CHAPTER 2

Analysis of Linear Deterministic Systems

Dynamic economics is concerned with the evolution of economic variables through time, that is, with the determination of the time paths or time functions of economic variables. It is well known that static economics is concerned with the determination of economic variables without specifying when their values will materialize. The foundation of a large body of static economic theory is the assumption that individual economic units maximize something. The individual consumer is assumed to maximize utility, and a static demand theory of the consumer is derived. The individual firm is assumed to maximize profit, and a static theory of the behavior of the firm is derived. The basis of many of the dynamic economic theories today is less rigorous.

Quite often a dynamic theory is obtained by introducing certain time delays into an otherwise static theory. The pattern of time delays is not necessarily derived from an assumption that certain individuals maximize something over time, although the static theory forming the basis of dynamics may be derived from a maximizing assumption. One example is the set of two equations that explains the cobweb phenomena in agricultural economics. A static linear demand equation relates the quantity demanded negatively to the price of the commodity of the same period. A dynamic linear supply equation relates the quantity supplied positively to the price of the commodity in the *last period*. By solving these two equations, assuming that quantities demanded and supplied are equal, we get an equation relating price at period t to the price at $t-1$. In this dynamic model the static theory of demand and supply is modified by introducing a time delay in the supply equation. We can find many examples in macroeconomics in which time delays are introduced into the

relations between consumption and income, between investment expenditures and investment plans (the latter possibly based on certain theory of maximizing behavior), and between demand for money and income. Whether the time delays themselves should *always* be justified by a theory of maximizing behavior over time, is a methodological, and somewhat philosophical, question that is not discussed in this book.

From the practical point of view we have dynamic macroeconomic models that are built from a combination of theory (of whatever kind) and statistical measurement, and these econometric models deserve to be studied by the methods of dynamics described. Furthermore, we introduce methods of optimization over time for the purpose of improving the dynamic performance of an economy—methods that will be useful for those theorists who wish to derive dynamic economic theories from optimization over time and presumably under uncertainty. There is no need to entertain the question whether the basis of static economics under certainty, namely maximizing behavior, should always be extended to the study of dynamic behavior over time under uncertainty. Possibly, or possibly not, maximizing behavior is a better approximation to human situations in which time and uncertainty are less relevant.

In this chapter we shall study dynamic models in the form of systems of linear difference equations. Linearity is assumed because of simplicity and because the most important concepts of dynamics can be understood via a linear model. Methods of linear dynamics can also be extended and modified for nonlinear systems, as in Chapter 6. Difference equations, rather than differential equations, are studied mainly because most existing macroeconometric models are in this form, and it would be difficult to cover the broad subject matter of this book by using both discrete-time and continuous-time models. The additional investment in mathematical skills to deal with models in continuous time, especially for the stochastic case, may not be worthwhile for most applied economists. Readers interested in the techniques of analysis and control of continuous-time systems may refer to Astrom (1970).

We confine ourselves to deterministic systems in this chapter. Solutions to systems of difference equations are provided and characterized. The solution to a univariate, first-order system is introduced first and then generalized to the multivariate, high-order case by matrix algebra. Properties of the solution are characterized and examined. Methods and concepts developed in this chapter will be useful for the study of dynamic properties of systems of stochastic difference equations in the next two chapters. This chapter may require several class periods to cover if the readers have no prior knowledge of linear difference equations. Readers who feel the need

for some economic applications of linear dynamic models may consult Baumol (1970) or Kenkel (1974) or read Sections 5.1, 5.2, and 5.3 at this point.

2.1 HOMOGENEOUS LINEAR DIFFERENCE EQUATION OF FIRST ORDER

To understand how an economic variable may evolve through time consider the simplest model of a first-order linear difference equation

$$y_t = ay_{t-1}. \tag{1}$$

It is *first order* because, in the determination of y_t, only y_{t-1} is used. If y_{t-m} is required, the equation is said to be of *order m*. In general, an equation is of order m if m is the largest difference in subscripts occurring in the equation. Equation 1 is also *homogeneous* because only lagged values of the endogenous variables are used, and no exogenous variables or constant or other given functions of time are present in the equation. The model (1) determines y_t simply by applying a factor a to its value of the last period. Given an initial value of the variable, say y_0 at time 0, the solution to (1) can be simply stated by repeated substitutions for the lagged variables on the right-hand side:

$$y_t = a^2y_{t-2} = a^ky_{t-k} = a^ty_0. \tag{2}$$

The solution y_0a^t, which is a function of time, can be easily characterized. First, whether it is *explosive* or *damped* depends on whether the absolute value of a is greater than or smaller than 1. It is a constant if a is exactly. 1. Second, whether it *oscillates* or not depends on whether a is negative or positive. If a is positive, the solution is a monotone function of time; if a is negative, the solution oscillates between successive time periods. Of course, the two characteristics can be combined. A solution may oscillate and be explosive if a is negative and has absolute value greater than 1; for example, if a equals -1.10. Thus the absolute value and the sign of a characterize the solution of (1).

2.2 HIGHER ORDER AND MULTIVARIATE HOMOGENEOUS LINEAR SYSTEMS

If a difference equation is of higher order, it is convenient for the purpose of analysis to rewrite it as a first-order system of many variables. Thus the

equation

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} \tag{3}$$

can be rewritten as

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix}. \tag{4}$$

This system explains a vector of two endogenous variables y_t and y_{t-1}. If the equation is of order m, a vector of $y_t, y_{t-1}, \ldots, y_{t-m+1}$ will be used in the system. The resulting system will be of first order like Equation 4 because we can denote the vector $(y_t, y_{t-1})'$ by z_t and the matrix on the right of (4) by A; then (4) can be written as $z_t = Az_{t-1}$ which is precisely in the same form as (1).

The same method of transforming a system into first order can be applied to a multivariate system of difference equations of higher order. Let y_t be a column *vector* of p endogenous variables that satisfy the mth-order difference equation

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_m y_{t-m}, \tag{5}$$

where $A_1, \ldots, A_m$ are $p \times p$ matrices of real coefficients. Equation 5 can be rewritten as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-m+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_m \\ I & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & I \; 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-m} \end{bmatrix}. \tag{6}$$

The first p components of (6) recaptures Equation 5. The remaining components of (6) are identities that explain the newly introduced endogenous variables $y_{t-1}, \ldots, y_{t-m+1}$. In a more compact form (6) can be rewritten as

$$y_t = Ay_{t-1}, \tag{7}$$

where y_t is redefined to represent the left-hand vector of the original equation (6) without the use of a new symbol and A stands for the right-hand matrix of (6). Difference equations in the form of (7) will be studied.

Analogous to (2), the solution to (7) can be obtained as

$$y_t = A^2 y_{t-2} = A^k y_{t-k} = A^t y_0, \tag{8}$$

where the product $A \cdot A$ is written as A^2, and so on. Given the initial value of the vector y_0, successive values of $y_t (t = 1, 2, \ldots)$ can be calculated by using $A^t y_0$. The solution in this form is not very informative, however. It would be desirable to characterize the time paths $A^t y_0$ to ascertain whether they will be explosive or damped and whether they will oscillate or fluctuate in some way.

2.3 CHARACTERIZATION OF THE SOLUTION TO A HOMOGENEOUS LINEAR SYSTEM

One useful approach to characterizing the solution (8) of a homogeneous linear system of difference equations is to decompose it into the solutions of many first-order univariate equations in the form of (1). This can be done by utilizing the *characteristic roots* and *characteristic vectors* (also called *eigenvalues* and *eigenvectors*) of the matrix A. A *characteristic root* of A is a scalar λ that satisfies

$$|A - \lambda I| = 0, \tag{9}$$

where the two vertical bars denote the determinant of the matrix inside and I is an identity matrix of order p. As an example, consider a 2×2 matrix $A = (a_{ij})$ in which a_{ij} are elements. Equation 9 becomes

$$|A - \lambda I| = \left| \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix}$$

$$= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0. \tag{10}$$

The two roots, say λ_1 and λ_2, of (10) are the characteristic roots or eigenvalues of the 2×2 matrix $A = (a_{ij})$. In general, Equation 9 is a polynomial equation of degree p in λ. It is called the *characteristic equation* of the $p \times p$ matrix A. Therefore the matrix A has p characteristic roots which can be found by solving a polynomial equation of degree p. Thus the computation of characteristic roots is equivalent to the solution of polynomial equations. The computational aspects of finding characteristic roots are not pursued in this book. Numerous computational algorithms for finding these roots and vectors and computer programs to implement them are readily available.

With each characteristic root λ_i there is associated a *right characteristic vector* b_i. A right characteristic vector b of a matrix A is a $p \times 1$ vector that satisfies the equation

$$Ab = \lambda b, \tag{11}$$

where λ is some scalar. In other words, the characteristic vector b has the property that, when premultiplied by the matrix A or when subject to the linear transformation A, it remains a scalar multiple of itself. Equation 11 is equivalent to

$$(A - \lambda I)b = 0 \tag{12}$$

and Equation 12 is satisfied, for b not equal to a zero vector, if and only if the matrix $(A - \lambda I)$ is singular, that is, if and only if the characteristic equation (9) is satisfied. The solution to (9) is the set of characteristic roots $\lambda_1, \ldots, \lambda_p$. For each λ_i, a characteristic vector b_i is defined by equation (12), that is,

$$(A - \lambda_i I)b_i = 0. \tag{13}$$

Note that the vector b_i is defined only up to a factor of proportionality; if b_i satisfies (13), so will $2b_i$, for instance. For the purpose of computing the vector b_i which corresponds to a real root λ_i we may arbitrarily set its first element b_{1i} equal to 1 and solve the resulting equations from (13) for the remaining unknowns $b_{2i}, \ldots, b_{pi}$. These equations will be a set of $(p-1)$ nonhomogeneous linear equations in these $(p-1)$ unknowns. Another convention, frequently applied in computer programs, normalizes the elements of b_i so that the sum of their squares is unity.

It will be useful to combine the p equations relating the characteristic roots λ_i and vectors $b_i (i = 1, \ldots, p)$ into one system of p vector equations; $Ab_1 = \lambda_1 b_1$, $Ab_2 = \lambda_2 b_2$, and so on, can be combined as

$$A(b_1, \ldots, b_p) = (\lambda_1 b_1, \ldots, \lambda_p b_p) = (b_1, \ldots, b_p) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}. \tag{14}$$

Denoting by B the matrix $(b_1, \ldots, b_p)$, whose columns are the characteristic vectors of A, and by D_λ the diagonal matrix which consists of the corresponding characteristic roots, we can write (14) as

$$AB = BD_\lambda. \tag{15}$$

To simplify the analysis, we assume that all the characteristic roots of A are distinct and ignore the case of multiple roots. (The seriousness of this assumption for practical purposes is discussed in Chapter 5.) Under this assumption B is known to be nonsingular. Postmultiplying both sides of (15) by B^{-1} gives

$$A = BD_\lambda B^{-1}. \tag{16}$$

Writing the matrix A in this form (16) is useful for the study of the solution $A^t y_0$ of the difference equations (7) because A^t can also be written in a convenient form. Note that $A^2 = (BD_\lambda B^{-1})(BD_\lambda B^{-1}) = BD_\lambda^2 B^{-1}$, and similarly that

$$A^t = BD_\lambda^t B^{-1}, \tag{17}$$

where D_λ^t is simply a diagonal matrix that has λ_i^t as its ith diagonal element. The solution (8), rewritten as

$$y_t = BD_\lambda^t B^{-1} y_0, \tag{18}$$

consists of functions of the characteristic roots of A. To write out the functions explicitly, let $B = (b_{ij})$ be a 3×3 matrix and let the $i-j$ element of B^{-1} be denoted by b^{ij}. Expanding (18) in scalar variables gives

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} b_{11}\lambda_1^t & b_{12}\lambda_2^t & b_{13}\lambda_3^t \\ b_{21}\lambda_1^t & b_{22}\lambda_2^t & b_{23}\lambda_3^t \\ b_{31}\lambda_1^t & b_{32}\lambda_2^t & b_{33}\lambda_3^t \end{bmatrix} \begin{bmatrix} b^{11}y_{10} + y^{12}y_{20} + b^{13}y_{30} \\ b^{21}y_{10} + b^{22}y_{20} + b^{23}y_{30} \\ b^{31}y_{10} + y^{32}y_{20} + b^{33}y_{30} \end{bmatrix}$$

$$= \begin{bmatrix} b_{11}z_{10}\lambda_1^t + b_{12}z_{20}\lambda_2^t + b_{13}z_{30}\lambda_3^t \\ b_{21}z_{10}\lambda_1^t + b_{22}z_{20}\lambda_2^t + b_{23}z_{30}\lambda_3^t \\ b_{31}z_{10}\lambda_1^t + b_{32}z_{20}\lambda_2^t + b_{33}z_{30}\lambda_3^t \end{bmatrix}, \tag{19}$$

where z_{i0} is defined as

$$z_{i0} = b^{il}y_{10} + b^{i2}y_{20} + b^{i3}y_{30}, \tag{20}$$

that is, as a linear combination of y_{10}, y_{20}, and y_{30}, using b^{il}, b^{i2}, and b^{i3} as weights. Written in scalar form, the solution y_{jt} is thus a linear combination of λ_1^t, λ_2^t, and λ_3^t.

2.4 CHARACTERIZATION OF THE SOLUTION USING CANONICAL VARIABLES

Because, by (19), the solution of each y_{jt} is a linear combination of λ_i^t and the solution of a univariate first-order equation $z_{it}=\lambda_i z_{i,t-1}$ is also proportional to λ_i^t, we can think of y_{jt} as a linear combination of a set of new variables z_{it}. Equation 20 suggests that a set of *canonical variables* z_{it} can be defined as

$$z_{it}=b^{i1}y_{1t}+b^{i2}y_{2t}+\cdots+b^{ip}y_{pt}, \qquad (i=1,\ldots,p). \tag{21}$$

The p canonical variables are linear combinations of the p original variables and vice persa. In vector form their relationship is given by

$$z_t=B^{-1}y_t \quad \text{or} \quad y_t=Bz_t. \tag{22}$$

Because of this relationship, the solution of one set of variables can be obtained as a linear combination of the solution of the other. Given the solutions for $z_{1t},\ldots,z_{pt}$, which are easy to obtain and interpret, the solution for y_{jt} is simply their linear combination, using the jth row of B as weights. Canonical variables are useful not only for deriving a solution for y_{jt} but also for treating an explosive system of difference equations which otherwise can hardly be characterized. We turn to the latter subject in Sections 6.1, 6.2, and 6.3.

To find the solution for z_{it} we rewrite the system (7) in terms of the canonical variables, using (16) and the definition (22):

$$z_t=B^{-1}y_t=B^{-1}(BD_\lambda B^{-1})y_{t-1}=D_\lambda z_{t-1}. \tag{23}$$

Each canonical variable z_{it} satisfies the difference equation

$$z_{it}=\lambda_i z_{i,t-1}. \tag{24}$$

Its solution is $z_{it}=z_{i0}\lambda_i^t$. The solution for y_{jt} is therefore

$$y_{jt}=b_{j1}(z_{10}\lambda_1^t)+b_{j2}(z_{20}\lambda_2^t)+\ldots+b_{jp}(z_{p0}\lambda_p^t). \tag{25}$$

Equation 25 agrees with the result of (19). Because the initial condition is often given in terms of y_{j0}, we need to calculate z_{i0} from y_{j0} by using (22) or (20).

To summarize, by defining the canonical variables z_t as linear combinations B^{-1} of y_t and observing that the solution of z_{it} is simply $z_{i0}\lambda_i^t$ we find that the solution for y_{jt} is a linear combination of the above solutions, using the jth row of B as weights. Accordingly, the solution for y_{jt} is a linear combination of the roots of A, each raised to a power t.

2.5 THE CASE OF COMPLEX ROOTS

The behavior of the solution can be quite interesting if some roots are complex. Complex numbers arise from solutions to polynomial equations such as $x^2+4=0$ and $x^2-x+2=0$. They will be required to express the solution of Equation 10 if $(a_{11}+a_{22})^2-4(a_{11}a_{22}-a_{12}a_{21})$ is negative. A *complex* number λ takes the form

$$\lambda=a+bi, \tag{26}$$

where $i=\sqrt{-1}$ and a and b are real numbers. It has two parts, the real part a and the imaginary part bi. By using a two-dimensional diagram to represent the real part along the horizontal axis and the imaginary part along the vertical axis we can represent λ by the point (a,b), as in Figure 2.1. The *absolute value* (also termed the *modulus*) of λ is the length of the vector (a,b) or the distance of the point (a,b) from the origin.

$$|\lambda|=\sqrt{a^2+b^2}\ . \tag{27}$$

Let θ denote the angle between the line from the origin to the point (a,b)

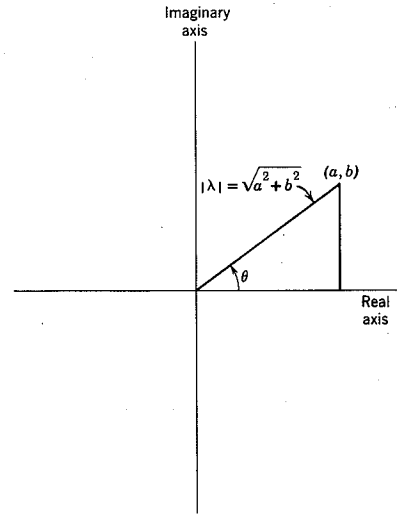

Figure 2.1 Diagrammatic representation of a complex number.

and the horizontal axis. By definition

$$a=|\lambda|\cos\theta; \qquad b=|\lambda|\sin\theta. \tag{28}$$

Using (28), we can rewrite the complex number (26) as

$$\lambda=|\lambda|(\cos\theta+i\sin\theta)=|\lambda|e^{i\theta}. \tag{29}$$

The second equality sign is due to the identity

$$e^{i\theta}=\cos\theta+i\sin\theta. \tag{30}$$

This identity can be made convincing, if not proved, by expanding the functions $e^{i\theta}$, $\cos\theta$, and $\sin\theta$ in Taylor series and matching the terms on both sides of (30). (See Problem 3.) Thus a complex number λ can be represented by the real and imaginary parts a and b or by the absolute value $|\lambda|$ and the angle θ, as in (29).

When complex roots of a matrix A of real numbers occur, they appear as pairs of *conjugate* numbers. The *conjugate* of the complex number λ, as defined by (26), is $a-bi$, or equivalently

$$|\lambda|(\cos\theta-i\sin\theta)=|\lambda|e^{-i\theta}.$$

It is denoted by $\bar{\lambda}$. The conjugate of $\bar{\lambda}$ is λ itself; λ and $\bar{\lambda}$ form a pair of conjugate complex numbers. Because the characteristic roots of A are the roots of a polynomial equation with real coefficients, they come in conjugate pairs when they are complex. This can be seen by factoring the polynomial equation (9) into

$$(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_p)=0,$$

where λ is the unknown and $\lambda_1,\lambda_2,\ldots,\lambda_p$ are the roots. If λ_1 is complex, there must be another root, say λ_2, such that the product

$$(\lambda-\lambda_1)(\lambda-\lambda_2)=\lambda^2-(\lambda_1+\lambda_2)\lambda+\lambda_1\lambda_2$$

will have real coefficients $(\lambda_1+\lambda_2)$ and $\lambda_1\lambda_2$, since the coefficients of the polynomial are real to begin with. These real coefficients imply that λ_1 and λ_2 are a pair of conjugate complex roots, as the reader may wish to verify in Problem 9 of this chapter.

In the solution to the system of difference equations, as given by (19), let λ_1 and λ_2 be a pair of conjugate complex roots, λ_3 being real. The contribution of the first two roots to the solution for y_{1t} is

$$b_{11}z_{10}\lambda_1^t+b_{12}z_{20}\lambda_2^t=c_{11}\lambda_1^t+c_{12}\lambda_2^t. \tag{31}$$

If this contribution is to be real, the coefficients c_{11} and c_{12} must be conjugate complex. The argument for this proposition is left to the reader as Problem 10 in this chapter. Let λ_1 and λ_2 be, respectively, $re^{i\theta}$ and $re^{-i\theta}$, r being the absolute value $|\lambda_1|=|\lambda_2|$. Let c_{11} and c_{12} be, respectively, $s_1e^{i\psi_1}$ and $s_1e^{-i\psi_1}$. The contribution of the first two roots to y_{1t} is then

$$\begin{aligned} c_{11}\lambda_1^t+c_{12}\lambda_2^t&=s_1r^te^{i(\psi_1+\theta t)}+s_1r^te^{-i(\psi_1+\theta t)}\\ &=2s_1r^t\cos(\psi_1+\theta t), \end{aligned} \tag{32}$$

where the second equality sign is due to the identity

$$\begin{aligned} e^{ix}+e^{-ix}&=\cos x+i\sin x+\cos x-i\sin x\\ &=2\cos x. \end{aligned} \tag{33}$$

It has been shown that when there is a pair of complex roots of the matrix A their contribution to the solution for each variable y_{it} can be written as a cosine function of time t, multiplied by a factor r^t. If the absolute value r of the roots is larger than 1, the cosine function will be magnified through time; if r is less than 1, the cosine function will be damped. If r is exactly 1, there will be no magnifying or damping effect. To obtain the entire solution for y_{it}, if there are more than two roots, this contribution will have to be added to the linear combination of the remaining roots raised to power t. The latter may involve a real root, such as $c_{13}\lambda_3^t$ for y_{1t} in (19). It may also involve another pair of complex roots. The solution would then be the sum of two magnified or damped cosine functions.

