2

Fundamental Concepts

In this chapter, we introduce some fundamental concepts that are necessary for proper understanding of time series models discussed in this book. We begin with a simple introduction to stochastic processes, the autocorrelation and partial autocorrelation functions, and the notion of white noise processes. We then discuss estimation of the mean, autocovariances, and the autocorrelation and partial autocorrelation functions. Thus, we can illustrate the sampling phenomena of time series models starting in Chapter 3, which, we believe, will enhance the appreciation of model identification discussed in later chapters. In addition to its own theoretical merits, the notion of the moving average and autoregressive representations of processes is useful for understanding the logic underlying parsimonious linear processes used in time series analysis. We also give a simple introduction to linear difference equations, with a special emphasis on the solution of homogeneous difference equations, which play a fundamental role in the parsimonious linear time series processes.

2.1 Stochastic Processes

A stochastic process is a family of time indexed random variables $Z(\omega, t)$, where ω belongs to a sample space and t belongs to an index set. For a fixed t, $Z(\omega, t)$ is a random variable. For a given ω, $Z(\omega, t)$, as a function of t, is called a sample function or realization. The population that consists of all possible realizations is called the ensemble in stochastic processes and time series analysis. Thus, a time series is a realization or sample function from a certain stochastic process. To have a proper appreciation for time series analysis, we introduce some essential concepts of stochastic processes in this section.

In our discussion, we assume that the index set is the set of all integers unless mentioned otherwise. Consider a finite set of random variables $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\}$ from a stochastic process $\{Z(\omega, t): t = 0, \pm 1, \pm 2, \ldots\}$. The n-dimensional distribution function is defined by

$$F_{Z_{t_1},\ldots,Z_{t_n}}(x_1, \ldots, x_n) = P\{\omega: Z_{t_1} \leq x_1, \ldots, Z_{t_n} \leq x_n\}, \tag{2.1.1}$$

where x_i, $i = 1, \ldots, n$ are any real numbers. A process is said to be first-order stationary in distribution if its one-dimensional distribution function is time invariant, i.e., if $F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1)$ for any integers t_1, k, and $t_1 + k$; second-order stationary in distribution if

$F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ for any integers t_1, t_2, k, $t_1 + k$, and $t_2 + k$; and nth-order stationary in distribution if

$$F_{Z_{t_1}, \ldots, Z_{t_n}}(x_1, \ldots, x_n) = F_{Z_{t_1+k}, \ldots, Z_{t_n+k}}(x_1, \ldots, x_n) \tag{2.1.2}$$

for any n-tuple $(t_1, \ldots, t_n)$ and k of integers. A process is said to be strictly stationary if (2.1.2) is true for any n, i.e., $n = 1, 2, \ldots$. The terms *strongly stationary* and *completely stationary* are also used to denote a strictly stationary process. Clearly, if (2.1.2) is true for $n = m$, it is also true for $n \leq m$ because the mth-order distribution function determines all distribution functions of lower order. Hence, a higher order of stationarity always implies a lower order of stationarity.

With proper understanding that a stochastic process, $Z(\omega, t)$, is a set of time indexed random variables defined on a sample space, we usually suppress the variable ω and simply write $Z(\omega, t)$ as $Z(t)$ or Z_t, just as we often denote a random variable by X rather than by $X(\omega)$. The process is called a real-valued process if it assumes only real values. Unless mentioned otherwise, processes discussed in this book are referred to as the real-valued processes. For a given real-valued process $\{Z_t : t = 0, \pm 1, \pm 2, \ldots\}$, we define the mean function of the process

$$\mu_t = E(Z_t), \tag{2.1.3}$$

the variance function of the process

$$\sigma_t^2 = E(Z_t - \mu_t)^2, \tag{2.1.4}$$

the covariance function between Z_{t_1} and Z_{t_2}

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}), \tag{2.1.5}$$

and the correlation function between Z_{t_1} and Z_{t_2}

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}. \tag{2.1.6}$$

For a strictly stationary process, since the distribution function is the same for all t, the mean function $\mu_t = \mu$ is a constant, provided $E(|Z_t|) < \infty$. Likewise, if $E(Z_t^2) < \infty$, then $\sigma_t^2 = \sigma^2$ for all t and hence is also a constant. Moreover, since $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ for any integers t_1, t_2, and k, we have

$$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k)$$

and

$$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k).$$

Letting $t_1 = t - k$ and $t_2 = t$, we get

$$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k \tag{2.1.7}$$

and

$$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k. \tag{2.1.8}$$

Thus, for a strictly stationary process with the first two moments finite, the covariance and the correlation between Z_t and Z_{t+k} depend only on the time difference k.

So far, we have discussed a strong sense of stationarity of a process in terms of its distribution function. A trivial example of a strictly stationary process is a sequence of independent identically distributed (i.i.d.) random variables. This sequence of independent random variables usually does not exist or renders no interest in time series. Other than this simple i.i.d. case, however, it is very difficult or impossible actually to verify a distribution function, particularly a joint distribution function from an observed time series. Thus, in time series analysis, we often use a weaker sense of stationarity in terms of the moments of the process.

A process is said to be nth-order weakly stationary if all its joint moments up to order n exist and are time invariant, i.e., independent of time origin. Therefore, a second-order weakly stationary process will have constant mean and variance, with the covariance and the correlation being functions of the time difference alone. Sometimes, the terms *stationary in the wide sense* or *covariance stationary* are also used to describe a second-order weakly stationary process. It follows from the definitions that a strictly stationary process with the first two moments finite is also a second-order weakly or covariance stationary process. Yet, a strictly stationary process may not have finite moments and hence may not be covariance stationary. A trivial example is the process consisting of a sequence of i.i.d. Cauchy random variables. The process is clearly strictly stationary, but it is not weakly stationary of any order because no joint moments exist.

EXAMPLE 2.1 Let Z_t be twice the value of a true die shown on the tth toss. If the die is tossed three times independently, we have a stochastic process $Z_t = Z(\omega, t)$ where t belongs to the index set $= \{1, 2, 3\}$ and ω belongs to the sample space

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

For a particular ω, say (1, 3, 2), the realization or sample function will be (2, 6, 4). The ensemble contains a total of 216 possible realizations. If a true die is tossed independently and repeatedly, then we have a stochastic process $Z_t = Z(\omega, t)$ with the index set being all positive integers and the sample space being

$$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \cdots.$$

A particular ω of (1, 4, 2, 3, . . .) would have the realization (2, 8, 4, 6, . . .). In this case, the total number of possible realizations in the ensemble is infinite. Clearly, the process in either case is strictly stationary as it relates to a sequence of i.i.d. random variables.

EXAMPLE 2.2 Consider the time sequence

$$Z_t = A\sin(\omega t + \theta), \tag{2.1.9}$$

where A is a random variable with a zero mean and a unit variance and θ is a random variable with a uniform distribution on the interval $[-\pi, \pi]$ independent of A. Then

$$\begin{aligned}
E(Z_t) &= E(A)E[\sin(\omega t + \theta)] = 0 \\
E(Z_t Z_{t+k}) &= E\{A^2 \sin(\omega t + \theta)\sin[\omega(t+k) + \theta]\} \\
&= E(A^2)E\left\{\frac{1}{2}[\cos(\omega k) - \cos(\omega(2t+k) + 2\theta)]\right\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}E\{\cos(\omega(2t+k) + 2\theta)\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}\int_{-\pi}^{\pi} \cos(\omega(2t+k) + 2\theta)\cdot\frac{1}{2\pi}d\theta \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{8\pi}[\sin(\omega(2t+k) + 2\theta)]_{-\pi}^{\pi} \\
&= \frac{1}{2}\cos(\omega k),
\end{aligned} \tag{2.1.10}$$

which depends only on the time difference k. Hence, the process is covariance stationary.

