, the partial correlation between X_t and X_{t-2} adjusted for X_{t-1} is not zero. The correlation between X_t and X_{t-1} is $\rho(1) = (1+\alpha^2)^{-1}\alpha$, which is the same as that between X_{t-1} and X_{t-2}. Therefore, the partial correlation between X_t and X_{t-2} adjusted for X_{t-1} is

$$\frac{\text{Cov}\{X_t - cX_{t-1}, X_{t-2} - cX_{t-1}\}}{\text{Var}\{X_t - cX_{t-1}\}} = -\frac{\alpha^2}{1+\alpha^2+\alpha^4},$$

where $c = (1+\alpha^2)^{-1}\alpha$.

In Figure 2.1.4, $X_t - 0.7(1.49)^{-1}X_{t-1}$ is plotted against $X_{t-2} - 0.7(1.49)^{-1}X_{t-1}$ for the time series of Figure 2.1.1. For this time series the theoretical partial autocorrelation is -0.283, and this is the slope of the line plotted in Figure 2.1.4. ▲▲

Example 2.1.2. One hundred observations from the second order moving average

$$X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$$

are displayed in Figure 2.1.5. The correlation function of this process is

$$\begin{aligned} \rho(h) &\doteq -0.65, & h &= 1, \\ &\doteq 0.15, & h &= 2, \\ &= 0, & h &\geq 3. \end{aligned}$$

Once again the nature of the correlation can be observed from the plot. Adjacent

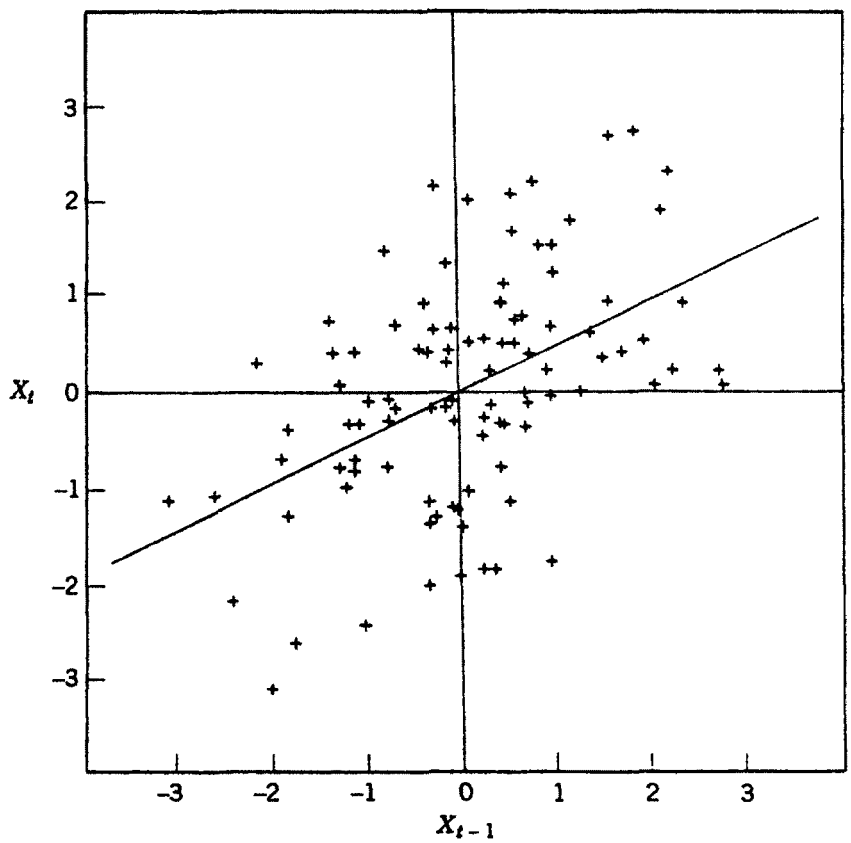

Figure 2.1.2. Plot of X_t against X_{t-1} for 100 observations from $X_t = e_t + 0.7e_{t-1}$.

observations are often on opposite sides of the mean, indicating that the first order autocorrelation is negative. This correlation is also apparent in Figure 2.1.6, where X_t is plotted against X_{t-1}. ▲▲

At the beginning of this section we defined a moving average time series to be a linear combination of a sequence of uncorrelated random variables. It is also common to speak of forming a moving average of an arbitrary time series X_t. Thus, Y_t defined by

$$Y_t = \sum_{i=-M}^{M} \alpha_i X_{t+i}$$

is a moving average of the time series X_t. The reader may prove:

Lemma 2.1.1. If $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ is a stationary time series with mean zero and covariance function $\gamma_X(h)$, then

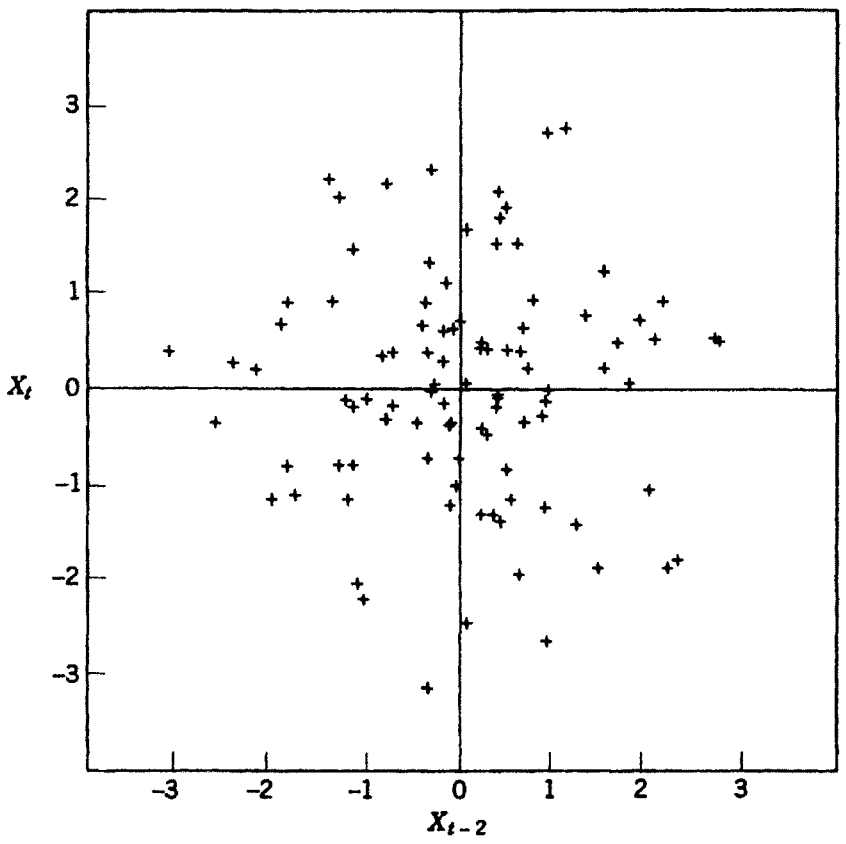

Figure 2.1.3. Plot of X_t against X_{t-2} for 100 observations from $X_t = e_t + 0.7e_{t-1}$.

$$Y_t = \sum_{i=-M}^{M} \alpha_i X_{t+i} ,$$

where M is finite and the α_i are real numbers, is a stationary time series with mean zero and covariance function

$$\gamma_Y(h) = \sum_{i=-M}^{M} \sum_{j=-M}^{M} \alpha_i \alpha_j \gamma_X(h + j - i) .$$

2.2. ABSOLUTELY SUMMABLE SEQUENCES AND INFINITE MOVING AVERAGES

Before investigating the properties of infinite moving average time series, we review some results on the convergence of partial sums of infinite sequences of real numbers.

Recall that an *infinite sequence* is a function whose domain is the set

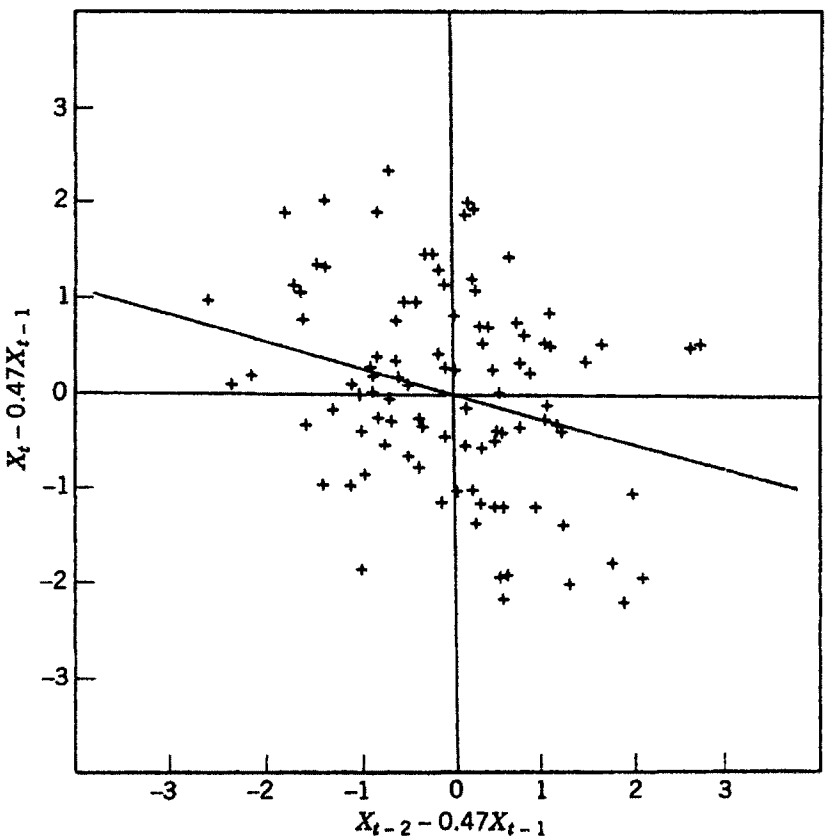

Figure 2.1.4. Plot of $X_t - 0.47X_{t-1}$ against $X_{t-2} - 0.47X_{t-1}$ for 100 observations from $X_t = e_t + 0.7e_{t-1}$.

$T_1 = (1, 2, \ldots)$ of natural numbers. A *doubly infinite sequence* is a function whose domain is the set $T_2 = (0, \pm 1, \pm 2, \ldots)$ of all integers. When no confusion will result, we shall use the word sequence for a function whose domain is either T_1 or T_2. We denote the infinite sequence by $\{a_j\}_{j=1}^{\infty}$ and the doubly infinite sequence by $\{a_j\}_{j=-\infty}^{\infty}$. When the domain is clear, the notation will be abbreviated to $\{a_j\}$ or perhaps to a_j.

The *infinite series* created from the sequence $\{a_j\}_{j=1}^{\infty}$ is the sequence of partial sums $\{s_j\}_{j=1}^{\infty}$ defined by

$$s_j = \sum_{i=1}^{j} a_i, \qquad j = 1, 2, \ldots. \tag{2.2.1}$$

The symbols $\sum_{j=1}^{\infty} a_j$ and $\Sigma\{a_j\}$ are often used to represent both the infinite sequence $\{s_j\}_{j=1}^{\infty}$ generated by the sequence $\{a_j\}$ and the limit

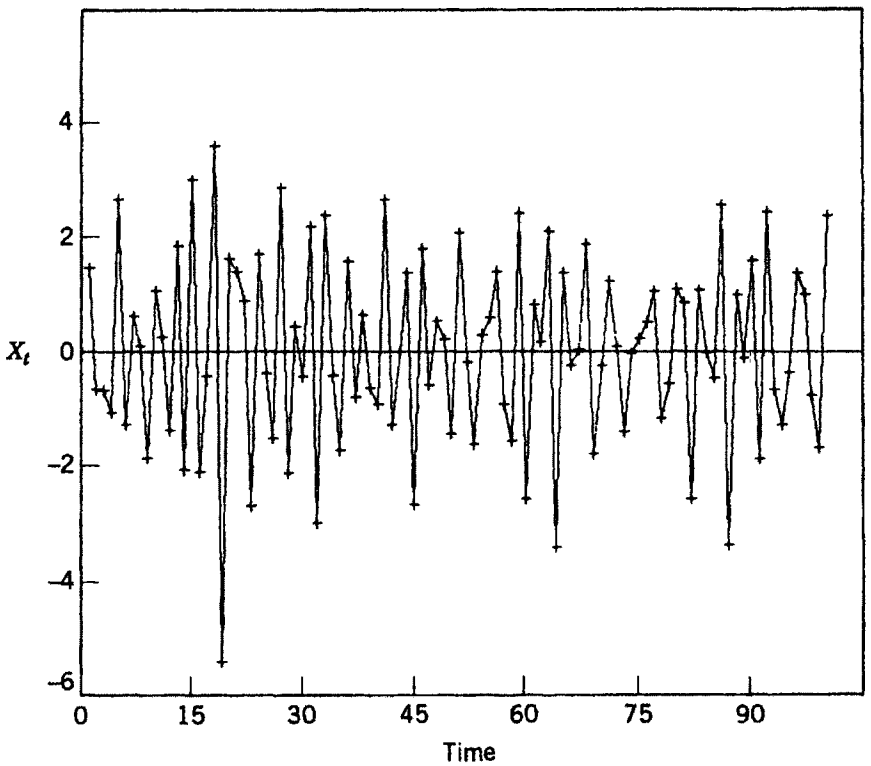

Figure 2.1.5. One hundred observations from the time series $X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$.

$$\lim_{n\to\infty} \sum_{j=1}^{n} a_j = \lim_{n\to\infty} s_n$$

when the limit is defined. If the sequence $\{s_j\}$ has a finite limit, we say that the series $\Sigma\{a_j\}$ is *convergent*. If the limit is not defined, the series is *divergent*.

If the limit

$$\lim_{n\to\infty} \sum_{j=1}^{n} |a_j|$$

is finite, the series $\Sigma\{a_j\}$ is said to be *absolutely convergent*, and we write $\Sigma_{j=1}^{\infty}|a_j| < \infty$. We also describe this situation by saying the sequence $\{a_j\}$ is *absolutely summable*. For a doubly infinite sequence, if the limit

$$\lim_{n\to\infty} \sum_{j=-n}^{n} |a_j|$$

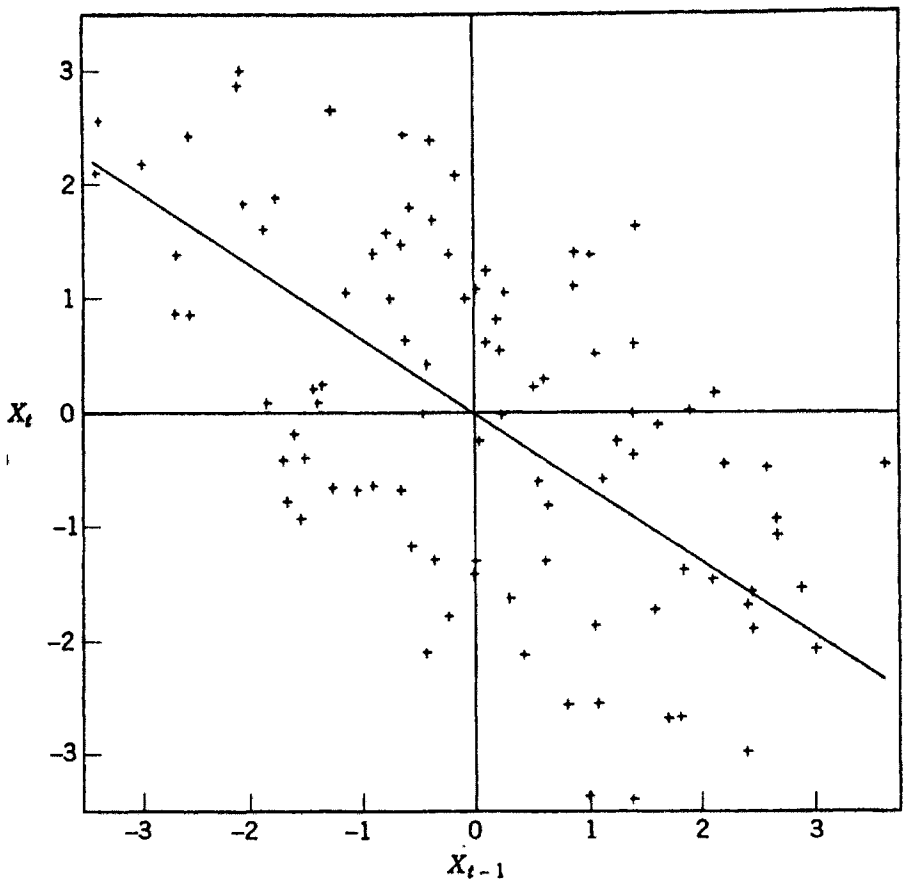

Figure 2.1.6. Plot of X_t against X_{t-1} for 100 observations from $X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$.

exists and is finite, then the series is absolutely convergent, and we write

$$\sum_{j=-\infty}^{\infty} |a_j| < \infty$$

and say that the sequence $\{a_j\}$ is absolutely summable.

The following elementary properties of absolutely summable sequences will be useful in our investigation of moving average time series.

1. If $\{a_j\}$ is absolutely summable, then

$$\sum_{j=-\infty}^{\infty} a_j^2 < \infty .$$

The reader can verify that the converse is not true by considering $\sum_{j=1}^{\infty} j^{-2}$ and $\sum_{j=1}^{\infty} j^{-1}$.

2. Given two absolutely summable sequences $\{a_j\}$ and $\{b_j\}$, then the sequences $\{a_j + b_j\}$ and $\{a_j b_j\}$ are absolutely summable:

$$\sum_{j=-\infty}^{\infty} |a_j + b_j| \leq \sum_{j=-\infty}^{\infty} (|a_j| + |b_j|) = \sum_{j=-\infty}^{\infty} |a_j| + \sum_{j=-\infty}^{\infty} |b_j| < \infty,$$

$$\sum_{j=-\infty}^{\infty} |a_j b_j| = \sum_{j=-\infty}^{\infty} (|a_j|\,|b_j|) \leq \sum_{j=-\infty}^{\infty} (|a_j| + |b_j|)^2 < \infty.$$

Of course, $\sum_{j=-\infty}^{\infty} (|a_j| + |b_j|)^2 < \infty$, since $\{(|a_j| + |b_j|)^2\}$ is the square of an absolutely summable sequence.

3. The *convolution* of two absolutely summable sequences $\{a_j\}$ and $\{b_j\}$, defined by

$$c_j = \sum_{k=-\infty}^{\infty} a_k b_{j-k},$$

is absolutely summable:

$$\sum_{j=-\infty}^{\infty} |c_j| \leq \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_k|\,|b_{j-k}| \leq \sum_{k=-\infty}^{\infty} |a_k| \sum_{s=-\infty}^{\infty} |b_s| < \infty.$$

This result generalizes, and, for example, if we define

$$d_j = \sum_{k=-\infty}^{\infty} a_k \sum_{m=-\infty}^{\infty} b_m c_{j-k-m},$$

where $\{a_j\}$, $\{b_j\}$, and $\{c_j\}$ are absolutely summable, then

$$\sum_{j=-\infty}^{\infty} |d_j| < \infty.$$

The infinite moving average time series

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

was introduced in equation (2.1.2) of Section 2.1. In investigating the behavior of infinite moving averages we will repeatedly interchange the summation and expectation operations. As justification for this procedure, we give the following theorems. The theorems may be skipped by those not familiar with real analysis. The organization of the proofs follows suggestions made by Torres (1986) and Pantula (1988a). We begin with a lemma on almost sure convergence of a sequence of random functions.

Lemma 2.2.1. Let $\{Z_j\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, P)$. Assume

$$\sum_{j=-\infty}^{\infty} E\{|Z_j|\} < \infty .$$

Then $\Sigma_{j=-\infty}^{\infty} Z_j$ is defined as an almost sure limit and

$$E\left\{\sum_{j=-\infty}^{\infty} Z_j\right\} = \sum_{j=-\infty}^{\infty} E\{Z_j\} .$$

Proof. Let a nondecreasing sequence of random variables be defined by

$$Y_k(\omega) = \sum_{j=-k}^{k} |Z_j(\omega)| .$$

Therefore, the limit as $k \to \infty$ is defined. By assumption, $E\{|Z_j|\}$ is finite for each j and it follows that

$$E\{Y_k\} < M < \infty$$

for some M and all k. Therefore, by the monotone convergence theorem,

$$E\left\{\sum_{j=-\infty}^{\infty} |Z_j|\right\} = \sum_{j=-\infty}^{\infty} E\{|Z_j|\} < \infty .$$

See, for example, Royden (1989, p. 265). Also $\Sigma_{j=-\infty}^{\infty} |Z_j| < \infty$ a.s. Therefore,

$$\lim_{n\to\infty} S_n = \sum_{j=-\infty}^{\infty} Z_j \qquad \text{a.s.}$$

and

$$|S_n| < \sum_{j=-\infty}^{\infty} |Z_j| < \infty \qquad \text{a.s.,}$$

where $S_n = \Sigma_{j=-n}^{n} Z_j$. It follows, by the dominated convergence theorem, that

$$E\left\{\sum_{j=-\infty}^{\infty} Z_j\right\} = \sum_{j=-\infty}^{\infty} E\{Z_j\} . \qquad \blacktriangle$$

Theorem 2.2.1. Let the sequence of real numbers $\{a_j\}$ and the sequence of random variables $\{Z_t\}$ satisfy

$$\sum_{j=-\infty}^{\infty} |a_j| = M < \infty$$

and $E\{Z_t^2\} \leq K$, $t = 0, \pm 1, \pm 2, \ldots$, for some finite K. Then there exists a sequence of random variables $\{X_t\}$ such that, for $t = 0, \pm 1, \pm 2, \ldots$,

$$X_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j} \qquad \text{a.s.},$$

$$\lim_{n \to \infty} E\left\{ \left| X_t - \sum_{j=-n}^{n} a_j Z_{t-j} \right|^2 \right\} = 0,$$

and $E\{X_t^2\} \leq M^2 K$.

Proof. By Lemma 2.2.1, X_t is well defined as an almost sure limit and

$$E\left\{ \sum_{j=-\infty}^{\infty} a_j Z_j \right\} = \sum_{j=-\infty}^{\infty} E\{a_j Z_j\}.$$

Because $(|Z_t| - |Z_j|)^2 \geq 0$, we have $E\{|Z_t|\,|Z_j|\} \leq K$ for all t, j. It follows that

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E\{|a_i Z_{t-i} a_j Z_{t-j}|\} \leq K\left(\sum_{j=-\infty}^{\infty} |a_j| \right)^2.$$

Letting

$$T_n = \sum_{i=-n}^{n} \sum_{j=-n}^{n} a_i a_j Z_{t-i} Z_{t-j}$$

and applying Lemma 2.2.1, we have that

$$T_n \to \left(\sum_{i=-\infty}^{\infty} a_i Z_{t-i} \right)^2 \qquad \text{a.s.}$$

and

$$E\left\{ \left(\sum_{i=-\infty}^{\infty} a_i Z_{t-i} \right)^2 \right\} \leq K\left(\sum_{j=-\infty}^{\infty} |a_j| \right)^2$$

for all $t = 0, \pm 1, \pm 2, \ldots$. It follows that

$$E\left[\left| X_t - \sum_{j=-n}^{n} a_j Z_{t-j} \right|^2 \right] = E\left\{ \left| \sum_{j=n+1}^{\infty} a_j Z_{t-j} + \sum_{j=-\infty}^{-n-1} a_j Z_{t-j} \right|^2 \right\} \leq K\left(\sum_{|j|>n} |a_j| \right)^2$$

and $\Sigma_{|j|>n} |a_j|$ converges to zero as $n \to \infty$. ▲

Because almost sure convergence implies convergence in probability, the sequence of random variables $\{X_t\}$ of Theorem 2.2.1 is defined as a limit in probability, as a limit in mean square, and as an almost sure limit. The second moment condition is required for convergence in mean square, and the first

absolute moment is required for almost sure convergence. Hereafter, we shall often use the infinite sum with no modifiers to represent the limit random variable.

Theorem 2.2.2. Let the sequence of real numbers $\{a_j\}$, $\{b_j\}$ be absolutely summable. Let $\{Z_t\}$ be a sequence of random variables such that $E\{Z_t^2\} \leq K$ for $t = 0, \pm 1, \pm 2, \ldots$, and for some finite K. Define

$$(X_t, Y_t) = \left(\sum_{j=-\infty}^{\infty} a_j Z_{t-j}, \sum_{k=-\infty}^{\infty} b_k Z_{t-k} \right).$$

Then

$$E\{X_t\} = \lim_{n \to \infty} \sum_{j=-n}^{n} a_j E\{Z_{t-j}\}$$

and

$$E\{X_t Y_t\} = \lim_{n \to \infty} \sum_{j=-n}^{n} \sum_{k=-n}^{n} a_j b_k E\{Z_{t-j} Z_{t-k}\}.$$

Proof. The expectation result follows from Lemma 2.2.1 and the proof of Theorem 2.2.1. By assumption,

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_j b_k| E\{|Z_{t-j} Z_{t-k}|\} \leq \left(\sum_{j=-\infty}^{\infty} |a_j| \right) \left(\sum_{k=-\infty}^{\infty} |b_k| \right) K < \infty$$

and the conclusion follows by Lemma 2.2.1. ▲

Corollary 2.2.2.1. Let the sequences of real numbers $\{a_j\}$ and $\{b_j\}$ be absolutely summable. Let $\{e_t\}$ be a sequence of uncorrelated $(0, \sigma^2)$ random variables, and define for $t = 0, \pm 1, \pm 2, \ldots$

$$(X_t, Y_t) = \left(\sum_{j=-\infty}^{\infty} a_j e_{t-j}, \sum_{k=-\infty}^{\infty} b_k e_{t-k} \right).$$

Then, $E\{X_t\} = 0$ and

$$E\{X_t Y_t\} = \sigma^2 \sum_{j=-\infty}^{\infty} a_j b_j .$$

Proof. The sequence $\{e_t\}$ satisfies the hypotheses of Theorem 2.2.2 with $K = \sigma^2$. Since $E\{e_t\} = 0$ and $E\{e_t e_j\} = 0$ for $t \neq j$, we have the results. ▲

Corollary 2.2.2.2. Let $\{X_t\}$ be given by

$$X_t = \sum_{j=-\infty}^{\infty} a_j e_{t-j} ,$$

and define

$$Y_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j},$$

where $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables and $\{a_j\}$ and $\{b_j\}$ are absolutely summable sequences of real numbers. Then $\{Y_t\}$ is a stationary moving average time series and

$$\begin{aligned}\gamma_Y(h) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \gamma_X(k-j+h) \\ &= \sum_{j=-\infty}^{\infty} c_j c_{j-h} \sigma^2, \end{aligned} \tag{2.2.2}$$

where $c_j = \sum_{k=-\infty}^{\infty} b_k a_{j-k}$. Furthermore,

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

Proof. By definition,

$$Y_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j} = \sum_{j=-\infty}^{\infty} b_j \sum_{k=-\infty}^{\infty} a_k e_{t-j-k}.$$

Since $\{a_j\}$ and $\{b_j\}$ are absolutely summable, we may interchange the order of summation (Fubini's theorem). Setting $m = j + k$, we have

$$\begin{aligned} Y_t &= \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j a_{m-j} e_{t-m} \\ &= \sum_{m=-\infty}^{\infty} c_m e_{t-m}, \end{aligned}$$

where $c_m = \sum_{j=-\infty}^{\infty} b_j a_{m-j}$. By result 3 on the convolution of absolutely summable sequences, $\{c_m\}$ is absolutely summable. Therefore,

$$\gamma_Y(h) = \sum_{m=-\infty}^{\infty} c_m c_{m-h} \sigma^2$$

is also absolutely summable.

