Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 4

Spectral Theory and Filtering

In Chapter 1 we discussed the correlation function as a way of characterizing a time series. In Chapter 2 we investigated representations of time series in terms of more elementary time series. These two descriptions are sometimes called descriptions in the *time domain* because of the obvious importance of the index set in the representations.

In Chapter 3 we introduced the Fourier transform of the correlation function. For certain correlation functions we demonstrated the uniqueness of the transform and showed that the correlation function is expressible as the inverse transform of the Fourier transform. The Fourier transform of the absolutely summable correlation function was called the spectral density. The spectral density furnishes another important representation of the time series. Because of the periodic nature of the trigonometric functions, the Fourier transform is often called the representation in the *frequency domain.*

4.1. THE SPECTRUM

In Chapter 1 we stated the result that the correlation function of a time series is analogous to a statistical characteristic function and may be expressed in the form

$$\rho(h) = \int_{-\pi}^{\pi} e^{ih\omega}\, dG(\omega)\,, \tag{4.1.1}$$

where the integral is a Lebesgue–Stieltjes integral and $G(\omega)$ is a statistical distribution function. In Theorem 3.1.9 we proved this result for time series with absolutely summable covariance functions.

Since the covariance function of a stationary time series is the correlation function multiplied by the variance of the process, we have

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\omega}\, dF(\omega)\,, \tag{4.1.1a}$$

where

$$dF(\omega) = \gamma(0)\, dG(\omega)\,.$$

Both of the functions $G(\omega)$ and $F(\omega)$ have been called the *spectral distribution function* in time series analysis. The spectral distribution function is a nondecreasing function that, for our purposes, can be assumed to be composed of the sum of two parts: an absolutely continuous portion and a step function.[1] We take (4.1.1a) as the definitional relationship between the spectral distribution function and the covariance function. The spectral distribution function is also sometimes called the *integrated spectrum.*

Let us assume $\gamma(h)$ is absolutely summable. Then, by Theorem 3.1.9, $f(\omega)$ defined by

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \omega h \end{aligned} \tag{4.1.2}$$

is a continuous nonnegative even function, and

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega h}\, d\omega\,.$$

Thus for time series with absolutely summable covariance function, $dF(\omega) = f(\omega)\, d\omega$, where $f(\omega)$ was introduced as the spectral density function in Section 3.3.

Recall that we have taken $\{e_t\}$ to be a time series of uncorrelated $(0, \sigma^2)$ random variables. The spectral density of $\{e_t\}$ is

$$\begin{aligned} f_e(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_e(h) \cos \omega h \\ &= \frac{1}{2\pi} \sigma^2\,, \end{aligned}$$

[1] Any statistical distribution function can be decomposed into three components: (1) a step function containing at most a countable number of finite jumps; (2) an absolutely continuous function; and (3) a "continuous singular" function. The third portion will be ignored in our treatment. See Tucker (1967, p. 15 ff.). While not formally correct, one may think of $F(\omega)$ as the sum of two parts, a step function with jumps at the points ω_j, $j = -M, -(M-1), \ldots, M-1, M$, and a function with continuous first derivative. Then the Lebesgue–Stieltjes integral $\int g(\omega)\, dF(\omega)$ is the sum of $\sum_{j=-M}^{M} g(\omega_j) l(\omega_j)$ and $\int g(\omega) f(\omega)\, d\omega$, where $l(\omega_j)$ is the height of the jump in $F(\omega)$ at the point ω_j, $f(\omega)$ is the derivative of the continuous portion of $F(\omega)$, and $\int g(\omega) f(\omega)\, d\omega$ is a Riemann integral.

which is positive, continuous, and trivially periodic. Similarly,

$$\begin{aligned}\gamma_e(h) &= \int_{-\pi}^{\pi} \frac{1}{2\pi}\sigma^2 e^{i\omega h}\, d\omega \\ &= \int_{-\pi}^{\pi} \frac{1}{2\pi}\sigma^2 \cos \omega h\, d\omega \\ &= \begin{cases}\sigma^2, & h = 0, \\ 0 & \text{otherwise}.\end{cases}\end{aligned}$$

As noted previously, the time series e_t is often called white noise. The reason for this description is now more apparent. The spectrum is a constant multiple of the variance and, in analogy to white light, one might say that all frequencies contribute equally to the variance.

If the spectral distribution function contains finite jumps, we can visualize the spectrum containing discrete "spikes" at the jump points in the same way that we view discrete probability distributions. We now investigate a time series whose spectral distribution function is a step function.

Let a time series be composed of a finite sum of simple processes of the form (1.6.3). That is, define the time series Y_t by

$$Y_t = \sum_{j=0}^{M} (A_j \cos \omega_j t + B_j \sin \omega_j t), \tag{4.1.3}$$

where the A_j and B_j are random variables with zero mean and

$$\begin{aligned} E\{A_j^2\} = E\{B_j^2\} &= \sigma_j^2, && j = 0, 1, 2, \ldots, M, \\ E\{B_j B_k\} = E\{A_j A_k\} &= 0, && j \neq k, \\ E\{A_j B_k\} &= 0 && \forall j, k, \end{aligned}$$

and ω_j, $j = 0, 1, \ldots, M$, are distinct frequencies contained in the interval $[-\pi, \pi]$. By (1.6.4), we have

$$\begin{aligned}\gamma_Y(h) = E\{Y_t Y_{t+h}\} &= \sum_{j=0}^{M} \sigma_j^2[\cos \omega_j t \cos \omega_j(t+h)] \\ &\quad + \sum_{j=0}^{M} \sigma_j^2[\sin \omega_j t \sin \omega_j(t+h)] \\ &= \sum_{j=0}^{M} \sigma_j^2 \cos \omega_j h\,. \end{aligned} \tag{4.1.4}$$

Since the function $\gamma_Y(h)$ is composed of a finite sum of cosine functions, the graph of σ_j^2 against ω_j (or j) will give us a picture of the relative contribution of the variance associated with frequency ω_j to the variance of the time series. This is

true because the variance of Y_t is given by

$$\gamma_Y(0) = \sum_{j=0}^{M} \sigma_j^2 .$$

While we permitted our original frequencies in (4.1.3) to lie anywhere in the interval $[-\pi, \pi]$, it is clear that with no loss of generality we could have restricted the frequencies to $[0, \pi]$. That is, the covariance function for

$$X_t = A_j \cos(-\omega_j)t + B_j \sin(-\omega_j)t$$

is the same as that of

$$X_t = A_j \cos \omega_j t + B_j \sin \omega_j t .$$

This suggests that for a covariance function of the form $\sigma_j^2 \cos \omega_j h$ we associate one half of the variance with the frequency $-\omega_j$ and one half with the frequency ω_j. To this end, we set

$$l(\omega_j) = l(-\omega_j) = \begin{cases} \frac{1}{2}\sigma_j^2, & \omega_j \neq 0, \\ \sigma_j^2, & \omega_j = 0. \end{cases}$$

To facilitate our representation, we assume that $\omega_0 = 0$ and then write the sum (4.1.4) as

$$\gamma_Y(h) = \sum_{j=-M}^{M} l(\omega_j) \cos \omega_j h , \tag{4.1.5}$$

where $\omega_{-j} = -\omega_j$.

We say that a time series with a covariance function of the form (4.1.5) has a *discrete spectrum* or a *line spectrum*. Equation (4.1.5) may be written as the Lebesgue–Stieltjes integral

$$\begin{aligned} \gamma_Y(h) &= \int_{-\pi}^{\pi} \cos \omega h \, dF(\omega) \\ &= \int_{-\pi}^{\pi} e^{i\omega h} \, dF(\omega) , \end{aligned} \tag{4.1.5a}$$

where $F(\omega)$ is a step function with jumps of height $\frac{1}{2}\sigma_j^2$ at the points ω_j and $-\omega_j$, $\omega_j \neq 0$, and a jump of height σ_0^2 at $\omega_0 = 0$. By construction, the jumps $l(\omega_j)$ are symmetric about zero, and we have expressed $\gamma_Y(h)$ in the general form (4.1.1a). We used this method of construction because the covariance function $\gamma_Y(h) = \sum_{j=0}^{M} \sigma_j^2 \cos \omega_j h$ is not absolutely summable and we could not directly apply (4.1.2).

For our purposes it is sufficient for us to be able to recognize the two types of autocovariance functions and associated spectra: (1) the autocovariance function

that is the sum of a finite number of cosine functions and is associated with a spectral distribution function, which is a step function; and (2) the absolutely summable autocovariance function that is associated with a continuous spectral density.

Example 4.1.1. Let us consider an example. Assume that the covariance function is given by (4.1.4) with the variances and frequencies specified by Table 4.1.1. Defining $I(\omega_j)$ as in (4.1.5), we have

$$I(0) = \tfrac{1}{2},$$
$$I(-\tfrac{3}{4}\pi) = I(\tfrac{3}{4}\pi) = \tfrac{1}{8},$$
$$I(-\tfrac{1}{10}\pi) = I(\tfrac{1}{10}\pi) = \tfrac{3}{8},$$
$$I(-\tfrac{3}{8}\pi) = I(\tfrac{3}{8}\pi) = \tfrac{5}{8},$$
$$I(-\tfrac{7}{8}\pi) = I(\tfrac{7}{8}\pi) = \tfrac{3}{8}.$$

The original variances are plotted against frequency in Figure 4.1.1, and the line spectrum is plotted in Figure 4.1.2. The associated spectral distribution function is given in Figure 4.1.3. ▲▲

Table 4.1.1. Examples of Variances for Time Series of Form (4.1.3)

j	σ_j^2	ω_j
0	$\frac{1}{2}$	0
1	$\frac{1}{4}$	$-\frac{3}{4}\pi$
2	$\frac{3}{4}$	$-\frac{1}{10}\pi$
3	$\frac{5}{4}$	$\frac{3}{8}\pi$
4	$\frac{3}{4}$	$\frac{7}{8}\pi$

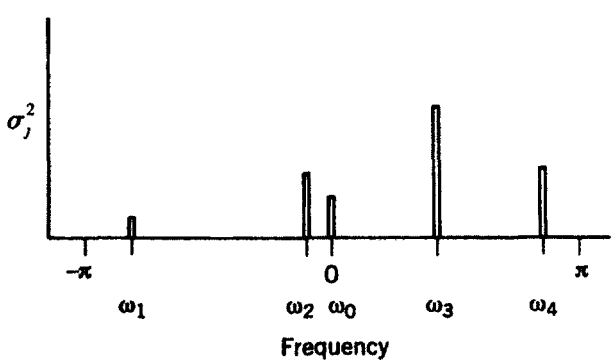

Figure 4.1.1. Graph of σ_j^2 for the time series of Table 4.1.1.

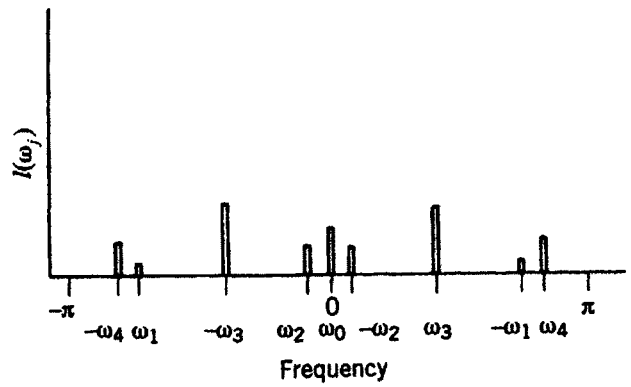

Figure 4.1.2. Graph of the line spectrum for the time series of Table 4.1.1.

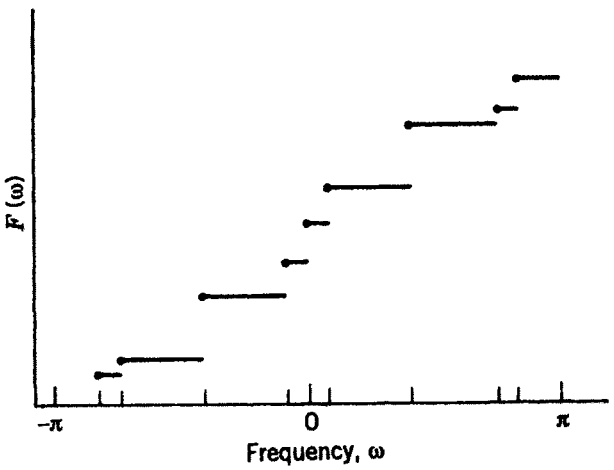

Figure 4.1.3. Spectral distribution function associated with the spectrum of Figure 4.1.2.

Example 4.1.2. As a second example, let X_t be defined by

$$X_t = \epsilon_1 \cos \frac{\pi}{2} t + \epsilon_2 \sin \frac{\pi}{2} t + e_t, \tag{4.1.6}$$

where the e_t, $t = 0, \pm 1, \pm 2, \ldots$, are independent $(0, 0.2\pi)$ random variables independent of the ϵ_j, $j = 1, 2$, which are independent $(0, 0.8\pi)$ random variables. Letting

$$Y_t = \epsilon_1 \cos \frac{\pi}{2} t + \epsilon_2 \sin \frac{\pi}{2} t,$$

it follows that

$$F_X(\omega) = F_Y(\omega) + F_e(\omega),$$
$$F_e(\omega) = 0.1\pi + 0.1\omega, \qquad -\pi \leq \omega \leq \pi,$$

and

$$F_Y(\omega) = \begin{cases} 0, & -\pi \leq \omega < -\pi/2, \\ 0.4\pi, & -\pi/2 \leq \omega < \pi/2, \\ 0.8\pi, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

Therefore,

$$F_X(\omega) = \begin{cases} 0.1\pi + 0.1\omega, & -\pi \leq \omega < -\pi/2, \\ 0.5\pi + 0.1\omega, & -\pi/2 \leq \omega < \pi/2, \\ 0.9\pi + 0.1\omega, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

The autocovariance function of X_t is

$$\begin{aligned} \gamma_X(h) &= \int_{-\pi}^{\pi} e^{i\omega h}\, dF_X(\omega) \\ &= \sum_{j=-1}^{1} I(\omega_j)\cos \omega_j h + \int_{-\pi}^{\pi} 0.1 \cos \omega h \, d\omega \\ &= \begin{cases} \pi, & h = 0, \\ 0.8\pi \cos \dfrac{\pi}{2} h & \text{otherwise}, \end{cases} \end{aligned}$$

where $\omega_{-1} = -\pi/2$ and $\omega_1 = \pi/2$. The reader may verify this expression by evaluating $E\{X_t X_{t+h}\}$ using the definition of X_t given in (4.1.6). ▲▲

4.2. CIRCULANTS—DIAGONALIZATION OF THE COVARIANCE MATRIX OF STATIONARY PROCESSES

In this section we investigate some properties of matrices encountered in time series analysis and use these properties to obtain the matrix that will approximately diagonalize the $n \times n$ covariance matrix of n observations from a stationary time series with absolutely summable covariance function. Let the $n \times n$ covariance matrix be denoted by Γ. Then

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \cdots & \gamma(0) \end{pmatrix}.$$

It is well known[2] that for any $n \times n$ positive semidefinite covariance matrix Γ, there exists an $\mathbf{M}$ satisfying $\mathbf{M}'\mathbf{M} = \mathbf{I}$ such that

$$\mathbf{M}'\Gamma\mathbf{M} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where λ_i, $i = 1, 2, \ldots, n$, are the characteristic roots of Γ. Our investigation will demonstrate that for large n the λ_i are approximately equal to $2\pi f(\omega_j)$, where $f(\omega)$ is the spectral density of X_t and $\omega_j = 2\pi j/n$, $j = 0, 1, 2, \ldots, n-1$. This permits us to interpret the spectral density as a multiple of the variance of the orthogonal random variables defined by the transformation $\mathbf{M}$. We initiate our study by introducing a matrix whose roots have a particularly simple representation.

A matrix of the form

$$\Gamma_c = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \gamma(n-1) & \gamma(0) & \cdots & \gamma(n-3) & \gamma(n-2) \\ \gamma(n-2) & \gamma(n-1) & \cdots & \gamma(n-4) & \gamma(n-3) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(2) & \gamma(3) & \cdots & \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) & \gamma(0) \end{bmatrix} \tag{4.2.1}$$

is called a *circular matrix* or *circulant*. Of course, in this definition $\gamma(j)$ may be any number, but we use the covariance notation, since our immediate application will be to a matrix whose elements are covariances.

The characteristic roots λ_j and vectors $\mathbf{x}_j$ of the matrix (4.2.1) satisfy the equation

$$\Gamma_c \mathbf{x}_j = \lambda_j \mathbf{x}_j, \qquad j = 1, 2, \ldots, n, \tag{4.2.2}$$

and we have

$$\begin{aligned} \gamma(0)x_{1j} + \gamma(1)x_{2j} + \cdots + \gamma(n-2)x_{n-1,j} + \gamma(n-1)x_{nj} &= \lambda_j x_{1j}, \\ \gamma(n-1)x_{1j} + \gamma(0)x_{2j} + \cdots + \gamma(n-3)x_{n-1,j} + \gamma(n-2)x_{nj} &= \lambda_j x_{2j}, \\ &\vdots \\ \gamma(1)x_{1j} + \gamma(2)x_{2j} + \cdots + \gamma(n-1)x_{n-1,j} + \gamma(0)x_{nj} &= \lambda_j x_{nj}, \end{aligned} \tag{4.2.3}$$

where x_{kj} is the kth element of the jth characteristic vector. Let r_j be a root of the scalar equation $r^n = 1$ and set $x_{kj} = r_j^k$. The system of equations (4.2.3) becomes

$$\begin{aligned} \gamma(0)r_j + \gamma(1)r_j^2 + \cdots + \gamma(n-2)r_j^{n-1} + \gamma(n-1)r_j^n &= \lambda_j r_j, \\ \gamma(n-1)r_j + \gamma(0)r_j^2 + \cdots + \gamma(n-3)r_j^{n-1} + \gamma(n-2)r_j^n &= \lambda_j r_j^2, \\ &\vdots \\ \gamma(1)r_j + \gamma(2)r_j^2 + \cdots + \gamma(n-1)r_j^{n-1} + \gamma(0)r_j^n &= \lambda_j r_j^n. \end{aligned} \tag{4.2.4}$$

[2] See, for example, Rao (1973, p. 39).

If we multiply the first equation of (4.2.4) by r_j^{n-1}, the second by r_j^{n-2}, and so forth, using $r_j^{n+k} = r_j^k$, we see that we shall obtain equality for each equation if

$$\lambda_j = \sum_{h=0}^{n-1} \gamma(h) r_j^h . \tag{4.2.5}$$

The equation $r^n = 1$ has n distinct roots, $e^{\epsilon 2\pi j/n}$, $j = 1, 2, \ldots, n$, which may also be expressed as $r_j = e^{-\epsilon 2\pi j/n}$, $j = 0, 1, \ldots, n-1$. Therefore, the n characteristic vectors associated with the n characteristic roots are given by

$$\mathbf{x}_j = \mathbf{g}_j^* = n^{-1/2}[1, e^{-\epsilon 2\pi j/n}, e^{-\epsilon 2\pi 2j/n}, \ldots, e^{-\epsilon 2\pi(n-1)j/n}]' , \qquad j = 0, 1, \ldots, n-1 , \tag{4.2.6}$$

where $\mathbf{g}_{j.}^*$ is the conjugate transpose of the row vector

$$\mathbf{g}_{j.} = n^{-1/2}[1, e^{\epsilon 2\pi j/n}, e^{\epsilon 2\pi 2j/n}, \ldots, e^{\epsilon 2\pi(n-1)j/n}] .$$

If we define the matrix $\mathbf{G}$ by

$$\mathbf{G} = (\mathbf{g}_{0.}', \mathbf{g}_{1.}', \ldots, \mathbf{g}_{n.}')' ,$$

then

$$\mathbf{G}\mathbf{G}^* = \mathbf{I} ,$$

$$\mathbf{G}\boldsymbol{\Gamma}_c\mathbf{G}^* = \operatorname{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1}) ,$$

where $\mathbf{G}^*$ is the conjugate transpose of $\mathbf{G}$.