EXAMPLE 2.3 Let Z_t be a sequence of independent random variables alternately following a standard normal distribution $N(0, 1)$ and a two-valued discrete uniform distribution with equal probability 1/2 of being 1 or -1. Clearly, $E(Z_t) = 0$ and $E(Z_t^2) = 1$ for all t. Now

$$E(Z_t Z_s) = \begin{cases} 0, & \text{if } t \neq s, \\ 1, & \text{if } t = s, \end{cases}$$

and

$$\rho(t, s) = \frac{E(Z_t Z_s)}{\sqrt{E(Z_t^2)}\sqrt{E(Z_s^2)}} = \begin{cases} 0, & \text{if } t \neq s, \\ 1, & \text{if } t = s. \end{cases}$$

Hence, the process is covariance stationary. The process, however, is not strictly stationary. It is, in fact, not stationary in distribution for any order.

From these discussions and examples, it is clear that "covariance stationary" is a much weaker form of stationarity than are "strictly stationary" or "stationary in distribution." We often work, however, with covariance stationary or second-order weakly stationary processes

in time series analysis because it is relatively simple to check the first two moments. Henceforth, unless mentioned otherwise, we use the term *stationary* to refer to all processes that are covariance stationary. With this, the following important remark is in order.

A stochastic process is said to be a normal or Gaussian process if its joint probability distribution is normal. Because a normal distribution is uniquely characterized by its first two moments, strictly stationary and weakly stationary are equivalent for a Gaussian process. Unless mentioned otherwise, processes which we discuss are assumed to be Gaussian. Like other areas in statistics, most time series results are established for Gaussian processes. Thus, the autocorrelation functions and the partial autocorrelation functions discussed in the next two sections become fundamental tools in time series analysis.

2.2 The Autocovariance and Autocorrelation Functions

For a stationary process $\{Z_t\}$, we have the mean $E(Z_t) = \mu$ and variance $\text{Var}(Z_t) = E(Z_t - \mu)^2 = \sigma^2$, which are constant, and the covariances $\text{Cov}(Z_t, Z_s)$, which are functions only of the time difference $|t - s|$. Hence, in this case, we write the covariance between Z_t and Z_{t+k} as

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu), \tag{2.2.1}$$

and the correlation between Z_t and Z_{t+k} as

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \tag{2.2.2}$$

where we note that $\text{Var}(Z_t) = \text{Var}(Z_{t+k}) = \gamma_0$. As functions of k, γ_k is called the autocovariance function and ρ_k is called the autocorrelation function (ACF) in time series analysis because they represent the covariance and correlation between Z_t and Z_{t+k} from the same process, separated only by k time lags.

It is easy to see that for a stationary process, the autocovariance function γ_k and the autocorrelation function ρ_k have the following properties:

1. $\gamma_0 = \text{Var}(Z_t)$; $\rho_0 = 1$.
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$ for all k, i.e., γ_k and ρ_k are even functions and hence symmetric about the lag $k = 0$. This property follows from the time difference between Z_t and Z_{t+k} and Z_t and Z_{t-k} being the same. Therefore, the autocorrelation function is often plotted only for the nonnegative lags such as the one shown in Figure 2.1. This plot is sometimes called a correlogram.
4. Another important property of the autocovariance function γ_k and the autocorrelation function ρ_k is that they are positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0 \tag{2.2.3}$$

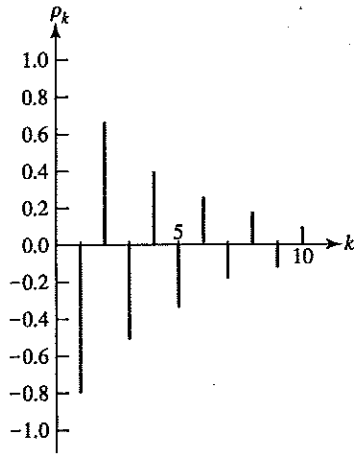

FIGURE 2.1 Example of autocorrelation function (ACF).

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_{|t_i-t_j|} \geq 0 \tag{2.2.4}$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any real numbers $\alpha_1, \alpha_2, \ldots,$ and α_n. By defining the random variable $X = \sum_{i=1}^{n}\alpha_i Z_{t_i}$, the result (2.2.3) follows from

$$0 \leq \mathrm{Var}(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \mathrm{Cov}(Z_{t_i}, Z_{t_j}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|}.$$

The similar result for ρ_k in (2.2.4) follows immediately by dividing the inequality (2.2.3) through by γ_0. Thus, it is important to know that not every arbitrary function satisfying properties (1) to (3) can be an autocovariance or autocorrelation function for a process. A necessary condition for a function to be the autocovariance or autocorrelation function of some process is that it be positive semidefinite.

2.3 The Partial Autocorrelation Function

In addition to the autocorrelation between Z_t and Z_{t+k}, we may want to investigate the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, \ldots,$ and Z_{t+k-1} has been removed. The conditional correlation

$$\mathrm{Corr}(Z_t, Z_{t+k}|Z_{t+1}, \ldots, Z_{t+k-1}) \tag{2.3.1}$$

is usually referred to as the partial autocorrelation in time series analysis.

Consider a stationary process $\{Z_t\}$ and, without loss of generality, we assume that $E(Z_t) = 0$. Let the linear dependence of Z_{t+k} on Z_{t+1}, Z_{t+2}, . . . , and Z_{t+k-1} be defined as the best linear estimate in the mean square sense of Z_{t+k} as a linear function of Z_{t+1}, Z_{t+2}, . . . , and Z_{t+k-1}. That is, if $\hat{Z}_{t+k}$ is the best linear estimate of Z_{t+k}, then

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+k-1} + \alpha_2 Z_{t+k-2} + \cdots + \alpha_{k-1} Z_{t+1}, \tag{2.3.2}$$

where $\alpha_i (1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained from minimizing

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 = E(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2. \tag{2.3.3}$$

The routine minimization method through differentiation gives the following linear system of equations

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \cdots + \alpha_{k-1} \gamma_{i-k+1} \qquad (1 \leq i \leq k-1). \tag{2.3.4}$$

Hence,

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \cdots + \alpha_{k-1} \rho_{i-k+1} \qquad (1 \leq i \leq k-1). \tag{2.3.5}$$

In terms of matrix notation, the above system of (2.3.5) becomes

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}. \tag{2.3.6}$$

Similarly,

$$\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \cdots + \beta_{k-1} Z_{t+k-1}, \tag{2.3.7}$$

where $\beta_i (1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained by minimizing

$$E(Z_t - \hat{Z}_t)^2 = E(Z_t - \beta_1 Z_{t+1} - \cdots - \beta_{k-1} Z_{t+k-1})^2. \tag{2.3.8}$$

Hence,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}, \tag{2.3.9}$$

which implies that $\alpha_i = \beta_i (1 \leq i \leq k-1)$.