2.6 A NOTE ON COSINE FUNCTIONS

Because the solution to a system of linear difference equations may require a cosine function of time, it may be useful to review some properties of this function. The function $\cos\theta t$ takes the same value 1 when $\theta t=0,2\pi,4\pi,6\pi$, etc., that is, when $t=0,2\pi/\theta,4\pi/\theta,6\pi/\theta$, etc. In fact, it takes an identical value for $\theta t=k,2\pi+k,4\pi+k$, etc., for any constant k, that is, when $t=k/\theta,2\pi/\theta+k/\theta,4\pi/\theta+k/\theta$, etc. Because the function repeats itself every $2\pi/\theta$ time units, where θ is measured in radians, it is called a *periodic function*, and the *length of the cycle* or the *period* of $\cos\theta t$ is $2\pi/\theta$ time units. This means that for each time unit the function will repeat itself $\theta/2\pi$ times, $\theta/2\pi$ being a fraction in most cases; $\theta/2\pi$ is called the

frequency of the function $\cos\theta t$ because it shows how frequently the function repeats itself per unit time. Frequency is the reciprocal of the length of the cycle.

The *amplitude* of the function $s\cos\theta t$ which is s shows the magnitude of the cyclical movements. The function may also be subject to a time shift, as given by $\cos(\theta t+\psi)$. Although $\cos\theta t$ starts with the value 1 at $t=0$, the function $\cos(\theta t+\psi)$ takes the value 1 at $t=-\psi/\theta$. Thus $\cos(\theta t+\psi)$ leads $\cos\theta t$ by ψ/θ time units, or $\cos\theta t$ lags behind $\cos(\theta t+\psi)$ by ψ/θ time units; ψ/θ is the *phase shift* in number of time units for the function $\cos(\theta t+\psi)$ and indicates a lead compared with $\cos\theta t$ if ψ is positive, a lag if ψ is negative. Note that the absolute value of ψ should be smaller than π. By convention, we do not say that $\cos(t+1.5\pi)$ leads $\cos t$ by 1.5π time units but rather that $\cos(t-.5\pi)$ lags behind $\cos t$ by $.5\pi$ time units.

To apply these concepts to the contribution (32) of a pair of complex roots $re^{i\theta}$ and $re^{-i\theta}$ to the solution of y_{1t} we note that this contribution is made up of a cosine function with cycle length equal to $2\pi/\theta$ time units or frequency equal to $\theta/2\pi$ per unit time. Its amplitude is a multiple of r^t that increases with t if the absolute value r of the roots is greater than 1 and decreases with t if the absolute value of r is smaller than 1. The phase ψ_1 of the function $\cos(\theta t+\psi_1)$ will depend on the initial conditions, discussed in the following section.

2.7 NUMERICAL EVALUATION OF THE SOLUTION

The solution of a homogeneous system of linear difference equations $y_t=Ay_{t-1}$ for the ith component of y_t has been found to be,

$$y_{it}=\sum_{k=1}^{p} b_{ik}z_{k0}\lambda_k^t=\sum_{k=1}^{p} c_{ik}\lambda_k^t, \tag{34}$$

where λ_k is the kth root of the $p\times p$ matrix A, b_{ik} is the ith element of the right characteristic vector b_k corresponding to the root λ_k, and z_{k0} is the initial value of the kth canonical variable which satisfies the univariate first-order difference equation $z_{kt}=\lambda_k z_{k,t-1}$. The solution y_{it} will be damped if all the roots $\lambda_k(k=1,\ldots,p)$ are smaller than 1 in absolute value. It will be explosive if any root λ_k (with nonzero coefficient c_{ik}) is larger than 1 in absolute value. There will be oscillations if some roots are negative or complex. A negative root gives rise to a component with two periods between successive peaks. A pair of complex roots gives rise to a component with $2\pi/\theta$ periods between peaks, θ being the angle (in radians) of the complex roots $re^{i\theta}$ and $re^{-i\theta}$. If all the roots are real and

positive, there can be some fluctuations in the linear combination (34) of the positive functions λ_k^t if the coefficients c_{ik} are partly positive and partly negative. But the fluctuations can hardly be prolonged because the behavior of the solution will eventually be dominated by the largest root as t increases. The behavior of the solution is thus mainly characterized by the characteristic roots of A as stated.

To compute the solution numerically three methods can be mentioned. The first is direct computation by using the equation $y_t=Ay_{t-1}$ successively, assuming that y_0 is known. This method requires repeated matrix multiplications. The second method employs the canonical variables discussed in Section 2.4. It is more complicated than the first, but it permits the decomposition of the solution into components of the form $c_{ik}\lambda_k^t(k=1,\ldots,p)$. For practical purposes the user needs only an efficient computer program to find the characteristic roots λ_k and the corresponding right characteristic vectors b_k of the real matrix A. The program should also provide B^{-1}. He will then have all b_{ik} and $z_{k0}=\sum_j b^{kj}y_{j0}$ to compute the coefficients c_{ik} used in the linear combination for the solution.

In the third method the solution is obtained in the form of (34) by finding the roots λ_k without going through the calculation of the characteristic vectors. By using this form, together with the initial vales of y_{it} directly, we can evaluate the coefficients c_{ik}. Consider the case of a 2×2 matrix A for illustrative purpose. The solution for y_{1t} is $c_{11}\lambda_1^t+c_{12}\lambda_2^t$. Given two values of y_{1t}, say y_{10} and y_{11}, we will have two equations,

$$c_{11}+c_{12}=y_{10},$$

$$\lambda_1 c_{11}+\lambda_2 c_{12}=y_{11},$$

for the two unknown c_{11} and c_{12}. If λ_1 and λ_2 are conjugate complex, being, say, $a+bi$ and $a-bi$, we can solve for the unknowns by letting $c_{11}=c+di$ and $c_{12}=c-di$ and substituting into the above equations.

$$c+di+c-di=y_{10}$$

$$(a+bi)(c+di)+(a-bi)(c-di)=y_{11},$$

or

$$ac-bd+(ad+bc)i+ac-bd-(ad+bc)i=y_{11}.$$

The solution is $c=y_{10}/2$ and $d=(ay_{10}-y_{11})/2b$. Similarly, if there are p roots, p initial values of y_{it} can be used to determine the p coefficients $c_{i1},\ldots,c_{ip}$.

2.8 A NUMERICAL EXAMPLE INVOLVING COMPLEX ROOTS

This section provides a numerical example of a first-order system of two difference equations with a pair of complex roots:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}. \tag{35}$$

The characteristic equation is

$$(1-\lambda)(-.80-\lambda)+1.62=\lambda^2-.20\lambda+0.82=0.$$

The characteristic roots are

$$\lambda_1=.1+.9i\cong.906e^{1.460i} \quad \text{and} \quad \lambda_2=.1-.9i\cong.906e^{-1.460i}.$$

To find the solution $y_t=A^t y_0=BD_\lambda^t B^{-1}y_0$ we apply the second method mentioned in Section 2.7 by finding the right characteristic vectors of A.

Denote the elements of the characteristic vector b_1 corresponding to λ_1 by $\alpha_1+\beta_1 i$ and $\alpha_2+\beta_2 i$; b_1 cannot be a real vector. If it were real, Ab_1 would also be real and could not be equal to $\lambda_1 b_1$ which would be complex. Writing out the equation for b_1, we have

$$\begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix} \begin{bmatrix} \alpha_1+\beta_1 i \\ \alpha_2+\beta_2 i \end{bmatrix} = (.1+.9i) \begin{bmatrix} \alpha_1+\beta_1 i \\ \alpha_2+\beta_2 i \end{bmatrix},$$

or

$$(\alpha_1+\alpha_2)+(\beta_1+\beta_2)i=(.1\alpha_1-0.9\beta_1)+(.9\alpha_1+.1\beta_1)i$$

$$(-1.62\alpha_1-.80\alpha_2)+(-1.62\beta_1-.80\beta_2)i=(.1\alpha_2-.9\beta_2)+(0.9\alpha_2+.1\beta_2)i,$$

implying four homogeneous linear equations in α_1, α_2, β_1, and β_2:

$$\begin{aligned} 0.9\alpha_1 + \alpha_2 + 0.9\beta_1 &= 0 \\ 0.9\alpha_1 - 0.9\beta_1 - \beta_2 &= 0 \\ 1.62\alpha_1 + .9\alpha_2 - .9\beta_2 &= 0 \\ .9\alpha_2 + 1.62\beta_1 + .9\beta_2 &= 0. \end{aligned}$$

These equations are not linearly independent because they have come from $(A-\lambda_1 I)b_1=0$, and $(A-\lambda_1 I)$ is singular. Note that .9 times the sum of the first and second equations gives the third equation; .9 times the difference of the second and first equations gives the fourth equation. Because a characteristic vector is determined only up to a factor of proportionality, we let $\alpha_1=1$ and $\beta_2=0$ and solve the first two equations to get $\beta_1=1$ and $\alpha_2=-1.8$.

Similarly, we can find a complex characteristic vector which corresponds to the root $\lambda_2=.1-.9i$. Its two elements are $1-i$ and -1.8, respectively. This vector is the complex conjugate of the vector associated with the first root $\lambda_1=.1+9i$, which means that each element is the complex conjugate of the corresponding element of the latter vector. In general, the *characteristic vectors corresponding to a pair of conjugate complex roots are conjugate complex*, which can be shown by doing Problem 14 of this chapter. Thus the matrix B which consists of the right characteristic vectors in our example is

$$B=\begin{bmatrix} 1+i & 1-i \\ -1.8 & -1.8 \end{bmatrix}.$$

To find the inverse of B we use the identity $B^{-1}B=I$. Denoting the first row of B^{-1} by $(\gamma_1+\delta_1 i,\ \gamma_2+\delta_2 i)$, we have

$$\gamma_1-\delta_1+(\gamma_1+\delta_1)i-1.8\gamma_2-1.8\delta_2 i=1,$$

$$\gamma_1+\delta_1+(-\gamma_1+\delta_1)i-1.8\gamma_2-1.8\delta_2 i=0,$$

which implies four linear equations in the unknowns γ_1, γ_2, δ_1, and δ_2:

$$\begin{aligned} \gamma_1 - 1.8\gamma_2 - \delta_1 &= 1, \\ \gamma_1 + \delta_1 - 1.8\delta_2 &= 0, \\ \gamma_1 - 1.8\gamma_2 + \delta_1 &= 0, \\ -\gamma_1 + \delta_1 - 1.8\delta_2 &= 0. \end{aligned}$$

The solution is $\gamma_1=0$, $\gamma_2=-3.6^{-1}$, $\delta_1=-.5$ and $\delta_2=-3.6^{-1}$. Similarly, we can obtain the second row of B^{-1} to yield the result

$$B^{-1}=\begin{bmatrix} -.5i & -3.6^{-1}-3.6^{-1}i \\ .5i & -3.6^{-1}+3.6^{-1}i \end{bmatrix}.$$

Note that the two rows of B^{-1} are conjugate complex.

Let the initial condition be $y_{10}=2.0$ and $y_{20}=3.6$. The initial condition for the canonical variables will then be

$$\begin{bmatrix} z_{10} \\ z_{20} \end{bmatrix} = \begin{bmatrix} -.5i & -3.6^{-1}-3.6^{-1}i \\ .5i & -3.6^{-1}+3.6^{-1}i \end{bmatrix} \begin{bmatrix} 2.0 \\ 3.6 \end{bmatrix} = \begin{bmatrix} -1-2i \\ -1+2i \end{bmatrix}.$$

The solution for $y_t = BD_\lambda^t z_0$ is therefore

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1+i & 1-i \\ -1.8 & -1.8 \end{bmatrix} \begin{bmatrix} (-1-2i)\lambda_1^t \\ (-1+2i)\lambda_2^t \end{bmatrix}$$

$$= \begin{bmatrix} (1-3i)\lambda_1^t+(1+3i)\lambda_2^t \\ (1.8+3.6i)\lambda_1^t+(1.8-3.6i)\lambda_2^t \end{bmatrix}$$

$$= \begin{bmatrix} 3.162e^{-1.249i}(.906e^{1.460i})^t+3.162e^{1.249i}(.906e^{1.460i})^t \\ 4.025e^{1.107i}(.906e^{1.460i})^t+4.025e^{1.107i}(.906e^{-1.460i})^t \end{bmatrix}$$

$$= \begin{bmatrix} 6.324(.906)^t\cos(1.460t-1.249) \\ 8.050(.906)^t\cos(1.460t+1.107) \end{bmatrix}.$$

Each time series is a damped cosine function with a cycle length equal to $2\pi/1.460$, or 4.30 time units, the damping factor being $(.906)^t$. The first series has smaller amplitudes than the second. It has a lag of $1.249/1.460$, or .855 time units, compared with the damped cosine function $(.906)^t \times \cos(1.460t)$. The second series has a lead of $1.107/1.460$, or .758 time units compared with the same. The first series thus lags behind the second by 1.613 time units.

In this chapter the solution to a homogeneous system of linear difference equations has been obtained and characterized in terms of the characteristic roots of the coefficient matrix A or of the solutions to the difference equations for the canonical variables. If the system is not homogeneous or if there are exogenous variables w_t affecting y_t by Bw_t, our solution $A^t y_0$ will have to be added to the cumulative effect

$$Bw_t + ABw_{t-1} + \cdots + A^{t-1}Bw_1$$

of these exogenous variables. In the special case in which w_t consists only of the dummy variable equal to 1 B will be a column vector. In any case the study of the solution to the homogeneous system is useful because it forms a part of the general solution to a nonhomogeneous system to which the effect of exogenous forces can be added. Before we consider the exogenous variables and the way to manipulate some of them to achieve desired policy objectives it is important to consider the combined effect of the random disturbance

$$u_t + Au_{t-1} + \cdots + A^{t-1}u_1$$

on the solution. This is the subject of the next two chapters.

Readers interested in pursuing the subject of nonstochastic difference equations may consult Goldberg (1958), Samuelson (1948), part II and Mathematical Appendix B, and Allen (1959). More elementary treatment with economic applications can be found in Baumol (1970) and Kenkel (1974).

PROBLEMS

1. Find the characteristic roots and the right characteristic vectors of the matrix

$$A = \begin{bmatrix} 5 & -1 \\ 2 & 2 \end{bmatrix}.$$

2. Obtain the solution to the following system of difference equations in terms of the relevant characteristic roots:

$$y_{1t} = 5y_{1,t-1} - y_{2,t-1},$$

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

Assuming that $y_{10}=0$ and $y_{20}=1$, graph the solution for y_{1t}, $t=1,2,\ldots,5$. Plot the solution for y_{2t}, $t=1,2,\ldots,5$.

3. Prove the identity (30) by Taylor series expansions of its components.

4. Under what conditions will the second-order difference equation (3) be explosive or damped and oscillatory? State your conditions in terms of the coefficients a_1 and a_2.

5. Construct a numerical example of the second-order difference equation (3) whose

solution is explosive and oscillatory. Construct an example whose solution is damped and oscillatory. Plot these solutions. Vary the initial conditions for these solutions and comment on the differences in the results.

6. Write out the characteristic equation of a 3×3 matrix $A=(a_{ij})$. Provide a numerical example and obtain the characteristic roots and associated right characteristic vectors.

7. Provide a multiplier-accelerator model that can be transformed into a second-order difference equation. Insert reasonable values for the parameters in the light of your knowledge of the economy of the United States. (If you are interested, you may actually estimate the values of the parameters by using some real data by whatever econometric method you deem appropriate.) Graph the solution by using initial conditions from actual data of the most recent periods. Discuss the nature of the solution. You may wish to consult Samuelson (1939) or Baumol (1970) for this problem.

8. Interpret the solution to Problem 2, 5, or 7 in terms of the canonical variables.

9. Let $\lambda_1=a+bi$ and $\lambda_2=c+di$, with $b\neq0$. Show that λ_1 and λ_2 are conjugate complex if both $\lambda_1+\lambda_2$ and $\lambda_1\lambda_2$ are real.

10. Let λ_1 and λ_2 be, respectively, $a+bi$ and $a-bi$, with $b\neq0$. If the linear combination $c_1\lambda_1^t+c_2\lambda_2^t$ is real for $t=1,2,\ldots$, show that c_1 and c_2 must be conjugate complex. *Hint.* Let $c_1=f_1+g_1i$ and $c_2=f_2+g_2i$. Show that $f_1=f_2$ and $g_1=-g_2$ if this linear combination is real for $t=0,1,2,\ldots$.

11. Let national income y_t be composed only of consumption y_{1t} and investment y_{2t}. Let these components satisfy, respectively, the following consumption and investment functions:

$$y_{1t}=a_1(y_{1t}+y_{2t})+b_1y_{1,t-1},$$

$$y_{2t}=a_2(y_{1t}+y_{2t})+b_2y_{2,t-1},$$

where all the coefficients a_i and b_i are positive and the sum a_1+a_2 of the marginal propensities is less than unity. Find the reduced form equations. Show that the roots of the system are real and positive and therefore that the system cannot have prolonged fluctuations. [A generalization of this proposition to the multivariate case can be found in Chow (1968).]

12. Let the investment function of the model of Problem 11 be changed to

$$y_{2t}=a_2(y_t-y_{t-1})+b_2y_{2,t-1},$$

the consumption function remaining the same as before. Construct a numerical example of this system which will give rise to a pair of complex roots. Graph your solution for five periods, using the most recent available data for consumption and investment (including government) expenditures as initial values.

13. By adding a new variable rewrite the nonhomogeneous system of difference equations $y_t=Ay_{t-1}+b$ as a homogeneous system. Obtain the solution of the nonhomogeneous system from the solution of the homogeneous system.

14. Show that the (right) characteristic vectors corresponding to a pair of conjugate complex roots $a+bi$ and $a-bi$ are conjugate complex. *Hint.* Follow the development in Section 2.8. Let the first characteristic vector be $\alpha+\beta i$, where α and β are real vectors. Write out the equation $A(\alpha+\beta i)=(a+bi)(\alpha+\beta i)$. Do the same for the second characteristic vector $\gamma+\delta i$.

15. Find the solution to the system

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}=\begin{bmatrix} 1 & -1.62 \\ 1 & -\ .80 \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}$$

by using the canonical variables, given the initial condition $y_{10}=3.6$ and $y_{20}=6.0$.

16. Find the solution to the system given by (35), using the third method described in Section 2.7, namely the initial conditions $y_{10}=2.0$ and $y_{20}=3.6$, directly without finding B or going through the canonical variables.

CHAPTER 3

Analysis of Linear Stochastic Systems: Time Domain

3.1 INTRODUCTION

In this chapter and the following we incorporate random disturbances as a part of the explanation of economic fluctuations and a part of dynamic economics. The importance of random disturbances in dynamic economics can hardly be exaggerated. All econometric models of the real world are built with explicit recognition of stochastic disturbances. It is therefore essential to incorporate this element into the theory of dynamic macroeconomics.

As early as the 1930s the Norwegian economist Ragnar Frisch emphasized the importance of random disturbances in the theory of business cycles. Partly for his contributions to stochastic dynamic economics, Frisch won the first Nobel Prize in Economic Science in 1969. The prize was shared by the Dutch economist Jan Tinbergen, who did pioneering work in econometrics and quantitative economic policy, the subject of Part 2 of this book. We quote from Frisch's classical paper [1933, pp. 197 and 202–203]:

> The examples we have discussed...show that when an [deterministic] economic system gives rise to oscillations, these will most frequently be damped. But in reality the cycles... are generally not damped. How can the maintenance of the swings be explained?... One way which I believe is particularly fruitful and promising is to study what would become of the solution of

> a determinate dynamic system if it were exposed to a stream of erratic shocks....
>
> Thus, by connecting the two ideas: (1) the continuous solution of a determinate dynamic system and (2) the discontinuous shocks intervening and supplying the energy that may maintain the swings—we get a theoretical setup which seems to furnish a rational interpretation of those movements which we have been accustomed to see in our statistical time data.

In the late 1950s, Irma Adelman and Frank Adelman (1959) used computer simulations to study the dynamic properties of the time series generated from an econometric model constructed by Klein and Goldberger (1955) and compared them with the properties of the economic time series actually observed in the United States and characterized by the work of the National Bureau of Economic Research. The method of computer simulations in economics is to solve the system of econometric equations for the values of the endogenous variables, given the values of the exogenous variables and random disturbances. If the random disturbances are excluded, it is called a nonstochastic simulation; otherwise, a stochastic simulation. The study by the Adelmans concluded that, without introducing random disturbances into the computer simulations, it was not possible to reproduce the dynamic characteristics of the economy by using the Klein-Goldberger model. With the random disturbances incorporated, the time series generated by the model looked remarkably similar to those actually observed from the viewpoint of measurements such as the mean time interval from peak to peak and trough to trough. In this chapter methods are provided to deduce some of these dynamic characteristics from a system of linear stochastic difference equations analytically rather than by computer simulations.

