Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 1

Introduction

The analysis of time series applies to many fields. In economics the recorded history of the economy is often in the form of time series. Economic behavior is quantified in such time series as the consumer price index, unemployment, gross national product, population, and production. The natural sciences also furnish many examples of time series. The water level in a lake, the air temperature, the yields of corn in Iowa, and the size of natural populations are all collected over time. Growth models that arise naturally in biology and in the social sciences represent an entire field in themselves.

The mathematical theory of time series has been an area of considerable activity in recent years. Applications in the physical sciences such as the development of designs for airplanes and rockets, the improvement of radar and other electronic devices, and the investigation of certain production and chemcial processes have resulted in considerable interest in time series analysis. This recent work should not disguise the fact that the analysis of time series is one of the oldest activities of scientific man.

A successful application of statistial methods to the real world requires a melding of statistical theory and knowledge of the material under study. We shall confine ourselves to the statistical treatment of moderately well-behaved time series, but we shall illustrate some techniques with real data.

1.1. PROBABILITY SPACES

When investigating outcomes of a game, an experiment, or some natural phenomenon, it is useful to have a representation for all possible outcomes. The individual outcomes, denoted by ω, are called *elementary events*. The set of all possible elementary events is called the *sure* event and is denoted by Ω. An example is the tossing of a die, where we could take $\Omega = \{\text{one spot shows, two spots show, } \ldots, \text{ six spots show}\}$ or, more simply, $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Let A be a subset of Ω, and let $\mathcal{A}$ be a collection of such subsets. If we observe the outcome ω and ω is in A, we say that A has occurred. Intuitively, it is possible

to specify $P(A)$, the probability that (or expected long-run frequency with which) A will occur. It is reasonable to require that the function $P(A)$ satisfy:

AXIOM 1. $P(A) \geq 0$ for every A in $\mathscr{A}$.

AXIOM 2. $P(\Omega) = 1$.

AXIOM 3. If $A_1, A_2, \ldots$ is a countable sequence from $\mathscr{A}$ and $A_i \cap A_j$ is the null set for all $i \neq j$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Using our die tossing example, if the die is fair we would take $P(A) = \frac{1}{6}$ [the number of elementary events ω in A]. Thus $P(\{1, 3, 5\}) = \frac{1}{6} \times 3 = \frac{1}{2}$. It may be verified that Axioms 1 to 3 are satisfied for $\mathscr{A}$ equal to $\mathscr{P}(\Omega)$, the collection of all possible subsets of Ω.

Unfortunately, for technical mathematical reasons, it is not always possible to define $P(A)$ for all A in $\mathscr{P}(\Omega)$ and also to satisfy Axiom 3. To eliminate this difficulty, the class of subsets $\mathscr{A}$ of Ω on which P is defined is restricted. The collection $\mathscr{A}$ is required to satisfy:

1. If A is in $\mathscr{A}$, then the complement A^c is also in $\mathscr{A}$.
2. If $A_1, A_2, \ldots$ is a countable sequence from $\mathscr{A}$, then $\bigcup_{i=1}^{\infty} A_i$ is in $\mathscr{A}$.
3. The null set is in $\mathscr{A}$.

A nonempty collection $\mathscr{A}$ of subsets of Ω that satisfies conditions 1 to 3 is said to be a *sigma-algebra* or *sigma-field*.

We are now in a position to give a formal definition of a probability space. A *probability space*, represented by $(\Omega, \mathscr{A}, P)$, is the sure event Ω together with a sigma-algebra $\mathscr{A}$ of subsets of Ω and a function $P(A)$ defined on $\mathscr{A}$ that satisfies Axioms 1 to 3.

For our purposes it is unnecessary to consider the subject of probability spaces in detail. In simple situations such as tossing a die, Ω is easy to enumerate, and P satisfies Axioms 1 to 3 for $\mathscr{A}$ equal to the set of all subsets of Ω.

Although it is conceptually possible to enumerate all possible outcomes of an experiment, it may be a practical impossibility to do so, and for most purposes it is unnecessary to do so. It is usually enough to record the outcome by some function that assumes values on the real line. That is, we assign to each outcome ω a real number $X(\omega)$, and if ω is observed, we record $X(\omega)$. In our die tossing example we could take $X(\omega) = 1$ if the player wins and -1 if the house wins.

Formally, a *random variable* X is a real valued function defined on Ω such that the set $\{\omega : X(\omega) \leq x\}$ is a member of $\mathscr{A}$ for every real number x. The function $F_X(x) = P(\{\omega : X(\omega) \leq x\})$ is called the *distribution function* of the random variable X.

The reader who wishes to explore further the subjects of probability spaces, random variables, and distribution functions for stochastic processes may read Tucker (1967, pp. 1–33). The preceding brief introduction will suffice for our purposes.

1.2. TIME SERIES

Let $(\Omega, \mathscr{A}, P)$ be a probability space, and let T be an index set. A real valued *time series* (or *stochastic process*) is a real valued function $X(t, \omega)$ defined on $T \times \Omega$ such that for each fixed t, $X(t, \omega)$ is a random variable on $(\Omega, \mathscr{A}, P)$. The function $X(t, \omega)$ is often written $X_t(\omega)$ or X_t, and a time series can be considered as a collection $\{X_t : t \in T\}$ of random variables.

For fixed ω, $X(t, \omega)$ is a real valued function of t. This function of t is called a *realization* or a *sample function*. If we look at a plot of some recorded time series such as the gross national product, it is important to realize that conceptually we are looking at a plot of $X(t, \omega)$ with ω fixed. The collection of all possible realizations is called the *ensemble* of functions or the ensemble of realizations.