Setting $\gamma(1) = \gamma(n-1)$, $\gamma(2) = \gamma(n-2), \ldots, \gamma(h) = \gamma(n-h), \ldots$ in $\boldsymbol{\Gamma}_c$, we obtain the circular symmetric matrix

$$\boldsymbol{\Gamma}_s = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(3) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(4) & \gamma(3) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(1) & \gamma(0) \end{bmatrix} . \tag{4.2.7}$$

Substituting into (4.2.5), we obtain for the roots of (4.2.7)

$$\lambda_j = \begin{cases} \displaystyle\sum_{h=-(n-1)/2}^{(n-1)/2} \gamma(h) e^{-\epsilon 2\pi hj/n} , & n \text{ odd} , \\ \displaystyle\sum_{h=-(n/2)+1}^{n/2} \gamma(h) e^{-\epsilon 2\pi hj/n} , & n \text{ even} , \end{cases} \tag{4.2.8}$$

where we have used the periodic property $e^{-\epsilon 2\pi j(n-h)/n} = e^{\epsilon 2\pi jh/n}$. Note that these roots are real, as they must be for a real symmetric matrix.

In some applications it is preferable to work with a real matrix rather than the complex matrix **G**. Consider first the case of n odd. Equation (4.2.8) may also be written

$$\lambda_j = \sum_{h=-(n-1)/2}^{(n-1)/2} \gamma(h)\cos\frac{2\pi}{n}hj\,, \qquad j = 0, 1, \ldots, n-1\,. \tag{4.2.9}$$

Since for $0 \leq m \leq 2\pi$ we have $\cos m = \cos(2\pi - m)$, we see that there is a root for $j = 0$ (or $j = n$) and $(n-1)/2$ roots of multiplicity two associated with $j = 1, 2, \ldots, (n-1)/2$. For each of these repeated roots we can find two real orthogonal vectors. These are chosen to be

$$2^{1/2}n^{-1/2}\left[1, \cos 2\pi\frac{j}{n}, \cos 2\pi\frac{2j}{n}, \cos 2\pi\frac{3j}{n}, \ldots, \cos 2\pi\frac{(n-1)j}{n}\right]$$

and

$$2^{1/2}n^{-1/2}\left[0, \sin 2\pi\frac{j}{n}, \sin 2\pi\frac{2j}{n}, \sin 2\pi\frac{3j}{n}, \ldots, \sin 2\pi\frac{(n-1)j}{n}\right]. \tag{4.2.10}$$

If we choose the vectors in (4.2.6) associated with the jth and $n-j$th roots of one, the vectors in (4.2.10) are given by

$$2^{-1/2}(\mathbf{g}_{j.} + \mathbf{g}_{n-j,.}) \quad \text{and} \quad 2^{-1/2}i(\mathbf{g}_{j.} - \mathbf{g}_{n-j,.})\,.$$

Much the same pattern holds for the roots of a circular symmetric matrix of dimension $n \times n$ where n is even. There is a characteristic vector $n^{-1/2}(1, 1, \ldots, 1)$ associated with $j = 0$, and a vector $n^{-1/2}(1, -1, 1, \ldots, -1)$ associated with $j = n/2$. The remaining $(n/2) - 1$ roots have multiplicity two, and the roots are given by (4.2.8).

Define the orthogonal matrix **Q** by setting $n^{1/2}2^{-1/2}\mathbf{Q}'$ equal to

$$\begin{bmatrix}
2^{-1/2} & 2^{-1/2} & 2^{-1/2} & \cdots & 2^{-1/2} \\
1 & \cos 2\pi\frac{1}{n} & \cos 2\pi\frac{2}{n} & \cdots & \cos 2\pi\frac{n-1}{n} \\
0 & \sin 2\pi\frac{1}{n} & \sin 2\pi\frac{2}{n} & \cdots & \sin 2\pi\frac{n-1}{n} \\
1 & \cos 4\pi\frac{1}{n} & \cos 4\pi\frac{2}{n} & \cdots & \cos 4\pi\frac{n-1}{n} \\
\vdots & \vdots & \vdots & & \vdots \\
0 & \sin\frac{n-1}{2}2\pi\frac{1}{n} & \sin\frac{n-1}{2}2\pi\frac{2}{n} & \cdots & \sin\frac{n-1}{2}2\pi\frac{n-1}{n}
\end{bmatrix}. \tag{4.2.11}$$

Note that **Q** is the matrix composed of the n characteristic vectors defined by

(4.2.10) and our illustration (4.2.11) is for odd n. Let X_t be a stationary time series with absolutely summable covariance function $\gamma(h)$. For odd n, define the $n \times n$ diagonal matrix $\mathbf{D}$ by

$$\mathbf{D} = \operatorname{diag}\{d_1, d_2, \ldots, d_n\}, \tag{4.2.12}$$

where

$$d_1 = f(0) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h),$$

$$d_{2j} = d_{2j+1} = f\left(2\pi \frac{j}{n}\right) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i2\pi hj/n}, \tag{4.2.13}$$

$j = 1, 2, \ldots, 0.5(n-1)$. For $\mathbf{\Gamma}_s$ defined by (4.2.7) the matrix $\mathbf{Q}'\mathbf{\Gamma}_s\mathbf{Q}$ is a diagonal matrix whose elements converge to $2\pi\mathbf{D}$ as n increases. This also holds for even n if the definition of $\mathbf{Q}$ is slightly modified. An additional row,

$$n^{-1/2}[1, -1, 1, \ldots, 1, -1],$$

which is the characteristic vector associated with $j = n/2$, is added to the $\mathbf{Q}'$ of (4.2.11) when n is even. The last entry in $\mathbf{D}$ for even n is

$$d_n = f(\pi) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \pi h.$$

The covariance matrix for n observations on X_t is given by

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}. \tag{4.2.14}$$

For $\gamma(h)$ absolutely summable, we now demonstrate that $\mathbf{Q}'\mathbf{\Gamma}\mathbf{Q}$ also converges to $2\pi\mathbf{D}$.

Let $\mathbf{q}_{\cdot i} = [q_{1i}, q_{2i}, \ldots, q_{ni}]'$ be the ith column of $\mathbf{Q}$. We have, for $\mathbf{\Gamma}$ defined in (4.2.14) and $\mathbf{\Gamma}_s$ defined in (4.2.7),

$$|\mathbf{q}_{\cdot i}'\mathbf{\Gamma}_s\mathbf{q}_{\cdot j} - \mathbf{q}_{\cdot i}'\mathbf{\Gamma}\mathbf{q}_{\cdot j}| = \left| \sum_{m=1}^{M} [\gamma(m) - \gamma(n-m)] \times \sum_{k=1}^{m} [q_{ki}q_{n-m+k,j} + q_{n-m+k,i}q_{kj}] \right|, \tag{4.2.15}$$

where $M = (n-1)/2$ if n is odd and $M = (n/2) - 1$ if n is even. Now (4.2.15) is less than

$$\frac{4}{n}\left[\sum_{m=1}^{M} m|\gamma(m)| + \sum_{m=1}^{M} m|\gamma(n-m)|\right] \leq \frac{4}{n}\left[\sum_{m=1}^{M} m|\gamma(m)| + \sum_{h=M+1}^{n} M|\gamma(h)|\right], \quad (4.2.16)$$

since $q_{ki}q_{rj} \leq 2/n$ for all $k, i, r, j \in (1, 2, \ldots, n)$. As n increases, the limit of the first term is zero by Lemma 3.1.4 and the limit of the second term is zero by the absolute summability of $\gamma(h)$. Therefore, the elements of $\mathbf{Q}'\boldsymbol{\Gamma}\mathbf{Q}$ converge to $2\pi\mathbf{D}$. We state the result as a theorem.

Theorem 4.2.1. Let $\boldsymbol{\Gamma}$ be the covariance matrix of n observations from a stationary time series with absolutely summable covariance function. Let $\mathbf{Q}$ be defined by (4.2.11), and take $\mathbf{D}$ to be the $n \times n$ diagonal matrix defined in (4.2.12). Then, given $\epsilon > 0$, there exists an n_ϵ such that for $n > n_\epsilon$ every element of the matrix

$$\mathbf{Q}'\boldsymbol{\Gamma}\mathbf{Q} - 2\pi\mathbf{D}$$

is less than ϵ in magnitude.

Corollary 4.2.1. Let $\boldsymbol{\Gamma}$ be the covariance matrix of n observations from a stationary time series with covariance function that satisfies

$$\sum_{h=-\infty}^{\infty} |h|\,|\gamma(h)| = L < \infty.$$

Let $\mathbf{Q}$ and $\mathbf{D}$ be as defined in Theorem 4.2.1. Then every element of the matrix $\mathbf{Q}'\boldsymbol{\Gamma}\mathbf{Q} - 2\pi\mathbf{D}$ is less than $4L/n$ in magnitude.

Proof. By (4.2.16) the difference between $\mathbf{q}_{\cdot i}'\boldsymbol{\Gamma}_s\mathbf{q}_{\cdot j}$ and $\mathbf{q}_{\cdot i}'\boldsymbol{\Gamma}\mathbf{q}_{\cdot j}$ is less than

$$\frac{4}{n}\left[\sum_{m=1}^{M} m|\gamma(m)| + \sum_{h=M+1}^{n} M|\gamma(h)|\right] \leq \frac{4}{n}\sum_{m=1}^{n} m|\gamma(m)| < \frac{2L}{n}.$$

Now $\mathbf{Q}'\boldsymbol{\Gamma}_s\mathbf{Q}$ is a diagonal matrix, and the difference between an element of $\mathbf{Q}'\boldsymbol{\Gamma}_s\mathbf{Q}$ and an element of $2\pi\mathbf{D}$ is

$$\left|\sum_{h=-(n-1)/2}^{(n-1)/2} \gamma(h)e^{-i2\pi hj/n} - \sum_{h=-\infty}^{\infty} \gamma(h)e^{-i2\pi hj/n}\right| \leq 2\sum_{h=(n-1)/2+1}^{\infty} |\gamma(h)| \leq \frac{4}{n}\sum_{h=(n+1)/2}^{\infty} h|\gamma(h)| < \frac{2L}{n}. \quad \blacktriangle$$

We have demonstrated that, asymptotically, the Q defined by (4.2.11) or the G defined by (4.2.6) will diagonalize *all* Γ-matrices associated with stationary time series with absolutely summable covariance functions. That is, the transformation Q applied to the n observations defines n new random variables that are "nearly" uncorrelated. Each of the new random variables is a linear combination of the n original observations with weights given in (4.2.10). The variances of both the $2j$th and $(2j+1)$th random variables created by the transformation are approximately $2\pi f(2\pi j/n)$.

For time series with absolutely summable covariance function the variance result holds for the random variable defined for any ω, not just for integer multiples of $2\pi/n$. That is, the complex random variable

$$Y_{\omega,n} = n^{-1/2} \sum_{t=1}^{n} X_t e^{i\omega t}$$

has variance given by

$$\begin{aligned} E\{|Y_{\omega,n}|^2\} &= \frac{1}{n} E\left\{ \sum_{t=1}^{n} X_t e^{-i\omega t} \sum_{t=1}^{n} X_t e^{i\omega t} \right\} \\ &= \frac{1}{n} E\left\{ \sum_{t=1}^{n} \sum_{j=1}^{n} X_t X_j e^{-i\omega(t-j)} \right\} \\ &= \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{n} \gamma(t-j) e^{-i\omega(t-j)} \\ &= \sum_{h=-(n-1)}^{n-1} n^{-1}(n-|h|)\gamma(h) e^{-i\omega h} . \end{aligned}$$

By Lemma 3.1.4,

$$\lim_{n\to\infty} E\{|Y_{\omega,n}|^2\} = 2\pi f_X(\omega) .$$

We shall return to this point in Chapter 7 and investigate functions of the random variables created by applying the transformation (4.2.10) or (4.2.6) to a finite number of observations from a realization.

4.3. THE SPECTRAL DENSITY OF MOVING AVERAGE AND AUTOREGRESSIVE TIME SERIES

By the Fourier integral theorem we know that the Fourier transform of an absolutely summable sequence is unique and that the sequence is given once again by the inverse transform of the transform. Therefore, we expect the spectral density of a particular kind of time series, say a finite moving average, to have a

distinctive form. We shall see that this is the case: the spectral densities of autoregressive and moving average processes are easily recognizable.

Stationary moving average, stationary autoregressive, and stationary autoregressive moving average time series are contained within the representation

$$X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \tag{4.3.0}$$

where

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

and the e_t are uncorrelated $(0, \sigma^2)$ random variables. Therefore, if we obtain the spectral density of a time series of the form (4.3.0), we shall have the spectral density for such types of time series. In fact, it is convenient to treat the case of an infinite weighted sum of a more general time series.

Theorem 4.3.1. Let X_t be a stationary time series with an absolutely summable covariance function and let $\{a_j\}_{j=-\infty}^{\infty}$ be absolutely summable. Then the spectral density of

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

is

$$f_Y(\omega) = (2\pi)^2 f_X(\omega) f_a(\omega) f_a^*(\omega), \tag{4.3.1}$$

where $f_X(\omega)$ is the spectral density of X_t,

$$f_a(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}$$

is the Fourier transform of a_j, and

$$f_a^*(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{i\omega j}$$

is the complex conjugate of the Fourier transform of a_j.

Proof. Given the absolute summability, we may interchange integration and summation and calculate the expectation term by term. Therefore, letting

$E\{X_t\} = 0$, we have

$$\begin{aligned} E\{Y_t Y_{t+h}\} &= E\left\{\left(\sum_{j=-\infty}^{\infty} a_j X_{t-j}\right)\left(\sum_{k=-\infty}^{\infty} a_k X_{t-k+h}\right)\right\} \\ &= E\left\{\sum_{j=-\infty}^{\infty} a_j \sum_{k=-\infty}^{\infty} a_k X_{t-j} X_{t-k+h}\right\} \\ &= \sum_{j=-\infty}^{\infty} a_j \sum_{k=-\infty}^{\infty} a_k \gamma_X(h-k+j). \end{aligned}$$

If we set $p = h - k + j$, then

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} a_j \sum_{p=-\infty}^{\infty} a_{j+h-p} \gamma_X(p). \tag{4.3.2}$$

Since $\gamma_Y(h)$ is absolutely summable, we may write

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} \gamma_Y(h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} \sum_{j=-\infty}^{\infty} a_j \sum_{p=-\infty}^{\infty} a_{j+h-p} \gamma_X(p) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} a_j a_{j+h-p} \gamma_X(p) e^{i\omega j} e^{-i\omega p} e^{-i\omega(j+h-p)} \\ &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{i\omega j} \sum_{p=-\infty}^{\infty} \gamma_X(p) e^{-i\omega p} \sum_{s=-\infty}^{\infty} a_s e^{-i\omega s}, \end{aligned}$$

where we have made the transformation $s = j + h - p$. ▲

The theorem is particularly useful, and we shall make repeated application of the result. If the X_t of the theorem are assumed to be uncorrelated random variables, we obtain the spectral density of a moving average time series.

Corollary 4.3.1.1. The spectral density of the moving average process

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

where $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables and the sequence $\{a_j\}$ is absolutely summable, is

$$f_X(\omega) = \frac{\sigma^2}{2\pi}\left(\sum_{j=-\infty}^{\infty} a_j e^{-ij\omega}\right)\left(\sum_{j=-\infty}^{\infty} a_j e^{ij\omega}\right). \tag{4.3.3}$$

Proof. The result follows immediately from Theorem 4.3.1. ▲

Let us use (4.3.3) to calculate the spectral density of the finite moving average

$$X_t = \sum_{j=0}^{p} a_j e_{t-j}, \tag{4.3.4}$$

We have

$$\begin{aligned}
f_X(\omega) &= \frac{\sigma^2}{2\pi}\left(\sum_{j=0}^{p} a_j e^{-\mathrm{i}j\omega}\right)\left(\sum_{j=0}^{p} a_j e^{\mathrm{i}j\omega}\right) \\
&= \frac{\sigma^2}{2\pi}(a_0 + a_1 e^{-\mathrm{i}\omega} + a_2 e^{-\mathrm{i}2\omega} + \cdots + a_p e^{-\mathrm{i}p\omega}) \\
&\quad \times (a_0 + a_1 e^{\mathrm{i}\omega} + a_2 e^{\mathrm{i}2\omega} + \cdots + a_p e^{\mathrm{i}p\omega}) \\
&= \frac{\sigma^2}{2\pi}[(a_0^2 + a_1^2 + a_2^2 + \cdots + a_p^2) \\
&\quad + (a_0 a_1 + a_1 a_2 + a_2 a_3 + \cdots + a_{p-1} a_p) e^{-\mathrm{i}\omega} \\
&\quad + (a_0 a_1 + a_1 a_2 + a_2 a_3 + \cdots + a_{p-1} a_p) e^{\mathrm{i}\omega} \\
&\quad + (a_0 a_2 + a_1 a_3 + \cdots + a_{p-2} a_p) e^{-\mathrm{i}2\omega} \\
&\quad + \cdots + a_0 a_p e^{-\mathrm{i}p\omega} + a_0 a_p e^{\mathrm{i}p\omega}] \\
&= \frac{1}{2\pi}[\gamma_X(0) + \gamma_X(1) e^{-\mathrm{i}\omega} + \gamma_X(1) e^{\mathrm{i}\omega} + \gamma_X(2) e^{-\mathrm{i}2\omega} \\
&\quad + \cdots + \gamma_X(p) e^{-\mathrm{i}p\omega} + \gamma_X(p) e^{\mathrm{i}p\omega}] \\
&= \frac{1}{2\pi}\left[\sum_{h=-p}^{p} \gamma_X(h) e^{-\mathrm{i}\omega h}\right].
\end{aligned} \tag{4.3.5}$$

Equation (4.3.5) is a proof of the corollary for finite p and serves to reinforce our confidence in the general result. The spectral density of a finite moving average may, alternatively, be expressed in terms of the roots of the auxiliary equation. The quantity

$$\sum_{j=0}^{p} a_j e^{-\mathrm{i}\omega j}$$

is seen to be a polynominal in $e^{-\mathrm{i}\omega}$. Therefore, it may be written as

$$\sum_{j=0}^{p} a_j e^{-\mathrm{i}\omega j} = e^{-\mathrm{i}\omega p} \sum_{j=0}^{p} a_j e^{\mathrm{i}\omega(p-j)} = e^{-\mathrm{i}\omega p} \prod_{j=1}^{p} (e^{\mathrm{i}\omega} - m_j),$$

where $a_0 = 1$ and the m_j are the roots of

$$m^p + a_1 m^{p-1} + a_2 m^{p-2} + \cdots + a_p = 0.$$

It follows that the spectral density of the $\{X_t\}$ defined by (4.3.4) is

$$f_X(\omega) = \frac{\sigma^2}{2\pi}\left[\prod_{j=1}^{p}(e^{-i\omega} - m_j)\right]\left[\prod_{j=1}^{p}(e^{i\omega} - m_j)\right]. \tag{4.3.6}$$

We can also use Corollary 4.3.1.1 to compute the spectral density of the first order autoregressive process

$$X_t = \sum_{j=0}^{\infty} \rho^j e_{t-j}, \qquad |\rho| < 1.$$

The transform of the weights $a_j = \rho^j$ is given by

$$\begin{aligned} g_a(\omega) &= \frac{1}{2\pi}\sum_{j=0}^{\infty} \rho^j e^{-ij\omega} \\ &= \frac{1}{2\pi}\sum_{j=0}^{\infty} (\rho e^{-i\omega})^j = \frac{1}{2\pi}\frac{1}{1-\rho e^{-i\omega}} \end{aligned}$$

and the complex conjugate by

$$g_a^*(\omega) = \frac{1}{2\pi}\sum_{j=0}^{\infty} \rho^j e^{ij\omega} = \frac{1}{2\pi}\frac{1}{1-\rho e^{i\omega}}.$$

Therefore,

$$\begin{aligned} f_X(\omega) &= (2\pi)^2 f_e(\omega) g_a(\omega) g_a^*(\omega) \\ &= \frac{\sigma^2}{2\pi}\left(\frac{1}{1-\rho e^{-i\omega}}\right)\left(\frac{1}{1-\rho e^{i\omega}}\right) \\ &= \frac{\sigma^2}{2\pi}\frac{1}{1-2\rho\cos\omega + \rho^2}, \end{aligned} \tag{4.3.7}$$

since $f_e(\omega)$, the spectral density of the uncorrelated sequence $\{e_t\}$, is $\sigma^2/2\pi$ for all ω.

We note that we could also find the spectral density of X_t by setting $Y_t = e_t$ in (4.3.1) and using the known spectral density of e_t. Applying this approach to the pth order autoregressive time series, we have the following corollary.