What are some of the dynamic characteristics of a stochastic time series that should be examined? A stochastic *time series* is a random function of time; that is, given time t, the time series is a random variable and the distribution of this random variable has time as a parameter. One important property is the mean, defined as a function of time, of each time series generated by a system of stochastic difference equations. It provides information on the trend of each series. The degree of variation around the mean should also be interesting. It may be measured by the standard deviation. Some measure of the length of the cycle may also be desirable. It could be the mean time interval from peak to peak or from trough to trough or the mean time interval when the time series crosses the trend, but there are other possible measures. We might wish to examine the correlations between successive time-series values. If the correlation between y_t

and y_{t-1} is high and positive, the time series may be considered slow moving or smooth; if the correlation is high and negative, the time series must oscillate. Besides their individual characteristics, we shall study the dynamic relations between time series. Does one time series grow faster than another? Does it fluctuate more than another? Is it made up of shorter cycles than another? Does it tend to lead or lag another or do the two tend to move up and down with approximately the same timing? These are some of the questions to be answered in this and the following chapter.

In this chapter we first study the simple case of a univariate first-order linear stochastic difference equation. Such an equation results from adding a random disturbance u_t to the right-hand side of the difference equation $y_t = ay_{t-1} + b$. Several useful concepts are defined. The discussion is generalized to a multivariate system of higher order linear difference equations. Canonical variables are also introduced to help characterize the solution, as in Chapter 2. Several dynamic characteristics concerning individual time series and the relations between them will be studied.

3.2 FIRST-ORDER LINEAR STOCHASTIC DIFFERENCE EQUATION

A simple example of a stochastic difference equation is

$$y_t = ay_{t-1} + b + u_t, \tag{1}$$

where y_t is a scalar and u_t is, for any integer t, a random variable with mean 0 and variance v; it is statistically independent of u_s for $t \neq s$. This model is not only simple but useful. In many applications of economic forecasting we may, as a crude approximation, predict the value of a variable by a linear function of its own value of the last period. The predictions from this model will almost always contain an error. The error, hopefully, is captured by the random variable u_t. Of course, we can make the model more complicated by introducing more lagged variables, y_{t-2}, y_{t-3}, and so on, and by specifying a more complicated random structure for u_t, but it is useful to consider the simplest case first. Besides being called a *first-order linear stochastic difference equation*, the model in (1) is also called a *first-order autoregression* or a *first-order autoregressive process*, for obvious reasons.

Going from a nonstochastic difference equation to a *stochastic* difference equation like (1), we would need different concepts for the solution. It is clear that the solution is no longer a deterministic function of time and the equation no longer specifies the time path of y_t exactly. It

does, however, give the probability distribution of the solution y_t for each time t. A *time series* y_t, being a stochastic function of time, is specified if we know the joint probability distribution of any subset of the random variables, say $(y_1,\ldots,y_k)$, in the same way that a scalar random variable y is specified if we know its probability distribution. Rather than dealing with all the parameters of the joint distribution of any subset of y_t that are of interest, we choose to concentrate on the *means*, the *variances*, and the *covariances*. If the joint distribution of any subset of y_t is multivariate normal, it is well known that these parameters are sufficient to determine the distribution completely. It is easy to see from Equation 2 that if u_t are normal the joint distribution of $y_1, y_2,\ldots,y_k$ (for any k), given y_0, is multivariate normal as a consequence of the theorem that linear combinations of normal random variables are jointly normal. In this case the means, variances, and covariances contain all the information required. Even if u_t are not normal, these parameters still contain much useful information that we exploit in our study of the properties of the time series.

To derive the mean of y_t, given y_0, we may first express y_t as a function of y_0 and the past u's by successive substitutions for the lagged y's on the right-hand side of (1):

$$\begin{aligned} y_t &= b + ab + a^2b + \cdots + a^{t-1}b + a^t y_0 + u_t + au_{t-1} + \cdots + a^{t-1}u_1 \\ &= b(1-a)^{-1}(1-a^t) + a^t y_0 + u_t + au_{t-1} + \cdots + a^{t-1}u_1 \end{aligned} \tag{2}$$

(provided $a \neq 1$; otherwise $bt + y_0$ replaces the sum of the first two items on the last line). Taking expectations on both sides of (2), we have

$$Ey_t = \bar{y}_t = b(1-a)^{-1}(1-a^t) + a^t y_0, \tag{3}$$

because the expectation of each u_j is 0. Alternatively, we could have obtained the mean $Ey_t \equiv \bar{y}_t$ by taking expectations on both sides of (1) to yield

$$\bar{y}_t = a\bar{y}_{t-1} + b \tag{4}$$

and then solving the nonstochastic difference equation (4). Note that $\bar{y}_0 \equiv Ey_0 = y_0$, for y_0 is treated as a given constant. Equation 3 or 4 shows that the mean function of the time series satisfying the stochastic difference equation (1) is identical with the solution to the nonstochastic difference equation $y_t = ay_{t-1} + b$, ignoring the random disturbances.

To find the variance of y_t and covariances between y_t and y_{t-k} we can

use (2) and subtract the mean $\bar{y}_t$ from both sides:

$$y_t - \bar{y}_t \equiv y_t^* = u_t + au_{t-1} + a^2u_{t-2} + \cdots + a^{t-1}u_1, \tag{5}$$

where y_t^* denotes the deviation of y_t from its mean. Similarly, we can express y_{t-k}^* as a weighted sum of u_{t-k}, u_{t-k-1},..., and u_1. Taking the expectation of the product $y_t^* y_{t-k}^*$ and noting $Eu_iu_j = 0$ for $i \neq j$, we get the covariance

$$\begin{aligned}\gamma(t,k) \equiv Ey_t^* y_{t-k}^* &= E(u_t + au_{t-1} + \cdots + a^{t-1}u_1) \\ &\quad \times (u_{t-k} + au_{t-k-1} + \cdots + a^{t-k-1}u_1) \\ &= E(a^k u_{t-k}^2 + a^{k+2}u_{t-k-1}^2 + a^{k+4}u_{t-k-2}^2 + \cdots + a^{k+2(t-k-1)}u_1^2) \\ &= va^k(1 + a^2 + a^4 + \cdots + a^{2(t-k-1)}) \\ &= v(1-a^2)^{-1}(1 - a^{2(t-k)})a^k. \end{aligned} \tag{6}$$

The variance of y_t is a special case of (6) when $k = 0$. Expression 6 is called the *autocovariance function* $\gamma(t,k)$ of the time series y_t, for it shows the covariance between y_t and its own value lagged k periods.

3.3 COVARIANCE STATIONARY TIME SERIES AND ITS AUTOCOVARIANCE FUNCTION

If the coefficient a in (1) is smaller than 1 in absolute value, the mean function (3) and the autocovariance function (6) will, as t increases, approach, respectively, the limits

$$\lim_{t\to\infty} \bar{y}_t = \bar{y} = b(1-a)^{-1}, \tag{7}$$

$$\lim_{t\to\infty} \gamma(t,k) = \gamma_k = v(1-a^2)^{-1}a^k. \tag{8}$$

In (7) and (8) we have dropped t as a subscript or argument in the mean and autocovariance functions because they are constant through time when t is sufficiently large.

A time series is called *weakly stationary* if its mean and autocovariance function are independent of time. A time series is *stationary* if the joint distribution of any subset of observations of the series remains unchanged

when the same constant is added to the time subscript of each observation. Stationarity implies weak stationarity. A time series is said to be at a *steady state* or in equilibrium if and only if it is stationary. This definition contrasts with the deterministic case in which *steady state* means unchanging values of the observations themselves. For a stochastic time series the steady state means unchanging probability distributions. A time series is *covariance stationary* if its autocovariance function is constant through time. The autocovariance, of course, is still a function of the time lag k between y_t and y_{t-k}. If the model (1) is modified by changing the intercept b into a function b_t of time, the mean function will be

$$\bar{y}_t = b_t + ab_{t-1} + \cdots + a^{t-1}b_1 + a^t y_0,$$

which may not approach a limit as t increases, but the time series will still have a constant covariance function around its mean.

Let us examine the autocovariance function γ_k given by (8) for the covariance stationary case. For $k = 0$, $\gamma_0 = v(1-a^2)^{-1}$ is the variance of the time series. It is a measure of the dispersion of the time series around its mean and can be used to provide interval prediction for the time series. The variance γ_0 is larger, the larger the variance v of the random disturbance u_t; in fact, γ_0 is proportional to v. The factor of proportionality $(1-a^2)^{-1}$ depends on the coefficient a of the stochastic difference equation (1); it is larger, the larger the absolute value of a (which should not exceed 1 for the time series to remain covariance stationary). This relationship is reasonable because the larger the coefficient a, the more the past variations in the time series will be carried over to the present; if a is 0, the only variations in y_t are from u_t, for y_t will be identical with the random series u_t itself.

If a is positive and large, say it equals .9, the covariance between y_t and its lagged value y_{t-1} will be large, being .9 of the variance of y_t. The covariance between y_t and y_{t-k} will be $(.9)^k$ of the variance; it decreases with the lag k at a slow rate; that is, although the covariance between y_t and y_{t-k} weakens as the time interval between the two observations increases, it weakens only slowly and retains substantial degrees of association between the observations at different times.

It is sometimes useful to speak of the correlation or correlation coefficient, rather than the covariance, between y_t and y_{t-k} in order to free the measure of association from the choice of units of measurement. The correlation is defined as the ratio of the covariance to the product of the standard deviations of the two variables. In the case of equal standard deviations it is the ratio of the covariance to the variance of either variable. Thus the *autocorrelation function* of a time series is the ratio of its

autovariance function to its variance. For the model in (1), in the covariance-stationary case, the autocorrelation function is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = a^k. \tag{9}$$

If a equals .9, the correlation coefficient between y_t and y_{t-k} is $(.9)^k$; it remains substantial for observations several time units apart. This means that the time series does not change rapidly from period to period. It tends to behave like a time path with long or slow-moving cycles. The meaning of this statement is clarified in Chapter 4.

If the coefficient a in (1) is small, say equals .05, the time series behaves like a random series. If a is very small, y_t is almost equal to the random disturbance u_t. From the viewpoint of the autocorrelation function y_t and y_{t-1} have a correlation coefficient of only .05, and y_t and y_{t-2} have a correlation coefficient of only .0025.

If the coefficient a is negative but large in absolute value, say equals $-.9$, successive values of y_t will tend to be highly negatively correlated, the correlation coefficient between y_t and y_{t-1} being $-.9$. The time series is certainly not random through time but rather shows a cycle of about two time units in length. It tends to go up in one period and down in the next. The movement is just opposite to that of the case in which $a = .9$. The time series now fluctuates rapidly with high frequencies rather than moving slowly. It reveals short cycles rather than long cycles. This example also suggests the important idea that *fluctuations in the autocovariance function can give indications to fluctuations in the time series itself.*

3.4* EXPECTED TIMES BETWEEN MEAN CROSSINGS AND BETWEEN MAXIMA

Because the autocovariance function provides information on the cyclical characteristics of a time series, it can be used to calculate certain measures of the average length of the cyclical movements. These measures are important because they can be used to compare the dynamic characteristics of a model with those calculated directly from economic time-series data for the purpose of checking the validity of the model. There are several ways to measure the mean length of the cycles of a stochastic time series. Two of these measures are discussed in this section. The first is the mean time interval when the time series crosses its mean function. The second is the mean time interval between successive maxima. It is certainly possible for the time series to have several maxima before crossing the

mean function so that the first measure may be expected to be greater than the second for the mean cycle length. Our discussion is confined to stationary time series. Only the results, not the proofs, of this section are required for the remainder of this book.

To obtain the expected time between successive down-crossings of the mean, that is, between crossings from an above-trend value to a below-trend value, we can find the probability at any discrete time point that the time series will change from an above-trend value to a below-trend value. If the probability is .2 per period for a time series to experience a downward movement crossing the trend, it will take an average of $1/.2$ or 5 periods to have a down-crossing. Thus the mean time for a down-crossing is the reciprocal of the probability of a down-crossing per period. To calculate the probability it is convenient to assume a 0 mean for the time series. Thus, instead of studying the time series of Equation 1, whose mean is $b(1-a)^{-1}$ in the stationary case, we shall study the deviations y_t^* from mean which satisfies the equation $y_t^* = ay_{t-1}^* + u_t$ and has a mean equal to 0. We have subtracted the mean from the time series and study the probability of down-crossing of 0 for the resulting series. The time series is assumed to be *normal.*

Let ρ_k be the autocorrelation function of a stationary time series which is normally distributed with 0 mean. In a normal time series with a constant mean covariance stationarity implies that the joint distribution of any subset of observations remains the same through time. The reason is that the mean and the autocovariance function specify the joint distribution completely. If the distribution of successive observations is invariant through time, we can consider any two successive observations, say, y_1 and y_2, and find the probability that y_1 is positive and y_2 is negative, using the bivariate normal distribution of y_1 and y_2.

$$P(y_1>0, y_2<0) = \left(2\pi\sqrt{1-\rho_1^2}\right)^{-1} \int_{-\infty}^{0}\int_{0}^{\infty} \exp\left\{-\left[2(1-\rho_1^2)\right]^{-1} \times (y_1^2 - 2\rho_1 y_1 y_2 + y_2^2)\right\} dy_1\, dy_2. \tag{10}$$

In equation 10 the variance of the time series is assumed to be 1. By changing the variables we can easily check that the double integral in (10) equals the integral for the general case in which the variance γ_0 is not 1. To argue verbally, if the original time series has a variance γ_0 not equal to 1, we divide by the standard deviation or change the unit of measurement to achieve a variance of 1. This should not affect the probability of 0 down-crossing.

To integrate the right-hand side of (10) with respect to y_1 we complete

the square for the exponent

$$y_1^2-2\rho_1 y_1 y_2+y_2^2=(y_1-\rho_1 y_2)^2+y_2^2-\rho_1^2 y_2^2 \tag{11}$$

and rewrite the integral

$$\int_0^\infty \exp\left\{-[2(1-\rho_1^2)]^{-1}(y_1-\rho_1 y_2)^2\right\}dy_1$$

$$=\sqrt{1-\rho_1^2}\int_{-\rho_1 y_2/\sqrt{1-\rho_1^2}}^\infty \exp(-\tfrac{1}{2}z^2)dz \tag{12}$$

by changing the variable y_1 to the standard normal deviate $z=(y_1-\rho_1 y_2)/\sqrt{1-\rho_1^2}$, with $dy_1=\sqrt{1-\rho_1^2}\,dz$. Substitution of (11) and (12) into (10) gives

$$P(y_1>0, y_2<0)=\frac{1}{2\pi}\int_{-\infty}^0\int_{-\rho_1 y_2/\sqrt{1-\rho_1^2}}^\infty \exp[-\tfrac{1}{2}(y_2^2+z^2)]dz\,dy_2. \tag{13}$$

Thus the desired result is the probability that the first random variable (whose value is denoted by z) will be larger than $-\rho_1/\sqrt{1-\rho_1^2}$ times the second random variable and that the second variable will be negative, given that the two variables are standard or unit normal and statistically independent.

If z is measured along the horizontal axis and y_2 along the vertical axis of a two-dimensional diagram, as in Figure 3.1, this event is represented by the set of points in the shaded area between the horizontal axis and the line $z=-\rho_1(1-\rho_1^2)^{-\frac{1}{2}}\cdot y_2$. The probability of this event is simply the ratio of the shaded area to the total area because, the two normal random variables being independent, the equal-probability contour lines are circles of different radii from the origin in Figure 3.1. In other words, the ratio of the angle θ (measured in radians) between the horizontal axis and the line $z=-\rho_1(1-\rho_1^2)^{-\frac{1}{2}}y_2$ to 2π will give the probability of down-crossing the mean. The angle θ is calculated by $\tan\theta=\rho_1^{-1}(1-\rho_1^2)^{\frac{1}{2}}$. By trigonometry it also satisfies $\cos\theta=\rho_1$. Therefore the probability of down-crossing the mean at any time unit is $\cos^{-1}\rho_1/2\pi$, and the expected time interval between *down-crossings of the mean for a stationary normal time series is* $2\pi/\cos^{-1}\rho_1$ *time units, where* ρ_1 *is the autocorrelation between* y_t *and* y_{t-1}. This expected time is obviously the same as the expected time for an up-crossing from a value below the mean to a value above because an up-crossing must be preceded by a down-crossing and vice versa.

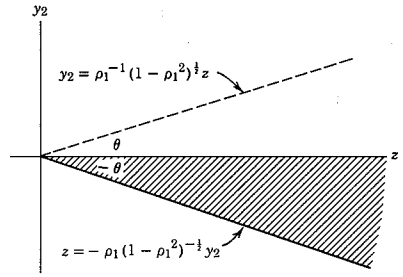

Figure 3.1 Probability of mean crossing.

Using the above formula, we find that a time series that obeys a first-order stochastic difference equation with coefficient .9 will take on the average, 13.93 time units between successive down-crossings. A coefficient of .8 would reduce the expected time between down-crossings to 9.76 time units.

A second measure of average cycle length is the expected time between relative maxima. A relative maximum occurs when the time series increases in the preceding period but decreases in the current period; that is, when $y_t>y_{t-1}$ and $y_t>y_{t+1}$. As before, we shall find the probability of this event, by assuming a multivariate normal distribution for the stationary time series with 0 mean, unit variance, and autocorrelation function ρ_k. If we define $z_1=y_1-y_0$ and $z_2=y_1-y_2$, the required probability is that of $z_1>0$ and $z_2>0$. First, the joint probability distribution of z_1 and z_2 must be derived. Being linear combinations of the normal random variables y_0, y_1, and y_2, the variables z_1 and z_2 are jointly normal. Their means are easily seen to be 0, as the means of the ys are all 0. The covariance matrix of z_1 and z_2 is

$$\operatorname{cov}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}=\begin{bmatrix} 2-2\rho_1 & 1-2\rho_1+\rho_2 \\ 1-2\rho_1+\rho_2 & 2-2\rho_1 \end{bmatrix}. \tag{14}$$

The proof (14) is left as an exercise for Problem 6 in this chapter. We now evaluate the probability that the two random variables that satisfy a bivariate normal distribution with means 0 and covariance matrix given by (14) are both positive. We can integrate the specified bivariate density over positive values for the two variables.

To simplify this procedure we change the units for the two variables z_1 and z_2 by dividing by their standard deviation, that is, by using $x_1 = z_1/(2-2\rho_1)^{\frac{1}{2}}$ and $x_2 = z_2/(2-2\rho_1)^{\frac{1}{2}}$. The probability that both z_1 and z_2 will be positive is the same as the probability that both x_1 and x_2 will be positive. The means of x_1 and x_2 are 0; their variances are 1 and their correlation coefficient is, according to (14),

$$r = \frac{1-2\rho_1+\rho_2}{2-2\rho_1}. \tag{15}$$

The required probability is therefore

$$P(x_1>0, x_2>0) = (2\pi\sqrt{1-r^2})^{-1} \int_0^\infty \int_0^\infty \exp\left\{-[2(1-r^2)]^{-1} \times (x_1^2 - 2rx_1x_2 + x_2^2)\right\} dx_1\, dx_2, \tag{16}$$

which can be evaluated by the same method as in (10). By completing the square for the exponent as in (11) and by changing the variable x_1 to $z = (x_1 - rx_2)/\sqrt{1-r^2}$ as in (12) we can rewrite (16) as the following expression, analogous to (13):

$$P(x_1>0, x_2>0) = \frac{1}{2\pi} \int_0^\infty \int_{-rx_2/\sqrt{1-r^2}}^\infty \exp\left[-\tfrac{1}{2}(x_2^2+z^2)\right] dz\, dx_2. \tag{17}$$

By the argument used for the evaluation of (13), the probability (17) is the ratio of the angle θ between the horizontal axis and the line $x_2 = -r^{-1} \times (1-r^2)^{\frac{1}{2}} z$ in the (z, x_2) diagram shown in Figure 3.2. The angle θ satisfies $\tan\theta = -r^{-1}(1-r^2)^{\frac{1}{2}}$ or, by trigonometry, $\cos\theta = -r$.

By using (15) for r we have obtained the probability $\cos^{-1}[(1-2\rho_1+\rho_2)/(2\rho_1-2)]/2\pi$ for a time series to be a maximum at any discrete time point. The *expected time between relative maxima is therefore* $2\pi/\cos^{-1}[(1-2\rho_1+\rho_2)/(2\rho_1-2)]$ *if the time series* y_t *is stationary normal and has autocorrelations* ρ_1 *and* ρ_2 *with* y_{t-1} *and* y_{t-2}, *respectively*. For a time series that satisfies (1) with $a = .9$ the expected time is 3.88, which is much shorter than the expected time between crossings of the mean.

