CHAPTER

5

Multivariate Spectral Models and Their Applications

5.1 INTRODUCTION

It is more the rule than the exception that real, time-varying physical processes require more than one measurement to adequately describe their behavior. Thus, the position or state of the process at each instant of time will be represented by a vector of time-dependent measurements

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \end{pmatrix},$$

called a *multivariate (vector, multidimensional) time series.* We will consider only those multivariate time series models for which the components are univariate time series, either all in continuous time or all in discrete time, possessing power spectra. As in Chapter 2, will will deal most extensively with stochastic models for which the components are stationary processes.

These models apply quite well to a variety of real phenomena. In geophysics they are used, for example, to describe wind velocities (two components of velocity at each of, say, n levels leads to a $2n$-dimensional series), the oscillation of the earth at a given location (the vertical and two components of horizontal motion produce a three-dimensional time series) and sea state over a given region (sea heights at p locations in the region generate a p-dimensional series). Economic systems characteristically require several descriptive variables such as price, available supply, and demand among

others. Under stable conditions, the "stationary" models are reasonably accurate. Even when conditions are not stable, these models provide useful results over restricted time periods. Early applications of multivariate time series analysis were made in engineering, particularly in the areas of communication and control theory. Since then, applications of these methods based on the models of this chapter have pervaded almost all of the physical and engineering sciences. A bibliography of applications of both the univariate and multivariate theory to the physical sciences and engineering is given by Tukey (1959).

More recently, these models have been applied to problems in the social, biological, and medical sciences as quantitative measurement techniques have been developed. Time series with large numbers of components are common in these areas. For example, the recording of EEG data for the study of brain function [see Walter *et al.* (1966)] often requires in excess of ten recording channels and results in a time series of as many dimensions. The future of time series analysis in these disciplines appears especially promising.

Time series models can be constructed with varying degrees of complexity depending on the purpose to which they are to be put and the knowledge available about the physical mechanism generating the data. At the purely descriptive level, a simple model can be fitted to the data much as polynomials are fitted to regression curves in statistics. The object is to fit the model to the data as closely as possible without trying to "understand" the underlying generating mechanism. For reasons that will be discussed in Chapter 7, the finite-parameter models defined therein and their multivariate counterparts are the ones most widely used for this purpose. Many important applications of time series analysis can be effectively made at this level. This is amply demonstrated by Box and Jenkins (1970), for example.

However, one of the more important uses of time series analysis is to improve our understanding, either qualitatively or quantitatively, of various properties of the generating mechanism. For this purpose, it is convenient to have the natural parameters of the time series model related in a simple way to the physical characteristics of the mechanism—or more precisely—to the parameters of a mathematical model of the mechanism. The spectra of the models to be discussed in this chapter will be seen to have this property in many applications. Since spectra can be readily estimated by the methods to be covered in Chapter 8, the validation of physical models by comparing predicted and measured spectra and the estimation of model parameters can be readily carried out.

We will see that by a simple application of the rules for operating with expectations summarized in Section 1.4, the spectra of time series derived from other time series via linear filters can be readily calculated. In this way a large and flexible class of models can be generated which has a number of

important applications and for which the parameters are easily computed. This will be illustrated by means of applications many of which have been taken from the time series literature.

5.2 THE SPECTRUM OF A MULTIVARIATE TIME SERIES—WIENER THEORY

Let

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \end{pmatrix}, \qquad -\infty < t < \infty,$$

be a vector of real-valued functions for which all limits of the form

$$C_{j,k}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x_j(t+\tau) x_k(t)\, dt \tag{5.1}$$

exist for $1 \leq j, k \leq p$, and $-\infty < \tau < \infty$. Although we treat the continuous time parameter case in this definition, all of the theory applies to discrete time series as well with the modifications spelled out in Chapter 3.

The functions $C_{j,j}(\tau)$ are the autocovariances of the time series $x_j(t)$ defined in Section 2.2. For $j \neq k$, $C_{j,k}(\tau)$ is called the *cross-covariance function* of $x_j(t)$ and $x_k(t)$. An application of the Schwarz inequality yields

$$|C_{j,k}(\tau)| \leq (C_{j,j}(0) C_{k,k}(0))^{1/2}, \qquad -\infty < \tau < \infty. \tag{5.2}$$

Consequently, all covariance functions are bounded. Moreover, by a change of variables it is easily seen that

$$C_{j,k}(-\tau) = C_{k,j}(\tau). \tag{5.3}$$

Thus, although the autocovariances are even functions, as was previously noted in Section 2.2, the cross-covariance functions are not. This has an important effect on the spectral representation of the cross-covariances.

For each $j \neq k$ there exists a (unique) measure $F_{j,k}(A)$, called the *cross-spectral distribution* of $x_j(t)$ and $x_k(t)$, such that

$$C_{j,k}(\tau) = \int e^{i\lambda\tau} F_{j,k}(d\lambda). \tag{5.4}$$

In order to deal with the discrete and continuous time parameter cases together, we will not explicitly specify the frequency range in expressions involving the frequency variable. Thus, the limits of integration are omitted

in (5.4) with the understanding that they are $-\infty$ and ∞ in the continuous-time case and $-\pi$ and π for discrete time.

As indicated in Section 2.2, in order for the Fourier transform of a real-valued function to be real-valued, it is necessary that the function be even. Since the cross-covariance function fails in this respect, *the cross-spectral distribution is generally complex-valued (with nonvanishing imaginary part)*. In this it differs from the spectral distributions $F_{j,j}(A)$.

A partial symmetry results from (5.3); namely,

$$F_{j,k}(-A) = F_{k,j}(A). \tag{5.5}$$

(Recall that $-A = \{-\lambda : \lambda \in A\}$.) To see this, note that

$$C_{j,k}(-\tau) = \int e^{i\lambda(-\tau)} F_{j,k}(d\lambda) = \int e^{i\mu\tau} F_{j,k}(-d\mu)$$

by the change of variables $\mu = -\lambda$. However, by (5.3) this is the spectral representation of $C_{k,j}(\tau)$. The uniqueness of the spectral distribution then yields (5.5).

Similarly, it can be shown that

$$F_{j,k}(A) = \overline{F_{k,j}(A)}. \tag{5.6}$$

This and other properties of the spectral distribution are easily demonstrated using the stochastic model of the next section. Consequently, we will defer the statement of these results until later.

The matrix representation of the auto- and cross-spectral distributions is quite useful. It is

$$\mathbf{F}(A) = \begin{bmatrix} F_{1,1}(A) & F_{1,2}(A) & \cdots & F_{1,p}(A) \\ F_{2,1}(A) & F_{2,2}(A) & \cdots & F_{2,p}(A) \\ \vdots & \vdots & & \vdots \\ F_{p,1}(A) & F_{p,2}(A) & \cdots & F_{p,p}(A) \end{bmatrix}. \tag{5.7}$$

Hereafter, we will denote such arrays more briefly by using the symbol $[a_{j,k}]$ to represent the matrix with element $a_{j,k}$ in the jth row and kth column. Relation (5.6) indicates that $\mathbf{F}(A)$ *is a Hermitian matrix*.

In the models we will consider, the spectral distribution matrix (5.7) or its equivalent, the matrix of auto- and cross-covariances,

$$\mathbf{C}(\tau) = [C_{j,k}(\tau)],$$

constitutes the complete set of parameters for the multivariate time series $\mathbf{x}(t)$. Consequently, the only new elements presented by the multivariate theory are the relationships between each pair of time series measured by the cross-spectra or the cross-covariances.

As we will see, this restriction to what are called *second-order parameters* still allows a rich and useful theory of interrelationships to be developed which very closely parallels the multivariate correlation theory of statistics. This will be discussed in Section 5.6.

The decomposition of $\mathbf{F}(A)$ into discrete and continuous components is accomplished as in the univariate case. The *spectral functions* are

$$p_{j,k}(\lambda) = F_{j,k}(\{\lambda\}) \tag{5.8}$$

and the *discrete spectral distributions* are then

$$F^{(\mathrm{d})}_{j,k}(A) = \sum_{\lambda_l \in A} p_{j,k}(\lambda_l) \tag{5.9}$$

where $\lambda_1, \lambda_2, \ldots$ are the frequencies for which $p_{j,j}(\lambda)\, p_{k,k}(\lambda) > 0$ for at least one pair of indices j, k. In matrix notation,

$$\mathbf{p}(\lambda) = [p_{j,k}(\lambda)]$$

and

$$\mathbf{F}^{(\mathrm{d})}(A) = [F^{(\mathrm{d})}_{j,k}(A)]. \tag{5.10}$$

As in Section 2.2, when the spectrum is discrete, the cross-spectral functions can be obtained from the cross-covariances by the expression

$$p_{j,k}(\lambda) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e^{-i\lambda\tau} C_{j,k}(\tau)\, d\tau$$

in the continuous-time case and by the corresponding limit

$$p_{j,k}(\lambda) = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{\tau=-L}^{L} e^{-i\lambda\tau} C_{j,k}(\tau)$$

for discrete time. The *spectral distribution functions* are, respectively,

$$F_{j,k}(\lambda) = F_{j,k}((-\infty, \lambda]) \quad \text{and} \quad F_{j,k}((-\pi, \lambda]).$$

The *spectral density functions* are then the derivatives of these distribution functions;

$$f_{j,k}(\lambda) = \frac{dF_{j,k}(\lambda)}{d\lambda} \tag{5.11}$$

and the continuous component of $\mathbf{F}(A)$ is

$$\mathbf{F}^{(c)}(A) = [F^{(c)}_{j,k}(A)], \tag{5.12}$$

where

$$F^{(c)}_{j,k}(A) = \int_A f_{j,k}(\lambda)\, d\lambda. \tag{5.13}$$

When $\int_{-\infty}^{\infty} |C_{j,k}(\tau)|\, d\tau < \infty$ or $\int_{-\infty}^{\infty} |C_{j,k}(\tau)|^2 d\tau < \infty$ in the case of continuous time, the spectral densities can be obtained by Fourier transformation (Example 1.4);

$$f_{j,k}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda\tau} C_{j,k}(\tau)\, d\tau. \tag{5.14}$$

In discrete time, if $\sum_{-\infty}^{\infty} |C_{j,k}(\tau)|^2 < \infty$, then

$$f_{j,k}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} e^{-i\lambda\tau} C_{j,k}(\tau). \tag{5.15}$$

(See Example 1.2.)

An important (multidimensional) parameter of the process is the spectral density matrix

$$\mathbf{f}(\lambda) = [f_{j,k}(\lambda)].$$

The standard measures of association will be defined only when the time series has a continuous spectrum and, thus, will be functions of the elements of $\mathbf{f}(\lambda)$. For this reason, the elements of this matrix will be the objects of interest for the statistical procedures to be derived in Chapter 8.

As in the univariate case, it should be stressed that the multivariate spectrum is defined for a large class of time series models both stochastic and nonstochastic and the weakly stationary processes form a rather small subclass of these models. Consequently, the idea of a spectrum and the various measures of association among time series to be derived in this chapter are much more widely applicable than one would surmise from our rather disproportionate coverage of weakly stationary processes. However, from the viewpoint of constructing probability models which have the general spectral structure of the Wiener theory, the weakly stationary processes play a central role. We consider them next.

5.3 MULTIVARIATE WEAKLY STATIONARY STOCHASTIC PROCESSES

Let $\{X_j(t);\ j = 1, 2, \ldots, p;\ -\infty < t < \infty\}$ be a family of real-valued random variables on the same probability space. In order for the vector of stochastic processes

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_p(t) \end{pmatrix}, \qquad -\infty < t < \infty,$$

to be a multivariate weakly stationary stochastic process, each univariate process $X_j(t)$ must be weakly stationary in the sense defined in Section 2.3 and, in addition, *the correlation between processes must be stationary*. These hypotheses can be summarized as follows.

(i) $EX_j(t) = m_j, \quad j = 1, 2, \ldots, p.$

As before, the constants m_j will all be set equal to zero by the expedient of replacing each process by the residual $X_j(t) - m_j$. Then we assume that the covariances are all finite and satisfy the conditions

(ii) $EX_j(t+\tau)X_k(t) = C_{j,k}(\tau), \quad -\infty < t, \tau < \infty, 1 \leq j, k \leq p.$

That is, the covariances depend on τ, the lag between the time arguments, but not on t. When $j = k$, this is the condition of covariance stationarity (2.20) which was basic to the theory of Chapter 2. When $j \neq k$, it is the condition of *stationary correlation* between processes.

With these assumptions, the covariance functions $C_{j,k}(\tau)$ enjoy all of the properties of the Wiener covariances defined by (5.1). Consequently, the spectral representation (5.4) and the subsequent properties of the spectrum, (5.5)–(5.13), are valid. A direct and useful method for deriving properties of the spectrum is obtained from the *multivariate version of the spectral representation theorem*. This method will be demonstrated in Section 5.4. As before, since each univariate series $X_j(t)$ is weakly stationary a spectral representation exists,

$$X_j(t) = \int e^{i\lambda t} Z_j(d\lambda).$$

The random spectral measures $Z_j(A)$ have as values complex-valued random variables for each j and each set A. They are related to the spectral distributions by the basic expression

$$EZ_j(A)\overline{Z_k(B)} = F_{j,k}(A \cap B). \tag{5.16}$$

As before, $EZ_j(A) = 0$ for all j and A.

By the Schwarz inequality, applied to the inner product

$$\langle Z_j(A), Z_k(A) \rangle = EZ_j(A)\overline{Z_k(A)},$$

we obtain the inequality

$$|F_{j,k}(A)| \leq \left(F_{j,j}(A)F_{k,k}(A)\right)^{1/2}. \tag{5.17}$$

It follows from this that the cross-spectral measure $F_{j,k}(A)$ can be nonzero only if both component processes $X_j(t)$ and $X_k(t)$ have positive power on A. Auxiliary inequalities which follow from this are

$$|p_{j,k}(\lambda)| \leq \left(p_{j,j}(\lambda)p_{k,k}(\lambda)\right)^{1/2} \tag{5.18}$$

and

$$|f_{j,k}(\lambda)| \leq (f_{j,j}(\lambda) f_{k,k}(\lambda))^{1/2} \tag{5.19}$$

for the spectral functions and spectral densities.

In terms of the symbolic intervals discussed in Section 2.5, expressions (5.16) and (5.17) can be combined to yield

$$EZ_j(d\lambda)\overline{Z_k(d\mu)} = \begin{cases} F_{j,k}(d\lambda), & \text{if } \mu = \lambda, \\ 0, & \text{if } \mu \neq \lambda. \end{cases} \tag{5.20}$$

This expression can be used operationally as indicated in the following statement:

Extension of the Basic Property *If $g(\lambda)$ and $h(\lambda)$ are complex-valued functions for which*

$$\int |g(\lambda)|^2 F_{j,j}(d\lambda) < \infty \qquad \text{and} \qquad \int |h(\lambda)|^2 F_{k,k}(d\lambda) < \infty,$$

then

$$\int g(\lambda) Z_j(d\lambda) \qquad \text{and} \qquad \int h(\lambda) Z_k(d\lambda)$$

are well-defined random variables with zero means and finite variances. Moreover,

$$E\left(\int g(\lambda) Z_j(d\lambda)\right)\left(\overline{\int h(\mu) Z_k(d\mu)}\right) = \iint g(\lambda)\overline{h(\mu)} EZ_j(d\lambda)\overline{Z_k(d\mu)}$$

$$= \int g(\lambda)\overline{h(\lambda)} F_{j,k}(d\lambda). \tag{5.21}$$

The equality of the first and last expressions in (5.21) is guaranteed by the theory outlined in the Appendix. Convention (5.20) leads to the correct result in this and other calculations in which it will be used later in the chapter. This will prove to be a most useful method for calculating spectra in applications of the theory. In particular, with $g(\lambda) = e^{i\lambda(t+\tau)}$ and $h(\lambda) = e^{i\lambda t}$, (5.21) yields the spectral representation (5.4) of the auto- and cross-covariances.

Properties (5.6) of the cross spectra follow easily from (5.16). Expression (5.5) requires the additional relation

$$Z_j(-A) = \overline{Z_j(A)}, \tag{5.22}$$

also noted in Section 2.4.

If we write the spectral measures in vector form

$$\mathbf{Z}(A) = \begin{pmatrix} Z_1(A) \\ Z_2(A) \\ \vdots \\ Z_p(A) \end{pmatrix},$$

then condition (5.16) implies that $\mathbf{Z}(A)$ is uncorrelated with $\mathbf{Z}(B)$ if A and B are disjoint sets. If the multivariate process $\mathbf{X}(t)$ is Gaussian, then with the additional condition $A \cap -B = \varnothing$, $\mathbf{Z}(A)$ and $\mathbf{Z}(B)$ are independent.