Using Theorem 2.2.2,

$$\begin{aligned} \gamma_Y(h) &= \sum_{m=-\infty}^{\infty} c_m c_{m-h} \sigma^2 \\ &= \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j a_{m-j} \sum_{k=-\infty}^{\infty} b_k a_{m-k-h} \sigma^2 \end{aligned}$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \sum_{m=-\infty}^{\infty} a_{m-j} a_{m-k-h} \sigma^2$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \gamma_X(k-j+h). \quad \blacktriangle$$

We may paraphrase this corollary by saying that an infinite moving average (where the coefficients are absolutely summable) of an infinite moving average stationary time series is itself an infinite moving average stationary time series. If the $\{a_j\}$ and $\{b_j\}$ are exponentially declining, then $\{c_m\}$ of Corollary 2.2.2.2 is exponentially declining.

Corollary 2.2.2.3. Let the $\{a_j\}$ and $\{b_j\}$ of Corollary 2.2.2.2 satisfy

$$|a_j| < M\lambda^{|j|} \quad \text{and} \quad |b_j| < M\lambda^{|j|}$$

for some $M < \infty$ and $0 < \lambda < 1$. Let Y_t be as defined in Corollary 2.2.2.2. Then

$$Y_t = \sum_{m=-\infty}^{\infty} c_m e_{t-m},$$

where $c_m = \sum_{j=-\infty}^{\infty} b_j a_{m-j}$ and $|c_m| < M_c \lambda^{|m|}$ for some $M_c < \infty$.

Proof. We have

$$|c_m| = \left| \sum_{j=-\infty}^{\infty} b_j a_{m-j} \right| \leq M^2 \lambda^{|m|} \left(m + 1 + 2 \sum_{j=1}^{\infty} \lambda^{2j} \right). \quad \blacktriangle$$

In Theorems 2.2.1 and 2.2.2, we made mild assumptions on the Z_t and assumed the sequence of coefficients, $\{a_j\}$, to be absolutely summable. In Theorem 2.2.3, we give a mean square result for the infinite sum of the elements of a time series under weaker assumptions on the coefficients.

Theorem 2.2.3. Let $\{a_j\}$ be a sequence of real numbers such that $\{a_j^2\}$ is summable. Let Z_t be the time series

$$Z_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j},$$

where $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables and $\{b_j\}$ is absolutely summable. Then there exists a sequence of random variables $\{X_t\}$ such that for $t = 0, \pm 1, \pm 2, \ldots,$

$$\lim_{n \to \infty} E\left\{ \left(X_t - \sum_{j=-n}^{n} a_j Z_{t-j} \right)^2 \right\} = 0,$$

$E\{X_t\} = 0$, and

$$E\{X_t X_{t+h}\} = \sigma^2 \sum_{j=-\infty}^{\infty} c_j c_{j+h}, \tag{2.2.3}$$

where $c_j = \Sigma_{k=-\infty}^{\infty} a_k b_{j-k}$. Furthermore,

$$\lim_{n\to\infty} E\left\{\left(X_t - \sum_{j=-n}^{n} c_j e_{t-j}\right)^2\right\} = 0.$$

Proof. By the Cauchy–Schwarz inequality, we have, for $n > m$ and $h > 0$,

$$\left|\sum_{j=m+1}^{n} a_j a_{j+h}\right| \leq \left[\left(\sum_{j=m+1}^{n} a_j^2\right)\left(\sum_{j=m+1}^{n} a_{j+h}^2\right)\right]^{1/2}. \tag{2.2.4}$$

Since $\Sigma a_j^2 < \infty$, given $\epsilon > 0$, there exists an N_0 such that, for all $n > m > N_0$,

$$\left|\sum_{j=m+1}^{n} a_j a_{j+h}\right| < \epsilon \tag{2.2.5}$$

uniformly in h. Let $Y_{nt} = \Sigma_{j=-n}^{n} a_j Z_{t-j}$. For $n > m > N_0$,

$$E\{(Y_{nt} - Y_{mt})^2\} \leq 4E\left\{\left(\sum_{j=-n}^{-m-1} a_j Z_{t-j}\right)^2\right\} + 4E\left\{\left(\sum_{j=m+1}^{n} a_j Z_{t-j}\right)^2\right\}. \tag{2.2.6}$$

Now

$$E\left\{\left(\sum_{j=m+1}^{n} a_j Z_{t-j}\right)^2\right\} = \left(\sum_{j=m+1}^{n} a_j^2\right)\gamma_Z(0) + 2\sum_{j=m+1}^{n-1}\sum_{k=j+1}^{n} a_j a_k \gamma_Z(k-j),$$

where

$$\left|\sum_{j=m+1}^{n-1}\sum_{k=j+1}^{n} a_j a_k \gamma_Z(k-j)\right| \leq \epsilon \sum_{h=1}^{\infty} |\gamma_Z(h)|$$

by (2.2.5). Because, by Corollary 2.2.2.2, $\Sigma_{h=1}^{\infty} |\gamma_Z(h)| < \infty$, it follows that the right side of (2.2.6) converges to zero in mean square as $m \to \infty$ and $n \to \infty$. Therefore, Y_{nt} is Cauchy in squared mean and there exists a random variable X_t such that $E\{X_t^2\} < \infty$ and

$$E\{(X_t - Y_{nt})^2\} \to 0 \qquad \text{as} \quad n \to \infty.$$

See, for example, Tucker (1967, p. 105). Also, $E\{X_t\} = \lim_{n\to\infty} E\{Y_{nt}\} = 0$ and $E\{X_t X_{t+h}\} = \lim_{n\to\infty} E\{Y_{nt} Y_{nt+h}\}$.

We write

$$Y_{nt} = \sum_{j=-\infty}^{\infty} d_{jn} Z_{t-j} = \sum_{j=-\infty}^{\infty} d_{jn} \sum_{l=-\infty}^{\infty} b_l e_{t-j-l},$$

where

$$d_{jn} = \begin{cases} a_j & \text{if } |j| \leq n, \\ 0 & \text{if } |j| > n. \end{cases}$$

Since d_{jn} and b_l are absolutely summable, we have

$$Y_{nt} = \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d_{jn} b_l e_{t-j-l} = \sum_{k=-\infty}^{\infty} c_{kn} e_{t-k},$$

where

$$c_{kn} = \sum_{j=-\infty}^{\infty} d_{jn} b_{k-j} = \sum_{j=-n}^{\infty} a_j b_{k-j}.$$

Also,

$$E\{Y_{nt} Y_{n,t+h}\} = \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2.$$

Therefore,

$$E\{X_t X_{t+h}\} = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2.$$

Now c_{kn} converges to $c_k = \sum_{j=-\infty}^{\infty} a_j b_{k-j}$ for all k. Also, $|c_{kn}| \leq m_k$ for all k, where $m_k = \sum_{j=-\infty}^{\infty} |a_j| \, |b_{k-j}|$, and

$$\sum_{k=-\infty}^{\infty} |m_k| \, |m_{k+h}| < \infty,$$

because $\{m_k\}$ is square summable. See Exercise 2.45. Therefore, by the dominated convergence theorem,

$$E\{X_t X_{t+h}\} = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2 = \sum_{k=-\infty}^{\infty} c_k c_{k+h} \sigma^2. \qquad \blacktriangle$$

Corollary 2.2.3. Let $\{c_j\}_{-\infty}^{\infty}$ be square summable and let $\{d_j\}_{-\infty}^{\infty}$ be absolutely summable. Let

$$Z_t = \sum_{j=-\infty}^{\infty} c_j e_{t-j},$$

where e_t is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Then there exists a sequence of random variables $\{X_t\}_{-\infty}^{\infty}$ such that

$$\lim_{n\to\infty} E\left\{\left|X_t - \sum_{j=-n}^{n} d_j Z_{t-j}\right|^2\right\} = 0,$$

$E\{X_t\} = 0$, and

$$E\{X_t X_{t+h}\} = \sigma^2 \sum_{j=-\infty}^{\infty} g_j g_{j+h},$$

where $g_j = \sum_{k=-\infty}^{\infty} c_k d_{j-k}$. Furthermore,

$$\lim_{n\to\infty} E\left\{\left|X_t - \sum_{j=-n}^{n} g_j e_{t-j}\right|^2\right\} = 0.$$

Proof. From Theorem 2.2.3, we have that Z_t is well defined in mean square with $E\{Z_t\} = 0$ and $\gamma_Z(h) = \sigma^2 \sum_{j=-\infty}^{\infty} c_j c_{j+h}$. Now, the existence of the sequence of random variables $\{X_t\}_{-\infty}^{\infty}$ such that $E\{X_t\} = 0$ and

$$\lim_{n\to\infty} E\left\{\left|X_t - \sum_{j=-n}^{n} d_j Z_{t-j}\right|^2\right\} = 0$$

follows from Theorem 2.2.1. Furthermore, from Theorem 2.2.2, we have

$$\begin{aligned} E\{X_t X_{t+h}\} &= \lim_{n\to\infty} \sum_{j=-n}^{n} \sum_{j=-n}^{n} d_j d_k E\{Z_{t-j} Z_{t+h-k}\} \\ &= \lim_{n\to\infty} \sum_{j=-n}^{n} \sum_{k=-n}^{n} d_j d_k \sum_{l=-\infty}^{\infty} c_l c_{l+j+h-k} \sigma^2 \\ &= \sigma^2 \sum_{j=-\infty}^{\infty} g_j g_{j+h}, \end{aligned}$$

where $g_j = \sum_{k=-\infty}^{\infty} c_k d_{j-k}$. Using the dominated convergence theorem, we have

$$\begin{aligned} \lim_{n\to\infty} E\left\{X_t \sum_{i=-n}^{n} g_i e_{t-i}\right\} &= \lim_{n\to\infty} \sum_{i=-n}^{n} g_i E\left\{\sum_{j=-\infty}^{\infty} d_j Z_{t-j} e_{t-i}\right\} \\ &= \lim_{n\to\infty} \sum_{i=-n}^{n} g_i \sum_{j=-\infty}^{\infty} d_j c_{i-j} \sigma^2 \\ &= \sigma^2 \sum_{i=-\infty}^{\infty} g_i^2 = E\{X_t^2\}. \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} E\left\{\left(X_t - \sum_{i=-n}^{n} g_i e_{t-i}\right)^2\right\} = E\{X_t^2\} + \lim_{n\to\infty} \sum_{i=-n}^{n} g_i^2\sigma^2 - 2\lim_{n\to\infty} E\left\{X_t \sum_{i=-n}^{n} g_i e_{t-i}\right\}$$
$$= 0. \qquad \blacktriangle$$

2.3. AN INTRODUCTION TO AUTOREGRESSIVE TIME SERIES

Many time series encountered in practice are well approximated by the representation

$$\sum_{i=0}^{p} \alpha_i X_{t-i} = e_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2.3.1}$$

where $\alpha_0 \neq 0$, $\alpha_p \neq 0$, and the e_t are uncorrelated $(0, \sigma^2)$ random variables. The sequence $\{X_t\}$ is called a *p*th *order autoregressive* time series. The defining equation (2.3.1) is sometimes called a *stochastic difference equation.*

To study such processes, let us consider the first order autoregressive time series, which we express in its most common form,

$$X_t = \rho X_{t-1} + e_t. \tag{2.3.2}$$

By repeated substitution of

$$X_{t-i} = \rho X_{t-i-1} + e_{t-i}$$

for $i = 1, 2, \ldots, N$ into (2.3.2), we obtain

$$X_t = \rho^N X_{t-N} + \sum_{i=0}^{N-1} \rho^i e_{t-i}. \tag{2.3.3}$$

Under the assumptions that $|\rho| < 1$ and $E\{X_t^2\} < K < \infty$, we have

$$\lim_{N\to\infty} E\left\{\left(X_t - \sum_{i=0}^{N} \rho^i e_{t-i}\right)^2\right\} = 0. \tag{2.3.4}$$

Thus, if e_t is defined for $t \in \{0, \pm 1, \pm 2, \ldots\}$ and X_t satisfies the difference equation (2.3.2) with $|\rho| < 1$, then we may express X_t as an infinite moving average of the e_t:

$$X_t = \sum_{i=0}^{\infty} \rho^i e_{t-i}.$$

It follows that $E\{X_t\} = 0$ for all t, and

$$\gamma(h) = E\{X_t X_{t+h}\} = E\left\{\left(\sum_{i=0}^{\infty} \rho^i e_{t-i}\right)\left(\sum_{j=0}^{\infty} \rho^j e_{t+h-j}\right)\right\}$$

$$= \sigma^2 \sum_{i=0}^{\infty} \rho^i \rho^{i+h} = \frac{\rho^h}{1-\rho^2}\sigma^2 \tag{2.3.5}$$

for $h = 0, 1, \ldots$. The covariance function for $h \neq 0$ is also rapidly obtained by making use of a form of (2.3.3),

$$X_t = \rho^h X_{t-h} + \sum_{j=0}^{h-1} \rho^j e_{t-j}, \qquad h = 1, 2, \ldots. \tag{2.3.6}$$

Since X_{t-h} is a function of e's with subscript less than or equal to $t-h$, X_{t-h} is uncorrelated with $e_{t-h+1}, e_{t-h+2}, \ldots$. Therefore, we have, after multiplying both sides of (2.3.6) by X_{t-h} and taking expectations,

$$\gamma(h) = E\{X_t X_{t-h}\} = E\{\rho^h X_{t-h}^2\} + E\left\{X_{t-h} \sum_{j=0}^{h-1} \rho^j e_{t-j}\right\}$$

$$= \rho^h \gamma(0), \qquad h = 1, 2, \ldots. \tag{2.3.7}$$

The correlation function is seen to be a monotonically declining function for $\rho > 0$, while for $\rho < 0$ the function is alternately positive and negative, the absolute value declining at a geometric rate.

Example 2.3.1. Figure 2.3.1 displays 100 observations from the first order process

$$X_t = 0.7X_{t-1} + e_t.$$

The current observation, X_t, is plotted against the preceding observation, X_{t-1}, in Figure 2.3.2 and against X_{t-2} in Figure 2.3.3. The correlation between X_t and X_{t-2} is $(0.7)^2 = 0.49$, but the partial correlation between X_t and X_{t-2} adjusted for X_{t-1} is zero, since

$$\text{Cov}\{X_t - \rho X_{t-1}, X_{t-2} - \rho X_{t-1}\} = (\rho^2 - \rho^2 - \rho^2 + \rho^2)\gamma(0) = 0.$$

The zero partial correlation is illustrated by Figure 2.3.4, which contains a plot of $X_t - 0.7X_{t-1}$ against $X_{t-2} - 0.7X_{t-1}$. In fact, for any $h \in (2, 3, \ldots)$, the partial correlation between X_t and X_{t-h} adjusted for X_{t-1} is zero. This follows because

$$\text{Cov}\{X_t - \rho X_{t-1}, X_{t-h} - \rho^{h-1} X_{t-1}\}$$

$$= (\rho^h - \rho^{h-1}\rho - \rho\rho^{h-1} + \rho^h)\gamma(0) = 0, \qquad h = 2, 3, \ldots.$$

This important property of the first order autoregressive time series can be

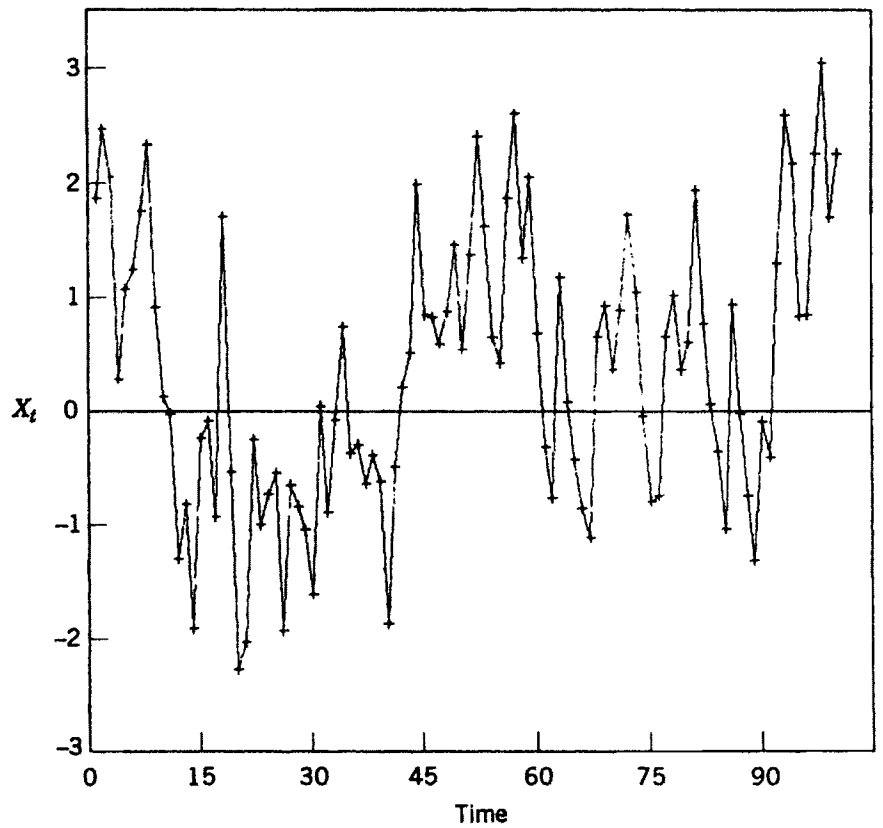

Figure 2.3.1. One hundred observations from the time series $X_t = 0.7X_{t-1} + e_t$.

characterized by saying that all the useful (correlational) information about X_t in the entire realization previous to X_t is contained in X_{t-1}. Compare this result with the partial correlational properties of the moving average process discussed in Section 2.1. ▲▲

Linear difference equations and their solutions play an important role in the study of autoregressive time series as well as in other areas of time series analysis. Therefore we digress to present the needed elementary results before studying higher order autoregressive processes. The initial development follows Goldberg (1958).

2.4. DIFFERENCE EQUATIONS

Given a sequence $\{y_t\}_{t=0}^{\infty}$, the *first difference* is defined by

$$\Delta y_t = y_t - y_{t-1}, \qquad t = 1, 2, \ldots, \tag{2.4.1}$$

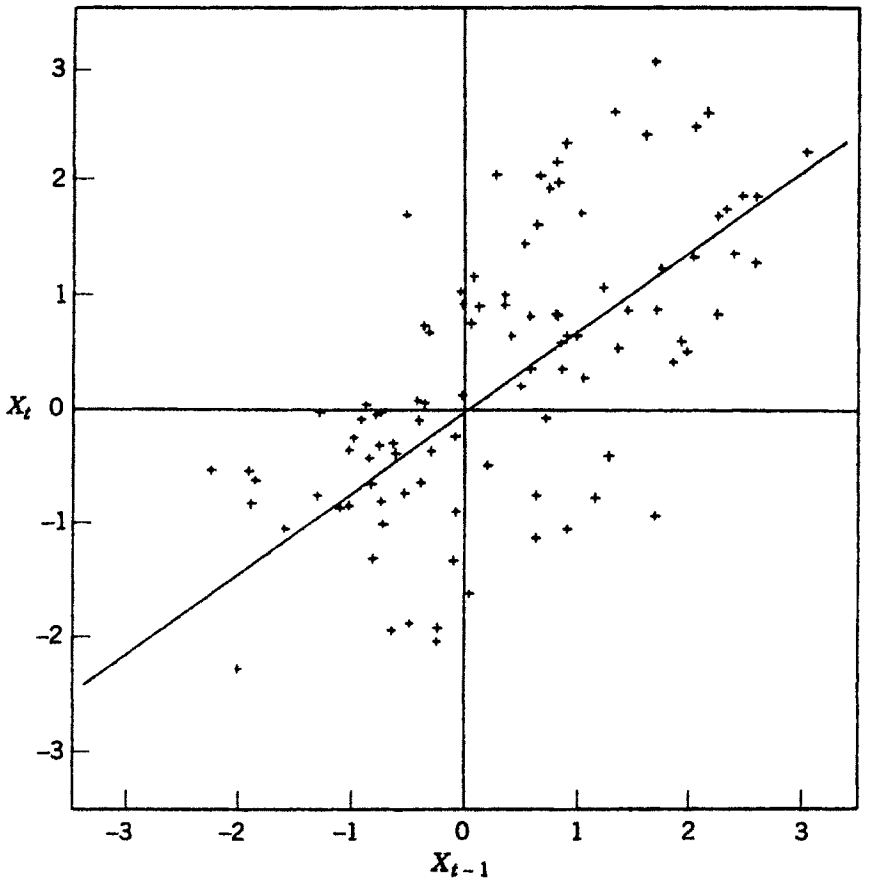

Figure 2.3.2. Plot of X_t against X_{t-1} for 100 observations from $X_t = 0.7X_{t-1} + e_t$.

and the *n*th *difference* is defined by

$$\Delta^n y_t = \Delta^{n-1} y_t - \Delta^{n-1} y_{t-1} = \sum_{r=0}^{n} (-1)^r \binom{n}{r} y_{t-r}, \qquad t = n, n+1, \ldots,$$

where

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

are the binomial coefficients. The equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = r_t, \qquad t = n, n+1, \ldots, \quad (2.4.2)$$

where the a_i are real constants, $a_n \neq 0$, and r_t is a real function of t, is a *linear difference equation of order n* with constant coefficients. The values $y_{t-1}, y_{t-2}, \ldots, y_{t-n}$ are sometimes called *lagged* values of y_t. Equation (2.4.2)

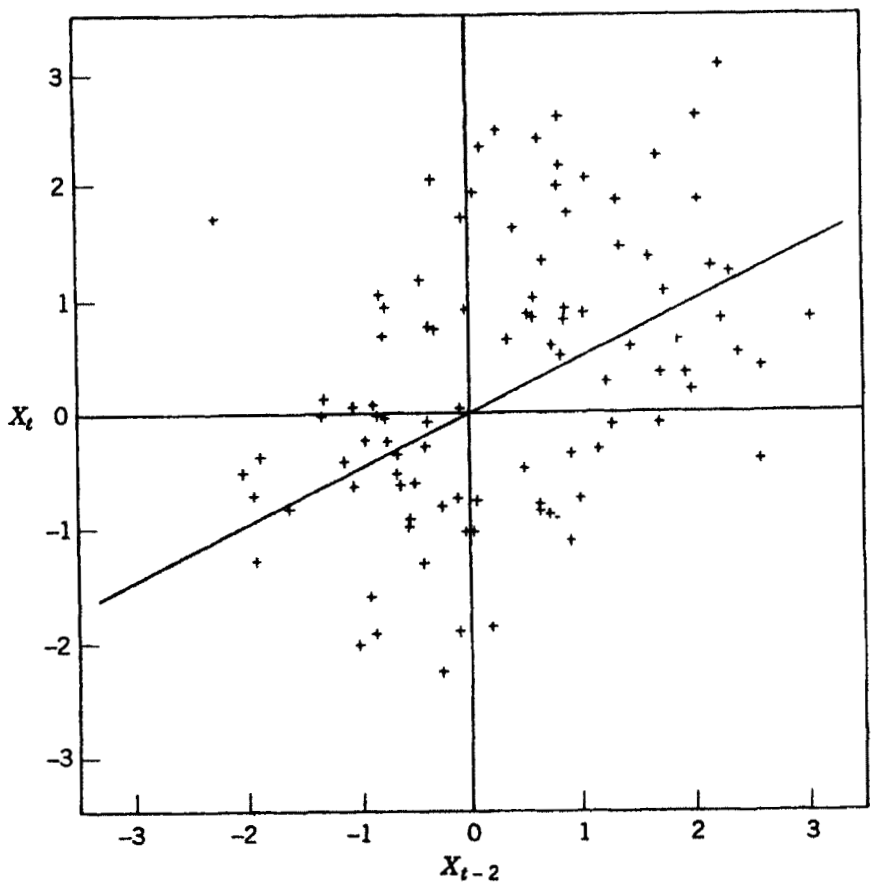

Figure 2.3.3. Plot of X_t against X_{t-2} for 100 observations from $X_t = 0.7X_{t-1} + e_t$.

could be expressed in terms of y_{t-n} and the differences of y_t; for example,

$$\Delta^n y_t + b_1 \Delta^{n-1} y_{t-1} + b_2 \Delta^{n-2} y_{t-2} + \cdots + b_{n-1} \Delta y_{t-n+1} + b_n y_{t-n} = r_t , \tag{2.4.3}$$

where the b_i's are linear functions of the a_i's.

A third representation[1] of (2.4.2) is possible. Let the symbol $\mathcal{B}$ denote the operation of replacing y_t by y_{t-1}; that is,

$$\mathcal{B} y_t = y_{t-1} .$$

[1] The reader may wonder at the benefits associated with several alternative notations. All are heavily used in the literature and all have advantages for certain operations.