There is a simple explanation why for many aggregate economic time series the average length of cycles, measured from peak to peak, is approximately four years. If the first difference of an annual time series is assumed to be serially independent and normally and identically distributed, the probability for the time series to be at a peak in any year, that

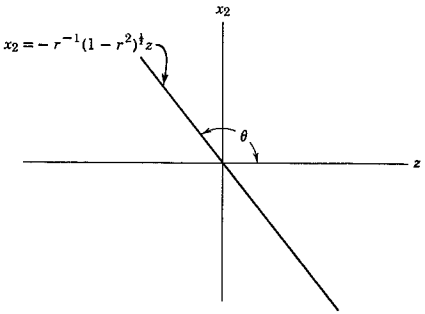

Figure 3.2 Probability of a relative maximum.

is, the probability of a positive first difference followed by a negative first difference, is one-half times one-half, or one quarter. Thus on the average a peak would be observed every four years. This result is a special case of the above formula, when $\rho_2 = \rho_1^2$ and ρ_1 is approximately equal to 1.

3.5 SYSTEMS OF LINEAR STOCHASTIC DIFFERENCE EQUATIONS

Having pointed out some important uses of the autocovariance function for a univariate time series, we shall now generalize the discussion to a situation that involves many time series generated by a system of linear stochastic difference equations:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_m y_{t-m} + b + u_t. \tag{18}$$

In (18) y_t is a multivariate or vector time series, A_i are matrices of real coefficients, b is a vector of intercepts, and u_t is a random vector with mean 0 and covariance matrix V which is statistically independent of u_s for $t \neq s$. We shall study the means of the time series as functions of time. For the study of linear models with autocorrelated disturbances u_t the reader is referred to Section 3.11.

As in the deterministic case in Chapter 2, it is convenient to rewrite (18)

as a first-order system,

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-m+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_m \\ I & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 \; I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-m} \end{bmatrix} + \begin{bmatrix} b \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{19}$$

and in a more compact form as

$$y_t = Ay_{t-1} + b + u_t. \tag{20}$$

The vector y_t in (20) stands for the vector on the left-hand side of (19), without the use of a new symbol. Similarly, b and u_t in (20) are also redefined. It is thus sufficient to study the first-order system (20).

The mean $Ey_t = \bar{y}_t$ of the time series that obeys (20) can be derived in two ways, as in the univariate time series in Equation 1. First, we eliminate the lagged y on the right-hand side of (20) successively to yield

$$y_t = b + Ab + A^2b + \cdots + A^{t-1}b + A^t y_0 + u_t + Au_{t-1} + \cdots + A^{t-1}u_1. \tag{21}$$

By taking mathematical expectations on both sides of (21) we obtain the mean function

$$Ey_t = \bar{y}_t = b + Ab + A^2b + \cdots + A^{t-1}b + A^t y_0. \tag{22}$$

Second, we take expectations of both sides of (20) and find that the mean function $\bar{y}_t$ satisfies the nonstochastic system of difference equation

$$\bar{y}_t = A\bar{y}_{t-1} + b. \tag{23}$$

The solution of (23) is (22), with $y_0 = Ey_0$.

The mean function (22) *will reach a steady state or a vector equilibrium value as t increases if and only if all the characteristic roots of the matrix A are smaller than* 1 *in absolute value*. To show this, we write the matrix A as $BD_\lambda B^{-1}$, following the discussion in Chapter 2. Here D_λ is a diagonal matrix consisting of the characteristic roots of A, and B is a matrix whose columns are the corresponding right characteristic vectors. Equation 22 can then be rewritten as

$$\begin{aligned} \bar{y}_t &= b + BD_\lambda B^{-1}b + BD_\lambda^2 B^{-1}b + \cdots + BD_\lambda^{t-1}B^{-1}b + BD_\lambda^t B^{-1}y_0 \\ &= B(I + D_\lambda + D_\lambda^2 + \cdots + D_\lambda^{t-1})B^{-1}b + BD_\lambda^t B^{-1}y_0. \end{aligned} \tag{24}$$

Note that the sum of the matrices in parentheses is a diagonal matrix with diagonal elements

$$1 + \lambda_i + \lambda_i^2 + \cdots + \lambda_i^{t-1}.$$

As t approaches infinity, the limits of the diagonal elements are $(1-\lambda_i)^{-1}$ if and only if the absolute values of all λ_i are smaller than 1. Under this assumption the matrix in parentheses in (24) will approach $(I - D_\lambda)^{-1}$ as the limit, and D_λ^t will approach a matrix of zeroes as t approaches infinity, making the last term of (24) vanish. The steady-state value of $\bar{y}_t$ is

$$\lim_{t\to\infty} \bar{y}_t = B(I - D_\lambda)^{-1}B^{-1}b. \tag{25}$$

An alternative, and simpler, way to write the steady state (25) is obtained by using the inverse of $B(I - D_\lambda)B^{-1} = I - A$, yielding

$$\lim_{t\to\infty} \bar{y}_t = (I - A)^{-1}b. \tag{26}$$

In connection with our proof of (25) and (26) it has been shown that if and only if all the characteristic roots of the matrix A are smaller than 1 in absolute value

$$\lim_{t\to\infty} (I + A + A^2 + \cdots + A^{t-1}) = B(I - D_\lambda)^{-1}B^{-1} = (I - A)^{-1} \tag{27}$$

and

$$\lim_{t\to\infty} A^t = 0. \tag{28}$$

The identity (27) can also be obtained in the following manner: let the finite sum of t terms $I + A + A^2 + \cdots + A^{t-1}$ be denoted by S_t. It is easy to see that $S_t - AS_t = I - A^t$, which implies that

$$S_t = I + A + A^2 + \cdots + A^{t-1} = (I - A)^{-1}(I - A^t), \tag{29}$$

itself a useful identity. By using (28) we find that the limit of (29) as t approaches infinity is $(I - A)^{-1}$.

3.6 AUTOCOVARIANCE MATRIX OF STOCHASTIC DIFFERENCE EQUATIONS

Having considered the means, we shall now study the variances and covariances of y_{it} and $y_{j,t-k}$. We have already dealt with the interpretation

and uses of the autocovariance function of an individual time series. As for the covariances between y_{it} and $y_{j,t-k}$ for different values of k, they indicate the degrees of association between the two time series at different times and are called *cross-covariances*. If the covariance between y_{it} and y_{jt} is large, the two time series observed at the same time are highly associated. To eliminate the arbitrary effects of the units of measurement *cross-correlations* between y_{it} and $y_{j,t-k}$ are used and defined, as usual, as ratios of the cross-variances to the products of the respective standard deviations, that is, to $\sqrt{\text{var}y_{it}\cdot\text{var}y_{j,t-k}}$. The cross-covariance or the cross-correlation function also reveals the lead-lag relationships between the two time series. Its value, for example, may be highest for $k=3$, which suggests that y_{it} and $y_{j,t-3}$ are most highly correlated or that y_{jt} perhaps leads y_{it} by about three time units.

To study the autocovariances and cross-covariances of a set of interrelated time series it is useful to define a matrix function whose $i-j$ element is $\text{cov}(y_{it}, y_{j,t-k})$; that is,

$$\Gamma(t,k)=[\text{cov}(y_{it},y_{j,t-k})]=[\gamma_{ij}(t,k)]. \tag{30}$$

The diagonal elements of (30) are autocovariance functions of the individual time series. Equation 30 is a matrix generalization of the scalar function defined by (6) and is called the *autocovariance matrix* of the vector time series y_t.

For the time series that obeys system (20) the autocovariance matrix can be derived as follows. First, define $y_t^*=y_t-\bar{y}_t$ as the random vector of deviations of the time series from their means. By subtracting (23) from (20) we find that

$$y_t^*=Ay_{t-1}^*+u_t, \tag{31}$$

which implies

$$y_t^*=u_t+Au_{t-1}+A^2u_{t-2}+\cdots+A^{t-1}u_1, \tag{32}$$

where we assume that $y_0=\bar{y}_0$ is a vector of constants, which implies $y_0^*=0$. The autocovariance matrix $\Gamma(t,k)$ for $k\geqslant 0$ is

$$\begin{aligned}E(y_t^*y_{t-k}^{*\prime})&=E(u_t+Au_{t-1}+\cdots+A^{t-1}u_1)\\&\quad\times(u_{t-k}'+u_{t-k-1}'A'+\cdots+u_1'A'^{t-k-1})\\&=E(A^ku_{t-k}u_{t-k}'+A^{k+1}u_{t-k-1}u_{t-k-1}'A'+\cdots+A^{t-1}u_1u_1'A'^{t-k-1})\\&=A^kV+A^{k+1}VA'+\cdots+A^{t-1}VA'^{t-k-1}\qquad(k\geqslant 0).\end{aligned} \tag{33}$$

In (33) the relations $Eu_tu_s'=0$ for $(t\neq s)$ and $Eu_tu_t'=V$ have been used.

A necessary and sufficient condition for the autocovariance matrix (33) to approach a limit for any nonnegative integer k, as t approaches infinity, is that all the characteristic roots of A are smaller than one in absolute value. To prove this theorem

$$\begin{aligned}\Gamma(t,k)&=E(y_t^*y_{t-k}^{*\prime})=A^k[V+AVA'+A^2VA'^2+\cdots+A^{t-k-1}VA'^{t-k-1}]\\&=A^k[V+BD_\lambda B^{-1}VB'^{-1}D_\lambda B'+\cdots+BD_\lambda^{t-k-1}B^{-1}VB'^{-1}D_\lambda^{t-k-1}B']\\&=A^k[B(B^{-1}VB'^{-1}+D_\lambda B^{-1}VB'^{-1}D_\lambda+\cdots\\&\qquad+D_\lambda^{t-k-1}B^{-1}VB'^{-1}D_\lambda^{t-k-1})B']\\&=A^k[B(W+D_\lambda WD_\lambda+D_\lambda^2WD_\lambda^2+\cdots+D_\lambda^{t-k-1}WD_\lambda^{t-k-1})B'],\end{aligned} \tag{34}$$

where we have defined

$$W=B^{-1}VB'^{-1}\quad\text{or}\quad BWB'=V. \tag{35}$$

Note that the $i-j$ element of the matrix in parentheses in the last line of (34) is

$$w_{ij}+w_{ij}\lambda_i\lambda_j+w_{ij}(\lambda_i\lambda_j)^2+\cdots+w_{ij}(\lambda_i\lambda_j)^{t-k-1}=w_{ij}(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^{t-k}\lambda_j^{t-k}). \tag{36}$$

The elements of this matrix will therefore approach limits equal to

$$w_{ij}(1-\lambda_i\lambda_j)^{-1} \tag{37}$$

if and only if all roots are smaller than 1 in absolute value. This proves the theorem.

While proving the important theorem of the last paragraph we have also provided an explicit expression for the autocovariance matrix $\Gamma(t,k)$ of a system of linear stochastic difference equations in the form of (20). To repeat the last line of (34), using (36) also, we write

$$\Gamma(t,k)=A^k\left\{B\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^{t-k}\lambda_j^{t-k})\right]B'\right\}\qquad(k\geqslant 0), \tag{38}$$

where w_{ij} is defined by (35) and where the expression in brackets stands for the $i-j$ element of the matrix; B, as before, denotes a matrix consisting of columns of the right characteristic vectors of A. For $k=0$, in particular, we have a covariance matrix of the contemporaneous variables $y_{1t},\ldots,y_{pt}$:

$$\Gamma(t,0)=B\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^t\lambda_j^t)\right]B'. \tag{39}$$

Equation 38 also shows that the autocovariance matrix satisfies the matrix difference equation in the variables t and k; that is,

$$\Gamma(t,k)=A\Gamma(t-1,k-1)=\cdots=A^k\Gamma(t-k,0), \quad (k\geqslant 0) \tag{40}$$

which can be used for the computation of $\Gamma(t,k)$ from $\Gamma(t-k,0)$, the latter given by (39).

If, and only if, all the roots of A are smaller than 1 in absolute value, the autocovariance matrix approaches a limit as t increases. The multivariate time series y_t is then *covariance-stationary*. The autocovariance function, at the steady state, can be written as

$$\lim_{t\to\infty}\Gamma(t,k)\equiv\Gamma_k=A^k\Gamma_0=A^k\left\{B\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}\right]B'\right\} \qquad (k\geqslant 0). \tag{41}$$

If we do not wish to use the characteristic roots and vectors of A to compute Γ_0, we can go back to the first line of (34) and use

$$\Gamma(t,0)=V+AVA'+\cdots+A^{t-1}VA'^{t-1}. \tag{42}$$

For the steady-state of a covariance-stationary vector time series, we use the infinite series

$$\lim_{t\to\infty}\Gamma(t,0)\equiv\Gamma_0=V+AVA'+A^2VA'^2+\cdots. \tag{43}$$

The additional terms on the right-hand side of (43) will get smaller and eventually become negligible if the roots of A are all smaller than 1 in absolute value. (43) can thus be used to compute Γ_0.

Explicit expressions have been given to evaluate the autocovariance matrix of a linear stochastic system. It suffices to evaluate the matrices for nonnegative integers k, using (40) and (42) for the general case and (41) and (43) for the stationary case. If k is negative, we can utilize the identity

$$\operatorname{cov}(y_{it},y_{j,t-k})=\operatorname{cov}(y_{j,t-k},y_{i,t-k-(-k)}),$$

or in matrix terms

$$\Gamma(t,k)=\Gamma'(t-k,-k), \tag{44}$$

and for the steady state

$$\Gamma_k=\Gamma'_{-k}. \tag{45}$$

In particular, for the steady state

$$\gamma_{ii,k}\equiv\operatorname{cov}(y_{it},y_{i,t-k})=\gamma_{ii,-k}. \tag{46}$$

Equation 46 simply means that for any individual time series at the steady state the covariance between y_{it} and $y_{i,t-k}$ is the same as the covariance between y_{it} and $y_{i,t+k}$ because the observations in either pair are k time units apart.

3.7 THE AUTOCOVARIANCE MATRIX VIA CANONICAL VARIABLES

The derivation of the autocovariance matrix of a system of linear stochastic difference equations is intimately tied up with characteristic roots and vectors and therefore not surprisingly with the canonical variables of the system. Recall the construction of canonical variables in Chapter 2 as linear combinations of the original variables y_t; that is,

$$z_t=B^{-1}y_t \quad \text{or} \quad y_t=Bz_t. \tag{47}$$

The construction can be applied whether y_t is stochastic or not.

For stochastic y_t, taking expectations on both sides of either equation in (47), we get

$$Ez_t\equiv\bar{z}_t=B^{-1}\bar{y}_t \quad \text{or} \quad \bar{y}_t=B\bar{z}_t; \tag{48}$$

that is, the mean vector of z_t consists of the same linear combinations of the elements of the mean vector of y_t as z_t itself of y_t. Also, the deviations of z_t from mean are the same linear combinations of the deviations of y_t from mean as

$$z_t^*\equiv z_t-\bar{z}_t=B^{-1}(y_t-\bar{y}_t)=B^{-1}y_t^* \quad \text{or} \quad y_t^*=Bz_t^*. \tag{49}$$

Therefore the autocovariance matrices of z_t and y_t are related by

$$\begin{aligned}&\Gamma_z(t,k)=Ez_t^*z_{t-k}^{*\prime}=B^{-1}(Ey_t^*y_{t-k}^{*\prime})B^{-1\prime}=B^{-1}\Gamma_y(t,k)B^{-1\prime};\\ &\Gamma_y(t,k)=B\Gamma_z(t,k)B'.\end{aligned} \tag{50}$$

We could have derived the autocovariance matrix of y_t from that of z_t by using (50). The latter matrix is easy to obtain. By premultiplying (31) by B^{-1} and using (49), we have

$$z_t^*=D_\lambda z_{t-1}^*+v_t, \tag{51}$$

where v_t is defined as

$$v_t=B^{-1}u_t \tag{52}$$

and has a covariance matrix

$$W \equiv Ev_t v_t' = B^{-1}(Eu_t u_t')B^{-1\prime} = B^{-1}VB^{-1\prime} = (w_{ij}). \tag{53}$$

Each component of (51) satisfies a first-order stochastic difference equation

$$z_{it}^* = \lambda_i z_{i,t-1}^* + v_{it} = v_{it} + \lambda_i v_{i,t-1} + \lambda_i^2 v_{i,t-2} + \cdots + \lambda_i^{t-1} v_{i1}. \tag{54}$$

By taking expectations of the products $z_{it}^* z_{j,t-k}^*$ the auto- and cross-covariances of the z_{it} and z_{jt} are

$$Ez_{it}^* z_{j,t-k}^* = w_{ij}\lambda_i^k\left(1 + \lambda_i\lambda_j + \lambda_i^2\lambda_j^2 + \cdots + \lambda_i^{t-1}\lambda_j^{t-1}\right)$$

$$= w_{ij}\lambda_i^k(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^t\lambda_j^t). \tag{55}$$

Therefore by using (55) and (50) we obtain the autocovariance matrix of the original variable y_t as

$$\Gamma_y(t,k) = B\left[w_{ij}\lambda_i^k(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^t\lambda_j^t)\right]B'$$

$$= BD_\lambda^k\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^t\lambda_j^t)\right]B'$$

$$= A^k B\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}(1-\lambda_i^t\lambda_j^t)\right]B', \tag{56}$$

which is identical with the result in (38) obtained without explicit mention of the canonical variables.

3.8 A RELATION BETWEEN STOCHASTIC AND NONSTOCHASTIC TIME SERIES

In Chapter 2 it was found that the solution of a system of nonstochastic linear difference equations is a linear combination of the characteristic roots λ_i, each raised to a power t. This is so because each of the solutions $z_{jt} = z_{j0}\lambda_j^t$ of the canonical variables is a multiple of λ_j^t and the solution of the original variable y_{it} is a linear combination of the above solutions, that is, $\sum_j b_{ij} z_{j0}\lambda_j^t$.

For a stochastic time series y_{it} we can no longer think in terms of a

deterministic time path for the variable itself. It would be useful to think of its autocovariance function, which is a function of the time lag k. If y_{it} is a weighted sum $\sum_j b_{ij} z_{jt}$ of the canonical variables, the covariance between y_{it} and $y_{i,t-k}$ will be a weighted sum of the auto- and cross-covariances of the canonical variables, all with lag k:

$$Ey_{it}^* y_{i,t-k}^* = \sum_{j,m} b_{ij} b_{im} Ez_{jt}^* z_{m,t-k}^*. \tag{57}$$

Equation 57 is simply the ith diagonal element of the second line of (50), but we know, from (55), that the auto- and cross-covariances between the jth and other canonical variables lagged k time units are multiples of λ_j^k. Therefore by (57) and (55) the autocovariance function of y_{it} is a linear combination of λ_j^k:

$$Ey_{it}^* y_{i,t-k}^* = \sum_{j,m} b_{ij} b_{im} w_{jm}(1-\lambda_j\lambda_m)^{-1}(1-\lambda_j^t\lambda_m^t)\lambda_j^k$$

$$= \sum_{j,m} b_{ij} b_{im}(Ez_{jt}^* z_{mt}^*)\cdot\lambda_j^k. \tag{58}$$

In the covariance stationary case the autocovariance function is reduced to

$$Ey_{it}^* y_{i,t-k}^* = \sum_{j,m} b_{ij} b_{im} w_{jm}(1-\lambda_j\lambda_m)^{-1}\cdot\lambda_j^k$$

$$= \sum_{j,m} b_{ij} b_{im}(Ez_{jt}^* z_{mt}^*)\cdot\lambda_j^k. \tag{59}$$

Therefore, whereas the ith deterministic time path from a system of nonstochastic difference equations is a linear combination of λ_j^t $(j=1,\ldots,p)$, the autocovariance function of the ith time series from a system of stochastic difference equations is a linear combination of λ_j^k $(j=1,\ldots,p)$.

3.9 PERIODICITY IN THE AUTOCOVARIANCE FUNCTION

The result just obtained with the autocovariance function has an interesting implication. If a pair of complex roots λ_1 and λ_2 exist, being, respectively, $re^{i\theta}$ and $re^{-i\theta}$, they would in the deterministic case contribute a component to the solution of the ith time path, which is a multiple of a cosine function with cycle length equal to $2\pi/\theta$, as we pointed out in Chapter 2. In the stochastic case the pair of complex roots would contribute a component to the autocovariance function of the ith time series, which is also a multiple of a cosine function with cycle length equal to

$2\pi/\theta$. By (58) or (59) the component is

$$b_{i1}\sum_m b_{im}(Ez_{1t}^*z_{mt}^*)\lambda_1^k + b_{i2}\sum_m b_{im}(Ez_{2t}^*z_{mt}^*)\lambda_2^k = 2q(t)r^k\cos[\theta k+\Phi(t)], \quad (60)$$

where the two coefficients of λ_1^k and λ_2^k are written as $q(t)e^{i\Phi(t)}$ and $q(t)e^{-i\Phi(t)}$.