The spectral representation of the process can be written in vector form as

$$\mathbf{X}(t) = \int e^{i\lambda t}\mathbf{Z}(d\lambda), \tag{5.23}$$

with the understanding that the integral is applied to each coordinate of the vector $e^{i\lambda t}\mathbf{Z}(d\lambda)$.

The correlational relationship (5.16) can be written succinctly in matrix form as

$$E\mathbf{Z}(A)\mathbf{Z}(B)^* = \mathbf{F}(A \cap B), \tag{5.24}$$

where $*$ denotes the combined operations of taking the transpose and forming the complex conjugates of the vector elements. In terms of this operation, the Hermetian relation (5.6) can be written

$$\mathbf{F}(A) = \mathbf{F}(A)^*.$$

The decomposition of the multivariate spectral distribution of $\mathbf{X}(t)$ into discrete and continuous components can be given in terms of the symbolic interval $d\lambda$ as

$$\mathbf{F}(d\lambda) = \mathbf{p}(\lambda) + \mathbf{f}(\lambda)\, d\lambda,$$

where, on the right-hand side, $d\lambda$ represents the (scalar) length of this interval.

The linearity property of expectation for random matrices given in Section 1.4 permits us to establish another important property of $\mathbf{F}(A)$: If $\mathbf{a}$ denotes any $p \times 1$ vector of complex numbers, then

$$E|\mathbf{a}^*\mathbf{Z}(A)|^2 \geq 0 \qquad \text{for every } A.$$

However,

$$\begin{aligned} |\mathbf{a}^*\mathbf{Z}(A)|^2 &= \mathbf{a}^*\mathbf{Z}(A)\overline{(\mathbf{a}^*\mathbf{Z}(A))} \\ &= \mathbf{a}^*\mathbf{Z}(A)(\mathbf{Z}(A)^*\mathbf{a}) = \mathbf{a}^*\mathbf{Z}(A)\mathbf{Z}(A)^*\mathbf{a}. \end{aligned}$$

Here, we have used the fact that $\overline{a^*\mathbf{Z}(A)} = (a^*\mathbf{Z}(A))^*$ and that $*$ operates on matrix products according to the rule

$$(\mathbf{UV})^* = \mathbf{V}^*\mathbf{U}^*.$$

Then taking expectations and using the fact that the expectation of a nonnegative random variable is nonnegative, we obtain

$$\mathbf{a}^*\mathbf{F}(A)\mathbf{a} \geq 0.$$

This establishes that the matrix of spectral distributions is *nonnegative definite* (Tucker, 1962, p. 144). (An alternate terminology is *positive semidefinite*). This is an important property which carries over to the matrices of spectral functions and spectral density functions $\mathbf{p}(\lambda)$ and $\mathbf{f}(\lambda)$ as well. Among the many implications of this property are inequalities (5.17)–(5.19). Various features of the multivariate spectral parameters to be defined later also depend on this property.

If two component processes $X_j(t)$, $X_k(t)$ are uncorrelated (Section 2.6), then

$$C_{j,k}(\tau) = 0, \qquad -\infty < \tau < \infty.$$

By the uniqueness of the spectral representation of the covariances (5.4), this implies that the cross-spectral distribution is everywhere zero,

$$F_{j,k}(A) = 0 \qquad \text{all } A.$$

More Than One Vector Process

Let $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ be $p \times 1$ and $q \times 1$ multivariate weakly stationary processes, respectively. When we consider more than one vector process it will be tacitly assumed that all components of all processes are stationarily correlated. In fact, all properties of such processes can be derived from the theory for a single weakly stationary vector process by forming the "stacked" process

$$\mathbf{W}(t) = \begin{pmatrix} \mathbf{X}(t) \\ \mathbf{Y}(t) \\ \vdots \end{pmatrix}.$$

In particular, the $p \times q$ matrix of covariances of $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ has elements

$$C^{X,Y}_{j,k}(\tau) = EX_j(t+\tau)Y_k(t).$$

In vector notation,

$$\mathbf{C}^{X,Y}(\tau) = E\mathbf{X}(t+\tau)\mathbf{Y}(t)^*. \tag{5.25}$$

If the spectral distribution of these processes is defined by

$$\mathbf{F}^{X,Y}(A \cap B) = E\mathbf{Z}^X(A)\mathbf{Z}^Y(B)^*,$$

then the matrix version of the calculation which yielded (5.21) can be used to derive the spectral representation of $\mathbf{C}^{X,Y}(\tau)$ as follows:

$$\begin{aligned}\mathbf{C}^{X,Y}(\tau) &= E\mathbf{X}(t+\tau)\mathbf{Y}(t)^* \\ &= E\left(\int e^{i\lambda(t+\tau)}\mathbf{Z}^X(d\lambda)\right)\left(\int e^{i\mu t}\mathbf{Z}^Y(d\mu)\right)^* \\ &= \iint e^{i\lambda\tau}e^{i(\lambda-\mu)t}E\mathbf{Z}^X(d\lambda)\mathbf{Z}^Y(d\mu)^* \\ &= \int e^{i\lambda\tau}\mathbf{F}^{X,Y}(d\lambda).\end{aligned}$$

The vector processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are said to be *uncorrelated* if every component of $\mathbf{X}(t)$ is uncorrelated with every component of $\mathbf{Y}(t)$. Equivalent conditions are $\mathbf{C}^{X,Y}(\tau) = \mathbf{0}$ for all τ, thus $\mathbf{F}^{X,Y}(A) = \mathbf{0}$ for every A, where $\mathbf{0}$ is the $p \times q$ matrix with all zero elements.

The decomposition

$$\mathbf{F}^{X,Y}(d\lambda) = \mathbf{p}^{X,Y}(\lambda) + \mathbf{f}^{X,Y}(\lambda)\,d\lambda$$

holds in the case of two vector processes, where $\mathbf{p}^{X,Y}(\lambda)$ and $\mathbf{f}^{X,Y}(\lambda)$ are obtained from $\mathbf{F}^{X,Y}(A)$ as in (5.8) and (5.11).

5.4 LINEAR FILTERS FOR MULTIVARIATE TIME SERIES

We again restrict our attention to the weakly stationary model, although the results we will derive for spectra are valid for the more inclusive model of Section 5.2. Let $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ be weakly stationary processes with p and q components, respectively. The new feature of a *multivariate linear filter* with input $\mathbf{X}(t)$ and output $\mathbf{Y}(t)$ is that each component of $\mathbf{Y}(t)$ will, in general, be influenced by every component of $\mathbf{X}(t)$. The "black box" representation of such a filter is pictured in Fig. 5.1. The nature of the influence of $X_k(t)$

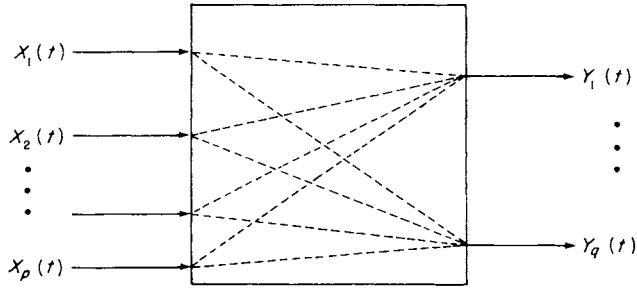

Fig. 5.1 *Representation of a multivariate linear filter as a black box with p input leads and q output leads. Each output is coupled with each input.*

on $Y_j(t)$ is that of an ordinary linear filter in the sense discussed in Chapter 4. Then, the total influence of all components of $\mathbf{X}(t)$ on $Y_j(t)$ is simply the sum of the contributions of each component. This can be stated more precisely by using the spectral representations of the input and output processes: Let $B_{j,k}(\lambda)$ denote the transfer function of the linear filter from $X_k(t)$ to $Y_j(t)$. This would be represented by the "black box" with all inputs except $X_k(t)$ turned off and measurements being made only at the $Y_j(t)$ terminal. With all inputs running, what would be observed at this terminal is

$$\begin{aligned} Y_j(t) &= \sum_{k=1}^{p} \int e^{i\lambda t} B_{j,k}(\lambda) Z_k^X(d\lambda) \\ &= \int e^{i\lambda t} \sum_{k=1}^{p} B_{j,k}(\lambda) Z_k^X(d\lambda), \end{aligned} \tag{5.26}$$

where $Z_k^X(d\lambda)$ is the random spectral measure of $X_k(t)$. That is, the spectral measure of $Y_j(t)$ is

$$Z_j^Y(d\lambda) = \sum_{k=1}^{p} B_{j,k}(\lambda) Z_k^X(d\lambda).$$

This can be written conveniently in matrix notation as

$$\mathbf{Z}^Y(d\lambda) = \mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda), \tag{5.27}$$

where $\mathbf{B}(\lambda)$ is the $(q \times p)$-dimensional matrix,

$$\mathbf{B}(\lambda) = [B_{j,k}(\lambda)].$$

If we let $\mathbf{L}$ denote the multivariate linear filter represented by the "black box," then $\mathbf{B}(\lambda)$ is its *transfer function* and (5.26) can be written in matrix form as

$$\mathbf{Y}(t) = \mathbf{L}(\mathbf{X}(t)) = \int e^{i\lambda t} \mathbf{B}(\lambda) \mathbf{Z}^X(d\lambda). \tag{5.28}$$

The multivariate filter will *match* the input if each term of the output in (5.26) has finite variance (power). That is, we require that

$$\int |B_{j,k}(\lambda)|^2 F_{k,k}^X(d\lambda) < \infty, \tag{5.29}$$

for $j = 1, 2, \ldots, q$ and $k = 1, 2, \ldots, p$. As we show in the Appendix, along with a more theoretical discussion of multivariate filters, this *matching condition* can be represented in matrix form as

$$\operatorname{tr} \int \mathbf{B}(\lambda) \mathbf{F}^X(d\lambda) \mathbf{B}(\lambda)^* < \infty, \tag{5.30}$$

where tr denotes the trace of the matrix. Moreover, the *spectral distribution of the output* can be calculated using matrix algebra and the linearity property

of expectation for random matrices (Section 1.4): From (5.27),

$$\begin{aligned}\mathbf{F}^Y(d\lambda) &= E\big(\mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda)\big)\big(\mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda)\big)^* \\ &= E\mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda)\mathbf{Z}^X(d\lambda)^*\mathbf{B}(\lambda)^* \\ &= \mathbf{B}(\lambda)\big(E\mathbf{Z}^X(d\lambda)\mathbf{Z}^X(d\lambda)^*\big)\mathbf{B}(\lambda).^*\end{aligned}$$

Thus,

$$\mathbf{F}^Y(d\lambda) = \mathbf{B}(\lambda)\mathbf{F}^X(d\lambda)\mathbf{B}(\lambda)^*. \tag{5.31}$$

This implies that the *spectral functions* and *spectral density functions* of input and output are

$$\mathbf{p}^Y(\lambda) = \mathbf{B}(\lambda)\mathbf{p}^X(\lambda)\mathbf{B}(\lambda)^* \tag{5.32}$$

and

$$\mathbf{f}^Y(\lambda) = \mathbf{B}(\lambda)\mathbf{f}^X(\lambda)\mathbf{B}(\lambda)^*. \tag{5.33}$$

Example 5.1 *Filters Which Transform Component Time Series Independently*

Suppose $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are both p-dimensional processes and let L_j be a univariate filter with transfer function $B_j(\lambda)$ such that

$$Y_j(t) = L_j\big(X_j(t)\big)$$

for $j = 1, 2, \ldots, p$. This situation occurs, for example, when each component has its own linear transmission channel which is completely separate from the channels of the other components. A prime example of this would be a telephone trunk line which carries thousands of individual, separate conversations within a single cable. Moreover, it is often the case that when multivariate time series are being prepared for spectrum analysis, different preprocessing filters are used on each component. The input and output spectra are then related by the expressions of this example. This model in some instances also applies to the same linear mechanism within which the component signals are polarized into noninterferring modes of transmission. Thus, two beams of light polarized at right angles to one another would be transformed simultaneously but independently by a piece of colored glass. To a great extent, the horizontal and vertical components of most types of seismic waves travel through the earth and are recorded independently. Coding techniques allow many signals to be mixed, transmitted, then unscrambled into their original form.

This type of filter has a simple matrix representation. The transfer function is

$$\mathbf{B}(\lambda) = \begin{bmatrix} B_1(\lambda) & 0 & \cdots & 0 \\ 0 & B_2(\lambda) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & B_p(\lambda) \end{bmatrix}.$$

That is, the one-dimensional transfer functions appear on the diagonal of the matrix and the off-diagonal elements are all zero. From the component representation

$$Z_j^Y(d\lambda) = B_j(\lambda)Z_j^X(d\lambda),$$

the output spectral distributions are seen to be

$$\begin{aligned} F_{j,k}^Y(d\lambda) &= EZ_j^Y(d\lambda)\overline{Z_k^Y(d\lambda)} \\ &= B_j(\lambda)\overline{B_k(\lambda)}F_{j,k}^X(d\lambda). \end{aligned} \tag{5.34}$$

This is also an immediate algebraic consequence of the form of $\mathbf{B}(\lambda)$ and expression (5.31). If $\mathbf{L}$ is the multivariate linear filter with this transfer function, then $\mathbf{L}$ matches any input $\mathbf{X}(t)$ for which L_j matches the component $X_j(t)$ for each $j = 1, 2, \ldots, p$.

A Method for Calculating Spectral Parameters

An important method for computing spectral parameters will be introduced in the next example.

Example 5.2 *The Multivariate Distributed Lag Model and a Method for Computing Spectra*

A discrete-time, weakly stationary, time series model that has been used in various fields, notably geophysics (Hamon and Hannan, 1963) and economics [see Fishman (1969) for a bibliography] can be written in the form

$$\mathbf{Y}(t) = \sum_{s=-r_1}^{r_2} \mathbf{a}(s)\mathbf{X}(t-s) + \boldsymbol{\eta}(t), \tag{5.35}$$

where $\mathbf{Y}(t)$ and $\boldsymbol{\eta}(t)$ are $(q \times 1)$-dimensional weakly stationary processes and $\mathbf{X}(t)$ is a $(p \times 1)$-dimensional process. This is the *multivariate distributed lag model.* In economics the coordinates of $\mathbf{X}(t)$ are called the *exogenous variables*, those of $\mathbf{Y}(t)$ the *endogeneous variables*, and $\boldsymbol{\eta}(t)$ is a *vector process of residuals.* For simplicity we assume all processes have zero means. A key assumption concerning this model is that the processes $\mathbf{X}(t)$ and $\boldsymbol{\eta}(t)$ are uncorrelated.

The objective of a statistical analysis of this model is to estimate, for given r_1 and r_2, the coefficients of the $r_1 + r_2 + 1$, $(q \times p)$-dimensional matrices $\mathbf{a}(s)$. Excellent treatments of this estimation problem are given by Fishman (1969) and Hannan (1970). We will be content to demonstrate a direct method for calculating the spectrum of $\mathbf{Y}(t)$ and the transfer function of the linear filter operating on $\mathbf{X}(t)$ which uses the spectral representations of the processes. This method of calculation is of fundamental importance for the determination of the spectral parameters of a large class of time series models.