If the index set contains exactly one element, the stochastic process is a single random variable and we have defined the distribution function of the process. For stochastic processes with more than one random variable we need to consider the *joint distribution function*. The joint distribution function of a finite set of random variables $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ from the collection $\{X_t : t \in T\}$ is defined by

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = P\{\omega : X(t_1, \omega) \leq x_{t_1}, \ldots, X(t_n, \omega) \leq x_{t_n}\}. \tag{1.2.1}$$

A time series is called *strictly stationary* if

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = F_{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}),$$

where the equality must hold for all possible (nonempty finite distinct) sets of indices $t_1, t_2, \ldots, t_n$ and $t_1 + h, t_2 + h, \ldots, t_n + h$ in the index set and all $(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$ in the range of the random variable X_t. Note that the indices $t_1, t_2, \ldots, t_n$ are not necessarily consecutive. If a time series is strictly stationary, we see that the distribution function of the random variable is the same at every point in the index set. Furthermore, the joint distribution depends only on the distance between the elements in the index set, and not on their actual values. Naturally this does not mean a particular realization will appear the same as another realization.

If $\{X_t : t \in T\}$ is a strictly stationary time series with $E\{|X_t|\} < \infty$, then the expected value of X_t is a constant for all t, since the distribution function is the same for all t. Likewise, if $E\{X_t^2\} < \infty$, then the variance of X_t is a constant for all t.

A time series is defined completely in a probabilistic sense if one knows the cumulative distribution (1.2.1) for any finite set of random variables $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$. However, in most applications, the form of the distribution function is not known. A great deal can be accomplished, however, by dealing

only with the first two moments of the time series. In line with this approach we define a time series to be *weakly stationary* if:

1. The expected value of X_t is a constant for all t.
2. The covariance matrix of $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is the same as the covariance matrix of $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})$ for all nonempty finite sets of indices $(t_1, t_2, \ldots, t_n)$, and all h such that $t_1, t_2, \ldots, t_n$, $t_1 + h$, $t_2 + h, \ldots, t_n + h$ are contained in the index set.

As before $t_1, t_2, \ldots, t_n$ are not necessarily consecutive members of the index set. Also, since the expected value of X_t is a constant, it may conveniently be taken as 0. The covariance matrix, by definition, is a function only of the distance between observations. That is, the covariance of X_{t+h} and X_t depends only on the distance, h, and we may write

$$\text{Cov}\{X_t, X_{t+h}\} = E\{X_t X_{t+h}\} = \gamma(h),$$

where $E\{X_t\}$ has been taken to be zero. The function $\gamma(h)$ is called the *autocovariance* of X_t. When there is no danger of confusion, we shall abbreviate the expression to *covariance*.

The terms *stationary in the wide sense*, *covariance stationary*, *second order stationary*, and *stationary* are also used to describe a weakly stationary time series. It follows from the definitions that a strictly stationary process with the first two moments finite is also weakly stationary. However, a strictly stationary time series may not possess finite moments and hence may not be covariance stationary.

Many time series as they occur in practice are not stationary. For example, the economies of many countries are developing or growing. Therefore, the typical economic indicators will be showing a "trend" through time. This trend may be in either the mean, the variance, or both. Such nonstationary time series are sometimes called evolutionary. A good portion of the practical analysis of time series is connected with the transformation of an evolving time series into a stationary time series. In later sections we shall consider several of the procedures used in this connection. Many of these techniques will be familiar to the reader because they are closely related to least squares and regression.

1.3. EXAMPLES OF STOCHASTIC PROCESSES

Example 1.3.1. Let the index set be $T = \{1, 2\}$, and let the space of outcomes be the possible outcomes associated with tossing two dice, one at "time" $t = 1$ and one at time $t = 2$. Then

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

Define

$$X(t, \omega) = t + [\text{value on die } t]^2 .$$

Therefore, for a particular ω, say $\omega_3 = (1, 3)$, the realization or sample function would be $(1 + 1, 2 + 9) = (2, 11)$. In this case, both Ω and T are finite, and we are able to determine that there are 36 possible realizations. ▲▲

Example 1.3.2. One of the most important time series is a sequence of uncorrelated random variables, say $\{e_t : t \in (0, \pm 1, \pm 2, \ldots)\}$, each with zero mean and finite variance, $\sigma^2 > 0$. We say that e_t is a sequence of uncorrelated $(0, \sigma^2)$ random variables.

This time series is sometimes called *white noise*. Note that the index set is the set of all integers. The set Ω is determined by the range of the random variables. Let us assume that the e_t are normally distributed and therefore have range $(-\infty, \infty)$. Then $\omega \in \Omega$ is a real valued infinite-dimensional vector with an element associated with each integer. The covariance function of $\{e_t\}$ is given by

$$\gamma_e(h) = \begin{cases} \sigma^2, & h = 0, \\ 0 & \text{otherwise}. \end{cases}$$

Because of the importance of this time series, we shall reserve the symbol e_t for a sequence of uncorrelated random variables with zero mean and positive finite variance. On occasion we shall further assume that the variables are independent and perhaps of a specified distribution. These additional assumptions will be stated when used.

In a similar manner the most commonly used index set will be the set of all integers. If the index set is not stated, the reader may assume that it is the set of all integers. ▲▲

Example 1.3.3. Consider the time series

$$X_t = \hat{\beta}_1 + \hat{\beta}_2 t, \qquad t \in [0, 1],$$

where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ is distributed as a bivariate normal random varible with mean $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and covariance matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Any realization of this process yields a function continuous on the interval [0, 1], and the process therefore is called *continuous*. Such a process might represent the outcome of an experiment to estimate the linear response to an input variable measured on the interval [0, 1]. The $\hat{\beta}$'s are then the estimated regression coefficients. The set Ω is the two-dimensional space of real numbers. Each

experiment conducted to obtain a set of estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ constitutes a realization of the process.