Corollary 4.3.1.2. The spectral density of the pth order autoregressive time series $\{X_t\}$ defined by

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

where $\alpha_0 = 1$, $\alpha_p \neq 0$, the e_t are uncorrelated $(0, \sigma^2)$ random variables, and the roots $\{m_j: j = 1, 2, \ldots, p\}$ of

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

are less than one in absolute value, is given by

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi}\left[\left(\sum_{j=0}^{p} \alpha_j e^{-i\omega j}\right)\left(\sum_{j=0}^{p} \alpha_j e^{i\omega j}\right)\right]^{-1} \\ &= \frac{\sigma^2}{2\pi}\prod_{j=1}^{p} \frac{1}{(1 - 2m_j \cos \omega + m_j^2)}. \end{aligned} \tag{4.3.8}$$

Proof. Apply Theorem 4.3.1 to the finite moving average time series

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

to obtain

$$(2\pi)^{-1}\sigma^2 = f_e(\omega) = T(\omega) f_X(\omega),$$

where

$$\begin{aligned} T(\omega) &= \left(\sum_{j=0}^{p} \alpha_j e^{-i\omega j}\right)\left(\sum_{j=0}^{p} \alpha_j e^{i\omega j}\right) \\ &= \left(e^{i\omega p}\sum_{j=0}^{p} \alpha_j e^{-i\omega j}\right)\left(e^{-i\omega p}\sum_{j=0}^{p} \alpha_j e^{i\omega j}\right) \\ &= R(e^{i\omega})R(e^{-i\omega}). \end{aligned}$$

Neither $R(e^{i\omega})$ nor $R(e^{-i\omega})$ can be zero, since the roots of $R(m)$ cannot have an absolute value of one. Therefore, $T(\omega)$ is not equal to zero for any ω, and we are able to write

$$f_X(\omega) = f_e(\omega)T^{-1}(\omega).$$

This is the first representation given in (4.3.8). The alternative form for $f_X(\omega)$ follows by writing $T(\omega)$ in the factored form

$$\begin{aligned} T(\omega) &= \left|\prod_{j=1}^{p} (1 - e^{-i\omega} m_j)\right|^2 \\ &= \prod_{j=1}^{p} (1 - 2m_j \cos \omega + m_j^2). \end{aligned}$$

▲

We can also write (4.3.8) in the factored form

$$f_X(\omega) = \frac{\sigma^2}{2\pi}\left[\prod_{j=1}^{p}(e^{-i\omega} - m_j)\prod_{j=1}^{p}(e^{i\omega} - m_j)\right]^{-1}. \tag{4.3.9}$$

The spectral density of the autoregressive moving average process is sometimes called a rational spectral density because it is the ratio of polynomials in $e^{i\omega}$.

Corollary 4.3.1.3. Let an autoregressive moving average process of order (p, q) be defined by

$$\sum_{j=0}^{p} \alpha_j X_{t-j} = \sum_{k=0}^{q} \beta_k e_{t-k},$$

where $\alpha_0 = \beta_0 = 1$, $\alpha_p \neq 0$, $\beta_q \neq 0$, the e_t are uncorrelated $(0, \sigma^2)$ random variables, and the roots $\{m_j: j = 1, 2, \ldots, p\}$ of

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

are less than one in absolute value. Then the spectral density of $\{X_t\}$ is given by

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi}\frac{\left(\sum_{k=0}^{q}\beta_k e^{i\omega k}\right)\left(\sum_{k=0}^{q}\beta_k e^{-i\omega k}\right)}{\left(\sum_{j=0}^{p}\alpha_j e^{i\omega j}\right)\left(\sum_{j=0}^{p}\alpha_j e^{-i\omega j}\right)} \\ &= \frac{\sigma^2}{2\pi}\frac{\prod_{k=1}^{q}(e^{i\omega} - r_k)\prod_{k=1}^{q}(e^{-i\omega} - r_k)}{\prod_{j=1}^{p}(e^{i\omega} - m_j)\prod_{j=1}^{p}(e^{-i\omega} - m_j)}, \end{aligned} \tag{4.3.10}$$

where $\{r_k: k = 1, 2, \ldots, q\}$ are the roots of

$$\sum_{k=0}^{q} \beta_k r^{q-k} = 0.$$

Proof. Reserved for the reader. ▲

In Theorem 2.6.4 we demonstrated that a finite moving average process with no roots of unit absolute value always has a representation with characteristic equation whose roots are less than one in absolute value. In Theorem 2.6.3 we obtained a moving average representation for a time series with $|\rho(1)| < 0.5$ and $\rho(h) = 0$, $h = \pm 2, \pm 3, \ldots$. We can now extend Theorem 2.6.3 to time series with a finite number of nonzero covariances.

Theorem 4.3.2. Let the stationary time series X_t have zero mean and spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-q}^{q} \gamma(h) e^{-i\omega h},$$

where $f(\omega)$ is strictly positive, $\gamma(h)$ is the covariance function of X_t, and $\gamma(q) \neq 0$. Then X_t has the representation

$$X_t = \sum_{j=1}^{q} \beta_j e_{t-j} + e_t = \sum_{j=0}^{q} \beta_j e_{t-j},$$

where $\beta_0 = 1$, the e_t are uncorrelated $(0, \sigma^2)$ random variables with $\sigma^2 = \gamma(0)[\Sigma_{j=0}^{q} \beta_j^2]^{-1}$, and the roots of

$$m^q + \sum_{j=1}^{q} \beta_j m^{q-j} = 0$$

are less than one in absolute value.

Proof. The function

$$A(z) = z^q[\gamma(q)]^{-1} \sum_{h=-q}^{q} \gamma(h) z^h = [\gamma(q)]^{-1} \sum_{h=0}^{2q} \gamma(h-q) z^h$$

is a polynomial of degree $2q$ in z with unit coefficient for z^{2q}. Therefore, it can be written in the factored form

$$A(z) = \prod_{r=1}^{2q} (z - m_r),$$

where the m_r are the roots of $A(z) = 0$. If $A(m_r) = 0$, then

$$\sum_{h=-q}^{q} \gamma(h) m_r^h = 0,$$

and by symmetry

$$\sum_{h=-q}^{q} \gamma(h) m_r^{-h} = 0.$$

Since $f(\omega)$ is strictly positive, none of the roots are of unit absolute value. Also, the coefficients are real, so that any complex roots occur as conjugate pairs. Therefore, we can arrange the roots in q pairs, $(m_r, m_{-r}) = (m_r, m_r^{-1})$, where multiple roots are repeated the proper number of times. We let m_r, $r = 1, 2, \ldots, q$, denote the roots less than one in absolute value. Using these roots, we define β_j, for

$j = 0, 1, 2, \ldots, q$, by

$$\prod_{r=1}^{q} (z - m_r) = \sum_{j=0}^{q} \beta_j z^{q-j}.$$

Therefore,

$$\begin{aligned} z^{-q}\gamma(q)A(z) &= z^{-q}\gamma(q) \prod_{r=1}^{q} (z - m_r) \prod_{r=1}^{q} (z - m_r^{-1}) \\ &= \gamma(q) \prod_{r=1}^{q} (z - m_r) \prod_{r=1}^{q} (1 - z^{-1} m_r^{-1}) \\ &= \gamma(q) \left(\sum_{j=0}^{q} \beta_j z^j \right) \left(\sum_{j=0}^{q} \beta_j z^{-j} \right) \prod_{r=1}^{q} (-m_r)^{-1}. \end{aligned}$$

It follows that the expression in $e^{-i\omega}$ defining $f(\omega)$ can be written as

$$f(\omega) = 2\pi\sigma^2 |f_\beta(\omega)|^2,$$

where $f_\beta(\omega)$ is the Fourier transform of $\{\beta_j\}$ and $\sigma^2 = \gamma(0)(\Sigma_{j=0}^{q} \beta_j^2)^{-1}$. Now define the sequence of random variables $\{e_t\}$ by $e_t = \Sigma_{j=0}^{\infty} c_j X_{t-j}$, where the sequence $\{c_j\}$ is defined in Theorem 2.6.2. These c_j are such that $X_t = \Sigma_{j=0}^{q} \beta_j e_{t-j}$, which means that $|f_c(\omega) f_\beta(\omega)|^2 = (2\pi)^{-4}$. Therefore, the spectral density of e_t is

$$\begin{aligned} f_e(\omega) &= (2\pi)^2 |f_c(\omega)|^2 f(\omega) \\ &= (2\pi)^2 |f_c(\omega)|^2 \left| \frac{\sigma^2 |2\pi f_\beta(\omega)|^2}{2\pi} \right| \\ &= \frac{\sigma^2}{2\pi}, \end{aligned}$$

and the e_t are uncorrelated with variance σ^2. ▲

By our earlier results in Fourier series we know that a continuous periodic function can be approximated by a trigonometric polynomial. This means that a time series with a continuous spectral density is "very nearly" a moving average process and also "very nearly" an autoregressive process.

Theorem 4.3.3. Let $g(\omega)$ be a nonnegative even continuous periodic function of period 2π. Then, given an $\epsilon > 0$, there is a time series with the representation

$$X_t = \sum_{j=0}^{q} \beta_j e_{t-j}$$

such that $|f_X(\omega) - g(\omega)| < \epsilon$ for all $\omega \in [-\pi, \pi]$ where $\beta_0 = 1$, q is a finite

integer, the e_t are uncorrelated $(0, \sigma^2)$ random variables, and

$$\sigma^2 = \left(\sum_{j=0}^{q} \beta_j^2 \right)^{-1} \int_{-\pi}^{\pi} g(\omega)\, d\omega .$$

Proof. The result is trivial if $g(\omega) \equiv 0$; hence, we assume $g(\omega) > 0$ for some ω. Given $\epsilon > 0$, let $\delta = 2\epsilon G(4\pi M + 3G)^{-1}$ and set

$$C(\omega) = \begin{cases} g(\omega) & \text{if } g(\omega) > \delta , \\ \delta & \text{otherwise} , \end{cases}$$

where

$$M = \max_{\omega} g(\omega)$$

$$G = \int_{-\pi}^{\pi} g(\omega)\, d\omega .$$

Then by Theorem 3.1.10, there is a q such that $|f_Y(\omega) - C(\omega)| < \frac{1}{2}\delta$ for all $\omega \in [-\pi, \pi]$, where

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{h=-q}^{q} \gamma(h) e^{-i\omega h} ,$$

$$\gamma(h) = q^{-1}(q - |h|) a_h ,$$

and

$$a_h = \int_{-\pi}^{\pi} e^{i\omega h} C(\omega)\, d\omega .$$

By Exercise 4.14, $\gamma(h)$ is positive semidefinite, and $f_Y(\omega)$ is of the same form as the spectral density in Theorem 4.3.2. Therefore, we may write

$$f_Y(\omega) = (2\pi)^{-1} \tau^2 \left| \sum_{j=0}^{q} \beta_j e^{-i\omega j} \right|^2 ,$$

where $\beta_0 = 1$, the roots of $\sum_{j=0}^{q} \beta_j m^{q-j}$ are less than one in absolute value, and $\tau^2 = [\sum_{j=0}^{q} \beta_j^2]^{-1} \int_{-\pi}^{\pi} C(\omega)\, d\omega$. Setting

$$f_X(\omega) = (2\pi)^{-1} \sigma^2 \left| \sum_{j=0}^{q} \beta_j e^{-i\omega j} \right|^2 ,$$

where

$$\sigma^2 = \tau^2 \left[\int_{-\pi}^{\pi} f_Y(\omega)\, d\omega \right]^{-1} \int_{-\pi}^{\pi} g(\omega)\, d\omega ,$$

we obtain the desired result. ▲

Theorem 4.3.4. Let $g(\omega)$ be a nonnegative even continuous periodic function of period 2π. Then, given an $\epsilon > 0$, there is a time series with the representation

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

such that $|f_Y(\omega) - g(\omega)| < \epsilon$ for all $\omega \in [-\pi, \pi]$, where $\alpha_0 = 1$, p is a finite integer, the roots of

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

are less than one in absolute value, and the e_t are uncorrelated $(0, \sigma^2)$ random variables.

Proof. The result is trivial if $g(\omega) \equiv 0$. Assuming $g(\omega) > 0$ for some ω define

$$d(\omega) = \begin{cases} g^{-1}(\omega) & \text{if } g(\omega) > \frac{1}{2}\epsilon, \\ 2\epsilon^{-1} & \text{otherwise.} \end{cases}$$

Also define $G = \max g(\omega)$, and let $0 < \delta < \epsilon[G(2G + \epsilon)]^{-1}$. Then, by Theorem 3.1.10, there is a finite p such that

$$\left| \sum_{h=-p}^{p} C_h e^{-i\omega h} - d(\omega) \right| < \delta$$

for all $\omega \in [-\pi, \pi]$. By Theorem 4.3.2,

$$\sum_{h=-p}^{p} C_h e^{-i\omega h} = M \sum_{j=0}^{p} \alpha_j e^{-i\omega j} \sum_{j=0}^{p} \alpha_j e^{i\omega j},$$

where $\alpha_0 = 1$, M is a constant, and the roots of $\Sigma_{j=0}^{p} \alpha_j m^{p-j} = 0$ are less than one in absolute value. Hence, by defining

$$f_Y(\omega) = M^{-1}\left[\left(\sum_{j=0}^{p} \alpha_j e^{-i\omega j}\right)\left(\sum_{j=0}^{p} \alpha_j e^{i\omega j}\right)\right]^{-1}$$

and $\sigma^2 = 2\pi M^{-1}$, we have the conclusion. ▲

Operations on a time series are sometimes called *filtering*. In some areas, particularly in electronics, physical devices are constructed to filter time series. Examples include the modification of radio waves and radar signals. The creation of a linearly weighted sum is the application of a *linear filter*, the weights being identified as the filter. If, as in (4.3.0), the weights are constant over time, the filter

is called *time invariant*.[3] Obviously such operations change the behavior of the time series, and in practice this is often the objective. By investigating the properties of the filter, one is able to make statements about the change in behavior.

It is clear that moving average and autoregressive processes are obtained from a white noise process by the application of a linear time invariant filter. Thus, using this terminology, we have been studying the effects of a linear time invariant filter on the spectral density of a time series.

To further illustrate the effects of the application of a linear filter, consider the simple time series

$$\begin{aligned} X_t &= 2\alpha \cos(\omega t + \beta) \\ &= \alpha[e^{i(\omega t+\beta)} + e^{-i(\omega t+\beta)}] \end{aligned}$$

and apply the absolutely summable linear filter a_j to obtain

$$\begin{aligned} Y_t &= \alpha \sum_{j=-\infty}^{\infty} a_j [e^{i\{\omega(t-j)+\beta\}} + e^{-i\{\omega(t-j)+\beta\}}] \\ &= \alpha e^{i(\omega t+\beta)} \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j} + \alpha e^{-i(\omega t+\beta)} \sum_{j=-\infty}^{\infty} a_j e^{i\omega j}. \end{aligned}$$

We may express the Fourier transform of a_j in a general complex form

$$\begin{aligned} 2\pi f_a(\omega) &= \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j} \overset{\text{(say)}}{=} u(\omega) + iv(\omega) \\ &= \psi(\omega)[\cos\varphi(\omega) + i\sin\varphi(\omega)] = \psi(\omega)e^{i\varphi(\omega)}, \end{aligned} \tag{4.3.11}$$

where

$$\begin{aligned} \psi(\omega) &= \{[u(\omega)]^2 + [v(\omega)]^2\}^{1/2}, \\ \varphi(\omega) &= \tan^{-1}\left[\frac{v(\omega)}{u(\omega)}\right]. \end{aligned}$$

The reader should note the convention used in defining the sign of $v(\omega)/u(\omega)$: $v(\omega)$ is the coefficient of i (not $-i$) in the original transform.

In this notation, the filtered cosine wave becomes

$$\begin{aligned} Y_t &= \alpha\psi(\omega)\{e^{i[\omega t+\beta+\varphi(\omega)]} + e^{-i[\omega t+\beta+\varphi(\omega)]}\} \\ &= 2\alpha\psi(\omega)\cos\{\omega t + \beta + \varphi(\omega)\}. \end{aligned}$$

Thus, the process Y_t is also a cosine function with the original period, but with amplitude $2\alpha\psi(\omega)$ and the phase angle $\beta + \varphi(\omega)$. The function $\psi(\omega)$ is called the

[3] We may define the time invariant property as follows. If the output of a filter g applied to X_t is Y_t, then the filter is time invariant if $g(X_{t+h}) = Y_{t+h}$ for all integer h.

gain of the filter, and the function $\varphi(\omega)$ is the *phase angle* (or simply *phase*) of the filter. The rationale for these terms is clear when one observes the effect of the filter on the cosine time series. The transform $2\pi f_a(\omega)$ is called the *frequency response function* or the *transfer function* of the filter. From (4.3.1) we know that the spectrum of a time series Y_t created by linearly filtering a time series X_t is the spectrum of the original series, $f_X(\omega)$, multiplied by the squared gain of the filter. The squared gain is sometimes called the *power transfer function.*

A particularly simple filter is the perfect *delay*. Let $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ be a stationary time series. The time series delayed or lagged by τ periods is defined by

$$Y_t = X_{t-\tau},$$

where the filter is defined by $a_\tau = 1$ and $a_j = 0$, $j \neq \tau$. Hence the frequency response function of the filter is $e^{-i\omega\tau} = \cos \omega\tau - i \sin \omega\tau$, the gain is $(\cos^2 \omega\tau + \sin^2 \omega\tau)^{1/2} = 1$, and the phase angle is $\tan^{-1}(-\sin \omega\tau / \cos \omega\tau) = -\omega\tau$. A cosine wave of frequency ω and corresponding period $2\pi\omega^{-1}$ completes $(2\pi)^{-1}\omega\tau$ cycles during a "time period" of length τ. Thus the phase of such a cosine wave lagged τ periods is shifted by the quantity $(2\pi)^{-1}\omega\tau$.

Theorem 4.3.1 is a very useful theorem for absolutely summable filters. In some situations, a mean square result applicable to a broader range of filters and time series is appropriate.

Theorem 4.3.5. Assume X_t is a zero mean stationary time series with spectral density $f_X(\omega)$. Assume $\{c_j\}_{j=-\infty}^{\infty}$ is a square summable sequence and that

$$\sum_{j=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} c_r c_{j+h} \gamma_X(j + h - r) \tag{4.3.12}$$

converges absolutely. Define

$$Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots.$$

Then

$$\gamma_Y(h) = \int_{-\pi}^{\pi} e^{i\omega h} \left| \sum_{j=-\infty}^{\infty} c_j e^{-i\omega j} \right|^2 f_X(\omega)\, d\omega$$

and

$$f_Y(\omega) = (2\pi)^2 f_X(\omega) f_c(\omega) f_c^*(\omega), \tag{4.3.13}$$

where

$$[f_c(\omega), f_c^*(\omega)] = (2\pi)^{-1} \left[\sum_{j=-\infty}^{\infty} c_j e^{-i\omega j}, \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \right].$$

Proof. Because (4.3.12) converges, we can interchange orders of integration and summation and we have

$$\begin{aligned} E\{Y_t Y_{t+h}\} &= E\left\{ \sum_{j=-\infty}^{\infty} c_j X_{t-j} \sum_{r=-\infty}^{\infty} c_r X_{t+h-r} \right\} \\ &= \sum_{j=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} c_j c_r \int_{-\pi}^{\pi} e^{i\omega(j+h-r)} f_X(\omega)\, d\omega \\ &= \int_{-\pi}^{\pi} \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \sum_{r=-\infty}^{\infty} c_r e^{-i\omega r} e^{i\omega h} f_X(\omega)\, d\omega \\ &= \int_{-\pi}^{\pi} e^{i\omega h} \left| \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \right|^2 f_X(\omega)\, d\omega . \end{aligned}$$

Because of the one-to-one correspondence in mean square between the Fourier coefficients of a square integrable function and the function, $f_Y(\omega)$ is the function specified in (4.3.13). ▲

In Section 2.11, we introduced the long memory processes that satisfy

$$(1 - \mathcal{B})^d Y_t = Z_t \tag{4.3.14}$$

or

$$Y_t = (1 - \mathcal{B})^{-d} Z_t , \tag{4.3.15}$$

where Z_t is a finite autoregressive moving average. By Theorem 3.1.6, there is a function $f_Y(\omega)$ defined as a limit in mean square,

$$f_Y(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_Y(h) e^{-i\omega h} ,$$

because $\gamma_Y(h)$ is square summable. Also,

$$\gamma_Y(h) = \int_{-\pi}^{\pi} f_Y(\omega) e^{i\omega h}\, d\omega .$$

Consider the simple case in which the Z_t of (4.3.14) is e_t, a time series of uncorrelated $(0, \sigma^2)$ random variables. Then, by Theorem 4.3.5,

$$\begin{aligned} f_Y(\omega) &= (1 - e^{-i\omega})^{-d} (1 - e^{i\omega})^{-d} f_e(\omega) \\ &= (2\pi)^{-1} |1 - e^{i\omega}|^{-2d} \sigma^2 \\ &= (2\pi)^{-1} [4 \sin^2(0.5\omega)]^{-d} \sigma^2 . \end{aligned}$$

Observe that $f_Y(\omega)$ is unbounded at $\omega = 0$ when $d > 0$. Because $x^{-1} \sin x \to 1$ as $x \to 0$,

$$f_Y(\omega) \to (2\pi)^{-1} \omega^{-2d} \sigma^2$$

as $\omega \to 0$. Sometimes a long memory process is defined to be a process whose spectral density is approximately equal to a multiple of ω^{-2d} for $-0.5 < d < 0.5$ and ω near zero.