It follows that the partial autocorrelation between Z_t and Z_{t+k} will equal the ordinary autocorrelation between $(Z_t - \hat{Z}_t)$ and $(Z_{t+k} - \hat{Z}_{t+k})$. Thus, letting P_k denote the partial autocorrelation between Z_t and Z_{t+k}, we have

$$P_k = \frac{\text{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{\text{Var}(Z_t - \hat{Z}_t)}\sqrt{\text{Var}(Z_{t+k} - \hat{Z}_{t+k})}}. \tag{2.3.10}$$

Now,

$$\begin{aligned}
\text{Var}(Z_{t+k} - \hat{Z}_{t+k}) &= E[(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})^2] \\
&= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&\quad - \alpha_1 E[Z_{t+k-1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&\quad - \cdots - \alpha_{k-1}E[Z_{t+1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})]
\end{aligned}$$

because all other remaining terms reduce to zero by virtue of Equation (2.3.4). Hence,

$$\text{Var}(Z_{t+k} - \hat{Z}_{t+k}) = \text{Var}(Z_t - \hat{Z}_t) = \gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}. \tag{2.3.11}$$

Next, using that $\alpha_i = \beta_i (1 \le i \le k-1)$, we have

$$\begin{aligned}
&\text{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1}Z_{t+k-1})(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1}Z_{t+1})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1}Z_{t+k-1})Z_{t+k}] \\
&= \gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1.
\end{aligned} \tag{2.3.12}$$

Therefore,

$$P_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}} = \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \cdots - \alpha_{k-1}\rho_{k-1}}. \tag{2.3.13}$$

Solving the system in (2.3.6) for α_i by Cramer's rule gives

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_{i-1} & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_{i-2} & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}} \tag{2.3.14}$$

as the ratio of two determinants. The matrix in the numerator is the same as the symmetric matrix in the denominator except for its ith column being replaced by $(\rho_1, \rho_2, \ldots, \rho_{k-1})$. Substituting α_i in (2.3.14) to (2.3.13) and multiplying both the numerator and denominator of (2.3.13) by the determinant

$$\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{vmatrix},$$

the resulting P_k in (2.3.13) can be easily seen to equal the ratio of the determinants in (2.3.15) each expanded by its last column,

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}. \tag{2.3.15}$$

The partial autocorrelation can also be derived as follows. Consider the regression model, where the dependent variable Z_{t+k} from a zero mean stationary process is regressed on k lagged variables $Z_{t+k-1}, Z_{t+k-2}, \ldots$ and Z_t, i.e.,

$$Z_{t+k} = \phi_{k1}Z_{t+k-1} + \phi_{k2}Z_{t+k-2} + \cdots + \phi_{kk}Z_t + e_{t+k}, \tag{2.3.16}$$

where ϕ_{ki} denotes the ith regression parameter and e_{t+k} is an error term with mean 0 and uncorrelated with Z_{t+k-j} for $j=1, 2, \ldots, k$. Multiplying Z_{t+k-j} on both sides of the above regression equation and taking the expectation, we get

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}; \tag{2.3.17}$$

hence,

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}. \tag{2.3.18}$$

For $j = 1, 2, \ldots, k$, we have the following system of equations:

$$\begin{aligned} \rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0. \end{aligned}$$

Using Cramer's rule successively for $k = 1, 2, \ldots$, we have

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$\vdots$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}. \tag{2.3.19}$$

Comparing Equation (2.3.19) with (2.3.15), we see that ϕ_{kk} equals P_k. Thus, the partial autocorrelation between Z_t and Z_{t+k} can also be obtained as the regression coefficient associated with Z_t when regressing Z_{t+k} on its k lagged variables $Z_{t+k-1}, Z_{t+k-2}, \ldots$, and Z_t as in (2.3.16). Because ϕ_{kk} has become a standard notation for the partial autocorrelation between Z_t and Z_{t+k} in time series literature, we use this notation in our book. As a function of k, ϕ_{kk} is usually referred to as the partial autocorrelation function (PACF).

2.4 White Noise Processes

A process $\{a_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $E(a_t) = \mu_a$, usually assumed to be 0, constant variance $\text{Var}(a_t) = \sigma_a^2$ and $\gamma_k = \text{Cov}(a_t, a_{t+k}) = 0$ for all $k \neq 0$. By definition, it immediately follows that a white noise process $\{a_t\}$ is stationary with the autocovariance function

$$\gamma_k = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{2.4.1}$$

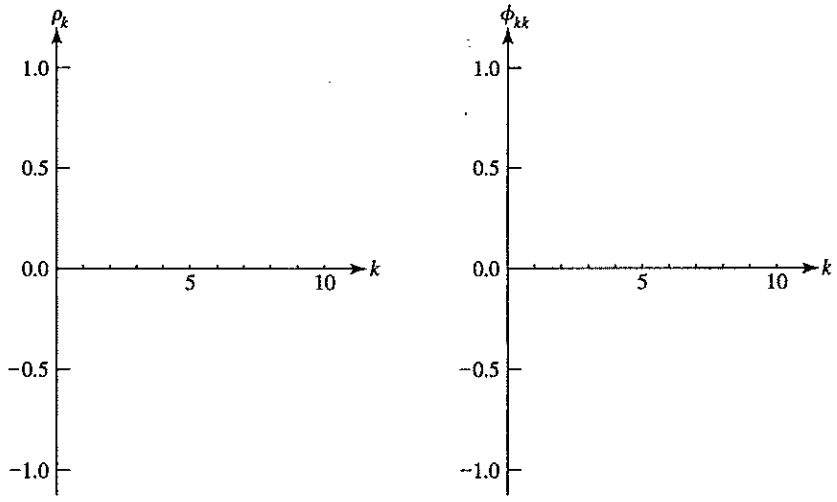

FIGURE 2.2 ACF and PACF of a white noise process: $Z_t = \mu + a_t$.

the autocorrelation function

$$\rho_k = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{2.4.2}$$

and the partial autocorrelation function

$$\phi_{kk} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases} \tag{2.4.3}$$

The ACF and PACF of a white noise process are shown in Figure 2.2.

Because by definition $\rho_0 = \phi_{00} = 1$ for any process, when we talk about the autocorrelation and partial autocorrelations, we refer only to ρ_k and ϕ_{kk} for $k \neq 0$. The basic phenomenon of the white noise process is that its ACF and PACF are identically equal to zero.

Although this process hardly ever occurs in practice, it plays an important role as a basic building block in the construction of time series models, just like the roles played by the sine and the cosine functions $\{\sin(nx), \cos(nx)\}$ in Fourier analysis. More precisely, it plays the role of an orthogonal basis in the general vector and function analysis.

A white noise process is Gaussian if its joint distribution is normal. In the following discussion, unless mentioned otherwise, $\{a_t\}$ is always referred to as a zero mean Gaussian white noise process.

2.5 Estimation of the Mean, Autocovariances, and Autocorrelations

A stationary time series is characterized by its mean μ, variance σ^2, autocorrelations ρ_k, and partial autocorrelations ϕ_{kk}. The exact values of these parameters can be calculated if the ensemble of all possible realizations is known. Otherwise, they can be estimated if

multiple independent realizations are available. In most applications, however, it is difficult or impossible to obtain multiple realizations. Most available time series constitute only a single realization, which makes it impossible to calculate the ensemble average. For a stationary process, however, we have a natural alternative of replacing the ensemble average by the time average. In the following discussion, we examine conditions under which we can estimate with good statistical properties the mean and autocovariances and hence the autocorrelations by using time averages.