The method is based on the following observations: Let

$$\mathbf{W}(t) = \mathbf{Y}(t) - \sum_{s=-r_1}^{r_2} \mathbf{a}(s)\mathbf{X}(t-s) - \boldsymbol{\eta}(t).$$

Then, $\mathbf{W}(t) \equiv 0$. This implies that the covariance matrix of $\mathbf{W}(t)$, $\mathbf{C}^W(\tau)$, is identically zero for all τ, thus

$$\mathbf{F}^W(A) = 0 \qquad \text{for all } A.$$

However, then,

$$\mathbf{Z}^W(A) = 0 \qquad \text{for all } A. \tag{5.36}$$

Now, applying the spectral representation (5.23) to both sides of (5.35), we obtain

$$\int e^{i\lambda t}\mathbf{Z}^Y(d\lambda) = \sum_{s=-r_1}^{r_2} \mathbf{a}(s) \int e^{i\lambda(t-s)}\mathbf{Z}^X(d\lambda) + \int e^{i\lambda t}\mathbf{Z}^\eta(d\lambda),$$

or

$$\mathbf{W}(t) = \int e^{i\lambda t}[\mathbf{Z}^Y(d\lambda) - \mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda) - \mathbf{Z}^\eta(d\lambda)], \tag{5.37}$$

where

$$\mathbf{B}(\lambda) = \sum_{s=-r_1}^{r_2} \mathbf{a}(s)e^{-i\lambda s}. \tag{5.38}$$

Now, the bracketed expression in (5.37) is $\mathbf{Z}^W(d\lambda)$. Thus, if we apply (5.36) to $\mathbf{Z}^W(d\lambda)$, *the original equation* (5.35) *is equivalent to the symbolic equation*

$$\mathbf{Z}^Y(d\lambda) = \mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda) + \mathbf{Z}^\eta(d\lambda) \qquad \textit{for all } \lambda. \tag{5.39}$$

This is the key equation of the method. All spectral parameters will be obtained from this expression by the application of standard rules of matrix algebra and expectation to $\mathbf{Z}^Y(d\lambda)$, $\mathbf{Z}^X(d\lambda)$ and $\mathbf{Z}^\eta(d\lambda)$ *as though they were vectors of valid, complex-valued random variables with finite variances.*

Note that this equation could have been written from (5.35) "by inspection." The form of the transfer function of the filter is a consequence of an easy multivariate extension of the rule for calculating transfer functions given in Section 4.2. Thus, it is generally a simple matter to obtain expressions such as (5.39) directly from the time domain representations of the processes.

We first calculate the spectral distribution $\mathbf{F}^{Y,X}(d\lambda)$ by multiplying both sides of (5.39) on the right by $\mathbf{Z}^X(d\lambda)^*$ and taking expectations:

$$\begin{aligned}\mathbf{F}^{Y,X}(d\lambda) &= E\mathbf{Z}^Y(d\lambda)\mathbf{Z}^X(d\lambda)^* \\ &= E[\mathbf{B}(\lambda)\mathbf{Z}^X(d\lambda)\mathbf{Z}^X(d\lambda)^* + \mathbf{Z}^\eta(d\lambda)\mathbf{Z}^X(d\lambda)^*] \\ &= \mathbf{B}(\lambda)\mathbf{F}^X(d\lambda).\end{aligned}$$

We have used the fact that $E\mathbf{Z}^{\eta}(d\lambda)\mathbf{Z}^{X}(d\lambda)^{*} = 0$ because $\mathbf{X}(t)$ and $\boldsymbol{\eta}(t)$ are uncorrelated processes.

Recall that

$$\mathbf{F}^{X}(d\lambda) = \mathbf{p}^{X}(\lambda) + \mathbf{f}^{X}(\lambda)\, d\lambda \qquad \text{and} \qquad \mathbf{F}^{X,Y}(d\lambda) = \mathbf{p}^{Y,X}(\lambda) + \mathbf{f}^{Y,X}(\lambda)\, d\lambda.$$

Thus, in the pure discrete case, the expression

$$\mathbf{B}(\lambda) = \mathbf{p}^{Y,X}(\lambda)\mathbf{p}^{X}(\lambda)^{-1}$$

determines the transfer function at those values of λ for which $\mathbf{p}^{X}(\lambda)$ is nonsingular. However, we will be primarily interested in the case of pure continuous spectra, since it is common practice to estimate and remove the discrete spectral component before analysis. Then,

$$\mathbf{B}(\lambda) = \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1} \tag{5.40}$$

whenever $\mathbf{f}^{X}(\lambda)$ is nonsingular. In this expression and all comparable expressions hereafter, when $d\lambda$ represents the length of the infinitesimal interval it is treated as though it were a positive number. Thus, it is simply cancelled when it appears to the same power in numerator and denominator of a ratio. Now, estimates of the spectral densities will provide an estimate $\hat{\mathbf{B}}(\lambda)$ of $\mathbf{B}(\lambda)$ at a finite set of frequencies and, since (5.38) represents this function as a Fourier series, estimates of the coefficient matrices can be obtained from a discrete version of the inversion formula,

$$\hat{\mathbf{a}}(s) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda s}\hat{\mathbf{B}}(\lambda)\, d\lambda.$$

This is the estimation procedure detailed by Hannan (1970, p. 475).

Returning to (5.39), the spectral distribution of $\mathbf{Y}(t)$ can be calculated as

$$\begin{aligned}\mathbf{F}^{Y}(d\lambda) &= E[\mathbf{B}(\lambda)\mathbf{Z}^{X}(d\lambda) + \mathbf{Z}^{\eta}(d\lambda)][\mathbf{B}(\lambda)\mathbf{Z}^{X}(d\lambda) + \mathbf{Z}^{\eta}(d\lambda)]^{*} \\ &= E\{\mathbf{B}(\lambda)\mathbf{Z}^{X}(d\lambda)\mathbf{Z}^{X}(d\lambda)^{*}\mathbf{B}(\lambda)^{*} + \mathbf{Z}^{\eta}(d\lambda)\mathbf{Z}^{X}(d\lambda)^{*}\mathbf{B}(\lambda)^{*} \\ &\quad + \mathbf{B}(\lambda)\mathbf{Z}^{X}(d\lambda)\mathbf{Z}^{\eta}(d\lambda)^{*} + \mathbf{Z}^{\eta}(d\lambda)\mathbf{Z}^{\eta}(d\lambda)^{*}\}.\end{aligned}$$

The expectation of the second and third terms are zero, since $\mathbf{X}(t)$ and $\boldsymbol{\eta}(t)$ are uncorrelated. Consequently, moving the expectation inside the braces, we obtain

$$\mathbf{F}^{Y}(d\lambda) = \mathbf{B}(\lambda)\mathbf{F}^{X}(d\lambda)\mathbf{B}(\lambda)^{*} + \mathbf{F}^{\eta}(d\lambda).$$

In the case of continuous spectra, this yields the following expression for the spectral densities,

$$\mathbf{f}^{Y}(\lambda) = \mathbf{B}(\lambda)\mathbf{f}^{X}(\lambda)\mathbf{B}(\lambda)^{*} + \mathbf{f}^{\eta}(\lambda).$$

Now, substituting (5.40) for $\mathbf{B}(\lambda)$ into this equation, we obtain

$$\mathbf{f}^{Y}(\lambda) = \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1}\mathbf{f}^{X,Y}(\lambda) + \mathbf{f}^{\eta}(\lambda),$$

where we have used the easily established result

$$\mathbf{f}^{Y,X}(\lambda)^{*} = \mathbf{f}^{X,Y}(\lambda).$$

This provides the following expression for the residual spectral density in terms of the spectral densities of the observed processes,

$$\mathbf{f}^{\eta}(\lambda) = \mathbf{f}^{Y}(\lambda) - \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1}\mathbf{f}^{X,Y}(\lambda). \tag{5.41}$$

Since $\boldsymbol{\eta}(t)$ measures the difference between $\mathbf{Y}(t)$ and the best "explanation" of $\mathbf{Y}(t)$ in terms of a linear function of $\mathbf{X}(t-s)$, $-r_1 \leq s \leq r_2$, the elements of $\mathbf{f}^{\eta}(\lambda)$ provide a measure of how good this "explanation" is as a function of frequency. A perfect fit, for example, would yield $\mathbf{f}^{\eta}(\lambda) = \mathbf{0}$ for all λ. An estimate of the residual spectral density can be obtained from estimates of the spectra of the observable processes by means of (5.41).

5.5 THE BIVARIATE SPECTRAL PARAMETERS, THEIR INTERPRETATIONS AND USES

Hereafter, we will restrict our attention to multivariate stochastic processes $\mathbf{X}(t)$ with continuous spectra, unless otherwise specified. Consequently, the basic (second-order) parameter is the spectral density matrix

$$\mathbf{f}(\lambda) = [f_{j,k}(\lambda)].$$

In this section we will be interested in the interrelationships between *pairs* of component series $X_j(t)$ and $X_k(t)$. These relationships are necessarily determined by the three *bivariate parameters* $f_{j,j}(\lambda)$, $f_{k,k}(\lambda)$, and $f_{j,k}(\lambda)$ for each $j \neq k$. We will take j and k to be fixed but arbitrary indices in this section. These parameters provide the information available in the correlation between $X_j(t)$ and $X_k(t)$ in the most useable form. Two sets of real-valued parameters will be defined which are equivalent to the three bivariate spectral parameters but which display the correlational information in different ways.

The two sets of parameters depend on the two standard representations of a complex number in terms of real numbers. The cartesian form leads to the representation

$$f_{j,k}(\lambda) = c_{j,k}(\lambda) - iq_{j,k}(\lambda), \tag{5.42}$$

where the real-valued functions $c_{j,k}(\lambda)$ and $q_{j,k}(\lambda)$ are called the *cospectral density* (*cospectrum*) and *quadrature spectral density* (*quadspectrum*), respectively. As we will see, the minus sign in (5.42) is a consequence of the

choice of the sign in the exponent of the spectral representation

$$C_{j,k}(\tau) = \int e^{i\lambda\tau} f_{j,k}(\lambda)\, d\lambda.$$

In all cases of practical interest, this expression can be inverted to yield

$$f_{j,k}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda\tau} C_{j,k}(\tau)\, d\tau \tag{5.43}$$

in the continuous-time case. Now, we expand the exponential to obtain $e^{-i\lambda\tau} = \cos\lambda\tau - i\sin\lambda\tau$ and write

$$C_{j,k}(\tau) = \frac{C_{j,k}(\tau) + C_{j,k}(-\tau)}{2} + \frac{C_{j,k}(\tau) - C_{j,k}(-\tau)}{2}.$$

The first term on the right-hand side of this expression is an even function and the second an odd function. Substituting these expressions into (5.43) we obtain

$$f_{j,k}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos\lambda\tau\left\{\frac{C_{j,k}(\tau) + C_{j,k}(-\tau)}{2}\right\} d\tau$$
$$- i\frac{1}{2\pi}\int_{-\infty}^{\infty} \sin\lambda\tau\left\{\frac{C_{j,k}(\tau) - C_{j,k}(-\tau)}{2}\right\} d\tau.$$

The products yielding odd functions integrate to zero in this result since the integrals are over symmetric intervals. Thus, we obtain the *continuous-time inversion formulas*

$$c_{j,k}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos\lambda\tau\left\{\frac{C_{j,k}(\tau) + C_{j,k}(-\tau)}{2}\right\} d\tau \tag{5.44}$$

and

$$q_{j,k}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \sin\lambda\tau\left\{\frac{C_{j,k}(\tau) - C_{j,k}(-\tau)}{2}\right\} d\tau. \tag{5.45}$$

Note that this computation and the definition of $q_{j,k}(\lambda)$ given by (5.45) account for the minus sign in (5.41).

The *discrete-time versions* of these formulas are

$$c_{j,k}(\lambda) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} \cos\lambda\tau\left\{\frac{C_{j,k}(\tau) + C_{j,k}(-\tau)}{2}\right\}, \tag{5.46}$$

$$q_{j,k}(\lambda) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} \sin\lambda\tau\left\{\frac{C_{j,k}(\tau) - C_{j,k}(-\tau)}{2}\right\}. \tag{5.47}$$

These expressions play an important role in the computation of cross-spectral density estimators for one of the methods of spectral estimation to be discussed in Chapter 8. Note that the real-valued parameters $c_{j,k}(\lambda)$, $q_{j,k}(\lambda)$, $f_{j,j}(\lambda)$, $f_{k,k}(\lambda)$ are equivalent to the original set of bivariate spectral parameters.

The polar representation of $f_{j,k}(\lambda)$ yields another set of spectral parameters. Write

$$f_{j,k}(\lambda) = |f_{j,k}(\lambda)| e^{i\vartheta_{j,k}(\lambda)}, \tag{5.48}$$

and let

$$\rho_{j,k}(\lambda) = |f_{j,k}(\lambda)| / \left(f_{j,j}(\lambda) f_{k,k}(\lambda)\right)^{1/2}. \tag{5.49}$$

The functions $\vartheta_{j,k}(\lambda)$ and $\rho_{j,k}(\lambda)$ are called the *phase* [*phase lead of* $X_j(t)$ *over* $X_k(t)$] and *coherence* (*coefficient of coherence*), respectively. The function

$$\gamma_{j,k}(\lambda) = f_{j,k}(\lambda) / \left(f_{j,j}(\lambda) f_{k,k}(\lambda)\right)^{1/2}$$

is called the *complex coherence*. It will play a role in the computation of higher order coherences in the next section. By property (5.19), the coherence satisfies the inequality

$$0 \leq \rho_{j,k}(\lambda) \leq 1 \tag{5.50}$$

wherever the ratio (5.49) is well defined. At frequencies for which $f_{j,j}(\lambda) f_{k,k}(\lambda) = 0$, we will define $\rho_{j,k}(\lambda) = 0$.

It is easily seen that $\rho_{j,k}(\lambda)$, $\vartheta_{j,k}(\lambda)$, $f_{j,j}(\lambda)$, and $f_{k,k}(\lambda)$ constitute a set of spectral parameters equivalent to the original set. These parameters are obtainable from the spectral densities, cospectrum, and quadrature spectrum by means of the relations

$$\rho_{j,k}(\lambda) = \left(\frac{c_{j,k}^2(\lambda) + q_{j,k}^2(\lambda)}{f_{j,j}(\lambda) f_{k,k}(\lambda)} \right)^{1/2} \tag{5.51}$$

and

$$\vartheta_{j,k}(\lambda) = -\operatorname{Arctan}\left(q_{j,k}(\lambda)/c_{j,k}(\lambda)\right). \tag{5.52}$$

[The minus sign again results from the fact that $q_{j,k}(\lambda) = -\operatorname{Im} f_{j,k}(\lambda)$.] By expanding (5.48) in cartesian form, the cospectrum and quadspectrum can be obtained from the spectral densities, phase and coherence by means of the expressions

$$c_{j,k}(\lambda) = \rho_{j,k}(\lambda) \cos \vartheta_{j,k}(\lambda) \left(f_{j,j}(\lambda) f_{k,k}(\lambda)\right)^{1/2} \tag{5.53}$$

$$q_{j,k}(\lambda) = -\rho_{j,k}(\lambda) \sin \vartheta_{j,k}(\lambda) \left(f_{j,j}(\lambda) f_{k,k}(\lambda)\right)^{1/2}. \tag{5.54}$$

The phase and coherence are, perhaps, the most useful parameters for measuring the relationship between $X_j(t)$ and $X_k(t)$ because the values of these functions can be interpreted quantitatively. We consider this next.

Interpretation of Phase and Coherence

Suppose that the random spectral measures of $X_j(t)$ and $X_k(t)$ could be represented in the form

$$Z_j(d\lambda) = e^{i\varphi_j(\lambda)}|Z_j(d\lambda)|, \qquad Z_k(d\lambda) = e^{i\varphi_k(\lambda)}|Z_k(d\lambda)|,$$

where $\varphi_j(\lambda)$ and $\varphi_k(\lambda)$ are nonrandom phase functions. That is, the phase functions are the same for all sample functions of the process. We would then have

$$f_{j,k}(\lambda)\,d\lambda = EZ_j(d\lambda)\overline{Z_k(d\lambda)} = e^{i(\varphi_j(\lambda)-\varphi_k(\lambda))}E|Z_j(d\lambda)Z_k(d\lambda)|,$$

from which it would follow that

$$\vartheta_{j,k}(\lambda) = \varphi_j(\lambda) - \varphi_k(\lambda).$$

That is, in this special case, the phase function is simply the difference in phase of the two time series at frequency λ or, more precisely, the phase lead of $X_j(t)$ over $X_k(t)$ at λ. In the usual situation wherein $\varphi_j(\lambda)$ and $\varphi_k(\lambda)$ are random quantities, this interpretation can be largely retained if now $\vartheta_{j,k}(\lambda)$ is thought of as an (ensemble) average of the random differences $\varphi_j(\lambda) - \varphi_k(\lambda)$. *Thus, $\vartheta_{j,k}(\lambda)$ is commonly interpreted as being the average phase lead of $X_j(t)$ over $X_k(t)$ at frequency λ.*

Recall from Section 4.2 that the phase shift of a linear filter at frequency λ can be given in time units by dividing by λ. In the same sense, the parameter

$$t_{j,k}(\lambda) = \vartheta_{j,k}(\lambda)/\lambda$$

is the time lead of the harmonic of $X_j(t)$ at frequency λ over that of $X_k(t)$.