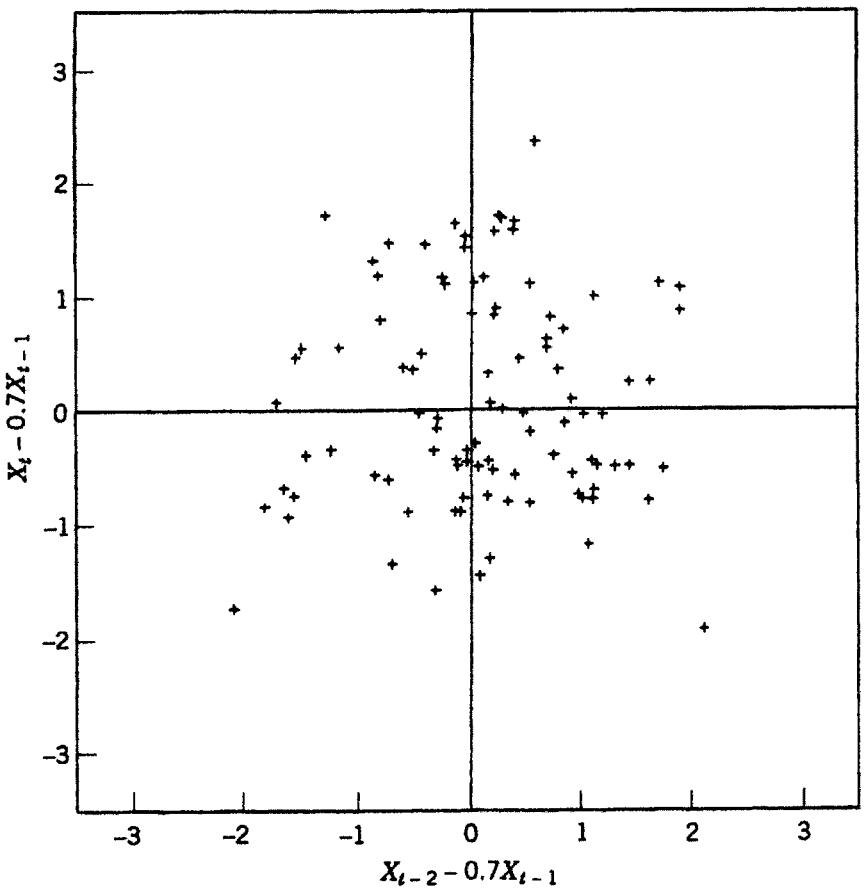

Figure 2.3.4. Plot of $X_t - 0.7X_{t-1}$ against $X_{t-2} - 0.7X_{t-1}$ for 100 observations from $X_t = 0.7X_{t-1} + e_t$.

Using the *backward shift operator* $\mathcal{B}$, we have

$$\Delta y_t = y_t - \mathcal{B}y_t = (1 - \mathcal{B})y_t .$$

As with the difference operator, we denote repeated application of the operator with the appropriate exponent; for example,

$$\mathcal{B}^3 y_t = \mathcal{B}\mathcal{B}\mathcal{B}y_t = y_{t-3} .$$

Therefore, we can use the operator symbol to write (2.4.2) as

$$(1 + a_1\mathcal{B} + a_2\mathcal{B}^2 + \cdots + a_n\mathcal{B}^n)y_t = r_t , \qquad t = n, n+1, \ldots .$$

Associated with (2.4.2) is the *reduced* or *homogeneous* difference equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = 0 . \tag{2.4.4}$$

From the linearity of (2.4.2) and (2.4.4) we have the following properties.

1. If $y^{(1)}$ and $y^{(2)}$ are solutions of (2.4.4), then $b_1 y^{(1)} + b_2 y^{(2)}$ is a solution of (2.4.4), where b_1 and b_2 are arbitrary constants.

2. If Y is a solution of (2.4.4) and $y^{\dagger}$ a solution of (2.4.2), then $Y + y^{\dagger}$ is a solution of (2.4.2).

$y^{\dagger}$ is called a *particular solution* and $Y + y^{\dagger}$ a *general solution* of the difference equation (2.4.2).

Equations (2.4.2) and (2.4.4) specify y_t as a function of the preceding n values of y. By using an inductive proof it is possible to demonstrate that the linear difference equation of order n has one and only one solution for which values at n consecutive t values are arbitrarily prescribed. This important result means that when we obtain n linearly independent solutions to the difference equation, we can construct the general solution.

The first order homogeneous difference equation

$$y_t = a y_{t-1}, \qquad t = 1, 2, \ldots, \tag{2.4.5}$$

has the unique solution

$$y_t = a^t y_0, \tag{2.4.6}$$

where y_0 is the value of y at $t = 0$. Note that if $|a| < 1$, the solution tends to zero as t increases. Conversely, if $|a| > 1$, the absolute value of the solution sequence increases without bound. If a is negative, the solution oscillates with alternately positive and negative values. The first order nonhomogeneous difference equation

$$y_t = a y_{t-1} + b \tag{2.4.7}$$

has the unique solution

$$y_t = \begin{cases} y_0 a^t + b\left(\dfrac{1 - a^t}{1 - a}\right), & a \neq 1, \\ y_0 + bt, & a = 1. \end{cases} \tag{2.4.8}$$

If the constant b in the difference equation (2.4.7) is replaced by a function of time, for example,

$$y_t = a y_{t-1} + b_t, \qquad t = 1, 2, \ldots, \tag{2.4.9}$$

then the solution is given by

$$y_t = y_0 a^t + \sum_{j=0}^{t-1} a^j b_{t-j}, \qquad t = 1, 2, \ldots, \tag{2.4.10}$$

where y_0 is the value of y_t at time 0.

The solutions (2.4.6), (2.4.8), and (2.4.10) are readily verified by differencing. The fact that the linear function is reduced by differencing to the constant function is of particular interest. The generalization of this result is important in the analysis of nonstationary time series.

Theorem 2.4.1. Let y_t be a polynomial of degree n whose domain is the integers. Then the first difference Δy_t is expressible as a polynomial of degree $n-1$ in t.

Proof. Since

$$y_t = \sum_{p=0}^{n} \beta_p t^p, \qquad \beta_n \neq 0,$$

and

$$y_{t-1} = \sum_{p=0}^{n} \beta_p (t-1)^p,$$

we have

$$\begin{aligned}
\Delta y_t &= \sum_{p=0}^{n} \beta_p [t^p - (t-1)^p] \\
&= \sum_{p=0}^{n} \beta_p \left[t^p - \sum_{s=0}^{p} \binom{p}{s} (-1)^s t^{p-s} \right] \\
&= \sum_{p=1}^{n} \beta_p \left[- \sum_{s=1}^{p} \binom{p}{s} (-1)^s t^{p-s} \right] \\
&\overset{\text{(say)}}{=} \sum_{p=0}^{n-1} \gamma_p t^p,
\end{aligned}$$

where $\gamma_{n-1} = n\beta_n$. ▲

It follows that by taking repeated differences, one can reduce a polynomial to the zero function. We state the result as a corollary.

Corollary 2.4.1. Let y_t be a polynomial of degree n whose domain is the integers. Then the nth difference $\Delta^n y_t$ is a constant function, and the $(n+1)$st difference $\Delta^{n+1} y_t$ is the zero function.

In investigating difference equations of higher order, the nature of the solution is a function of the *auxiliary equation*, sometimes called the *characteristic equation*. For equation (2.4.2) the auxiliary equation is

$$m^n + a_1 m^{n-1} + \cdots + a_n = 0. \tag{2.4.11}$$

This polynomial equation will have n (not necessarily distinct) roots, and the behavior of the solution is intimately related to these roots.

Consider the second order linear homogeneous equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = 0, \tag{2.4.12}$$

where $a_2 \neq 0$. The two roots of the auxiliary equation

$$m^2 + a_1 m + a_2 = 0 \tag{2.4.13}$$

will fall into one of three categories:

1. Real and distinct.
2. Real and equal.
3. A complex conjugate pair.

The general solution of the homogeneous difference equation (2.4.12) for the three categories is:

1.

$$y_t = b_1 m_1^t + b_2 m_2^t, \tag{2.4.14}$$

where b_1 and b_2 are constants determined by the initial conditions, and m_1 and m_2 are the roots of (2.4.13).

2.

$$y_t = (b_1 + b_2 t) m^t, \tag{2.4.15}$$

where m is the repeated root of (2.4.13).

3.

$$y_t = b_1^* m_1^t + b_1 m_2^t, \tag{2.4.16}$$

where b_1^* is the complex conjugate of b_1 and $m_2 = m_1^*$ is the complex conjugate of m_1.

If the complex roots are expressed as

$$r(\cos\theta \pm i \sin\theta),$$

where $r = a_2^{1/2} = |m_1|$ and

$$(\cos\theta, \sin\theta) = (2r)^{-1}(a_1, |a_1^2 - 4a_2|^{1/2}),$$

then the solution can be expressed as

$$\begin{aligned} y_t &= r^t(d_1 \cos t\theta + d_2 \sin t\theta) \\ &= g_1 r^t \cos(t\theta + g_2), \end{aligned} \tag{2.4.17}$$

where d_1, d_2, g_1, and g_2 are real constants.

Substitution of (2.4.14) into (2.4.12) immediately establishes that (2.4.14) is a solution and, in fact, that m_1^t and m_2^t are independent solutions when $m_1 \neq m_2$. For a repeated root, substitution of (2.4.15) into (2.4.12) gives

$$(b_1 + b_2 t)m^{t-2}(m^2 + a_1 m + a_2) - b_2 m^{t-2}(a_1 m + 2a_2) = 0.$$

The first term is zero because m is a root of (2.4.13). For m a repeated root, $ma_1 + 2a_2 = 0$, proving that (2.4.15) is the general solution.

As with unequal real roots, substitution of (2.4.16) into (2.4.12) demonstrates that (2.4.16) is the solution in the case of complex roots.

The solution for a linear homogeneous difference equation of order n can be obtained using the roots of the auxiliary equation. The solution is a sum of n terms where:

1. For every real and distinct root m, a term of the form bm^t is included.
2. For every real root of order p (a root repeated p times), a term of the form

$$(b_1 + b_2 t + b_3 t^2 + \cdots + b_p t^{p-1})m^t$$

is included.

3. For each pair of unrepeated complex conjugate roots, a term of the form

$$\alpha r^t \cos(t\theta + \beta)$$

is included, where r and θ are defined following (2.4.16).

4. For a pair of complex conjugate roots occurring p times, a term of the form

$$r^t[\alpha_1 \cos(t\theta + \beta_1) + \alpha_2 t \cos(t\theta + \beta_2) + \cdots + \alpha_p t^{p-1} \cos(t\theta + \beta_p)]$$

is included.

For the nonhomogeneous difference equation (2.4.2), with initial conditions $y_0, y_1, \ldots, y_{n-1}$, a particular solution is

$$y_t^\dagger = \sum_{j=0}^{t} w_j r_{t-j}, \qquad t = n, n+1, \ldots,$$

where $w_0 = 1$,

$$\sum_{i=0}^{j} a_i w_{j-i} = 0, \qquad j = 1, 2, \ldots, n-1,$$

$$\sum_{i=0}^{n} a_i w_{j-i} = 0, \qquad j = n, n+1, \ldots,$$

and $r_0, r_1, \ldots, r_{n-1}$ are defined by

$$y_t = \sum_{j=0}^{t} w_j r_{t-j}, \qquad t = 0, 1, \ldots, n-1.$$

Observe that the weights w_j satisfy the same homogeneous difference equation as the y_t. Therefore, w_j can be expressed as a sum of the four types of terms that define the solution to the homogeneous difference equation.

Systems of difference equations in more than one function of time arise naturally in many models of the physical and economic sciences. To extend our discussion to such models, consider the system of difference equations in the two functions y_{1t} and y_{2t}:

$$\begin{aligned} y_{1t} &= a_{11} y_{1,t-1} + a_{12} y_{2,t-1} + b_{1t}, \\ y_{2t} &= a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + b_{2t}, \end{aligned} \tag{2.4.18}$$

where a_{11}, a_{12}, a_{21}, a_{22} are real constants, b_{1t} and b_{2t} are real functions of time, and $t = 1, 2, \ldots$. The system may be expressed in matrix notation as

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{b}_t, \tag{2.4.19}$$

where

$$\mathbf{y}_t' = (y_{1t}, y_{2t}),$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

and $\mathbf{b}_t' = (b_{1t}, b_{2t})$.

Proceeding from the solution (2.4.6) of the first order homogeneous scalar difference equation, we might postulate that the homogeneous equation

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} \tag{2.4.20}$$

would have the solution

$$\mathbf{y}_t = \mathbf{A}^t \mathbf{y}_0, \tag{2.4.21}$$

and that the equation (2.4.19) would have the solution

$$\mathbf{y}_t = \mathbf{A}^t \mathbf{y}_0 + \sum_{j=0}^{t-1} \mathbf{A}^j \mathbf{b}_{t-j}, \tag{2.4.22}$$

where we understand that $\mathbf{A}^0 = \mathbf{I}$. Direct substitution of (2.4.21) into (2.4.20) and (2.4.22) into (2.4.19) verifies that these are the solutions of the matrix analogs of the first order scalar equations.

In the scalar case the behavior of the solution depends in a critical manner on the magnitude of the coefficient a for the first order equation and on the magnitude of the roots of the characteristic equation for the higher order equations. If the

roots are all less than one in absolute value, then the effect of the initial conditions is transient, the initial conditions being multiplied by the product of powers of the roots and polynomials in t.

In the vector case the quantities analogous to the roots of the characteristic equation are the roots of the determinantal equation

$$|\mathbf{A} - m\mathbf{I}| = 0 .$$

For the moment, we assume that the roots of the determinantal equation are distinct. Let the two roots of the matrix $\mathbf{A}$ of (2.4.19) be m_1 and m_2, and define the vectors $\mathbf{q}_{\cdot 1}$ and $\mathbf{q}_{\cdot 2}$ by the equations

$$\begin{aligned}(\mathbf{A} - m_1\mathbf{I})\mathbf{q}_{\cdot 1} &= 0 ,\\ (\mathbf{A} - m_2\mathbf{I})\mathbf{q}_{\cdot 2} &= 0 ,\\ \mathbf{q}_{\cdot 1}'\mathbf{q}_{\cdot 1} = \mathbf{q}_{\cdot 2}'\mathbf{q}_{\cdot 2} &= 1 .\end{aligned}$$

By construction, the matrix $\mathbf{Q} = (\mathbf{q}_{\cdot 1}, \mathbf{q}_{\cdot 2})$ is such that

$$\mathbf{AQ} = \mathbf{QM} , \tag{2.4.23}$$

where $\mathbf{M} = \operatorname{diag}(m_1, m_2)$. Since the roots are distinct, the matrix $\mathbf{Q}$ is nonsingular. Hence,

$$\mathbf{Q}^{-1}\mathbf{AQ} = \mathbf{M} . \tag{2.4.24}$$

Define $\mathbf{z}_t' = (z_{1t}, z_{2t})$ and $\mathbf{c}_t' = (c_{1t}, c_{2t})$ by

$$\begin{aligned}\mathbf{z}_t &= \mathbf{Q}^{-1}\mathbf{y}_t ,\\ \mathbf{c}_t &= \mathbf{Q}^{-1}\mathbf{b}_t .\end{aligned}$$

Then, multiplying (2.4.19) by $\mathbf{Q}^{-1}$, we have

$$\mathbf{z}_t = \mathbf{Q}^{-1}\mathbf{AQ}\mathbf{z}_{t-1} + \mathbf{c}_t$$

or

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} c_{1t} \\ c_{2t} \end{pmatrix} .$$

The transformation $\mathbf{Q}^{-1}$ reduces the system of difference equations to two simple first order difference equations with solutions

$$\begin{aligned} z_{1t} &= \sum_{j=0}^{t} m_1^j c_{1,t-j} ,\\ z_{2t} &= \sum_{j=0}^{t} m_2^j c_{2,t-j} ,\end{aligned} \tag{2.4.25}$$

where we have set $y_{10} = b_{10}$ and $y_{20} = b_{20}$ for notational convenience. Therefore, the effect of the initial conditions is transient if $|m_1| < 1$ and $|m_2| < 1$. The solutions (2.4.25) may be expressed in terms of the original variables as

$$\mathbf{Q}^{-1}\mathbf{y}_t = \sum_{j=0}^{t} \mathbf{M}^j \mathbf{Q}^{-1}\mathbf{b}_{t-j}$$

or

$$\begin{aligned}\mathbf{y}_t &= \sum_{j=0}^{t} \mathbf{Q}\mathbf{M}^j\mathbf{Q}^{-1}\mathbf{b}_{t-j} \\ &= \sum_{j=0}^{t} \mathbf{A}^j\mathbf{b}_{t-j}, \end{aligned} \tag{2.4.26}$$

where the last representation follows from $\mathbf{A}^2 = \mathbf{Q}\mathbf{M}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{M}\mathbf{Q}^{-1} = \mathbf{Q}\mathbf{M}^2\mathbf{Q}^{-1}$, etc.

If the $n \times n$ matrix $\mathbf{A}$ has multiple roots, it cannot always be reduced to a diagonal by a transformation matrix $\mathbf{Q}$. However, $\mathbf{A}$ can be reduced to a matrix whose form is slightly more complicated.

Theorem 2.4.2. Let $\mathbf{\Lambda}_i$ be the $k_i \times k_i$ matrix

$$\mathbf{\Lambda}_i = \begin{bmatrix} m_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & m_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & m_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & m_i \end{bmatrix},$$

where $\mathbf{\Lambda}_i = m_i$ when $k_i = 1$. Then there exists a matrix $\mathbf{Q}$ such that

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Lambda}_r \end{bmatrix} = \mathbf{\Lambda}, \tag{2.4.27}$$

where $\sum_{i=1}^{r} k_i = n$ and the m_i, $i = 1, 2, \ldots, r$, are the characteristic roots of $\mathbf{A}$. If m_i is a repeated root, it may appear in more than one block of $\mathbf{\Lambda}$, but the total number of times it appears is equal to its multiplicity.

Proof. See Finkbeiner (1960) or Miller (1963). ▲

The representation Λ is called the *Jordan canonical form*. The powers of the matrix Λ_i are given by

$$\Lambda_i^j = \begin{pmatrix} m_i^j & \binom{j}{1} m_i^{j-1} & \binom{j}{2} m_i^{j-2} & \cdots & \binom{j}{k_i - 1} m_i^{j-k_i+1} \\ 0 & m_i^j & \binom{j}{1} m_i^{j-1} & \cdots & \binom{j}{k_i - 2} m_i^{j-k_i+2} \\ 0 & 0 & m_i^j & \cdots & \binom{j}{k_i - 3} m_i^{j-k_i+3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & m_i^j \end{pmatrix}. \quad (2.4.28)$$

Therefore, the solution of the first order vector difference equation with initial conditions $\mathbf{y}_0 = \mathbf{b}_0$ is given by

$$\mathbf{y}_t = \sum_{j=0}^{t} \mathbf{Q}\Lambda^j \mathbf{Q}^{-1} \mathbf{b}_{t-j}, \quad (2.4.29)$$

and the effect of the initial conditions goes to zero as $t \to \infty$ if all of the roots of $|\mathbf{A} - m\mathbf{I}| = 0$ are less than one in absolute value.

To further illustrate the solution, consider a system of dimension two with repeated root that has been reduced to the Jordan canonical form:

$$y_{1t} = m y_{1,t-1} + y_{2,t-1},$$
$$y_{2t} = m y_{2,t-1}.$$

It follows that

$$y_{2t} = y_{20} m^t$$

and y_{1t} is given as the solution of the nonhomogeneous equation

$$y_{1t} = m y_{1,t-1} + y_{20} m^{t-1},$$

whence

$$y_{1t} = y_{10} m^t + y_{20} t m^{t-1}.$$

This approach may be extended to treat higher order vector difference equations. For example, consider the second order vector difference equation

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2}, \quad (2.4.30)$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are $k \times k$ matrices and $\mathbf{y}_t' = (y_{1t}, y_{2t}, \ldots, y_{kt})$. The system of equations may also be written as

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}, \tag{2.4.31}$$

where $\mathbf{x}_t' = (\mathbf{y}_t', \mathbf{y}_{t-1}')$ and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix}. \tag{2.4.32}$$

Thus we have converted the original second order vector difference equation of dimension k to a first order difference equation of dimension $2k$. The properties of the solution will depend on the nature of the Jordan canonical form of $\mathbf{A}$. In particular the effect of the initial conditions will be transient if all of the roots of $|\mathbf{A} - m\mathbf{I}| = 0$ are less than one in absolute value. This condition is equivalent to the condition that the roots of

$$|\mathbf{I}m^2 - \mathbf{A}_1 m - \mathbf{A}_2| = 0$$

are less than one in absolute value. To prove this equivalence, form the matrix product

$$\begin{pmatrix} \mathbf{I} & \mathbf{A}_2 m^{-1} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} A_1 - m\mathbf{I} & \mathbf{A}_2 \\ \mathbf{I} & -m\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ m^{-1}\mathbf{I} & \mathbf{I} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{A}_1 - m\mathbf{I} + \mathbf{A}_2 m^{-1} & \mathbf{0} \\ \mathbf{0} & -m\mathbf{I} \end{pmatrix}$$

and take the determinant of both sides.

Let the nth order vector difference equation be given by

$$y_t + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \cdots + A_n \mathbf{y}_{t-n} = \mathbf{r}_t, \qquad t = 1, 2, \ldots, \tag{2.4.33}$$

where $\mathbf{y}_t$ is a k-vector and $\mathbf{r}_t$ is a real vector valued function of time. The associated A-matrix is

$$\mathbf{A} = \begin{pmatrix} -\mathbf{A}_1 & -\mathbf{A}_2 & \cdots & -\mathbf{A}_{n-1} & -\mathbf{A}_n \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}.$$

Then the solution is

$$\mathbf{y}_t = \sum_{j=0}^{t-1} \mathbf{W}_j \mathbf{r}_{t-j} + \tau_t,$$

where the $\mathbf{W}_j$ are $k \times k$ matrices defined by

$$\begin{aligned} \mathbf{W}_0 &= \mathbf{I}, \\ \mathbf{W}_1 + \mathbf{A}_1 &= \mathbf{0}, \\ \mathbf{W}_2 + \mathbf{A}_1\mathbf{W}_1 + \mathbf{A}_2 &= \mathbf{0}, \\ &\vdots \\ \mathbf{W}_j + \mathbf{A}_1\mathbf{W}_{j-1} + \cdots + \mathbf{A}_n\mathbf{W}_{j-n} &= \mathbf{0}, \qquad j = n, n+1, \ldots, \end{aligned}$$

τ_t is the vector composed of the first k entries in the nk-vector $\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0$, and $\mathbf{x}_0' = (\mathbf{y}_0', \mathbf{y}_{-1}', \ldots, \mathbf{y}_{1-n}')$. It follows that $\mathbf{W}_j$ is the upper left block of the matrix $\mathbf{A}^j$.

2.5. THE SECOND ORDER AUTOREGRESSIVE TIME SERIES

We consider the stationary second order autoregressive time series defined by the stochastic difference equation

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = e_t, \qquad t \in (0, \pm 1, \pm 2, \ldots), \tag{2.5.1}$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables, $\alpha_2 \neq 0$, and the roots of $m^2 + \alpha_1 m + \alpha_2 = 0$ are less than one in absolute value. Recalling that we were able to express the first order autoregressive process as a weighted average of past values of e_t, we postulate that X_t can be expressed as

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}, \tag{2.5.2}$$

where the w_j are functions of α_1 and α_2. We display this representation as

$$\begin{aligned} X_t &= w_0 e_t + w_1 e_{t-1} + w_2 e_{t-2} + w_3 e_{t-3} + \cdots, \\ X_{t-1} &= w_0 e_{t-1} + w_1 e_{t-2} + w_2 e_{t-3} + \cdots, \\ X_{t-2} &= w_0 e_{t-2} + w_1 e_{t-3} + \cdots. \end{aligned}$$

If the X_t are to satisfy the difference equation (2.5.1), then we must have

$$\begin{aligned} w_0 &= 1, \\ w_1 + \alpha_1 w_0 &= 0, \\ w_j + \alpha_1 w_{j-1} + \alpha_2 w_{j-2} &= 0, \qquad j = 2, 3, \ldots. \end{aligned} \tag{2.5.3}$$

Thus the w_j are given by the general solution to the second order linear homogeneous difference equation subject to the initial conditions specified by the equations in w_0 and w_1. Therefore, if the roots are distinct, we have

$$w_j = (m_1 - m_2)^{-1} m_1^{j+1} + (m_2 - m_1)^{-1} m_2^{j+1}, \qquad j = 0, 1, 2 \ldots,$$

and if the roots are equal,

$$w_j = (1+j)m^j, \qquad j = 0, 1, 2, \ldots .$$

In Theorem 2.6.1 of the next section, we prove that the representation (2.5.2) holds for a stationary time series satisfying (2.5.1).

The covariance function for the second order autoregressive time series defined in (2.5.2) is

$$\gamma(h) = \sigma^2 \sum_{j=h}^{\infty} w_j w_{j-h}, \qquad h = 0, 1, 2, \ldots . \tag{2.5.4}$$

We shall investigate the covariance function in a slightly different manner. Multiply (2.5.1) by X_{t-h} $(h \geq 0)$ to obtain

$$X_t X_{t-h} + \alpha_1 X_{t-1} X_{t-h} + \alpha_2 X_{t-2} X_{t-h} = e_t X_{t-h} . \tag{2.5.5}$$

Because X_t is expressible as a weighted average of e_t and previous e's,

$$E\{X_{t-j} e_t\} = 0, \qquad j \geq 1 .$$

Thus, taking the expectation of both sides of (2.5.5), we have

$$\gamma(h) + \alpha_1 \gamma(h-1) + \alpha_2 \gamma(h-2) = \begin{cases} 0, & h > 0, \\ \sigma^2, & h = 0. \end{cases} \tag{2.5.6}$$

That is, the covariance function, for $h > 0$, satisfies the homogeneous difference equation associated with the stochastic difference equation defining the time series.

The equations (2.5.6) are called the *Yule–Walker equations*. With the aid of these equations we can obtain a number of alternative expressions for the autocovariances, autocorrelations, and coefficients of the original representation. Using the two equations associated with $h = 1$ and $h = 2$, we may solve for α_1 and α_2 as functions of the autocovariances:

$$\alpha_1 = \frac{\gamma(1)\gamma(2) - \gamma(0)\gamma(1)}{\gamma^2(0) - \gamma^2(1)}, \qquad \alpha_2 = \frac{\gamma^2(1) - \gamma(0)\gamma(2)}{\gamma^2(0) - \gamma^2(1)} . \tag{2.5.7}$$

If we use the three equations (2.5.6) associated with $h = 0, 1, 2$, we can obtain expressions for the autocovariances in terms of the coefficients:

$$\begin{aligned} \gamma(0) &= \frac{(1+\alpha_2)\sigma^2}{(1-\alpha_2)[(1+\alpha_2)^2 - \alpha_1^2]}, \\ \gamma(1) &= \frac{-\alpha_1 \sigma^2}{(1-\alpha_2)[(1+\alpha_2)^2 - \alpha_1^2]}, \\ \gamma(2) &= \frac{[\alpha_1^2 - \alpha_2(1+\alpha_2)]\sigma^2}{(1-\alpha_2)[(1+\alpha_2)^2 - \alpha_1^2]} . \end{aligned} \tag{2.5.8}$$

Using $\gamma(0)$ and $\gamma(1)$ of (2.5.8) as the initial conditions together with the general solution of the homogeneous equation, we can obtain the expression for the covariance function.