If the absolute value r of the roots is greater than 1, the contribution to the time path in the deterministic case will be magnified by r^t through time; if the absolute value is smaller than 1, the combination will be damped, as pointed out in Section 2.5. In the stochastic case, if r is greater than 1, the contribution (60) to the autocovariance function of y_{it} will be a function of t because the terms $Ez_{1t}^*z_{mt}^*$ and $Ez_{2t}^*z_{mt}^*$ in (58) and (60) are in general functions of t. If r is smaller than 1 and all other roots are also smaller than 1 in absolute value, the contribution (60) to the autocovariance function is constant through time because the terms $1-\lambda_l^t\lambda_m^t$ and $1\lambda_2^t\lambda_m^t$ in (58) will approach unity. From (60) this contribution will be a damped cosine function $2qr^k\cos(\theta k+\Phi)$ of the lag k. Thus the angle θ of a pair of complex roots tells something about the length of the cycles in both the deterministic and the stochastic cases. If the autocovariance function shows high degrees of association between observations of a time series $2\pi/\theta$ time units apart, we may be tempted to say that the time series itself has important cycles of lengths approximately equal to $2\pi/\theta$. The statement implies another possible measure to the length of cycles for a stochastic time series via the angle of the complex roots. The elaboration of this statement, and a general discussion of time series in terms of their cyclical components will be found in the following chapter.

3.10 A NUMERICAL EXAMPLE

This section provides a numerical example of an autocovariance matrix for a first-order system of two stochastic difference equations:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \quad (61)$$

where the covariance matrix V of u_{1t} and u_{2t} is assumed to be the identity matrix. This example is obtained by adding the random disturbances to the nonstochastic equations in (35) in Section 2.8. We shall evaluate the

autocovariance matrix Γ_k by $BD_\lambda^k B^{-1}\Gamma_0$, as Section 2.8 already contains the roots λ_1 and λ_2 and the matrices B and B^{-1}.

To obtain Γ_0 we can use (43):

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix}\begin{bmatrix} 1 & -1.62 \\ 1 & -.80 \end{bmatrix}$$

$$+\begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix}^2\begin{bmatrix} 1 & -1.62 \\ 1 & -.80 \end{bmatrix}^2 + \cdots$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -2.42 \\ -2.42 & 3.2644 \end{bmatrix} + \begin{bmatrix} .4244 & .00488 \\ .00488 & 1.065376 \end{bmatrix}$$

$$+\begin{bmatrix} 1.499{,}536 & -1.551{,}638 \\ -1.551{,}638 & 1.808{,}285 \end{bmatrix} + \cdots.$$

This series will converge, but not too rapidly because the absolute value of the roots is .906. We can apply (56) alternatively by first finding the covariance matrix W of the residuals in the autoregressions for the canonical variables:

$$W = B^{-1}VB^{-1\prime} = \begin{bmatrix} -.5i & -3.6^{-1}-3.6^{-1}i \\ .5i & -3.6^{-1}+3.6^{-1}i \end{bmatrix}$$

$$\times\begin{bmatrix} -.5i & .5i \\ -3.6^{-1}-3.6^{-1}i & -3.6^{-1}+3.6^{-1}i \end{bmatrix}$$

$$= \begin{bmatrix} -.25+.1543i & .4043 \\ .4043 & -.25-.1543i \end{bmatrix},$$

where the matrix B^{-1} is taken from Section 2.8. Because the residuals are complex, their variances can be complex. At the steady state the

covariance matrix of the canonical variables is

$$\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}\right]=\begin{bmatrix}(-.25+.1543i)(1.80-.18i)^{-1} & .4043(.18)^{-1}\\ .4043(.18)^{-1} & (-.25-.1543i)(1.80+.18i)^{-1}\end{bmatrix}$$

$$=\begin{bmatrix}.2938e^{-.5530i}(1.809e^{-.09967i})^{-1} & 2.246\\ 2.246 & .2938e^{.5530i}(1.809e^{.09967i})^{-1}\end{bmatrix}$$

$$=\begin{bmatrix}.1624e^{-.4533i} & 2.246\\ 2.246 & .1624e^{.4533i}\end{bmatrix}.$$

The autocovariance matrix of y_{1t} and y_{2t} is $BD_\lambda^k\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}\right]B'$ by the second line of (56):

$$\begin{bmatrix}1+i & 1-i\\ -1.8 & -1.8\end{bmatrix}\begin{bmatrix}(.1460-.07112i)\lambda_1^k & 2.246\lambda_1^k\\ 2.246\lambda_2^k & (.1460+.07112i)\lambda_2^k\end{bmatrix}\begin{bmatrix}1+i & -1.8\\ 1-i & -1.8\end{bmatrix}$$

$$=\begin{bmatrix}(4.634+.292i)\lambda_1^k & (-4.434-4.178i)\lambda_1^k\\ +(4.634-.292i)\lambda_2^k & +(-4.434+4.178i)\lambda_2^k\\ (-4.434+3.908i)\lambda_1^k & (7.750-.230i)\lambda_1^k\\ +(-4.434-3.908i)\lambda_2^k & +(7.750+.230i)\lambda_2^k\end{bmatrix}.$$

The autocovariance function $\gamma_{11,k}$ of the first time series is

$$Ey_{1t}^*y_{1,t-k}^*=4.643e^{.0629i}(.9055e^{1.460i})^k+4.643e^{-.0629i}(.9055e^{-1.460i})^k$$

$$=9.286(.9055)^k\cos(1.460k+.0629)\qquad(k\geqslant 0).$$

The autocovariance function $\gamma_{22,k}$ of the second time series is

$$Ey_{2t}^*y_{2,t-k}^*=7.754e^{-.0297i}(.9055e^{1.460i})^k+7.754e^{.0297i}(.9055e^{-1.460i})^k$$

$$=15.508(.9055)^k\cos(1.460k-.0297)\qquad(k\geqslant 0).$$

By evaluating these functions at $k=0$ we obtain the variances of the two time series, 9.268 and 15.501, respectively. Both autocovariance functions are damped cosine functions with frequency $1.460/2\pi$ or cycle length of 4.3 time units. The periodicity is the same as that of the damped cosine functions showing the time paths of the deterministic series in Section 2.8.

The cross-covariance function $\gamma_{12,k}$ is

$$Ey_{1t}^*y_{2,t-k}^*=6.092e^{.756i}(.9055e^{1.460i})^k+6.092e^{-.756i}(.9055e^{-.1460i})^k$$

$$=12.184(.9055)^k\cos(1.460k+.756)\qquad(k\geqslant 0).$$

The cross-covariance function $\gamma_{21,k}$ is

$$Ey_{2t}^*y_{1,t-k}^*=5.910e^{-.722i}(.9055e^{1.460i})^k+5.910e^{.722i}(.9055e^{-1.460i})^k$$

$$=11.820(.9055)^k\cos(1.460k-.722)\qquad(k\geqslant 0).$$

By evaluating either of these functions at $k=0$ we obtain the contemporaneous covariance 8.87, between the two time series. The correlation coefficient between these two series is 8.87 $(9.268\times 15.501)^{\frac{1}{2}}$, or .740. This fairly high correlation is not surprising in view of the way the two series are interrelated in the model (61).

3.11 TREATMENT OF AUTOCORRELATED RESIDUALS

This section provides a method of converting a system of linear stochastic difference equations with autocorrelated residuals u_t to a system with serially uncorrelated residuals. This method of conversion can be applied not only for the study of the dynamic properties in this and later chapters but also for the solution to optimal control problems from Chapter 7 on. It is therefore convenient to consider the linear model

$$y_t=Ay_{t-1}+Cx_t+b+u_t, \tag{62}$$

which is made more general than the model (20) by including a vector of exogenous variables x_t, some of which may be subject to the control of a policy maker. Assume that a higher order system to begin with has been transformed to first-order and that lagged exogenous variables are also eliminated by identities (see Section 7.2). The following method has been used in Pagan (1973).

Consider first the case of a residual vector u_t which satisfies an autore-

gressive system

$$u_t = \Phi_1 u_{t-1} + \cdots + \Phi_q u_{t-q} + e_t = (\Phi_1 \cdots \Phi_q) \begin{bmatrix} u_{t-1} \\ \vdots \\ u_{t-q} \end{bmatrix} + e_t, \tag{63}$$

where e_t is serially uncorrelated. The system (63) can be transformed into first-order by writing

$$\begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-q+1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_q \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \quad 0 \end{bmatrix} \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-q} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{64}$$

and by using (64) to define the first-order system

$$v_t = F v_{t-1} + e_t^*. \tag{65}$$

The definition of v_t also permits us to write

$$u_t = (\Phi_1 \cdots \Phi_q) v_{t-1} + e_t = \Phi v_{t-1} + e_t. \tag{66}$$

By substituting the right-hand side of (66) for u_t in the original system (62), we have

$$y_t = A y_{t-1} + \Phi v_{t-1} + C x_t + b + e_t, \tag{67}$$

and by combining this result with (65) we form a first-order system with serially uncorrelated residuals:

$$\begin{bmatrix} y_t \\ v_t \end{bmatrix} = \begin{bmatrix} A & \Phi \\ 0 & F \end{bmatrix} \begin{bmatrix} y_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} C \\ 0 \end{bmatrix} x_t + \begin{bmatrix} b \\ 0 \end{bmatrix} + \begin{bmatrix} e_t \\ e_t^* \end{bmatrix}. \tag{68}$$

If the residual vector u_t satisfies a moving average process

$$u_t = e_t + \Phi_1 e_{t-1} + \cdots + \Phi_q e_{t-q} = e_t + \Phi v_{t-1}^*, \tag{69}$$

where the e_t are serially uncorrelated as before, we observe that v_t^* is specified by

$$v_t^* = \begin{bmatrix} e_t \\ e_{t-1} \\ \vdots \\ e_{t-q+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \quad 0 \end{bmatrix} \begin{bmatrix} e_{t-1} \\ e_{t-2} \\ \vdots \\ e_{t-q} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{70}$$

or by the first-order equation

$$v_t^* = F^* v_{t-1}^* + e_t^*. \tag{71}$$

Substituting the right-hand side of (69) for u_t in (62) and combining the result with (71) will form a first-order system in y_t and v_t^* with serially uncorrelated residuals.

Numerous references to the subject of time series relevant to various parts of this chapter could be cited. Among the many books on time series analysis in general perhaps mention should be made of Anderson (1971) and Box and Jenkins (1970). Both books cover much more ground than this chapter and contain in particular material on the statistical estimation of parameters in time series models, a subject that has been ignored here, and other kinds of time series models than systems of stochastic difference equations. One important early contribution to time series analysis is Quenouille (1947), but it was not written as a textbook.

On the subject of expected times between down-crossings of the mean and between relative maxima the reader may refer to Kendall (1945). These measurements were used by Johnston (1955) to study the average duration of business cycles in the United States. Howrey (1968) provided an economic application of several measurements of average cycle length, and Chow (1968) made an attempt to integrate stochastic elements into business cycle theory, using the tools of this chapter.

PROBLEMS

1. Plot the mean function of the time series $y_t = .9y_{t-1} + 1 + u_t$, given that the variance of u_t is $\frac{1}{2}$ and given $y_0 = 10$. What is the stationary value for the mean? When will this stationary value be a good approximation to the mean.

2. Plot the autocovariance function $\gamma(t,k)$, $0 \leq k \leq 3$, $t = 1,2,3,\ldots$, for the time series specified by Problem 1. What is the autocovariance function when the steady state is reached? How long does it take to reach the steady state, approximately? What is the autocorrelation function at the steady state?

3. Plot the stationary autocorrelation function for the time series $y_t = ay_{t-1} + u_t$, given that the variance of u_t is 1, for $a = .8$, $a = .2$, $a = 0$, $a = -.2$, and $a = -.8$. Comment on the differences.

4. What are the expected times between down-crossings of the mean for the time series specified in Problem 3, assuming u_t to be normal?

5. What are the expected times between maxima for the time series specified in Problem 3, assuming u_t to be normal?

6. Let y_0, y_1, and y_2 be jointly normal with means 0 and variances 1. The correlation between y_0 and y_1 and between y_1 and y_2 are both ρ_1; the correlation between y_0 and y_2 is ρ_2. Define $z_1 = y_1 - y_0$ and $z_2 = y_1 - y_2$. Show that the covariance matrix of z_1 and z_2 is that given by Equation 14.

7. Let y_t satisfy the second-order stochastic difference equation

$$y_t = .9y_{t-1} - .8y_{t-2} + u_t,$$

where u_t has mean 0, variance 1, and is uncorrelated with u_s for $t \neq s$. Find the autocorrelation function of y_t in the steady state, using (41) and (43). Plot this function for $k = 0, \pm 1, \pm 2, \ldots$, and other selected values.

8. Find the canonical variables of the equation in Problem 7.

9. Express the autocorrelation function of the time series in Problem 7 in terms of the characteristic roots and as a modified cosine function of the lag k.

10. What is the expected time interval between down-crossings of the mean for the time series specified in Problem 7?

11. What is the expected time interval between relative maxima for the time series specified in Problem 7?

12. What is your estimate of the approximate length of cycles for the time series specified in Problem 7, using the angle of the pair of complex roots?

13. Compare the answers to Problems 10, 11, and 12, or any pair of them. Comment on the differences.

14. Compare the behavior of the time series

$$y_t = .99y_{t-1} - .99y_{t-2} + u_t$$

with the time series specified in Problem 7 in terms of any one of the following:

a. the autocorrelation function;
b. the associated canonical variables and the difference equations for them;
c. the length of cycles in terms of the angle of the roots;
d. the expected time between down-crossings of mean;
e. the expected time between maxima.

15. Construct a time series which, when compared with that of Problem 7, will have longer cycles. Check your answer.

16. Find the autocovariance matrix of the system

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -1.62 \\ 1 & -.80 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix},$$

where the covariance matrix V of u_{1t} and u_{2t} is assumed to be the identity matrix.

CHAPTER 4

Analysis of Linear Stochastic Systems: Frequency Domain

In Chapter 3 it was pointed out that periodic or approximately periodic movements in the autocovariance function may reveal periodic movements in the time series itself. In this chapter the discussion of the dynamic properties of stochastic time series stresses their periodic or cyclical movements, both individually and in relation to one another. Because the notion of the autocorrelation function is already familiar, it is used first to derive a spectral density function for the time series. This function shows the importance of different periodic movements of which the time series is composed. A time series is viewed as a weighted sum of many cosine or sine functions of time with different periodicities or frequencies. The spectral density, being a function of frequency, measures the importance of the cosine function of that frequency as a component of the time series. The analysis of a time series through its periodic components of different frequencies is said to be an analysis in the *frequency domain*.

Having obtained a spectral density function by way of the autocovariance function of a time series, we then perform a similar operation on the cross-covariance function to obtain a cross-spectral density function to show the relation between the cyclical movements of two time series. A time series is defined as a weighted sum of many periodic or frequency components. Spectral and cross-spectral densities are defined directly in terms of the time series and the relations between two time series without using the autocovariance and cross-covariance functions. The two definitions are shown to be equivalent. The spectral and cross-spectral densities

of time series generated by a system of linear stochastic difference equations are also derived. Throughout the discussion, the time series is assumed to be covariance stationary, implying, in the case of linear stochastic difference equations, that all roots are smaller than 1 in absolute value. Traditionally, spectral and cross-spectral densities are defined only for covariance stationary time series. We postpone the definition and derivation of these functions for linear systems with roots greater than 1 in absolute value until Chapter 6. It will then be possible to study stochastic systems of cyclical growth by the useful concepts of spectral and cross-spectral densities.

4.1 SPECTRAL DENSITY FUNCTION VIA THE AUTOCOVARIANCE FUNCTION

As suggested in Chapter 3, periodic or approximately periodic movements in the autocovariance function reveal periodic movements of similar frequencies in the time series itself. A striking illustration is the time series $y_t = -.9y_{t-1} + u_t$, where u_t is random, with mean 0 and variance 1, and uncorrelated with u_s for $t \neq s$. We can imagine that this series moves up and down from one time unit to the next, exhibiting an approximately periodic movement of about two time units in length. It is only approximately periodic because there are random disturbances contributing to the series and because only .9 of the preceding observation repeats itself, with opposite sign, in each time unit. Accordingly, it is not precisely a periodic function of exactly two time units in cycle length but only approximately so.

What about its autocovariance function γ_k? By the method in Chapter 3 $\gamma_k = (-.9)^k \gamma_0 (k = 0, 1, 2, \ldots)$. Thus γ_k, as a function of the time lag k, satisfies the difference equation $\gamma_k = -.9\gamma_{k-1}$. This difference equation is identical to the difference equation $y_t = -.9y_{t-1}$ for the original variable y_t in time t omitting the random disturbance u_t; γ_k must behave like y_t of the nonstochastic difference equation $y_t = -.9y_{t-1}$. Note that the initial conditions γ_0 and y_0 differ: γ_0, the variance in the stochastic time series at the steady state is $(1-.9^2)^{-1}$ for the above example; y_0 is the value of the deterministic series at time 0, whatever it may be. In any case, periodicities in the autocovariance function do reveal similar periodicities in the time series itself.

This point can be generalized to a time series generated by a system of linear stochastic difference equations such as the first component y_{1t} of the vector time series $y_t = Ay_{t-1} + u_t$. The deterministic system omitting the random disturbances satisfies $y_t = Ay_{t-1}$. The autocovariance matrix of the

stochastic system satisfies the matrix difference equation $\Gamma_k = A\Gamma_{k-1}$, implying, for the first column and with $\gamma_{i1,k}$ denoting $\text{cov}(y_{it}, y_{1,t-k})$,

$$\begin{bmatrix} \gamma_{11,k} \\ \gamma_{21,k} \\ \vdots \\ \gamma_{p1,k} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} \gamma_{11,k-1} \\ \gamma_{21,k-1} \\ \vdots \\ \gamma_{p1,k-1} \end{bmatrix}.$$

Thus the autocovariance function for the first variable $\gamma_{11,k}$, along with other functions $\gamma_{21,k}, \ldots, \gamma_{p1,k}$, satisfies the same system of difference equations in the time lag k as the variable y_{1t} of the deterministic system $y_t = Ay_{t-1}$, along with other variables $y_{2t}, \ldots, y_{pt}$. The initial conditions of the two systems are different. The first is a set of covariances $\gamma_{11,0}$, $\gamma_{21,0}, \ldots, \gamma_{p1,0}$. The second is a set of initial observations of the multivariate time series itself, $y_{10}, \ldots, y_{p0}$, but the two sets of variables should have similar cyclical characteristics.

To extract the periodic movements in the autocovariance function $\gamma_{11,k}$ we can weight it by the periodic function $\cos\omega k$ and form the weighted sum

$$f_{11}(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_{11,k} \cos\omega k. \tag{1}$$

If the periodic movements of $\gamma_{11,k}$ and of $\cos\omega k$ coincide, this weighted sum will be large. In particular, imagine that $\gamma_{11,k}$ is itself a cosine function of the same frequency $\omega/2\pi$ as the weighting function. When $\gamma_{11,k}$ equals 1 (for $k = 0, 2\pi/\omega, 4\pi/\omega, \ldots$), the weighting function will also be 1. When $\gamma_{11,k}$ equals -1 (for $k = \pi/\omega, 3\pi/\omega, 5\pi/\omega, \ldots$), the weighting function will be -1, giving a product of 1. On the other hand, if the periodic movements of $\gamma_{11,k}$ and of $\cos\omega k$ do not coincide, the weighted sum (1) will be small, for when $\gamma_{11,k}$ is large $\cos\omega k$ may be small or even negative and when $\gamma_{11,k}$ is small or negative $\cos\omega k$ may be large, so that the products of the two functions for different values of k will cancel out to yield a small sum. In other words, if $\gamma_{11,k}$ includes an important periodic movement of frequency $\omega/2\pi$, the weighted sum (1) will be large; otherwise the sum will be small. To the extent that periodic movements in the autocovariance function reflect similar periodic movements in the time series itself, we can say that if (1) is large the time series itself will contain important periodic movements of frequency $\omega/2\pi$ or of cycle length $2\pi/\omega$.

The function (1) is called the *power spectrum* of the time series y_{1t}. It is a

weighted sum, over k, of the autocovariance function $\gamma_{11,k}$ with $\cos\omega k$ as weights. The power spectrum is a function of ω. A large value of the function at $\omega=\omega_1$ indicates that the time series contains important cycles of frequency $\omega_1/2\pi$ or of cycle length equal to $2\pi/\omega_1$ time units. For a discrete time series the argument ω can range from 0 to π, as the corresponding length of the cycle $2\pi/\omega$ ranges from an infinitely large value to 2. Observations in discrete time, with t an integer, cannot reveal cycles shorter than two time units or frequencies larger than $\frac{1}{2}$. If we divide the power spectrum by the variance of the time series, we will obtain the *spectral density function* of the time series. In other words, the spectral density function of y_{1t} is the weighted sum of the autocorrelation function $\rho_{11,k}$, for $k=-\infty$ to $+\infty$, using $\cos\omega k$ as weights. In the literature, however, the term spectral density function is often used more generally to denote both the power spectrum (1) and the normalized power spectrum that results from dividing (1) by the variance $\gamma_{11,0}$ of the time series. This more general use of the term spectral density is adopted in this book. If the spectral density $f_{11}(\omega)$ is normalized, the integral $\int_0^{\pi} f_{11}(\omega)d\omega$ will be equal to 1. That is why π^{-1} appears in the definition (1). For a proof of this normalization constant see (28) below, with ρ_k replacing γ_k and $k=0$.