An important property of the phase parameter is the following: If the univariate processes $X_j(t)$ and $X_k(t)$ are independently passed through linear filters with transfer functions $B_j(\lambda)$ and $B_k(\lambda)$ and if $Y_j(t)$ and $Y_k(t)$ are the outputs, then

$$f^Y_{j,k}(\lambda) = B_j(\lambda)\overline{B_k(\lambda)}f^X_{j,k}(\lambda)$$

by (5.34). Thus, if $\vartheta_j(\lambda)$ and $\vartheta_k(\lambda)$ are the phase shifts of the two filters, the phase of the output is related to that of the input by the expression

$$\vartheta^Y_{j,k}(\lambda) = \vartheta^X_{j,k}(\lambda) + \vartheta_j(\lambda) - \vartheta_k(\lambda).$$

Consequently, if it is important that the phase relationships remain undisturbed when the time series are processed by linear filters, filters with the same phase shift must be used on both series. A class of symmetric filters, called *nonnegative definite filters*, are important from this viewpoint in that they have zero phase shift for all frequencies. This is nearly true of all properly designed symmetric filters as well. See the discussion of this topic in Section A6.2.

The interpretation of coherence is somewhat more involved, but correspondingly more important than that of phase. As might be guessed, the term "coherence" is borrowed from the study of light. The term was first applied to time series by Wiener (1930). Because of the symbolic representation

$$\rho_{j,k}(\lambda) = |EZ_j(d\lambda)\overline{Z_k(d\lambda)}|/(E|Z_j(d\lambda)|^2 E|Z_k(d\lambda)|^2)^{1/2}$$

[see (5.20) and (5.49)], it is reasonable to expect that the coefficient of coherence will have the properties of the absolute value of a correlation coefficient at each frequency λ. In particular, the extreme values, zero and one, should correspond to complete lack of correlation and the maximum degree of correlation possible for a definition of correlation that makes sense in the time series context. Moreover, values of coherence other than zero and one should have some reasonable quantitative interpretation. We investigate the basis for such an interpretation in what follows.

The correlation coefficient measures the degree of *linear* association between two random variables. That is, it represents the degree to which one random variable can be represented as a linear function of the other. The term "association" rather than "dependence" is used because a large correlation need not indicate a causal relationship between the random variables. Similarly, the term "regression" is used to describe a *directed* relationship of one random variable with another. That is, whereas we speak of the linear association between X_1 and X_2, treating X_1 and X_2 on equal footing, when we speak of the linear regression of X_1 on X_2 we think of X_1 as the dependent and X_2 the independent variable in some sort of functional relationship. This distinction marks the difference between correlation theory and regression theory. The form of regression considered here differs from the theory more commonly taught in statistics wherein the independent variable is assumed to be nonrandom. The terms and interpretations of statistical correlation and regression theory will carry over to the time series context with little change. We do not distinguish between the two topics in this section in order to concentrate on the interpretations of the parameters. In the multivariate theory of the next section, these topics will be treated separately.

It is reasonable to expect that the coefficient of coherence will, in some sense, measure the degree of *linear* association between the time series $X_j(t)$ and $X_k(t)$. The only concept of linear function we have available for time series is that of a linear filter. Consequently, we can ask the question: To what degree can $X_k(t)$ be represented as the output of a linear filter with input $X_j(t)$? If mean-squared error is taken as the measure of the difference between $X_k(t)$ and filtered versions of $X_j(t)$, this question can be rephrased to read: How small can we make

$$\sigma_L^2 = E(X_k(t) - L(X_j(t)))^2$$

by proper choice of the linear filter L? [Note that this expression does not depend upon t, since $X_k(t) - L(X_j(t))$ is weakly stationary.] If the linear filter $\tilde{L}$ which minimizes ${\sigma_L}^2$ can be found, then the degree of linear association between $X_j(t)$ and $X_k(t)$ can be assessed by comparing ${\sigma_{\tilde{L}}}^2$, the power in the residual time series $X_k(t) - \tilde{L}(X_j(t))$, with the power $E(X_k(t))^2$ of $X_k(t)$ itself. The coefficient of coherence actually makes this comparison possible frequency by frequency.

To see this, it is first necessary to construct the linear filter $\tilde{L}$. Since filters are uniquely determined by their transfer functions, it suffices to compute the transfer function $\tilde{B}_{j,k}(\lambda)$ of $\tilde{L}$. This can easily be done by virtue of the following observation. Let L be any linear filter which matches $X_j(t)$ and let $B(\lambda)$ be its transfer function. Then by the spectral representations of $X_k(t)$ and $L(X_j(t))$,

$$\begin{aligned}{\sigma_L}^2 &= E\left\{\int e^{i\lambda t}[Z_k(d\lambda) - B(\lambda)Z_j(d\lambda)]\right\}\overline{\left\{\int e^{i\mu t}[Z_k(d\mu) - B(\mu)Z_j(d\mu)]\right\}} \\ &= \int E|Z_k(d\lambda) - B(\lambda)Z_j(d\lambda)|^2.\end{aligned}$$

Now, the minimum value of ${\sigma_L}^2$ *will be achieved for the linear filter whose transfer function minimizes the integrand of this expression for all frequencies* λ. That is, the problem of minimizing ${\sigma_L}^2$ is reduced to an infinite number of identical minimization problems, each involving the determination of a complex number $\tilde{B}_{j,k}(\lambda)$ which minimizes the quantity

$$E|Z_k(d\lambda) - B(\lambda)Z_j(d\lambda)|^2$$

among all possible complex numbers $B(\lambda)$. However, viewing $Z_k(d\lambda)$ and $Z_j(d\lambda)$ as random variables in $L_2(P)$ as before, this quantity is the squared distance between $Z_k(d\lambda)$ and an element in the linear subspace generated by $Z_j(d\lambda)$, since all such elements are simply complex multiples of $Z_j(d\lambda)$. The minimum distance is achieved by the orthogonal projection of $Z_k(d\lambda)$ onto this subspace. By the criterion for determining orthogonal projections given in Section 1.3, this means that the projection $\tilde{B}_{j,k}(\lambda)Z_j(d\lambda)$ has the property

$$(Z_k(d\lambda) - \tilde{B}_{j,k}(\lambda)Z_j(d\lambda)) \perp Z_j(d\lambda),$$

since $Z_j(d\lambda)$ is the only generator of the subspace. That is,

$$E(Z_k(d\lambda) - \tilde{B}_{j,k}(\lambda)Z_j(d\lambda))\overline{Z_j(d\lambda)} = 0.$$

From this it follows that

$$\tilde{B}_{j,k}(\lambda) = EZ_k(d\lambda)\overline{Z_j(d\lambda)}/E|Z_j(\lambda)|^2.$$

Because of (5.20) and the restriction to continuous spectra, this becomes

$$\tilde{B}_{j,k}(\lambda) = f_{k,j}(\lambda)/f_{j,j}(\lambda). \tag{5.55}$$

This is the transfer function of the filter $\tilde{L}$ which minimizes $\sigma_L{}^2$. Note that $\tilde{L}$ matches $X_j(t)$, since the condition $\rho_{j,k}(\lambda) \leq 1$ implies

$$\int |\tilde{B}_{j,k}(\lambda)|^2 f_{j,j}(\lambda)\, d\lambda = \int \left(|f_{k,j}(\lambda)|^2 / f_{j,j}(\lambda)\right) d\lambda$$
$$= \int \rho_{j,k}^2(\lambda) f_{k,k}(\lambda)\, d\lambda < \infty.$$

The gain and phase shift functions of $\tilde{L}$ are

$$|\tilde{B}_{j,k}(\lambda)| = |f_{k,j}(\lambda)|/f_{j,j}(\lambda) = \rho_{j,k}(\lambda)\left(f_{k,k}(\lambda)/f_{j,j}(\lambda)\right)^{1/2} \tag{5.56}$$

and

$$\tilde{\vartheta}_{j,k}(\lambda) = \arg\left(f_{k,j}(\lambda)\right) = -\vartheta_{j,k}(\lambda). \tag{5.57}$$

Thus, another interpretation of the phase angle $\vartheta_{j,k}(\lambda)$ is that it is the negative of the phase shift of the filter which transforms $X_j(t)$ into the best linear approximation to $X_k(t)$. In this sense, $X_j(t)$ leads $X_k(t)$ at frequency λ by the angle $\vartheta_{j,k}(\lambda)$.

In actual fact, this is the angle by which $X_j(t)$ leads $\tilde{X}_k(t) = \tilde{L}\left(X_j(t)\right)$ at frequency λ and the degree to which it is a reasonable measure of the angular separation of $X_j(t)$ and $X_k(t)$ will depend on how close $\tilde{X}_k(t)$ and $X_k(t)$ are at this frequency. To determine this, let

$$U_k(t) = X_k(t) - \tilde{X}_k(t).$$

This is a weakly stationary time series with spectral measure

$$Z_k{}^U(d\lambda) = Z_k{}^X(d\lambda) - \tilde{B}_{j,k}(\lambda) Z_j{}^X(d\lambda).$$

Applying the Pythagorean theorem to the orthogonal elements $Z_k{}^U(d\lambda)$ and $\tilde{B}_{j,k}(\lambda) Z_j{}^X(d\lambda) = Z_k{}^{\tilde{X}}(d\lambda)$, we obtain

$$E|Z_k{}^X(d\lambda)|^2 = E|Z_k{}^U(d\lambda) + \tilde{B}_{j,k}(\lambda) Z_j{}^X(d\lambda)|^2$$
$$= E|Z_k{}^U(d\lambda)|^2 + |\tilde{B}_{j,k}(\lambda)|^2 E|Z_j{}^X(d\lambda)|^2.$$

Consequently, the spectral densities of the processes $X_k(t)$, $\tilde{X}_k(t)$, and $U_k(t)$ are related by the expression

$$f_{k,k}^X(\lambda) = f_{k,k}^U(\lambda) + f_{k,k}^{\tilde{X}}(\lambda), \tag{5.58}$$

where

$$f_{k,k}^{\tilde{X}}(\lambda) = |\tilde{B}_{j,k}(\lambda)|^2 f_{j,j}^X(\lambda). \tag{5.59}$$

Now, by the definition of $\sigma_{\tilde{L}}{}^2$, the minimum value of the error is

$$\sigma_{\tilde{L}}{}^2 = \int f_{k,k}^U(\lambda)\, d\lambda.$$

Therefore, a frequency by frequency assessment of the magnitude of this error relative to the power in $X_k(t)$ can be obtained by means of the ratio of spectral densities $f^U_{k,k}(\lambda)/f^X_{k,k}(\lambda)$. With only minor abuse of language, this ratio can be described as the proportion of the power in $X_k(t)$ at frequency λ which *cannot* be explained by the linear regression of $X_k(t)$ on $X_j(t)$ [i.e., by $\tilde{L}(X_j(t))$]. Because of (5.58), the proportion of the power at frequency λ which *can* be so explained is $f^{\tilde{X}}_{k,k}(\lambda)/f^X_{k,k}(\lambda)$. However, from (5.56) and (5.59) we easily obtain

$$f^{\tilde{X}}_{k,k}(\lambda) = \rho^2_{j,k}(\lambda) f^X_{k,k}(\lambda).$$

Thus, this proportion is $\rho^2_{j,k}(\lambda)$. Moreover, repeating the above computations with j and k interchanged, the symmetry of the coefficient of coherence in j and k implies that $\rho^2_{j,k}(\lambda)$ is also the proportion of the power in $X_j(t)$ at frequency λ which can be explained by the linear regression of $X_j(t)$ on $X_k(t)$. *Thus, the squared coefficient of coherence $\rho^2_{j,k}(\lambda)$ can be interpreted as the proportion of the power at frequency λ in either time series $X_j(t)$, $X_k(t)$ which can be explained by its linear regression on the other.*

This means that linear filters can be constructed by which $100\,\rho^2_{j,k}(\lambda)\%$ of the power at frequency λ can be removed from either series by subtracting off the appropriate linearly filtered version of the other. This is the most useful interpretation of coherence.

In addition, by (5.58) the proportion of the power in the residual series at frequency λ will be $1 - \rho^2_{j,k}(\lambda)$. This yields the useful expression

$$f^U_{k,k}(\lambda) = \big(1 - \rho^2_{j,k}(\lambda)\big) f^X_{k,k}(\lambda) \tag{5.60}$$

for the residual spectral density.

Note that if $\rho_{j,k}(\lambda) = 0$, the two series are completely unrelated at frequency λ in the sense that no linearly filtered version of one can be used to reduce the power in the other at that frequency. On the other hand, if $\rho_{j,k}(\lambda) = 1$, one series is exactly a linearly filtered version of the other at frequency λ. In fact, if $\rho_{j,k}(\lambda) = 1$ for all λ, both time series can be represented as the output of a linear filter with the other as input. For then, $f^U_{k,k}(\lambda) \equiv 0$ which implies that $\sigma_L{}^2 = E\big(X_k(t) - \tilde{L}(X_j(t))\big)^2 = 0$. It follows that

$$X_k(t) = \tilde{L}\big(X_j(t)\big), \qquad -\infty < t < \infty.$$

The same argument applies with j and k interchanged. More generally, the coherence will vary with frequency indicating a changing pattern of linear association. Regions of high coherence usually have special significance in applications. We will now look at some examples in which the phase and coherence play a special role.

Example 5.3 *An Experiment in Optics*

An experiment traditionally performed in courses in optics is the demonstration of the nature of light coherence. This phenomenon, which makes the modern laser possible, has an analog in the mathematical model we have developed and, in fact, the mathematical coherence parameter is a useful quantitative description of light coherence. For simplicity we will consider two time series $X_1(t)$, $X_2(t)$ which represent the amplitudes of corresponding components of the electromagnetic vectors for two monochromatic light beams. The two beams emanate from a common source and are constrained to travel different paths to a photometer as pictured in Fig. 5.2. The principle of

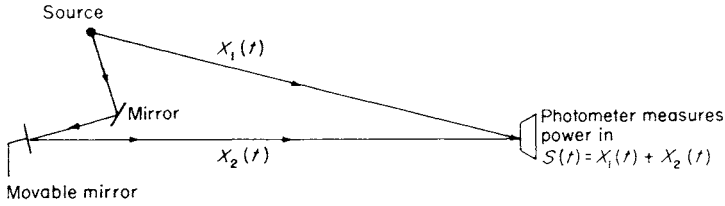

Fig. 5.2 *Schematic model of optical experiment to demonstrate coherence.*

this simple mechanism is identical to that of the Michelson interferometer. The actual operation of the Michelson interferometer is more complex, however [see, e.g., Jenkins and White (1950, p. 239)].

At the point where the two light beams are picked up by the photometer, the amplitudes are added and the power in the resulting sum $S(t) = X_1(t) + X_2(t)$ is measured. If λ denotes the frequency (velocity $\times$ wave number) of the light rays, then the photometer (effectively) records the spectral density $f_S(\lambda)$ of $S(t)$. In terms of our model, the random spectral measure of $S(t)$ is the sum of the spectral measures of $X_1(t)$ and $X_2(t)$;

$$Z_S(d\lambda) = Z_1(d\lambda) + Z_2(d\lambda).$$

Thus, by the properties of expectation,

$$\begin{aligned} f_S(\lambda)\,d\lambda &= EZ_S(d\lambda)\overline{Z_S(d\lambda)} \\ &= EZ_1(d\lambda)\overline{Z_1(d\lambda)} + EZ_1(d\lambda)\overline{Z_2(d\lambda)} + EZ_2(d\lambda)\overline{Z_1(d\lambda)} + EZ_2(d\lambda)\overline{Z_2(d\lambda)} \\ &= [f_{1,1}(\lambda) + f_{1,2}(\lambda) + f_{2,1}(\lambda) + f_{2,2}(\lambda)]\,d\lambda. \end{aligned}$$

However, since $f_{2,1}(\lambda) = \overline{f_{1,2}(\lambda)}$ and $\bar{u} + u = 2 \operatorname{Re} u$, we obtain

$$f_S(\lambda) = f_{1,1}(\lambda) + f_{2,2}(\lambda) + 2c_{1,2}(\lambda).$$

If we assume that the two light beams are adjusted to have the same power, it will follow that $f_{1,1}(\lambda) = f_{2,2}(\lambda)$. Then because of relation (5.53), we have

$$f_S(\lambda) = 2f_{1,1}(\lambda)[1 + \rho_{1,2}(\lambda) \cos \vartheta_{1,2}(\lambda)].$$

Now, if the source consists of two separate light producers, for example, two different light bulbs, then the two light rays will be completely incoherent (in theory) which implies that $\rho_{1,2}(\lambda) = 0$. The light power recorded at the photometer would then be

$$f_S(\lambda) = 2f_{1,1}(\lambda).$$

That is, the power would simply be twice that of each component ray. On the other hand, if the two light rays were perfectly coherent, a property which could be obtained by splitting a single ray into two beams, say, we would have $\rho_{1,2}(\lambda) = 1$ and

$$f_S(\lambda) = 2f_{1,1}(\lambda)[1 + \cos \vartheta_{1,2}(\lambda)].$$

Now, by varying the length of the path of $X_2(t)$ by operating the moveable mirror, the phase angle $\vartheta_{1,2}(\lambda)$, which measures the lead of $X_1(t)$ over $X_2(t)$, will vary. When the two amplitudes are perfectly in phase, which occurs when $\vartheta_{1,2}(\lambda) = \pi k$ for an even integer k, we will have

$$f_S(\lambda) = 4f_{1,1}(\lambda).$$

Thus, the power will be twice that for incoherent light. On the other hand, if the amplitudes are 180° out of phase [$\vartheta_{1,2}(\lambda) = \pi k$ for an odd integer k], the amplitudes will "cancel" and we will observe

$$f_S(\lambda) = 0.$$

Thus, as the mirror is moved, the observed light power will vary from 0 to $4f_{1,1}(\lambda)$.