2.5.1 Sample Mean

With only a single realization, a natural estimator for the mean $\mu = E(Z_t)$ of a stationary process is the sample mean

$$\overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t, \tag{2.5.1}$$

which is the time average of n observations. The question becomes whether the above estimator is a valid or good estimator. Clearly,

$$E(\overline{Z}) = \frac{1}{n}\sum_{t=1}^{n} E(Z_t) = \frac{1}{n}\cdot n\mu = \mu, \tag{2.5.2}$$

which implies that $\overline{Z}$ is an unbiased estimator of μ. It can also be easily shown that

$$\begin{aligned} \mathrm{Var}(\overline{Z}) &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} \mathrm{Cov}(Z_t, Z_s) = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_{(t-s)} \qquad (2.5.3)\\ &= \frac{\gamma_0}{n^2}\sum_{k=-(n-1)}^{n-1}(n - |k|)\rho_k \\ &= \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1}\left(1 - \frac{|k|}{n}\right)\rho_k, \qquad (2.5.4) \end{aligned}$$

where we let $k = (t - s)$. Thus, if

$$\lim_{n\to\infty}\left[\sum_{k=-(n-1)}^{n-1}\left(1 - \frac{|k|}{n}\right)\rho_k\right]$$

is finite, then $\mathrm{Var}(\overline{Z}) \to 0$ as $n \to \infty$, and $\overline{Z}$ is a consistent estimator for μ. That is,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} Z_t = \mu \tag{2.5.5}$$

in mean square. The process is said to be ergodic for the mean if the result in (2.5.5) holds. A sufficient condition for this result to hold is that $\rho_k \to 0$ as $k \to \infty$, because $\rho_k \to 0$ as $k \to \infty$ implies that for any $\epsilon > 0$, we can choose an N such that $|\rho_k| < \epsilon/4$ for all $k > N$. Hence, for $n > (N + 1)$, we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\rho_k\right| &\leq \frac{2}{n}\sum_{k=0}^{n-1}|\rho_k| \\ &= \frac{2}{n}\sum_{k=0}^{N}|\rho_k| + \frac{2}{n}\sum_{k=N+1}^{n-1}|\rho_k| \\ &\leq \frac{2}{n}\sum_{k=0}^{N}|\rho_k| + \frac{1}{2}\epsilon \\ &\leq \epsilon, \end{aligned} \tag{2.5.6}$$

where we choose an n large enough so that the first term in the next to last inequality above is also less than $\epsilon/2$. Thus, when $\rho_k \to 0$ as $k \to \infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\rho_k = 0,$$

which implies that, in Equation (2.5.4),

$$\lim_{n\to\infty}\mathrm{Var}(\overline{Z}) = 0. \tag{2.5.7}$$

Intuitively, these results simply say that if Z_t and Z_{t+k} sufficiently far apart are almost uncorrelated, then some useful new information can be continually added so that the time average will approach the ensemble average.

2.5.2 Sample Autocovariance Function

Similarly, with only a single realization, we employ the following estimators using the time average to estimate the autocovariance function γ_k, i.e.,

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z}) \tag{2.5.8}$$

or

$$\hat{\gamma}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z}). \tag{2.5.9}$$

Now,

$$\begin{aligned}\sum_{t=1}^{n-k}(Z_t-\bar{Z})(Z_{t+k}-\bar{Z}) &= \sum_{t=1}^{n-k}[(Z_t-\mu)-(\bar{Z}-\mu)][(Z_{t+k}-\mu)-(\bar{Z}-\mu)] \\ &= \sum_{t=1}^{n-k}(Z_t-\mu)(Z_{t+k}-\mu)-(\bar{Z}-\mu)\sum_{t=1}^{n-k}(Z_t-\mu) \\ &\quad -(\bar{Z}-\mu)\sum_{t=1}^{n-k}(Z_{t+k}-\mu)+(n-k)(\bar{Z}-\mu)^2 \\ &\approx \sum_{t=1}^{n-k}(Z_t-\mu)(Z_{t+k}-\mu)-(n-k)(\bar{Z}-\mu)^2, \qquad (2.5.10)\end{aligned}$$

where we approximate the terms $\sum_{t=1}^{n-k}(Z_t-\mu)$ and $\sum_{t=1}^{n-k}(Z_{t+k}-\mu)$ by $(n-k)(\bar{Z}-\mu)$. Hence,

$$E(\hat{\gamma}_k) \simeq \gamma_k - \frac{k}{n}\gamma_k - \left(\frac{n-k}{n}\right)\mathrm{Var}(\bar{Z}), \qquad (2.5.11)$$

$$E(\hat{\hat{\gamma}}_k) \simeq \gamma_k - \mathrm{Var}(\bar{Z}). \qquad (2.5.12)$$

It is clear that both of these estimators are biased. When we ignore the term $\mathrm{Var}(\bar{Z})$ representing the effect of estimating μ, $\hat{\hat{\gamma}}_k$ becomes unbiased but $\hat{\gamma}_k$ is still biased. In general, $\hat{\gamma}_k$ has a larger bias than $\hat{\hat{\gamma}}$, especially when k is large with respect to n. Hence, for a given n, often at most $n/4$ estimates are suggested to be calculated in time series analysis. If $\rho_k \to 0$ as $k \to \infty$ and hence the process is ergodic for the mean and $\lim_{n\to\infty}\mathrm{Var}(\bar{Z}) = 0$ as shown in (2.5.7), however, then both estimators $\hat{\gamma}_k$ and $\hat{\hat{\gamma}}_k$ are asymptotically unbiased. Some may argue that because both $\hat{\gamma}_k$ and $\hat{\hat{\gamma}}_k$ are biased, it would be more appropriate to compare their mean square errors. It can be shown that for certain types of processes $\hat{\gamma}_k$ has smaller mean square error than $\hat{\hat{\gamma}}_k$ (see, e.g., Parzen [1961b]). In addition, the estimate $\hat{\gamma}_k$ is always positive semidefinite like γ_k, but $\hat{\hat{\gamma}}_k$ is not necessarily so. As a result, we use $\hat{\gamma}_k$ in (2.5.8) as the sample autocovariance function to estimate the autocovariance function γ_k.

When the process $\{Z_t\}$ is Gaussian, Bartlett (1946) has shown the following approximate results:

$$\mathrm{Cov}(\hat{\gamma}_k, \hat{\gamma}_{k+j}) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\gamma_i\gamma_{i+j} + \gamma_{i+k+j}\gamma_{i-k}) \qquad (2.5.13)$$

and

$$\mathrm{Var}(\hat{\gamma}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\gamma_i^2 + \gamma_{i+k}\gamma_{i-k}). \qquad (2.5.14)$$

Similarly,

$$\text{Cov}(\hat{\hat{\gamma}}_k, \hat{\hat{\gamma}}_{k+j}) \simeq \frac{1}{n-k}\sum_{i=-\infty}^{\infty}(\gamma_i\gamma_{i+j} + \gamma_{i+k+j}\gamma_{i-k}), \tag{2.5.15}$$

and

$$\text{Var}(\hat{\hat{\gamma}}_k) \simeq \frac{1}{n-k}\sum_{i=-\infty}^{\infty}(\gamma_i^2 + \gamma_{i+k}\gamma_{i-k}). \tag{2.5.16}$$

Thus, the variance of $\hat{\hat{\gamma}}_k$ is larger than the variance of $\hat{\gamma}_k$. In fact, from (2.5.16) we see that the variance of $\hat{\hat{\gamma}}_k$, $\text{Var}(\hat{\hat{\gamma}}_k)$, can be substantial for large k resulting in an unstable and erratic estimate.