It follows from the spectral density that

$$\gamma_Y(h) = \int_{-\pi}^{\pi} e^{i\omega h} f_Y(\omega)\, d\omega$$
$$= \frac{(-1)^h \Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-h-d)}\sigma^2, \qquad h = 0, 1, 2, \ldots .$$

4.4. VECTOR PROCESSES

The spectral representation of vector time series follows in a straightforward manner from that of scalar time series. We denote the cross covariance of two zero mean stationary time series X_{jt} and X_{mt} by $\gamma_{jm}(h) = \{\Gamma(h)\}_{jm} = E\{X_{jt}X_{m,t+h}\}$ and assume that $\{\gamma_{jm}(h)\}$ is absolutely summable. Then

$$\frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \gamma_{jm}(h)e^{-i\omega h} = f_{jm}(\omega) \tag{4.4.1}$$

is a continuous periodic function of ω, which we call the *cross spectral function* of X_{jt} and X_{mt}. Since $\gamma_{jm}(h)$ may not be symmetric about 0, $f_{jm}(\omega)$ is, in general, a complex valued function. As such, it can be written as

$$f_{jm}(\omega) = c_{jm}(\omega) - iq_{jm}(\omega),$$

where $c_{jm}(\omega)$ and $q_{jm}(\omega)$ are real valued functions of ω. The function $c_{jm}(\omega)$ is called the *coincident spectral density* or simply the *cospectrum*. The function $q_{jm}(\omega)$ is called the *quadrature spectral density*. The function $c_{jm}(\omega)$ is the cosine portion of the transform and is an even function of ω, and $q_{jm}(\omega)$ is the sine portion and is an odd function of ω. Thus we may define these quantities as the transforms

$$c_{jm}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \tfrac{1}{2}[\gamma_{jm}(h) + \gamma_{jm}(-h)]e^{-i\omega h},$$
$$-iq_{jm}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \tfrac{1}{2}[\gamma_{jm}(h) - \gamma_{jm}(-h)]e^{-i\omega h},$$

or, in real terms,

$$c_{jm}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \tfrac{1}{2}[\gamma_{jm}(h) + \gamma_{jm}(-h)]\cos \omega h,$$
$$q_{jm}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \tfrac{1}{2}[\gamma_{jm}(h) - \gamma_{jm}(-h)]\sin \omega h. \tag{4.4.2}$$

By the Fourier integral theorem,

$$\gamma_{jm}(h) = \int_{-\pi}^{\pi} e^{i\omega h} f_{jm}(\omega)\, d\omega .$$

If we let $\mathbf{f}(\omega)$ denote the matrix with typical element $f_{jm}(\omega)$, we have the matrix representations

$$\mathbf{\Gamma}(h) = \int_{-\pi}^{\pi} e^{i\omega h} \mathbf{f}(\omega)\, d\omega , \tag{4.4.3}$$

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} \mathbf{\Gamma}(h) . \tag{4.4.4}$$

For a general stationary time series we can write

$$\mathbf{\Gamma}(h) = \int_{-\pi}^{\pi} e^{i\omega h}\, d\mathbf{F}(\omega) \tag{4.4.5}$$

in complete analogy to (4.1.1a).

Let us investigate some of the properties of the matrix $\mathbf{f}(\omega)$.

Definition 4.4.1. A square complex valued matrix $\mathbf{B}$ is called a *Hermitian matrix* if it is equal to its conjugate transpose, that is,

$$\mathbf{B} = \mathbf{B}^* ,$$

where the jmth element of $\mathbf{B}^*$ is the complex conjugate of the mjth element of $\mathbf{B}$.

Definition 4.4.2. A Hermitian matrix $\mathbf{B}$ is *positive definite* if for any complex vector $\mathbf{w}$ such that $\mathbf{w}^*\mathbf{w} > 0$,

$$\mathbf{w}^*\mathbf{B}\mathbf{w} > 0 ,$$

and it is *positive semidefinite* if

$$\mathbf{w}^*\mathbf{B}\mathbf{w} \geq 0 .$$

Lemma 4.4.1. For stationary vector time series of dimension k satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{jm}(h)| < \infty$$

for $j, m = 1, 2, \ldots, k$, the matrix $\mathbf{f}(\omega)$ is a positive semidefinite Hermitian matrix for all ω in $[-\pi, \pi]$.

Proof. The matrix $\mathbf{f}(\omega)$ is Hermitian, since, for all ω in $[-\pi, \pi]$ and for

$j, m = 1, 2, \ldots, k$,

$$\begin{aligned} f_{jm}^*(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{jm}^*(h) e^{i\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{mj}(-h) e^{i\omega h} \\ &= \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_{mj}(r) e^{-i\omega r} \\ &= f_{mj}(\omega) . \end{aligned}$$

Consider the complex valued time series

$$Z_t = \boldsymbol{\alpha}' \mathbf{X}_t = \sum_{j=1}^{k} \alpha_j X_{jt} ,$$

where $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is a vector of arbitrary complex numbers. The autocovariance function of Z_t is given by

$$\gamma_Z(h) = \sum_{j=1}^{k} \sum_{r=1}^{k} \alpha_j \alpha_r^* \gamma_{jr}(h) .$$

Now $\gamma_Z(h)$ is positive semidefinite, and hence, for any n,

$$\sum_{m=1}^{n} \sum_{q=1}^{n} \gamma_Z(m-q) e^{-i m \omega} e^{i q \omega} \geq 0$$

and

$$\sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \left[\sum_{j=1}^{k} \sum_{r=1}^{k} \alpha_j \alpha_r^* \gamma_{jr}(h) \right] e^{-i\omega h} \geq 0 .$$

Taking the limit as $n \to \infty$, we have

$$\sum_{j=1}^{k} \sum_{r=1}^{k} \alpha_j \alpha_r^* f_{jr}(\omega) \geq 0 ,$$

which establishes that $\mathbf{f}(\omega)$ is positive semidefinite. ▲

It follows immediately from Lemma 4.4.1 that the determinant of any two by two matrix of the form

$$\begin{pmatrix} f_{jj}(\omega) & f_{jm}(\omega) \\ f_{mj}(\omega) & f_{mm}(\omega) \end{pmatrix}$$

is nonnegative. Hence,

$$|f_{jm}(\omega)|^2 = f_{jm}(\omega)f^*_{jm}(\omega) = f_{jm}(\omega)f_{mj}(\omega) \leq f_{jj}(\omega)f_{mm}(\omega)\,. \tag{4.4.6}$$

The quantity

$$\mathcal{K}^2_{jm}(\omega) = \frac{|f_{jm}(\omega)|^2}{f_{jj}(\omega)f_{mm}(\omega)} = \frac{c^2_{jm}(\omega) + q^2_{jm}(\omega)}{f_{jj}(\omega)f_{mm}(\omega)} \tag{4.4.7}$$

is called the *squared coherency function*. The spectral density may be zero at certain frequencies, in which case $\mathcal{K}^2_{jm}(\omega)$ is of the form 0/0. We adopt the convention of assigning zero to the coherency in such situations. The inequality (4.4.6), sometimes written as

$$\mathcal{K}^2_{jm}(\omega) \leq 1\,, \tag{4.4.8}$$

is called the *coherency inequality*.

To further appreciate the properties of $\mathbf{f}(\omega)$, consider the time series

$$\begin{aligned} X_{1t} &= A_1 \cos rt + B_1 \sin rt\,, \\ X_{2t} &= A_2 \cos rt + B_2 \sin rt\,, \end{aligned} \tag{4.4.9}$$

where $r \in (0, \pi)$ and $(A_1, B_1, A_2, B_2)'$ is distributed as a multivariate normal with zero mean and covariance matrix

$$\begin{pmatrix} \sigma_{11} & 0 & \sigma_{13} & \sigma_{14} \\ 0 & \sigma_{11} & -\sigma_{14} & \sigma_{13} \\ \sigma_{13} & -\sigma_{14} & \sigma_{33} & 0 \\ \sigma_{14} & \sigma_{13} & 0 & \sigma_{33} \end{pmatrix}.$$

Then

$$\begin{aligned} \gamma_{11}(h) &= \sigma_{11} \cos rh\,, \\ \gamma_{22}(h) &= \sigma_{33} \cos rh\,, \\ \gamma_{12}(h) &= E\{A_1A_2[\cos rt]\cos r(t+h) + A_1B_2[\cos rt]\sin r(t+h) \\ &\qquad + B_1A_2[\sin rt]\cos r(t+h) + B_1B_2[\sin rt]\sin r(t+h)\} \\ &= \sigma_{13} \cos rh + \sigma_{14} \sin rh\,, \end{aligned}$$

and

$$\gamma_{21}(h) = \gamma_{12}(-h)\,.$$

The matrix analog of the spectral distribution function introduced in Section 4.1

is

$$\mathbf{F}(\omega) = \begin{pmatrix} F_{11}(\omega) & F_{12}(\omega) \\ F_{21}(\omega) & F_{22}(\omega) \end{pmatrix},$$

where $F_{12}(\omega)$ is the cross spectral distribution function. For the example (4.4.9), $F_{11}(\omega)$ is a step function with jumps of height $\frac{1}{2}\sigma_{11}$ at $\pm r$, $F_{22}(\omega)$ is a step function with jumps of height $\frac{1}{2}\sigma_{33}$ at $\pm r$, $F_{12}(\omega)$ is a complex valued function where $\operatorname{Re} F_{12}(\omega)$ is a step function with a jump of height $\frac{1}{2}\sigma_{13}$ at $\pm r$, and $\operatorname{Im} F_{12}(\omega)$ is a step function with a jump of $\frac{1}{2}\sigma_{14}$ at $-r$ and a jump of $-\frac{1}{2}\sigma_{14}$ at r. Since the elements of $\mathbf{F}(\omega)$ are pure jump functions, we have a pure line spectrum. The real portion of the cross line spectrum is one-half the covariance between the coefficients of the cosine portions of the two time series, which is also one-half the covariance between the coefficients of the sine portions. The absolute value of the imaginary portion of the cross line spectrum is one-half the covariance between the coefficient of the cosine portion of the first time series and the coefficient of the sine portion of the second time series. This is one-half the negative of the covariance between the coefficient of the sine of the first time series and the coefficient of the cosine of the second. To consider the point further, we write the time series X_{2t} as a sine wave,

$$X_{2t} = \psi \sin(rt + \varphi),$$

where

$$\psi = (A_2^2 + B_2^2)^{1/2},$$

$$\varphi = \tan^{-1} \frac{A_2}{B_2}.$$

Let X_{3t} be the cosine with the same amplitude and phase,

$$\begin{aligned} X_{3t} &= \psi \cos(rt + \varphi) \\ &= B_2 \cos rt - A_2 \sin rt. \end{aligned}$$

It follows that X_{2t} is uncorrelated with X_{3t}. The covariance between X_{1t} and X_{2t} is

$$\begin{aligned} E\{X_{1t}X_{2t}\} &= E\{(A_1 \cos rt + B_1 \sin rt)(A_2 \cos rt + B_2 \sin rt)\} \\ &= \sigma_{13}(\cos^2 rt + \sin^2 rt) = \sigma_{13}, \end{aligned}$$

and the covariance between X_{1t} and X_{3t} is

$$\begin{aligned} E\{X_{1t}X_{3t}\} &= E\{(A_1 \cos rt + B_1 \sin rt)(B_2 \cos rt - A_2 \sin rt)\} \\ &= \sigma_{14}. \end{aligned}$$

The covariance between X_{1t} and X_{2t} is proportional to the real part of the cross spectrum at r. The covariance between X_{1t} and X_{3t} is proportional to the imaginary

portion of the cross spectrum at r. The fact that X_{3t} is X_{2t} "shifted" by a phase angle of $\pi/2$ explains why the real portion of the cross spectrum is called the cospectrum and the imaginary part is called the quadrature spectrum. The squared coherency is the multiple correlation between X_{1t} and the pair X_{2t}, X_{3t}, that is,

$$\mathcal{K}^2_{12}(r) = \frac{\sigma^2_{13} + \sigma^2_{14}}{\sigma_{33}\sigma_{11}}.$$

We now introduce some cross spectral quantities useful in the analysis of input–output systems. Let the bivariate time series $(X_t, Y_t)'$ have absolutely summable covariance function. We may then write the cross spectral function as

$$f_{XY}(\omega) = A_{XY}(\omega)e^{i\varphi_{XY}(\omega)}, \tag{4.4.10}$$

where

$$\varphi_{XY}(\omega) = \tan^{-1}\left[\frac{-q_{XY}(\omega)}{c_{XY}(\omega)}\right],$$

$$A_{XY}(\omega) = [c^2_{XY}(\omega) + q^2_{XY}(\omega)]^{1/2}.$$

We use the convention of setting $\varphi_{XY}(\omega) = 0$ when both $c_{XY}(\omega)$ and $q_{XY}(\omega)$ are zero. The quantity $\varphi_{XY}(\omega)$ is called the *phase spectrum*, and $A_{XY}(\omega)$ is called the *cross amplitude spectrum*. The *gain* of Y_t over X_t is defined by

$$\psi_{XY}(\omega) = \frac{A_{XY}(\omega)}{f_{XX}(\omega)} \tag{4.4.11}$$

for those ω where $f_{XX}(\omega) > 0$.

Let us assume that an absolutely summable linear filter is applied to an input time series X_t with zero mean, absolutely summable covariance function, and everywhere positive spectral density to yield an output time series

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}. \tag{4.4.12}$$

The cross covariance function is

$$\gamma_{XY}(h) = E\{X_t Y_{t+h}\} = \sum_{j=-\infty}^{\infty} a_j \gamma_{XX}(h-j), \qquad h = 0, \pm 1, \pm 2, \ldots,$$

and, by Corollary 3.4.1.1, the cross spectral function is

$$f_{XY}(\omega) = 2\pi f_a(\omega) f_{XX}(\omega), \tag{4.4.13}$$

where

$$f_a(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j},$$

$$f_{XX}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{XX}(h) e^{-i\omega h}.$$

It follows that $f_{XY}(\omega)/f_{XX}(\omega)$ is the transfer function of the filter.

Recall that the phase $\varphi(\omega)$ and gain $\psi(\omega)$ of the filter $\{a_j\}$ were defined by

$$2\pi f_a(\omega) = \psi(\omega) e^{i\varphi(\omega)}.$$

Since $f_{XY}(\omega)$ is the product of $2\pi f_a(\omega)$ and $f_{XX}(\omega)$, where $f_{XX}(\omega)$ is a real function of ω, it follows that the phase spectrum is the same as the phase of the filter. The cross amplitude spectrum is the product of the spectrum of the input time series and the gain of the filter. That is,

$$A_{XY}(\omega) = |f_{XY}(\omega)| = f_{XX}(\omega)|2\pi f_a(\omega)| \tag{4.4.14}$$

and the gain of Y_t over X_t is simply the gain of the filter.

By Theorem 4.3.1, the spectral density of Y_t is

$$f_{YY}(\omega) = (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{XX}(\omega),$$

and it follows that the squared coherency, for $|f_a(\omega)| > 0$, is

$$\mathcal{K}_{XY}^2(\omega) = \frac{|f_{XY}(\omega)|^2}{f_{YY}(\omega) f_{XX}(\omega)} = \frac{|2\pi f_a(\omega) f_{XX}(\omega)|^2}{(2\pi)^2 |f_a(\omega)|^2 |f_{XX}(\omega)|^2} = 1. \tag{4.4.15}$$

This is an interesting result, because it shows that the squared coherency between an output time series created by the application of a linear filter and the original time series is one at all frequencies. The addition of an error (noise) time series to the output will produce a coherency less than one in a linear system. For example, consider the bivariate time series

$$\begin{aligned} X_{1t} &= \beta X_{1,t-1} + e_{1t} \\ X_{2t} &= \alpha_1 X_{1t} + \alpha_2 X_{1,t-1} + e_{2t}, \end{aligned} \tag{4.4.16}$$

where $|\beta| < 1$ and $\{(e_{1t}, e_{2t})'\}$ is a sequence of uncorrelated vector random variables with $E\{e_{1t}^2\} = \sigma_{11}$, $E\{e_{2t}^2\} = \sigma_{22}$, and $E\{e_{1t} e_{2,t+h}\} = 0$ for all t and h.

The input time series X_{1t} is a first order autoregressive time series, and the output X_{2t} is a linear function of X_{1t}, $X_{1,t-1}$, and e_{2t}. The autocovariance and cross

covariance functions are therefore easily computed:

$$\gamma_{11}(h) = E\{X_{1,t}X_{1,t+h}\} = \frac{\beta^{|h|}}{1-\beta^2}\sigma_{11},$$

$$\gamma_{22}(h) = \begin{cases} (\alpha_1^2 + \alpha_2^2)\gamma_{11}(0) + 2\alpha_1\alpha_2\gamma_{11}(1) + \sigma_{22}, & h = 0, \\ (\alpha_1^2 + \alpha_2^2)\gamma_{11}(h) + \alpha_1\alpha_2[\gamma_{11}(h+1) + \gamma_{11}(h-1)], & h \neq 0, \end{cases}$$

$$\gamma_{12}(h) = E\{X_{1,t}X_{2,t+h}\} = \alpha_1\gamma_{11}(h) + \alpha_2\gamma_{11}(h-1).$$

The cross covariance and cross correlation functions for parameters $\alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\gamma_{11}(0) = 1.0$, $\sigma_{22} = 0.5$, and $\beta = 0.8$ are displayed in Table 4.4.1.