In actual experiments it is difficult to achieve perfectly coherent light rays, so, in fact, we will have $0 < \rho_{1,2}(\lambda) < 1$ and the light power will vary over the less extreme range $2f_{1,1}(\lambda)(1 - \rho_{1,2}(\lambda))$ to $2f_{1,1}(\lambda)(1 + \rho_{1,2}(\lambda))$ as the path length of $X_2(t)$ is changed.

Example 5.4 *Computing the Transfer Function of an In-Service Linear Filter. Degree of Linearity, Signal-to-Noise Ratio*

The standard laboratory method for determining the transfer function $B(\lambda)$ of a linear filter L parallels our procedure for calculating transfer functions. For several values of λ, sinusoids $e^{i\lambda t}$ are fed into the filter and the amplitudes and phases of the outputs $B(\lambda)e^{i\lambda t}$ are measured. When the filter cannot be taken out of service or if the filter is a complicated physical system, such as the earth between two observation points, it is necessary to be able to determine the transfer function by observing the normal input $X_1(t)$ and

output $X_2(t)$ of the filter while it is in operation. This can be done if the input is sufficiently "broad-band"; that is if $f_{1,1}(\lambda) > 0$ over a wide range of frequencies λ. To see this, note that

$$Z_2(d\lambda) = B(\lambda)Z_1(d\lambda).$$

Thus,

$$\begin{aligned} f_{2,1}(\lambda)\, d\lambda &= EZ_2(d\lambda)\overline{Z_1(d\lambda)} \\ &= B(\lambda)E|Z_1(d\lambda)|^2 = B(\lambda)f_{1,1}(\lambda)\, d\lambda. \end{aligned}$$

It follows that

$$B(\lambda) = f_{2,1}(\lambda)/f_{1,1}(\lambda) \qquad \text{wherever} \quad f_{1,1}(\lambda) > 0.$$

Consequently, if the spectra are estimated from observed lengths of the input and output series, the gain and phase function of L can be estimated by means of the expressions

$$\begin{aligned} |B(\lambda)| &= |f_{1,2}(\lambda)|/f_{1,1}(\lambda) = \left(c_{1,2}^2(\lambda) + q_{1,2}^2(\lambda)\right)^{1/2}/f_{1,1}(\lambda), \\ \vartheta(\lambda) &= -\vartheta_{1,2}(\lambda) = \operatorname{Arctan}\left(q_{1,2}(\lambda)/c_{1,2}(\lambda)\right), \end{aligned}$$

where $f_{1,1}(\lambda)$, $c_{1,2}(\lambda)$, and $q_{1,2}(\lambda)$ are replaced by the corresponding estimates (see Chapter 8). For an example of the determination of the transfer function of a linear system with servo control by cross-spectral methods; see Goodman and Katz (1958).

By the theory of coherence given above, if $X_2(t) = L\left(X_1(t)\right)$, where L is a linear filter, then $\rho_{1,2}(\lambda) = 1$ for all λ. It follows that deviations of $\rho_{1,2}(\lambda)$ from unity indicate the presence of nonlinearity in the system. (We use the term *nonlinear* to mean that it is not true that $X_2(t) = L\left(X_1(t)\right)$ for a linear filter L. This is a mild abuse of standard terminology.) There are an infinite variety of types of nonlinearity and coherence does little to indicate which type is operating in the system. However, if the type of nonlinearity is known, its strength can often be related to the value of the coefficient of coherence.

An important nonlinear model is that of a linear system L contaminated additively by random noise. This model can be represented by the expression

$$X_2(t) = Y(t) + N(t),$$

where $Y(t) = L\left(X_1(t)\right)$ and the input $X_1(t)$, and noise $N(t)$ are uncorrelated time series. A schematic representation of this system is given in Fig. 5.3.

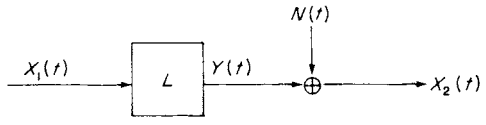

Fig. 5.3 *Schematic representation of a linear system contaminated by additive noise.*

We will assume that both $X_1(t)$ and $N(t)$ are weakly stationary. An important measure of the output signal fidelity is the signal-to-noise density ratio

$$\alpha(\lambda) = f_Y(\lambda)/f_N(\lambda),$$

which provides a frequency-by-frequency comparison of the power in the output signal relative to that of the noise. We will now show that the coefficient of coherence between the "observable" time series $X_1(t)$ and $X_2(t)$ is related to $\alpha(\lambda)$ in a simple way.

The spectral measure of the process $X_2(t)$ can be expressed as

$$Z_2(d\lambda) = B(\lambda)Z_1(d\lambda) + Z_N(d\lambda),$$

where $Z_1(d\lambda)$ and $Z_N(d\lambda)$ are the spectral measures of $X_1(t)$ and $N(t)$. Thus,

$$EZ_2(d\lambda)\overline{Z_1(d\lambda)} = B(\lambda)E|Z_1(d\lambda)|^2 + EZ_N(d\lambda)\overline{Z_1(d\lambda)}$$

and

$$E|Z_2(d\lambda)|^2 = |B(\lambda)|^2E|Z_1(d\lambda)|^2 + B(\lambda)EZ_1(d\lambda)\overline{Z_N(d\lambda)} + \overline{B(\lambda)}EZ_N(d\lambda)\overline{Z_1(d\lambda)} + E|Z_N(d\lambda)|^2.$$

Since $X_1(t)$ and $N(t)$ are assumed to be uncorrelated,

$$EZ_N(d\lambda)\overline{Z_1(d\lambda)} = EZ_1(d\lambda)\overline{Z_N(d\lambda)} = 0.$$

Thus, the cross-spectral density of $X_2(t)$ and $X_1(t)$ and spectral density of $X_2(t)$ are

$$f_{2,1}(\lambda) = B(\lambda)f_{1,1}(\lambda) \tag{5.61}$$

and

$$f_{2,2}(\lambda) = |B(\lambda)|^2f_{1,1}(\lambda) + f_N(\lambda).$$

It follows that the squared coefficient of coherence of $X_1(t)$ and $X_2(t)$ is

$$\begin{aligned}\rho^2_{1,2}(\lambda) &= |f_{2,1}(\lambda)|^2/f_{2,2}(\lambda)f_{1,1}(\lambda) \\ &= |B(\lambda)|^2f_{1,1}(\lambda)/\big(|B(\lambda)|^2f_{1,1}(\lambda) + f_N(\lambda)\big) = \alpha(\lambda)/\big(1 + \alpha(\lambda)\big).\end{aligned}$$

We have used the relation $f_Y(\lambda) = |B(\lambda)|^2f_{1,1}(\lambda)$ and the definition of $\alpha(\lambda)$ to obtain the last expression. Thus,

$$\alpha(\lambda) = \rho^2_{1,2}(\lambda)/\big(1 - \rho^2_{1,2}(\lambda)\big).$$

Since $\rho^2_{1,2}(\lambda)$ can be estimated from samples of the processes $X_1(t)$ and $X_2(t)$ by techniques to be given in Chapter 8, the signal-to-noise ratio can be estimated. Note that by using (5.61), the characteristics of the linear filter can also be estimated in terms of the observable processes as before. For regions of the spectrum for which $\rho_{1,2}(\lambda)$ is close to one, thus $\alpha(\lambda)$ is large, the

output of the linear filter is the dominant component of $X_2(t)$. Thus, a signal with improved *overall* signal-to-noise ratio $\int f_Y(\lambda)\,d\lambda / \int f_N(\lambda)\,d\lambda$ can be achieved by band-pass filtering $X_2(t)$ to eliminate bands of the spectrum for which $\alpha(\lambda)$ is relatively small. An application of these ideas to a problem in seismology was given by Koopmans (1961). We next look at a classical time series problem which is easily solved by the methods developed above.

Example 5.5 *The Historical Filtering Problem*

Let $X(t)$ and $Y(t)$ be stationarily correlated univariate time series where $Y(t)$ is thought of as being an observable process which is a distorted version of the unobservable process $X(t)$. The problem is to construct the linear filter $\tilde{L}$ which, when applied to $Y(t)$, best reproduces $X(t)$. By "best" we will again mean that $\tilde{L}$ minimizes

$$E\big(X(t) - L(Y(t))\big)^2$$

among all linear filters L which match $Y(t)$. Since $\tilde{L}$ will be allowed to operate on $Y(t)$ for all times t, $-\infty < t < \infty$, in theory, a complete sample function must be available before the process $\tilde{X}(t) = \tilde{L}\big(Y(t)\big)$ can be constructed. Consequently, this problem is called the *historical filtering problem* in contrast to the more difficult, *real-time filtering problem* in which the filter is restricted to operate on the past and present of the process. We will consider this problem in Chapter 7.

By the argument given earlier in this section, if $C(\lambda)$ is the transfer function of L, then

$$E\big(X(t) - L(Y(t))\big)^2 = \int E\,|Z_X(d\lambda) - C(\lambda)Z_Y(d\lambda)|^2,$$

and $\tilde{C}(\lambda)$, the transfer function of $\tilde{L}$, minimizes

$$E\,|Z_X(d\lambda) - C(\lambda)Z_Y(d\lambda)|^2$$

for each λ. The derivation leading to (5.55) produces

$$\tilde{C}(\lambda) = f_{X,Y}(\lambda)/f_Y(\lambda)$$

in this case, where $f_{X,Y}(\lambda)$ and $f_Y(\lambda)$ are the cross-spectral densities of $X(t)$ and $Y(t)$ and the spectral density of $Y(t)$, respectively. Moreover, if $\eta(t) = X(t) - \tilde{X}(t)$, then the degree to which the optimal filter fails to reproduce $X(t)$ is

$$E\big(\eta(t)\big)^2 = \int f_\eta(\lambda)\,d\lambda,$$

where $f_\eta(\lambda)$ is the spectral density of $\eta(t)$. However, this is the analog of the residual spectral density (5.60), thus can be expressed in the form

$$f_\eta(\lambda) = \big(1 - \rho^2_{X,Y}(\lambda)\big) f_X(\lambda),$$

where $\rho_{X,Y}(\lambda)$ is the coefficient of coherence between $X(t)$ and $Y(t)$ and $f_X(\lambda)$ is the spectral density of $X(t)$. Thus, as would be expected, the reproducibility of $X(t)$ by a linear filtered version of $Y(t)$ depends on the degree of linear association between the two series as measured by the coherence.

If

$$Y(t) = X(t) + N(t), \tag{5.62}$$

where $N(t)$ is a noise process with spectral density $f_N(\lambda)$, uncorrelated with $X(t)$, then it is easy to check that

$$f_{X,Y}(\lambda) = f_X(\lambda) \qquad \text{and} \qquad f_Y(\lambda) = f_X(\lambda) + f_N(\lambda).$$

Thus,

$$\tilde{C}(\lambda) = f_X(\lambda)/\big(f_X(\lambda) + f_N(\lambda)\big).$$

Note that this transfer function is real-valued. Hence, the filter $\tilde{L}$ is symmetric. Thus it is clear in this special case that this filter cannot be used in real-time filtering. When $Y(t)$ has form (5.62), alternate forms for the transfer function are easily seen to be

$$\tilde{C}(\lambda) = \alpha(\lambda)/\big(1 + \alpha(\lambda)\big),$$

where $\alpha(\lambda)$ is the signal-to-noise density ratio $\alpha(\lambda) = f_X(\lambda)/f_N(\lambda)$ and

$$\tilde{C}(\lambda) = \rho^2_{X,Y}(\lambda).$$

Both of these expressions make sense, intuitively, since where the signal-to-noise ratio is high, and thus $X(t)$ and $Y(t)$ are nearly equal, the filter $\tilde{L}$ passes $Y(t)$ virtually unchanged. However, where the signal-to-noise ratio is low, which means that $N(t)$ is the dominant component of $Y(t)$, the filter suppresses the noise by passing almost none of the power of $Y(t)$.

Since $X(t)$ and $N(t)$ are generally not observable separately, it is not always possible to estimate the transfer function $\tilde{C}(\lambda)$ from actual data. An exception is provided by the model of the last example in which the signal $X(t)$ is the output of a linear system with observable input $W(t)$. Then, even if the transfer characteristics of the linear system are unknown, the linear filter which, when applied to $Y(t)$, best reproduces $X(t)$ has the transfer function

$$\tilde{C}(\lambda) = \rho^2_{W,Y}(\lambda).$$

This is an immediate consequence of a very useful property of coherence (to be considered next) which implies that $\rho^2_{X,Y}(\lambda) = \rho^2_{W,Y}(\lambda)$. Thus, $\tilde{C}(\lambda)$ can be estimated from observations on $W(t)$ and $Y(t)$ and the impulse response function of the corresponding convolution filter can be obtained by Fourier transformation. Consequently, the best noise suppression filter can be constructed to a good degree of approximation in this case.

An Invariance Property of Coherence

We saw in Section 1.4 that the correlation coefficient is a generalization of the cosine of the angle between vectors to the "angle" between random variables. Since the coefficient of coherence $\rho^X_{j,k}(\lambda)$ is essentially the correlation coefficient of the random spectral measures $Z_j{}^X(d\lambda)$ and $Z_k{}^X(d\lambda)$ of two time series, it will also have the properties of the cosine. Cosines are invariant under scale changes of the component vectors, which corresponds to multiplying $Z_j{}^X(d\lambda)$ and $Z_k{}^X(d\lambda)$ by complex quantities $B_j(\lambda)$ and $B_k(\lambda)$, respectively. However, this is equivalent to the modification of the spectral measures which results from the application of independent linear filters to $X_j(t)$ and $X_k(t)$. *That is, if* $Y_j(t) = L_j\big(X_j(t)\big)$ *and* $Y_k(t) = L_k\big(X_k(t)\big)$, *where* L_j *and* L_k *are any linear filters matching the corresponding inputs, then*

$$\rho^Y_{j,k}(\lambda) = \rho^X_{j,k}(\lambda).$$

This equation is valid for all λ *for which the transfer functions of both filters are nonzero.* This property of coherence, which we will call *invariance under linear filtering*, can be easily established by appealing directly to the definition of coherence and to expression (5.34). It has important practical consequences: Often, it is necessary to compute the degree of association between two time series which have been passed through a series of linear filters with possibly unknown characteristics. Because of this invariance property, the coefficient of coherence of the outputs is the same as that of the inputs. The next example illustrates the usefulness of this property.