The mean function of X_t is given by

$$E\{X_t\} = E\{\hat{\beta}_1 + \hat{\beta}_2 t\} = \beta_1 + \beta_2 t\,,$$

and the covariance function by

$$\begin{aligned} &E\{[X_t - E\{X_t\}][X_{t+h} - E\{X_{t+h}\}]\} \\ &\quad = E\{[(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)t][(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)(t+h)]\} \\ &\quad = \sigma_{11} + \sigma_{12}(t+h) + \sigma_{12}t + \sigma_{22}t(t+h)\,, \qquad t, t+h \in [0,1]\,. \end{aligned}$$

It is clear that this process is not stationary. ▲▲

Example 1.3.4. For the reader familiar only with the study of multivariate statistics, the idea of a random variable that is both continuous and random in time may require a moment of reflection. On the other hand, such processes occur in nature. For example, the water level of a lake or river can be plotted as a continuous function of time. Futhermore, such a plot might appear so smooth as to support the idea of a derivative or instantaneous rate of change in level.

Consider such a process, $\{X_t : t \in (-\infty, \infty)\}$, and let the covariance of the process be given by

$$\gamma(h) = E\{(X_t - \mu)(X_{t+h} - \mu)\} = Ae^{-\alpha h^2}\,, \qquad \alpha > 0\,.$$

Thus, for example, if time is measured in hours and the river level at 7:00 A.M. is reported daily, then the covariance between daily reports is $Ae^{-576\alpha}$. Likewise, the covariance between the change from 7:00 A.M. to 8:00 A.M. and the change from 8:00 A.M. to 9:00 A.M. is given by

$$E\{(X_{t+1} - X_t)(X_{t+2} - X_{t+1})\} = -A[1 - 2e^{-\alpha} + e^{-4\alpha}]\,.$$

Note that the variance of the change from t to $t+h$ is

$$\mathrm{Var}\{X_{t+h} - X_t\} = 2A[1 - e^{-\alpha h^2}]\,,$$

and

$$\lim_{h \to 0} \mathrm{Var}\{X_{t+h} - X_t\} = 0\,.$$

A process with this property is called *mean square continuous.*

We might define the rate of change per unit of time by

$$R_t(h) = \frac{X_{t+h} - X_t}{h}\,.$$

For a fixed $h > 0$, $R_t(h)$ is a well-defined random variable and

$$\text{Var}\{R_t(h)\} = \frac{2A[1 - e^{-\alpha h^2}]}{h^2}.$$

Furthermore, by L'Hospital's rule,

$$\lim_{h \to 0} \text{Var}\{R_t(h)\} = \lim_{h \to 0} \frac{2A[2\alpha h e^{-\alpha h^2}]}{2h}$$
$$= 2A\alpha.$$

Stochastic processes for which this limit exists are called *mean square differentiable*. ▲▲

1.4. PROPERTIES OF THE AUTOCOVARIANCE AND AUTOCORRELATION FUNCTIONS

To compare the basic properties of time series, it is often useful to have a function that is not influenced by the units of measurement. To this end, we define the *autocorrelation function* of a stationary time series by

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

Thus the autocorrelation function is the autocovariance function normalized to be one at $h = 0$. As with the autocovariance function, we shall abbreviate the expression to correlation function when no confusion will result.

The autocovariance and autocorrelation functions of stationary time series possess several distinguishing characteristics. A function $f(x)$ defined for $x \in \chi$ is said to be *positive semidefinite* if it satisfies

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k f(t_k - t_j) \geq 0 \tag{1.4.1}$$

for any set of real numbers $(a_1, a_2, \ldots, a_n)$ and for any $(t_1, t_2, \ldots, t_n)$ such that $t_i - t_j$ is in χ for all (i, j). Some authors reserve the term positive semidefinite for a function that is zero for at least one nonzero vector $(a_1, a_2, \ldots, a_n)$ and use *nonnegative definite* to describe functions satisfying (1.4.1). We shall use both terms for functions satisfying (1.4.1).

Theorem 1.4.1. The covariance function of the stationary time series $\{T_t : t \in T\}$ is nonnegative definite.

Proof. Without loss of generality, let $E\{X_t\} = 0$. Let $(t_1, t_2, \ldots, t_n) \in T$, let $(a_1, a_2, \ldots, a_n)$ be any set of real numbers, and let $\gamma(t_k - t_j)$ be the covariance

between X_{t_j} and X_{t_k}. We know that the variance of a random variable, when it is defined, is nonnegative. Therefore,

$$0 \leq \operatorname{Var}\left\{\sum_{j=1}^{n} a_j X_{t_j}\right\} = E\left\{\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k X_{t_j} X_{t_k}\right\}$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(t_k - t_j). \qquad \blacktriangle$$

If we set $n = 2$ in (1.4.1), we have

$$0 \leq a_1^2\gamma(0) + a_2^2\gamma(0) + 2a_1a_2\gamma(t_1 - t_2),$$

which implies

$$0.5(a_1^2 + a_2^2) \geq -a_1a_2[\gamma(0)]^{-1}\gamma(t_1 - t_2).$$

For $t_1 - t_2 = h$ we set $-a_1 = a_2 = 1$ and then set $a_1 = a_2 = 1$ to obtain the well-known property of correlations:

$$|\rho(h)| \leq 1. \tag{1.4.2}$$

The concepts of *even* and *odd* functions (about zero) will prove useful in our study. For example, $\cos t$ is an *even function*. The use of this description is apparent when one notes the symmetry of $\cos t$ on the interval $[-\pi, \pi]$. Similarly, $\sin t$ is an *odd function*. In general, an even function, $f(t)$, defined on a domain T, is a function that satisfies $f(t) = f(-t)$ for all t and $-t$ in T. An odd function $g(t)$ is a function satisfying $g(t) = -g(-t)$ for all t and $-t$ in T. Many simple properties of even and odd functions follow immediately. For example, if $f(t)$ is an even function and $g(t)$ is an odd function, where both are integrable on the interval $[-A, A]$, then, for $0 \leq b \leq A$,

$$\int_{-b}^{b} g(t)\,dt = 0,$$

$$\int_{-b}^{b} f(t)\,dt = 2\int_{0}^{b} f(t)\,dt.$$

As an exercise, the reader may verify that the product of two even functions is even, the product of two odd functions is even, the product of an odd and an even function is odd, the sum of even functions is even, and the sum of odd functions is odd.