However, it is somewhat simpler to derive the correlation function first. To obtain $\rho(1)$ as a function of the coefficients, we divide equation (2.5.6) for $h = 1$ by $\gamma(0)$, giving

$$\rho(1) + \alpha_1 + \alpha_2\rho(1) = 0$$

and

$$\rho(1) = -\frac{\alpha_1}{1+\alpha_2} = \frac{m_1 + m_2}{1 + m_1 m_2}. \tag{2.5.9}$$

Using (2.5.9) and $\rho(0) = 1$ as initial conditions, we have, for unequal roots (real or complex),

$$\rho(h) = [(m_1 - m_2)(1 + m_1 m_2)]^{-1}[m_1^{h+1}(1 - m_2^2) - m_2^{h+1}(1 - m_1^2)], \tag{2.5.10}$$

$$h = 0, 1, 2, \ldots .$$

If the roots are complex, the autocorrelation function may also be expressed as

$$\rho(h) = \frac{r^h \sin(h\theta + \delta)}{\sin \delta}, \tag{2.5.11}$$

where

$$r = \alpha_2^{1/2} = (m_1 m_2)^{1/2},$$

$$\cos\theta = -\frac{\alpha_1}{2\alpha_2^{1/2}} = \frac{m_1 + m_2}{2(m_1 m_2)^{1/2}},$$

$$\tan\delta = \frac{1+\alpha_2}{1-\alpha_2}\tan\theta .$$

For roots real and equal the autocorrelation function is given by

$$\rho(h) = \left[1 + h\left(\frac{1-m^2}{1+m^2}\right)\right]m^h, \qquad h = 0, 1, 2, \ldots . \tag{2.5.12}$$

If we substitute $\alpha_1 = -(m_1 + m_2)$ and $\alpha_2 = m_1 m_2$ into (2.5.8), we can express the variance as a function of the roots:

$$\gamma(0) = \frac{1 + m_1 m_2}{(1 - m_1 m_2)(1 - m_1^2)(1 - m_2^2)}\sigma^2 . \tag{2.5.13}$$

Equation (2.5.13) together with (2.5.10) and (2.5.12) may be used to express the covariance function in terms of the roots.

Example 2.5.1. Figure 2.5.1 contains a plot of the correlation function, sometimes called the *correlogram*, for the time series

$$X_t = 1.40X_{t-1} - 0.48X_{t-2} + e_t .$$

The roots of the auxiliary equation are 0.8 and 0.6. The correlogram has much the same appearance as that of a first order autoregressive process with large ρ. The empirical correlograms of many economic time series have this general appearance, the first few correlations being quite large.

The correlogram in Figure 2.5.2 is that associated with the time series

$$X_t = X_{t-1} - 0.89X_{t-2} + e_t .$$

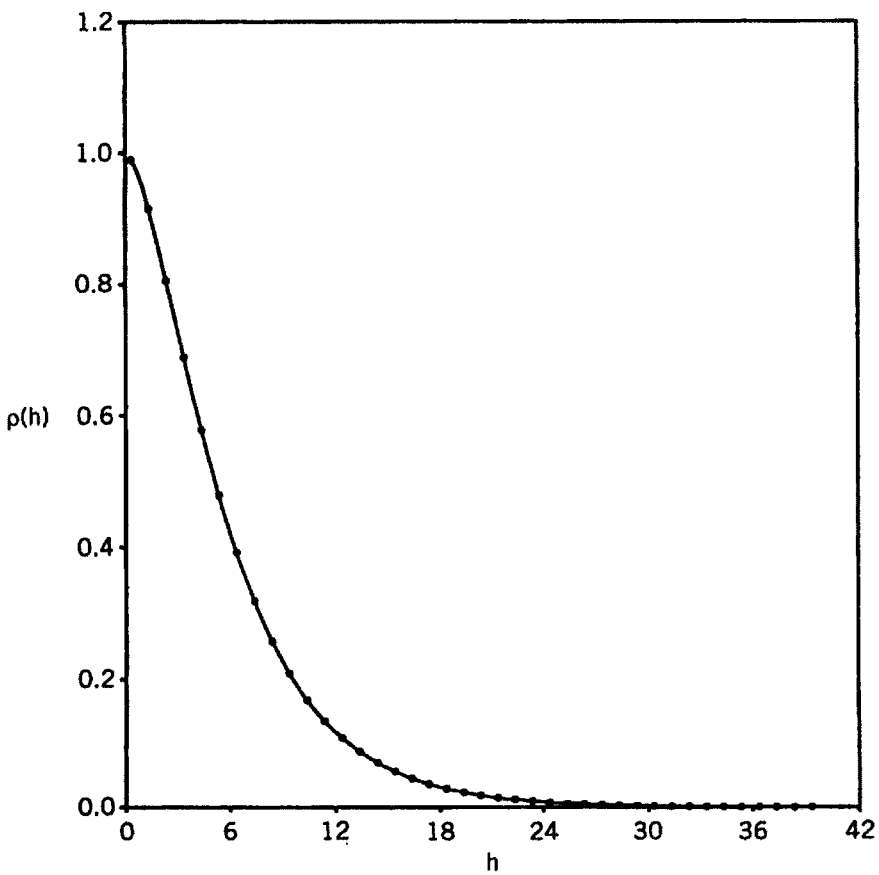

Figure 2.5.1. Correlogram for $X_t = 1.40X_{t-1} - 0.48X_{t-2} + e_t$.

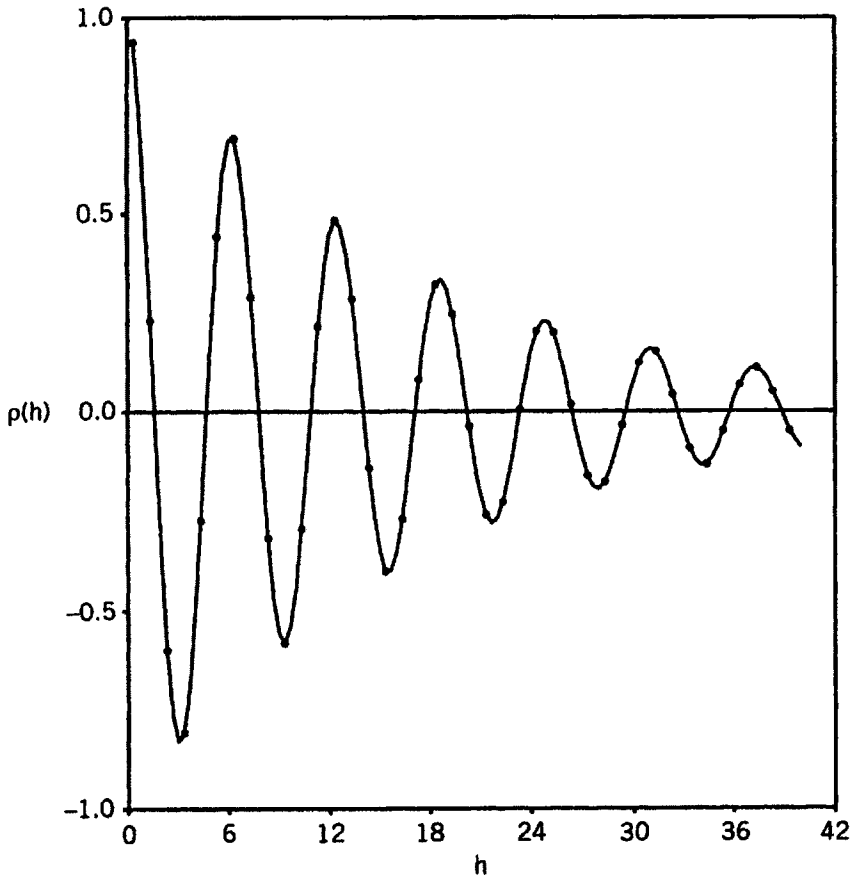

Figure 2.5.2. Correlogram for $X_t = X_{t-1} - 0.89X_{t-2} + e_t$.

The roots of the auxiliary equation are the complex pair $0.5 \pm 0.8i$. In this case the correlogram has the distinctive "declining cyclical" appearance associated with complex roots. This correlogram illustrates how such a process can generate a time series with the appearance of moderately regular "cycles." In the present case $\cos \theta \doteq 0.53$ and $\theta \doteq 0.322\ \pi$, where $\cos \theta$ is defined in (2.5.11). Thus the apparent period of the cyclical behavior would be $2(0.322)^{-1} \doteq 6.2$ time units. ▲▲

2.6. ALTERNATIVE REPRESENTATIONS OF AUTOREGRESSIVE AND MOVING AVERAGE PROCESSES

We have argued that the stationary second order autoregressive time series can also be represented as an infinite moving average time series. We now prove that a time series satisfying a pth order difference equation has a representation as an infinite moving average.

Theorem 2.6.1. Let $\{X_t\}$ be a time series defined on the integers with $E\{X_t^2\} < K$ for all t. Suppose X_t satisfies

$$X_t + \sum_{j=1}^{p} \alpha_j X_{t-j} = e_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2.6.1}$$

where $\{e_t\}$, $t = 0, \pm 1, \pm 2, \ldots$, is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Let $m_1, m_2, \ldots, m_p$ be the roots of

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0, \tag{2.6.2}$$

and assume $|m_i| < 1$, $i = 1, 2, \ldots, p$. Then X_t is covariance stationary. Furthermore, X_t is given as a limit in mean square by

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}, \tag{2.6.3}$$

where $\{w_j\}_{j=0}^{\infty}$ is the unique solution of the homogeneous difference equation

$$w_j + \alpha_1 w_{j-1} + \cdots + \alpha_p w_{j-p} = 0, \qquad j = p, p+1, \ldots, \tag{2.6.4}$$

subject to the boundary conditions $w_0 = 1$ and

$$w_j + \sum_{i=1}^{j} \alpha_i w_{j-i} = 0, \qquad j = 1, 2, \ldots, p-1.$$

The limit (2.6.3) is also an almost sure limit.

Proof. Let $\mathbf{Y}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})'$, $\boldsymbol{\epsilon}_t = (e_t, 0, \ldots, 0)'$, and

$$\mathbf{A} = \begin{pmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & \cdots & -\alpha_{p-1} & -\alpha_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{2.6.5}$$

Then

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t = \mathbf{A}^n \mathbf{Y}_{t-n} + \sum_{j=0}^{n-1} \mathbf{A}^j \boldsymbol{\epsilon}_{t-j}$$

for $n \geq 1$ and $t = 0, \pm 1, \pm 2, \ldots$. Therefore,

$$X_t = A^n_{(11)}X_{t-n} + A^n_{(12)}X_{t-n-1} + \cdots + A^n_{(1p)}X_{t-n-p+1} + \sum_{j=0}^{n-1} A^j_{(11)}e_{t-j}$$

$$= \sum_{j=1}^{p} A^n_{(1j)}X_{t-n-j+1} + \sum_{j=0}^{n-1} A^j_{(11)}e_{t-j},$$

where $A^r_{(ij)}$ is the ijth element of the matrix $\mathbf{A}^r$. Consider the determinant

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda + \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ -1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & -1 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & \lambda \end{vmatrix}.$$

Beginning with column 1, successively multiply column i by λ and add the result to column $i + 1$ to obtain

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda + \alpha_1 & \sum_{j=0}^{2} \lambda^{2-j}\alpha_j & \sum_{j=0}^{3} \lambda^{3-j}\alpha_j & \cdots & \sum_{j=0}^{p} \lambda^{p-j}\alpha_j \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{vmatrix}$$

$$= (\lambda^p + \alpha_1\lambda^{p-1} + \cdots + \alpha_{p-1}\lambda + \alpha_p)(-1)^{2p-1}.$$

By the Cayley–Hamilton theorem,

$$\mathbf{A}^p + \alpha_1\mathbf{A}^{p-1} + \cdots + \alpha_{p-1}\mathbf{A} + \alpha_p\mathbf{I} = \mathbf{0}. \tag{2.6.6}$$

For $j \geq p$ multiply the matrix equation (2.6.6) by $\mathbf{A}^{j-p}$ to obtain

$$\mathbf{A}^j + \alpha_1\mathbf{A}^{j-1} + \cdots + \alpha_{p-1}\mathbf{A}^{j-p+1} + \alpha_p\mathbf{A}^{j-p} = 0, \quad j \geq p. \tag{2.6.7}$$

Set $w_j = A^j_{(11)}$, the first element of $\mathbf{A}^j$. Then

$$w_j + \alpha_1 w_{j-1} + \alpha_2 w_{j-2} + \cdots + \alpha_p w_{j-p} = 0, \qquad j \geq p.$$

It is readily verified that $w_j = A^j_{(11)}$ satisfies the initial conditions associated with (2.6.4) in the statement of the theorem. Therefore,

$$X_t = \sum_{j=1}^{p} A^n_{(1j)} X_{t-n-j+1} + \sum_{j=0}^{n-1} w_j e_{t-j}$$

and

$$E\left[\left(X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right)^2\right] = E\left[\left(\sum_{j=1}^{p} A^n_{(1j)} X_{t-n-j+1}\right)^2\right]$$

for $n = 1, 2, \ldots$. Because each element of $\mathbf{A}^n$ satisfies the homogeneous difference equation (2.6.7), we have

$$A^j_{(1i)} + \alpha_1 A^{j-1}_{(1i)} + \cdots + \alpha_p A^{j-p}_{(1i)} = 0 \qquad \text{for} \quad j \geq p,$$

and there exists a c such that $|A^j_{(1i)}| < c\lambda^j$ for $i = 1, 2, \ldots, p$, where $1 > \lambda > M$ and M is the largest of the absolute values of the roots m_i. See Exercise 2.24. It follows that

$$E\left[\left(X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right)^2\right] \leq (cp\lambda^n)^2 K \to 0$$

as $n \to \infty$, because $0 < \lambda < 1$. Because $\sum_{j=0}^{\infty} |w_j| < \infty$, X_t is covariance stationary. Now,

$$E\left\{\left|X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right|\right\} \leq \sum_{j=1}^{p} |A^n_{(1j)}| E\{|X_{t-n-j+1}|\} \leq cp\lambda^n K^{1/2},$$

and it follows that the limit (2.6.3) holds almost surely. ▲

Because X_t can be written as a weighted sum of current and past e_t's, it follows that e_t is uncorrelated with X_{t-h} for $h = 1, 2, \ldots$. This enables us to derive many useful properties of the autoregressive time series and of the autocovariance function of the autoregressive time series.

Corollary 2.6.1.1. Let X_t be stationary and satisfy

$$X_t + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} = e_t,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables and the roots of the characteristic polynomial

$$m^p + \alpha_1 m^{p-1} + \cdots + \alpha_p = 0$$

are less than one in absolute value. Then

$$\gamma(0) + \alpha_1\gamma(1) + \cdots + \alpha_p\gamma(p) = \sigma^2$$

and

$$\gamma(h) + \alpha_1\gamma(h-1) + \cdots + \alpha_p\gamma(h-p) = 0\,, \qquad h = 1, 2, \ldots\,. \tag{2.6.8}$$

Proof. Reserved for the reader. See equation (2.5.6). ▲

The system (2.6.8) defines the α's in terms of the γ's and vice versa. For example, for a stationary third order autoregressive process, we have

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} -\gamma(1) \\ -\gamma(2) \\ -\gamma(3) \end{pmatrix} \tag{2.6.9}$$

and

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1 & 1+\alpha_2 & \alpha_3 & 0 \\ \alpha_2 & \alpha_1+\alpha_3 & 1 & 0 \\ \alpha_3 & \alpha_2 & \alpha_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.6.10}$$

Using the representation results, we show that the partial autocorrelation function, introduced in equation (1.4.12), of a pth order autoregressive process is zero for $k > p$.

Corollary 2.6.1.2. Let X_t be a stationary pth order autoregressive process with all of the roots of the characteristic polynomial less than one in absolute value, and let $\phi(k)$ be the partial autocorrelation function. Then

$$\phi(k) = 0\,, \qquad k = p+1, p+2, \ldots\,.$$

Proof. By the definition of the partial autocorrelation, $\phi(p+1)$ is the correlation between the residual obtained in the population regression of X_t on $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ and the residual obtained in the population regression of X_{t-p-1} on $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$. By the definition of the autoregressive process, the residual obtained in the population regression of X_t on $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ is e_t, and the coefficients are $-\alpha_i$, $i = 1, 2, \ldots, p$. By the representation of Theorem 2.6.1, e_t is uncorrelated with X_{t-i}, $i \geq 1$ and hence with a linear combination of the

X_{t-i}, $i \geq 1$. Therefore, θ_{kk} of (1.4.11), and hence $\phi(k)$, is zero for $k = p + 1$. If $\theta_{ik} = -\alpha_i$ for $i = 1, 2, \ldots, p$ and $\theta_{ik} = 0$ for $i = p + 1, \ldots, k$, then $\theta_{k+1,k+1} = 0$ and the result follows. ▲

Whenever discussing stationary autoregressive time series, we have assumed that the roots of the auxiliary equation were less than one in absolute value. This is because we have explicitly or implicitly visualized the time series as being created in a forward manner. If a time series is created in a forward manner, the effect of the initial conditions will go to zero only if the roots are less than one in absolute value. For example, if we define

$$X_t = \begin{cases} a_0 e_0, & t = 0, \\ \rho X_{t-1} + e_t, & t = 1, 2, \ldots, \end{cases} \tag{2.6.11}$$

where it is understood that X_t is formed by adding e_t to ρX_{t-1} and e_t is a sequence of uncorrelated $(0, \sigma^2)$ random variables, then $\{X_t : t = 0, 1, 2, \ldots\}$ is stationary for $|\rho| < 1$ and $a_0 = (1 - \rho^2)^{-1/2}$. However, for $|\rho| \geq 1$, the time series formed by adding e_t to ρX_{t-1} is nonstationary for all a_0.

On the other hand, there is a stationary time series that satisfies the difference equation

$$X_t = \rho X_{t-1} + e_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{2.6.12}$$

for $|\rho| > 1$. To see this, let us consider the stationary time series

$$\begin{aligned} X_r &= 0.8 X_{r-1} + \epsilon_r \\ &= \sum_{j=0}^{\infty} (0.8)^j \epsilon_{r-j}, \qquad r = 0, \pm 1, \pm 2, \ldots, \end{aligned} \tag{2.6.13}$$

where ϵ_r are uncorrelated $(0, \sigma^2)$ random variables. If we change the direction in which we count on the integers so that $-r = t$ and $r - 1 = -t + 1$, and divide (2.6.13) by 0.8, we have

$$X_{t+1} = 1.25 X_t - 1.25 \epsilon_t \tag{2.6.14}$$

and

$$X_t = \sum_{j=0}^{\infty} (0.8)^j \epsilon_{t+j}. \tag{2.6.15}$$

By setting $e_{t+1} = -1.25\epsilon_t$, equation (2.6.14) can be written in the form (2.6.12) with $\rho = 1.25$.

The covariance function for the time series (2.6.15) is the same as the covariance function of the time series (2.6.13). Thus, if a time series $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ has an autocorrelation function of the form $\rho^{|h|}$, where $0 < |\rho| < 1$, it can be written as a forward moving average of uncorrelated random variables

or as a backward moving average of uncorrelated random variables. Likewise,

$$X_t - \rho X_{t-1} = e_t \tag{2.6.16}$$

defines a sequence of uncorrelated random variables, as does

$$X_t - \rho^{-1} X_{t-1} = Z_t . \tag{2.6.17}$$

From the representations (2.6.13) and (2.6.15), one sees that the only stationary time series with a unit root is the trivial time series that is a constant (with probability one) for all t. See Exercise 2.19.

While both the e_t of (2.6.16) and the Z_t of (2.6.17) are uncorrelated random variables, the variance of Z_t is larger than the variance of the e_t by a factor of ρ^{-2}. This explains why the representation (2.6.16) is the one that appears in the applications of stationary autoregressive processes. That is, one typically chooses the ρ in an equation such as (2.6.16) to minimize the variance of e_t. It is worth mentioning that nonstationary autoregressive representations with roots greater than or equal to one in absolute value have appeared in practice. Estimation for such time series is discussed in Chapter 10. That the stationary autoregressive time series can be given either a forward or a backward representation also finds some use, and we state the result before proceeding.

Corollary 2.6.1.3. Let the covariance function of the time series $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ with zero mean satisfy the difference equation

$$\sum_{j=0}^{p} \alpha_j \gamma_X(h-j) = 0 , \qquad h = 1, 2, \ldots ,$$

where $\alpha_0 = 1$ and the roots of the characteristic equation

$$\sum_{j=0}^{p} \alpha_j m^{p-j} = 0$$

are less than one in absolute value. Then X_t satisfies the stochastic difference equation

$$\sum_{j=0}^{p} \alpha_j X_{t-j} = e_t ,$$

where $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables, and also satisfies the stochastic difference equation

$$\sum_{j=0}^{p} \alpha_j X_{t+j} = v_t ,$$

where $\{v_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables.

Proof. Omitted. ▲

Having demonstrated that a stationary finite autoregressive time series can be given an infinite moving average representation, we now obtain an alternative representation for the finite moving average time series. Because the finite moving average time series can be viewed as a difference equation in e_t, we have a result parallel to that of Theorem 2.6.1.

Theorem 2.6.2. Let the time series $\{X_t: t \in (0, \pm 1, \ldots)$ be defined by

$$X_t = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q}, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

where $b_q \neq 0$, the roots of the characteristic equation

$$m^q + b_1 m^{q-1} + b_2 m^{q-2} + \cdots + b_q = 0$$

are less than one in absolute value, and $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Then X_t can be expressed as an infinite autoregressive process

$$\sum_{j=0}^{\infty} c_j X_{t-j} = e_t, \tag{2.6.18}$$

where the coefficients c_j satisfy the homogeneous difference equation

$$c_j + b_1 c_{j-1} + b_2 c_{j-2} + \cdots + b_q c_{j-q} = 0, \qquad j = q, q+1, \ldots, \tag{2.6.19}$$

with the initial conditions $c_0 = 1$, $c_1 = -b_1$,

$$c_2 = -b_1 c_1 - b_2,$$

$$\vdots$$

$$c_{q-1} = -b_1 c_{q-2} - b_2 c_{q-3} - \cdots - b_{q-1}.$$

Proof. Reserved for the reader. ▲

The autocorrelation function of the pth autoregressive process is nonzero for some integer greater than N_0 for all N_0. See Exercise 2.33. By Corollary 2.6.1.2, the partial autocorrelation function of the pth order autoregressive process is zero for argument greater than p. The opposite conditions hold for the autocorrelations of the moving average process.

Corollary 2.6.2.1. Let the time series X_t be the qth order moving average defined in Theorem 2.6.2. Let N_0 be given. Then there is some $k > N_0$ such that the partial autocorrelation function $\phi(k) \neq 0$.

Proof. By our definition, $X_t = \Sigma_{i=0}^{q} b_i e_{t-i}$, where $b_0 = 1$ and $b_q \neq 0$. If $\phi(k) =$

0 for all $k > N_0$, then the coefficients c_j in Theorem 2.6.2 must be zero for $j > N_0$. By Theorem 2.6.2 the coefficients c_j satisfy the difference equation (2.6.19). If all c_j are equal to zero for $j > N_0$, then c_{N_0} must be zero if it is to satisfy (2.6.19). Because $b_q \neq 0$, this leads to the conclusion that all c_j are zero, which contradicts the initial condition for $(c_0, c_1, \ldots, c_{q-1})$. ▲

In discussing finite moving averages we placed no restrictions on the coefficients, and, for example, the time series

$$Y_t = e_t - e_{t-1} \tag{2.6.20}$$

is clearly stationary. The root of the auxiliary equation for this difference equation is one, and therefore the condition of Theorem 2.6.2 is not met. An attempt to express e_t as an autoregressive process using that theorem will fail because the remainder associated with an autoregressive representation of order n will be e_{t-n-1}.

Time series satisfying the conditions of Theorem 2.6.2 are sometimes called *invertible* moving averages. From the example (2.6.20) we see that not all moving average processes are invertible.

In our earlier discussion of the moving average time series we demonstrated that we could always assign the value one to the coefficient of e_t. We were also able to obtain all autocorrelation functions of the first order moving average type for an α restricted to the range $[-1, 1]$. We are now in a position to generalize this result.

Theorem 2.6.3. Given a time series X_t with zero mean and autocorrelation function

$$\rho_X(h) = \begin{cases} 1, & h = 0, \\ \rho(1), & h = 1, \\ 0, & h > 1, \end{cases}$$

where $|\rho(1)| < 0.5$, there exists an α, $|\alpha| < 1$, and a sequence of uncorrelated random variables $\{e_t\}$ such that X_t is defined by

$$X_t = e_t + \alpha e_{t-1}.$$

Proof. The equation $\rho(1) = (1 + \alpha^2)^{-1}\alpha$ in α has one root that is less than one in absolute value and one that is greater than one in absolute value. The root of smaller absolute value is chosen, and we define e_t by

$$e_t = \sum_{j=0}^{\infty} (-\alpha)^j X_{t-j}.$$

By Theorem 2.2.1 this random variable is well defined as a limit in squared mean,

and by Theorem 2.2.2,

$$\gamma_e(0) = E\{e_t^2\} = \sum_{j=0}^{\infty} (-\alpha)^{2j}\gamma_X(0) - 2\alpha \sum_{j=0}^{\infty} (-\alpha)^{2j}\gamma_X(1)$$

$$= \left(\frac{1}{1-\alpha^2} - \frac{2\alpha}{1-\alpha^2}\frac{\alpha}{1+\alpha^2}\right)\gamma_X(0) = \frac{\gamma_X(0)}{1+\alpha^2},$$

$$\gamma_e(h) = \sum_{j=0}^{\infty} (-\alpha)^{h+2j}\gamma_X(0) + \sum_{j=0}^{\infty} (1+\alpha^2)(-\alpha)^{h-1+2j}\gamma_X(1)$$

$$= \left(\frac{(-\alpha)^h}{1-\alpha^2} + \frac{(1+\alpha^2)(-\alpha)^{h-1}}{1-\alpha^2}\frac{\alpha}{1+\alpha^2}\right)\gamma_X(0)$$

$$= 0, \qquad h > 0. \qquad \blacktriangle$$

The dualities between autoregressive and moving average representations established in Theorems 2.6.1 and 2.6.2 can also be described using formal operations with the backward shift operator. We recall that the first order autoregressive time series

$$Y_t - \rho Y_{t-1} = (1 - \rho\mathcal{B})Y_t = e_t, \qquad |\rho| < 1,$$

can be written as

$$Y_t = \sum_{j=0}^{\infty} \rho^j e_{t-j} = \left[\sum_{j=0}^{\infty} (\mathcal{B}\rho)^j\right] e_t = \frac{1}{1-\rho\mathcal{B}} e_t,$$

where it is understood that $\mathcal{B}^0$ is the identity operator, and we have written

$$(1 - \rho\mathcal{B})^{-1} = \sum_{j=0}^{\infty} (\mathcal{B}\rho)^j$$

in analogy to the expansion that holds for a real number $|\rho\mathcal{B}| < 1$.