Example 5.6 *The Detection of a Coherent Source of Ocean Waves*

Munk *et al.* (1959) considered the possibility of detecting and locating sources of distant underwater disturbances or storm centers by computing the spectral parameters of low-frequency ocean wave amplitudes at a number of stations. For convenience, we will deal with only two stations and will let $X_1(t)$ and $X_2(t)$ denote the recorded amplitudes at these stations. A simple but reasonable model of the mechanisms generating these time series is the following. At the source of the disturbance the energy impressed on the sea surface can be represented by a time series $Y(t)$. The resulting waves travel different paths to the two recording stations and if the sea is assumed to transmit energy linearly, the contributions to the variations in sea height due to the disturbance at the two stations will be $U_1(t) = L_1\big(Y(t)\big)$ and $U_2(t) = L_2\big(Y(t)\big)$, respectively. Because both filters have the same input, the invariance of coherence under linear filtering implies that

$$\rho^U_{1,2}(\lambda) \equiv 1.$$

Now it is reasonable to assume that

$$X_j(t) = U_j(t) + N_j(t), \qquad j = 1, 2,$$

where $N_j(t)$ is the contribution to sea height due to causes other than the distant source modeled by $Y(t)$. Moreover, by the definition of the $N_j(t)$'s it is reasonable to assume that $U_j(t)$ and $N_k(t)$ are uncorrelated for $j, k = 1, 2$. Then, any correlation between $X_1(t)$ and $X_2(t)$ can be attributed either to the correlation between $U_1(t)$ and $U_2(t)$ or to that between $N_1(t)$ and $N_2(t)$. The simplest situation is that of rather widely separated recording stations for which $N_1(t)$ and $N_2(t)$ are locally generated, thus, uncorrelated. This precludes the existence of a second distant coherent source, for example. Then the presence of the original source would make itself known by a coefficient of coherence $\rho^X_{1,2}(\lambda)$ near unity over regions of the spectrum for which the power in $U_1(t)$ and $U_2(t)$ dominates that of $N_1(t)$ and $N_2(t)$. Since the power in locally generated sea waves will generally occupy a higher frequency range than that of waves which have traveled a very long distance, the computed coefficient of coherence provides an effective indicator for detecting distant coherent sources.

Under the given simplifying assumptions, the coefficient of coherence between $X_1(t)$ and $X_2(t)$ can be shown to be

$$\rho^X_{1,2}(\lambda) = \left(\alpha_1(\lambda)\alpha_2(\lambda)/(1 + \alpha_1(\lambda))(1 + \alpha_2(\lambda))\right)^{1/2},$$

where $\alpha_j(\lambda)$ is the signal-to-noise density ratio for $U_j(t)$ and $N_j(t)$,

$$\alpha_j(\lambda) = f^U_{j,j}(\lambda)/f^N_{j,j}(\lambda), \qquad j = 1, 2.$$

Consequently, it is possible to interpret the values of the coefficient of coherence directly in terms of the relative signal and noise strengths.

Example 5.7 *An Economics Model and Its Spectral Parameters*

As another illustration of the use of the method for calculating spectral parameters given in the last section we will calculate the spectra, coherence, and phase for discrete time series $X(t)$ and $Y(t)$, which represent the deviations from their mean values of the price and available quantity of a given commodity, respectively. These series are assumed to be related by the equations

$$Y(t) = \alpha X(t-1) + U(t), \qquad X(t) = -\beta Y(t) + V(t), \tag{5.63}$$

where $U(t)$ and $V(t)$ are taken to be zero-mean white noise processes, uncorrelated with one another, with variances $\sigma_U{}^2$, $\sigma_V{}^2$. To obtain a reasonably realistic model, the "loading parameters" α and β are both assumed to be positive. A further restriction will be indicated presently.

The equations for the spectral measures of the various processes can be read off directly from (5.63),

$$Z_Y(d\lambda) = \alpha e^{-i\lambda} Z_X(d\lambda) + Z_U(d\lambda), \qquad Z_X(d\lambda) = -\beta Z_Y(d\lambda) + Z_V(d\lambda).$$

Here, we have used the fact that $e^{-i\lambda}$ is the transfer function of the linear filter which takes $X(t)$ into $X(t-1)$. This is immediate from the usual rule for calculating transfer functions. Rewriting these equations as a pair of linear equations in the unknowns $Z_X(d\lambda)$ and $Z_Y(d\lambda)$, it is easy to obtain the solutions

$$Z_X(d\lambda) = \frac{-\beta Z_U(d\lambda) + Z_V(d\lambda)}{1 + \alpha\beta e^{-i\lambda}}, \quad Z_Y(d\lambda) = \frac{\alpha e^{-i\lambda} Z_V(d\lambda) + Z_U(d\lambda)}{1 + \alpha\beta e^{-i\lambda}}.$$

Thus, the spectral density of $X(t)$ is

$$f_X(\lambda)\, d\lambda = E|Z_X(d\lambda)|^2 = \frac{E[-\beta Z_U(d\lambda) + Z_V(d\lambda)]\overline{[-\beta Z_U(d\lambda) + Z_V(d\lambda)]}}{|1 + \alpha\beta e^{-i\lambda}|^2}$$

$$= \frac{1}{2\pi} \frac{\beta^2 {\sigma_U}^2 + {\sigma_V}^2}{|1 + \alpha\beta e^{-i\lambda}|^2}\, d\lambda.$$

This yields,

$$f_X(\lambda) = \frac{1}{2\pi} \frac{\beta^2 {\sigma_U}^2 + {\sigma_V}^2}{1 + \alpha^2\beta^2 + 2\alpha\beta \cos\lambda}.$$

Similarly,

$$f_Y(\lambda) = \frac{1}{2\pi} \frac{\alpha^2 {\sigma_V}^2 + {\sigma_U}^2}{1 + \alpha^2\beta^2 + 2\alpha\beta \cos\lambda}$$

and

$$f_{X,Y}(\lambda) = \frac{1}{2\pi} \frac{-\beta {\sigma_U}^2 + \alpha e^{i\lambda} {\sigma_V}^2}{|1 + \alpha\beta e^{-i\lambda}|^2}$$

$$= \frac{1}{2\pi} \frac{-\beta {\sigma_U}^2 + \alpha {\sigma_V}^2 \cos\lambda + i\alpha {\sigma_V}^2 \sin\lambda}{1 + \alpha^2\beta^2 + 2\alpha\beta \cos\lambda}.$$

In these calculations we have used the relation

$$EZ_U(d\lambda)\overline{Z_V(d\lambda)} = 0$$

and the expressions

$$f_U(\lambda) = {\sigma_U}^2/2\pi, \qquad f_V(\lambda) = {\sigma_V}^2/2\pi, \qquad -\pi < \lambda \leq \pi,$$

for the spectral densities of the white noise processes.

Now to guarantee that these spectral densities do not "blow up" for some value of λ, that is, the denominators do not go to zero, it is necessary to take $\alpha\beta \neq 1$. As we will show in our discussion of finite parameter models in Chapter 7, it is actually necessary to take $|\alpha\beta| < 1$ in order for $X(t)$ and $Y(t)$ to be realizable processes, certainly a desirable attribute in a realistic model.

Consequently, combining this with the earlier restriction on the parameters, we assume that $0 < \alpha\beta < 1$.

Finally, the squared coefficient of coherence of $X(t)$ and $Y(t)$ is

$$\rho^2_{X,Y}(\lambda) = \frac{|f_{X,Y}(\lambda)|^2}{f_X(\lambda)f_Y(\lambda)} = \frac{(-\beta\sigma_U{}^2 + \alpha\sigma_V{}^2 \cos\lambda)^2 + (\alpha\sigma_V{}^2 \sin\lambda)^2}{(\beta^2\sigma_U{}^2 + \sigma_V{}^2)(\alpha^2\sigma_V{}^2 + \sigma_U{}^2)}$$

$$= \frac{\beta^2\sigma_U{}^4 + \alpha^2\sigma_V{}^4 - 2\alpha\beta\sigma_U{}^2\sigma_V{}^2 \cos\lambda}{\beta^2\sigma_U{}^4 + \alpha^2\sigma_V{}^4 + (1 + \alpha^2\beta^2)\sigma_U{}^2\sigma_V{}^2},$$

and the phase angle is

$$\vartheta_{X,Y}(\lambda) = \operatorname{Arctan} \frac{\operatorname{Im} f_{X,Y}(\lambda)}{\operatorname{Re} f_{X,Y}(\lambda)} = \operatorname{Arctan}\left(\frac{\alpha\sigma_V{}^2 \sin\lambda}{-\beta\sigma_U{}^2 + \alpha\sigma_V{}^2 \cos\lambda}\right).$$

This model predicts the highest coherence between price and quantity in the high frequencies where the power in both series is greatest. This would lead to highly oscillatory price and supply series, which are characteristic of a rather unstable economy. For values of λ beyond a point determined by α, β, $\sigma_U{}^2$, and $\sigma_V{}^2$, $\vartheta_{X,Y}(\lambda)$ would be negative but close to zero, indicating that price follows the rise and fall of supply. Thus, this simple model displays a number of features one might expect of a real economic system. To determine the adequacy of the model for a real system, it would be possible to compare the predicted spectral parameters computed above with estimates of the coherence, phase, and spectral densities based on actual observations of the two processes. The estimation of these parameters will be taken up in Chapter 8.

5.6 THE MULTIVARIATE SPECTRAL PARAMETERS, THEIR INTERPRETATIONS AND USES

There is a close relationship between the spectral parameters of time series analysis and the correlational parameters of multivariate statistics due to the virtually identical representations of inner products and linear transformations for the vectors of spectral measures $Z_1(d\lambda), \ldots, Z_p(d\lambda)$ of a multivariate time series on one hand and vectors of zero-mean random variables with finite variances on the other. Thus, as we saw, the coefficient of coherence is essentially the modulus of a correlation coefficient with the properties and interpretation one would expect of this parameter. The remaining multivariate spectral parameters are defined to retain this parallelism and, in fact, the relationship is so close that most of the spectral theory can be taken, with only minor modifications, from texts on multivariate analysis such as the one by Anderson (1958). See also Cramér (1951a). We will cover the essential features

of the theory without going into great depth. Additional details are covered by Hannan (1970) and Koopmans (1964b). Due to an important difference in emphasis, we will treat the correlation theory and regression theory separately.

Multivariate Correlation Analysis

If $\mathbf{X}(t)$ is a multivariate weakly stationary process with $p > 2$ components, then it is often important to account for the interaction among several of the components when the power in one of them is to be determined or when the association between two of the components is to be assessed. It may be the case, for example, that most of the power in a given series can be removed by subtracting off a function of various other of the components. This would indicate a relationship among the components which might arise, for example, as the result of a common "driving mechanism."

Again the type of association to be considered is *linear association*. Thus, we will account for the influence of components $X_{m_1}(t), X_{m_2}(t), \ldots, X_{m_q}(t)$ on $X_j(t)$ by constructing the linear function of these series which best approximates $X_j(t)$. That is, if

$$Y_j(t) = \mathbf{L}\big(X_{m_1}(t), \ldots, X_{m_q}(t)\big) \tag{5.64}$$

denotes a multivariate linear filter with q inputs and one output, as described in Section 5.4, then the filter which best approximates $X_j(t)$ is the one which minimizes

$$E\big(X_j(t) - Y_j(t)\big)^2. \tag{5.65}$$

Let $\tilde{X}_{j\cdot\mathbf{m}}(t)$ denote the output of the minimizing filter, where $\mathbf{m} = (m_1, m_2, \ldots, m_q)'$. Then, to measure the strength of the linear regression of $X_j(t)$ on $X_{m_1}(t), \ldots, X_{m_q}(t)$ we can compare the power of $\tilde{X}_{j\cdot\mathbf{m}}(t)$ with the power of $X_j(t)$. This comparison can be made frequency by frequency by using the coefficient of coherence of the two time series. This parameter, called the *multiple coherence*, is defined in terms of the spectral densities and cross-spectral densities of $X_j(t)$ and $\tilde{X}_{j\cdot\mathbf{m}}(t)$. Consequently, it is important to be able to calculate these densities.

The problem of describing the unconditional linear relationship between two component time series $X_j(t)$ and $X_k(t)$ was solved in the last section by the introduction of the coefficient of coherence. When the vector time series has other components, it is likely that part of the coherence between $X_j(t)$ and $X_k(t)$ can be attributed to the linear relationship each has with some or all of the others. To determine the coherence between $X_j(t)$ and $X_k(t)$ which is *not* attributable to the regression of these processes on $X_{m_1}(t), \ldots, X_{m_q}(t)$, say, it is reasonable to adjust each of the two processes by subtracting off its best "explanation" as a linear function of the $X_{m_j}(t)$'s. Thus, the *partial*

coherence is the coefficient of coherence of the residual processes $X_j(t) - \tilde{X}_{j\cdot\mathbf{m}}(t)$ and $X_k(t) - \tilde{X}_{k\cdot\mathbf{m}}(t)$. Again, it is necessary to compute the spectral densities from which this parameter is calculated. To do this, we will use the spectral representation of $\mathbf{X}(t)$ and that of a multidimensional linear filter to obtain the random spectral measure of $\tilde{X}_{j\cdot\mathbf{m}}(t)$.

In spectral form, Eq. (5.64) can be written

$$Z_j^Y(d\lambda) = \sum_{r=1}^{q} B_{j,r}(\lambda) Z_{m_r}^X(d\lambda), \tag{5.66}$$

where the $1 \times q$ vector $\mathbf{B}_j(\lambda) = [B_{j,r}(\lambda)]$ is the transfer function of $\mathbf{L}$. The argument of the last section can again be applied to show that

$$E(X_j(t) - Y_j(t))^2 = \int E\left| Z_j^X(d\lambda) - \sum_{r=1}^{q} B_{j,r}(\lambda) Z_{m_r}^X(d\lambda) \right|^2.$$

Thus, the minimum of this expression occurs at the transfer function $\tilde{\mathbf{B}}_j(\lambda)$ which determines the projection of $Z_j^X(d\lambda)$ on the linear subspace generated by $Z_{m_1}^X(d\lambda), \ldots, Z_{m_q}^X(d\lambda)$ for each λ. By the criterion for determining projections given in Section 1.3, the elements of the transfer function must satisfy the conditions

$$E\left[Z_j^X(d\lambda) - \sum_{r=1}^{q} \tilde{B}_{j,r}(\lambda) Z_{m_r}^X(d\lambda) \right] \overline{[Z_{m_s}^X(d\lambda)]} = 0, \qquad s = 1, 2, \ldots, q.$$

In terms of the spectral density functions, these equations are

$$\sum_{r=1}^{q} \tilde{B}_{j,r}(\lambda) f_{m_r, m_s}^X(\lambda) = f_{j,m_s}^X(\lambda), \qquad s = 1, 2, \ldots, q \tag{5.67}$$

or, in vector form,

$$\tilde{\mathbf{B}}_j(\lambda) \mathbf{f}_{\mathbf{m}}^X(\lambda) = \mathbf{f}_{j,\mathbf{m}}^X(\lambda),$$

where $\mathbf{f}_{\mathbf{m}}^X(\lambda)$ and $\mathbf{f}_{j,\mathbf{m}}^X(\lambda)$ denote the $q \times q$ matrix and $1 \times q$ vector of spectral densities indicated in (5.67). Thus, when the inverse exists,

$$\tilde{\mathbf{B}}_j(\lambda) = \mathbf{f}_{j,\mathbf{m}}^X(\lambda) \mathbf{f}_{\mathbf{m}}^X(\lambda)^{-1}. \tag{5.68}$$

When the inverse does not exist, it is still possible to define the transfer function by a formula of this type with the inverse replaced by what is known as a *pseudo inverse* [see, e.g., Koopmans (1964b)]. We will not be concerned with this, since the case in which the inverse exists is by far the more important one in practice.