The special matrix has elements

$$f_{11}(\omega) = g_{11}(\omega)(1 - \beta e^{-i\omega})^{-1}(1 - \beta e^{i\omega})^{-1},$$

$$f_{22}(\omega) = (\alpha_1 + \alpha_2 e^{-i\omega})(\alpha_1 + \alpha_2 e^{i\omega})f_{11}(\omega) + g_{22}(\omega),$$

$$f_{12}(\omega) = (\alpha_1 + \alpha_2 e^{-i\omega})f_{11}(\omega),$$

$$f_{21}(\omega) = f_{12}^*(\omega) = (\alpha_1 + \alpha_2 e^{i\omega})f_{11}(\omega),$$

where $g_{11}(\omega)$ is the spectral density of $e_{1,t}$, and $g_{22}(\omega)$ is the spectral density of $e_{2,t}$; that is,

$$g_{11}(\omega) = (2\pi)^{-1}\sigma_{11},$$

$$g_{22}(\omega) = (2\pi)^{-1}\sigma_{22}.$$

Table 4.4.1. Autocovariance and Cross Covariance Functions of the Time Series Defined in (4.4.16)

h	$\gamma_{11}(h)$	$\gamma_{22}(h)$	$\gamma_{12}(h)$	$\rho_{12}(h)$
−6	0.262	0.596	0.341	0.213
−5	0.328	0.745	0.426	0.267
−4	0.410	0.932	0.532	0.333
−3	0.512	1.165	0.666	0.417
−2	0.640	1.456	0.832	0.521
−1	0.800	1.820	1.040	0.651
0	1.000	2.550	1.300	0.814
1	0.800	1.820	1.400	0.877
2	0.640	1.456	1.120	0.701
3	0.512	1.165	0.896	0.561
4	0.410	0.932	0.717	0.449
5	0.328	0.745	0.573	0.359
6	0.262	0.596	0.459	0.287

The cospectrum and quadrature spectral density are

$$c_{12}(\omega) = (\alpha_1 + \alpha_2 \cos \omega) f_{11}(\omega),$$
$$q_{12}(\omega) = \alpha_2 [\sin \omega] f_{11}(\omega),$$

the phase spectrum and cross amplitude spectrum are

$$\varphi_{12}(\omega) = \tan^{-1}\left(\frac{-\alpha_2 \sin \omega}{\alpha_1 + \alpha_2 \cos \omega}\right),$$
$$A_{12}(\omega) = f_{11}(\omega)[\alpha_1^2 + 2\alpha_1\alpha_2 \cos \omega + \alpha_2^2]^{1/2},$$

and the squared coherency is

$$\mathcal{K}_{12}^2(\omega) = \frac{|2\pi f_a(\omega)|^2 |f_{11}(\omega)|^2}{f_{11}(\omega)[|2\pi f_a(\omega)|^2 f_{11}(\omega) + g_{22}(\omega)]}$$
$$= \frac{1}{1 + \eta(\omega)},$$

where

$$\eta(\omega) = \frac{g_{22}(\omega)}{|2\pi f_a(\omega)|^2 f_{11}(\omega)}.$$

Since $g_{22}(\omega)$ is positive at all frequencies, the squared coherency is strictly less than one at all frequencies. The quantity $\eta(\omega)$ is sometimes called the *noise to signal ratio* in physical applications. Although the presence of noise reduces the squared coherency, the ratio of the cross spectrum to the spectrum of the input series still gives the transfer function of the filter. This is because the time series e_{2t} is uncorrelated with the input X_{1t}. The quantity

$$f_{22}(\omega)[1 - \mathcal{K}_{12}^2(\omega)] = f_{22}(\omega) - f_{21}(\omega) f_{11}^{-1}(\omega) f_{12}(\omega)$$

is sometimes called the *error spectral density* or *error spectrum*. We see that for a model such as (4.4.16),

$$g_{22}(\omega) = f_{22}(\omega) - f_{21}(\omega) f_{11}^{-1}(\omega) f_{12}(\omega).$$

For the example of Table 4.4.1 the elements of the spectral matrix are

$$f_{11}(\omega) = \frac{0.36}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})},$$
$$f_{22}(\omega) = \frac{(0.5 + 1.0e^{-i\omega})(0.5 + 1.0e^{i\omega})(0.36)}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})} + \frac{0.5}{2\pi}$$
$$= \frac{1.23(1 - 0.17e^{-i\omega})(1 - 0.17e^{i\omega})}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})},$$

$$f_{12}(\omega) = \frac{(0.5 + 1.0e^{-i\omega})(0.36)}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})}.$$

It is interesting that the spectral density of X_{2t} is that of an autoregressive moving average (1, 1) process. Note that if X_{2t} had been defined as the simple sum of X_{1t} and e_{2t}, the spectral density would also have been that of an autoregressive moving average (1, 1) process but with different parameters.

The noise to signal ratio is

$$\eta(\omega) = \frac{g_{22}(\omega)}{|2\pi f_\alpha(\omega)|^2 f_{11}(\omega)} = \frac{(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})}{(0.72)(0.5 + e^{-i\omega})(0.5 + e^{i\omega})},$$

and the squared coherency is

$$\mathcal{K}_{12}^2(\omega) = \frac{0.45 + 0.36\cos\omega}{1.27 - 0.44\cos\omega}.$$

Let us consider the example a bit further. The input time series X_{1t} is autocorrelated. Let us filter both the input and output with the same filter, choosing the filter so that the input time series becomes a sequence of uncorrelated random variables. Thus we define

$$X_{3t} = X_{1t} - \beta X_{1,t-1} = e_{1t},$$
$$X_{4t} = X_{2t} - \beta X_{2,t-1},$$

and it follows that

$$\begin{aligned} X_{4t} &= \alpha_1(X_{1t} - \beta X_{1,t-1}) + \alpha_2(X_{1,t-1} - \beta X_{1,t-2}) + e_{2t} - \beta e_{2,t-1} \\ &= \alpha_1 e_{1t} + \alpha_2 e_{1,t-1} + e_{2t} - \beta e_{2,t-1}. \end{aligned}$$

The cross covariance function of X_{3t} and X_{4t} then has a particularly simple form:

$$\gamma_{34}(h) = \begin{cases} \alpha_1\sigma_{11}, & h = 0, \\ \alpha_2\sigma_{11}, & h = 1, \\ 0 & \text{otherwise}. \end{cases}$$

By transforming the input series to white noise, the cross covariance function is transformed into the coefficients of the function (or linear filter) that defines X_{4t} as a function of X_{3t}. The spectral matrix of X_{3t} and X_{4t} has elements

$$f_{33}(\omega) = \frac{\sigma_{11}}{2\pi},$$

$$\begin{aligned} f_{44}(\omega) &= (\alpha_1 + \alpha_2 e^{-i\omega})(\alpha_1 + \alpha_2 e^{i\omega})\frac{\sigma_{11}}{2\pi} \\ &\quad + (1 - \beta e^{-i\omega})(1 - \beta e^{i\omega})\frac{\sigma_{22}}{2\pi}, \end{aligned}$$

$$f_{34}(\omega) = (\alpha_1 + \alpha_2 e^{-i\omega})\frac{\sigma_{11}}{2\pi}.$$

The reader may verify that

$$\varphi_{34}(\omega) = \tan^{-1}\left(\frac{-\alpha_2 \sin \omega}{\alpha_1 + \alpha_2 \cos \omega}\right)$$

and

$$A_{34}(\omega) = \frac{\sigma_{11}}{2\pi}(\alpha_1^2 + 2\alpha_1\alpha_2 \cos \omega + \alpha_2^2)^{1/2}.$$

As we expected from (4.4.11) and (4.4.13), the phase spectrum is unchanged by the transformation, since we transformed both input and output with the same filter. The filter changed the input spectrum, and as a result, that portion of the cross amplitude spectrum associated with the input was altered. Also, the error spectral density

$$f_{44}(\omega) - f_{43}(\omega)f_{33}^{-1}(\omega)f_{34}(\omega) = (1 - \beta e^{-i\omega})(1 - \beta e^{i\omega})\frac{\sigma_{22}}{2\pi}$$

is that of a moving average with parameter $-\beta$. The reader may verify that $\mathcal{K}_{34}^2(\omega)$ is the same as $\mathcal{K}_{12}^2(\omega)$.

We have introduced and illustrated the ideas of phase spectrum and amplitude spectrum using the input–output model. Naturally, these quantities can be computed, in much the same manner as we compute the correlation for a bivariate normal distribution, without recourse to this model. The same is true of squared coherency and error spectrum, which have immediate generalizations to higher dimensions.

The effect of the application of a matrix filter to a vector time series is summarized in Theorem 4.4.1.

Theorem 4.4.1. Let $\mathbf{X}_t$ be a real k-dimensional stationary time series with absolutely summable covariance matrix and let $\{\mathbf{A}_j\}_{j=-\infty}^{\infty}$ be a sequence of real $k \times k$ matrices that is absolutely summable. Then the spectral density of

$$\mathbf{Y}_t = \sum_{j=-\infty}^{\infty} \mathbf{A}_j' \mathbf{X}_{t-j}$$

is

$$\mathbf{f}_{\mathbf{Y}}(\omega) = (2\pi)^2 \mathbf{f}_{\mathbf{A}}^*(\omega)\mathbf{f}_{\mathbf{X}}(\omega)\mathbf{f}_{\mathbf{A}}(\omega),$$

where

$$\mathbf{f}_{\mathbf{X}}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \mathbf{\Gamma}_{\mathbf{X}}(h)e^{-i\omega h},$$

$$\mathbf{f}_{\mathbf{A}}(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \mathbf{A}_j e^{-i\omega j},$$

and $\mathbf{f}_{\mathbf{A}}^{*}(\omega)$ is the conjugate transpose of $\mathbf{f}_{\mathbf{A}}(\omega)$,

$$\mathbf{f}_{\mathbf{A}}^{*}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbf{A}_j' e^{i\omega j} .$$

Proof. The proof parallels that of Theorem 4.3.1. We have

$$\begin{aligned} \mathbf{f}_{\mathbf{Y}}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{A}_j' \boldsymbol{\Gamma}_{\mathbf{X}}(h-s+j)\mathbf{A}_s e^{-i\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{A}_j' e^{i\omega j} \boldsymbol{\Gamma}_{\mathbf{X}}(h-s+j) e^{-i\omega(h-s+j)} \mathbf{A}_s e^{-i\omega s} \end{aligned}$$

and the result follows. ▲

As in the scalar case, the spectral matrix of autoregressive and moving average time series follows immediately from Theorem 4.4.1.

Corollary 4.4.1.1. The spectral density of the moving average process

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

where $\{\mathbf{e}_t\}$ is a sequence of uncorrelated $(\mathbf{0}, \boldsymbol{\Sigma})$ vector random variables and the sequence $\{\mathbf{B}_j\}$ is absolutely summable, is

$$\begin{aligned} \mathbf{f}_{\mathbf{X}}(\omega) &= \frac{1}{2\pi} \left(\sum_{j=-\infty}^{\infty} \mathbf{B}_j e^{ij\omega} \right) \boldsymbol{\Sigma} \left(\sum_{j=-\infty}^{\infty} \mathbf{B}_j' e^{-ij\omega} \right) \\ &= (2\pi)^2 \mathbf{f}_{\mathbf{B}'}^{*}(\omega) \boldsymbol{\Sigma} \mathbf{f}_{\mathbf{B}'}(\omega) . \end{aligned}$$

Corollary 4.4.1.2. Define the vector autoregressive process $\mathbf{X}_t$ by

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

where $\mathbf{A}_0 = \mathbf{I}$, $\mathbf{A}_p \neq \mathbf{0}$, the $\mathbf{e}_t$ are uncorrelated $(\mathbf{0}, \boldsymbol{\Sigma})$ vector random variables, and the roots of

$$\left| \sum_{j=0}^{p} \mathbf{A}_j m^{p-j} \right| = 0$$

are less than one in absolute value. Then the spectral density of $\mathbf{X}_t$ is

$$\mathbf{f}_{\mathbf{X}}(\omega) = 2\pi [\mathbf{f}_{\mathbf{A}'}^{*}(\omega)]^{-1} \boldsymbol{\Sigma} [\mathbf{f}_{\mathbf{A}'}(\omega)]^{-1} ,$$

where $\mathbf{f}_{\mathbf{A}'}(\omega)$ is defined in Theorem 4.4.1.

4.5. MEASUREMENT ERROR – SIGNAL DETECTION

In any statistical model the manner in which the "errors" enter the model is very important. In models of the "input–output" or "independent-variable–dependent-variable" form, measurement error in the output or dependent variable is relatively easy to handle. On the other hand, measurement error in the input or independent variable typically introduces additional complexity into the analysis. In the simple regression model with normal errors the presence of normal measurement error in the independent variable requires additional information, such as knowledge of the variance of the measurement error, before one can estimate the slope of the regression line.

In time series analysis the distinction between independent variable and dependent variable may be blurred. For example, in predicting a future observation in the realization, the past observations play the role of independent variables, while the future observation plays the role of (unknown) dependent variable. As a result, considerable care is required in specifying and treating measurement errors in such analyses. One important problem where errors of observation play a central role is the estimation of the values of an underlying time series that is observed with measurement error.

To introduce the problem, let $\{X_t\colon t \in (0, \pm 1, \pm 2, \ldots)\}$ be a stationary time series with zero mean. Because of measurement error we do not observe X_t directly. Instead, we observe

$$Y_t = X_t + u_t\,, \tag{4.5.1}$$

where $\{u_t\colon t \in (0, \pm 1, \pm 2, \ldots)\}$ is a time series with u_t independent of X_j for all t, j. The u_t are the measurement errors or the noise in the system. A problem of *signal measurement* or signal detection is the construction of an estimator of X_t given a realization on Y_t (or a portion of the realization). We assume the covariance functions of X_t and u_t are known.

We first consider the problem in the time domain and restrict ourselves to finding a linear filter that minimizes the mean square error of our estimator of X_t. Thus we desire weights $\{a_j\colon j = -L, -(L-1), \ldots, M-1, M\}$, for $L \geqslant 0$ and $M \geqslant 0$ fixed, such that

$$E\left\{\left(X_t - \sum_{j=-L}^{M} a_j Y_{t-j}\right)^2\right\} = E\left\{\left[X_t - \sum_{j=-L}^{M} a_j (X_{t-j} + u_{t-j})\right]^2\right\} \tag{4.5.2}$$

is a minimum. We set the derivatives of (4.5.2) with respect to the a_j equal to zero and obtain the system of equations

$$\sum_{j=-L}^{M} a_j[\gamma_{XX}(j-r) + \gamma_{uu}(j-r)] = \gamma_{XX}(r)\,,$$
$$r = -L,\ -(L-1), \ldots, M-1, M\,. \tag{4.5.3}$$

For modest choices of L and M this system of liner equations is easily solved for the a_j. The mean square error of the estimator can then be obtained from (4.5.2).

To obtain the set of weights one would use if one had available the entire realization, we investigate the problem in the frequency domain. This permits us to establish the bound on the mean square error and to use this bound to evaluate the performance of a filter with a finite number of weights. We assume that u_t and X_t have continuous spectral densities with bounded derivatives. The spectral density of Y_t and the cross spectral density follow immediately from the independence:

$$f_{YY}(\omega) = f_{XX}(\omega) + f_{uu}(\omega),$$
$$f_{XY}(\omega) = f_{XX}(\omega).$$

We assume that $f_{YY}(\omega)$ is strictly positive and that $f_{YY}^{-1}(\omega)$ has bounded first derivative.

We shall search in the class of absolutely summable filters $\{a_j\}$ for that filter such that our estimator of X_t,

$$\hat{X}_t = \sum_{j=-\infty}^{\infty} a_j Y_{t-j},$$

will have minimum mean square error. We write

$$f_{\hat{X}\hat{X}}(\omega) = (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{YY}(\omega),$$
$$f_{X\hat{X}}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{X\hat{X}}(h) e^{-ih\omega}$$
$$= 2\pi f_a(\omega) f_{XY}(\omega).$$

Now the mean square error is

$$E\{(X_t - \hat{X}_t)^2\} = E\left\{\left(X_t - \sum_{j=-\infty}^{\infty} a_j Y_{t-j}\right)^2\right\}, \tag{4.5.4}$$

and $W_t = X_t - \Sigma_{j=-\infty}^{\infty} a_j Y_{t-j}$ is a stationary time series with a spectral density, say $f_{WW}(\omega)$. Hence, the variance of W_t is

$$\gamma_{WW}(0) = E\{W_t^2\} = \int_{-\pi}^{\pi} f_{WW}(\omega)\, d\omega$$
$$= \int_{-\pi}^{\pi} \{f_{XX}(\omega) - 2\pi[f_a(\omega) f_{XY}(\omega) + f_a^*(\omega) f_{YX}(\omega)]$$
$$+ (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{YY}(\omega)\}\, d\omega. \tag{4.5.5}$$

We have converted the problem of finding the a_j that minimize (4.5.4) to the problem of finding the $f_a(\omega)$ that minimizes (4.5.5). This problem retains the same form and appearance as the classical regression problem with $f_a(\omega)$ playing the role

of the vector of coefficients. Therefore, we take as our candidate solution

$$f_a(\omega) = \frac{1}{2\pi} \frac{f_{YX}(\omega)}{f_{YY}(\omega)}, \tag{4.5.6}$$

which gives

$$\begin{aligned} \gamma_{WW}(0) &= \int_{-\pi}^{\pi} \left[f_{XX}(\omega) - \frac{f_{XY}(\omega) f^*_{XY}(\omega)}{f_{YY}(\omega)} \right] d\omega \\ &= \int_{-\pi}^{\pi} f_{XX}(\omega)[1 - \mathcal{K}^2_{XY}(\omega)]\, d\omega . \end{aligned} \tag{4.5.7}$$

The weights a_j are given by the inverse transform of $f_a(\omega)$,

$$a_j = \int_{-\pi}^{\pi} f_a(\omega) e^{i\omega j}\, d\omega , \qquad j = 0, \pm 1, \pm 2, \ldots .$$

To demonstrate that these weights yield the minimum value for the mean square error, we consider an alternative to (4.5.6):

$$f_b(\omega) = \frac{f_{YX}(\omega)}{2\pi f_{YY}(\omega)} + \frac{f_d(\omega)}{2\pi f_{YY}(\omega)},$$

where $f_b(\omega)$ must be in the class of functions such that the integral defining the mean square error exists. The mean square error is then

$$\begin{aligned} \int_{-\pi}^{\pi} &\{ f_{XX}(\omega) - 2\pi [f_b(\omega) f^*_{YX}(\omega) + f^*_b(\omega) f_{YX}(\omega)] \\ &\quad + (2\pi)^2 f_b(\omega) f^*_b(\omega) f_{YY}(\omega) \}\, d\omega \\ &= \int_{-\pi}^{\pi} \left[f_{XX}(\omega) - \frac{|f_{XY}(\omega)|^2}{f_{YY}(\omega)} + \frac{|f_d(\omega)|^2}{f_{YY}(\omega)} \right] d\omega . \end{aligned} \tag{4.5.8}$$

Since $|f_d(\omega)|^2 / f_{YY}(\omega)$ is nonnegative, we conclude that the $f_a(\omega)$ of (4.5.6) yields the minimum value for (4.5.5).

Note that $2\pi f_a(\omega) = f^{-1}_{YY}(\omega) f_{YX}(\omega)$ is a real valued symmetric function of ω because $f_{YX}(\omega) = f_{XX}(\omega)$ is a real valued symmetric function of ω. Therefore, the weights a_j are also symmetric about zero.

We summarize in Theorem 4.5.1.

Theorem 4.5.1. Let $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ and $\{u_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ be independent zero mean stationary time series, and define $Y_t = X_t + u_t$. Let $f_{YY}(\omega)$, $f^{-1}_{YY}(\omega)$, and $f_{XX}(\omega)$ be continuous with bounded first derivatives. Then the best linear filter for extracting X_t from a realization of Y_t is given by

$$a_j = \int_{-\pi}^{\pi} f_a(\omega) e^{i\omega j}\, d\omega ,$$

where

$$f_a(\omega) = \frac{f_{YX}(\omega)}{2\pi f_{YY}(\omega)}.$$

Furthermore, the mean square error of $\Sigma_{j=-\infty}^{\infty} a_j Y_{t-j}$ as an estimator for X_t is

$$\int_{-\pi}^{\pi} \{f_{XX}(\omega) - [f_{YY}(\omega)]^{-1}|f_{XY}(\omega)|^2\}\, d\omega = \int_{-\pi}^{\pi} f_{XX}(\omega)[1 - \mathcal{K}_{XY}^2(\omega)]\, d\omega .$$

Proof. Since the derivatives of $f_{YY}^{-1}(\omega)$ and $f_{YX}(\omega)$ are bounded, the derivative of $f_{YY}^{-1}(\omega)f_{YX}(\omega)$ is square integrable. Therefore, by Theorem 3.1.8, the Fourier coefficients of $f_{YY}^{-1}(\omega)f_{YX}(\omega)$, the a_j, are absolutely summable, and by Theorem 2.2.1, $\Sigma_{j=-\infty}^{\infty} a_j Y_{t-j}$ is a well-defined random variable. That $\{a_j\}$ is the best linear filter follows from (4.5.8), and the mean square error of the filter follows from (4.5.7). ▲

Example 4.5.1. To illustrate the ideas of this section, we use some data on the sediment suspended in the water of the Des Moines River at Boone, Iowa. A portion of the data obtained by daily sampling of the water during 1973 are displayed in Table 4.5.1. The data are the logarithm of the parts per million of suspended sediment. Since the laboratory determinations are made on a small sample of water collected from the river, the readings can be represented as

$$Y_t = X_t + u_t ,$$

where Y_t is the recorded value, X_t is the "true" average sediment in the river water, and u_t is the measurement error introduced by sampling and laboratory determination. Assume that X_t can be represented as a first order autoregressive process

$$X_t - 5.28 = 0.81(X_{t-1} - 5.28) + e_t ,$$

where the e_t are independent (0, 0.172) random variables. Assume further that u_t is a sequence of independent (0, 0.053) random variables independent of X_j for all t, j.