The fact that the operator can be formally manipulated as if it were a real number with absolute value 1 furnishes a useful way of obtaining the alternative expressions for moving average and autoregressive processes whose characteristic equations have roots less than one in absolute value. Thus the second order autoregressive time series

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = e_t$$

can also be written as

$$(1 + a_1\mathcal{B} + a_2\mathcal{B}^2)X_t = e_t$$

or

$$(1 - m_1\mathcal{B})(1 - m_2\mathcal{B})X_t = e_t$$

or

$$X_t = [(1 - m_1 \mathcal{B})(1 - m_2 \mathcal{B})]^{-1} e_t ,$$

where m_1 and m_2 are the roots of the characteristic equation

$$m^2 + a_1 m + a_2 = 0 .$$

Since the roots of the quadratic in $\mathcal{B}$

$$1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 = 0$$

are the reciprocals of the roots of the characteristic equation, the restriction on the roots is often stated as the requirement that the roots of

$$1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 = 0$$

be greater than one in absolute value.

Using the backward shift operator, the invertible second order moving average

$$Y_t = e_t + b_1 e_{t-1} + b_2 e_{t-2}$$

can be given the representation

$$Y_t = (1 + b_1 \mathcal{B} + b_2 \mathcal{B}^2) e_t$$

or

$$Y_t = (1 - g_1 \mathcal{B})(1 - g_2 \mathcal{B}) e_t ,$$

where g_1 and g_2, the roots of

$$g^2 + b_1 g + b_2 = 0 ,$$

are less than one in absolute value. The autoregressive representation of the moving average process is then given by

$$e_t = [(1 - g_1 \mathcal{B})(1 - g_2 \mathcal{B})]^{-1} Y_t .$$

The backward shift representation of the moving average process is useful in illustrating the fact that any finite moving average process whose characteristic equation has some roots greater than one and some less than one can be given a representation whose characteristic equation has all roots less than one in absolute value.

Theorem 2.6.4. Let $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ have the representation

$$X_t = \prod_{i=1}^{p} (1 - m_i \mathcal{B}) e_t ,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables and none of the m_i, $i = 1, 2, \ldots, p$, are of unit absolute value. Let $m_1, m_2, \ldots, m_L$, $0 < L \leq p$, be greater than one in absolute value. Then X_t also has the representation

$$X_t = \begin{cases} \prod_{i=1}^{L} (1 - m_i^{-1}\mathcal{B}) \prod_{i=L+1}^{p} (1 - m_i\mathcal{B})\epsilon_t, & 0 < L < p \\ \prod_{i=1}^{p} (1 - m_i^{-1}\mathcal{B})\epsilon_t, & L = p, \end{cases}$$

where the ϵ_t are uncorrelated $(0, \sigma^2[\Pi_{i=1}^{L} m_i]^2)$.

Proof. We define the polynomial $\mathcal{Q}(\mathcal{B})$ in $\mathcal{B}$ by

$$\mathcal{Q}(\mathcal{B}) = \begin{cases} \prod_{i=1}^{L} (1 - m_i^{-1}\mathcal{B}) \prod_{i=L+1}^{p} (1 - m_i\mathcal{B}), & 0 < L < p, \\ \prod_{i=1}^{p} (1 - m_i^{-1}\mathcal{B}), & L = p. \end{cases}$$

The roots of $\mathcal{Q}(\mathcal{B})$ are greater than one in absolute value, and the roots of $m^p\mathcal{Q}(m^{-1})$ are less than one in absolute value. Therefore the time series

$$\epsilon_t = \sum_{j=0}^{\infty} w_j X_{t-j},$$

where the w_j satisfy the difference equation

$$\mathcal{Q}(\mathcal{B})w_j = 0, \qquad j = p, p+1, \ldots,$$

and the initial conditions outlined in Theorem 2.6.2, is well defined. Let m_1 be a real root greater than one in absolute value, and define Z_t by

$$Z_t = \prod_{i=L+1}^{p} (1 - m_i\mathcal{B})e_t.$$

Then the two time series $Z_t - m_1 Z_{t-1}$ and $m_1(Z_t - m_1^{-1}Z_{t-1})$ have the same covariance function. Likewise if m_1 and $m_2 = m_1^*$ are a complex conjugate pair of roots with absolute value greater than one, $Z_t - (m_1 + m_2)Z_{t-1} + |m_1|^2 Z_{t-2}$ has the same covariance function as $|m_1|^2[Z_t - (m_1^{-1} + m_2^{-1})Z_{t-1} + |m_1|^{-2}Z_{t-2}]$.

By repeated application of these arguments, we conclude that the original time series

$$X_t = \prod_{i=1}^{p} (1 - m_i\mathcal{B})e_t$$

and the time series

$$Y_t = \left(\prod_{i=1}^{L} m_i\right) \mathcal{Q}(\mathcal{B}) e_t$$

have the same covariance function. Therefore,

$$\epsilon_t = \mathcal{Q}^{-1}(\mathcal{B}) X_t$$

and

$$\left(\prod_{i=1}^{L} m_i\right) e_t = \mathcal{Q}^{-1}(\mathcal{B}) Y_t$$

have the same covariance function. ▲

As an example of this theorem, the moving average

$$\begin{aligned} X_t &= e_t - 3.2e_{t-1} - 3.2e_{t-2} \\ &= (1 - 4.0\mathcal{B})(1 + 0.8\mathcal{B})e_t , \end{aligned}$$

where the e_t are uncorrelated $(0, \sigma^2)$, also has the representation

$$\begin{aligned} X_t &= \epsilon_t + 0.55\epsilon_{t-1} - 0.20\epsilon_{t-2} \\ &= (1 - 0.25\mathcal{B})(1 + 0.8\mathcal{B})\epsilon_t , \end{aligned}$$

where the ϵ_t are uncorrelated $(0, 16\sigma^2)$. The reader may check that $\gamma_X(0)$, $\gamma_X(1)$, and $\gamma_X(2)$ are $21.48\sigma^2$, $7.04\sigma^2$, and $-3.20\sigma^2$, respectively, for this moving average.

2.7. AUTOREGRESSIVE MOVING AVERAGE TIME SERIES

Having considered autoregressive and moving average processes, it is natural to investigate time series defined by the combination of low order autoregressive and moving average components. The sequence $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$, defined by

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q} , \qquad (2.7.1)$$

where $a_p \neq 0$, $b_q \neq 0$, and $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables, is called an *autoregressive moving average time series of order* (p, q), which we shall abbreviate to autoregressive moving average (p, q). The notation ARMA(p, q) is also commonly used.

The autoregressive moving average (1, 1),

$$X_t - \theta_1 X_{t-1} = e_t + b_1 e_{t-1} , \qquad (2.7.2)$$

where $|\theta_1| < 1$, has furnished a useful approximation for some time series encountered in practice. Note that we have defined the coefficient on X_{t-1} as $-\theta_1$ to simplify the representations to follow. If we let

$$u_t = e_t + b_1 e_{t-1}$$

and write

$$X_t = \theta_1 X_{t-1} + u_t ,$$

we can express X_t of (2.7.2) as an infinite moving average of the u_t,

$$X_t = \sum_{j=0}^{\infty} \theta_1^j u_{t-j} ,$$

and hence of the e_t,

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} \theta_1^j e_{t-j} + b_1 \sum_{j=0}^{\infty} \theta_1^j e_{t-j-1} \\ &= e_t + (\theta_1 + b_1) \sum_{j=1}^{\infty} \theta_1^{j-1} e_{t-j} , \end{aligned} \tag{2.7.3}$$

where the random variables are defined by Theorem 2.2.1.

If $|b_1| < 1$, we can also express e_t as an infinite moving average of the u_t,

$$e_t = \sum_{j=0}^{\infty} (-b_1)^j u_{t-j} ,$$

and hence of the X_t,

$$e_t = X_t - (\theta_1 + b_1) \sum_{j=1}^{\infty} (-b_1)^{j-1} X_{t-j} , \tag{2.7.4}$$

If we let Z_t denote the first order autoregressive process

$$Z_t = \sum_{j=0}^{\infty} \theta_1^j e_{t-j} ,$$

then

$$X_t = Z_t + b_1 Z_{t-1} . \tag{2.7.5}$$

That is, the time series can be expressed as a first order moving average of a first order autoregressive time series. Using this representation, we express the autocovariance function of X_t in terms of that of Z_t,

$$\gamma_X(h) = (1 + b_1^2)\gamma_Z(h) + b_1 \gamma_Z(h-1) + b_1 \gamma_Z(h+1) ,$$

to obtain

$$\gamma_X(h) = \left[\frac{1+b_1^2}{1-\theta_1^2}\theta_1^{|h|} + \frac{b_1}{1-\theta_1^2}(\theta_1^{|h-1|} + \theta_1^{|h+1|}) \right] \sigma^2 . \tag{2.7.6}$$

Hence,

$$\gamma_X(h) = \begin{cases} \dfrac{1+b_1^2+2b_1\theta_1}{1-\theta_1^2}\sigma^2, & h=0, \\ \dfrac{(1+b_1\theta_1)(\theta_1+b_1)}{1-\theta_1^2}\theta_1^{h-1}\sigma^2, & h=1,2,\ldots, \end{cases} \tag{2.7.7}$$

and

$$\rho_X(h) = \begin{cases} 1, & h=0, \\ \dfrac{(1+b_1\theta_1)(\theta_1+b_1)}{1+b_1^2+2b_1\theta_1}\theta_1^{h-1}, & h=1,2,\ldots. \end{cases}$$

Thus, for $h \geq 1$, the autocorrelation function of the autoregressive moving average (1, 1) has the same appearance as that of a first order autoregressive time series in that it is declining at a geometric rate where the rate is θ_1.

From equation (2.7.7) we see that $\gamma_X(h)$ is zero for $h \geq 1$ if $b_1 = -\theta_1$. The autoregressive moving average (1, 1) process reduces to a sequence of uncorrelated random variables in such a case, which is also clear from the representation in (2.7.3).

To avoid this kind of degeneracy, the autoregressive moving average of order (p, q) is often defined with the condition that none of the roots of

$$m^p + a_1 m^{p-1} + \cdots + a_p = 0 \tag{2.7.8}$$

are roots of

$$m^q + b_1 m^{q-1} + \cdots + b_q = 0 . \tag{2.7.9}$$

The autoregressive and moving average representations of (2.7.3) and (2.7.4) generalize to higher order processes in the obvious manner. We state the generalizations as theorems.

Theorem 2.7.1. Let the stationary time series X_t satisfy

$$\sum_{j=0}^{p} a_j X_{t-j} = \sum_{i=0}^{q} b_i e_{t-i} , \tag{2.7.10}$$

where $a_0 = b_0 = 1$, the roots of (2.7.8) are less than one in absolute value, and $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Then X_t has the

representation

$$X_t = \sum_{j=0}^{\infty} v_j e_{t-j}, \tag{2.7.11}$$

where $v_0 = 1$, $v_j = 0$ if $j < 0$, and

$$v_j = b_j - \sum_{i=1}^{p} a_i v_{j-i}$$

with $b_j = 0$ for $j > q$.

Proof. Writing

$$\sum_{j=0}^{p} a_j X_{t-j} = u_t,$$

we obtain

$$X_t = \sum_{j=0}^{\infty} w_j u_{t-j} = \sum_{j=0}^{\infty} w_j \left(\sum_{i=0}^{q} b_i e_{t-j-i} \right)$$
$$\overset{\text{(say)}}{=} \sum_{r=0}^{\infty} v_r e_{t-r},$$

where $u_t = \sum_{i=0}^{q} b_i e_{t-i}$ and the w_j are defined in Theorem 2.6.1. Since the w_j are absolutely summable, this representation is well defined as a limit in mean square. We then write

$$\sum_{i=0}^{q} b_i e_{t-i} = \sum_{j=0}^{p} a_j X_{t-j} = \sum_{j=0}^{p} a_j \left(\sum_{r=0}^{\infty} v_r e_{t-j-r} \right)$$

and obtain the result by equating coefficients of e_{t-i}, $i = 0, 1, \ldots$. ▲

The first few v_i of Theorem 2.7.1 can be used to compute the first few covariances of the process. If we multiply the defining equation of Theorem 2.7.1 by X_{t-r} for $r = 0, 1, \ldots, p$ and take expectations, we have

$$\sum_{j=0}^{p} a_j \gamma_X(r-j) = \sum_{i=r}^{q} b_i v_{i-r} \sigma^2,$$

where it is understood that the summation on the right of the equality is zero for $r > q$.

Theorem 2.7.2. Let the stationary time series X_t satisfy

$$\sum_{j=0}^{p} a_j X_{t-j} = \sum_{i=0}^{q} b_i e_{t-i},$$

where $a_0 = b_0 = 1$, the roots of (2.7.8) and of (2.7.9) are less than one in absolute value, and $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Then X_t has the representation

$$e_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \tag{2.7.12}$$

where $d_0 = 1$, $d_j = 0$ for $j < 0$, and

$$d_j = -\sum_{i=1}^{q} b_i d_{j-i} + a_j,$$

with $a_j = 0$ for $j > p$.

Proof. Reserved for the reader. ▲

The backward shift operator is often used to give a compact representation for the autoregressive moving average. Let

$$A(m) = 1 + a_1 m + \cdots + a_p m^p$$

and

$$B(m) = 1 + b_1 m + \cdots + b_q m^q$$

be the polynomials associated with the autoregressive moving average (2.7.10). Then we can write

$$A(\mathcal{B})X_t = B(\mathcal{B})e_t \tag{2.7.13}$$

or, if the roots of $A(m) = 0$ are greater than one in absolute value,

$$X_t = A^{-1}(\mathcal{B})B(\mathcal{B})e_t \tag{2.7.14}$$

or, if the roots of $B(m) = 0$ are also greater than one in absolute value,

$$e_t = B^{-1}(\mathcal{B})A(\mathcal{B})X_t. \tag{2.7.15}$$

The representation (2.7.14) corresponds to (2.7.11), and the representation (2.7.15) corresponds to (2.7.12). Also see Exercise 2.26.

The equation (2.7.13) defines an autoregressive moving average constructed with e_t variables. By (2.7.14), we could apply the same moving average operator to any sequence. Thus, if $C(m)$ and $D(m)$ are polynomials whose roots are greater

than one in absolute value, and if X_t is the time series of (2.7.13), we might define the new time series Y_t by

$$C(\mathcal{B})Y_t = D(\mathcal{B})X_t$$

or

$$\begin{aligned} Y_t &= C^{-1}(\mathcal{B})D(\mathcal{B})X_t \\ &= C^{-1}(\mathcal{B})D(\mathcal{B})A^{-1}(\mathcal{B})B(\mathcal{B})e_t \\ &= G^{-1}(\mathcal{B})H(\mathcal{B})e_t , \end{aligned} \tag{2.7.16}$$

where $G(\mathcal{B}) = C(\mathcal{B})A(\mathcal{B})$ and $H(\mathcal{B}) = D(\mathcal{B})B(\mathcal{B})$. If there are any common factors in $G(\mathcal{B})$ and $H(\mathcal{B})$, they can be removed to obtain the standard representation. On the basis of this discussion, we might say that an autoregressive moving average of an autoregressive moving average is an autoregressive moving average.

2.8. VECTOR PROCESSES

From an operational standpoint the generalizations of moving average and autoregressive representations to the vector case are obtained by substituting matrix and vector expressions for the scalar expressions of the preceding sections.

Consider a first order moving average process of dimension two:

$$\begin{aligned} X_{1t} &= b_{11}e_{1,t-1} + b_{12}e_{2,t-1} + e_{1t} , \\ X_{2t} &= b_{21}e_{1,t-1} + b_{22}e_{2,t-1} + e_{2t} . \end{aligned} \tag{2.8.1}$$

The system may be written in matrix notation as

$$\mathbf{X}_t = \mathbf{B}\mathbf{e}_{t-1} + \mathbf{e}_t ,$$

where $\{\mathbf{e}_t\}$ is a sequence of uncorrelated $(\mathbf{0}, \mathbf{\Sigma})$ vector random variables,

$$\mathbf{\Sigma} = E\{\mathbf{e}_t\mathbf{e}_t'\} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} ,$$

and we assume $\mathbf{B}$ contains at least one nonzero element. Then the autocovariance function of $\mathbf{X}_t$ is

$$\mathbf{\Gamma}(h) = E\{\mathbf{X}_t\mathbf{X}_{t+h}'\} = E\{(\mathbf{B}\mathbf{e}_{t-1} + \mathbf{e}_t)(\mathbf{e}_{t+h-1}'\mathbf{B}' + \mathbf{e}_{t+h}')\} ,$$

and it follows that

$$\Gamma(h) = \begin{cases} \mathbf{B}\Sigma\mathbf{B}' + \Sigma, & h = 0, \\ \Sigma\mathbf{B}', & h = 1, \\ \mathbf{B}\Sigma, & h = -1, \\ \mathbf{0} & \text{otherwise}. \end{cases}$$

Note that $\Gamma(h) = \Gamma'(-h)$, as we would expect from Lemma 1.7.1.

The qth order moving average time series of dimension k is defined by

$$\mathbf{X}_t = \sum_{j=0}^{q} \mathbf{B}_j \mathbf{e}_{t-j},$$

where the $\mathbf{B}_j$ are $k \times k$ matrices with $\mathbf{B}_0 = \mathbf{I}$ and at least one element of $\mathbf{B}_q$ not zero, and the $\mathbf{e}_t$ are uncorrelated $(\mathbf{0}, \Sigma)$ random variables. Then

$$\Gamma(h) = \begin{cases} \sum_{j=0}^{q-h} \mathbf{B}_j \Sigma \mathbf{B}'_{j+h}, & 0 \leq h \leq q, \\ \sum_{j=0}^{q+h} \mathbf{B}_{j-h} \Sigma \mathbf{B}'_j, & -q \leq h \leq 0, \\ \mathbf{0}, & |h| > q. \end{cases} \tag{2.8.2}$$

As with the scalar process, the autocovariance matrix is zero for $|h| > q$.

In investigating the properties of infinite moving average and autoregressive processes in the scalar case, we made extensive use of absolutely summable sequences. We now extend this concept to matrix sequences.

Definition 2.8.1. Let $\{\mathbf{G}_j\}_{j=1}^{\infty}$ be a sequence of $k \times r$ matrices where the elements of $\mathbf{G}_j$ are $g_{im}(j)$, $i = 1, 2, \ldots, k$, $m = 1, 2, \ldots, r$. If each of the kr sequences $\{g_{im}(j)\}_{j=1}^{\infty}$ is absolutely summable, we say that the sequence $\{\mathbf{G}_j\}_{j=1}^{\infty}$ is absolutely summable.

The infinite moving average

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{B}_j \mathbf{e}_{t-j},$$

where the $\mathbf{e}_t$ are uncorrelated $(\mathbf{0}, \Sigma)$ random variables, will be well defined as an almost sure limit if $\{\mathbf{B}_j\}$ is absolutely summable. Thus, for example, if $\mathbf{X}_t$ is a vector stationary time series of dimension k with zero mean and absolutely summable covariance function and $\{\mathbf{W}_j\}_{j=-\infty}^{\infty}$ is an absolutely summable sequence of $k \times k$ matrices, then

$$\mathbf{Y}_t = \sum_{j=-\infty}^{\infty} \mathbf{W}_j \mathbf{X}_{t-j}$$

is well defined as an almost sure limit and

$$\mathbf{\Gamma}_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{W}_j \mathbf{\Gamma}_X(h-s+j)\mathbf{W}_s' .$$

The vector autoregressive process of order p and dimension k is defined by

$$\sum_{j=0}^{p} \mathbf{A}_j \mathbf{X}_{t-j} = \mathbf{e}_t , \tag{2.8.3}$$

where the $\mathbf{e}_t$ are uncorrelated k-dimensional $(\mathbf{0}, \mathbf{\Sigma})$ random variables and the $\mathbf{A}_j$ are $k \times k$ matrices with $\mathbf{A}_0 = \mathbf{I}$ and $\mathbf{A}_p \neq \mathbf{0}$.

Any autoregressive process can be expressed as a first order vector process. Let $\mathbf{Y}_t = (\mathbf{X}_t', \mathbf{X}_{t-1}', \ldots, \mathbf{X}_{t-p+1}')'$ and $\boldsymbol{\epsilon}_t = (\mathbf{e}_t', \mathbf{0}, \ldots, \mathbf{0})'$, where the $\mathbf{X}_t$ satisfy (2.8.3). Then we can write

$$\mathbf{Y}_t + \mathbf{A}\mathbf{Y}_{t-1} = \boldsymbol{\epsilon}_t ,$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I} & \mathbf{0} \end{pmatrix} .$$

This representation for the scalar autoregressive process was used in the proof of Theorem 2.6.1.

The process (2.8.3) can be expressed as an infinite one-sided moving average of the $\mathbf{e}_t$ when the roots of the characteristic equation are less than one in absolute value.

Theorem 2.8.1. Let the stationary time series $\{\mathbf{X}_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ satisfy (2.8.3), and let the roots of the determinantal equation

$$|\mathbf{A}_0 m^p + \mathbf{A}_1 m^{p-1} + \cdots + \mathbf{A}_p| = 0 \tag{2.8.4}$$

be less than one in absolute value. Then $\mathbf{X}_t$ can be expressed as an infinite moving average of the $\mathbf{e}_t$, say

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{t-j} ,$$

where $\mathbf{K}_0 = \mathbf{I}$, $\mathbf{K}_1 = -\mathbf{A}_1$,

$$\begin{aligned} \mathbf{K}_2 &= -\mathbf{A}_1\mathbf{K}_1 - \mathbf{A}_2, \\ &\vdots \\ \mathbf{K}_s &= -\sum_{j=1}^{p} \mathbf{A}_j\mathbf{K}_{s-j}, \qquad s = p, p+1, \ldots. \end{aligned}$$

Proof. Essentially the same argument as used at the beginning of Section 2.5 may be used to demonstrate that $\Sigma_{j=0}^{\infty} \mathbf{K}_j\mathbf{e}_{t-j}$ is the solution of the difference equation. Since the roots of the determinantal equation are less than one in absolute value, the sequence of matrices $\{\mathbf{K}_j\}$ is absolutely summable and the random variables are well defined as almost sure limits. ▲

Given Theorem 2.8.1, we can construct the multivariate Yule–Walker equations for the autocovariances of the process. Multiplying (2.8.3) on the right by $\mathbf{X}'_{t-s}$, $s \geq 0$ and taking expectations, we obtain

$$\begin{aligned} \sum_{j=0}^{p} \mathbf{A}_j\boldsymbol{\Gamma}(j) &= \boldsymbol{\Sigma}, \qquad s = 0, \\ \sum_{j=0}^{p} \mathbf{A}_j\boldsymbol{\Gamma}(j-s) &= \mathbf{0}, \qquad s = 1, 2, \ldots, \end{aligned} \tag{2.8.5}$$

where we have used

$$E\{\mathbf{e}_t\mathbf{X}'_{t-s}\} = \begin{cases} \boldsymbol{\Sigma}, & s = 0, \\ \mathbf{0} & \text{otherwise}. \end{cases}$$

These equations are of the same form as those given in equation (2.5.6) and Corollary 2.6.1.1, but they contain a larger number of elements.

Theorem 2.8.2. Let the time series $\mathbf{X}_t$ be defined by

$$\mathbf{X}_t = \sum_{j=0}^{q} \mathbf{B}_j\mathbf{e}_{t-j},$$

where the $\mathbf{e}_t$ are uncorrelated k-dimensional vector random variables with zero mean and covariance matrix $\boldsymbol{\Sigma}$, the $\mathbf{B}_j$ are $k \times k$ matrices with $\mathbf{B}_0 = \mathbf{I}$, $\mathbf{B}_q \neq 0$, and the kq roots of the determinantal equation

$$|\mathbf{B}_0\mathbf{m}^q + \mathbf{B}_1\mathbf{m}^{q-1} + \cdots + \mathbf{B}_q| = 0 \tag{2.8.6}$$

are less than one in absolute value. Then $\mathbf{X}_t$ can be expressed as an infinite

autoregressive process

$$\mathbf{e}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{X}_{t-j},$$

where the $\mathbf{C}_j$ satisfy $\mathbf{C}_0 = \mathbf{I}$, $\mathbf{C}_1 = -\mathbf{B}_1$,

$$\mathbf{C}_2 = -\mathbf{B}_1\mathbf{C}_1 - \mathbf{B}_2,$$

$$\vdots$$

$$\mathbf{C}_s = -\sum_{j=1}^{q} \mathbf{B}_j \mathbf{C}_{s-j}, \qquad s = q, q+1, \ldots.$$

Proof. Reserved for the reader. ▲

It can also be demonstrated that an autoregressive moving average of the form

$$\sum_{j=0}^{p} \mathbf{A}_j \mathbf{X}_{t-j} = \sum_{j=0}^{q} \mathbf{B}_j \mathbf{e}_{t-j},$$

where the roots of (2.8.4) and of (2.8.6) are less than one in absolute value, can be given either an infinite autoregressive or an infinite moving average representation with matrix weights of the same form as the scalar weights of Theorems 2.7.1 and 2.7.2.

2.9. PREDICTION

One of the important problems in time series analysis is the following: Given n observations on a realization, predict the $(n+s)$th observation in the realization, where s is a positive integer. The prediction is sometimes called the *forecast* of the $(n+s)$th observation. Because of the functional nature of a realization, prediction is also called *extrapolation.*

In some areas of statistics the term *predictor* is applied to a function of observations used to approximate an unknown *random* quantity, and the term *estimator* is applied to a function of observations used to approximate an unknown *fixed* quantity. Some authors use the term estimator to describe the function in both situations. In time series, the term *filter* is often used for the function of observations used to approximate an unknown value of the random process at time t. If only values of the process prior to t are used in the approximating function for the time series, the function is called a predictor. We adopt this use and apply the term predictor to functions used to approximate a future term of the realization. We will use estimator as a generic term for a function of observations used to approximate either a fixed or a random quantity.

In order to discuss alternative predictors, it is necessary to have a criterion by which the performance of a predictor is measured. The usual criterion, and the one that we adopt, is the mean square error of the predictor. For example, if $\hat{Y}_{n+s}(Y_1, \ldots, Y_n)$ is the predictor of Y_{n+s} based on the n observations Y_1, $Y_2, \ldots, Y_n$, then the mean square error (MSE) of the predictor is

$$\text{MSE}\{\hat{Y}_{n+s}(Y_1, \ldots, Y_n)\} = E\{[Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n)]^2\}. \tag{2.9.1}$$

Generally, the problem of determining optimal predictors requires that the class of predictors be restricted. We shall investigate *linear* predictors.

The best linear predictor for a stationary time series with known covariance function and known mean is easily obtained.