Theorem 1.4.2. The covariance function of a real valued stationary time series is an even function of h. That is, $\gamma(h) = \gamma(-h)$.

Proof. We assume, without loss of generality, that $E\{X_t\} = 0$. By stationarity,

$$E\{X_t X_{t+h}\} = \gamma(h)$$

for all t and $t + h$ contained in the index set. Therefore, if we set $t_0 = t_1 - h$,

$$\gamma(h) = E\{X_{t_0} X_{t_0+h}\} = E\{X_{t_1-h} X_{t_1}\} = \gamma(-h)\,. \qquad \blacktriangle$$

Given this theorem, we shall often evaluate $\gamma(h)$ for real valued time series for nonnegative h only. Should we fail to so specify, the reader may always safely substitute $|h|$ for h in a covariance function.

In the study of statistical distribution functions the characteristic function of a distribution function is defined by

$$\varphi(h) = \int e^{ixh}\, dG(x)\,,$$

where the integral is a Lebesgue–Stieltjes integral, $G(x)$ is the distribution function, and e^{ixh} is the complex exponential defined by $e^{ixh} = \cos xh + i \sin xh$. It is readily established that the function $\varphi(h)$ satisfies:

1. $\varphi(0) = 1$.
2. $|\varphi(h)| \leq 1$ for all $h \in (-\infty, \infty)$.
3. $\varphi(h)$ is uniformly continuous on $(-\infty, \infty)$.

See, for example, Tucker (1967, p. 42 ff.) or Gnedenko (1967, p. 266 ff.).

It can be shown that a continuous function $\rho(h)$ with $\rho(0) = 1$ is a characteristic function if and only if it is nonnegative definite. For example, see Gnedenko (1967, pp. 290, 387).

Theorem 1.4.3. The real valued function $\rho(h)$ is the correlation function of a mean square continuous stationary real valued time series with index set $T = (-\infty, \infty)$ if and only if it is representable in the form

$$\rho(h) = \int_{-\infty}^{\infty} e^{ixh}\, dG(x)\,,$$

where $G(x)$ is a symmetric distribution function, and the integral is a Lebesgue–Stieltjes integral.

For the index set $T = \{0, \pm 1, \pm 2, \ldots\}$, the corresponding theorem is:

Theorem 1.4.4. The real valued function $\rho(h)$ is the correlation function of a real valued stationary time series with index set $T = \{0, \pm 1, \pm 2, \ldots\}$ if and only if

it is representable in the form

$$\rho(h) = \int_{-\pi}^{\pi} e^{ixh}\, dG(x),$$

where $G(x)$ is a symmetric distribution function.

This theorem will be discussed in Chapter 3.

Associated with the autocorrelation function of a time series is the partial autocorrelation function. Before discussing ideas of partial correlation for time series, we recall the partial correlation coefficient of multivariate analysis. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$ be a p-variate random variable with nonsingular covariance matrix. Then the partial correlation between Y_1 and Y_2 after Y_3 is

$$\rho_{12\cdot 3} = [(1-\rho_{13}^2)(1-\rho_{23}^2)]^{-1/2}(\rho_{12}-\rho_{13}\rho_{23}), \tag{1.4.3}$$

where $\rho_{ij} = (\sigma_{ii}\sigma_{jj})^{-1/2}\sigma_{ij}$ and σ_{ij} is the covariance between Y_i and Y_j. An alternative definition of $\rho_{12\cdot 3}^2$ in terms of regression coefficients is

$$\rho_{12\cdot 3}^2 = \beta_{12\cdot 3}\beta_{21\cdot 3}, \tag{1.4.4}$$

where $\beta_{ij\cdot k}$ is the partial regression coefficient for Y_j in the regression of Y_i on Y_j and Y_k. For example,

$$\begin{pmatrix}\beta_{12\cdot 3}\\ \beta_{13\cdot 2}\end{pmatrix} = \begin{pmatrix}\sigma_{22} & \sigma_{23}\\ \sigma_{32} & \sigma_{33}\end{pmatrix}^{-1}\begin{pmatrix}\sigma_{12}\\ \sigma_{13}\end{pmatrix},$$

$$\begin{pmatrix}\beta_{21\cdot 3}\\ \beta_{23\cdot 1}\end{pmatrix} = \begin{pmatrix}\sigma_{11} & \sigma_{13}\\ \sigma_{31} & \sigma_{33}\end{pmatrix}^{-1}\begin{pmatrix}\sigma_{21}\\ \sigma_{23}\end{pmatrix}.$$

Recall that $\beta_{ij\cdot k}$ is the simple population regression coefficient for the regression of

$$Y_i - \beta_{ik}Y_k \quad \text{on} \quad Y_j - \beta_{jk}Y_k,$$

where $\beta_{ik} = \sigma_{kk}^{-1}\sigma_{ik}$. Therefore, the partial correlation between Y_i and Y_j after Y_k is the simple correlation between $Y_i - \beta_{ik}Y_k$ and $Y_j - \beta_{jk}Y_k$. Also note that the sign of $\beta_{ij\cdot k}$ is always equal to the sign of $\beta_{ji\cdot k}$. The multiple correlation associated with the regression of Y_i on Y_j and Y_k is denoted by $R_{i(jk)}^2$ and is defined by the equation

$$1 - R_{i(jk)}^2 = (1-\rho_{ij}^2)(1-\rho_{ik\cdot j}^2) = (1-\rho_{ik}^2)(1-\rho_{ij\cdot k}^2). \tag{1.4.5}$$