As an example, let us find the spectral density function of the time series $y_t=ay_{t-1}+u_t$. For this time series the autocorrelation function is $\rho_k=a^k$ for $k=0,1,2,\ldots$, and $\rho_k=\rho_{-k}$; that is, $\rho_k=a^{|k|}$. We evaluate the weighted sum of $\rho_k=a^{|k|}$ by rewriting the spectral density function, using complex weights:

$$f(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_k\cos\omega k=\frac{1}{\pi}\left(\sum_{k=-\infty}^{\infty}\rho_k e^{-i\omega k}+i\sum_{k=-\infty}^{\infty}\rho_k\sin\omega k\right)$$

$$=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_k e^{-i\omega k}, \tag{2}$$

where we have used the identities

$$e^{-i\omega k}=\cos\omega k-i\sin\omega k$$

and

$$\sum_{k=-\infty}^{\infty}\rho_k\sin\omega k=0,$$

because $\rho_{-k}=\rho_k$ and $\sin\omega(-k)=-\sin\omega k$. The last expression in (2) serves as an alternative definition of spectral density. For $\rho_k=a^{|k|}$ it can be

evaluated thus:

$$f(\omega)=\frac{1}{\pi}\left(\sum_{k=0}^{\infty}\rho_k e^{-i\omega k}+\sum_{k=0}^{\infty}\rho_{-k}e^{i\omega k}-\rho_0\right)$$

$$=\frac{1}{\pi}\left(\sum_{k=0}^{\infty}a^k e^{-i\omega k}+\sum_{k=0}^{\infty}a^k e^{i\omega k}-1\right)$$

$$=\frac{1}{\pi}\left(\frac{1}{1-ae^{-i\omega}}+\frac{1}{1-ae^{i\omega}}-1\right)$$

$$=\frac{1}{\pi}\cdot\frac{2-ae^{i\omega}-ae^{-i\omega}-(1-ae^{-i\omega})(1-ae^{i\omega})}{(1-ae^{-i\omega})(1-ae^{i\omega})}$$

$$=\frac{1}{\pi}\cdot\frac{1-a^2}{1+a^2-a(e^{-i\omega}+e^{i\omega})}=\frac{1}{\pi}\cdot\frac{1-a^2}{(1+a^2-2a\cos\omega)}. \tag{3}$$

This is the spectral density function for the autoregressive time series $y_t=ay_{t-1}+u_t$.

The shapes of the spectral density functions for different values of the coefficient a can be asertained from the last expression in (3). Please see Problem 1 in this chapter. For $0<\omega<\pi$, $\cos\omega$ is a decreasing function of ω. *If a is positive*, the denominator $(1+a^2-2a\cos\omega)$ increases with ω and therefore the spectral density function $f(\omega)$ *decreases with* ω. This means that the components of the time series with the smaller frequencies $\omega/2\pi$ or the longer cycles dominate those with greater frequencies or shorter cycles. A time series satisfying $y_t=.9y_{t-1}+u_t$ is a slow moving time series in which 90 percent of its value in the last period carries over to the present. The series does not change much from one period to the next. Therefore long cycles dominate short cycles. *If a is negative*, the denominator $(1+a^2-2a\cos\omega)$ decreases with ω and the spectral density function $f(\omega)$ *increases with* ω. In this case short cycles dominate long cycles. The time series shows fluctuations with high frequencies. The cycles with the highest frequency $\frac{1}{2}$ or with the shortest length of two time units are the most pronounced. In the extreme case in which $a=.99$, say, the function $f(\omega)$ has densities heavily concentrated for ω near π or for cycles with lengths close to two time units. When a equals 0, y_t is just the *random series* u_t and $f(\omega)=1/\pi$ *is a constant*. It shows that cycles of different lengths are equally important. Because $f(\omega)$ is normalized, the area under the function from $\omega=0$ to $\omega=\pi$ is exactly 1.

4.2 SPECTRAL DENSITY FUNCTIONS OF A BIVARIATE SYSTEM

The spectral density function of a time series that satisfies a first-order autoregression $y_t = ay_{t-1} + u_t$ does not have a maximum at ω between 0 and π. In other words, either very long or very short cycles dominate, depending on whether a is positive or negative. The spectral density function of a time series generated by $y_t = Ay_{t-1} + u_t$, A being a 2×2 matrix, can have a relative maximum at a value of ω between 0 and π. To pursue this point and to set the stage for the discussion of spectral density functions of a multivariate stochastic system we consider next the spectral density functions of time series that satisfy a first-order bivariate system of stochastic difference equations.

The autocorrelation function ρ_k of a time series generated by a bivariate system of first-order stochastic difference equations takes the form $d_1\lambda_1^k + d_2\lambda_2^k$ $(k \geqslant 0)$ and $\rho_{-k} = \rho_k$, as we saw in Chapter 3. Here λ_1 and λ_2 are the characteristic roots of the system; d_1 and d_2 are suitable weights, which are conjugate complex if λ_1 and λ_2 are conjugate complex. To derive the spectral density function from this autocorrelation function we can use the result from (3) that for any λ with absolute value smaller than 1

$$\sum_{k=-\infty}^{\infty} \lambda^{|k|} e^{-i\omega k} = \frac{1-\lambda^2}{(1-\lambda e^{i\omega})(1-\lambda e^{-i\omega})} = \frac{1-\lambda^2}{1+\lambda^2-2\lambda\cos\omega}, \tag{4}$$

and obtain

$$f(\omega) = \frac{1}{\pi}\sum_{k=-\infty}^{\infty} (d_1\lambda_1^{|k|} + d_2\lambda_2^{|k|})e^{-i\omega k}$$

$$= \frac{1}{\pi}\left[\frac{d_1(1-\lambda_1^2)}{1+\lambda_1^2-2\lambda_1\cos\omega} + \frac{d_2(1-\lambda_2^2)}{1+\lambda_2^2-2\lambda_2\cos\omega}\right]. \tag{5}$$

This is the spectral density function of a time series that satisfies a bivariate system of stochastic difference equations.

To find out whether the time series contains any periodic components of particular importance we can try to locate a relative maximum for the spectral density function (5) for $0<\omega<\pi$. If a relative maximum exists at ω_1, periodic movements at frequencies near $\omega_1/2\pi$ are more important than at other frequencies. Of course, how pronounced these periodic

movements are depends on the sharpness of the peak. To search for a relative maximum we differentiate (5) with respect to ω and obtain

$$\frac{df(\omega)}{d\omega} = \frac{1}{\pi}\left[\frac{-d_1(1-\lambda_1^2)\cdot 2\lambda_1\sin\omega}{(1+\lambda_1^2-2\lambda_1\cos\omega)^2} + \frac{-d_2(1-\lambda_2^2)\cdot 2\lambda_2\sin\omega}{(1+\lambda_2^2-2\lambda_2\cos\omega)^2}\right] = 0.$$

Because $\sin\omega > 0$ for $0<\omega<\pi$, the foregoing necessary condition can be restated as

$$d_1\lambda_1(1-\lambda_1^2)(1+\lambda_2^2-2\lambda_2\cos\omega)^2 + d_2\lambda_2(1-\lambda_2^2)(1+\lambda_1^2-2\lambda_1\cos\omega)^2 = 0. \tag{6}$$

Equation 6 is a quadratic equation in $\cos\omega$. In order for a relative maximum to exist a solution of (6) must take a value between 1 and -1 for $\cos\omega$.

We illustrate the spectral density (5) with the numerical example in Section 3.10. The system of stochastic difference equations is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1.62 & -.80 \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \tag{7}$$

where the covariance matrix V of u_{1t} and u_{2t} is assumed to be the identity matrix. According to Section 3.10, the autocorrelation functions of the two time series are, respectively,

$$\rho_{11,k} = d_{11}\lambda_1^k + d_{12}\lambda_2^k$$

$$= .5010e^{.0629i}(.9055e^{1.460i})^k + .5010e^{-.0629i}(.9055e^{-1.460i})^k$$

$$= 1.0020(.9055)^k\cos(1.460k + .0629) \qquad (k \geqslant 0),$$

and

$$\rho_{22,k} = d_{21}\lambda_1^k + d_{22}\lambda_2^k$$

$$= .5002e^{-.0297i}(.9055e^{1.460i})^k + .5002e^{.0297i}(.9055e^{-1.460i})^k$$

$$= 1.0004(.9055)^k\cos(1.460k - .0297) \qquad (k \geqslant 0).$$

As we have pointed out, both autocorrelation functions are damped cosine functions with periodicity $2\pi/1.460$ or 4.30 time units. If either function is weighted by $\cos 1.460k$ and the sum over k from $-\infty$ to $+\infty$ is formed, we would expect the result to be larger than the sum obtained by a weighting function $\cos\omega k$ whose argument ω is quite different from 1.460. In other words, we can expect the spectral density to have a relative maximum near $\omega = 1.460$. The relative maximum will not occur at exactly 1.460 because the autocorrelation function is not strictly a periodic function, having been blurred by the factor $(.9055)^k$.

The spectral density function for the first time series is

$$f_{11}(\omega) = \frac{1}{\pi}\left[\frac{.5010e^{.0629i}(1.8090e^{-.0997i})}{.2691e^{.7328i} - 1.811e^{1.460i}\cos\omega} + \frac{.5010e^{-.0629i}(1.8090e^{.0997i})}{.2691e^{-.7328i} - 1.811e^{-1.460i}\cos\omega}\right]$$

For any ω between 0 and π the two components of the sum in square brackets are complex conjugates. Therefore the spectral density is real. To find a relative maximum we use (6):

$$\lambda_1 d_{11}(1-\lambda_1^2)\left[(1+\lambda_2^2)^2 - 4(1+\lambda_2^2)\lambda_2\cos\omega + 4\lambda_2^2\cos^2\omega\right] + \text{complex conjugate} = 0.$$

Substituting the values for λ_1, d_{11}, and λ_2 and dividing the equation by the absolute value or modulus of $\lambda_1 d_{11}(1-\lambda_1^2)$, we have

$$e^{1.4232i}(.07241e^{-1.4656i} - .9746e^{-2.1928i}\cos\omega + 3.2797e^{-2.920i}cos^2\omega) + \text{complex conjugate} = 0.$$

The addition of each complex term to its conjugate by (33) in Chapter 2 yields

$$.07234 - .69995\cos\omega + .24246\cos^2\omega = 0.$$

The solutions to this quadratic equation are 2.780 and .1074. The first is greater than 1, hence unacceptable. The second implies $\omega = 1.463$. This figure is close to the value 1.460 of ω in the autocorrelation function. It

implies a cycle length of $2\pi/1.463$ or 4.29 time units, compared with 4.30 units for the latter value. This example shows how a periodic weighting function $e^{-i\omega k}$ or $\cos\omega k$ picks up the periodic movements in the autocorrelation function.

4.3 CROSS-SPECTRAL DENSITY FUNCTION VIA THE CROSS-CORRELATION FUNCTION

In the same way that a spectral density function is obtained as a weighted sum of the autocorrelation function by using periodic weights the cross-spectral density function is a periodically weighted sum of the cross-correlation function. There is one difference, however, in using the periodic weights. For the autocorrelation function $\cos\omega k$ can serve as the weighting function. There is no possibility that these two functions will be out of phase because both functions reach a maximum at $k=0$ and both take equal values at k and $-k$. One only has to match the two functions by their periodicities or frequencies and does not have to be concerned with their phase differences. The cross-correlation function $\rho_{12,k}$ may not have a maximum at $k=0$ and, in general, $\rho_{12,k}$ does not equal $\rho_{12,-k}$. The function $\rho_{12,k}$ may have a maximum at $k=2$, for instance, which suggests that $y_{2,t-2}$ is more highly correlated with y_{1t} than any $y_{2,t-k}$ for $k \neq 2$. This may indicate that the second time series leads the first. In any case, the weighting function applied to the cross-correlation function should be $\cos(\omega k - \psi)$ for some ψ yet to be determined rather than just $\cos\omega k$.

To measure the importance of the periodic component with frequency $\omega/2\pi$ in the cross-correlation function $\rho_{12,k}$ we use the weighting function $\cos(\omega k - \psi)$, where ψ is chosen to make the weighted sum

$$\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_{12,k}\cos(\omega k - \psi) = \frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_{12,k}(\cos\omega k\cdot\cos\psi + \sin\omega k\cdot\sin\psi) \quad (8)$$

as large as possible. By maximizing (8) with respect to ψ we find the appropriate phase shift for the periodic weighting function with frequency $\omega/2\pi$. Setting the derivative of (8) with respect to ψ equal to 0, we get

$$\frac{1}{\pi}\left(-\sin\psi\sum_{k=-\infty}^{\infty}\rho_{12,k}\cos\omega k + \cos\psi\sum_{k=-\infty}^{\infty}\rho_{12,k}\sin\omega k\right) = 0,$$

which implies

$$\tan\psi_{12}(\omega) = \frac{q_{12}(\omega)}{c_{12}(\omega)}, \quad (9)$$

where

$$c_{12}(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_{12,k}\cos\omega k, \tag{10}$$

and

$$q_{12}(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_{12,k}\sin\omega k. \tag{11}$$

The solution of (9) for $\psi_{12}(\omega)$ is that phase shift in the weighting function $\cos(\omega k-\psi)$ which will make the weighted sum (8) as large as possible. It indicates that the periodic component of frequency $\omega/2\pi$ in the auto-correlation function $\rho_{12,k}$ has a phase shift of $\psi_{12}(\omega)$. This means that, as far as the periodic components of frequency $\omega/2\pi$ are concerned, the first time series is most highly correlated with the second time series lagged $k=[\psi_{12}(\omega)]/\omega$ periods. In other words, the cyclical component of frequency $\omega/2\pi$ in the first time series lags behind the corresponding cyclical component in the second time series by $[\psi_{12}(\omega)]/\omega$ periods or by $\psi_{12}(\omega)/2\pi$ cycles; $\psi_{12}(\omega)$, defined by (9), is called the *phase-difference cross-spectral density* of the two series. It shows that the first series lags behind the second series by $\psi_{12}(\omega)/\omega$ periods in relation to the cyclical components of frequency $\omega/2\pi$. It is computed from the functions $c_{12}(\omega)$ and $q_{12}(\omega)$, defined respectively by (10) and (11). The former is called the *in-phase cross-spectral density* and the latter the *out-of-phase cross-spectral density*. If the out-of-phase cross-spectral density $q_{12}(\omega)$ is 0, the phase-difference cross-spectral density will also be 0. In this case there is no phase shift required in the weighting function $\cos\omega k$.

Having obtained the appropriate phase shift $\psi_{12}(\omega)$ for the weighting function $\cos(\omega k-\psi)$, we can use it to compute the weighted sum (8). The result will measure the correlation between the periodic components of frequency $\omega/2\pi$ in the two time series, once their phase-difference is ironed out. Substituting (9) for ψ in (8) and using the definitions (10) and (11), we obtain

$$\begin{aligned}\frac{1}{\pi}&\left[\cos\psi_{12}(\omega)\cdot\sum_{k=-\infty}^{\infty}\rho_{12,k}\cos\omega k+\sin\psi_{12}(\omega)\cdot\sum_{k=-\infty}^{\infty}\rho_{12,k}\sin\omega k\right]\\&=\frac{c_{12}(\omega)}{\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}}\cdot c_{12}(\omega)+\frac{q_{12}(\omega)}{\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}}\cdot q_{12}(\omega)\\&=\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}\ .\end{aligned} \tag{12}$$

Equation 12 is called the *cross-amplitude spectral density*. It shows the degree of association between the periodic component of frequency $\omega/2\pi$ in the first series and the corresponding component in the second series after appropriate adjustment has been made of the phase difference.

The above development of the phase-difference cross-spectral density and the cross-amplitude spectral density helps to explain and motivate the definition of the *cross-spectral* density as

$$f_{12}(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\rho_{12,k}e^{-i\omega k}, \tag{13}$$

where the complex weighting function $e^{-i\omega k}$ is used. The resulting function $f_{12}(\omega)$ is in general complex. It can be written as

$$\begin{aligned}f_{12}(\omega)&=\frac{1}{\pi}\left[\sum_{k=-\infty}^{\infty}\rho_{12,k}\cos\omega k-i\sum_{k=-\infty}^{\infty}\rho_{12,k}\sin\omega k\right]\\&=c_{12}(\omega)-iq_{12}(\omega)\\&=\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}\cdot e^{-i\psi_{12}(\omega)}.\end{aligned} \tag{14}$$

The *absolute value of* $f_{12}(\omega)$ *is* $\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}$ or the *cross amplitude spectral density. The angle* $\psi_{12}(\omega)$ *of* $f_{12}(\omega)$ *is* $\tan^{-1}[q_{12}(\omega)/c_{12}(\omega)]$ or the *phase-difference cross-spectral density*.

For a pair of time series satisfying a first-order system of stochastic difference equations the cross-spectral density function $f_{12}(\omega)$ can be derived from the cross-correlation function

$$\rho_{12,k}=c_{11}\lambda_1^k+c_{12}\lambda_2^k \qquad (k\geqslant 0),$$

$$\rho_{12,-k}=\rho_{21,k}=c_{21}\lambda_1^k+c_{22}\lambda_2^k \qquad (k\geqslant 0).$$

Because $\rho_{12,-k}\neq\rho_{12,k}$, the derivation of the cross-spectral density function does not simplify to the same extent as that of (5) for the spectral density function. The result can be written as

$$\begin{aligned}f_{12}(\omega)&=\frac{1}{\pi}\left(\sum_{k=0}^{\infty}\rho_{12,k}e^{-i\omega k}+\sum_{k=0}^{\infty}\rho_{12,-k}e^{i\omega k}-\rho_{12,0}\right)\\&=\frac{1}{\pi}\left[\frac{c_{11}}{1-\lambda_1e^{-i\omega}}+\frac{c_{12}}{1-\lambda_2e^{-i\omega}}+\frac{c_{21}}{1-\lambda_1e^{i\omega}}+\frac{c_{22}}{1-\lambda_2e^{i\omega}}-(c_{11}+c_{12})\right].\end{aligned} \tag{15}$$

Note that $c_{11}+c_{12}=c_{21}+c_{22}$ because $\rho_{12,0}=\rho_{21,0}$. Equation 15 is in general complex. It can be used to obtain the cross-amplitude spectral density and the phase-difference spectral density as recommended in Problem 5 in this chapter. We shall study the applications of the cross-spectral density functions in Sections 4.8 and 5.8.

4.4 DECOMPOSITION OF TIME SERIES DATA INTO PERIODIC COMPONENTS

Our discussion has hinted to the possibility of decomposing a time series into periodic components of cosine and sine functions. This decomposition provides a useful way to characterize a time series. We first perform the decomposition for a *set of data* y_{1t} $(t=1,\ldots,N)$, which are simply N *arbitrarily given numbers*. This decomposition will be suggestive to, and enhance the understanding of, a similar decomposition for a time series that is, as we have defined it, *a random function of time*. The first decomposition amounts to manipulating sample data. The second is concerned with analyzing population values and should be distinguished from the first; it is the subject of Section 4.5.

Let N arbitrary numbers $y_{11},y_{12},\ldots,y_{1N}$ be given. Imagine fitting this set of data by a weighted sum of sine and cosine functions. These functions should have cycle lengths equal to N, $\frac{1}{2}N$, $\frac{1}{3}N,\ldots,2$, so that the N points will cover one cycle, two cycles, and so on, of the functions. The functions are $\cos\omega_j t$ and $\sin\omega_j t$, where $\omega_j=(2\pi/N)\cdot j$ $(j=1,\ldots,N/2)$. For convenience let N be even or $N/2=n$. For $j=n$ we have $\omega_n=\pi$ and therefore $\sin\omega_n t=0$. The decomposition is

$$y_{1t}=a_{10}+\sum_{j=1}^{n-1}(a_{1j}\cos\omega_j t+b_{1j}\sin\omega_j t)+a_{1n}\cos\omega_n t. \tag{16}$$

The N unknown coefficients a_{1j} and b_{1j} can be determined by performing a least-squares regression of y_{1t} on the explanatory variables $\cos\omega_j t$ $(j=1,\ldots,n)$ and $\sin\omega_j t$ $(j=1,\ldots,n-1)$.