To construct a filter $\{a_{-2}, a_{-1}, a_0, a_1, a_2\}$ that will best estimate X_t using the observations $\{Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}\}$, we solve the system of equations

$$\begin{pmatrix} 0.553 & 0.405 & 0.328 & 0.266 & 0.215 \\ 0.405 & 0.553 & 0.405 & 0.328 & 0.266 \\ 0.328 & 0.405 & 0.553 & 0.405 & 0.328 \\ 0.266 & 0.328 & 0.405 & 0.553 & 0.405 \\ 0.215 & 0.266 & 0.328 & 0.405 & 0.553 \end{pmatrix} \begin{pmatrix} a_{-2} \\ a_{-1} \\ a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.328 \\ 0.405 \\ 0.500 \\ 0.405 \\ 0.328 \end{pmatrix}$$

to obtain

$$\mathbf{a} = \{a_{-2}, a_{-1}, a_0, a_1, a_2\} = \{0.023, 0.120, 0.702, 0.120, 0.023\} .$$

Table 4.5.1. Logarithm of Sediment Suspended in Des Moines River, Boone, Iowa, 1973

Daily Observations	Smoothed Observations	
	Two-Sided Filter	One-Sided Filter
5.44	—	—
5.38	—	—
5.43	5.40	—
5.22	5.26	5.26
5.28	5.27	5.28
5.21	5.22	5.22
5.23	5.25	5.23
5.33	5.37	5.31
5.58	5.63	5.53
6.18	6.08	6.04
6.16	6.14	6.10
6.07	6.13	6.04
6.56	6.38	6.42
5.93	5.96	5.99
5.70	5.69	5.74
5.36	5.39	5.42
5.17	5.24	5.22
5.35	5.36	5.33
5.51	5.51	5.47
5.80	5.67	5.72
5.29	5.36	5.37
5.28	5.29	5.30
5.27	—	5.27
5.17	—	5.19

SOURCE: U.S. Department of Interior Geological Survey—Water Resources Division, Sediment Concentration Notes, Des Moines River, Boone, Iowa.

The mean square error of the filtered time series $5.28 + \sum_{j=-2}^{2} a_j(Y_{t-j} - 5.28)$ as an estimator of X_t is $0.500 - \mathbf{a}[0.328, 0.405, 0.500, 0.405, 0.328]' = 0.0372$. The data transformed by this filter are displayed in the second column of Table 4.5.1. Note that the filtered data are "smoother" in that the variance of changes from one period to the next is smaller for the filtered data than for the original data.

We now obtain a one-sided filter that can be used to estimate the most recent value of X_t using only the most recent and the four preceding values of Y_t. The estimator of X_t is given by

$$5.28 + b_0(Y_t - 5.28) + b_1(Y_{t-1} - 5.28) + b_2(Y_{t-2} - 5.28) + b_3(Y_{t-3} - 5.28) + b_4(Y_{t-4} - 5.28),$$

where

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} 0.553 & 0.405 & 0.328 & 0.266 & 0.215 \\ 0.405 & 0.553 & 0.405 & 0.328 & 0.266 \\ 0.328 & 0.405 & 0.553 & 0.405 & 0.328 \\ 0.266 & 0.328 & 0.405 & 0.553 & 0.405 \\ 0.215 & 0.266 & 0.328 & 0.405 & 0.553 \end{pmatrix}^{-1} \begin{pmatrix} 0.500 \\ 0.405 \\ 0.328 \\ 0.266 \\ 0.215 \end{pmatrix} = \begin{pmatrix} 0.790 \\ 0.134 \\ 0.023 \\ 0.005 \\ 0.000 \end{pmatrix}.$$

This one-sided estimator of X_t has a mean square error of 0.0419. The filtered data using this filter are given in the last column of Table 4.5.1.

To obtain the minimum value for the mean square error of the two-sided filter, we evaluated (4.5.7), where

$$f_{XX}(\omega) = \frac{0.172}{2\pi(1-0.81e^{-i\omega})(1-0.81e^{i\omega})},$$

$$f_{YY}(\omega) = \frac{0.172}{2\pi(1-0.81e^{-i\omega})(1-0.81e^{i\omega})} + \frac{0.053}{2\pi}$$

$$= \frac{0.2524(1-0.170e^{i\omega})(1-0.170e^{-i\omega})}{2\pi(1-0.81e^{-i\omega})(1-0.81e^{i\omega})},$$

obtaining

$$\gamma_{WW}(0) = \int_{-\pi}^{\pi} f_{XX}(\omega)\,d\omega - \int_{-\pi}^{\pi} \frac{(0.172)^2\,d\omega}{2\pi(0.2524)|1-0.170e^{-i\omega}|^2|1-0.81e^{-i\omega}|^2}$$

$$= 0.500 - 0.463 = 0.037.$$

The infinite sequence of weights is given by the inverse transform of

$$\frac{0.172}{2\pi(0.2524)(1-0.17e^{-i\omega})(1-0.17e^{i\omega})},$$

which yields

$$(\ldots, 0.1193, 0.7017, 0.1193, 0.0203, 0.0034, \ldots).$$

While it is possible to use spectral methods to evaluate the minimum mean square error of a one-sided filter [see, for example, Yaglom (1962, p. 97)], we examine the problem in a slightly different manner. Since Y_t is an autoregressive moving average (1, 1) process, the methods of Section 2.9 can be used to obtain a one-period-ahead predictor $\hat{Y}_t(Y_{t-1}, Y_{t-2}, \ldots)$ based on an infinite past. As $Y_t = X_t + u_t$, where u_t is a sequence of independent random variables, the best predictor of X_t based on $Y_{t-1}, Y_{t-2}, \ldots$ must be the same as the best predictor of Y_t based on $Y_{t-1}, Y_{t-2}, \ldots$. Furthermore, given the predictor, the partial correlation between any Y_{t-j}, $j > 0$, and X_t is zero. Therefore, to obtain the best filter for X_t using $Y_t, Y_{t-1}, \ldots$, we find the optimal linear combination of $\hat{Y}_t(Y_{t-1}, Y_{t-2}, \ldots)$ and Y_t.

Denote the linear combination by $c_0 Y_t + c_1 \hat{Y}_t(Y_{t-1}, Y_{t-2}, \ldots)$, where

$$\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 0.5530 & 0.3006 \\ 0.3006 & 0.3006 \end{pmatrix}^{-1} \begin{pmatrix} 0.5000 \\ 0.3006 \end{pmatrix} = \begin{pmatrix} 0.7900 \\ 0.2100 \end{pmatrix},$$

the matrix

$$\begin{pmatrix} 0.5530 & 0.3006 \\ 0.3006 & 0.3006 \end{pmatrix}$$

is the covariance matrix of $[Y_t, \hat{Y}_t(Y_{t-1}, Y_{t-2}, \ldots)]$, and $(0.5000, 0.3006)$ is the vector of covariances between $[Y_t, \hat{Y}_t(Y_{t-1}, Y_{t-2}, \ldots)]$ and X_t. It follows that the minimum mean square error for a one-sided predictor is 0.0419. ▲▲

4.6. STATE SPACE MODELS AND KALMAN FILTERING

We begin our discussion under the assumption that the univariate time series X_t is a stationary, zero mean, first order autoregressive process. We write

$$X_t = \alpha X_{t-1} + e_t, \qquad t = 1, 2, \ldots, \tag{4.6.1}$$

where e_t are independent identically distributed $(0, \sigma_e^2)$ random variables, denoted by $e_t \sim \text{II}(0, \sigma_e^2)$. Assume we are unable to observe X_t directly. Instead, we observe Y_t, where

$$Y_t = X_t + u_t, \qquad t = 1, 2, \ldots, \tag{4.6.2}$$

and u_t is the measurement error. We assume $u_t \sim \text{II}(0, \sigma_u^2)$ and that u_t is independent of e_j for all t and j.

The model (4.6.1)–(4.6.2) is a special case of the *state space* representation of time series. Equation (4.6.1) is called the *state equation* or the *transition equation*, and X_t is called the *state* of the system at time t. Equation (4.6.2) is called the *measurement equation* or the *observation equation*. The model (4.6.1)–(4.6.2) was introduced in the Des Moines River example of Section 4.5. In that example it is very natural to think of the unknown level of the river as the true "state" of nature.

In Section 4.5, we constructed linear filters to estimate the values of the time series X_t that is observed subject to measurement error. The filters were designed to minimize the mean square error of the estimation error for a particular set of observations. In many applications, the observation set is composed of all previous observations plus the current observation. In some engineering applications, it is important to have an efficient method of computing the current estimated value requiring as little storage of information as possible. Such computational methods have been developed by Kalman (1960, 1963) and others. The methods are often

called *Kalman filters*. In this section, we study the state space model and the Kalman filter.

Let the model (4.6.1) and (4.6.2) hold, and assume that an initial estimator for X_0, denoted by $\hat{X}_0$, is available. Let

$$\hat{X}_0 = X_0 + v_0 , \tag{4.6.3}$$

where v_0 is a $(0, \sigma_{v0}^2)$ random variable, independent of (u_t, e_t), $t = 1, 2, \ldots$. The initial estimator $\hat{X}_0$ and parameters σ_u^2, σ_e^2, σ_{v0}^2, and α are assumed known. Equation (4.6.3) is called the *initial condition equation* (or *starting equation*). A possible choice for $\hat{X}_0$ for the model (4.6.1) is $\hat{X}_0 = 0$. With $\hat{X}_0 = 0$, we have $\sigma_{v0}^2 = \gamma_X(0) = (1 - \alpha^2)^{-1}\sigma_e^2$.

At time $t = 1$, we have the observation $(Y_1, \hat{X}_0)$ and knowledge of σ_u^2, σ_e^2, and α to use in constructing an estimator (predictor) of X_1. On the basis of the specification, we have

$$\begin{aligned} Y_1 &= X_1 + u_1 , \\ \alpha\hat{X}_0 &= X_1 + w_1 , \end{aligned} \tag{4.6.4}$$

where $w_1 = \alpha v_0 - e_1$. The system of equations (4.6.4) can be written in matrix form as

$$\mathbf{Z}_1 = \mathbf{J}X_1 + \boldsymbol{\epsilon}_1 ,$$

where $\mathbf{Z}_1 = (Y_1, \alpha\hat{X}_0)'$, $\mathbf{J} = (1, 1)'$, and $\boldsymbol{\epsilon}_1 = (u_1, w_1)'$. Because u_1, e_1, and v_0 are mutually uncorrelated, the covariance matrix of $\boldsymbol{\epsilon}_1$ is $\operatorname{diag}(\sigma_u^2, \sigma_{w1}^2)$, where $\sigma_{w1}^2 = \sigma_e^2 + \alpha^2\sigma_{v0}^2$. Therefore, it is natural to construct an estimator of X_1 as the weighted average of Y_1 and $\alpha\hat{X}_0$, where the weights are proportional to the inverses of the variances of u_1 and w_1. Thus,

$$\hat{X}_1 = (\sigma_u^{-2} + \sigma_{w1}^{-2})^{-1}(\sigma_u^{-2}Y_1 + \sigma_{w1}^{-2}\alpha\hat{X}_0) . \tag{4.6.5}$$

The estimator (4.6.5) is constructed by analogy to linear regression theory. In the problem formulation (4.6.4), the information about the unknown random value, X_1, is contained in the second equation of system (4.6.4). See Exercise 4.19. The same approach can be used to construct succeeding estimators.

Let the error in $\hat{X}_1$ as an estimator of X_1 be v_1, where

$$v_1 = \hat{X}_1 - X_1 = (\sigma_u^{-2} + \sigma_{w1}^{-2})^{-1}(\sigma_u^{-2}u_1 + \sigma_{w1}^{-2}w_1) . \tag{4.6.6}$$

The variance of v_1 is

$$\sigma_{v1}^2 = (\sigma_u^{-2} + \sigma_{w1}^{-2})^{-1} . \tag{4.6.7}$$

At time $t = 2$, it is desired to estimate X_2 using Y_2, Y_1, and $\hat{X}_0$. Now $\alpha\hat{X}_1$ is the best

predictor of X_2 constructed from the data $(Y_1, \hat{X}_0)$. Therefore, we need only combine the information in $\alpha\hat{X}_1$ with that in Y_2 to obtain the best estimator of X_2. Using (4.6.1)–(4.6.2) and the identity

$$\alpha\hat{X}_1 = \alpha X_1 + \alpha(\hat{X}_1 - X_1) = \alpha X_1 + \alpha v_1 ,$$

we have the system of two equations containing X_2,

$$\begin{aligned} Y_2 &= X_2 + u_2 , \\ \alpha\hat{X}_1 &= X_2 + w_2 , \end{aligned} \tag{4.6.8}$$

where $w_2 = \alpha v_1 - e_2$. Because the vectors (u_t, e_t) are uncorrelated, we have

$$E\{(u_2, e_2, v_1)'(u_2, e_2, v_1)\} = \operatorname{diag}(\sigma_u^2, \sigma_e^2, \sigma_{v1}^2)$$

and

$$\mathrm{V}\{(u_2, w_2)\} = \operatorname{diag}(\sigma_u^2, \sigma_e^2 + \alpha^2\sigma_{v1}^2) .$$

It follows that the best estimator of X_2, given $(\hat{X}_0, Y_1, Y_2)$, is

$$\hat{X}_2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}(\sigma_u^{-2}Y_2 + \sigma_{w2}^{-2}\alpha\hat{X}_1) , \tag{4.6.9}$$

where $\sigma_{w2}^2 = \sigma_e^2 + \alpha^2\sigma_{v1}^2$. Letting the error $\hat{X}_2 - X_2$ be

$$v_2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}(\sigma_u^{-2}u_2 + \sigma_{w2}^{-2}w_2) ,$$

the variance of v_2 is $\sigma_{v2}^2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}$.

Because $\hat{X}_{t-1}$ contains all of the information about X_t available from the previous observations, the estimator of X_t for general t is

$$\hat{X}_t = (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1}(\sigma_u^{-2}Y_t + \sigma_{wt}^{-2}\alpha\hat{X}_{t-1}) , \tag{4.6.10}$$

where $w_t = \alpha v_{t-1} - e_t$, $v_t = \hat{X}_t - X_t$,

$$\sigma_{wt}^2 = \sigma_e^2 + \alpha^2\sigma_{v,t-1}^2 ,$$

and

$$\sigma_{vt}^2 = (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1} .$$

Equation (4.6.10) can be rearranged to yield

$$\hat{X}_t = \alpha\hat{X}_{t-1} + (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1}\sigma_u^{-2}(Y_t - \alpha\hat{X}_{t-1}) \tag{4.6.11}$$

or

$$\hat{X}_t = \alpha\hat{X}_{t-1} + \sigma_{wt}^2(\sigma_u^2 + \sigma_{wt}^2)^{-1}(Y_t - \alpha\hat{X}_{t-1}) . \tag{4.6.12}$$

The first term on the right of (4.6.12) is the estimator of X_t based upon $\hat{X}_{t-1}$. The second term is the negative of an estimator of the error made in predicting X_t with $\alpha\hat{X}_{t-1}$. We can give a direct justification of (4.6.12) as follows. Because $u_t - w_t = Y_t - \alpha\hat{X}_{t-1}$ is a linear combination of Y_t and $\hat{X}_{t-1}$, the linear estimator of X_t based on

$$(u_t - w_t, \hat{X}_{t-1}) = (Y_t - \alpha\hat{X}_{t-1}, \hat{X}_{t-1})$$

is the same as the linear estimator based on $(Y_t, \hat{X}_{t-1})$. Now $u_t - w_t$ is uncorrelated with $\hat{X}_{t-1}$. It follows, by the properties of regression (or of Hilbert spaces), that the best linear estimator of X_t based on $(\hat{X}_{t-1}, Y_t)$ is the best linear estimator of X_t based on $\hat{X}_{t-1}$, plus the best linear estimator of $X_t - \alpha\hat{X}_{t-1}$ based on $Y_t - \alpha\hat{X}_{t-1}$. Hence, the best estimator of X_t is given by (4.6.12). See Exercise 4.25.

An equation equivalent to (4.6.12) is

$$\hat{X}_t = Y_t - \sigma_u^2(\sigma_u^2 + \sigma_{wt}^2)^{-1}(Y_t - \alpha\hat{X}_{t-1}) . \tag{4.6.13}$$

The form (4.6.12) appears in the original signal extraction literature, and the form (4.6.13) appears in the random model prediction literature. In the form (4.6.13), an estimator of the error u_t is subtracted from Y_t to obtain the estimator of X_t.

The variance of v_t satisfies the equation

$$\sigma_{vt}^2 = \sigma_{wt}^2 - (\sigma_{wt}^2)^2(\sigma_u^2 + \sigma_{wt}^2)^{-1} . \tag{4.6.14}$$

Hence, the data required at time $t+1$ to construct the estimator of X_{t+1} and to construct the variance of the estimation error are the elements of $(Y_{t+1}, \hat{X}_t, \sigma_{vt}^2)$. The equations (4.6.12) and (4.6.14) are sometimes called the *updating equations* of the Kalman filter.

The results for the simple model (4.6.1)–(4.6.2) generalize to p-dimensional vector time series $\mathbf{X}_t$ and to the situation wherein the observed vector $\mathbf{Y}_t$ is the sum of a known linear function of $\mathbf{X}_t$ and measurement error. Let

$$\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{e}_t , \qquad t = 1, 2, \ldots , \tag{4.6.15}$$

$$\mathbf{Y}_t = \mathbf{H}_t\mathbf{X}_t + \mathbf{u}_t , \qquad t = 1, 2, \ldots , \tag{4.6.16}$$

$$\hat{\mathbf{X}}_0 = \mathbf{X}_0 + \mathbf{v}_0 , \tag{4.6.17}$$

where $\mathbf{X}_t$ is a p-dimensional column vector, $\mathbf{Y}_t$ is an r-dimensional column vector, $\{\mathbf{H}_t\}$ is a sequence of known $r \times p$ matrices, $\mathbf{A}_t$ is a sequence of known $p \times p$ matrices, $\hat{\mathbf{X}}_0$ is a known initial vector, and $\{(\mathbf{u}_t', \mathbf{e}_t')\}$ is a sequence of uncorrelated,

zero mean, vector random variables with known covariance matrix

$$E\{(\mathbf{u}_t', \mathbf{e}_t')'(\mathbf{u}_t', \mathbf{e}_t')\} = \text{block diag}(\boldsymbol{\Sigma}_{uutt}, \boldsymbol{\Sigma}_{eett}) .$$

Equation (4.6.15) is the state equation, equation (4.6.16) is the measurement equation, and equation (4.6.17) is the initial equation. As before, we assume $\mathbf{v}_0$ of the initial equation to be uncorrelated with $\mathbf{u}_t$ and $\mathbf{e}_t$, $t = 1, 2, \ldots$. Considerable generalization of the model is obtained by permitting the variances to be functions of t and by the inclusion of the matrices $\mathbf{H}_t$ in the measurement equation. Many different forms of state space representations appear in the literature.