Using (1.4.4), the extension of the definition to higher order partial correlations is straightforward. The squared partial correlation between Y_i and Y_j after adjusting for Y_k and Y_l is

$$\rho_{ij\cdot kl}^2 = \beta_{ij\cdot kl}\beta_{ji\cdot kl}, \tag{1.4.6}$$

where

$$\begin{pmatrix} \beta_{ij\cdot kl} \\ \beta_{ik\cdot jl} \\ \beta_{il\cdot jk} \end{pmatrix} = \begin{pmatrix} \sigma_{jj} & \sigma_{jk} & \sigma_{jl} \\ \sigma_{kj} & \sigma_{kk} & \sigma_{kl} \\ \sigma_{lj} & \sigma_{lk} & \sigma_{ll} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{ij} \\ \sigma_{ik} \\ \sigma_{il} \end{pmatrix}. \tag{1.4.7}$$

The partial correlation $\rho_{ij\cdot kl}$ can also be interpreted as the simple correlation between $Y_i - \beta_{ik\cdot l}Y_k - \beta_{il\cdot k}Y_l$ and $Y_j - \beta_{jk\cdot l}Y_k - \beta_{jl\cdot k}Y_l$. The multiple correlation associated with the regression of Y_i on the set (Y_j, Y_k, Y_l) satisfies

$$\begin{aligned} 1 - R^2_{i(jkl)} &= (1 - \rho^2_{ij})(1 - \rho^2_{ik\cdot j})(1 - \rho^2_{il\cdot jk}) \\ &= (1 - \rho^2_{ik})(1 - \rho^2_{il\cdot k})(1 - \rho^2_{ij\cdot kl}). \end{aligned} \tag{1.4.8}$$

The squared partial correlation between Y_1 and Y_2 after adjusting for $(Y_3, Y_4, \ldots, Y_p)$ is

$$\rho^2_{12\cdot 3,4,\ldots,p} = \beta_{12\cdot 3,4,\ldots,p}\beta_{21\cdot 3,4,\ldots,p},$$

where $\beta_{12\cdot 3,4,\ldots,p}$ is the population regression coefficient for Y_2 in the regression of Y_1 on $Y_2, Y_3, \ldots, Y_p$, and $\beta_{21\cdot 3,4,\ldots,p}$ is the population regression coefficient for Y_1 in the regression of Y_2 on $Y_1, Y_3, Y_4, \ldots, Y_p$. The multiple correlation between Y_1 and the set $(Y_2, Y_3, \ldots, Y_p)$ satisfies the equation

$$1 - R^2_{1(2,3,\ldots,p)} = (1 - \rho^2_{12})(1 - \rho^2_{13\cdot 2})(1 - \rho^2_{14\cdot 2,3})\cdots(1 - \rho^2_{1p\cdot 2,3,\ldots,p-1}). \tag{1.4.9}$$

For a covariance stationary time series Y_t with $R^2_{1(2,3,\ldots,r-1)} < 1$, the *partial autocorrelation function* is denoted by $\phi(h)$ and, for $0 < h \leq r$, is the partial correlation between Y_t and Y_{t+h} after adjusting for $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+h-1}$. It is understood that $\phi(0) = 1$ and $\phi(1) = \rho(1)$. The partial autocorrelation function is defined in an analogous way for $h < 0$. Because $\rho(h)$ is symmetric, $\phi(h)$ is symmetric. Let θ_{ih} be the population regression coefficient for Y_{t-i}, $1 \leq i \leq h$, in the regression of Y_t of $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-h}$. The population regression equation is

$$Y_t = \theta_{0h} + \theta_{1h}Y_{t-1} + \theta_{2h}Y_{t-2} + \cdots + \theta_{hh}Y_{t-h} + a_{ht}, \tag{1.4.10}$$

where

$$\begin{pmatrix} \theta_{1h} \\ \theta_{2h} \\ \vdots \\ \theta_{hh} \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix}, \tag{1.4.11}$$

the matrix of correlations is positive definite by the assumption that $R^2_{1(2,3,\ldots,h-1)} < 1$, and a_{hl} is the population regression residual. From (1.4.11),

$$\phi(h) = \theta_{hh} \tag{1.4.12}$$

because the coefficient for Y_{t-h} in the regression of Y_t on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-h}$ is equal to the coefficient for Y_t in the regression of Y_{t-h} on $Y_{t-h+1}, Y_{t-h+2}, \ldots, Y_t$. The equality follows from the symmetry of the autocorrelation function $\rho(h)$. From (1.4.9) we have

$$E\{a_{hl}^2\} = \gamma_Y(0) \prod_{i=1}^{h} [1 - \phi^2(i)] . \tag{1.4.13}$$

1.5. COMPLEX VALUED TIME SERIES

Occasionally it is advantageous, from a theoretical point of view, to consider complex valued time series. Letting X_t and Y_t be two real valued time series, we define the complex valued time series Z_t by

$$Z_t = X_t + iY_t . \tag{1.5.1}$$

The expected value of Z_t is given by

$$E\{Z_t\} = E\{X_t\} + iE\{Y_t\} , \tag{1.5.2}$$

and we note that

$$E^*\{Z_t\} = E\{Z_t^*\} = E\{X_t\} - iE\{Y_t\} , \tag{1.5.3}$$

where the symbol "*" is used to denote the complex conjugate.

The covariance of Z_t and Z_{t+h} is defined as

$$\begin{aligned} \mathrm{Cov}\{Z_t, Z_{t+h}\} &= E\{(Z_t - E\{Z_t\})(Z_{t+h}^* - E\{Z_{t+h}^*\})\} \\ &= \mathrm{Cov}\{X_t, X_{t+h}\} + i\,\mathrm{Cov}\{Y_t, X_{t+h}\} \\ &\quad - i\,\mathrm{Cov}\{X_t, Y_{t+h}\} + \mathrm{Cov}\{Y_t, Y_{t+h}\} . \end{aligned} \tag{1.5.4}$$

Note that the variance of a complex valued process is always real and nonnegative, since it is the sum of the variances of two real valued random variables.