Now, by the definition of $\tilde{X}_{j\cdot\mathbf{m}}(t)$ and by (5.66), the spectral measure of this process is

$$Z_{j\cdot\mathbf{m}}^{\tilde{X}}(d\lambda) = \mathbf{f}_{j,\mathbf{m}}^X(\lambda) \mathbf{f}_{\mathbf{m}}^X(\lambda)^{-1} \mathbf{Z}_{\mathbf{m}}^X(d\lambda), \tag{5.69}$$

where

$$\mathbf{Z}_{\mathbf{m}}{}^{X}(d\lambda) = \big(Z^{X}_{m_1}(d\lambda), \ldots, Z^{X}_{m_q}(d\lambda)\big)'.$$

Moreover,

$$\begin{aligned} E\,|\,Z^{\hat{X}}_{j\,\cdot\,\mathbf{m}}(d\lambda)\,|^2 &= E\big(\mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\mathbf{Z}_{\mathbf{m}}{}^{X}(d\lambda)\big)\big(\mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\mathbf{Z}_{\mathbf{m}}{}^{X}(d\lambda)\big)^{*} \\ &= f^{X}_{j,\,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)^{*}\,d\lambda, \end{aligned}$$

since

$$E\mathbf{Z}_{\mathbf{m}}{}^{X}(d\lambda)\mathbf{Z}_{\mathbf{m}}{}^{X}(d\lambda)^{*} = \mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)\,d\lambda.$$

It follows, that the spectral density function of $\tilde{X}_{j\,\cdot\,\mathbf{m}}(t)$ is

$$f^{\tilde{X}}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda) = \mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\mathbf{f}_{j,\,\mathbf{m}}(\lambda)^{*}. \tag{5.70}$$

Thus, the multiple coherence of $X_j(t)$ on $X_{m_1}(t), \ldots, X_{m_q}(t)$, which is the proportion of the power (density) at frequency λ attributable to the linear regression of $X_j(t)$ on $X_{m_1}(t), \ldots, X_{m_q}(t)$, is given by

$$\begin{aligned} R^2_{j\,\cdot\,\mathbf{m}}(\lambda) &= f^{\tilde{X}}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda)/f^{X}_{j,\,j}(\lambda) \\ &= \mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\mathbf{f}^{X}_{j,\,\mathbf{m}}(\lambda)^{*}/f^{X}_{j,\,j}(\lambda). \end{aligned} \tag{5.71}$$

By the Pythagorean theorem, since $\big(Z_j{}^{X}(d\lambda) - Z^{\tilde{X}}_{j\,\cdot\,\mathbf{m}}(d\lambda)\big) \perp Z^{\tilde{X}}_{j\,\cdot\,\mathbf{m}}(d\lambda)$, we have

$$E\,|\,Z_j{}^{X}(d\lambda)\,|^2 = E\,|\,Z^{\tilde{X}}_{j\,\cdot\,\mathbf{m}}(d\lambda)\,|^2 + E\,|\,Z_j{}^{X}(d\lambda) - Z^{\tilde{X}}_{j\,\cdot\,\mathbf{m}}(d\lambda)\,|^2.$$

Thus,

$$f^{X}_{j,\,j}(\lambda) = f^{\tilde{X}}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda) + f^{U}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda), \tag{5.72}$$

where $f^{U}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda)$ is the spectral density of the residual process

$$U_{j\,\cdot\,\mathbf{m}}(t) = X_j(t) - \tilde{X}_{j\,\cdot\,\mathbf{m}}(t).$$

It follows from (5.71) and (5.72) that $0 \le R^2_{j\,\cdot\,\mathbf{m}}(\lambda) \le 1$ as one would expect, and, moreover, *the proportion of the power (density) in $X_j(t)$ at frequency λ not attributable to its linear regression on $X_{m_1}(t), \ldots, X_{m_q}(t)$ is*

$$f^{U}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda)/f^{X}_{j,\,j}(\lambda) = 1 - R^2_{j\,\cdot\,\mathbf{m}}(\lambda).$$

This also yields the useful expression

$$f^{U}_{j,\,j\,\cdot\,\mathbf{m}}(\lambda) = \big(1 - R^2_{j\,\cdot\,\mathbf{m}}(\lambda)\big)f^{X}_{j,\,j}(\lambda) \tag{5.73}$$

for the residual spectral density.

The *partial coherence* of $X_j(t)$ and $X_k(t)$ with the regression on $X_{m_1}(t), \ldots, X_{m_q}(t)$ removed is the coefficient of coherence of $U_{j\,\cdot\,\mathbf{m}}(t)$ and $U_{k\,\cdot\,\mathbf{m}}(t)$. This parameter can now be computed. Moreover, the computation can be simplified by the observation that $Z^{U}_{j\,\cdot\,\mathbf{m}}(d\lambda) \perp Z^{\tilde{X}}_{k\,\cdot\,\mathbf{m}}(d\lambda)$, since $Z^{\tilde{X}}_{k\,\cdot\,\mathbf{m}}(d\lambda)$

is in the linear subspace generated by $Z^X_{m_1}(d\lambda), \ldots, Z^X_{m_q}(d\lambda)$. We can use this fact to write

$$EZ^U_{j\cdot\mathbf{m}}(d\lambda)\overline{Z^U_{k\cdot\mathbf{m}}(d\lambda)} = E\big(Z_j{}^X(d\lambda) - Z^{\tilde{X}}_{j\cdot\mathbf{m}}(d\lambda)\big)\overline{\big(Z_k{}^X(d\lambda)\big)}$$
$$= \big(f^X_{j,k}(\lambda) - \mathbf{f}^X_{j,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^X(\lambda)^{-1}\mathbf{f}^X_{k,\mathbf{m}}(\lambda)^*\big)\,d\lambda.$$

Thus, we obtain

$$f^U_{j,k\cdot\mathbf{m}}(\lambda) = f^X_{j,k}(\lambda) - \mathbf{f}^X_{j,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^X(\lambda)^{-1}\mathbf{f}^X_{k,\mathbf{m}}(\lambda)^*. \tag{5.74}$$

The $(p-q)\times(p-q)$ matrix $\mathbf{f}_{\mathbf{m}}{}^U(\lambda)$ with elements given by (5.74) as j and k range over the indices $1, 2, \ldots, p$ excluding $m_1, m_2, \ldots, m_q$, is called the *residual spectral matrix*. The *complex partial coherence* is

$$\gamma_{j,k\cdot\mathbf{m}}(\lambda) = f^U_{j,k\cdot\mathbf{m}}(\lambda)/\big(f^U_{j,j\cdot\mathbf{m}}(\lambda)f^U_{k,k\cdot\mathbf{m}}(\lambda)\big)^{1/2}$$
$$= \big(1 - R^2_{j,k\cdot\mathbf{m}}(\lambda)\big)\gamma_{j,k}(\lambda)/\big((1 - R^2_{j\cdot\mathbf{m}}(\lambda))(1 - R^2_{k\cdot\mathbf{m}}(\lambda))\big)^{1/2}, \tag{5.75}$$

where we have let

$$R^2_{j,k\cdot\mathbf{m}}(\lambda) = f^{\tilde{X}}_{j,k\cdot\mathbf{m}}(\lambda)/f^X_{j,k}(\lambda) = \mathbf{f}^X_{j,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^X(\lambda)^{-1}\mathbf{f}^X_{k,\mathbf{m}}(\lambda)^*/f^X_{j,k}(\lambda), \tag{5.76}$$

and $\gamma_{j,k}(\lambda)$ is the complex coherence of $X_j(t)$ and $X_k(t)$. The partial coherence is then

$$\rho_{j,k\cdot\mathbf{m}}(\lambda) = |1 - R^2_{j,k\cdot\mathbf{m}}(\lambda)|\,\rho_{j,k}(\lambda)/\big((1 - R^2_{j\cdot\mathbf{m}}(\lambda))(1 - R^2_{k\cdot\mathbf{m}}(\lambda))\big)^{1/2}. \tag{5.77}$$

The phase angle between the residual processes $U_j(t)$ and $U_k(t)$ is

$$\vartheta_{j,k\cdot\mathbf{m}}(\lambda) = \psi_{j,k\cdot\mathbf{m}}(\lambda) + \vartheta_{j,k}(\lambda), \tag{5.78}$$

where

$$\psi_{j,k\cdot\mathbf{m}}(\lambda) = \arg\big(1 - R^2_{j,k\cdot\mathbf{m}}(\lambda)\big).$$

It can be seen from expressions (5.71), (5.76), and (5.77) that the multiple and partial coherence are elementary functions of the elements of the spectral density matrix $\mathbf{f}^X(\lambda)$. The standard procedure for obtaining statistical estimates of these parameters is to first obtain estimates of the elements of $\mathbf{f}^X(\lambda)$, then enter these estimates in the above formulas in place of the corresponding parameters. Since this must be done for a variety of values of λ and, often, for several selections of the indices j, k, and elements in $\mathbf{m}$, the number of computations required can be quite large. Fortunately, convenient digital computer programs exist for carrying out these calculations.

For example, matrices of estimated spectra and cross spectra can be calculated by means of the program BMDX92 (Dixon, 1969). This is used as input to the multiple time series spectral analysis program BMDX68 which computes the estimates of the residual spectral matrix, coherences, multiple coherences, and matrix transfer function (5.68), among other things.

Example 5.8 *Calculation of the Spectral Parameters in a Simple Case*

When $\mathbf{m}$ consists of a single index, the expressions for the spectral parameters are especially simple. Suppose $R^2_{1\cdot 3}(\lambda)$ and $\rho_{1,2\cdot 3}(\lambda)$ are to be calculated. Now,

$$\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda) = [f^X_{3,3}(\lambda)]$$

and it follows that

$$\begin{aligned} R^2_{1\cdot 3}(\lambda) &= f^X_{1,\mathbf{m}}(\lambda)\mathbf{f}_{\mathbf{m}}{}^{X}(\lambda)^{-1}\overline{f^X_{1,\mathbf{m}}(\lambda)}/f^X_{1,1}(\lambda) \\ &= |f^X_{1,3}(\lambda)|^2/f^X_{1,1}(\lambda)f^X_{3,3}(\lambda) = \rho^2_{1,3}(\lambda). \end{aligned}$$

The equality of the multiple and ordinary coherence agrees with the fact that the interpretations of the two parameters are the same in this case. From (5.76) and the definition of complex coherence,

$$\begin{aligned} &[1 - R^2_{1,2\cdot 3}(\lambda)]\gamma_{1,2}(\lambda) \\ &\quad = \left(f^X_{3,3}(\lambda)f^X_{1,2}(\lambda) - f^X_{1,3}(\lambda)f^X_{3,2}(\lambda)\right)/f^X_{3,3}(\lambda)\left(f^X_{1,1}(\lambda)f^X_{2,2}(\lambda)\right)^{1/2}. \end{aligned}$$

Thus, from the second expression on the right-hand side of (5.75), we obtain

$$\gamma_{1,2\cdot 3}(\lambda) = \left(\gamma_{1,2}(\lambda) - \gamma_{1,3}(\lambda)\overline{\gamma_{2,3}(\lambda)}\right)/\left((1 - \rho^2_{1,3}(\lambda)\right)\left(1 - \rho^2_{2,3}(\lambda)\right)^{1/2}.$$

This is the complex analog of a well-known expression for correlation coefficients. The partial coefficient of coherence is, then,

$$\rho_{1,2\cdot 3}(\lambda) = |\gamma_{1,2\cdot 3}(\lambda)|.$$

Example 5.9 *An Application of Partial Coherence to a Biomedical Problem*

Gersh and Goddard (1970) considered the problem of determining the location of an epileptic focus in the brain of a cat based on EEG records from electrodes implanted in six deep sites in the brain. The goal of the study was to determine whether any one of the six sites could be interpreted as "driving" the others, based on recordings made during an induced epileptic seizure. The data, then, consisted of a six-dimensional time series recorded over the period of the seizure—some 8 seconds in the given record.

If, for example, site 1 is "driving" the remaining sites 2–6, one might postulate a model of the type pictured in Fig. 5.4. The noise processes $N_j(t)$, $j = 2, \ldots, 6$ are taken to be uncorrelated with each other and with $X_1(t)$. This model could be "identified" over regions of the spectrum in which the noise power is relatively low by the fact that the pairwise coherences $\rho_{j,k}(\lambda)$ would all be reasonably large as would all partial coherences $\rho_{j,k\cdot l}(\lambda)$ regressed on single series other than $X_1(t)$, i.e., for $l \neq 1$. This is true, since the dominant variation in each pair of series would be due to $X_1(t)$. On the other hand, $\rho_{j,k\cdot 1}(\lambda)$ would be relatively small, since upon

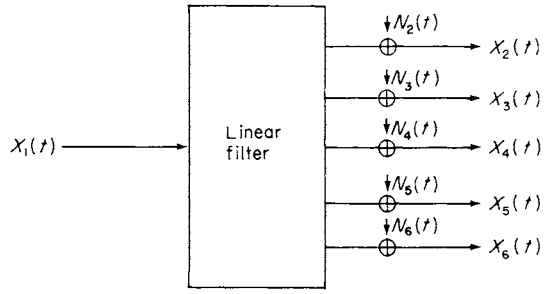

Fig. 5.4 *Schematic representation of a six-dimensional time series in which* $X_1(t)$ "*drives*" *the remaining components.*

removing the effects of $X_1(t)$, the comparison would be, principally, between the incoherent residual noises.

By computing the coherences for each pair of series and for all pairs regressed on each single series, Gersh and Goddard were able to identify one of the series as having the characteristics of the "driving" series in this model.

Multivariate Spectral Regression Analysis

In the regression context, the component processes of the multivariate time series are partitioned into two categories—the dependent processes and the independent processes. To emphasize this distinction, the dependent components will be assembled into a $(q \times 1)$-dimensional vector process $\mathbf{Y}(t)$ and the independent processes will constitute a $(p \times 1)$-vector process $\mathbf{X}(t)$. It is now assumed that

$$\mathbf{Y}(t) = \mathbf{L}(\mathbf{X}(t)) + \boldsymbol{\eta}(t), \tag{5.79}$$

where $\mathbf{L}$ is a multivariate linear filter with unknown $(q \times p)$-dimensional transfer function $\mathbf{B}(\lambda)$ and $\boldsymbol{\eta}(t)$ is an unobservable $(q \times 1)$-dimensional process uncorrelated with $\mathbf{X}(t)$. Thus $\mathbf{Y}(t)$ is assumed to arise from a fixed but unknown linear transformation of $\mathbf{X}(t)$ which is disturbed by an error process $\boldsymbol{\eta}(t)$. The extent of the deviation of $\mathbf{Y}(t)$ from a linear function of $\mathbf{X}(t)$ is measured by the unknown spectral density matrix $\mathbf{f}^{\eta}(\lambda)$. This matrix and the transfer function $\mathbf{B}(\lambda)$, which indicates how the linear dependence is parcelled out to the various input and output time series, are the principal parameters of interest.

Except for the more general form of the linear transformation, model (5.79) is identical to the distributed lag model of Example 5.2. Precisely the same argument can be used to calculate the transfer function of $\mathbf{L}$ and the spectral density function of $\boldsymbol{\eta}(t)$. Recall from Eqs. (5.40) and (5.41) that

$$\mathbf{B}(\lambda) = \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1} \tag{5.80}$$

and

$$\mathbf{f}^{\eta}(\lambda) = \mathbf{f}^{Y}(\lambda) - \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1}\mathbf{f}^{X,Y}(\lambda), \tag{5.81}$$

where

$$\mathbf{f}^{X}(\lambda)\,d\lambda = E\mathbf{Z}^{X}(d\lambda)\mathbf{Z}^{X}(d\lambda)^{*}, \qquad \mathbf{f}^{Y}(\lambda)\,d\lambda = E\mathbf{Z}^{Y}(d\lambda)\mathbf{Z}^{Y}(d\lambda)^{*},$$

and

$$\mathbf{f}^{X,Y}(\lambda)\,d\lambda = E\mathbf{Z}^{X}(d\lambda)\mathbf{Z}^{Y}(d\lambda)^{*}.$$

In the regression context, processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are observable. Consequently, estimates of $\mathbf{f}^{X}(\lambda)$, $\mathbf{f}^{Y}(\lambda)$, and $\mathbf{f}^{X,Y}(\lambda)$ can be obtained by the technique to be discussed in Chapter 8. By replacing the matrices of parameters in (5.80) and (5.81) by their estimators, estimates of $\mathbf{B}(\lambda)$ and $\mathbf{f}^{\eta}(\lambda)$ can be computed. We will discuss some of the properties of these estimators in Chapter 8.

A matrix parameter comparable to the coefficient of coherence can be defined which provides a more convenient measure of the extent of the linear regression of $\mathbf{Y}(t)$ on $\mathbf{X}(t)$ than does $\mathbf{f}^{\eta}(\lambda)$. If these processes were one-dimensional, the complex coherence would be

$$\gamma(\lambda) = f^{X,Y}(\lambda)/\big(f^{X}(\lambda)f^{Y}(\lambda)\big)^{1/2}.$$

The fact that $\mathbf{f}^{X}(\lambda)$ and $\mathbf{f}^{Y}(\lambda)$ are nonnegative definite matrices (see Section 5.3) makes it possible to define a comparable expression in the multidimensional case. We need a few facts about nonnegative definite (Hermetian) matrices which can be found, for example, in the book by Graybill (1969): The inverse of a nonnegative definite matrix, when it exists, is again nonnegative definite. The eigenvalues of such a matrix are all real and nonnegative. Thus, the inverse will fail to exist only when one or more of the eigenvalues are zero. Nonnegative definite matrices have well-defined square roots. That is, if $\mathbf{A}$ is a nonnegative definite matrix, there is a nonnegative definite matrix $\mathbf{C}$ such that $\mathbf{CC} = \mathbf{A}$.