For the vector model, the system of equations analogous to (4.6.8) is

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{H}_t\mathbf{X}_t + \mathbf{u}_t , \\ \mathbf{A}_t\hat{\mathbf{X}}_{t-1} &= \mathbf{X}_t + \mathbf{w}_t \end{aligned} \tag{4.6.18}$$

for $t = 1, 2, \ldots$, where $\mathbf{w}_t = \mathbf{A}_t\mathbf{v}_{t-1} - \mathbf{e}_t$ and $\mathbf{v}_t = \hat{\mathbf{X}}_t - \mathbf{X}_t$. If we assume that $\boldsymbol{\Sigma}_{uutt}$ is nonsingular, the best estimator of $\mathbf{X}_t$ in (4.6.18) is

$$\hat{\mathbf{X}}_t = (\mathbf{H}_t'\boldsymbol{\Sigma}_{uutt}^{-1}\mathbf{H}_t + \boldsymbol{\Sigma}_{wwtt}^{-1})^{-1}(\mathbf{H}_t'\boldsymbol{\Sigma}_{uutt}^{-1}\mathbf{Y}_t + \boldsymbol{\Sigma}_{wwtt}^{-1}\mathbf{A}_t\hat{\mathbf{X}}_{t-1}) , \tag{4.6.19}$$

where

$$\boldsymbol{\Sigma}_{wwtt} = E\{\mathbf{w}_t\mathbf{w}_t'\} = \boldsymbol{\Sigma}_{eett} + \mathbf{A}_t\boldsymbol{\Sigma}_{vv,t-1,t-1}\mathbf{A}_t' , \tag{4.6.20}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{vvtt} &= E\{(\hat{\mathbf{X}}_t - \mathbf{X}_t)(\hat{\mathbf{X}}_t - \mathbf{X}_t)'\} = (\mathbf{H}_t'\boldsymbol{\Sigma}_{uutt}^{-1}\mathbf{H}_t + \boldsymbol{\Sigma}_{wwtt}^{-1})^{-1} \\ &= \boldsymbol{\Sigma}_{wwtt} - \boldsymbol{\Sigma}_{wwtt}\mathbf{H}_t'\mathbf{D}_t^{-1}\mathbf{H}_t\boldsymbol{\Sigma}_{wwtt} , \end{aligned} \tag{4.6.21}$$

$$\mathbf{D}_t = \boldsymbol{\Sigma}_{uutt} + \mathbf{H}_t\boldsymbol{\Sigma}_{wwtt}\mathbf{H}_t' . \tag{4.6.22}$$

The estimator (4.6.19) can also be written as

$$\hat{\mathbf{X}}_t = \mathbf{A}_t\hat{\mathbf{X}}_{t-1} + \boldsymbol{\Sigma}_{wwtt}\mathbf{H}_t'\mathbf{D}_t^{-1}(\mathbf{Y}_t - \mathbf{H}_t\mathbf{A}_t\hat{\mathbf{X}}_{t-1}) , \tag{4.6.23}$$

where $\mathbf{D}_t$ is the covariance matrix of $\mathbf{Y}_t - \mathbf{H}_t\mathbf{A}_t\hat{\mathbf{X}}_{t-1}$ and $\boldsymbol{\Sigma}_{wwtt}\mathbf{H}_t'$ is the covariance between $\mathbf{w}_t$ and $\mathbf{Y}_t - \mathbf{H}_t\mathbf{A}_t\hat{\mathbf{X}}_{t-1}$. Therefore, the estimator (4.6.23) is the difference between the unbiased estimator $\mathbf{A}_t\hat{\mathbf{X}}_{t-1}$ of $\mathbf{X}_t$ and an unbiased estimator of the error $\mathbf{w}_t$ in $\mathbf{A}_t\hat{\mathbf{X}}_{t-1}$.

Equation (4.6.20), the second equation of (4.6.21), and equation (4.6.23) form a set of equations that can be used to construct $\hat{\mathbf{X}}_t$ given $\mathbf{Y}_t$, $\hat{\mathbf{X}}_{t-1}$, and $\boldsymbol{\Sigma}_{vv,t-1,t-1}$. If this system of equations is used for updating, only the matrix $\mathbf{D}_t$ is required to be nonsingular. The matrix $\mathbf{D}_t$ should always be nonsingular because there is little reason for the subject matter specialist to consider singular observation vectors $\mathbf{Y}_t$.

The vector of updating equations analogous to (4.6.13) is

$$\hat{\mathbf{X}}_t = \mathbf{Y}_t - \boldsymbol{\Sigma}_{uutt}\mathbf{D}_t^{-1}(\mathbf{Y}_t - \mathbf{H}_t\mathbf{A}_t\hat{\mathbf{X}}_{t-1})$$

and an alternative expression for $\boldsymbol{\Sigma}_{vvtt}$ is

$$\boldsymbol{\Sigma}_{vvtt} = \boldsymbol{\Sigma}_{uutt} - \boldsymbol{\Sigma}_{uutt}\mathbf{D}_t^{-1}\boldsymbol{\Sigma}_{uutt}.$$

Example 4.6.1. In this example, we construct the Kalman filter for the Des Moines River example of Section 4.5. The model for the data is

$$\begin{aligned} X_t - 5.28 &= 0.81(X_{t-1} - 5.28) + e_t, \\ Y_t &= X_t + u_t, \end{aligned} \tag{4.6.24}$$

where $e_t \sim \text{II}(0, 0.172)$ and $u_t \sim \text{II}(0, 0.053)$. The sediment time series has a nonzero mean, so the difference $X_t - 5.28$ plays the role of X_t of (4.6.1).

We begin with the first observation of Table 4.6.1, which we call Y_1. If X_t is a stationary process, as we assume for the Des Moines River example, we can use the population mean to initiate the filter. If we use the population mean as an estimator of X_0, the variance of the estimation error is 0.500, which is the variance of the X_t process. The system of equations (4.6.4) becomes

$$\begin{aligned} Y_1 - 5.28 = 0.16 &= (X_1 - 5.28) + u_1, \\ 0 &= (X_1 - 5.28) + w_1, \end{aligned}$$

Table 4.6.1. Estimates of Sediment Constructed with Kalman Filter

t	Y_t	$\hat{X}_t$	σ_{vt}^2	σ_{wt}^2
1	5.44	5.42467	0.04792	0.50000
2	5.38	5.38355	0.04205	0.20344
3	5.43	5.41613	0.04188	0.19959
4	5.22	5.25574	0.04187	0.19948
5	5.28	5.27588	0.04187	0.19947
6	5.21	5.22399	0.04187	0.19947
7	5.23	5.23097	0.04187	0.19947
8	5.33	5.31117	0.04187	0.19947
9	5.58	5.52232	0.04187	0.19947
10	6.18	6.03227	0.04187	0.19947
11	6.16	6.10318	0.04187	0.19947
12	6.07	6.04413	0.04187	0.19947
13	6.56	6.42123	0.04187	0.19947
14	5.93	5.98760	0.04187	0.19947
15	5.70	5.73215	0.04187	0.19947

where $\mathbf{V}\{(u_1, w_1)\} = \text{diag}(0.053, 0.500)$. Therefore, by (4.6.10),

$$\hat{X}_1 = 5.28 + (18.868 + 2.000)^{-1}[18.868(Y_1 - 5.28) + 0]$$
$$= 5.28 + 0.145 = 5.425.$$

The reader may verify that the coefficient for $Y_1 - 5.28$ is the covariance between X_1 and Y_1 divided by the variance of Y_1. The variance of the error in the estimator of X_1 is $\sigma_{v1}^2 = 0.0479$.

The estimate of X_2 is

$$\hat{X}_2 = 5.28 + [(0.053)^{-1} + (0.2034)^{-1}]^{-1}$$
$$\times [(0.053)^{-1}(Y_2 - 5.28) + (0.2034)^{-1}(0.81)(\hat{X}_1 - 5.28)]$$
$$= 0.2075 + 0.7933Y_2 + 0.1674\hat{X}_1$$
$$= 5.3836,$$

where $\sigma_{w2}^2 = \sigma_e^2 + \alpha^2\sigma_{v1}^2 = 0.2034$. The variance of $\hat{X}_2 - X_2$ is

$$\sigma_{v2}^2 = 0.2034 - (0.2034)^2(0.053 + 0.2034)^{-1}$$
$$= 0.0420.$$

The estimate for X_3 is

$$\hat{X}_3 = 5.28 + \frac{(0.053)^{-1}(Y_3 - 5.28) + (4.0581)(\hat{X}_2 - 5.28)}{(0.053)^{-1} + (0.1996)^{-1}}$$
$$= 5.4161,$$

and the variance of the estimation error is

$$\sigma_{v3}^2 = 0.1996 - (0.1996)^2(0.053 + 0.1996)^{-1} = 0.0419.$$

The estimates and variances for the remaining observations are given in Table 4.6.1. Note that the variance of the estimation error is approaching 0.0419. This limiting variance, denoted by σ_v^2, was derived in Example 4.5.1 as the mean square error based on an infinite past. The variance of w_t stabilizes at

$$\sigma_w^2 = \sigma_e^2 + \alpha^2\sigma_v^2 = 0.172 + (0.81)^2(0.0419)$$
$$= 0.1995.$$

It follows that equation (4.6.12) stabilizes at

$$\hat{X}_t = 5.28 + 0.81(\hat{X}_{t-1} - 5.28) + 0.7901[Y_t - 5.28 - 0.81(\hat{X}_{t-1} - 5.28)],$$

where $0.7901 = (\sigma_u^2 + \sigma_w^2)^{-1}\sigma_w^2$ and $\alpha = 0.81$. ▲▲

Example 4.6.2. We investigate estimation of X_t for the Des Moines River

example under the assumption that the mean of the process is unknown. We retain the assumption that the other parameters of the model are known. We write the model as

$$\begin{aligned} Z_t &= \alpha Z_{t-1} + e_{t1}, \\ \mu_t &= \mu, \\ Y_t &= \mu + Z_t + u_t, \end{aligned} \tag{4.6.25}$$

where μ is the mean of the process. The first two equations of (4.6.25) are the state equations and the last equation is the measurement equation. In terms of the model (4.6.15)–(4.6.17), $\sigma_{uutt} = \sigma_{uu} = 0.053$,

$$\begin{aligned} \mathbf{A}_t = \mathbf{A} &= \operatorname{diag}(\alpha, 1) = \operatorname{diag}(0.81, 1), \\ \mathbf{\Sigma}_{eett} = \mathbf{\Sigma}_{ee} &= \operatorname{diag}(0.172, 0), \end{aligned}$$

$\mathbf{X}_t' = (Z_t, \mu)$, $\mathbf{e}_t' = (e_{t1}, 0)$, and $\mathbf{H}_t = (1, 1)$. Under the model, each Y_t is unbiased for μ, and the variance of $Y_t - \mu$ is $\sigma_{ZZ} + \sigma_{uu}$.

To initiate the filter, we use the knowledge that Z_t is a random variable with mean zero and variance $\sigma_{ZZ} = \sigma_Z^2$. Letting Y_0 be the first observation of Table 4.6.2, we can form the system of equations

$$0 = Z_0 + v_0, \tag{4.6.26}$$

$$Y_0 = \mu + Z_0 + u_0. \tag{4.6.27}$$

We are treating μ as a fixed unknown parameter, so we do not have an initial equation of the type (4.6.26) for μ. The first real observation furnishes the first information about μ. From the system (4.6.26)–(4.6.27), we obtain the estimator

$$\hat{\mathbf{X}}_0' = (\hat{Z}_0, \hat{\mu}_0) = (0, Y_0)$$

with covariance matrix

$$\mathbf{\Sigma}_{vv00} = \begin{pmatrix} \sigma_{ZZ} & -\sigma_{ZZ} \\ -\sigma_{ZZ} & \sigma_{ZZ} + \sigma_{uu} \end{pmatrix},$$

where $\sigma_{ZZ} = 0.500$ and $\sigma_{uu} = 0.053$.

Using $\hat{\mathbf{X}}_0' = (0, 5.44)$ and equation (4.6.19), we have

$$\begin{aligned} \begin{pmatrix} \hat{Z}_1 \\ \hat{\mu}_1 \end{pmatrix} &= \begin{pmatrix} 0.46953 & -0.45252 \\ -0.45252 & 0.47901 \end{pmatrix} \begin{pmatrix} 18.868Y_1 + 19.5884 \\ 18.868Y_1 + 24.1832 \end{pmatrix} \\ &= (-0.0193, 5.4100)', \end{aligned}$$

and from (4.6.20),

$$\mathbf{\Sigma}_{ww11} = \begin{pmatrix} 0.172 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.328 & -0.405 \\ -0.405 & 0.553 \end{pmatrix} = \begin{pmatrix} 0.500 & -0.405 \\ -0.405 & 0.553 \end{pmatrix}.$$

If we use (4.6.23),

$$\hat{\mathbf{X}}_1 = \begin{pmatrix} 0.81 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 5.44 \end{pmatrix} + \begin{pmatrix} 0.500 & -0.405 \\ -0.405 & 0.553 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}(0.1430)^{-1}(5.38 - 5.44)$$
$$= (-0.0193, 5.4100)',$$

where

$$D_1 = 0.053 + (1, 1)\boldsymbol{\Sigma}_{ww11}(1, 1)' = 0.29605.$$

It follows that

$$\boldsymbol{\Sigma}_{vv11} = \begin{pmatrix} 0.46953 & -0.45252 \\ -0.45252 & 0.47901 \end{pmatrix}.$$

The estimate of the sediment at $t = 1$ is

$$\hat{S}_1 = \hat{Z}_1 + \hat{\mu}_1 = -0.0193 + 5.4100 = 5.3907.$$

The variance of the error in the estimated sediment at time one is

$$(1, 1)\hat{\boldsymbol{\Sigma}}_{vv11}(1, 1)' = 0.04351.$$

The estimates for $t = 2$ are

$$\begin{pmatrix} \hat{Z}_2 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} 0.43387 & -0.41230 \\ -0.41230 & 0.43367 \end{pmatrix}\begin{pmatrix} 123.116 \\ 129.558 \end{pmatrix} = \begin{bmatrix} -0.0004 \\ 5.4247 \end{bmatrix},$$

where

$$\boldsymbol{\Sigma}_{ww22} = \begin{pmatrix} 0.4801 & -0.3665 \\ -0.3665 & 0.4790 \end{pmatrix}.$$

The estimates of $S_t = Z_t + \mu$, based on $Y_t, Y_{t-1}, \ldots, Y_0$, are given in Table 4.6.2. Note that the estimation of the mean contributes modestly to the variance of $\hat{S}_t - S_t$. The variance of $\hat{\mu}_t$ is declining approximately at the rate t^{-1}. While the variance of the estimator of $S_t = \mu + Z_t$ will eventually approach 0.0419, the approach is slowed by the estimation error in $\hat{\mu}_t$. ▲▲

Example 4.6.3. In the previous examples, we applied the Kalman filter to a stationary time series. Stationarity made it relatively easy to find starting values for the filter. In this example, we consider filtering for a time series in which the autoregressive part has a unit root. Let the model be

$$Z_t = \begin{cases} 0, & t = 0, \\ Z_{t-1} + e_t, & t = 1, 2, \ldots, \end{cases} \tag{4.6.28}$$

$$Y_t = \theta + Z_t + u_t, \qquad t = 0, 1, \ldots, \tag{4.6.29}$$

Table 4.6.2. Estimates of Sediment Constructed with Kalman Filter, Unknown Mean

t	Y_t	$\hat{Z}_t + \hat{\mu}_t = \hat{S}_t$	$\hat{\mu}_t$	$V\{\hat{S}_t - S_t\}$	$V\{\hat{\mu}_t - \mu\}$
0	5.44	5.44000	5.44000	0.05300	0.55300
1	5.38	5.39074	5.41001	0.04351	0.47901
2	5.43	5.42324	5.42436	0.04293	0.43367
3	5.22	5.25915	5.35070	0.04280	0.39757
4	5.28	5.27933	5.35185	0.04272	0.36722
5	5.21	5.22621	5.32613	0.04266	0.34121
6	5.23	5.23298	5.32174	0.04261	0.31864
7	5.33	5.31422	5.34347	0.04256	0.29887
8	5.58	5.52856	5.40987	0.04252	0.28141
9	6.18	6.04632	5.57235	0.04249	0.26588
10	6.16	6.11947	5.61890	0.04246	0.25198
11	6.07	6.06090	5.62881	0.04243	0.23945
12	6.56	6.44377	5.74903	0.04240	0.22811
13	5.93	6.00652	5.67363	0.04238	0.21780
14	5.70	5.74886	5.62767	0.04236	0.20838

where $(e_t, u_t)' \sim \text{NI}[0, \text{diag}(\sigma_e^2, \sigma_u^2)]$. We treat the true part of Y_0, denoted by θ, as a fixed unknown constant to be estimated. The Y-data of Table 4.6.3 were generated by the model with $(\sigma_e^2, \sigma_u^2) = (0.25, 0.16)$. In terms of the model (4.6.15)–(4.6.17), $\sigma_{uu|t} = \sigma_{uu} = 0.16$, $\mathbf{A}_t = \mathbf{A} = \mathbf{I}$,

$$\mathbf{X}_t' = (Z_t, \theta), \qquad \mathbf{e}_t' = (e_t, 0),$$

$$\mathbf{H}_t = \mathbf{H} = (1, 1), \quad \text{and} \quad \boldsymbol{\Sigma}_{ee} = \text{diag}(0.25, 0).$$

Table 4.6.3. Kalman Filter Applied to a Unit Root Process

t	Y_t	$\hat{y}_t = \hat{\theta}_t + \hat{Z}_t$	$\hat{\theta}_t$	$V\{\hat{y}_t - y_t\}$	$V\{\hat{\theta}_t\}$	$C\{\hat{\theta}_t, \hat{y}_t - y_t\}$
0	2.30797	2.30797	2.30797	0.16000	0.16000	0.16000
1	2.54141	2.47588	2.37350	0.11509	0.11509	0.04491
2	3.08044	2.89622	2.42521	0.11125	0.11125	0.01369
3	1.35846	1.83049	2.38483	0.11089	0.11089	0.00420
4	1.55019	1.63629	2.38257	0.11085	0.11085	0.00129
5	2.34068	2.12430	2.38432	0.11085	0.11085	0.00040
6	1.33786	1.57945	2.38372	0.11085	0.11085	0.00012
7	0.98497	1.16759	2.38358	0.11085	0.11085	0.00004
8	1.17314	1.17143	2.38358	0.11085	0.11085	0.00001
9	0.65385	0.81285	2.38357	0.11085	0.11085	0.00000
10	0.35140	0.49315	2.38357	0.11085	0.11085	0.00000
11	0.47546	0.48090	2.38357	0.11085	0.11085	0.00000
12	−0.56643	−0.24470	2.38356	0.11085	0.11085	0.00000
13	0.04359	−0.04497	2.38356	0.11085	0.11085	0.00000
14	−0.25374	−0.18961	2.38356	0.11085	0.11085	0.00000

The model specifies $Z_0 = 0$ and $Y_0 = \theta + u_0$. Therefore, we initiate the filter with

$$\hat{\mathbf{X}}_0' = (\hat{Z}_0, \hat{\theta}_0) = (0, Y_0) = (0, 2.308)$$

and

$$\mathbf{\Sigma}_{vv00} = \text{diag}(0, 0.16).$$

Then

$$\begin{aligned}\mathbf{\Sigma}_{ww11} &= \begin{pmatrix} 0.25 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \text{diag}(0.25, 0.16), \\ D_1 &= 0.16 + (1, 1)[\text{diag}(0.25, 0.16)](1, 1)' = 0.57,\end{aligned}$$

and

$$\begin{aligned}\begin{pmatrix} \hat{Z}_1 \\ \hat{\theta}_1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 2.308 \end{pmatrix} + \begin{pmatrix} 0.4386 \\ 0.2807 \end{pmatrix}(Y_1 - \mathbf{H}\hat{\mathbf{X}}_0) \\ &= (0.1024, 2.3735)'.\end{aligned}$$

The estimate of $\theta + Z_1$ is $\hat{y}_1 = \hat{\theta}_1 + \hat{Z}_1 = 2.4759$.

The remaining estimates and the variances of the estimators are given in Table 4.6.3. There are several interesting aspects of the results. First, the variance of the estimator of θ stabilizes rather quickly. Observations after the first five add very little information about the true value at time zero. Associated with the stabilization of the variance of $\hat{\theta}_t$ is the fact that the estimate of $\theta + Z_t$ at time t depends very little on values of Y_τ for $t - \tau < 5$. Also, the variance of the estimator of θ at time t is equal to the variance of the estimator of $\theta + Z_t$. This is because the estimator of θ, expressed as a function of $(Y_0, Y_1, \ldots, Y_t)$, is the mirror image of the estimator of $\theta + Z_t$.

After a few initial observations, the estimation equation stabilizes. For this problem, the limiting updating equation for $y_t = \theta + Z_t$ can be written as

$$\hat{y}_t = \hat{y}_{t-1} + c(Y_t - \hat{y}_{t-1}), \tag{4.6.30}$$

where

$$\begin{aligned} c &= (\sigma_u^2 + \sigma_w^2)^{-1}\sigma_w^2, \\ \sigma_w^2 &= 0.5\sigma_{ee11}[1 + (1 + 4\sigma_{ee11}^{-1}\sigma_u^2)^{1/2}], \\ \sigma_v^2 &= \sigma_w^2 - (\sigma_u^2 + \sigma_w^2)^{-1}\sigma_w^4. \end{aligned}$$

The expression for σ_w^2 was obtained from the expression following (4.6.10) by setting $\sigma_w^2 = \sigma_{wt}^2$ and $\sigma_{vt}^2 = \sigma_{v,t-1}^2$.