Next, we would like to ask when the process is ergodic for the autocovariance function so that in mean square we have

$$\lim_{n\to\infty}\hat{\gamma}_k = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) = \gamma_k. \tag{2.5.17}$$

A rigorous proof of the above statement is complicated. For our purpose, it suffices to note that for any given k, because the sample autocovariance $\hat{\gamma}_k$ is an asymptotically unbiased estimator of γ_k, hence a sufficient condition for $\hat{\gamma}_k$ to be mean square consistent and the process to be ergodic for the autocovariances is that the autocovariance is absolutely summable, i.e., $\sum_{-\infty}^{\infty}|\gamma_i| < \infty$ and hence $\lim_{n\to\infty}\text{Var}(\hat{\gamma}_k) = 0$. For relevant readings, interested readers are referred to Gnedenko (1962), Hannan (1970, p. 201), and Fuller (1996, p. 308), among others. Ergodicity is assumed to hold for the remainder of this book.

2.5.3 Sample Autocorrelation Function

For a given observed time series $Z_1, Z_2, \ldots, Z_n$, the sample ACF is defined as

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}, \qquad k = 0, 1, 2, \ldots \tag{2.5.18}$$

where $\bar{Z} = \sum_{t=1}^{n}Z_t/n$, the sample mean of the series. A plot of $\hat{\rho}_k$ versus k is sometimes called a sample correlogram.

For a stationary Gaussian process, Bartlett (1946) has shown that for $k > 0$ and $k + j > 0$,

$$\text{Cov}(\hat{\rho}_k, \hat{\rho}_{k+j}) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i\rho_{i+j} + \rho_{i+k+j}\rho_{i-k} - 2\rho_k\rho_i\rho_{i-k-j} - 2\rho_{k+j}\rho_i\rho_{i-k} + 2\rho_k\rho_{k+j}\rho_i^2). \tag{2.5.19}$$

For large n, $\hat{\rho}_k$ is approximately normally distributed with mean ρ_k and variance

$$\text{Var}(\hat{\rho}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2). \tag{2.5.20}$$

For processes in which $\rho_k = 0$ for $k > m$, Bartlett's approximation of (2.5.20) becomes

$$\text{Var}(\hat{\rho}_k) \simeq \frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_m^2). \tag{2.5.21}$$

In practice, $\rho_i (i = 1, 2, \ldots, m)$ are unknown and are replaced by their sample estimates $\hat{\rho}_i$, and we have the following large-lag standard error of $\hat{\rho}_k$:

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_m^2)}. \tag{2.5.22}$$

To test a white noise process, we use

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}. \tag{2.5.23}$$

EXAMPLE 2.4 To illustrate the computation of the sample ACF, consider the following ten values of a time series:

t	Z_t	Z_{t+1}	Z_{t+2}	Z_{t+3}	$\cdots;$	Z_{t-1}	Z_{t-2}
1	13	8	15	4			
2	8	15	4	4		13	
3	15	4	4	12		8	13
4	4	4	12	11		15	8
5	4	12	11	7		4	15
6	12	11	7	14		4	4
7	11	7	14	12		12	4
8	7	14	12			11	12
9	14	12				7	11
10	12					14	7

The sample mean of these ten values is $\bar{Z} = 10$. Thus,

$$\hat{\rho}_1 = \frac{(13-10)(8-10) + (8-10)(15-10) + \cdots + (7-10)(14-10) + (14-10)(12-10)}{(13-10)^2 + (8-10)^2 + \cdots + (14-10)^2 + (12-10)^2}$$

$$= \frac{-27}{144} = -.188$$

$$\hat{\rho}_2 = \frac{(13-10)(15-10)+(8-10)(4-10)+\cdots+(11-10)(14-10)+(7-10)(12-10)}{144}$$

$$= \frac{-29}{144} = -.201$$

$$\hat{\rho}_3 = \frac{(13-10)(4-10)+(8-10)(4-10)+\cdots+(12-10)(14-10)+(11-10)(12-10)}{144}$$

$$= \frac{26}{144} = .181$$

$$\vdots$$

and we note that

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Z_t-\bar{Z})(Z_{t+k}-\bar{Z})}{\sum_{t=1}^{n}(Z_t-\bar{Z})^2} = \frac{\sum_{t=k+1}^{n}(Z_t-\bar{Z})(Z_{t-k}-\bar{Z})}{\sum_{t=1}^{n}(Z_t-\bar{Z})^2} = \hat{\rho}_{-k}. \tag{2.5.24}$$

In other words, the sample ACF is also symmetric about the origin $k = 0$.

2.5.4 Sample Partial Autocorrelation Function

The sample PACF $\hat{\phi}_{kk}$ is obtained by substituting ρ_i by $\hat{\rho}_i$ in Equation (2.3.19). Instead of calculating the complicated determinants for large k in (2.3.19), a recursive method starting with $\hat{\phi}_{11} = \hat{\rho}_1$ for computing $\hat{\phi}_{kk}$ has been given by Durbin (1960) as follows:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k}\hat{\phi}_{kj}\hat{\rho}_j} \tag{2.5.25}$$

and

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}, \qquad j = 1,\ldots,k. \tag{2.5.26}$$

The method holds also for calculating the theoretical PACF ϕ_{kk}.

It was shown by Quenouille (1949) that on the hypothesis that the underlying process is a white noise sequence, the variance of $\hat{\phi}_{kk}$ can be approximated by

$$\mathrm{Var}(\hat{\phi}_{kk}) \simeq \frac{1}{n}. \tag{2.5.27}$$

Hence, $\pm 2/\sqrt{n}$ can be used as critical limits on ϕ_{kk} to test the hypothesis of a white noise process.

EXAMPLE 2.5 Using the data in Example 2.4, we have from (2.3.19), (2.5.25), and (2.5.26)

$$\hat{\phi}_{11} = \hat{\rho}_1 = -.188$$
$$\hat{\phi}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{-.201 - (-.188)^2}{1 - (-.188)^2} = -.245$$
$$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22} \cdot \hat{\phi}_{11} = (-.188) - (-.245)(-.188) = -.234.$$

Thus, by (2.5.25) we have

$$\hat{\phi}_{33} = \frac{\hat{\rho}_3 - \hat{\phi}_{21}\hat{\rho}_2 - \hat{\phi}_{22}\hat{\rho}_1}{1 - \hat{\phi}_{21}\hat{\rho}_1 - \hat{\phi}_{22}\hat{\rho}_2}$$
$$= \frac{.181 - (-.234)(-.201) - (-.245)(-.188)}{1 - (-.234)(-.188) - (-.245)(-.201)} = \frac{.088}{.907} = .097.$$

Other $\hat{\phi}_{kk}$ can be calculated similarly.