Combining these results, we can define the *matrix complex coherence* by

$$\boldsymbol{\gamma}(\lambda) = \mathbf{f}^{X}(\lambda)^{-1/2}\mathbf{f}^{X,Y}(\lambda)\mathbf{f}^{Y}(\lambda)^{-1/2}, \tag{5.82}$$

where $\mathbf{A}^{-1/2}$ denotes the square root of the inverse of the matrix $\mathbf{A}$. Now the matrix parameter

$$\boldsymbol{\rho}^{2}(\lambda) = \boldsymbol{\gamma}(\lambda)^{*}\boldsymbol{\gamma}(\lambda) = \mathbf{f}^{Y}(\lambda)^{-1/2}\mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1}\mathbf{f}^{X,Y}(\lambda)\mathbf{f}^{Y}(\lambda)^{-1/2} \tag{5.83}$$

is comparable to the squared coherence. It can be shown to be a nonnegative definite matrix, consequently, all its eigenvalues are nonnegative. Moreover, from (5.81) we have

$$\mathbf{f}^{\eta}(\lambda) = \mathbf{f}^{Y}(\lambda)^{1/2}\big(\mathbf{I} - \boldsymbol{\rho}^{2}(\lambda)\big)\mathbf{f}^{Y}(\lambda)^{1/2}, \tag{5.84}$$

where $\mathbf{I}$ is the $q \times q$ identity matrix. When $\boldsymbol{\eta}(t)$ has no power in any component at frequency λ, i.e., $\mathbf{f}^{\eta}(\lambda) = \mathbf{0}$, we have

$$\boldsymbol{\rho}^2(\lambda) = \mathbf{I}$$

[provided $\mathbf{f}^{Y}(\lambda)$ is nonsingular]. In this case all eigenvalues of $\boldsymbol{\rho}^2(\lambda)$ are equal to 1. Now, it is easily shown from (5.84) that $\mathbf{I} - \boldsymbol{\rho}^2(\lambda)$ is also nonnegative definite. Since the eigenvalues of this matrix are one minus the eigenvalues of $\boldsymbol{\rho}^2(\lambda)$, it follows that the eigenvalues of $\boldsymbol{\rho}^2(\lambda)$ are all between zero and one. Thus the complete linear dependence of $\mathbf{Y}(t)$ on $\mathbf{X}(t)$ at frequency λ corresponds to the extreme case in which all eigenvalues are one. On the other hand, if all eigenvalues of $\boldsymbol{\rho}^2(\lambda)$ are zero, then $\boldsymbol{\rho}^2(\lambda)$ is the zero matrix and (5.84) implies that $\mathbf{f}^{\eta}(\lambda) = \mathbf{f}^{Y}(\lambda)$. In this case $\mathbf{Y}(t)$ is completely "explained" by the error term $\boldsymbol{\eta}(t)$ at frequency λ and no linear relationship between $\mathbf{Y}(t)$ and $\mathbf{X}(t)$ exists.

More generally, (5.84) is seen to be the matrix version of (5.60). The eigenvalues of $\boldsymbol{\rho}^2(\lambda)$ can be used to assess the relative degree of linear regression of $\mathbf{Y}(t)$ on $\mathbf{X}(t)$. Statistical estimates of these eigenvalues can be used to test various hypotheses about this regression. For example, the hypothesis of no linear regression at frequency λ can be tested by comparing the estimated largest eigenvalue or the sum of the eigenvalues of $\boldsymbol{\rho}^2(\lambda)$ with zero. The second possibility is equivalent to testing the trace of $\boldsymbol{\rho}^2(\lambda)$ to be zero. This hypothesis can also be framed as a hypothesis concerning $\mathbf{B}(\lambda)$, since it is easily seen that

$$\boldsymbol{\rho}^2(\lambda) = \mathbf{f}^{Y}(\lambda)^{-1/2}\mathbf{B}(\lambda)\mathbf{f}^{X}(\lambda)\mathbf{B}(\lambda)^{*}\mathbf{f}^{Y}(\lambda)^{-1/2}.$$

If $\mathbf{f}^{X}(\lambda)$ and $\mathbf{f}^{Y}(\lambda)$ are nonsingular, the hypothesis of no regression at frequency λ is equivalent to the hypothesis $\mathbf{B}(\lambda) = \mathbf{0}$.

When the $\mathbf{Y}(t)$ series is one-dimensional, i.e., $q = 1$, expression (5.83) yields

$$\boldsymbol{\rho}^2(\lambda) = \mathbf{f}^{Y,X}(\lambda)\mathbf{f}^{X}(\lambda)^{-1}\mathbf{f}^{Y,X}(\lambda)^{*}/f^{Y}(\lambda).$$

Comparing this with (5.71), it is seen that $\boldsymbol{\rho}^2(\lambda)$ is simply the multiple coherence of $Y(t)$ with $\mathbf{X}(t)$ as we would expect. Thus, the multiple coherence is an important regression parameter in this special case. When $p = 1$ it can be shown that

$$\boldsymbol{\rho}^2(\lambda) = \mathbf{f}^{X,Y}(\lambda)\mathbf{f}^{Y}(\lambda)^{-1}\mathbf{f}^{Y,X}(\lambda)/f^{X}(\lambda), \tag{5.85}$$

the multiple coherence of $X(t)$ with $\mathbf{Y}(t)$. This indicates that the matrix parameter $\boldsymbol{\rho}^2(\lambda)$ possesses a symmetry in the time series $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ roughly analogous to that displayed by the coefficient of coherence.

Finally, although we have derived the above regression theory model based on the assumption that all time series are weakly stationary processes, it is important to note that it is also possible to derive a statistical regression theory based on a model in which the $\mathbf{X}(t)$ time series is nonrandom. In this

theory, which has been developed by Brillinger (1970), the randomness is assumed to be due entirely to the error process $\boldsymbol{\eta}(t)$ appearing in (5.79). The resulting model is closer to the standard regression theory of statistics than is the model we have discussed. In particular, although the spectral density function of the error process $\mathbf{f}^{\eta}(\lambda)$ is well defined, none of the other spectral densities need be. However, expressions (5.80) and (5.81) are still used to construct estimates of $\mathbf{B}(\lambda)$ and $\mathbf{f}^{\eta}(\lambda)$ based on the same functions of the observed time series as would be used if the $\mathbf{X}(t)$ series were stochastic. Although the estimation procedures are essentially identical, the statistical distributions of the estimates are different [see Brillinger (1970)].

Some Spectral Regression Examples

The following three examples will give some idea of the scope of applicability of spectral regression methods. Complete descriptions of the studies and their conclusions are not feasible here. Further details can be found in the indicated references.

Example 5.10 *An Application to Metallurgy*

The earliest time series study in which spectral regression techniques were used is attributed to Tick (1955) in which the variability of the hot metal output of a blast furnace was evaluated as a function of (i.e., was regressed on) such variables as hot blast temperature, wind rate, and amounts of ore, coke, and limestone. These variables unavoidably vary with time and are correlated with one another to different degrees making the time series model a natural choice. The dependence of the variability (power) of hot metal output on the regression time series was evaluated by integrating the appropriate estimated residual spectral densities over frequency.

Example 5.11 *A Study of the Relationship between Sun Spots and Meteorological Data*

A contribution to the long-standing controversy concerning the influence of solar energy indicators, such as sun-spot numbers, on terrestrial time series, such as temperature and rainfall data, was made by Brillinger (1969). In this study the independent time series $\mathbf{X}(t)$ was taken to be the one-dimensional series of monthly relative sun-spot numbers. Three different sets of dependent series $\mathbf{Y}(t)$ were used; (i) Santa Fe, New Mexico rainfall ($p = 1$), (ii) English rainfall ($p = 1$), and (iii) temperatures at 14 European stations ($p = 14$). The degree of linear regression of each dependent series on sun-spot numbers was evaluated by estimating the generalized coherence in form (5.85). Although the estimates showed a considerable variability with frequency, they were smaller than 0.5 at all frequencies in all three cases. The

sun-spot coherence was larger with temperature at almost all frequencies than with rainfall which supports the conclusion of other investigators that little relationship between rainfall and sun-spot numbers exists. A comprehensive summary of evidence for and against the existence of relationships between solar energy parameters and various terrestrial series is given by Monin and Vulis (1971).

Example 5.12 *A Study of Sea Level Data*

Groves and Hannan (1968) studied records of sea level, surface atmospheric pressure, and wind velocity components at Kwajelein and Eniwetok in the Marshall Islands for the purpose of determining oceanic influences on sea level records free from local weather noise. To do this, the two sea level series were taken as the dependent series and were regressed on the six series of pressure and wind velocities. The residual series of sea levels, with weather effects accounted for, were then compared. It was found that the partial coherence of sea level records at the two islands with weather effects removed was smaller than the ordinary coherence of these records. This indicates that most of the coherence in sea level records is probably due to coherent weather patterns rather than to coherent patterns of water movement from non-weather-induced phenomena.

As with all studies of this type, the interpretation of results is subject to the statistical uncertainty of the estimates of the parameters. We will consider methods for evaluating and controlling this uncertainty in Chapters 8 and 9.

APPENDIX TO CHAPTER 5

A5.1 The Multidimensional Spectral Representation

The mathematical setting for the spectral representation of a multivariate weakly stationary process is somewhat more involved than that for the representation of one-dimensional processes, but many of the formulas carry over with little change. We will touch on only a few of the details here. More extensive analyses are given by Koopmans (1964a,b) and Wiener and Masani (1957, 1958).

A vector analog of the space $L_2(P)$ is required. Take the process $\mathbf{X}(t)$ to be p-dimensional. For each t the vector $\mathbf{X}(t)$ is an element of the product space $\mathbf{L}_2(P) = [L_2(P)]^p$ of column vectors $\mathbf{X} = (X_1, \ldots, X_p)'$ with $EX_j = 0$ and $E|X_j|^2 < \infty$. This vector space is endowed with the norm

$$\|\mathbf{X}\| = \left[\sum_{j=1}^{p} E|X_j|^2\right]^{1/2}.$$

Starting with finite linear combinations of the form

$$\mathbf{Y} = \sum_k \mathbf{A}_k \mathbf{X}(t_k), \tag{A5.1}$$

where the $\mathbf{A}_k$'s are $p \times p$ matrices of complex numbers, and forming all possible limits of Cauchy sequences of such elements in the vector norm, we obtain the *space generated by the process* $\mathscr{M}^{\mathbf{X}}$. It is easily seen that

$$\mathscr{M}^{\mathbf{X}} = [\mathscr{M}^{\mathbf{X}}]^p,$$

where $\mathscr{M}^{\mathbf{X}}$ is the linear subspace of $L_2(P)$ generated by the elements of the component processes of $\mathbf{X}(t)$. On the other hand, if $\mathbf{Z}(A)$ is the vector of spectral measures of the component processes, (A5.1) becomes

$$\begin{aligned}\mathbf{Y} &= \sum_k \mathbf{A}_k \int e^{i\lambda t_k} \mathbf{Z}(d\lambda) \\ &= \int \mathbf{B}(\lambda)\mathbf{Z}(d\lambda),\end{aligned}$$

where

$$\mathbf{B}(\lambda) = \sum_k e^{i\lambda t_k} \mathbf{A}_k .$$

In the limit, every element of $\mathscr{M}^{\mathbf{X}}$ can be represented in the form

$$\mathbf{Y} = \int \mathbf{B}(\lambda)\mathbf{Z}(d\lambda) \tag{A5.2}$$

for some $p \times p$ matrix of complex-valued functions $\mathbf{B}(\lambda)$.

Next we define a generalized inner product, the *Grammian matrix*, by

$$\langle\!\langle \mathbf{X}, \mathbf{Y} \rangle\!\rangle_P = E\mathbf{X}\mathbf{Y}^*.$$

Then, by (A5.2), if $\mathbf{X} = \int \mathbf{B}(\lambda)\mathbf{Z}(d\lambda)$ and $\mathbf{Y} = \int \mathbf{C}(\lambda)\mathbf{Z}(d\lambda)$, we have

$$\begin{aligned}\langle\!\langle \mathbf{X}, \mathbf{Y} \rangle\!\rangle_P &= E\left(\int \mathbf{B}(\lambda)\mathbf{Z}(d\lambda)\right)\left(\int \mathbf{C}(\lambda)\mathbf{Z}(d\lambda)\right)^* \\ &= \int \mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{C}(\lambda)^*.\end{aligned}$$

Note that $\|\mathbf{X}\|^2 = \operatorname{tr} \langle\!\langle \mathbf{X}, \mathbf{X} \rangle\!\rangle_P = \operatorname{tr} \int \mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{B}(\lambda)^*$.

Now, if $\mathbf{L}_2(\mathbf{F})$ is the class of all $p \times p$ matrix-valued functions $\mathbf{B}(\lambda)$ with complex entries such that

$$\operatorname{tr} \int \mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{B}(\lambda)^* < \infty,$$

and if the Grammian of $\mathbf{L}_2(\mathbf{F})$ is defined by

$$\langle\!\langle \mathbf{B}(\lambda), \mathbf{C}(\lambda) \rangle\!\rangle_{\mathbf{F}} = \int \mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{C}(\lambda)^*,$$

then the mapping $\mathbf{Y} \leftrightarrow \mathbf{B}(\lambda)$ defined by (A5.2) is a one-to-one correspondence between $\mathscr{M}^{\mathbf{X}}$ and $\mathbf{L}_2(\mathbf{F})$ which preserves the Grammian. This is the desired multivariate generalization of the spectral representation defined in the appendix to Chapter 2.

A5.2 Multivariate Linear Filters

To simplify the discussion of linear filters we will only consider filters with the same number of inputs and outputs. The generalized shift operator $\mathbf{U}_t$ is taken to be the $p \times p$ matrix operator with the univariate operator U_t, which was defined in Section A4.1, repeated down the main diagonal and 0's in the off-diagonal positions. Then a multivariate linear filter $\mathbf{L}$ is a linear operator on a domain $\mathscr{D}(\mathbf{L}) \subset \mathscr{M}^{\mathbf{X}}$ to $\mathscr{M}^{\mathbf{X}}$ which satisfies the condition

$$\mathbf{L}\mathbf{U}_t = \mathbf{U}_t\mathbf{L}, \qquad -\infty < t < \infty.$$

By the same argument as the one given in Section A4.1, $\mathbf{L}$ is completely determined by its value at $\mathbf{X}(0)$. Since $\mathbf{L}(\mathbf{X}(0)) \in \mathscr{M}^{\mathbf{X}}$ it follows that

$$\mathbf{L}(\mathbf{X}(0)) = \int \mathbf{B}(\lambda)\mathbf{Z}(d\lambda).$$

The matrix function $\mathbf{B}(\lambda)$ is, as before, the transfer function of the filter.

Now, $\mathbf{U} \in \mathscr{M}^{\mathbf{X}}$ is in $\mathscr{D}(\mathbf{L})$ provided

$$\operatorname{tr} \int \mathbf{C}(\lambda)\mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{B}(\lambda)^*\mathbf{C}(\lambda)^* < \infty,$$

where $\mathbf{U} = \int \mathbf{C}(\lambda)\mathbf{Z}(d\lambda)$. Then, as in the one-dimensional case, it can be shown that

$$\mathbf{L}(\mathbf{U}) = \int \mathbf{C}(\lambda)\mathbf{B}(\lambda)\mathbf{Z}(d\lambda).$$

In particular, since it is easily seen that $\mathbf{X}(t) \leftrightarrow e^{i\lambda t}\mathbf{I}$, where $\mathbf{I}$ is the $p \times p$ identity matrix, we have

$$\mathbf{L}(\mathbf{X}(t)) = \int e^{i\lambda t}\mathbf{B}(\lambda)\mathbf{Z}(d\lambda).$$

This is the spectral representation of the filter output. The *matching condition*

$$\operatorname{tr} \int \mathbf{B}(\lambda)\mathbf{F}(d\lambda)\mathbf{B}(\lambda)^* < \infty$$

is now seen to be simply the requirement that $\mathbf{X}(t) \in \mathscr{D}(\mathbf{L})$.

Again, there is a unique correspondence between the class of all linear filters on $\mathscr{M}^{\mathbf{X}}$ and $\mathbf{L}_2(\mathbf{F})$. Every element of this collection is the transfer function of some multidimensional linear filter. The various operations on filters detailed in Chapter 4 have natural extensions to the multivariate case but we will not pursue this topic further.