The definitions of stationarity for complex time series are completely analogous to those for real time series. Thus, a complex time series Z_t is weakly stationary if the expected value of Z_t is a constant for all t and the covariance matrix of $(Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n})$ is the same as the covariance matrix of $(Z_{t_1+h}, Z_{t_2+h}, \ldots, Z_{t_n+h})$, where all indices are contained in the index set.

From (1.5.4), the autocovariance of a stationary complex time series Z_t with zero mean is given by

$$\gamma_Z(h) = [E\{X_tX_{t+h}\} + E\{Y_tY_{t+h}\}] + i[E\{Y_tX_{t+h}\} - E\{X_tY_{t+h}\}]$$
$$\overset{\text{(say)}}{=} g_1(h) + ig_2(h). \tag{1.5.5}$$

We see that $g_1(h)$ is a symmetric or even function of h, and $g_2(h)$ is an odd function of h. By the definition of the complex conjugate, we have

$$\gamma_Z^*(h) = \gamma_Z(-h). \tag{1.5.6}$$

Therefore, the autocovariance function of a complex time series is *skew symmetric*, where (1.5.6) is the definition of a skew symmetric function.

A complex valued fucntion $\gamma(\cdot)$, defined on the integers, is positive semidefinite if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} v_j v_k^* \gamma(t_k - t_j) \geq 0 \tag{1.5.7}$$

for any set of n complex numbers $(v_1, v_2, \ldots, v_n)$ and any integers $(t_1, t_2, \ldots, t_n)$. Thus, as in the real valued case, we have:

Theorem 1.5.1. The covariance function of a stationary complex valued time series is positive semidefinite.

It follows from Theorem 1.5.1 that the correlation inequality holds for complex random variables; that is,

$$\rho_Z(h)\rho_Z^*(h) = \frac{\gamma_Z(h)\gamma_Z^*(h)}{[\gamma_Z(0)]^2} \leq 1.$$

In the sequel, if we use the simple term "time series," the reader may assume that we are speaking of a real valued time series. All complex valued time series will be identified as such.

1.6. PERIODIC FUNCTIONS AND PERIODIC TIME SERIES

Periodic functions play an important role in the analysis of empirical time series. We define a function $f(t)$ with domain T to be *periodic* if there exists an $H > 0$ such that for all t, $t + H \in T$,

$$f(t + H) = f(t),$$

where H is the *period* of the function. That is, the function $f(t)$ takes on all of its

possible values in an interval of length H. For any positive integer, k, kH is also a period of the function. While examples of perfectly periodic functions are rare, there are situations where observed time series may be decomposed into the sum of two time series, one of which is periodic or nearly periodic. Even casual observation of many economic time series will disclose seasonal variation wherein peaks and troughs occur at approximately the same month each year. Seasonal variation is apparent in many natural time series, such as the water level in a lake, daily temperature, wind speeds and velocity, and the levels of the tides. Many of these time series also display regular daily variation that has the appearance of rough "cycles."

The trigonometric functions have traditionally been used to approximate periodic behavior. A function obeying the sine–cosine type periodicity is completely specified by three parameters. Thus, we write

$$f(t) = A \sin(\lambda t + \varphi), \tag{1.6.1}$$

where the *amplitude* is the absolute value of A, λ is the *frequency*, and φ is the *phase angle*. The frequency is the number of times the function repeats itself in a period of length 2π. The phase angle is a "shift parameter" in that it determines the points ($-\lambda^{-1}\varphi$ plus integer multiples of π) where the function is zero. A parametrization that is more useful in the estimation of such a function can be constructed by using the trigonometric identity

$$\sin(\lambda t + \varphi) = \sin \lambda t \cos \varphi + \cos \lambda t \sin \varphi .$$

Thus (1.6.1) becomes

$$f(t) = B_1 \cos \lambda t + B_2 \sin \lambda t , \tag{1.6.2}$$

where $B_1 = A \sin \varphi$ and $B_2 = A \cos \varphi$.

Let us consider a simple type of time series whose realizations will display perfect periodicity. Define $\{X_t : t \in (0, \pm 1, \pm 2, \ldots)\}$ by

$$X_t = e_1 \cos \lambda t + e_2 \sin \lambda t , \tag{1.6.3}$$

where e_1 an e_2 are independent drawings from a normal (0, 1) population. Note that the realization is completely determined by the two random variables e_1 and e_2. The amplitude and phase angle vary from realization to realization, but the period is the same for every realization.

The stochastic properties of this time series are easily derived:

$$\begin{aligned} E\{X_t\} &= E\{e_1 \cos \lambda t + e_2 \sin \lambda t\} \\ &= 0 ; \\ \gamma(h) &= E\{(e_1 \cos \lambda t + e_2 \sin \lambda t)[e_1 \cos \lambda(t+h) + e_2 \sin \lambda(t+h)]\} \\ &= E\{e_1^2 \cos \lambda t \cos \lambda(t+h) + e_2^2 \sin \lambda t \sin \lambda(t+h) \end{aligned} \tag{1.6.4}$$

$$+ e_1 e_2 \cos \lambda t \sin \lambda(t+h) + e_2 e_1 \sin \lambda t \cos \lambda(t+h)\}$$
$$= \cos \lambda t \cos \lambda(t+h) + \sin \lambda t \sin \lambda(t+h)$$
$$= \cos \lambda h .$$

We see that the process is stationary. This example also serves to emphasize the fact that the covariance function is obtained by averaging the product $X_t X_{t+h}$ over realizations.