2.6 Moving Average and Autoregressive Representations of Time Series Processes

In time series analysis, there are two useful representations to express a time series process. One is to write a process Z_t as a linear combination of a sequence of uncorrelated random variables, i.e.,

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \tag{2.6.1}$$

where $\psi_0 = 1$, $\{a_t\}$ is a zero mean white noise process, and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Here and in the following an infinite sum of random variables is defined as the limit in quadratic mean (mean square) of the finite partial sums. Thus, Z_t in (2.6.1) is defined such that

$$E\left[\left(\dot{Z}_t - \sum_{j=0}^{n} \psi_j a_{t-j}\right)^2\right] \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty,$$

where $\dot{Z}_t = Z_t - \mu$. By introducing the backshift operator $B^j x_t = x_{t-j}$, we can write (2.6.1) in the compact form

$$\dot{Z}_t = \psi(B) a_t, \tag{2.6.2}$$

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

It is easy to show that for the process in (2.6.1)

$$E(Z_t) = \mu, \tag{2.6.3}$$

$$\text{Var}(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2, \tag{2.6.4}$$

and

$$E(a_t Z_{t-j}) = \begin{cases} \sigma_a^2, & \text{for } j = 0, \\ 0, & \text{for } j > 0. \end{cases} \tag{2.6.5}$$

Hence,

$$\begin{aligned} \gamma_k &= E(\dot{Z}_t \dot{Z}_{t+k}) \\ &= E\left(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t+k-j} \right) \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \end{aligned} \tag{2.6.6}$$

and

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}. \tag{2.6.7}$$

Clearly, the autocovariance and autocorrelation functions in (2.6.6) and (2.6.7) are functions of the time difference k only. Because they involve infinite sums, however, to be stationary we have to show that γ_k is finite for each k. Now,

$$|\gamma_k| = |E(\dot{Z}_t \dot{Z}_{t+k})| \leq [\text{Var}(Z_t)\, \text{Var}(Z_{t+k})]^{1/2} = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2.$$

Hence, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is a required condition for the process in (2.6.1) to be stationary.

The form in (2.6.1) is called a moving average (MA) representation of a process. Wold (1938) proved that a stationary process that is purely nondeterministic (i.e., a process contains no deterministic component that can be forecast or predicted exactly from its own past) can always be expressed in the form of (2.6.1). Hence, the representation is also known as Wold's representation in the literature, and any process that can be represented in this form is called a nondeterministic process. Sometimes the term *linear process* is also used to refer to the process in (2.6.1).

For a given sequence of autocovariances γ_k, $k = 0 \pm 1, \pm 2, \ldots$, the autocovariance generating function is defined as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k, \tag{2.6.8}$$

where the variance of the process, γ_0, is the coefficient of B^0 and the autocovariance of lag k, γ_k, is the coefficient of both B^k and B^{-k}. Using (2.6.6) and stationarity, we can write (2.6.8) as

$$\begin{aligned}\gamma(B) &= \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} \\ &= \sigma_a^2 \psi(B)\psi(B^{-1}),\end{aligned} \tag{2.6.9}$$

where we let $j = i + k$ and note that $\psi_j = 0$ for $j < 0$. This method is a convenient way of calculating the autocovariances for some linear processes. The corresponding autocorrelation generating function will be

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}. \tag{2.6.10}$$

Another useful form is to write a process Z_t in an autoregressive (AR) representation, in which we regress the value of Z at time t on its own past values plus a random shock, i.e.,

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \cdots + a_t$$

or, equivalently,

$$\pi(B)\dot{Z}_t = a_t, \tag{2.6.11}$$

where $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, and $1 + \sum_{j=1}^{\infty} |\pi_j| < \infty$. The autoregressive representation is useful in understanding the mechanism of forecasting. Box and Jenkins (1976) call a process invertible if it can be written in this form. They argue that in forecasting, a noninvertible process is meaningless. It is easily seen that not every stationary process is invertible. For a linear process $Z_t = \psi(B)a_t$ to be invertible so that it can be written in terms of the AR representation, the roots of $\psi(B) = 0$ as a function of B must lie outside of the unit circle. That is, if β is a root of $\psi(B)$, then $|\beta| > 1$, where $|\ |$ is the standard Euclidian metric. When β is a real number, $|\beta|$ is equal to the absolute value of β and when β is a complex number, $\beta = c + id$, then $|\beta| = \sqrt{c^2 + d^2}$.

It should be noted that an invertible process is not necessarily stationary. By Wold's result, it follows that for the process presented in (2.6.11) to be stationary, the process must be able to be rewritten in a MA representation, i.e.,

$$\dot{Z}_t = \frac{1}{\pi(B)} a_t = \psi(B) a_t, \tag{2.6.12}$$

such that the condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is satisfied. To achieve that, the required condition is that the roots of $\pi(B) = 0$ all lie outside the unit circle. That is, if δ is a root of $\pi(B)$, we have $|\delta| > 1$.

Although the autoregressive and moving average representations are useful, they are not the model forms we use at the beginning stage of model building because they contain an infinite number of parameters that are impossible to estimate from a finite number of available observations. Instead, in modeling a phenomenon, we construct models with only a finite number of parameters.

In the autoregressive representation of a process, if only a finite number of π weights are nonzero, i.e., $\pi_1 = \phi_1, \pi_2 = \phi_2, \ldots, \pi_p = \phi_p$ and $\pi_k = 0$ for $k > p$, then the resulting process is said to be an autoregressive process of order p. This process is written as

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \cdots - \phi_p \dot{Z}_{t-p} = a_t. \tag{2.6.13}$$

Similarly, in the moving average representation, if only a finite number of ψ weights are nonzero, i.e., $\psi_1 = -\theta_1, \psi_2 = -\theta_2, \ldots, \psi_q = -\theta_q$ and $\psi_k = 0$ for $k > q$, then the resulting process is said to be a moving average process of order q, which is given by

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \tag{2.6.14}$$

If we restrict ourselves to the above finite-order autoregressive models in (2.6.13) and moving average models in (2.6.14), however, the number of parameters may still be prohibitively large. A natural alternative is the mixed autoregressive moving average model

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \cdots - \phi_p \dot{Z}_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \tag{2.6.15}$$

For a fixed number of observations, the more parameters in a model, the less efficient is the estimation of the parameters. Other things being equal, in general, we choose a simpler model to describe the phenomenon. This modeling criteria is the principle of parsimony in model building recommended by Tukey (1967) and Box and Jenkins (1976). In the following chapters, we discuss some very useful parsimonious time series models and their properties.

2.7 Linear Difference Equations

Linear difference equations play an important role in the time series models discussed in this book. In fact, all models of finite number of parameters mentioned above relate an output Z_t to an input a_t in terms of linear difference equations. Thus, they are sometimes referred to as

linear difference equation models. The properties of these models often depend on the characteristics of the roots of these difference equations. To better understand these models, we give a brief introduction to linear difference equations, especially to the method of solving these equations. A general nth-order linear difference equation with constant coefficients is given by

$$C_0Z_t + C_1Z_{t-1} + C_2Z_{t-2} + \cdots + C_nZ_{t-n} = e_t, \tag{2.7.1}$$

where C_i, $i = 0, 1, \ldots, n$, are constants. Without loss of generality, we can set $C_0 = 1$. The function e_t in (2.7.1) is called a forcing function. Equation (2.7.1) is said to be nonhomogeneous (or complete) if $e_t \neq 0$, and homogeneous if $e_t = 0$.