Theorem 2.9.1. Given n observations on a realization of a time series with zero mean and known covariance matrix, the minimum mean square error linear predictor of the $(n + s)$th, $s = 1, 2, \ldots$, observation is

$$\hat{Y}_{n+s}(Y_1, \ldots, Y_n) = \mathbf{y}'\mathbf{b}_{ns}, \tag{2.9.2}$$

where $\mathbf{y}' = (Y_1, Y_2, \ldots, Y_n)$, $\mathbf{V}_{nn} = E\{\mathbf{y}\mathbf{y}'\}$,

$$\mathbf{b}_{ns} = \mathbf{V}_{nn}^{+}\mathbf{V}_{ns},$$
$$\mathbf{V}_{ns}' = E\{Y_{n+s}\mathbf{y}'\},$$

and $\mathbf{V}_{nn}^{+}$ is a generalized inverse of $\mathbf{V}_{nn}$. [See Rao (1973) for a discussion of generalized inverses.] The mean of the prediction error is zero, and the variance of the prediction error is

$$\tau_{ns}^2 = V\{Y_{n+s}\} - \mathbf{b}_{ns}'\mathbf{V}_{ns}. \tag{2.9.3}$$

Proof. To find the minimum mean square error linear predictor, we minimize

$$E\{[Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n)]^2\} = E\{[Y_{n+s} - \mathbf{y}'\mathbf{b}]^2\}$$
$$= (1, -\mathbf{b}')\begin{pmatrix} V\{Y_{n+s}\} & \mathbf{V}_{ns}' \\ \mathbf{V}_{ns} & \mathbf{V}_{nn} \end{pmatrix}\begin{pmatrix} 1 \\ -\mathbf{b} \end{pmatrix}$$

with respect to the vector of coefficients $\mathbf{b}$. Differentiating the quadratic form with respect to $\mathbf{b}$ and equating the derivative to zero, we find that a $\mathbf{b}$ that minimizes the expectation is $\mathbf{b}_{ns} = \mathbf{V}_{nn}^{+}\mathbf{V}_{ns}$. While $\mathbf{b}_{ns}$ is not necessarily unique, the predictor $\mathbf{y}'\mathbf{b}_{ns}$ is unique. See Exercise 2.42. The mean of the prediction error is zero because the mean of the process is zero and

$$E\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n)\} = E\{Y_{n+s} - \mathbf{y}'\mathbf{b}_{ns}\} = 0.$$

The variance of the prediction error follows immediately from least squares regression theory. ▲

To understand why the generalized inverse was required in Theorem 2.9.1, consider the time series

$$X_t = e_1 \cos \pi t, \tag{2.9.4}$$

where e_1 is normally distributed with zero mean and variance σ^2. This time series is perfectly periodic with period two and covariance function $\gamma_X(h) = \sigma^2 \cos \pi h$. The matrix $\mathbf{V}_{nn}$ for a sample of three observations is

$$\mathbf{V}_{33} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \sigma^2 \tag{2.9.5}$$

which is easily seen to be singular. Thus there are many possible linear combinations of past values that can be used to predict future values. For example, one can predict X_t by $-X_{t-1}$ or by X_{t-2} or by $-\frac{1}{2}X_{t-1} + \frac{1}{2}X_{t-2}$.

The time series of equation (2.9.4) is a member of the class of time series defined by

$$X_t = e_1 \cos \lambda t + e_2 \sin \lambda t, \tag{2.9.6}$$

where e_i, $i = 1, 2$, are independent $(0, \sigma^2)$ random variables and $\lambda \in (0, \pi]$. Given λ and two observations, it is possible to predict perfectly all future observations in a given realization. That is, for $n \geq 2$,

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{bmatrix} \sum_{t=1}^{n} \cos^2 \lambda t & \sum_{t=1}^{n} \cos \lambda t \sin \lambda t \\ \sum_{t=1}^{n} \cos \lambda t \sin \lambda t & \sum_{t=1}^{n} \sin^2 \lambda t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} X_t \cos \lambda t \\ \sum_{t=1}^{n} X_t \sin \lambda t \end{bmatrix},$$

and all future observations can be predicted by substituting these e's into the defining equation (2.9.6).

This example illustrates that in predicting future observations of a realization from past observations of that realization, the element $\omega \in \Omega$ is fixed. In the current example, $\omega \in \Omega$ is a two-dimensional vector and can be perfectly identified for a particular realization by a few observations on that realization. In the more common situation, $\omega \in \Omega$ is infinite-dimensional, and we can never determine it perfectly from a finite number of observations on the realization.

Time series for which it is possible to use the observations $Y_t, Y_{t-1}, \ldots$ of a realization to predict future observations with zero mean square error are called *deterministic*. They are also sometimes called *singular*, a term that is particularly meaningful when one looks at the covariance matrix (2.9.5). Time series that have a nonzero lower bound on the prediction error are called *regular* or *nondeterministic*.

Definition 2.9.1. A time series is *nonsingular* (*regular, nondeterministic*) if the sequence of mean square errors of one-period prediction, τ_{n1}^2, is bounded away from zero. A time series is *singular* (*deterministic*) if

$$\lim_{n\to\infty} \tau_{n1}^2 = 0 .$$

The vector of coefficients in Theorem 2.9.1 for one-period predictions can be computed recursively for nondeterministic time series. The recursive procedure is known as the Durbin–Levinson algorithm. See Durbin (1960) and Morettin (1984).

Theorem 2.9.2. Let Y_t be a zero mean, covariance stationary, nondeterministic time series defined on the integers. Then the coefficients defining the predictor of Y_{n+1} given in (2.9.2) of Theorem 2.9.1 satisfy

$$\begin{pmatrix} b_{n,1,2} \\ b_{n,1,3} \\ \vdots \\ b_{n,1,n} \end{pmatrix} = \begin{pmatrix} b_{n-1,1,1} \\ b_{n-1,1,2} \\ \vdots \\ b_{n-1,1,n-1} \end{pmatrix} - b_{n,1,1} \begin{pmatrix} b_{n-1,1,n-1} \\ b_{n-1,1,n-2} \\ \vdots \\ b_{n-1,1,1} \end{pmatrix}, \tag{2.9.7}$$

where $b_{0,1,1} = 0$, $\tau_{11}^2 = \gamma(0)$, $\gamma(h) = \mathrm{Cov}\{Y_t, Y_{t+h}\}$,

$$b_{n,1,1} = \left[\gamma(n) - \sum_{j=1}^{n-1} b_{n-1,1,j}\gamma(n-j) \right] \tau_{n-1,1}^{-2} , \tag{2.9.8}$$

and

$$\tau_{n1}^2 = \tau_{n-1,1}^2 (1 - b_{n,1,1}^2) , \tag{2.9.9}$$

for $n = 1, 2, \ldots$.

Proof. It follows from the symmetry of the covariance function of a stationary process that the coefficients of the predictor of Y_n based upon $(Y_1, \ldots, Y_{n-1})$ are the coefficients of the predictor of Y_1 based upon $(Y_n, \ldots, Y_2)$. That is,

$$\hat{Y}_n(Y_1, \ldots, Y_{n-1}) = \sum_{j=1}^{n-1} b_{n-1,1,j} Y_j$$

and

$$\hat{Y}_1(Y_n, \ldots, Y_2) = \sum_{j=1}^{n-1} b_{n-1,1,j} Y_{n+1-j} ,$$

where $\hat{Y}_1(Y_n, \ldots, Y_2)$ is the predictor of Y_1 based on $(Y_2, \ldots, Y_n)$ and

$$\mathbf{b}_{n-1,1} = (b_{n-1,1,1}, b_{n-1,1,2}, \ldots, b_{n-1,1,n-1})'$$

is the vector of (2.9.2) for $s = 1$. The coefficient vector $\mathbf{b}_{n-1,1}$ is unique because the covariance matrix of a nondeterministic time series is positive definite.

The best predictor of Y_{n+1} based on $(Y_1, \ldots, Y_n)$ can be constructed from the regression of Y_{n+1} on the n-dimensional vector $[Y_1 - \hat{Y}_1(Y_n, \ldots, Y_2), Y_2, \ldots, Y_n]$. By stationarity, the coefficients for the regression of Y_{n+1} on $(Y_2, \ldots, Y_n)$ are the same as the coefficients for the regression of Y_n on $(Y_1, \ldots, Y_{n-1})$. Because $\hat{Y}_1(Y_n, \ldots, Y_2)$ is the best linear predictor, $Y_1 - \hat{Y}_1(Y_n, \ldots, Y_2)$ is uncorrelated with $(Y_2, \ldots, Y_n)$, and the regression coefficient for $Y_1 - \hat{Y}_1(Y_n, \ldots, Y_2)$ is

$$b_{n,1,1} = \tau_{n-1,1}^{-2}[\gamma(n) - \sum_{j=1}^{n-1} b_{n-1,1,j}\gamma(n-j)] .$$

Therefore,

$$\begin{aligned}\hat{Y}_{n+1}(Y_1, \ldots, Y_n) &= \sum_{j=1}^{n-1} b_{n-1,1,j}Y_{1+j} + b_{n,1,1}\left(Y_1 - \sum_{j=1}^{n-1} b_{n-1,1,j}Y_{n+1-j}\right) \\ &= b_{n,1,1}Y_1 + \sum_{j=1}^{n-1} (b_{n-1,1,j} - b_{n,1,1}b_{n-1,1,n-j})Y_{1+j} ,\end{aligned}$$

and we have the result (2.9.7). The error mean square for the prediction of Y_{n+1} based on $Y_1, Y_2, \ldots, Y_n$ is equal to the prediction variance for the prediction of Y_{n+1} based on $Y_2, Y_3, \ldots, Y_n$ reduced by the sum of squares due to Y_1 after adjusting for $Y_2, Y_3, \ldots, Y_n$. The variance of the prediction error for $\hat{Y}_{n+1}(Y_2, \ldots, Y_n)$ is $\tau_{n-1,1}^2$, and $b_{n,1,1}^2$ is the squared correlation between $Y_{n+1} - \hat{Y}_{n+1}(Y_2, \ldots, Y_n)$ and $Y_1 - \hat{Y}_1(Y_n, \ldots, Y_2)$, giving the result (2.9.9). ▲

Theorem 2.9.1 gives a quite general solution to the problem of linear prediction, but the computation of $n \times n$ generalized inverses can become inconvenient if one has a large number of predictions to make. Theorem 2.9.2 presents a recursive method of computing the coefficients of the lagged Y's required for prediction. However, in general, these computations also become sizable as n increases. Fortunately, the forecast equations for autoregressive processes are very simple, and those for invertible moving average processes are fairly simple.

We begin with the first order autoregressive process, assuming the parameter to be known. Let the time series Y_t be defined by

$$Y_t = \rho Y_{t-1} + e_t , \qquad |\rho| < 1 ,$$

where the e_t are independent $(0, \sigma^2)$ random variables. We wish to construct a predictor of $Y_{n+1} = \rho Y_n + e_{n+1}$. Clearly, the "best predictor" of ρY_n is ρY_n.

Therefore, we need only predict e_{n+1}. But, by assumption, e independent of $Y_n, Y_{n-1}, \ldots$. Hence, the best predictor of e_{n+1} is zero, si edicting e_{n+1} is equivalent to predicting a random selection from a distribution with zero mean and finite variance. Therefore, we choose as a predictor of Y_{n+1} at time n

$$\hat{Y}_{n+1}(Y_1, \ldots, Y_n) = \rho Y_n . \tag{2.9.10}$$

Notice that we constructed this predictor by finding the expected value of Y_{n+1} given $Y_n, Y_{n-1}, \ldots, Y_1$. Because of the nature of the time series, knowledge of $Y_{n-1}, Y_{n-2}, \ldots, Y_1$ added no information beyond that contained in knowledge of Y_n. It is clear that the mean square error of this predictor is

$$E\{[Y_{n+1} - \hat{Y}_{n+1}(Y_1, \ldots, Y_n)]^2\} = E\{e_{n+1}^2\} = \sigma^2 .$$

To predict more than one period into the future, we recall the representation

$$Y_{n+s} = \rho^s Y_n + \sum_{j=1}^{s} \rho^{s-j} e_{n+j} , \qquad s = 1, 2, \ldots . \tag{2.9.11}$$

It follows that, for independent e_t, the conditional expectation of Y_{n+s} given $Y_1, Y_2, \ldots, Y_n$ is

$$E\{Y_{n+s} \mid Y_1, \ldots, Y_n\} = \rho^s Y_n , \tag{2.9.12}$$

and this is the best predictor for Y_{n+s}.

These ideas generalize immediately to higher order processes. Let the pth order autoregressive process be defined by

$$Y_t = \sum_{j=1}^{p} \theta_j Y_{t-j} + e_t ,$$

where the roots of

$$m^p - \sum_{j=1}^{p} \theta_j m^{p-j} = 0$$

are less than one in absolute value and the e_t are uncorrelated $(0, \sigma^2)$ random variables. On the basis of the arguments for the first order process we choose as the one-period-ahead linear predictor for the stationary pth order autoregressive process

$$\hat{Y}_{n+1}(Y_1, \ldots, Y_n) = \sum_{j=1}^{p} \theta_j Y_{n+1-j} . \tag{2.9.13}$$

It is clear that this predictor minimizes (2.9.1), since the prediction error is e_{n+1} and this is uncorrelated with Y_n and all previous Y_t's. If the e_t are independent random variables, the predictor is the expected value of Y_{n+1} conditional on $Y_n, Y_{n-1}, \ldots, Y_1$. If the e_t are only uncorrelated, we cannot make this statement. (See

Exercise 2.16). To obtain the best two-period linear prediction, we note that

$$\begin{aligned} Y_{n+2} &= \theta_1\left[\sum_{j=1}^{p}\theta_j Y_{n-j+1} + e_{n+1}\right] + \sum_{j=2}^{p}\theta_j Y_{n-j+2} + e_{n+2} \\ &= \theta_1[\hat{Y}_{n+1}(Y_1,\ldots,Y_n) + e_{n+1}] + \sum_{j=2}^{p}\theta_j Y_{n-j+2} + e_{n+2} \qquad (2.9.14) \\ &= \sum_{j=1}^{p}(\theta_1\theta_j + \theta_{j+1})Y_{n-j+1} + \theta_1 e_{n+1} + e_{n+2}\,, \end{aligned}$$

where it is understood that $\theta_{p+i} = 0$ for $i = 1, 2, \ldots$ and, to simplify the subscripting, we have taken $p \geq 2$. Since e_{n+2} and e_{n+1} are uncorrelated with $(Y_1, Y_2, \ldots, Y_n)$, the two-period predictor can be constructed as

$$\begin{aligned} \hat{Y}_{n+2}(Y_1,\ldots,Y_n) &= \theta_1\hat{Y}_{n+1}(Y_1,\ldots,Y_n) + \sum_{j=2}^{p}\theta_j Y_{n-j+2} \\ &= \sum_{j=1}^{p}(\theta_1\theta_j + \theta_{j+1})Y_{n-j+1}\,. \end{aligned}$$

In general, we can build predictors for s periods ahead by substituting the predictors for earlier periods into (2.9.13). Hence, for $p \geq 2$,

$$\begin{aligned} \hat{Y}_{n+s}(Y_1,\ldots,Y_n) &= \sum_{j=1}^{s-1}\theta_j\hat{Y}_{n-j+s}(Y_1,\ldots,Y_n) + \sum_{j=s}^{p}\theta_j Y_{n-j+s}\,, \qquad s = 2, 3, \ldots, p\,, \\ &= \sum_{j=1}^{p}\theta_j\hat{Y}_{n-j+s}(Y_1,\ldots,Y_n)\,, \qquad s = p+1,\ p+2, \ldots\,. \end{aligned} \qquad (2.9.15)$$

We now turn to prediction for the qth order moving average time series

$$Y_t = \sum_{j=1}^{q}\beta_j e_{t-j} + e_t\,. \qquad (2.9.16)$$

We assume β_j, $j = 1, 2, \ldots, q$, are known. The pth order autoregressive process expresses Y_t as a linear combination of p previous Y's and e_t, while the qth order moving average process expresses Y_t as a linear combination of q previous e's and e_t. In both cases, e_t is uncorrelated with previous Y's and uncorrelated with previous e's. Suppose we knew the e_t for $t = n,\ n-1, \ldots, n-q+1$ for the process (2.9.16). Then, by arguments analogous to those used for the autoregressive process, the best linear predictor for s periods ahead would be

$$\hat{Y}_{n+s}(e_1,\ldots,e_n) = \begin{cases} \displaystyle\sum_{j=s}^{q}\beta_j e_{n+s-j}\,, & 1 \leq s \leq q\,, \\ 0\,, & s > q\,. \end{cases} \qquad (2.9.17)$$

If the e_t's are independent,

$$\hat{Y}_{n+s}(e_1, \ldots, e_n) = E\{Y_{n+s} \mid e_1, e_2, \ldots, e_n\}.$$

Because we do not know the e_t, we must develop predictors based upon estimators of the e_t, where the estimators of the e_t are functions of the observed Y_t. We first show that the predictions of Theorem 2.9.1 can be expressed in terms of the one-period prediction errors of previous predictions when Y_t is nondeterministic. The recursive computation is based on the Gram–Schmidt orthogonalization. Brockwell and Davis (1991, p. 171) call the procedure the *innovation algorithm*, where the Z_j are the innovations. We give the proof for stationary time series, but the result holds for more general covariance functions.

Theorem 2.9.3. Let Y_t be a zero mean, stationary, nondeterministic time series defined on the integers. Let

$$Z_t = Y_t - \sum_{j=1}^{t-1} c_{tj} Z_j, \qquad t = 2, 3, \ldots, \tag{2.9.18}$$

where $Z_1 = c_{11} Y_1$, $c_{11} = 1$, $c_{21} = \kappa_1^{-2} \gamma_Y(1)$, $\kappa_1^2 = \gamma_Y(0)$,

$$\begin{aligned} c_{t1} &= \kappa_1^{-2} \gamma_Y(t-1), \\ c_{ti} &= \kappa_i^{-2} \left[\gamma_Y(t-i) - \sum_{j=1}^{i-1} c_{ij} c_{ij} \kappa_j^2 \right], \qquad i = 2, 3, \ldots, t-1, \end{aligned} \tag{2.9.19}$$

and

$$\kappa_t^2 = \gamma_Y(0) - \sum_{j=1}^{t-1} c_{tj}^2 k_j^2 \qquad t = 2, 3, \ldots. \tag{2.9.20}$$

Then the best predictor of Y_{n+s} given $(Y_1, \ldots, Y_n)$ is

$$\hat{Y}_{n+s}(Y_1, \ldots, Y_n) = \sum_{j=1}^{n} c_{n+s,j} Z_j, \tag{2.9.21}$$

and the variance of the prediction error is

$$\tau_{ns}^2 = \begin{cases} \kappa_{n+1}^2, & s = 1, \\ \kappa_{n+s}^2 + \sum_{j=1}^{s-1} c_{n+s,n+j}^2 \kappa_{n+j}^2, & s > 1. \end{cases} \tag{2.9.22}$$

Proof. Because Z_t is the residual from the regression of Y_t on $(Z_{t-1},$

$Z_{i-2}, \ldots, Z_1$), Z_i is uncorrelated with $(Z_{i-1}, Z_{i-2}, \ldots, Z_1)$, and hence the Z's are mutually uncorrelated. The covariance between Y_t and Z_i for $t > i$ is

$$\begin{aligned}
\mathrm{Cov}(Y_t, Z_i) &= \mathrm{Cov}\left(Y_t, Y_i - \sum_{j=1}^{i-1} c_{ij} Z_j\right) \\
&= \gamma_Y(t-i) - \mathrm{Cov}\left(Y_t, \sum_{j=1}^{i-1} c_{ij} Z_j\right) \\
&= \gamma_Y(t-i) - \mathrm{Cov}\left(Z_t + \sum_{r=1}^{t-1} c_{tr} Z_r, \sum_{j=1}^{i-1} c_{ij} Z_j\right) \\
&= \gamma_Y(t-i) - \sum_{j=1}^{i-1} c_{ij} c_{ij} \kappa_j^2 ,
\end{aligned}$$

where we have used the zero correlation property of the Z_i. Dividing $\mathrm{Cov}(Y_t, Z_i)$ by the variance of Z_i, we obtain the expression for c_{ti}. The variance of the prediction error is the regression error mean square, and by the zero correlation property we have (2.9.20).

Because of the zero correlation property, the coefficients of $(Z_n, \ldots, Z_1)$ in the regression of Y_{n+s} on $(Z_n, \ldots, Z_1)$ are the same as the coefficients of $(Z_n, \ldots, Z_1)$ in the regression of Y_{n+s} on $(Z_{n+s-1}, \ldots, Z_1)$. The result (2.9.21) follows. The result (2.9.22) is an immediate consequence of (2.9.21) and (2.9.20). ▲

The recursive equations of Theorem 2.9.3 are particularly effective for predictions of finite moving average processes. Because the autocovariances of the qth order moving average are zero for $h > q$, the c_{ti} of Theorem 2.9.3 are zero for $i < t - q$.

Corollary 2.9.3. Let Y_t be the qth order moving average

$$Y_t = e_t + \sum_{j=1}^{q} \beta_j e_{t-j} ,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables. Assume that the Z_j, c_{ji}, and κ_j^2 of Theorem 2.9.3 have been constructed for $j = 1, 2, \ldots, t-1$, where $t - 1 > q$. Then the c_{ti} of Theorem 2.9.3 are

$$\begin{aligned}
c_{ti} &= 0 , \qquad i < t - q , \\
c_{t,t-q} &= \kappa_{t-q}^{-2} \gamma_Y(q) ,
\end{aligned}$$

and

$$c_{ti} = \kappa_i^{-2} \left[\gamma_Y(t-i) - \sum_{j=t-q}^{i-1} c_{ij} c_{ij} \kappa_j^2 \right]$$

for $i = t - q + 1, \ldots, t - 1$.

Proof. Omitted. ▲

If the roots of the moving average process are less than one in absolute value, then the Z_t of Theorem 2.9.3 converges to e_t, κ_t^2 converges to σ^2, and $c_{t,t-i}$ converges to β_i as t increases. For invertible moving averages, the coefficients are often approximated by the β_i. Let $(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_q)$ be initial estimates of $(e_1, e_2, \ldots, e_q)$, and let

$$\tilde{e}_t = Y_t - \sum_{j=1}^{q} \beta_j \tilde{e}_{t-j}, \qquad t = q+1, q+2, \ldots. \tag{2.9.23}$$

Then $\tilde{e}_n$ for $n > q$ is

$$\tilde{e}_n = \sum_{j=n-q}^{n-1} c_j \tilde{e}_{n-j} + \sum_{j=0}^{n-1-q} c_j Y_{n-j},$$

where the c_j are defined in Theorem 2.6.2. For an invertible moving average, the $|c_j|$ converge to zero as j increases, and $\tilde{e}_n$ converges to e_n in mean square as n increases. The initial values of (2.9.23) can be computed using Theorem 2.9.3 or by setting $\tilde{e}_{-q+1}, \tilde{e}_{-q+2}, \ldots, \tilde{e}_0$ equal to zero and using (2.9.23) for $t = 1, 2, \ldots, q$. Therefore, an easily computed s-period predictor for the qth order invertible moving average, appropriate for large n, is

$$\tilde{Y}_{n+s}(Y_1, \ldots, Y_n) = \begin{cases} \sum_{j=s}^{q} \beta_j \tilde{e}_{n+s-j}, & 1 \leq s \leq q, \\ 0, & s > q, \end{cases} \tag{2.9.24}$$

where $\tilde{e}_j$ is defined in (2.9.23). The prediction error for $\tilde{Y}_{n+1}(Y_1, \ldots, Y_n)$ goes to σ^2 as n increases.

The procedures for autoregressive and moving average processes are easily extended to the autoregressive moving average time series defined by

$$Y_t = \sum_{j=1}^{p} \theta_j Y_{t-j} + \sum_{i=1}^{p} \beta_i e_{t-i} + e_t,$$

where the roots of

$$m^p - \sum_{j=1}^{p} \theta_j m^{p-j} = 0$$

and of

$$m^q + \sum_{i=1}^{p} \beta_i m^{q-i} = 0$$

are less than one in absolute value and the e_t are uncorrelated $(0, \sigma^2)$ random variables. For such processes the predictor for large n is given by

$$\hat{Y}_{n+s}(Y_1, \ldots, Y_n) = \begin{cases} \sum_{j=1}^{p} \theta_j Y_{n-j+1} + \sum_{i=1}^{q} \beta_i \tilde{e}_{n-i+1}, & s = 1, \\ \sum_{j=1}^{s-1} \theta_j \hat{Y}_{n-j+s}(Y_1, \ldots, Y_n) + \sum_{j=s}^{p} \theta_j Y_{n-j+s} & \\ \quad + \sum_{i=s}^{q} \beta_i \tilde{e}_{n-i+s}, & s = 2, 3, \ldots, \end{cases} \tag{2.9.25}$$

where

$$\tilde{e}_t = \begin{cases} 0, & t = p, p-1, \ldots, \\ Y_t - \sum_{j=1}^{p} \theta_j Y_{t-j} - \sum_{i=1}^{q} \beta_i \tilde{e}_{t-i}, & t = p+1, p+2, \ldots, n, \end{cases}$$

and it is understood that $\theta_j = 0$ for $j > p$ and

$$\sum_{i=s}^{q} \beta_i \tilde{e}_{n-i+s} = 0, \qquad s = q+1, q+2, \ldots,$$

$$\sum_{j=s}^{p} \theta_j Y_{n-j+s} = 0, \qquad s = p+1, p+2, \ldots.$$

The stationary autoregressive invertible moving average, which contains the autoregressive and invertible moving average processes as special cases, can be expressed as

$$Y_t = \sum_{j=0}^{\infty} v_j e_{t-j}, \tag{2.9.26}$$

where $v_0 = 1$, $\Sigma_{j=0}^{\infty} |v_j| < \infty$, and the e_t are uncorrelated $(0, \sigma^2)$ random variables. The variances of the forecast errors for predictions of more than one period are readily obtained when the time series is written in this form.

Theorem 2.9.4. Given a stationary autoregressive invertible moving average with the representation (2.9.26) and the predictor

$$\hat{Y}_{n+s}(Y_1, \ldots, Y_n) = \sum_{j=s}^{n+s-1} v_j \tilde{e}_{n-j+s},$$

where $\tilde{e}_t$, $t = 1, 2, \ldots, n$, is defined following (2.9.25), then

$$\lim_{n \to \infty} E\left\{ \left[Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n) - \sum_{j=0}^{s-1} v_j e_{n-j+s} \right]^2 \right\} = 0.$$

Proof. The result follows immediately from Theorem 2.2.1. ▲

Using Theorem 2.9.4, we can construct the covariance matrix of the s predictors for $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+s}$. Let $\mathbf{Y}'_{n.s} = [Y_{n+1}, Y_{n+2}, \ldots, Y_{n+s}]$, and let

$$\hat{\mathbf{Y}}'_{n.s} = [\hat{Y}_{n+1}(Y_1, \ldots, Y_n), \hat{Y}_{n+2}(Y_1, Y_2, \ldots, Y_n), \ldots, \hat{Y}_{n+s}(Y_1, \ldots, Y_n)] .$$

Then

$$\lim_{n\to\infty} E\{(\mathbf{Y}_{n.s} - \hat{\mathbf{Y}}_{n.s})(\mathbf{Y}_{n.s} - \hat{\mathbf{Y}}_{n.s})'\}$$

$$= \begin{pmatrix} 1 & v_1 & \cdots & v_{s-1} \\ v_1 & \sum_{j=0}^{1} v_j^2 & \cdots & \sum_{j=0}^{1} v_j v_{s+j-2} \\ \vdots & \vdots & & \vdots \\ v_{s-1} & \sum_{j=0}^{1} v_j v_{s+j-2} & \cdots & \sum_{j=0}^{s-1} v_j^2 \end{pmatrix} \sigma^2 . \quad (2.9.27)$$

Note that the predictors we have considered are unbiased: that is,

$$E\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n)\} = 0 .$$

Therefore, the mean square error of the predictor is the variance. To use this information to establish confidence limits for the predictor, we require knowledge of the distribution of the e_t. If the time series is normal, confidence intervals can be constructed using normal theory. That is, a $1 - \eta$ level confidence interval is given by

$$\hat{Y}_{n+s}(Y_1, \ldots, Y_n) \pm t_\eta [\mathrm{Var}\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \ldots, Y_n)\}]^{1/2} ,$$

where t_η is the value such that η of the probability of the standard normal distribution is beyond $\pm t_\eta$.