It is clear that any time series defined by the finite sum

$$X_t = \sum_{j=1}^{M} e_{1j} \cos \lambda_j t + \sum_{j=1}^{M} e_{2j} \sin \lambda_j t , \tag{1.6.5}$$

where e_{1j} and e_{2j} are independent drawings from a normal $(0, \sigma_j^2)$ population and $\lambda_j \neq \lambda_k$ for $k \neq j$, will display periodic behavior. The representation (1.6.5) is a useful approximation for portions of some empirical time series, and we shall see that ideas associated with this representation are important for the theoretical study of time series as well.

1.7. VECTOR VALUED TIME SERIES

Most of the concepts associated with univariate time series have immediate generalizations to vector valued time series. We now introduce representations for multivariate processes.

The k-dimensional time series $\{\mathbf{X}_t : t = 0, \pm 1, \pm 2, \ldots\}$ is defined by

$$\mathbf{X}_t = [X_{1t}, X_{2t}, \ldots, X_{kt}]' , \tag{1.7.1}$$

where $\{X_{it} : t = 0, \pm 1, \pm 2, \ldots\}$, $i = 1, 2, \ldots, k$, are scalar time series. The expected value of $\mathbf{X}_t'$ is

$$E\{\mathbf{X}_t'\} = [E\{X_{1t}\}, E\{X_{2t}\}, \ldots, E\{X_{kt}\}] .$$

Assuming the mean is zero, we define the covariance matrix of $\mathbf{X}_t$ and $\mathbf{X}_{t+h}$ by

$$E\{\mathbf{X}_t \mathbf{X}_{t+h}'\} = \begin{bmatrix} E\{X_{1t}X_{1,t+h}\} & E\{X_{1t}X_{2,t+h}\} & \cdots & E\{X_{1t}X_{k,t+h}\} \\ E\{X_{2t}X_{1,t+h}\} & E\{X_{2t}X_{2,t+h}\} & \cdots & E\{X_{2t}X_{k,t+h}\} \\ \vdots & \vdots & \vdots & \vdots \\ E\{X_{kt}X_{1,t+h}\} & E\{X_{kt}X_{2,t+h}\} & \cdots & E\{X_{kt}X_{k,t+h}\} \end{bmatrix} . \tag{1.7.2}$$

As with scalar time series, we define $\mathbf{X}_t$ to be covariance stationary if:

1. The expected value of $\mathbf{X}_t$ is a constant function of time.
2. The covariance matrix of $\mathbf{X}_t$ and $\mathbf{X}_{t+h}$ is the same as the covariance matrix of $\mathbf{X}_j$ and $\mathbf{X}_{j+h}$ for all t, $t+h$, j, $j+h$ in the index set.

If a vector time series is stationary, then every component scalar time series is stationary. However, a vector of scalar stationary time series is not necessarily a vector stationary time series.

The second stationarity condition means that we can express the covariance matrix as a function of h only, and, assuming $E\{\mathbf{X}_t\} = 0$, we write

$$\Gamma(h) = E\{\mathbf{X}_t \mathbf{X}'_{t+h}\}$$

for stationary time series. Note that the diagonal elements of this matrix are the autocovariances of the X_{jt}. The off-diagonal elements are the *cross covariances* of X_{it} and X_{jt}. The element $E\{X_{it}X_{j,t+h}\}$ is not necessarily equal to $E\{X_{i,t+h}X_{jt}\}$, and hence $\Gamma(h)$ is not necessarily equal to $\Gamma(-h)$. For example, let

$$\begin{aligned} X_{1t} &= e_t\,, \\ X_{2t} &= e_t + \beta e_{t-1}\,. \end{aligned} \tag{1.7.3}$$

Then

$$\begin{aligned} \gamma_{12}(1) &= E\{X_{1t}X_{2,t+1}\} = \beta\sigma^2\,, \\ \gamma_{12}(-1) &= E\{X_{1,t+1}X_{2t}\} = 0\,. \end{aligned}$$

However,

$$\begin{aligned} \gamma_{21}(1) &= E\{X_{2t}X_{1,t+1}\} = 0 = \gamma_{12}(-1)\,, \\ \gamma_{21}(-1) &= E\{X_{2t}X_{1,t-1}\} = \beta\sigma^2 = \gamma_{12}(1)\,, \end{aligned}$$

and it is clear that

$$\Gamma(h) = \Gamma'(-h)\,. \tag{1.7.4}$$

It is easy to verify that (1.7.4) holds for all vector stationary time series, and we state the result as a lemma.

Lemma 1.7.1. The autocovariance matrix of a vector stationary time series satisfies $\Gamma(h) = \Gamma'(-h)$.

The nonnegative definite property of the scalar autocovariance function is maintained essentially unchanged for vector processes.

Lemma 1.7.2. The covariance function of a vector stationary time series $\{\mathbf{X}_t : t \in T\}$ is a nonnegative definite function in that

$$\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{a}_j'\boldsymbol{\Gamma}(t_i - t_j)\mathbf{a}_i \geq 0$$

for any set of real vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ and any set of indices $\{t_1, t_2, \ldots, t_n\} \in T$.

Proof. The result can be obtained by evaluating the variance of

$$\sum_{j=1}^{n} \mathbf{a}_j'\mathbf{X}_{t_j}. \qquad \blacktriangle$$

We define the *correlation matrix* of $\mathbf{X}_t$ and $\mathbf{X}_{t+h}$ by

$$\mathbf{P}(h) = \mathbf{D}_0^{-1}\boldsymbol{\Gamma}(h)\mathbf{D}_0^{-1}, \tag{1.7.5}$$

where $\mathbf{D}_0$ is a diagonal matrix with the square root of the variances of the X_{jt} as diagonal elements; that is,

$$\mathbf{D}_0^2 = \operatorname{diag}\{\gamma_{11}(0), \gamma_{22}(0), \ldots, \gamma_{kk}(0)\}.$$

The ijth element of $\mathbf{P}(h)$, $\rho_{ij}(h)$, is called the *cross correlation* of X_{it} and X_{jt}.