Using the backshift operator and letting $C(B) = (1 + C_1B + C_2B^2 + \ldots + C_nB^n)$, we can rewrite (2.7.1) as

$$C(B)Z_t = e_t. \tag{2.7.2}$$

As a function of B, $C(B) = 0$ is called the auxiliary equation associated with the given linear difference equation. The solution of linear difference equations rests primarily on the following lemmas, which can be easily proved using the definition of a solution and the property of linearity.

Lemma 2.7.1. If $Z_t^{(1)}$ and $Z_t^{(2)}$ are solutions of the homogeneous equation, then $b_1Z_t^{(1)} + b_2Z_t^{(2)}$ is also a solution for any arbitrary constants b_1 and b_2. ∎

Lemma 2.7.2. If $Z_t^{(H)}$ is a solution to the homogeneous equation and $Z_t^{(P)}$ is a particular solution of the nonhomogeneous equation, then $Z_t^{(H)} + Z_t^{(P)}$ is the general solution of the complete equation. ∎

The particular solution of a nonhomogeneous difference equation depends on the form of the forcing function. The solution of a homogeneous difference equation depends on the roots of the associated auxiliary equation. As will be seen, most of the time series models considered in this book are of the difference equation type and the behavior of the time series generated by such models is governed by the nature of the roots of the associated auxiliary equation. In the remainder of this section, we concentrate on the solution of a general homogeneous linear difference equation. When B is used as an operator, it operates on time index t.

Lemma 2.7.3. Let $(1 - B)^m Z_t = 0$. Then a solution is given by $Z_t = bt^j$, where b is any constant and j is a nonnegative integer less than m.

Proof For $m = 1$, $Z_t = bt^0 = b$. Clearly, we have $(1-B)Z_t = (1 - B)b = b - b = 0$. Now assume that $(1 - B)^{m-1} Z_t = 0$, where $Z_t = bt^j, j < (m - 1)$. Then for $Z_t = bt^j, j < m$,

$$\begin{aligned}(1 - B)^m Z_t &= (1 - B)^{m-1}(1 - B)bt^j \\ &= (1 - B)^{m-1}b\{t^j - (t - 1)^j\} \\ &= (1 - B)^{m-1}\left\{-b\sum_{i=0}^{j-1}\binom{j}{i}(-1)^{j-i}t^i\right\}.\end{aligned}$$

Now each term in the last expression involving t contains only integer powers less than $(m - 1)$. Thus, by our induction hypothesis, each term is reduced to zero by the operator $(1 - B)^{m-1}$. The lemma is proved. ∎

It follows from Lemmas 2.7.1 and 2.7.3 that $(1 - B)^m\left[\sum_{j=0}^{m-1} b_j t^j\right] = 0$ for any constants $b_0, b_1, \ldots$ and b_{m-1}.

Lemma 2.7.4 Let $(1-RB)^m Z_t = 0$. Then a solution is given by $Z_t = t^j R^t$, where j is any nonnegative integer less than m, and the general solution is given by $Z_t = \left(\sum_{j=0}^{m-1} b_j t^j\right)R^t$, where b_j is a constant.

Proof First, we note that

$$\begin{aligned}(1 - RB)Z_t &= (1 - RB)t^jR^t = t^jR^t - R(t - 1)^jR^{t-1} \\ &= [(1 - B)t^j]R^t.\end{aligned}$$

Repeated application of the above result gives

$$(1 - RB)^mZ_t = (1 - RB)^mt^jR^t = [(1 - B)^mt^j]R^t,$$

which equals zero because $(1 - B)^m t^j = 0$ by Lemma 2.7.3. The result then follows immediately from Lemma 2.7.1. ∎

Finally, we have the following main result.

Theorem 2.7.1. Let $C(B)Z_t = 0$ be a given homogeneous linear difference equation where $C(B) = 1 + C_1B + C_2B^2 + \cdots + C_nB^n$. If $C(B) = \prod_{i=1}^{N}(1 - R_iB)^{m_i}$, where $\sum_{i=1}^{N} m_i = n$ and $B_i = R_i^{-1}$ $(i = 1, 2, \ldots, N)$ are roots of multiplicity m_i of $C(B) = 0$, then $Z_t = \sum_{i=1}^{N}\sum_{j=0}^{m_i-1} b_{ij} t^jR_i^t$. In particular, if $m_i = 1$ for all i and R_i^{-1} $(i = 1, 2, \ldots, n)$ are all distinct, we have $Z_t = \sum_{i=1}^{n} b_iR_i^t$.

Proof The result follows immediately from Lemmas 2.7.1, 2.7.3, and 2.7.4. ∎

Note that for a real-valued linear difference equation, a complex root of $C(B) = 0$ must appear in pairs. That is, if $(c + di)$ is a root, then its complex conjugate $(c + di)^* = (c - di)$ is also a root. A general complex number can always be written in the polar form, i.e.,

$$(c \pm di) = \alpha(\cos \phi \pm i \sin \phi), \tag{2.7.3}$$

where

$$\alpha = (c^2 + d^2)^{1/2} \tag{2.7.4}$$

and

$$\phi = \tan^{-1}(d/c). \tag{2.7.5}$$

Because $(c \pm di)^t = \alpha^t(\cos \phi t \pm i \sin \phi t)$, it follows that for each pair of complex roots of multiplicity m, the solution of the homogeneous difference equation should contain the sequences $\alpha^t \cos \phi t$, $\alpha^t \sin \phi t$; $t\alpha^t \cos \phi t$, $t\alpha^t \sin \phi\, t$; . . . ; and $t^{m-1}\alpha^t \cos \phi t$, $t^{m-1}\alpha^t \sin \phi t$.

For an illustration, consider the second-order difference equation with the following auxiliary equation

$$(1 - C_1B - C_2B^2) = (1 - R_1B)(1 - R_2B) = 0$$

with

$$R_1 = c + di = \alpha(\cos \phi + i \sin \phi) \qquad \text{and} \qquad R_2 = c - di = \alpha(\cos \phi - i \sin \phi).$$

By Theorem 2.7.1, we have

$$Z_t = e_1(c + di)^t + e_2(c - di)^t = e_1[\alpha(\cos \phi + i \sin \phi)]^t + e_2[\alpha(\cos \phi - i \sin \phi)]^t,$$

where e_1 and e_2 are any (complex) constants. For a real-valued process, we claim that we can always write

$$Z_t = b_1\alpha^t \cos \phi t + b_2\alpha^t \sin \phi t,$$

where b_1 and b_2 are real constants. This result follows because we can choose e_2 to be the complex conjugate of e_1. Thus, if $e_1 = x + iy$, then $e_2 = x - iy$, and

$$\begin{aligned} Z_t &= (e_1 + e_2)(\alpha^t \cos \phi t) + (e_1 - e_2)i(\alpha^t \sin \phi t) \\ &= b_1\alpha^t \cos \phi t + b_2\alpha^t \sin \phi t, \end{aligned}$$

where b_1 and b_2 can be easily seen as real.

EXAMPLE 2.6 Let $Z_t - 2Z_{t-1} + Z_{t-2} = 0$. Find the closed form solution for Z_t.