To obtain the least-squares estimates the sums of squares and cross-products of the explanatory variables (measured from their means) are used. Note first that the sum of each variable over t is 0 because the positive and negative values of each function cancel out for every cycle and there are an integral number of cycles. Thus

$$\sum_{t=1}^{N}\cos\omega_j t=\sum_{t=1}^{N}\sin\omega_j t=0 \qquad \left(\omega_j=\frac{2\pi}{N}\cdot j;\ j=1,\ldots,\frac{N}{2}\right). \tag{17}$$

The explanatory variables have 0 mean and we can deal directly with the sums of squares and cross-products:

$$\sum_{t=1}^{N}\cos^2\omega_j t=\sum_{t=1}^{N}\sin^2\omega_j t=\frac{N}{2} \qquad (j=1,\ldots,n-1), \tag{18}$$

$$\sum_{t=1}^{N}\cos^2\omega_n t=N,$$

$$\sum_{t=1}^{N}(\cos\omega_j t)(\sin\omega_k t)=0 \qquad (j,k=1,\ldots,n-1), \tag{19}$$

$$\sum_{t=1}^{N}(\cos\omega_j t)(\cos\omega_k t)=\sum_{t=1}^{N}(\sin\omega_j t)(\sin\omega_k t)=0 \qquad (j\neq k).$$

For the proofs of (18) and (19) consult Problems 6 and 7 at the end of this chapter. Equations 18 and 19 are important identities. They show that each periodic component $\cos\omega_j t$ or $\sin\omega_j t$ $(j=1,\ldots,n-1)$ has the same variance and that each component is uncorrelated with any other component.

By the method of least squares the normal equations are

$$\begin{bmatrix} \frac{N}{2} & 0 & \cdots & & 0 \\ 0 & \frac{N}{2} & \cdots & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & \cdots & \frac{N}{2} & 0 \\ 0 & & \cdots & 0 & N \end{bmatrix} \begin{bmatrix} a_{11} \\ b_{11} \\ \vdots \\ b_{1,n-1} \\ a_{1n} \end{bmatrix} = \begin{bmatrix} \sum_t y_{1t}\cos\omega_1 t \\ \sum_t y_{1t}\sin\omega_1 t \\ \vdots \\ \sum_t y_{1t}\sin\omega_{n-1} t \\ \sum_t y_{1t}\cos\pi t \end{bmatrix} \tag{20}$$

Solving these equations and using a well-known formula for the intercept a_{10}, we obtain

$$a_{1j}=\frac{2}{N}\sum_{t=1}^{N}y_{1t}\cos\omega_j t, \qquad b_{1j}=\frac{2}{N}\sum_{t=1}^{N}y_{1t}\sin\omega_j t \qquad (j=1,\ldots,n-1) \tag{21}$$

$$a_{10}=\frac{1}{N}\sum_{t=1}^{N}y_{1t}=\bar{y}_1; \qquad a_{1n}=\frac{1}{N}\sum_{t=1}^{N}(-1)^t y_{1t}.$$

Because the number of coefficients in (16) equals the number of data points, the equation will fit the observations y_{1t} $(t=1,\ldots,N)$ exactly without leaving any residuals. A given set of data has thus been decomposed into a weighted sum of cosine and sine functions that are mutually uncorrelated.

To measure the importance of each periodic component in y_{1t} it seems natural to use the contribution of that component to the sample variance of y_{1t}. Squaring $y_{1t}-a_{10}$ from (16), summing over t and using (19) and (18), we have

$$\sum_{t=1}^{N}(y_{1t}-\bar{y}_1)^2=\sum_{j=1}^{n-1}\left(a_{1j}^2\sum_{t=1}^{N}\cos^2\omega_j t+b_{1j}^2\sum_{t=1}^{N}\sin^2\omega_j t\right)+a_{1n}^2\sum_{t=1}^{N}\cos^2\omega_n t$$

$$=\frac{N}{2}\sum_{j=1}^{n-1}(a_{1j}^2+b_{1j}^2)+Na_{1n}^2. \tag{22}$$

The contribution of the periodic component of frequency $\omega_j/2\pi$ to the sample variance of y_{1t} is therefore $\frac{1}{2}(a_{1j}^2+b_{1j}^2)$.

4.5 DECOMPOSITION OF THEORETICAL TIME SERIES INTO PERIODIC COMPONENTS

The idea of the last section can be extended and modified for the decomposition of a time series that is defined as a random function of time. No probability considerations were involved in the decomposition of (16). Now let y_{1t} *be a random function of* t defined by

$$y_{1t}=\sum_j[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t], \tag{23}$$

where $\alpha_1(\omega_j)$ and $\beta_1(\omega_j)$ are random variables with 0 mean and equal variance $E[\alpha_1^2(\omega_j)]=E[\beta_1^2(\omega_j)]$ for each j and all of these random variables are mutually uncorrelated. By introducing randomness in the coefficients $\alpha_1(\omega_j)$ and $\beta_1(\omega_j)$ we attempt to model a stochastic time series. Here ω_j may range over many values between 0 and π, so that there may be many and even an infinite number of periodic components. What is the contribution of the periodic component of frequency $\omega_j/2\pi$ to the total variance of y_{1t}? To

answer this question we decompose the variance of y_{1t}:

$$Ey_{1t}^2=E\left\{\sum_j[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t]\right\}^2$$

$$=\sum_j\{[E\alpha_1^2(\omega_j)]\cos^2\omega_j t+[E\beta_1^2(\omega_j)]\sin^2\omega_j t\}$$

$$=\sum_j[E\alpha_1^2(\omega_j)]=\sum_j\operatorname{var}[\alpha_1(\omega_j)], \tag{24}$$

where use has been made of $E[\alpha_1^2(\omega_j)]=E[\beta_1^2(\omega_j)]$ and the 0 correlations of the random coefficients. Thus the total variance of the time series is the sum of the variances of the periodic components. The contribution of the jth component is the common variance of the coefficient of $\cos\omega_j t$ or of $\sin\omega_j t$. Please compare this contribution with the contribution given in the last section by the jth component to the sample variance of an arbitrary set of data. In the latter case we have estimated coefficients a_{1j} and b_{1j}. The common variance of these coefficients can be estimated by $\frac{1}{2}(a_{1j}^2+b_{1j}^2)$, which is analogous to $E\alpha_1^2(\omega_j)$ of (24).

For the time series (23) let us *redefine the power spectrum $f_{11}(\omega_j)$ as the common variance $E\alpha_1^2(\omega_j)$ of the random coefficient $\alpha_1(\omega_j)$ or $\beta_1(\omega_j)$ or equivalently as the contribution of the random periodic component of frequency $\omega_j/2\pi$ to the total variance of the time series.* In order to make this definition applicable to more general situations we have to allow the variable ω to be continuous. Thus ω_j in (23), even if countably infinite, will have to be replaced by a continuous variable ω. The sum of (23) can indeed be replaced by an appropriate integral such that all covariance stationary time series can be so represented. This can be done without abandoning the essential structure of the αs and the βs as specified above. We can then speak of a power spectrum or a spectral density function $f_{11}(\omega)$ for continuous ω between 0 and π. The area $\int_a^b f_{11}(\omega)\,d\omega$ under the spectral density function between two points a and b measures the contribution of the periodic components with ω between a and b to the total variance of the time series. If the spectral density is normalized, the contribution to total variance is given as a fraction and the total area under the function $\int_0^\pi f_{11}(\omega)\,d\omega$ is equal to 1.

This definition of the spectral density function enables us to understand the concept in terms of regression theory. As pointed out in Section 4.4, we can imagine decomposing a time series into many periodic components as performing a regression of the time series on many explanatory variables which are cosine and sine functions of time. The variances of the regres-

sion coefficients are the spectral densities, but we have to think of an infinite number of explanatory variables and of regression coefficients. These coefficients are also random coefficients.

4.6* EQUIVALENCE OF TWO DEFINITIONS OF SPECTRAL DENSITY

In Sections 4.1 and 4.5 two definitions have been given to the term spectral density. The first is through the autocovariance function. The second is in terms of the variances of the periodic components of the time series itself. Both are useful ways to view the concept of spectral density and should be equivalent.

Before we show the equivalence of the two definitions, one further mathematical relationship has to be developed from the first definition. In Equation 1 the spectral density function $f_{11}(\omega)$ is derived from the autocovariance function $\gamma_{11,k}$. It is possible to invert the operation, that is, to obtain the autocovariance function from the spectral density function. The inverse operation will be derived in a more general setting by the use of complex weights $e^{-i\omega k}$ and $e^{i\omega k}$ and by defining $g(\omega)=g(-\omega)=\frac{1}{2}f(\omega)$ $(0\leqslant\omega\leqslant\pi)$ to extend the domain of the spectral density function to negative values of ω. Thus we propose to invert the operation

$$g(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k} \qquad (-\pi\leqslant\omega\leqslant\pi). \tag{25}$$

The inversion can be achieved by taking the integral

$$\int_{-\pi}^{\pi} g(\omega)e^{i\omega k}\,d\omega=\int_{-\pi}^{\pi}\left(\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\gamma_m e^{-i\omega m}\right)e^{i\omega k}\,d\omega$$

$$=\frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\gamma_m\int_{-\pi}^{\pi}e^{i(k-m)\omega}\,d\omega$$

$$=\frac{1}{2\pi}\gamma_k\int_{-\pi}^{\pi}e^{i(k-k)\omega}\,d\omega=\gamma_k, \tag{26}$$

where it was observed that the integral of

$$e^{i(k-m)\omega}=\cos(k-m)\omega+i\sin(k-m)\omega$$

over ω from $-\pi$ to π is 0 for $m\neq k$ because the positive and the negative parts from these cosine and sine functions exactly cancel out. The function $g(\omega)$ defined by (25) is the *Fourier transform of* γ_k. The *inverse Fourier transform* of $g(\omega)$ is given by (26) and equals γ_k. Thus the spectral density function and the autocovariance function are a pair of Fourier transforms to each other. One can be defined in terms of the other if we adopt our earlier definition of the spectral density function of Section 4.1. It is also known that the pair of Fourier transforms is unique. Given either one of the transforms, the other is uniquely determined.

The development of (25) and (26) implies that the *cosine transform of the autocovariance function*

$$f(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\gamma_k\cos\omega k=2g(\omega) \qquad (0\leqslant\omega\leqslant\pi), \tag{27}$$

can be inverted by the *cosine transform of the spectral density*

$$\int_0^{\pi}f(\omega)\cos\omega k\,d\omega=2\int_0^{\pi}g(\omega)\cos\omega k\,d\omega$$

$$=\int_{-\pi}^{\pi}g(\omega)\cos\omega k\,d\omega=\int_{-\pi}^{\pi}g(\omega)e^{i\omega k}\,d\omega=\gamma_k, \tag{28}$$

where the integral of $g(\omega)\sin\omega k$ vanishes because $g(\omega)\sin\omega k=-g(-\omega)\sin(-\omega k)$. Hence the spectral density function and the autocovariance function are related by a pair of cosine transforms. One can be obtained from the other.

To check the equivalence of the two definitions of the spectral density function for a model like (23) we evaluate the autocovariance function of the time series defined by (23). If the result is a cosine transform (28) of the spectral density function, as specified by the second definition, the two definitions are equivalent because of the uniqueness of the cosine transforms. The autocovariance function of y_{1t} by (23) is, with $\alpha_1(\omega_j)$ and $\beta_1(\omega_j)$

abbreviated by α_j and β_j,

$$\begin{aligned}\gamma_{11,k} &= Ey_{1t}y_{1t-k}\\ &= E\Big[\sum_j(\alpha_j\cos\omega_j t+\beta_j\sin\omega_j t)\Big]\\ &\quad\times\Big\{\sum_j[\alpha_j\cos(\omega_j t-\omega_j k)+\beta_j\sin(\omega_j t-\omega_j k)]\Big\}\\ &= \sum_j E(\alpha_j\cos\omega_j t+\beta_j\sin\omega_j t)[\alpha_j\cos(\omega_j t-\omega_j k)+\beta_j\sin(\omega_j t-\omega_j k)]\\ &= \sum_j E(\alpha_j\cos\omega_j t+\beta_j\sin\omega_j t)(\alpha_j\cos\omega_j t\cos\omega_j k+\beta_j\sin\omega_j t\cos\omega_j k\\ &\qquad +\alpha_j\sin\omega_j t\sin\omega_j k-\beta_j\cos\omega_j t\sin\omega_j k)\\ &= \sum_j(E\alpha_j^2\cos^2\omega_j t+E\beta_j^2\sin^2\omega_j t)\cos\omega_j k\\ &= \sum_j[E\alpha_1^2(\omega_j)]\cos\omega_j k. \end{aligned}\tag{29}$$

In (29) we used the assumption that α_i, α_j, β_i and β_j are all uncorrelated and observed, after the fourth equality sign, that certain terms with opposite signs cancel out. Presumably, if proper care is taken to define (23) as an integral and to make ω continuous, the result in (29) will be an integral like the first part of (28), with $E\alpha_1^2(\omega)$ identified with the spectral density $f_{11}(\omega)$. Thus the two definitions of the spectral density function coincide. Although this discussion is far from being a rigorous mathematical argument, it is useful for an intuitive understanding of the subject.

4.7* A SECOND DEFINITION OF CROSS-SPECTRAL DENSITY

Besides the time series (23), let there be another time series defined by

$$y_{2t}=\sum_j[\alpha_2(\omega_j)\cos\omega_j t+\beta_2(\omega_j)\sin\omega_j t].\tag{30}$$

The coefficients $\alpha_2(\omega_j)$ and $\beta_2(\omega_j)$ are also assumed to be random with 0 mean, uncorrelated, and with the same variance for each j. In addition, it is specified that

$$\begin{aligned}E[\alpha_1(\omega_j)\alpha_2(\omega_j)]&=E[\beta_1(\omega_j)\beta_2(\omega_j)];\\ E[\beta_1(\omega_j)\alpha_2(\omega_j)]&=-E[\alpha_1(\omega_j)\beta_2(\omega_j)].\end{aligned}\tag{31}$$

These are covariances because the means are all 0.

Consider the covariance between the component of frequency $\omega_j/2\pi$ in y_{1t} and the corresponding component in $y_{2,t-k}$:

$$\begin{aligned}&\operatorname{cov}[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t,\alpha_2(\omega_j)\cos(\omega_j t-\omega_j k)+\beta_2(\omega_j)\sin(\omega_j t-\omega_j k)]\\ &\quad=E\{[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t]\times[\alpha_2\cos\omega_j t\cdot\cos\omega_j k\\ &\qquad+\alpha_2\sin\omega_j t\cdot\sin\omega_j k+\beta_2\sin\omega_j t\cdot\cos\omega_j k-\beta_2\cos\omega_j t\sin\omega_j k]\}\\ &\quad=\cos\omega_j kE[(\alpha_1\cos\omega_j t+\beta_1\sin\omega_j t)(\alpha_2\cos\omega_j t+\beta_2\sin\omega_j t)]\\ &\qquad+\sin\omega_j kE[(\alpha_1\cos\omega_j t+\beta_1\sin\omega_j t)(\alpha_2\sin\omega_j t-\beta_2\cos\omega_j t)]\end{aligned}\tag{32}$$

Define the *in-phase cross-spectral density* as

$$\begin{aligned}c_{12}(\omega_j)&=\operatorname{cov}[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t,\alpha_2(\omega_j)\cos\omega_j t+\beta_2(\omega_j)\sin\omega_j t]\\ &=E[\alpha_1(\omega_j)\alpha_2(\omega_j)]=E[\beta_1(\omega_j)\beta_2(\omega_j)],\end{aligned}\tag{33}$$

where the relations in (31) have been applied. Define the *out-of-phase cross-spectral density* as

$$\begin{aligned}q_{12}(\omega_j)&=\operatorname{cov}[\alpha_1(\omega_j)\cos\omega_j t+\beta_1(\omega_j)\sin\omega_j t,\alpha_2(\omega_j)\sin\omega_j t-\beta_2(\omega_j)\cos\omega_j t]\\ &=E[\beta_1(\omega_j)\alpha_2(\omega_j)]=-E[\alpha_1(\omega_j)\beta_2(\omega_j)].\end{aligned}\tag{34}$$

The covariance (32) of the two components then becomes, for any ω,

$$\cos\psi c_{12}(\omega)+\sin\psi q_{12}(\omega),\tag{35}$$

where we have defined $\psi=\omega k$.

To find the phase shift ψ that will maximize the covariance (35) we set the derivative equal to 0 to obtain

$$\tan\psi_{12}(\omega)=\frac{q_{12}(\omega)}{c_{12}(\omega)}.\tag{36}$$

The phase shift $\psi_{12}(\omega)$ so defined shows that the ω-component of the second time series leads the ω-component of the first time series by $k=\psi_{12}(\omega)/\omega$ time units; $\psi_{12}(\omega)$ is the *phase-difference cross-spectral density*. Once this optimum phase shift is found, the covariance between the ω-component of the first series and the appropriately timed ω-component of the second series is, by (35) and (36),

$$\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}\ .\tag{37}$$

This is the *cross-amplitude spectral density*.

By using (36) and (37) we can define the *cross-spectral density* as the complex function

$$f_{12}(\omega)=\sqrt{c_{12}^2(\omega)+q_{12}^2(\omega)}\; e^{-i\psi_{12}(\omega)}=c_{12}(\omega)-iq_{12}(\omega). \tag{38}$$

In sum, the cross-spectral density shows in two parts the magnitude of the covariance between corresponding periodic components in the two time series and their relative lead or lag.

To check whether this definition of cross-spectral density is consistent with the one given in Section 4.3 via the cross-covariance function an analysis similar to that in Section 4.6 can be performed. By the same argument used in (25) and (26) the cross-spectral density function and the cross-covariance functions are a pair of Fourier transforms:

$$g_{12}(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_{12,k}e^{-i\omega k}, \qquad (-\pi\leqslant\omega\leqslant\pi); \tag{39}$$

where $g_{12}(\omega)=\frac{1}{2}f_{12}(\omega)$ is seen to be the complex conjugate of $g_{12}(-\omega)$ for $0\leqslant\omega\leqslant\pi$;

$$\begin{aligned}\gamma_{12,k}&=\int_{-\pi}^{\pi}g_{12}(\omega)e^{i\omega k}\,d\omega\\&=\int_0^{\pi}g_{12}(\omega)e^{i\omega k}\,d\omega+\int_0^{\pi}g_{12}(-\omega)e^{-i\omega k}\,d\omega.\end{aligned} \tag{40}$$

Equation 40 implies that $\gamma_{12,k}$ is twice the real part of either of the last two integrals. For the two definitions to agree the covariance between y_{1t} and $y_{2,t-k}$, as given by (23) and (30), should be equal to the $\gamma_{12,k}$ of (40). To evaluate the $\gamma_{12,k}$ of (40) we take the real part of

$$\begin{aligned}&\int_0^{\pi}f_{12}(\omega)e^{i\omega k}\,d\omega\\&=\int_0^{\pi}[c_{12}(\omega)-iq_{12}(\omega)](\cos\omega k+i\sin\omega k)d\omega\\&=\int_0^{\pi}c_{12}(\omega)\cos(\omega k)d\omega+\int_0^{\pi}q_{12}(\omega)\sin(\omega k)d\omega+\text{imaginary part.}\end{aligned} \tag{41}$$

If the definition proposed in this section is to agree with the former definition $c_{12}(\omega)$ and $q_{12}(\omega)$ in the above integrals should be, respectively, $E[\alpha_1(\omega)\alpha_2(\omega)]$ and $E[\beta_1(\omega)\alpha_2(\omega)]$, as given by (33) and (34). The reader may wish to check this by doing Problem 10.

4.8 GAIN AND COHERENCE

The variances and covariances of periodic components of time series have been defined as spectral and cross-spectral densities. It is natural to utilize these concepts to study the simple regression problem for the periodic components. Consider the regression of the ω-component of the first series on the ω-component of the second. The regression coefficient is the ratio of the covariance to the variance of the second series; that is,

$$G_{12}(\omega)=\frac{|f_{12}(\omega)|}{f_{22}(\omega)}, \tag{42}$$

where $|f_{12}(\omega)|$ stands for the absolute value or the modulus of the cross-spectral density function. The regression coefficient $G_{12}(\omega)$ is called the *gain*. The squared correlation coefficient between the two components is

$$R_{12}^2(\omega)=\frac{|f_{12}(\omega)|^2}{f_{11}(\omega)\cdot f_{22}(\omega)}, \tag{43}$$

called the *coherence*.

The gain and coherence have the same interpretations as the regression coefficient and correlation squared, respectively, as they are applied to the relationship between the corresponding periodic components of two time series. They are functions of ω. It is possible for some components of one time series to be highly dependent on the corresponding components of a second series, but not for others; for example, the slow-moving components of consumption expenditures may have large coefficients in the regressions on the corresponding components of income, but the fast-moving components of consumption expenditures may not. We can extend the regression and correlation analyses via the periodic components to more than two variables. Multiple regression and partial correlations can be defined in a fairly straightforward manner, but we shall not pursue them in this book.

4.9 SPECTRAL DENSITY MATRIX OF STOCHASTIC DIFFERENCE EQUATIONS

Having defined and interpreted spectral and cross-spectral densities and derived them for univariate and bivariate systems of stochastic difference equations of the first-order, we now derive them for a multivariate system.

Define the *spectral density matrix* for a vector time series y_t as

$$F(\omega)=[f_{ij}(\omega)]. \tag{44}$$

The diagonal elements of $F(\omega)$ are the spectral densities and the off-diagonal elements are the cross-spectral densities. The spectral-density matrix is related to the autocovariance matrix $\Gamma_k=\Gamma'_{-k}$ by

$$F(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}\Gamma_k e^{-i\omega k}$$

$$=\frac{1}{\pi}\left(\sum_{k=0}^{\infty}\Gamma_k e^{-i\omega k}+\sum_{k=0}^{\infty}\Gamma'_k e^{i\omega k}-\Gamma_0\right). \tag{45}$$

In each term of a sum the weight $e^{-i\omega k}$ is to multiply every element of the autocovariance matrix Γ_k or Γ'_k. A sum of these matrices is obtained by adding corresponding elements as usual.