For the time series of (1.7.3) we have

$$\mathbf{P}(h) = \begin{cases} \begin{pmatrix} 1 & (1+\beta^2)^{-1/2} \\ (1+\beta^2)^{-1/2} & 1 \end{pmatrix}, & h = 0, \\ \begin{pmatrix} 0 & (1+\beta^2)^{-1/2}\beta \\ 0 & (1+\beta^2)^{-1}\beta \end{pmatrix}, & h = 1, \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & |h| > 1. \end{cases}$$

REFERENCES

Chung (1968), Gnedenko (1967), Loève (1963), Rao (1973), Tucker (1967), Yaglom (1962).

EXERCISES

1. Determine the mean and covariance function for Example 1 of Section 1.3.

2. Discuss the stationarity of the following time series:

(a) $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\} =$ Value of a randomly chosen observation from a normal distribution with mean $\frac{1}{2}$ and variance $\frac{1}{4}$ (t odd)
$= 1$ if toss of a true coin results in a head
$= 0$ if toss of a true coin results in a tail (t even)

(b) $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ is a time series of independent identically distributed random variables whose distribution function is that of Student's t-distribution with one degree of freedom.

(c) $X_t = \begin{cases} ce_0, & t = 0, \\ \rho X_{t-1} + e_t, & t = 1, 2, \ldots, \end{cases}$

where c is a constant, $|\rho| < 1$, and the e_t are iid$(0, 1)$ random variables.

3. Which of the following processes is covariance stationary for $T = \{t: t \in (0, 1, 2, \ldots)\}$, where the e_t are independent identically distributed $(0, 1)$ random variables and a_1, a_2 are fixed real numbers?
   (a) $e_1 + e_2 \cos t$.
   (b) $e_1 + e_2 \cos t + e_3 \sin t$.
   (c) $a_1 + e_1 \cos t$.
   (d) $a_1 + e_1 \cos t + e_2 \sin t$.
   (e) $e_t + a_1 \cos t$.
   (f) $a_1 + e_1 a_2^t + e_2$, $0 < a_2 < 1$.

4. (a) Prove the following theorem.

   Theorem. If X_t and Y_t are independent covariance stationary time series, then $aX_t + bY_t$, where a and b are real numbers, is a covariance stationary time series.

   (b) Let $Z_t = X_t + Y_t$. Give an example to show that Z_t covariance stationary does not imply X_t is covariance stationary.

5. Give a covariance stationary time series such that $\rho(h) \neq 0$ for all h.

6. Which of the following functions is the covariance function of a stationary time series? Explain why or why not.

(a) $g(h) = 1 + |h|, \quad h = 0, \pm 1, \pm 2, \ldots;$

(b) $g(h) = \begin{cases} 1, & h = 0, \\ -\frac{1}{2}, & h = \pm 1, \\ 0 & \text{otherwise}; \end{cases}$

(c) $g(h) = 1 + \frac{1}{4}\sin 4h, \quad h = 0, \pm 1, \pm 2, \ldots;$

(d) $g(h) = 1 + \frac{1}{4}\cos 4h, \quad h = 0, \pm 1, \pm 2, \ldots;$

(e) $g(h) = \begin{cases} 1, & h = 0, \pm 1, \\ 0 & \text{otherwise}. \end{cases}$

7. Let e_t be time series of normal independent $(0, \sigma^2)$ random variables. Define

$$Y_t = \begin{cases} e_t & \text{if } t \text{ is odd}, \\ -e_t & \text{if } t \text{ is even}. \end{cases}$$

Is the complex valued time series $Z_t = e_t + eY_t$ stationary?

8. Let

$$\rho(h) = \begin{cases} 1, & h = 0, \\ a, & h = \pm 1, \\ 0 & \text{otherwise}. \end{cases}$$

(a) For what values of a is $\rho(h)$ the autocorrelation function of a stationary time series?

(b) Compute the partial autocorrelation function associated with $\rho(h)$, given that a is such that $\rho(h)$ is an autocorrelation function.

9. Let the time series Y_t be defined by

$$Y_t = e_1 \cos 0.5\pi t + e_2 \sin 0.5\pi t + e_t$$

for $t = 3, 4, \ldots$, where $e_t \sim \text{NI}(0, 1)$ and $\sim$NI(0, 1) denotes normally and independently distributed with zero mean and unit variance. Compute the partial autocorrelation function of Y_t for $h = 1, 2, 3, 4$.

10. Show that if $\beta_{12\cdot 3}$ of (1.4.4) is zero, then $\beta_{21\cdot 3}$ is also zero. Extend this result to the pth order partial regression coefficients $\beta_{1p\cdot 2,\ldots,p-1}$ and $\beta_{p1\cdot 2,\ldots,p-1}$.

11. Assume that for the stationary time series Y_t, there is some r such that $R^2_{1(2,3,\ldots,r)} = 1$, and $R^2_{1(2,3,\ldots,r-1)} < 1$, where $R^2_{1(2,3,\ldots,r)}$ is the squared multiple correlation between Y_1 and $(Y_2, \ldots, Y_r)$. Show that $|\phi(r-1)| = 1$ and that the correlation matrix in (1.4.11) is singular for $h \geq r$.

12. Let $\{Z_t\}$ be a sequence of normal independent (0, 1) random variables defined on the integers.
    (a) Give the mean and covariance function of the time series (i) $Y_t = Z_t Z_{t-1}$ and (ii) $X_t = Z_t + Z_{t-1}$.
    (b) Is the time series

$$W_t = \begin{cases} Z_t, & t \text{ even}, \\ 2^{-1/2}(Z_t^2 - 1), & t \text{ odd} \end{cases}$$

    covariance stationary? Is it strictly stationary?