Example 2.9.1. In Example 6.3.1, we demonstrate that the U.S. quarterly seasonally adjusted unemployment rate for the period 1948–1972 behaves much like the time series

$$Y_t - 4.77 = 1.54(Y_{t-1} - 4.77) - 0.67(Y_{t-2} - 4.77) + e_t ,$$

where the e_t are uncorrelated $(0, 0.115)$ random variables. Let us assume that we know that this is the proper representation. The four observations for 1972 are 5.83, 5.77, 5.53, and 5.30. The predictions for 1973 are

$$\begin{aligned} \hat{Y}_{73-1}(Y_{72-1}, \ldots, Y_{72-4}) &= 4.77 + 1.54(0.53) - 0.67(0.76) \\ &= 5.08 , \end{aligned}$$

$$\hat{Y}_{73-2}(Y_{72-1}, \ldots, Y_{72-4}) = 4.77 + 1.54(0.3070) - 0.67(0.53)$$
$$= 4.89,$$
$$\hat{Y}_{73-3}(Y_{72-1}, \ldots, Y_{72-4}) = 4.77 + 1.54(0.1177) - 0.67(0.3070)$$
$$= 4.75,$$
$$\hat{Y}_{73-4}(Y_{72-1}, \ldots, Y_{72-4}) = 4.77 + 1.54(-0.0245) - 0.67(0.1177)$$
$$= 4.65.$$

The covariance matrix of the prediction errors can be obtained by using the representation $Y_t = \sum_{j=0}^{\infty} w_j e_{t-j}$, where the w's are defined in Theorem 2.6.1. We have $a_1 = -1.54$, $a_2 = 0.67$, and

$$w_0 = 1,$$
$$w_1 = 1.54,$$
$$w_2 = (1.54)^2 - 0.67 = 1.702,$$
$$w_3 = 1.54(1.702) - 0.67(1.54) = 1.589.$$

Therefore, the covariance matrix of equation (2.9.27) given by the product

$$0.115\mathbf{H}'\mathbf{H} = \begin{pmatrix} 0.115 & 0.177 & 0.196 & 0.183 \\ 0.177 & 0.388 & 0.478 & 0.477 \\ 0.196 & 0.478 & 0.720 & 0.789 \\ 0.183 & 0.477 & 0.789 & 1.011 \end{pmatrix},$$

where

$$\mathbf{H}' = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 1.54 & 1.00 & 0 & 0 \\ 1.70 & 1.54 & 1.00 & 0 \\ 1.59 & 1.70 & 1.54 & 1.00 \end{pmatrix}.$$

We observe that there is a considerable increase in the variance of the predictor as s increases from 1 to 4. Also, there is a high positive correlation between the predictors. ▲▲

Example 2.9.2. In Table 2.9.1 are seven observations from the time series

$$Y_t = 0.9Y_{t-1} + 0.7e_{t-1} + e_t,$$

where the e_t are normal independent (0, 1) random variables. The first 11 autocorrelations for this time series are, approximately, 1.00, 0.95, 0.85, 0.77, 0.69, 0.62, 0.56, 0.50, 0.45, 0.41, and 0.37. The variance of the time series is about 14.4737.

To construct the linear predictor for Y_8, we obtain the seven weights by solving

Table 2.9.1. Observations and predictions for the time series $Y_t = 0.9Y_{t-1} + 0.7e_{t-1} + e_t$

t	Observation	Prediction
1	−1.58	—
2	−2.86	—
3	−3.67	—
4	−2.33	—
5	0.42	—
6	2.87	—
7	3.43	—
8	—	3.04
9	—	2.73
10	—	2.46

the system of equations

$$\begin{pmatrix} 1.00 & 0.95 & 0.85 & \cdots & 0.56 \\ 0.95 & 1.00 & 0.95 & \cdots & 0.62 \\ 0.85 & 0.95 & 1.00 & \cdots & 0.69 \\ \vdots & \vdots & \vdots & & \vdots \\ 0.56 & 0.62 & 0.69 & \cdots & 1.00 \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{12} \\ b_{13} \\ \vdots \\ b_{17} \end{pmatrix} = \begin{pmatrix} 0.50 \\ 0.56 \\ 0.62 \\ \vdots \\ 0.95 \end{pmatrix}. \tag{2.9.28}$$

This system of equations gives the weights for the one-period-ahead forecast. The vector of weights $\mathbf{b}'_{7,1} = (0.06, -0.18, 0.32, -0.50, 0.75, -1.10, 1.59)$ is applied to $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)$. [In obtaining the solution more digits were used than are displayed in equation (2.9.28).] The predictor for two periods ahead is obtained by replacing the right side of (2.9.28) by $(0.45, 0.50, 0.56, 0.62, 0.77, 0.85)'$, and the weights for the three-period-ahead predictor are obtained by using $(0.41, 0.45, 0.50, 0.56, 0.62, 0.69, 0.77)'$ as the right side. The predictions using these weights are displayed in Table 2.9.1.

The covariance matrix of the prediction errors is

$$E\{(\mathbf{Y}_{7,3} - \hat{\mathbf{Y}}_{7,3})(\mathbf{Y}_{7,3} - \hat{\mathbf{Y}}_{7,3})'\} = \begin{pmatrix} 1.005 & 1.604 & 1.444 \\ 1.604 & 3.563 & 3.907 \\ 1.444 & 3.907 & 5.637 \end{pmatrix},$$

where $\hat{\mathbf{Y}}_{7,3}$ and $\mathbf{Y}_{7,3}$ are defined following Theorem 2.9.4. The covariance matrix of prediction errors was obtained as $\mathbf{AVA}'$, where $\mathbf{V}$ is the 10×10 covariance matrix for 10 observations and

$$\mathbf{A} = \begin{bmatrix} -0.06 & 0.18 & -0.32 & 0.50 & -0.75 & 1.10 & -1.59 & 1 & 0 & 0 \\ -0.05 & 0.16 & -0.29 & 0.45 & -0.68 & 0.99 & -1.44 & 0 & 1 & 0 \\ -0.05 & 0.14 & -0.26 & 0.41 & -0.61 & 0.89 & -1.29 & 0 & 0 & 1 \end{bmatrix}.$$

If we had an infinite past to use in prediction, the variance of the one-period-ahead prediction error would be one, the variance of e_t. Thus, the loss in prediction efficiency from having only seven observations is about $\frac{1}{2}\%$. Since the time series is normal, we can construct 95% confidence intervals for the predictions. We obtain the intervals (1.08, 5.00), (−0.97, 6.43), and (−2.19, 7.11) for the predictions one, two, and three periods ahead, respectively.

To construct a predictor using the $\tilde{e}$-method, we set $\tilde{e}_1 = 0$ and calculate

$$\tilde{e}_2 = Y_2 - 0.9Y_1 - 0.7\tilde{e}_1 = -2.86 - 0.9(-1.58) = -1.438,$$
$$\tilde{e}_3 = Y_3 - 0.9Y_2 - 0.7\tilde{e}_2 = -3.67 - 0.9(-2.86) - 0.7(-1.438) = -0.089,$$
$$\tilde{e}_4 = -2.33 - 0.9(-3.67) - 0.7(-0.089) = 1.035,$$
$$\tilde{e}_5 = 0.42 - 0.9(-2.33) - 0.7(1.035) = 1.793,$$
$$\tilde{e}_6 = 2.87 - 0.9(0.42) - 0.7(1.793) = 1.237,$$
$$\tilde{e}_7 = 3.43 - 0.9(2.86) - 0.7(1.227) = -0.003.$$

Therefore, the predictor of Y_8 is

$$\hat{Y}_8(Y_1, \ldots, Y_7) = 0.9Y_7 + 0.7\tilde{e}_7 = 3.085,$$

and

$$\hat{Y}_9(Y_1, \ldots, Y_7) = 0.9\hat{Y}_8(Y_1, \ldots, Y_7) = 2.776,$$
$$\hat{Y}_{10}(Y_1, \ldots, Y_7) = 0.9\hat{Y}_9(Y_1, \ldots, Y_7) = 2.498.$$

In this case the prediction using the approximate method of setting $\tilde{e}_1 = 0$ is quite close to the prediction obtained by the method of Theorem 2.9.1, even though the sample size is small. If the coefficient of e_{t-1} had been closer to one, the difference would have been larger.

By Theorem 2.7.1, we may write our autoregressive moving average process as

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j},$$

where

$$\psi_0 = b_0 = 1,$$
$$\psi_1 = b_1 - a_1 = 0.7 + 0.9 = 1.6,$$
$$\psi_j = -a_1\psi_{j-1} = (0.9)\psi_{j-1}, \qquad j = 2, 3, \ldots.$$

Therefore, the large sample covariance matrix of prediction errors for predictors one, two, and three periods ahead is

$$\begin{pmatrix} 1.00 & 1.60 & 1.44 \\ 1.60 & 3.56 & 3.90 \\ 1.44 & 3.90 & 5.63 \end{pmatrix},$$

which is very close to the covariance matrix of the best predictor based on $n = 7$ observations. ▲▲

2.10 THE WOLD DECOMPOSITION

We obtained moving average representations for stationary autoregressive and autoregressive moving average processes in Sections 2.6 and 2.7. In this section, we give a moving average representation, due to Wold (1938), that is appropriate for any stationary process. We first investigate the properties of the sequence of one-period prediction errors of a stationary nondeterministic process. We are particularly interested in the limit as n becomes large.

Theorem 2.10.1. Let $\{Y_t\}$ be a stationary, nondeterministic time series defined on the integers. Then there exists a unique (a.s.) time series $\{e_t\}$ defined as the limit in mean square of the one-period prediction error as n becomes large. Furthermore, $\{e_t\}$ is a sequence of uncorrelated random variables with mean zero and common variance σ^2, where

$$\sigma^2 = \lim_{n\to\infty} \tau_{n1}^2 , \tag{2.10.1}$$

and τ_{n1}^2 is the variance of the one-period prediction error defined in Theorem 2.9.1.

Proof. Without loss of generality, let the mean of Y_t be zero. Define the new time series

$$\begin{aligned} W_{1t} &= Y_{t-1} , \\ W_{2t} &= Y_{t-2} - [\gamma_Y(0)]^{-1}\gamma_Y(1)Y_{t-1} , \\ &\vdots \\ W_{it} &= Y_{t-i} - \hat{Y}_{t-i}(Y_{t-i+1}, \ldots, Y_{t-1}) \\ &= Y_{t-i} - \sum_{j=1}^{i-1} b_{i-1,1,j}Y_{t-j} \qquad i = 2, 3, \ldots , \end{aligned} \tag{2.10.2}$$

where

$$\mathbf{b}_{i-1,1} = (b_{i-1,1,1}, b_{i-1,1,2}, \ldots, b_{i-1,1,i-1})' , \tag{2.10.3}$$

$$E\{W_{it}^2\} = \tau_{i-1,1}^2 , \qquad i = 2, 3, \ldots , \tag{2.10.4}$$

and $\mathbf{b}_{n1}$ and τ_{n1}^2 are defined in Theorem 2.9.1. Because Y_t is regular, $E\{W_{it}^2\}$ is bounded below and the vectors $\mathbf{b}_{i-1,1}$ are unique. Each W_{it} is a stationary time

series, and

$$E\{W_{it}W_{jt}\} = 0\,, \qquad i \neq j\,. \tag{2.10.5}$$

Let

$$e_t(n) = Y_t - \sum_{i=1}^{n} c_i W_{it}\,, \tag{2.10.6}$$

where

$$\begin{aligned}(c_1, c_2, \ldots, c_n)' &= [E\{\mathbf{w}_{nt}\mathbf{w}_{nt}'\}]^{-1}E\{\mathbf{w}_{nt}Y_t\}\,,\\ \mathbf{w}_{nt}' &= (W_{1t}, W_{2t}, \ldots, W_{nt})\,.\end{aligned} \tag{2.10.7}$$

Then

$$E\{e_t^2(n)\} = \gamma_Y(0) - \sum_{j=1}^{n} c_j^2\tau_{j-1,1}^2 = \tau_{n1}^2\,. \tag{2.10.8}$$

By (2.10.8), τ_{n1}^2 is monotone decreasing in n, and by the assumption that Y_t is regular, τ_{n1}^2 is bounded below. Therefore, the limit

$$\lim_{n\to\infty} \tau_{n1}^2 = \sigma^2 \tag{2.10.9}$$

is well defined and

$$\lim_{n\to\infty} \sum_{j=1}^{n} c_j^2\tau_{j-1,1}^2 = \gamma_Y(0) - \sigma^2\,.$$

By the definitions,

$$\begin{aligned}E\{[e_t(n) - e_t(n+m)]^2\} &= \sum_{j=1}^{m} c_{n+j}^2\tau_{n+j-1,1}^2\\ &\leqslant \sum_{j=n+1}^{\infty} c_j^2\tau_{j-1,1}^2\,,\end{aligned}$$

which converges to zero uniformly in m as $n \to \infty$. Therefore, by Exercise 4.4 of Tucker (1967, p. 106) there exists a random variable e_t such that $E\{e_t^2\} < \infty$ and $e_t(n)$ converges to e_t in mean square. For any fixed h,

$$\begin{aligned}E\{|e_te_{t+h} - e_t(n)e_{t+h}(n)|\} &= E\{|[e_t - e_t(n)]e_{t+h} + e_t(n)[e_{t+h} - e_{t+h}(n)]|\}\\ &\leqslant (E\{[e_t - e_t(n)]^2\}E\{e_{t+h}^2\})^{1/2}\\ &\quad + (E\{e_t^2(n)\}E\{[e_{t+h} - e_{t+h}(n)]^2\})^{1/2}\\ &\to 0 \qquad \text{as} \quad n \to \infty\,,\end{aligned}$$

because $E\{e_t^2(n)\} < \gamma_Y(0)$. Therefore, by Exercise 4.14 of Royden (1989, p. 93),

$$E\{e_t e_{t+h}\} = \lim_{n\to\infty} E\{e_t(n)e_{t+h}(n)\}.$$

Similarly,

$$E\{e_t\} = \lim_{n\to\infty} E\{e_t(n)\} = 0.$$

The covariance between $e_t(n)$ and $e_{t+h}(n)$, for $n > h > 0$, is

$$\begin{aligned} E\{e_t(n)e_{t+h}(n)\} &= E\left\{e_{t+h}(n)\left(Y_t - \sum_{i=1}^{n-h} c_i W_{it} - \sum_{i=n-h+1}^{n} c_i W_{it}\right)\right\} \\ &= -E\left\{e_{t+h}(n) \sum_{i=n-h+1}^{n} c_i W_{it}\right\}, \end{aligned}$$

where we have used the fact that $e_{t+h}(n)$ is uncorrelated with Y_{t+h-1}, $Y_{t+h-2}, \ldots, Y_{t+h-n}$. Because

$$\begin{aligned} |E\{e_t(n)e_{t+h}(n)\}|^2 &\leq \tau_{n1}^2 \sum_{i=n-h+1}^{n} c_i^2 \tau_{i-1,1}^2 \\ &\leq \gamma_Y(0) \sum_{i=n-h+1}^{\infty} c_i^2 \tau_{i-1,1}^2 \end{aligned}$$

we have

$$\lim_{n\to\infty} E\{e_t(n)e_{t+h}(n)\} = E\{e_t e_{t+h}\} = 0.$$

Similarly,

$$E(e_t^2) = \lim_{n\to\infty} E[e_t^2(n)] = \sigma^2. \qquad \blacktriangle$$

We now give Wold's theorem for the representation of a stationary time series. The theorem states that it is possible to express every stationary time series as the sum of two time series: an infinite moving average time series and a singular time series.

Theorem 2.10.2 (Wold decomposition). Let Y_t be a covariance stationary time series defined on $T = \{0, \pm 1, \pm 2, \ldots\}$. Then Y_t has the representation

$$Y_t = X_t + Z_t, \tag{2.10.10}$$

where

$$X_t = \sum_{i=0}^{\infty} a_i e_{t-i},$$

$a_0 = 1$, $\{Z_t\}$ is singular, $\{Z_t\}$ and $\{X_t\}$ are uncorrelated, $\{e_t\}$ is the sequence of

uncorrelated $(0, \sigma^2)$ random variables defined in Theorem 2.10.1, and X_t is defined as a limit in mean square.

Proof. If Y_t is singular, then $Y_t = Z_t$ and $X_t = 0$. Suppose Y_t is nonsingular. Let $\{e_t\}$ be the time series defined by Theorem 2.10.1. Let

$$X_t(n) = \sum_{i=0}^{n} a_i e_{t-i},$$

where

$$(a_0, a_1, \ldots, a_n) = [\gamma_e(0)]^{-1}[\gamma_{eY}(0), \gamma_{eY}(1), \ldots, \gamma_{eY}(n)],$$

and

$$\gamma_{eY}(h) = \operatorname{Cov}(e_t, Y_{t+h}) = \operatorname{Cov}(Y_t, e_{t-h}).$$

We have

$$\gamma_Y(0) \geq \gamma_{X_t(n)}(0) = \sum_{i=0}^{n} a_i^2 \gamma_e(0) = \sigma^2 \sum_{i=0}^{n} a_i^2$$

and $\sum_{i=0}^{\infty} a_i^2 < \infty$. It follows that

$$\sum_{i=0}^{n} a_i e_{t-i} \to \sum_{i=0}^{\infty} a_i e_{t-i} = X_t$$

in mean square. Let $Z_t = Y_t - X_t$. Using the arguments of Theorem 2.10.1, it follows that

$$\operatorname{Cov}(X_{t+h}, e_t) = a_h \sigma^2 = \operatorname{Cov}(Y_{t+h}, e_t), \qquad h \geq 0,$$

$$\operatorname{Cov}(Y_t, e_{t+h}) = \lim_{n \to \infty} \operatorname{Cov}\left(Y_t, Y_{t+h} - \sum_{i=1}^{n} c_i W_{i,t+h}\right) = 0, \qquad h > 0,$$

$$\operatorname{Cov}(X_t, e_{t+h}) = \operatorname{Cov}\left(\sum_{i=0}^{\infty} a_i e_{t-i}, e_{t+h}\right) = 0, \qquad h > 0.$$

Therefore, the processes $\{Z_t\}$ and $\{X_t\}$ are uncorrelated. For any set of real numbers $\{\delta_i : i = 0, 1, \ldots, p\}$ with $\delta_0 = 1$,

$$V\left\{\sum_{i=0}^{p} \delta_i Y_{t-i}\right\} = V\left\{\sum_{i=0}^{p} \delta_i Z_{t-i}\right\} + V\left\{\sum_{i=0}^{p} \delta_i X_{t-1}\right\},$$

where we abbreviate $\operatorname{Var}\{X_t\}$ to $V\{X_t\}$. By (2.10.9), given $\epsilon > 0$, there is some p_ϵ

and a set of δ_i defined by (2.10.6) such that

$$V\left\{\sum_{i=0}^{p_\epsilon} \delta_i Y_{t-i}\right\} < \sigma^2 + \epsilon .$$

Also,

$$\sum_{i=0}^{p} \delta_i X_{t-i} = e_t + (\delta_1 + a_1)e_{t-1} + (\delta_2 + \delta_1 a_1 + a_2)e_{t-2} + \cdots$$

and $V\{\sum_{i=0}^{p} \delta_i X_{t-i}\} \geq \sigma^2$. Therefore, given $\epsilon > 0$, there exist p_ϵ and δ_i such that

$$V\left\{\sum_{i=0}^{p_\epsilon} \delta_i Z_{t-i}\right\} < \epsilon ,$$

and we conclude that $\{Z_t\}$ is singular. ▲

Note that the mean can be included in the singular part of the representation (2.10.10). Proofs of the Wold decomposition using Hilbert space arguments are available in Anderson (1971, p. 420) and Brockwell and Davis (1991, p. 187).

2.11. LONG MEMORY PROCESSES

In Section 2.7, we demonstrated that the autocorrelation function of an autoregressive moving average process declines at an exponential rate. By the results of Section 2.10, any nondeterministic stationary process can be represented as an infinite moving average of uncorrelated random variables. The coefficients of the sum decline exponentially if the original process is an autoregressive moving average. Because of this rapid rate of decline, finite autoregressive moving averages are sometimes called *short memory* processes.

Consider a weighted average of uncorrelated $(0, \sigma^2)$ random variables

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{2.11.1}$$

in which the coefficients of e_{t-j} are declining at the rate j^{d-1} for some $d \in (-0.5, 0.5)$. Such a time series is well defined as a limit in mean square by Theorem 2.2.3. For the special case $\psi_j = (j+1)^{d-1}$, we have

$$E\{Y_t^2\} = \sum_{j=0}^{\infty} (j+1)^{2(d-1)}\sigma^2 \doteq [1 - (2d-1)^{-1}(1.5)^{2d-1}]\sigma^2 , \tag{2.11.2}$$

$$\gamma_Y(h) = E\{Y_t Y_{t+h}\} = \sum_{i=1}^{\infty} i^{d-1}(i+h)^{d-1}\sigma^2 , \qquad h > 0 , \tag{2.11.3}$$

where we have used the integral approximation to the sum in (2.11.2). For large h, $\gamma_Y(h)$ is approximately equal to ch^{2d-1}, where c is a constant. See Exercise 2.44. We say that a process X_t is a *long memory* process if $\rho_X(h)$ is approximately proportional to h^{2d-1} for $d \in (-0.5, 0.5)$.

A time series that has received considerable attention is

$$X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}, \tag{2.11.4}$$

where $\psi_0 = 1$,

$$\begin{aligned} \psi_j &= [\Gamma(j+1)\Gamma(d)]^{-1}\Gamma(j+d) \\ &= \prod_{i=1}^{j} i^{-1}(i-1+d), \qquad j = 1, 2, \ldots, \end{aligned} \tag{2.11.5}$$

and $\Gamma(\cdot)$ is the gamma function. Using Stirling's approximation that $\Gamma(s+1) \doteq (2\pi s)^{1/2} e^{-s} s^s$ for large s, we have

$$\psi_j \doteq [\Gamma(d)]^{-1} j^{d-1}$$

for large j. It can be shown that the autocovariance, autocorrelation, and partial autocorrelation of the time series (2.11.4) are

$$\gamma_X(h) = \sigma^2[\Gamma(h-d+1)\Gamma(1-h-d)]^{-1}(-1)^h\Gamma(1-2d), \qquad h = 0, 1, 2, \ldots, \tag{2.11.6}$$

$$\rho_X(h) = [\Gamma(h-d+1)\Gamma(d)]^{-1}\Gamma(h+d)\Gamma(1-d) \tag{2.11.7}$$

$$= \prod_{i=1}^{h} (i-d)^{-1}(i-1+d), \qquad h = 1, 2, \ldots,$$

and

$$\phi_X(h) = (h-d)^{-1}d, \qquad h = 1, 2, \ldots. \tag{2.11.8}$$

Using Stirling's approximation,

$$\rho_X(h) \doteq [\Gamma(d)]^{-1}\Gamma(1-d)h^{2d-1}.$$

Thus, the moving average with weights (2.11.5) is a long memory process.

The time series (2.11.4) can be written

$$X_t = (1-\mathcal{B})^{-d} e_t, \tag{2.11.9}$$

where $(1-\mathcal{B})^r$ is defined by the expansion

$$(1-\mathcal{B})^r = 1 - r\mathcal{B} - (2!)^{-1}r(1-r)\mathcal{B}^2 - (3!)^{-1}r(1-r)(2-r)\mathcal{B}^3 - \cdots$$

for $-0.5 < r < 0$ and $0 < r < 0.5$. For $r = 0$, the operator is the identity operator. The operator $(1-\mathcal{B})^d$ for $d \in (-0.5, 0.5)$ is called the *fractional difference operator*. The jth coefficient in the expansion of $(1-\mathcal{B})^{-d}$ is the v_j of (2.11.5).

For the time series (2.11.4), we can also write

$$(1-\mathcal{B})^d X_t = e_t \tag{2.11.10}$$

or

$$X_t + \sum_{j=1}^{\infty} \kappa_j(d) X_{t-j} = e_t, \tag{2.11.11}$$

where

$$\kappa_j(d) = [\Gamma(j+1)\Gamma(-d)]^{-1}\Gamma(j-d) = \prod_{i=1}^{j} i^{-1}(i-1-d). \tag{2.11.12}$$

If the time series satisfies (2.11.10) with $d \in (-0.5, 0.5)$, then it has an infinite moving average representation in which the coefficients decline at the rate j^{d-1} and an infinite autoregressive representation in which the coefficients decline at the

Table 2.11.1. Coefficients for fractional differenced time series

	$d = 0.4$			$d = -0.4$		
Index	v_j	κ_j	$\rho(j)$	v_j	κ_j	$\rho(j)$
1	0.4000	−0.4000	0.6667	−0.4000	0.4000	−0.2857
2	0.2800	−0.1200	0.5833	−0.1200	0.2800	−0.0714
3	0.2240	−0.0640	0.5385	−0.0640	0.2240	−0.0336
4	0.1904	−0.0416	0.5085	−0.0416	0.1904	−0.0199
5	0.1676	−0.0300	0.4864	−0.0300	0.1676	−0.0132
6	0.1508	−0.0230	0.4691	−0.0230	0.1508	−0.0095
7	0.1379	−0.0184	0.4548	−0.0184	0.1379	−0.0072
8	0.1275	−0.0152	0.4429	−0.0152	0.1275	−0.0057
9	0.1190	−0.0128	0.4326	−0.0128	0.1190	−0.0046
10	0.1119	−0.0110	0.4236	−0.0110	0.1119	−0.0038
15	0.0881	−0.0062	0.3906	−0.0062	0.0881	−0.0018
20	0.0743	−0.0041	0.3688	−0.0041	0.0743	−0.0011
25	0.0650	−0.0030	0.3527	−0.0030	0.0650	−0.0007
30	0.0583	−0.0023	0.3400	−0.0023	0.0583	−0.0005
35	0.0532	−0.0019	0.3297	−0.0019	0.0532	−0.0004
40	0.0491	−0.0015	0.3210	−0.0015	0.0491	−0.0003

rate j^{-d-1}. Thus, if $d > 0$, the moving average coefficients are square summable and the autoregressive coefficients are absolutely summable. If $d < 0$, the moving average coefficients are absolutely summable and the autoregressive coefficients are square summable. Examples of the coefficients are given in Table 2.11.1.