By using the results on Γ_k from Chapter 3 we can derive the spectral density matrix for the multivariate stochastic system

$$y_t=Ay_{t-1}+b+u_t, \tag{46}$$

where u_t has a covariance matrix V and is serially uncorrelated. Recall the matrix B that consists of the right characteristic vectors of A and the diagonal matrix D_λ that consists of the corresponding characteristic roots, with $A=BD_\lambda B^{-1}$. The autocovariance matrix has been found to be $\Gamma_k=BD_\lambda^k B^{-1}\Gamma_0$ for $k\geqslant 0$, where $\Gamma_0=B\{w_{ij}(1-\lambda_i\lambda_j)^{-1}\}B'$, $(w_{ij})=W=B^{-1}VB^{-1'}$ being the covariance matrix of the residuals of the autoregressions for the canonical variables. The canonical variables were defined by $z_t=B^{-1}y_t$. Their autocovariance matrix is

$$\Gamma_k^z=D_\lambda^k\Gamma_0^z=D_\lambda^k\left[w_{ij}(1-\lambda_i\lambda_j)^{-1}\right], \qquad (k\geqslant 0), \tag{47}$$

and the autocovariance matrix of y_t is

$$\Gamma_k=BD_\lambda^k\Gamma_0^z B'=BD_\lambda^k B^{-1}\Gamma_0, \qquad (k\geqslant 0). \tag{48}$$

By applying the first part of (48) in (45) and interchanging summation

and matrix multiplications in each of the two sums we have

$$F(\omega)=\frac{1}{\pi}\left[B\left(\sum_{k=0}^{\infty}D_\lambda^k\Gamma_0^z e^{-i\omega k}\right)B'+B\left(\sum_{k=0}^{\infty}\Gamma_0^z D_\lambda^k e^{i\omega k}\right)B'-\Gamma_0\right]$$

$$=\frac{1}{\pi}B\left(\sum_{k=0}^{\infty}D_\lambda^k\Gamma_0^z e^{-i\omega k}+\sum_{k=0}^{\infty}\Gamma_0^z D_\lambda^k e^{i\omega k}-\Gamma_0^z\right)B'. \tag{49}$$

The $i-j$ element of the matrix in parentheses in the second line of (49) is

$$\sum_{k=0}^{\infty}\lambda_i^k\gamma_{ij,0}^z e^{-i\omega k}+\sum_{k=0}^{\infty}\gamma_{ij,0}^z\lambda_j^k e^{i\omega k}-\gamma_{ij,0}^z$$

$$=\gamma_{ij,0}^z\left(\frac{1}{1-\lambda_i e^{-i\omega}}+\frac{1}{1-\lambda_j e^{i\omega}}-1\right)$$

$$=\frac{w_{ij}}{1-\lambda_i\lambda_j}\left[\frac{1-\lambda_i\lambda_j}{(1-\lambda_i e^{-i\omega})(1-\lambda_j e^{i\omega})}\right]$$

$$=\frac{w_{ij}}{(1-\lambda_i e^{-i\omega})(1-\lambda_j e^{i\omega})}. \tag{50}$$

The spectral density matrix of y_t is therefore

$$F(\omega)=\frac{1}{\pi}B\left[\frac{w_{ij}}{(1-\lambda_i e^{-i\omega})(1-\lambda_j e^{i\omega})}\right]B'. \tag{51}$$

4.10* SPECTRAL DENSITY MATRIX IN TERMS OF CANONICAL VARIABLES

The derivation of the spectral density matrix in (51) is based on and intimately related to the idea of canonical variables. To make the idea explicit recall that the autocovariance matrix Γ_k^z of the canonical variables is related to the autocovariance matrix Γ_k of y_t by

$$\Gamma_k=B\Gamma_k^z B'. \tag{52}$$

Because the spectral density matrix is defined in (45) as a weighted sum of Γ_k and because summation and matrix multiplications can be interchanged, the spectral density matrices of the two sets of variables are related by

$$F(\omega) = BF^z(\omega)B'. \tag{53}$$

Once the spectral density matrix $F^z(\omega)$ of the canonical variables is found the required result for $F(\omega)$ can be obtained by matrix multiplications as specified by (53).

To derive the cross-spectral density for two canonical variables z_{it} and z_{jt} we first obtain their cross-covariance function:

$$\begin{aligned}
\gamma^z_{ij,k} &= Ez_{it}z_{j,t-k}\cdot \\
&= E(v_{it}+\lambda_i v_{i,t-1}+\lambda_i^2 v_{i,t-2}+\cdots)(v_{j,t-k}+\lambda_j v_{j,t-k-1}+\cdots) \\
&= (Ev_{it}v_{jt})(\lambda_i^k+\lambda_i^{k+1}\lambda_j+\lambda_i^{k+2}\lambda_j^2+\cdots) \\
&= \frac{w_{ij}}{1-\lambda_i\lambda_j}\cdot\lambda_i^k \\
&= \gamma^z_{ij,0}\lambda_i^k, \qquad (k\geqslant 0); \\
\gamma^z_{ij,-k} &= \gamma^z_{ji,k} = \gamma^z_{ij,0}\lambda_j^k, \qquad (k\geqslant 0).
\end{aligned} \tag{54}$$

The Fourier transform of (54) can be performed as it was in (50) to give the cross-spectral density

$$f^z_{ij}(\omega) = \frac{1}{\pi}\sum_{k=-\infty}^{\infty}\gamma^z_{ij,k}e^{-i\omega k} = \frac{1}{\pi}\cdot\frac{w_{ij}}{(1-\lambda_i e^{-i\omega})(1-\lambda_j e^{i\omega})} \tag{55}$$

for the ith and jth canonical variables. Application of (55) to (53) yields the desired result (51).

Equation 55 gives the spectral or cross-spectral density functions $f^z_{ij}(\omega)$ of the canonical variables. If the spectral density function of the ith canonical variable is *normalized* and denoted by $g_i(\omega)$ [not the $g(\omega)$ of (25)], it is

$$\begin{aligned}
g_i(\omega) &= \frac{1}{\gamma^z_{ii,0}}f^z_{ii}(\omega) = \frac{1}{\pi}\frac{1-\lambda_i^2}{(1-\lambda_i e^{-i\omega})(1-\lambda_i e^{i\omega})} \\
&= \frac{1}{\pi}\frac{1-\lambda_i^2}{(1+\lambda_i^2-2\lambda_i\cos\omega)}.
\end{aligned} \tag{56}$$

This is the same result that we have obtained for the first-order autoregression in Section 4.1. If a root λ_i is real, the spectral density of the corresponding canonical variable is real. If λ_i is positive, the spectral density is a decreasing function of ω; if λ_i is negative, it is an increasing function that shows the relative importance of short cycles. It would be interesting to inquire into the relation between the spectral density of an original variable y_{it} and the spectral densities of the canonical variables.

By (51) the spectral density of y_{it} is a quadratic form in the ith row $b_{i\cdot}$ of B:

$$f_{ii}(\omega) = b_{i\cdot}F^z(\omega)b'_{i\cdot} = \tfrac{1}{2}b_{i\cdot}[F^z(\omega)+F^z(\omega)']b'_{i\cdot}, \tag{57}$$

where the $j-m$ element of the matrix in square brackets is, by (50),

$$\begin{aligned}
&f^z_{jm}(\omega)+f^z_{mj}(\omega) \\
&\quad = \frac{1}{\pi}\gamma^z_{jm,0}\left(\frac{1}{1-\lambda_j e^{-i\omega}}+\frac{1}{1-\lambda_m e^{i\omega}}+\frac{1}{1-\lambda_m e^{-i\omega}}+\frac{1}{1-\lambda_j e^{i\omega}}-2\right) \\
&\quad = \gamma^z_{jm,0}[\,g_j(\omega)+g_m(\omega)].
\end{aligned} \tag{58}$$

Denoting by D_g the diagonal matrix with the normalized spectral densities $g_j(\omega)$ of the canonical variables on the diagonal, we can combine (57) and (58) to write

$$\begin{aligned}
f_{ii}(\omega) &= \tfrac{1}{2}b_{i\cdot}D_g\Gamma^z_0 b'_{i\cdot}+\tfrac{1}{2}b_{i\cdot}\Gamma^z_0 D_g b'_{i\cdot}. \\
&= b_{i\cdot}D_g\Gamma^z_0 b'_{i\cdot} = \sum_j\sum_m b_{ij}b_{im}\gamma^z_{jm,0}g_j(\omega).
\end{aligned} \tag{59}$$

The result is that the spectral density function of y_{it} is a linear combination of the normalized spectral densities $g_j(\omega)$ of the canonical variables. In fact, this is the same linear combination as the autocovariance function $\gamma_{ii,k}$ of y_{it} is of the autocorrelation functions λ_j^k of the canonical variables. Recall the latter relationship given by (59) in Chapter 3:

$$\gamma_{ii,k} = \sum_j\sum_m b_{ij}b_{im}\gamma^z_{jm,0}\lambda_j^k. \tag{60}$$

By taking the Fourier transform on both sides of (60) we could have obtained (59) in this chapter directly. Thus we see that the spectral density function of a time series generated by a system of linear stochastic difference equations is a linear combination of the spectral densities of first-order univariate autoregressions with the form (56).

If all the roots in a linear deterministic system are real and positive, the time paths cannot have prolonged oscillations. An interesting question is

whether the time series generated by a linear stochastic system can have pronounced cycles in the form of a peak in the spectral density function if all roots are real and positive. The answer is yes because a linear combination of the spectral densities $g_j(\omega)$ formed by real and positive roots λ_j can have a relative maximum for $0<\omega<\pi$. Consider the following example of a bivariate first-order system.

$$\lambda_1=.1 \quad \lambda_2=.9$$

$$W=\begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix} \tag{61}$$

$$(b_{11}\ b_{12})=(1 \quad -.01)$$

The first time series y_{1t} will have a spectral density equal to π^{-1} times

$$[1 \quad -.01]\begin{bmatrix} g_1(\omega) & 0 \\ 0 & g_2(\omega) \end{bmatrix}\begin{bmatrix} \frac{1}{.99} & \frac{.8}{.91} \\ \frac{.8}{.91} & \frac{1}{.19} \end{bmatrix}\begin{bmatrix} 1 \\ -.01 \end{bmatrix}$$

$$=\frac{.9913}{1.01-.2\cos\omega}-\frac{.001570}{1.81-1.8\cos\omega}. \tag{62}$$

Selected values of the function (62) are

$\omega/2\pi$	0	1/32	1/16	2/16	3/16	4/16	5/16	6/16	7/16	8/16
$\pi f_{11}(\omega)$	1.067	1.183	1.191	1.138	1.061	.981	.912	.860	.829	.819

The peak of the spectral density at $\omega/2\pi$ approximately equal to $\frac{1}{20}$ can be discerned. This example shows how much the incorporation of random disturbances can alter the conclusion from a deterministic theory.

4.11* A NOTE ON SPECTRAL ANALYSIS

The concepts of spectral and cross-spectral densities are useful for studying cyclical properties of econometric models. In addition, and more prevalently, they are useful for extracting cyclical properties of observed economic time series without the intervention of an econometric model.

Thus, assuming that certain time series are covariance stationary, or approximately so, probably after adjustments for trends by fitting trend functions or by taking first differences and the like, we may wish to estimate the spectral and cross-spectral densities directly from the data. This process of estimation is called *spectral analysis* of time series data. No econometric models in the form of systems of interdependent dynamic equations are needed. The only hypothesis employed is concerned with the form of the trend, if trend is taken out, with the appropriate transformation of variables such as taking logarithm and with the smoothness of the spectral density function. Little theory is required. Spectral analysis can also be applied, not to economic data directly but to data generated from stochastic simulations of an econometric model estimated from economic data. This will enable us to study certain dynamic properties of the econometric model. Such an undertaking differs from spectral analysis of raw data and also from the analytical derivation of spectral properties from a model as we have done in this chapter.

Given time series data for a variable y_{1t} $(t=1,\ldots,N)$, it seems reasonable to utilize the periodic regression coefficients a_{1j} and b_{1j} of (21) in Section 4.4 to estimate the spectral density at ω_j. The spectral density at ω_j is the common variance of the coefficients a_{1j} and b_{1j}. There are $N/2$ of these variances that can be calculated from a sample of size N. Each variance has to cover a range of ω values equal in width to $2\pi/N$ if ω is a continuous variable from 0 to π. Therefore a possible estimate of the spectral density at ω_j is the variance $\frac{1}{2}(a_{1j}^2+b_{1j}^2)$ divided by the width $2\pi/N$:

$$\begin{aligned} I(\omega_j)&=\tfrac{1}{2}\left(a_{1j}^2+b_{1j}^2\right)\frac{N}{2\pi} \\ &=\frac{1}{\pi N}\left[\left(\sum_{t=1}^{N} y_{1t}\cos\omega_j t\right)^2+\left(\sum_{t=1}^{N} y_{1t}\sin\omega_j t\right)^2\right] \\ &=\frac{1}{\pi N}\left\{\sum_{t=1}^{N} y_{1t}y_{1t}(\cos\omega_j t\cos\omega_j t+\sin\omega_j t\sin\omega_j t)\right. \\ &\quad+\sum_{t=1}^{N-1} y_{1t}y_{1,t+1}[\cos\omega_j t\cos(\omega_j t+\omega_j)+\sin\omega_j t\sin(\omega_j t+\omega_j)] \\ &\quad\left.+\sum_{t=1}^{N-2} y_{1t}y_{1,t+2}[\cos\omega_j t\cos(\omega_j t+2\omega_j)+\sin\omega_j t\sin(\omega_j t+2\omega_j)]+\cdots\right\} \\ &=\frac{1}{\pi}\sum_{k=-N+1}^{N-1} c_{11,k}\cdot\cos\omega_j k, \end{aligned} \tag{63}$$

where $c_{11,k}$ is the sample autocovariance

$$c_{11,k}=c_{11,-k}=\frac{1}{N}\sum_{t=1}^{N-k}y_{1t}y_{1,t+k}. \tag{64}$$

Thus the use of the arithmetic mean of the squares of the regression coefficients a_{1j} and b_{1j} amounts to the same thing as the application of a cosine transform to the sample autocovariance $c_{11,k}$.

What is the sampling property of the estimate of spectral density given by (63)? If y_{1t} is normally and independently distributed, each regression coefficient a_{1j} and b_{1j} computed by the method of least squares, using (20), will also be normally and independently distributed. Each will have a variance equal to $2/N$ times the variance of y_{1t}, according to least-squares theory. The variables $a_{1j}\sqrt{N/2}$ and $b_{1j}\sqrt{N/2}$ are independently normal and have variance equal to the variance of y_{1t}. The estimate $I(\omega_j)$ in (63) can be written as

$$\left[\left(a_{ij}\sqrt{\frac{N}{2}}\right)^2+\left(b_{2j}\sqrt{\frac{N}{2}}\right)^2\right](2\pi)^{-1}.$$

Except for the factor $(2\pi)^{-1}$, it is the sum of the squares of two independent normal random variables with a common variance. Therefore, except for a factor, it is distributed as a chi-square variable with two degrees of freedom. The main point is that, no matter how large the sample size N, the estimate $I(\omega_j)$ for each ω_j will be distributed as a chi-square random variable with two degrees of freedom and thus will not converge to a constant. This means that a consistent estimate of the spectral density cannot be obtained by using $I(\omega_j)$.

In order to obtain a consistent estimate of the spectral density, it is required to modify (63) by applying a set of weights w_k to the sample autocovariance functions $c_{11,k}$ before performing the cosine transform:

$$\hat{f}_{11}(\omega)=\frac{1}{\pi}\sum_{k=-\infty}^{\infty}w_k c_{11,k}\cdot\cos\omega k. \tag{65}$$

The weighting function w_k is called the *lag window*. Specialists in spectral analysis have proposed various windows and investigated their properties. A simple one is the Bartlett window. For some m smaller than N this

window is

$$w_k=\begin{cases}1-\dfrac{|k|}{m}, & 0\leqslant|k|\leqslant m,\\ 0, & \text{otherwise.}\end{cases}$$

It ignores those sample autocovariances $c_{11,k}$ with lag k larger than m that are computed from fewer and fewer observations as k increases in absolute value. For this particular window the weights w_k are linearly declining with the absolute value of k. Others have proposed windows that are different decreasing functions of $|k|$, but there is no need for us to pursue the subject here. These windows can also be applied to the sample cross-covariance functions before the appropriate cosine and sine transforms are formed for the consistent estimation of cross-spectral densities. This process is termed *cross-spectral analysis*. Computer programs are available to perform spectral analysis and cross-spectral analysis of time series data.

In this chapter our main concern has been to introduce the important tools of spectral and cross-spectral densities for the analysis of stochastic dynamic systems and to derive these functions for linear systems of stochastic difference equations. We turn to some applications of these tools in economics in Chapter 5.

Readers interested in studying time series analysis from the frequency point of view may consult Anderson (1971), Box and Jenkins (1970), Cox and Miller (1965), Dhrymes (1970), Hannan (1960), Kendall and Stuart (1966), and Whittle (1963). Blackman and Tukey (1958) and Parzen (1961) contain early contributions to spectral analysis. On applications of spectral analysis in economics important references include Fishman (1969), Granger and Hatanaka (1964), Granger and Morgenstern (1970), and Nerlove (1964), as well as many other studies of special economic problems. Part of the material in this chapter is based on Chow (1968).

PROBLEMS

1. Plot the spectral densities of the time series $y_t=ay_{t-1}+u_t$ for $a=.8,\ .2,\ .0,-.2$ and $-.8$.
2. Find and plot the spectral densities of the time series

$$y_t=.9y_{t-1}-.8y_{t-2}+u_t.$$

3. Consider y_t and y_{t-1} as two separate time series that satisfy a first-order system of stochastic difference equations obtained by transforming the equation $y_t=.9y_{t-1}-.8y_{t-2}+u_t$. Plot the cross-spectral density function for these two time series. Comment on your answer.
4. Find and plot the spectral density function of y_{2t} as defined by (7).
5. Calculate two values at $\omega=0$ and $\omega=\pi/2$ for the phase-difference spectral density

function and the cross-amplitude spectral density function of the two time series defined by (7), using (15) and the result from Section 3.10. Interpret your answer.

6. Prove Equations 18 in this chapter. *Hint*: You may write $\cos\omega_j t = \frac{1}{2}(e^{i\omega_j t} + e^{-i\omega_j t})$; $\sin\omega_j t = -(i/2)(e^{i\omega_j t} - e^{-i\omega_j t})$. Use (17). Once it is shown that either sum is equal to $N/2$, the other sum can be evaluated by using $\sin^2\omega_j t + \cos^2\omega_j t = 1$.

7. Prove (19) in this chapter. *Hint*: See the hint in Problem 6.

8. Obtain a quarterly series of real GNP in recent decades consisting of an even number of N (> 40) observations. Use a computer to decompose the series into cosine and sine functions as in (16). Plot the contribution of each periodic component to the sample variance of the series.

9. Fit the model $y_t = \alpha y_{t-1} + u_t$ to the data obtained in Problem 8. Plot the power spectrum of the time series generated by the fitted model. Compare it with the graph of Problem 8, if you have done that problem.

10. Show that the covariance between y_{1t} and $y_{2,t-k}$ defined by (23) and (30) equals

$$\sum_j E[\alpha_1(\omega_j)\alpha_2(\omega_j)]\cos\omega_j k + \sum_j E[\beta_1(\omega_j)\alpha_2(\omega_j)]\sin\omega_j k.$$

11. Calculate the gain $G_{12}(\omega)$ for the two time series defined by (7) at $\omega = 0$ and $\omega = \pi/2$. Interpret your answer.

12. Calculate the coherence for the two time series defined by (7) at $\omega = 0$ and $\omega = \pi/2$. Interpret your answer.

13. Calculate the gain $G_{21}(\omega)$ for the two time series defined by (7) at $\omega = 0$ and $\omega = \pi/2$. Compare your answer with that of Problem 11.

14. What is the relation between the gains $G_{12}(\omega)$ and $G_{21}(\omega)$ and the coherence $R^2_{12}(\omega)$?

15. What are the gains $G_{12}(\omega)$ and $G_{21}(\omega)$ and the coherence $R^2_{12}(\omega)$ for the two time series defined by Problem 3.

16. Specify the canonical variables for the system $y_t = .9y_{t-1} - .8y_{t-2} + u_t$. Find their cross-covariance function and from that their cross-spectral density function.

17. Specify the canonical variables for the system of (7). Find their cross-covariance function and from that their cross-spectral density function.

18. Provide numerically a bivariate first-order system of stochastic difference equations from which the spectral density function of (62) for y_{1t} can be derived. Evaluate the spectral density of y_{2t} from this system at several values of ω. Does it also have a peak? Explain.