The auxiliary equation is given by

$$C(B) = (1 - 2B + B^2) = (1 - B)^2 = 0.$$

$R^{-1} = 1$ is a root of multiplicity 2. Hence, by Theorem 2.7.1,

$$Z_t = (b_1 + b_2t)1^t = b_1 + b_2t.$$

EXAMPLE 2.7 Find the solution for $Z_t - 2Z_{t-1} + 1.5Z_{t-2} - .5Z_{t-3} = 0$.

The auxiliary equation is given by

$$C(B) = (1 - 2B + 1.5B^2 - .5B^3) = (1 - B + .5B^2)(1 - B) = 0.$$

Hence, $B_1 = R_1^{-1} = 1$ is a root of multiplicity one; and

$$B_2 = R_2^{-1} = \frac{1 + \sqrt{1 - 4(.5)}}{2(.5)} = 1 + i$$

and $B_3 = R_3^{-1} = 1 - i$ are a pair of complex roots of multiplicity one.

Now $R_1 = 1$ and $R_2 = (1 - i)/2$ and $R_3 = (1 + i)/2$. To express R_2 and R_3 in the polar form, we have from (2.7.4) and (2.7.5)

$$\alpha = (c^2 + d^2)^{1/2} = \sqrt{\frac{1}{2}}$$

$$\phi = \tan^{-1}\left(\frac{d}{c}\right) = \tan^{-1}(1) = \frac{\pi}{4}.$$

Hence, by the remarks following Theorem 2.7.1, we have

$$Z_t = b_1 + b_2\left(\sqrt{\frac{1}{2}}\right)^t \cos\left(\frac{\pi}{4}t\right) + b_3\left(\sqrt{\frac{1}{2}}\right)^t \sin\left(\frac{\pi}{4}t\right).$$

Remark The R_i used in the solution of $C(B) = 0$ in Theorem 2.7.1 is the inverse of the root B_i, i.e., $R_i = B_i^{-1}$. Thus, as shown in Example 2.7, to find R_i, we first find the root B_i of $C(B) = 0$ and then calculate its inverse. This cumbersome procedure can be avoided by noting that if we let $R = B^{-1}$ and multiply the equation

$$1 - C_1B - C_2B^2 - \cdots - C_nB^n = 0, \tag{2.7.6}$$

by $R^n = (B^{-1})^n$, we get

$$R^n - C_1R^{n-1} - C_2R^{n-2} - \cdots - C_n = 0. \tag{2.7.7}$$

It can be easily seen that B_i is a root of (2.7.6) if and only if R_i is a root of (2.7.7). Thus, sometimes we can compute $R_i = B_i^{-1}$ needed in Theorem 2.7.1 more easily from solving (2.7.7). The reader can easily verify that each R_i obtained in Example 2.7 is indeed a solution of $R^3 - 2R^2 + 1.5R - .5 = (R^2 - R + .5)(R - 1) = 0$.

EXAMPLE 2.8 Solve $(1 - \phi B)(1 - B)^2Z_t = 0$.

Because $R_1 = \phi$, and $R_2 = 1$ of multiplicity 2, we have from Theorem 2.7.1, $Z_t = b_1\phi^t + b_2 + b_3t$.

EXERCISES

2.1 Let Z_t, where t is even, be a sequence of independent random variables defined as $Z_t = +1$ or -1 with equal probability of 1/2, and $Z_t = Z_{t-1}$ if t is odd, where t is an integer.

(a) Is the process first order stationary in distribution?

(b) Is it second order stationary in distribution?

2.2 Let $Z_t = U \sin(2\pi t) + V \cos(2\pi t)$, where U and V are independent random variables, each with mean 0 and variance 1.

(a) Is Z_t strictly stationary?

(b) Is Z_t covariance stationary?

2.3 Prove or disprove that the following process is covariance stationary:

(a) $Z_t = A \sin(2\pi t + \theta)$, where A is a constant, and θ is a random variable that is uniformly distributed on $[0, 2\pi]$

(b) $Z_t = A \sin(2\pi t + \theta)$, where A is a random variable with zero mean and unit variance, and θ is a constant.

(c) $Z_t = (-1)^t A$, where A is a random variable with zero mean and unit variance.

2.4 Is the following a valid autocorrelation function for a real-valued covariance stationary process? Why?

$$\rho_k = \begin{cases} 1, & \text{if } k = 0, \\ \phi, & \frac{1}{2} < |\phi| < 1, \quad \text{if } |k| = 1, \\ 0, & \text{if } |k| \geq 2. \end{cases}$$

2.5 Verify the following properties for the autocorrelation function of a stationary process:

(a) $\rho_0 = 1$,

(b) $|\rho_k| \leq 1$,

(c) $\rho_k = \rho_{-k}$.

2.6 It is important to note that not every sequence satisfying the properties specified in Exercise 2.5 is an ACF of a process. Show that the following function satisfies the properties stated in Exercise 2.5, but is not an ACF for any stationary process:

$$\rho_k = \begin{cases} 1, & k = 0, \\ .8, & k = \pm 1, \\ .1, & k = \pm 2, \\ 0, & \text{otherwise.} \end{cases}$$

2.7 Given the time series 53, 43, 66, 48, 52, 42, 44, 56, 44, 58, 41, 54, 51, 56, 38, 56, 49, 52, 32, 52, 59, 34, 57, 39, 60, 40, 52, 44, 65, 43:

(a) Plot the series.

(b) Can you guess an approximate value for the first lag autocorrelation coefficient ρ_1 based on the plot of the series?

(c) Plot Z_t against Z_{t+1}, and try again to guess the value of ρ_1.

(d) Calculate and plot the sample ACF, $\hat{\rho}_k$, for $k = 0, 1, 2, 3, 4, 5$.
(e) Calculate and plot the sample PACF, $\hat{\phi}_{kk}$, for $k = 0, 1, 2, 3, 4, 5$.

2.8 Show that the estimate $\hat{\gamma}_k$ is always positive semi-definite, but $\hat{\hat{\gamma}}_k$ is not necessarily so.

2.9 Consider a stationary series with theoretical autocorrelation function

$$\rho_k = \phi^{|k|}, |\phi| < 1, \quad k = 1, 2, 3, \ldots.$$

Find the variance of $\hat{\rho}_k$ using Bartlett's approximation.

2.10 Let $Z_t = \mu + \sum_{j=0}^{\infty}\psi_j a_{t-j}$ or, equivalently, $Z_t = \mu + \psi(B)a_t$, where $\psi(B) = \sum_{j=0}^{\infty}\psi_j B^j$, $\psi_0 = 1$, $\sum_{j=0}^{\infty}|\psi_j| < \infty$, and the a_t are white noise with mean 0 and variance σ_a^2. Prove that $\sum_{j=-\infty}^{\infty}|\gamma_j| < \infty$, where γ_j is the jth autocovariance of the process Z_t.

2.11 Prove that $n\,\mathrm{Var}(\bar{z}) \to \sum_{k=-\infty}^{\infty}\gamma_k$ as $n \to \infty$, if $\sum_{k=-\infty}^{\infty}|\gamma_k| < \infty$.

2.12 Find a closed form solution for the following difference equations:
(a) $Z_t - 1.1Z_{t-1} + .3Z_{t-2} = 0$,
(b) $Z_t - Z_{t-1} + Z_{t-2} - Z_{t-3} = 0$,
(c) $Z_t - 1.8Z_{t-1} + .81Z_{t-2} = 0$.