To generalize the class of time series with long memory properties, we assume Y_t satisfies

$$(1 - \mathcal{B})^d Y_t = Z_t, \tag{2.11.13}$$

where Z_t is the stationary autoregressive moving average satisfying

$$\sum_{j=0}^{p} \alpha_j Z_{t-j} = \sum_{i=0}^{q} \beta_i e_{t-i}, \tag{2.11.14}$$

e_t are uncorrelated $(0, \sigma^2)$ random variables, the roots of the characteristic equations associated with (2.11.14) are less than one in absolute value, and $d \in (-0.5, 0.5)$. The time series Y_t is sometimes called a fractionally differenced autoregressive moving average. The representation is sometimes abbreviated to ARIMA(p, d, q). By Theorem 2.2.2 and Theorem 2.2.3, the time series Y_t can be given an infinite moving average or an infinite autoregressive representation.

REFERENCES

Sections 2.1, 2.3, 2.5–2.9. Anderson (1971), Bartlett (1966), Box, Jenkins and Reinsel (1994), Hannan (1970), Kendall and Stuart (1966), Pierce (1970a), Wold (1938), Yule (1927).

Section 2.2. Bartle (1964), Pantula (1988a), Royden (1989), Torres (1986), Tucker (1967).

Section 2.4. Bellman (1960), Finkbeiner (1960), Goldberg (1958), Hildebrand (1968), Kempthorne and Folks (1971), Miller (1963, 1968).

Section 2.10. Anderson (1971), Brockwell and Davis (1991), Royden (1989), Tucker (1967), Wold (1938).

Section 2.11. Beran (1992), Cox (1984), Deo (1995), Granger (1980), Granger and Joyeux (1980), Hosking (1981, 1982), Mandelbrot (1969).

EXERCISES

1. Express the coefficients b_i of (2.4.3) in terms of the coefficients a_i of (2.4.2) for $n = 3$.

2. Given that $\{e_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ is a sequence of independent $(0, \sigma^2)$ random variables, define

$$(X_t, Y_t) = \left(\sum_{i=-2}^{2} e_{t-i}, \sum_{i=-1}^{1} X_{t-i} \right).$$

(a) Derive and graph $\gamma_X(h)$ and $\gamma_Y(h)$.

(b) Express Y_t as a moving average of e_t.

(c) Is it possible to express e_t as a finite moving average of X_t?

3. For the second order autoregressive process, obtain an expression for $\rho(h)$ as a function of α_1 and α_2 analogous to (2.5.10).

4. Given a second order moving average time series, what are the largest and smallest possible values for $\rho(1)$? For $\rho(2)$? For $\rho(3)$? Give an example of a moving average time series such that $\gamma(0) = 2$, $\gamma(1) = 0$, $\gamma(2) = -1$, and $\gamma(h) = 0$ for $|h| > 2$.

5. Consider the difference equation

$$y_t - 1.6y_{t-1} + 0.89y_{t-2} = 0, \qquad t = 2, 3, \ldots .$$

(a) Give the characteristic equation, and find its roots.

(b) Give the expression for y_t if $y_0 = 0.8$ and $y_1 = 0.9$.

(c) If $e_t \sim \text{NI}(0, \sigma^2)$, $\sigma^2 > 0$, $t = \{0, \pm 1, \pm 2, \ldots\}$, is there a stationary time series satisfying

$$Y_t - 1.6Y_{t-1} + 0.89Y_{t-2} = e_t?$$

Explain.

(d) Let the time series Y_t, $t = 2, 3, 4, \ldots$, be defined by

$$Y_t = 1.6Y_{t-1} - 0.89Y_{t-2} + e_t,$$

where $(Y_0, Y_1) = (0, 0)$ and $e_t \sim \text{NI}(0, \sigma^2)$. Give the covariance matrix for (Y_2, Y_3, Y_4).

(e) Define (Y_0, Y_1) so that Y_t of (d) is a stationary time series.

6. Find four (admittedly similar) time series with the covariance function

$$\gamma(h) = \begin{cases} 1, & h = 0, \\ 0.3, & |h| = 1, \\ 0 & \text{otherwise}. \end{cases}$$

7. Draw pictures illustrating the differences among the theoretical correlation functions for

(a) A (finite) moving average process.

(b) A (finite) autoregressive process.

(c) A strictly periodic process of the form

$$X_t = \sum_{i=1}^{M} \{e_{1i} \cos \lambda_i t + e_{2i} \sin \lambda_i t\}, \qquad M < \infty,$$

where e_{1i} and e_{2i} are independent $(0, \sigma^2)$ random variables independent of e_{1j} and e_{2j} for all $i \neq j$, and λ_i, $i = 1, 2, \ldots, M$, are distinct frequencies.

8. Consider the time series $\{X_t\}$ defined by

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} = e_t,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables and the roots of

$$m^3 + \alpha_1 m^2 + \alpha_2 m + \alpha_3 = 0$$

are less than one in absolute value. Develop a system of equations defining $\gamma(0)$ as a function of the α's.

9. Let the time series Y_t be defined by

$$Y_t = \beta_0 + \beta_1 t + X_t, \qquad t = 1, 2, \ldots,$$

where

$$X_t = e_t + 0.6e_{t-1},$$

$\{e_t : t \in (0, 1, 2, \ldots)\}$ is a sequence of normal independent $(0, \sigma^2)$ random variables, and β_0 and β_1 are fixed numbers. Give the mean and covariance functions for the time series.

10. The second order autoregressive time series

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = e_t,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables and the roots of $m^2 + a_1 m + a_2 = 0$ are less than one in absolute value, can be expressed as the infinite moving average

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}.$$

For the following three time series find and plot w_j for $j = 0, 1, \ldots, 8$ and $\rho(h)$ for $h = 0, 1, \ldots, 8$:

(a) $X_t + 1.6X_{t-1} + 0.64X_{t-2} = e_t$.

(b) $X_t - 0.4X_{t-1} - 0.45X_{t-2} = e_t$.

(c) $X_t - 1.2X_{t-1} + 0.85X_{t-2} = e_t$.

11. Given the time series

$$Y_t - 0.8Y_{t-1} = e_t + 0.7e_{t-1} + 0.6e_{t-2},$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables, find and plot $\rho(h)$ for $h = 0, 1, 2, \ldots, 8$.

12. Prove Theorem 2.1.1.

13. Find a second order autoregressive time series with complex roots whose absolute value is 0.8 that will give realizations with an apparent cycle of 20 time periods.

14. Given that X_t is a first order autoregressive process with parameter $\rho = 0.8$, use Theorem 2.9.1 to find the best one- and two-period-ahead predictions based on three observations.

15. Let the time series X_t satisfy

$$(X_t - 4) - 0.8(X_{t-1} - 4) = e_t,$$

where the e_t are uncorrelated $(0, 7)$ random variables. Five observations from a realization of the time series are given below:

t	X_t
1	5.9
2	4.9
3	2.2
4	2.0
5	4.9

Predict X_t for $t = 6$, 7, 8, and estimate the covariance matrix for your predictions, that is, estimate the matrix $E\{(\mathbf{X}_{5,3} - \hat{\mathbf{X}}_{5,3})(\mathbf{X}_{5,3} - \hat{\mathbf{X}}_{5,3})'\}$.

16. Let Y_t be defined by

$$Y_t = \rho Y_{t-1} + e_t, \qquad |\rho| < 1,$$

where $\{e_t, t \in (0, \pm 1, \pm 2, \ldots)\}$ is obtained from the sequence $\{u_t\}$ of normal independent $(0, 1)$ random variables as follows:

$$e_t = \begin{cases} u_t, & t = 0, \pm 2, \pm 4, \ldots, \\ 2^{-1/2}(u_{t-1}^2 - 1), & t = \pm 1, \pm 3, \ldots. \end{cases}$$

Is this a strictly stationary process? Is it a covariance stationary process? Given a sample of ten observations $(Y_1, Y_2, \ldots, Y_{10})$, what is the best linear predictor of Y_{11}? Can you construct a better predictor of Y_{11}?

17. Let Y_t be defined by

$$Y_t = e_t + \beta e_{t-1}, \qquad |\beta| < 1.$$

Show that the predictor of Y_{n+1} given in (2.9.17) is equivalent to the predictor $\hat{Y}_{n+1}(Y_1, \ldots, Y_n) = -\sum_{j=1}^{n} (-\beta)^j Y_{n-j+1}$.

18. Let $\{e_t\}$ be a sequence of independent $(0, 1)$ random variables. Let the time series X_t be defined by:

(a) $X_t = e_0$.

(b) $X_t = e_t - 1.2e_{t-1} + 0.36e_{t-2}$.

(c) $X_t = e_1 \cos 0.4\pi t + e_2 \sin 0.4\pi t + 6$.

(d) $X_t = e_t + 1.2e_{t-1}$.

Assume the covariance structure is known and we have available an infinite past for prediction, that is, $\{X_t : t \in (\ldots, -1, 0, 1, \ldots, n)\}$ is known. What is the variance of the prediction error for a predictor of X_{n+1} for each case above?

19. Show that if a stationary time series satisfies the difference equation

$$Y_t - Y_{t-1} = e_t,$$

then $E\{e_t^2\} = 0$.

20. Let $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ be defined by

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = e_t,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables, and the roots r_1 and r_2 of

$$m^2 + \alpha_1 m + \alpha_2 = 0$$

are less than one in absolute value.

(a) Let r_1 and r_2 be real. Show that $X_t - r_1 X_{t-1}$ is a first order autoregressive process with parameter r_2.

(b) If the roots r_1 and r_2 form a complex conjugate pair, how would you describe the time series $Y_t = X_t - r_1 X_{t-1}$?

21. Assume that X_t is a stationary, zero mean, normal time series with autocorrelation function $\rho_X(h)$. Let $Y_t = X_t^2$. Show that the autocorrelation function of Y_t is $\rho_Y(h) = \rho_X^2(h)$.

22. Let $\Delta y_t = a_1 + a_2 \cos \omega t$, $t = 1, 2, \ldots$, where a_1, a_2, and ω are real numbers, and $y_0 = 0$. Find y_t.

23. Let $Y_t = e_t - e_{t-1}$, where the e_t are normal independent $(0, \sigma^2)$ random variables. Given n observations from a realization, predict the $(n+1)$st observation in the realization. What is the mean square error of your predictor? *Hint*: Consider the time series

$$Z_t = \sum_{j=1}^{t} Y_j$$

and predict Z_{n+1}. How would your answer change if e_0 were considered fixed?

24. Prove the following lemma.

Lemma. Let $\lambda > \beta \geq 0$ and let p be a nonnegative integer. Then there exists an M such that $(t+1)^p \beta^t < M\lambda^t$ for all $t \geq 0$.

25. Let

$$Y_t - 0.8Y_{t-1} = e_t - 0.9e_{t-4},$$

where the e_t are uncorrelated $(0, 1)$ random variables. Using Theorem 2.7.1 obtain $E\{Y_t e_{t-4}\}$ and $E\{Y_{t-1} e_{t-4}\}$. Multiply the equation defining Y_t by Y_t and Y_{t-1}, and take expectations, to obtain a system of equations defining $\gamma_Y(0)$ and $\gamma_Y(1)$. Solve for the variance of Y_t.

26. Using the backward shift operator, we can write the autoregressive moving average process of order (p, q) as

$$\left(\sum_{j=0}^{p} a_j \mathcal{B}^j\right) X_t = \left(\sum_{i=0}^{q} b_i \mathcal{B}^i\right) e_t .$$

Show that the coefficients of the moving average representation

$$X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = \left(\sum_{j=0}^{\infty} \psi_j \mathcal{B}^j\right) e_t$$

given in Theorem 2.7.1 can be obtained from the quotient $(\Sigma_{j=0}^{p} a_j \mathcal{B}^j)^{-1} (\Sigma_{i=0}^{q} b_i \mathcal{B}^i)$ by long division. Obtain the autoregressive representation of Theorem 2.7.2 in the same manner.

27. Demonstrate how the nth order scalar difference equation $y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = 0$ can be written as a first order vector difference equation of dimension n. If $\mathbf{A}$ denotes the matrix of coefficients of the resulting first order difference equation, show that the roots of $|\mathbf{A} - m\mathbf{I}| = 0$ are the same as the roots of

$$m^n + a_1 m^{n-1} + \cdots + a_n = 0 .$$

28. Find $\Gamma(h)$ for the vector autoregressive time series

$$X_{1t} = 0.9X_{1,t-1} + e_{1t},$$
$$X_{2t} = 0.8X_{1,t-1} + 0.3X_{2,t-1} + e_{2t},$$

where $\mathbf{e}_t = (e_{1t}, e_{2t})'$ is a sequence of normal independent $(\mathbf{0}, \boldsymbol{\Sigma})$ random variables and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Hint: Express the covariance matrix as a linear function of the covariance functions of the two scalar autoregressive processes associated with the canonical form. Assume that we have observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ with $\mathbf{X}_n' = (3, 2)$. Predict $\mathbf{X}_{n+1}$ and $\mathbf{X}_{n+2}$ and obtain the covariance matrix of the prediction errors.

29. The cobweb theory has been suggested as a model for the price behavior of some agricultural commodities. The model is composed of a demand equation and a supply equation:

$$\text{demand} \quad Q_t = \alpha_1 + \beta_1 P_t + U_{1t},$$
$$\text{supply} \quad Q_t = \alpha_2 + \beta_2 P_{t-1} + U_{2t},$$

where Q_t is the quantity produced for time t, P_t is the price at time t, and $\mathbf{U}_t' = (U_{1t}, U_{2t})$ is a stationary vector time series. Assuming $\mathbf{U}_t$ is a sequence of uncorrelated $(\mathbf{0}, \boldsymbol{\Sigma})$ vector random variables, express the model as a vector autoregressive process. Under what conditions will the process be stationary? If $U_{1t} = e_{1t} + b_{11}e_{1,t-1}$ and $U_{2t} = e_{2t} + b_{22}e_{2,t-1}$ where $\mathbf{e}_t = (e_{1t}, e_{2t})'$ is a sequence of uncorrelated $(\mathbf{0}, \sigma^2\mathbf{I})$ vector random variables, how would you describe the vector time series $(Q_t, P_t)'$?

30. Let e_t be a sequence of independent $(0, \sigma^2)$ random variables, and let $\{a_j\}$ be an absolutely summable sequence of real numbers. Use Theorem 2.2.2 to show that

$$X_t = \sum_{j=0}^{\infty} a_j e_{t-j}$$

is defined as an almost sure limit and X_t is a stationary process with zero mean and

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+h}.$$

31. Show that the assumption that $E\{W_{it}^2\}$ is bounded below guarantees the existence of the inverse of $\mathbf{V}_{it}$ in Theorem 2.9.1 used to define (2.10.3).

32. Let the stationary pth order autoregressive process satisfy

$$\sum_{j=0}^{p} a_j X_{t-j} = e_t ,$$

where $a_0 = 1$ and e_t is a sequence of uncorrelated $(0, \sigma^2)$ variables. The equation can also be written

$$\Phi(\mathcal{B})X_t = e_t ,$$

where $\mathcal{B}$ is the back shift operator defined in Section 2.4, and

$$\Phi(\mathcal{B}) = 1 + a_1\mathcal{B} + a_2\mathcal{B}^2 + \cdots + a_p\mathcal{B}^p .$$

Some authors define a stationary process Y_t to be *causal* if it is possible to represent Y_t as

$$Y_t = \sum_{j=0}^{\infty} w_j e_{t-j} , \qquad t = 0, \pm 1, \pm 2, \ldots ,$$

where $\sum_{j=0}^{\infty} |w_j| < \infty$. Show that the following statements are equivalent:

(i) X_t is causal.

(ii) The roots of the characteristic equation

$$m^p + a_1 m^{p-1} + \cdots + a_p = 0$$

are less than one in absolute value.

(iii) $\Phi(z) \neq 0$ for all complex numbers z such that $|z| \leq 1$.

(iv) All roots of $\Phi(z) = 0$ are greater than one in absolute value.

33. Using (2.6.8) and the fact that $\alpha_p \neq 0$, show that the autocorrelation function of the pth order autoregressive process is nonzero for some integer greater than N_0 for all N_0.

34. Let

$$Y_t = e_0 \cos \pi t + Z_t , \qquad t = 1, 2, \ldots ,$$

where $e_0 \sim N(0, \sigma_{ee})$, independent of $Z_t \sim \mathrm{NI}(0, \sigma_{zz})$. Let σ_{ee} and σ_{zz} be known.

(a) Given three observations (Y_1, Y_2, Y_3), what is the minimum mean square error predictor of Y_4? Give the variance of the prediction error.

(b) What is the limiting expression for the predictor of Y_{n+1} given $(Y_1, Y_2, \ldots, Y_n)$, where the limit is taken as n increases? What is the limit of the variance of the prediction error?

35. Let $(Y_1, \ldots, Y_n)$ be observations on the first order moving average,

$$Y_t = e_t + \beta_1 e_{t-1},$$

where $e_t \sim \text{NI}(0, \sigma^2)$ and $|\beta_1| < 1$. Construct the best predictor $\hat{e}_0(Y_1, Y_2, Y_3)$ of e_0. What is the variance of the prediction error of $\hat{e}_0(Y_1, Y_2, Y_3)$? What is the limit of the variance of the prediction error of $\hat{e}_0(Y_1, \ldots, Y_n)$ as $n \to \infty$?

36. Let the stationary vector autoregressive process $\mathbf{Y}_t$ satisfy

$$\mathbf{Y}_t + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{Y}_{t-j} = \mathbf{e}_t,$$

where the $\mathbf{e}_t$ are independent $(\mathbf{0}, \boldsymbol{\Sigma}_{ee})$ random vectors. Show that the process also satisfies an equation of the form

$$\mathbf{Y}_t + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{Y}_{t+j} = \mathbf{a}_t,$$

where the $\mathbf{a}_t$ are uncorrelated $(\mathbf{0}, \boldsymbol{\Sigma}_{aa})$ random vectors. Give the equations defining the $\mathbf{B}_j$, $j = 1, 2, \ldots, p$, as a function of the $\boldsymbol{\Gamma}(h)$. Show that if $p = 1$, $\mathbf{A}_1 = \mathbf{A}_1'$, and $\boldsymbol{\Sigma}_{ee} = \mathbf{I}$, then $\boldsymbol{\Sigma}_{aa} = \mathbf{I}$.

37. Let a_t be a sequence of uncorrelated $(0, 1)$ random variables. Put the following time series in canonical form as moving averages of uncorrelated random variables e_t: (a) $a_t + 4a_{t-1}$, (b) $a_{t-1} + 4a_t + a_{t+1}$, (c) $a_t - 2.5a_{t-1} + a_{t-2}$. Give the variance of e_t in each case.

38. Let $\mathbf{V}_n$ denote the $n \times n$ covariance matrix of n observations $(Y_1, Y_2, \ldots, Y_n)$ on the first order moving average process

$$Y_t = e_t + \beta e_{t-1},$$

where the e_t are uncorrelated $(0, 1)$ random variables. Let D_n be the determinant of $\mathbf{V}_n$. Show that

$$D_n = (1 + \beta^2) D_{n-1} - \beta^2 D_{n-2}$$

for $n = 2, 3, \ldots$, with $D_0 = 1$ and $D_1 = 1 + \beta^2$. Hence, show that $D_n = (1 - \beta^2)^{-1}(1 - \beta^{2(n+1)})$, $n = 1, 2, \ldots$, for $|\beta| < 1$, and $D_n = n + 1$, $n = 1, 2, \ldots$, for $|\beta| = 1$.

39. Let Y_t be a zero mean stationary time series with $\gamma(0) > 0$, $\lim_{n \to \infty} \gamma(h) = 0$, and $\phi(h) = 0$ for $h > K$. Show that Y_t is nondeterministic.

40. Consider the model $Y_t = \mathbf{x}_t' \boldsymbol{\beta} + Z_t$, $t = 1, \ldots, n$, where Z_t is a process with mean zero and known covariance function $\text{Cov}(Z_t, Z_j) = V_{ij}$.

(a) Show that the best linear unbiased estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_G = (\mathbf{X}'\mathbf{V}_{nn}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{nn}^{-1}\mathbf{y},$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, $\mathbf{y} = (Y_1, \ldots, Y_n)'$, and $\mathbf{V}_{nn} = \mathbf{V}\{\mathbf{y}\}$ is nonsingular.

(b) Given $Y_1, \ldots, Y_n$, $\mathbf{X}'_{n+s}$, and $\mathbf{V}_{ns} = \mathrm{Cov}(\mathbf{y}, Y_{n+s})$, show that the best linear unbiased predictor of $Y_{n+s} = \mathbf{x}'_{n+s}\boldsymbol{\beta} + Z_{n+s}$ is

$$\mathbf{x}'_{n+s}\hat{\boldsymbol{\beta}}_G + \mathbf{b}'_{ns}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_G),$$

where $\mathbf{b}_{ns} = \mathbf{V}_{nn}^{-1}\mathbf{V}_{ns}$.

41. Consider a stationary and invertible time series Y_t given by

$$Y_t = \sum_{j=1}^{\infty} w_j e_{t-j} + e_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + e_t,$$

where $\Sigma_{j=1}^{\infty} |w_j| < \infty$, $\Sigma_{j=1}^{\infty} |\pi_j| < \infty$, and the e_t are uncorrelated $(0, \sigma^2)$ random variables. Show that

$$\lim_{n\to\infty} E\left[\left(\hat{Y}_{t,n} - \sum_{j=1}^{n} \pi_j Y_{t-j}\right)^2\right] = 0,$$

where $\hat{Y}_{t,n} = \hat{Y}_t(Y_{t-1}, \ldots, Y_{t-n})$ is the best linear predictor of Y_t given $Y_{t-1}, \ldots, Y_{t-n}$.

42. In Theorem 2.9.1, the fact that the system of equations $\mathbf{V}_{nn}\mathbf{b} = \mathbf{V}_{ns}$ is consistent is used. Show that this system is always consistent and that $\mathbf{y}'\mathbf{b}_{ns}$ in (2.9.2) is unique a.e. for different choices of generalized inverses of $\mathbf{V}_{nn}$.

43. Consider a sequence satisfying $e_t = Z_t(\beta_0 + \beta_1 e_{t-1}^2)^{1/2}$, where $Z_t \sim \mathrm{NI}(0, 1)$, $\beta_0 > 0$ and $0 \leq \beta_1 < 1$. That is, the distribution of e_t given the past is $\mathrm{N}(0, h_t)$, where $h_t = \beta_0 + \beta_1 e_{t-1}^2$. Thus, the conditional variance depends on the past errors. The model

$$Y_t = \alpha Y_{t-1} + e_t$$

is a special case of autoregressive conditionally heteroscedastic models called an ARCH(1) model. See Engle (1982). Note that

$$\begin{aligned} e_t^2 = Z_t^2(\beta_0 + \beta_1 e_{t-1}^2) &= Z_t^2[\beta_0 + \beta_1 Z_{t-1}^2(\beta_0 + \beta_1 e_{t-2}^2)] \cdots \\ &= \beta_0 \sum_{j=0}^{\infty} \beta_1^j \left(\prod_{i=0}^{j} Z_{t-j}^2\right) \quad \text{a.s.}, \end{aligned}$$

where it is assumed that the process started with finite variance in the indefinite past.

(a) Show that $\{e_t\}$ is a sequence of uncorrelated $[0, (1-\beta_1)^{-1}\beta_0]$ random variables.

(b) Show that if $3\beta_1^2 < 1$, then $E\{e_t^4\}$ exists. Find $E\{e_t^4\}$.

(c) Consider $X_t = e_t^2$. Assuming $3\beta_1^2 < 1$, show that X_t is stationary. Give its autocorrelation function.

(d) Consider the stationary ARCH(1) model, $Y_t = \alpha_1 Y_{t-1} + e_t$, where $|\alpha_1| < 1$. Assume $(\alpha_1, \beta_0, \beta_1)$ are known.

(i) What is the best predictor $\hat{Y}_{n+s}$ for Y_{n+s} given $Y_1, \ldots, Y_n$?

(ii) Find $V\{Y_{n+s} - \hat{Y}_{n+s}\}$, the unconditional forecast error variance.

(iii) Find $V\{(Y_{n+s} - \hat{Y}_{n+s}) | (e_1, \ldots, e_n)\}$, the conditional forecast error variance.

(iv) Show that $V\{(Y_{n+2} - \hat{Y}_{n+2}) | (e_1, \ldots, e_n)\}$ may be less than $V\{(Y_{n+1} - \hat{Y}_{n+1}) | (e_1, \ldots, e_n)\}$ for some β_0, β_1, α_1, and e_n^2. (That is, the conditional forecast error variance for two-step-ahead forecasting may be less than that for one-step-ahead forecasting.)

(v) Show that

$$\lim_{s\to\infty} V\{Y_{n+s} - \hat{Y}_{n+s} | (e_1, \ldots, e_n)\} = (1-\alpha_1^2)^{-1}\sigma^2 \qquad \text{a.s.},$$

where $\sigma^2 = (1-\beta_1)^{-1}\beta_0$.

44. Use the facts that, for $-0.5 < d < 0.5$ and $h > 0$,

$$2^{d-1} i^{d-1} < i^{d-1}(1+h^{-1}i)^{d-1} < i^{d-1} \qquad \text{for} \quad 1 \le i \le h,$$
$$(i+h)^{2d-2} < i^{d-1}(i+h)^{d-1} < 2^{d-1} i^{2d-2} \qquad \text{for} \quad i > h$$

to show that $\gamma_Y(h)$ of (2.11.3) is bounded above by $c_1 h^{2d-1}$ and bounded below by $c_2 h^{2d-1}$, where c_1 and c_2 are constants.

45. Let $\{a_j\}_{-\infty}^{\infty}$ and $\{b_j\}_{-\infty}^{\infty}$ be sequences of real numbers satisfying

$$\sum_{j=-\infty}^{\infty} |a_j| = M < \infty \quad \text{and} \quad \sum_{j=-\infty}^{\infty} b_j^2 < \infty.$$

Show that $\sum_{r=-\infty}^{\infty} c_r^2 < \infty$, where

$$c_r = \sum_{i=-\infty}^{\infty} b_r a_{r-i} = \sum_{j=-\infty}^{\infty} a_r b_{r-j}.$$