The $\hat{y}_t$ of (4.6.30) is a convex combination of $\hat{y}_{t-1}$ and Y_t. Hence, the sum of the weights on the Y_{t-j}, $j = 0, 1, \ldots, t$, that define $\hat{y}_t$ is one. This means that estimates constructed with this model maintain the level of the original series. If the true Y_t contains a positive time trend, the predictor (4.6.30) will have a negative bias. ▲▲

In our discussion to this point, we have used information on the Y-process through Y_t to construct an estimator for $\mathbf{X}_t$. It follows from (4.6.15) that predictors of future values can be constructed from the estimator of $\mathbf{X}_t$. For example, the predictor of $\mathbf{X}_{t+1}$ constructed with data through $\mathbf{Y}_t$ is

$$\hat{\mathbf{X}}_{t+1|t} = \mathbf{A}_{t+1}\hat{\mathbf{X}}_t \tag{4.6.31}$$

and the variance of the prediction error $\mathbf{V}\{\hat{\mathbf{X}}_{t+1|t} - \mathbf{X}_{t+1}\}$ is

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{\Sigma}_{ww,t+1,t+1} = \mathbf{A}_{t+1}\boldsymbol{\Sigma}_{vvtt}\mathbf{A}'_{t+1} + \boldsymbol{\Sigma}_{ee,t+1,t+1}\,. \tag{4.6.32}$$

The formula (4.6.31) can be applied recursively to obtain predictions for any number of periods. Thus,

$$\hat{\mathbf{X}}_{t+l|t} = \mathbf{A}_{t+l-1}\hat{\mathbf{X}}_{t+l-1|t}\,,$$

and the variance of the prediction error is

$$\boldsymbol{\Sigma}_{t+l|t} = \mathbf{A}_{t+l}\boldsymbol{\Sigma}_{t+l-1|t}\mathbf{A}'_{t+l} + \boldsymbol{\Sigma}_{ee,t+l,t+l}\,.$$

The prediction formulas can also be used in constructing estimators of $\mathbf{X}_t$ when data are missing. Assume that $(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{r-1}, \mathbf{Y}_{r+1})$ is available, that the Kalman filter has been initiated prior to time $r-1$, and that the objective is to estimate $\mathbf{X}_{r+1}$. At time r, the best estimator of $\mathbf{X}_r$ is

$$\hat{\mathbf{X}}_r = \hat{\mathbf{X}}_{r|r-1} = \mathbf{A}_r\hat{\mathbf{X}}_{r-1}\,,$$

and $\mathbf{V}\{\mathbf{X}_r - \mathbf{A}_r\hat{\mathbf{X}}_{r-1}\} = \boldsymbol{\Sigma}_{wwrr}$. Because there is no Y-information at time r, we have $\boldsymbol{\Sigma}_{wwrr} = \boldsymbol{\Sigma}_{vvrr}$. That is, the error in the estimator of $\mathbf{X}_r$ constructed with $\hat{\mathbf{X}}_{r-1}$ is the final estimation error. At time $r+1$, when $\mathbf{Y}_{r+1}$ is available, the best estimator of $\mathbf{X}_{r+1}$ is

$$\hat{\mathbf{X}}_{r+1} = \mathbf{A}_{r+1}\hat{\mathbf{X}}_r + \boldsymbol{\Sigma}_{ww,r+1,r+1}\mathbf{H}'_{r+1}\mathbf{D}^{-1}_{r+1}(\mathbf{Y}_{r+1} - \mathbf{H}_{r+1}\mathbf{A}_{r+1}\hat{\mathbf{X}}_r)\,,$$

where $\boldsymbol{\Sigma}_{ww,r+1,r+1}$, $\mathbf{D}_{r+1}$, and $\boldsymbol{\Sigma}_{vv,r+1,r+1}$ are given by (4.6.20), (4.6.22), and (4.6.21), respectively.

Example 4.6.4. We use the Kalman filter and the data of Example 4.6.1 to construct estimators in the presence of missing data. We assume that the mechanism that causes data to be missing is independent of the (Y_t, X_t) process.

Table 4.6.4 contains the data of Table 4.6.1 with observations 7, 11, and 12

Table 4.6.4. Estimates of Sediment Constructed with Kalman Filter

t	Y_t	$\hat{X}_t$	σ^2_{vt}	σ^2_{wt}
1	5.44	5.42467	0.04792	0.50000
2	5.38	5.38355	0.04205	0.20344
3	5.43	5.41613	0.04188	0.19959
4	5.22	5.25574	0.04187	0.19948
5	5.28	5.27588	0.04187	0.19947
6	5.21	5.22399	0.04187	0.19947
7		5.23463	0.19947	0.19947
8	5.33	5.31708	0.04511	0.30287
9	5.58	5.52380	0.04197	0.20160
10	6.18	6.03256	0.04188	0.19954
11		5.88957	0.19948	0.19948
12		5.77375	0.30288	0.30288
13	6.56	6.44992	0.04637	0.37072
14	5.93	5.99176	0.04200	0.20242
15	5.70	5.73285	0.04188	0.19956

missing. Recall that the model of Example 4.6.1 is

$$X_t - 5.28 = 0.81(X_{t-1} - 5.28) + e_t,$$
$$Y_t = X_t + u_t,$$

where $(e_t, u_t)' \sim \text{II}[\mathbf{0}, \text{diag}(0.172, 0.053)]$. Given the previous data, the predictor of X_7 is

$$\hat{X}_7 = 5.28 + 0.81(5.22399 - 5.28) = 5.23463,$$

and the variance of the estimation error is $\sigma^2_{v7} = 0.19947$. The estimator for X_8 is

$$\hat{X}_8 = 5.28 + 0.81(-0.04537) + (0.30287)(0.35587)^{-1}(0.05 + 0.03675)$$
$$= 5.31708,$$

where $Y_8 - 5.28 = 0.05$,

$$\sigma^2_{w8} = \sigma^2_e + \alpha^2\sigma^2_{w7} = 0.172 + (0.81)^2 0.19947 = 0.30287,$$
$$D_8 = 0.053 + 0.30287 = 0.35587,$$
$$\sigma^2_{v8} = 0.30287 - (0.30287)^2(0.35587)^{-1} = 0.04511.$$

The calculations are analogous for $\hat{X}_{11}$. The estimator for X_{12} is

$$\hat{X}_{12} = 5.28 + 0.81(5.88957 - 5.28) = 5.77375,$$

where

$$\sigma^2_{w,12} = \sigma^2_e + \alpha^2\sigma^2_{w,11} = 0.172 + (0.81)^2 0.19948 = 0.30288 .$$

The estimator for X_{13} given in Table 4.6.4 is constructed with

$$(\sigma^2_{w,13}, D_{13}, \sigma^2_{v,13}) = (0.37072, 0.42372, 0.04637) .$$

These calculations illustrate the importance of Y_t in the estimation of X_t for these data. ▲▲

In some situations, it is useful to express an autoregressive moving average in the state space form. The vector first order autoregressive process is in the form (4.6.15) with a fixed matrix **A** and zero measurement error. Therefore, because any autoregressive model can be put in the vector first order form (see Section 2.8), any pure autoregressive model can be put in the state space form. We shall see that there are alternative state space representations.

Akaike (1974) suggested a state space representation for autoregressive moving averages. To introduce the ideas, consider a second order moving average

$$Y_t = \beta_2\epsilon_{t-2} + \beta_1\epsilon_{t-1} + \epsilon_t , \tag{4.6.33}$$

where the ϵ_t are independent $(0, \sigma^2_\epsilon)$ random variables. Let

$$\begin{aligned} \mathbf{X}'_t &= (E\{Y_t \mid t\}, E\{Y_{t+1} \mid t\}, E\{Y_{t+2} \mid t\}) , \\ &= (Y_t, \beta_2\epsilon_{t-1} + \beta_1\epsilon_t, \beta_2\epsilon_t) , \end{aligned} \tag{4.6.34}$$

where $E\{Y_\tau \mid t\}$ is the expected value of Y_τ given an infinite past $(Y_t, Y_{t-1}, \ldots)$. The vector $\mathbf{X}_t$ contains information equivalent to the vector $(\epsilon_t, \epsilon_{t-1})$, which is the information required for any future predictions of Y_t. Furthermore, $\mathbf{X}_t$ satisfies the first order vector autoregression

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t , \tag{4.6.35}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

and $\mathbf{e}'_t = (1, \beta_1, \beta_2)\epsilon_t$. The state space representation of (4.6.18) is completed by adding the "observation equation"

$$Y_t = (1, 0, 0)\mathbf{X}_t = \mathbf{H}\mathbf{X}_t \tag{4.6.36}$$

to (4.6.35).

To initiate the filter for the second order moving average, we begin with the

vector $\hat{\mathbf{X}}_0' = (0, 0, 0)$. The covariance matrix of $\hat{\mathbf{X}}_0 - \mathbf{X}_0$ is

$$\boldsymbol{\Sigma}_{vv00} = \begin{pmatrix} 1+\beta_1^2+\beta_2^2 & \beta_1+\beta_1\beta_2 & \beta_2 \\ \beta_1+\beta_1\beta_2 & \beta_2^2+\beta_1^2 & \beta_1\beta_2 \\ \beta_2 & \beta_1\beta_2 & \beta_2^2 \end{pmatrix} \sigma_\epsilon^2 .$$

The reader may verify that

$$\begin{aligned} \boldsymbol{\Sigma}_{ww11} &= \boldsymbol{\Sigma}_{\epsilon\epsilon} + \mathbf{A}\boldsymbol{\Sigma}_{vv00}\mathbf{A}' = \boldsymbol{\Sigma}_{vv00} , \\ D_1 &= \mathbf{H}\boldsymbol{\Sigma}_{ww11}\mathbf{H}' = (1+\beta_1^2+\beta_2^2)\sigma_\epsilon^2 , \\ \boldsymbol{\Sigma}_{\epsilon\epsilon} &= (1, \beta_1, \beta_2)'(1, \beta_1, \beta_2)\sigma_\epsilon^2 , \\ \hat{\mathbf{X}}_1' &= (1+\beta_1^2+\beta_2^2)^{-1}(1+\beta_1^2+\beta_2^2, \beta_1+\beta_1\beta_2, \beta_2)Y_1 , \end{aligned} \tag{4.6.37}$$

and

$$\hat{\mathbf{X}}_{2|1} = \mathbf{A}\hat{\mathbf{X}}_1 = (1+\beta_1^2+\beta_2^2)^{-1}(\beta_1+\beta_1\beta_2, \beta_2, 0)'Y_1 . \tag{4.6.38}$$

The vector $\hat{\mathbf{X}}_1'$ is the best predictor of (Y_1, Y_2, Y_3) given Y_1, and the vector $\hat{\mathbf{X}}_{2|1}$ is the best predictor of (Y_2, Y_3, Y_4) given Y_1. The predictor of (Y_2, Y_3, Y_4) could have been obtained directly from the regression of (Y_2, Y_3, Y_4) on Y_1. The predictor for $\mathbf{X}_2$ given (Y_2, Y_1), and subsequent predictions can be constructed using the equations (4.6.20), (4.6.21), and (4.6.23).

We now give a state space representation for a stationary autoregressive moving average of order (p, q). Let

$$Y_t = \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t , \tag{4.6.39}$$

where the ϵ_t are independent $(0, \sigma_\epsilon^2)$ random variables. The vector of conditional expectations, $\mathbf{X}_t$, of (4.6.34) becomes

$$\mathbf{X}_t = (E\{Y_t \mid t\}, E\{Y_{t+1} \mid t\}, \ldots, E\{Y_{t+m-1} \mid t\})' , \tag{4.6.40}$$

where $m = \max(p, q+1)$. From Theorem 2.7.1, we have

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} , \tag{4.6.41}$$

where the ψ_i are defined in that theorem. It follows that

$$E\{Y_{t+j} \mid t\} = \sum_{i=j}^{\infty} \psi_i \epsilon_{t-i+j} , \tag{4.6.42}$$

for $j = 1, 2, \ldots, m-1$. From equations (4.6.41) and (4.6.42), we have

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t, \tag{4.6.43}$$

where

$$\mathbf{e}_t = (1, v_1, v_2, \ldots, v_{m-1})'\epsilon_t,$$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_m & \alpha_{m-1} & \alpha_{m-2} & \cdots & \alpha_1 \end{pmatrix},$$

and it is understood that $\alpha_m, \ldots, \alpha_{p+1}$ are zero if $m > p$. Equation (4.6.43) and the equation

$$Y_t = (1, 0, \ldots, 0)\mathbf{X}_t \tag{4.6.44}$$

form a state space representation for a stationary autoregressive moving average process.

Example 4.6.5. In this example, we use the Kalman filter to construct predictions for the stationary autoregressive moving average model

$$Y_t = 0.80Y_{t-1} + \epsilon_t + 0.60\epsilon_{t-1} + 0.58\epsilon_{t-2}, \tag{4.6.45}$$

where $\epsilon_t \sim \text{NI}(0, 1)$. The second column of Table 4.6.5 contains observations generated by the process (4.6.45). By Theorem 2.7.1,

$$Y_t = \sum_{i=0}^{\infty} v_i \epsilon_{t-i},$$

where $(v_0, v_1, v_2) = (1.00, 1.40, 1.70)$. The state space representation of the model (4.6.45) is

$$\begin{aligned} \mathbf{X}_t &= \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t, \\ Y_t &= (1, 0, 0)\mathbf{X}_t, \end{aligned} \tag{4.6.46}$$

where

$$\begin{aligned} \mathbf{X}_t' &= (Y_t, E\{Y_{t+1} \mid t\}, E\{Y_{t+2} \mid t\}), \\ \mathbf{e}_t' &= (1.00, 1.40, 1.70)\epsilon_t, \end{aligned}$$

Table 4.6.5. Kalman Filter Used to Construct Predictors for an Autoregressive Moving Average

t	Y_t	$\hat{X}_{t+1,1\mid t}$	$\hat{X}_{t+1,2\mid t}$	$V\{\hat{X}_{t+1,1\mid t}-Y_{t+1}\}$	$V\{\hat{X}_{t+1,2\mid t}-Y_{t+2}\}$
1	3.240	3.008	2.578	1.515	4.033
2	1.643	0.699	0.036	1.162	3.200
3	2.521	2.457	2.875	1.150	3.170
4	3.122	3.779	3.359	1.049	3.074
5	3.788	3.371	2.702	1.032	3.001
6	2.706	1.782	1.052	1.024	3.001
7	4.016	4.169	4.601	1.008	2.977
8	5.656	6.680	6.199	1.007	2.969
9	6.467	5.903	4.600	1.004	2.968
10	7.047	6.201	5.622	1.002	2.963
11	4.284	2.939	1.241	1.002	2.962
12	2.587	0.749	0.395	1.001	2.961
13	−0.421	−1.242	−1.671	1.000	2.960
14	0.149	0.276	1.028	1.000	2.960
15	−1.012	−0.776	−1.368	1.000	2.960

NOTE: $\hat{X}_{t+1,1\mid t}=\hat{Y}_{t+1}(Y_1,\ldots,Y_t)$ and $\hat{X}_{t+1,2\mid t}=\hat{Y}_{t+2}(Y_1,\ldots,Y_t)$.

and

$$\mathbf{A}=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0.8\end{pmatrix}.$$

Multiplying (4.6.45) by Y_t and Y_{t-1}, and taking expectations, we have

$$\gamma_Y(0)-0.80\gamma_Y(1)=2.8260,$$
$$\gamma_Y(1)-0.80\gamma_Y(0)=1.4120.$$

It follows that $[\gamma_Y(0),\gamma_Y(1)]=[10.9878, 10.2022]$. If we initiate the filter with

$$\hat{\mathbf{X}}_0'=(0,0,0),$$

the covariance matrix of $\hat{\mathbf{X}}_0-\mathbf{X}_0$ is the variance of $[Y_t, E\{Y_{t+1}\mid t\}, E\{Y_{t+2}\mid t\}]$, namely,

$$\mathbf{\Sigma}_{vv00}=\begin{pmatrix}10.988 & 10.202 & 8.742\\ 10.202 & 9.988 & 8.802\\ 8.742 & 8.802 & 8.028\end{pmatrix}.$$

To construct the estimator of $\mathbf{X}_1$ given $\hat{\mathbf{X}}_0$ and Y_1, we compute

$$\mathbf{\Sigma}_{ww11}=\mathbf{\Sigma}_{\epsilon\epsilon}+\mathbf{A}\mathbf{\Sigma}_{vv00}\mathbf{A}'=\mathbf{\Sigma}_{vv00},$$
$$D_1=(1,0,0)\mathbf{\Sigma}_{ww11}(1,0,0)'=\gamma_Y(0)=10.988,$$

and

$$\hat{\mathbf{X}}_1' = (1.000, 0.929, 0.796)\mathbf{Y}_1$$
$$= (3.240, 3.008, 2.578). \tag{4.6.47}$$

The reader can verify that the coefficients in (4.6.47) are $\gamma_Y^{-1}(0)[\gamma_Y(0), \gamma_Y(1), \gamma_Y(2)]$. The covariance matrix of $\hat{\mathbf{X}}_1 - \mathbf{X}_1$ is

$$\mathbf{\Sigma}_{vv11} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.515 & 0.685 \\ 0 & 0.685 & 1.073 \end{pmatrix}.$$

The predictor of $\mathbf{X}_2$ given $\hat{\mathbf{X}}_1$ is

$$\hat{\mathbf{X}}_{2|1} = \mathbf{A}\hat{\mathbf{X}}_1,$$

and the covariance matrix of the prediction error is

$$\mathbf{\Sigma}_{2|1} = \begin{pmatrix} 1.515 & 2.085 & 2.248 \\ 2.085 & 3.033 & 3.238 \\ 2.248 & 3.238 & 3.577 \end{pmatrix},$$

where $\mathbf{\Sigma}_{2|1} = \mathbf{A}\mathbf{\Sigma}_{vv11}\mathbf{A}' + \mathbf{\Sigma}_{ee}$. The quantities required to construct $\hat{\mathbf{X}}_2$ are

$$\mathbf{\Sigma}_{ww22} = \begin{pmatrix} 1.515 & 2.085 & 2.248 \\ 2.085 & 3.033 & 3.238 \\ 2.248 & 3.238 & 3.577 \end{pmatrix},$$

and

$$D_2 = \mathbf{H}_2\mathbf{\Sigma}_{ww22}\mathbf{H}_2' = 1.515.$$

The predictions and variances of the prediction errors are given in the third and fifth columns, respectively, of Table 4.6.5. As the initial effects die out, the estimator of $\mathbf{X}_t$ stabilizes at

$$\hat{\mathbf{X}}_t = \mathbf{A}\hat{\mathbf{X}}_{t-1} + (1.0, 1.4, 1.7)'[Y_t - (1, 0, 0)\mathbf{A}\hat{\mathbf{X}}_{t-1}]$$
$$= \mathbf{A}\mathbf{X}_{t-1} + (1.0, 1.4, 1.7)'\boldsymbol{\epsilon}_t.$$

The predictor stabilizes at

$$\hat{\mathbf{X}}_{t+1|t} = \mathbf{A}\hat{\mathbf{X}}_t = \mathbf{A}\mathbf{X}_t,$$

and the covariance matrix of the prediction error stabilizes at

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{V}\{\hat{\mathbf{X}}_{t+1|t} - \mathbf{X}_{t+1}\} = \mathbf{\Sigma}_{ee}.$$

The second entry in $\hat{\mathbf{X}}_{t+1|t}$ is the predictor of $E\{Y_{t+2} \mid t+1\}$, based on

observations through time t. Thus, it is also the predictor of Y_{t+2} based on observations through time t. The limiting variance of $\hat{X}_{t+1,2|t} - Y_{t+2}$ is $(1 + v_1^2)\sigma^2$, while the second entry on the diagonal of Σ_{ee} is $v_1^2\sigma^2$. In the last column of Table 4.6.5 we give

$$V\{\hat{X}_{t+1,2|t} - Y_{t+2}\} = V\{\hat{X}_{t+1,2|t} - X_{t+1,2}\} + 1\,,$$

where $V\{\hat{X}_{t+1,2|t} - X_{t+1,2}\}$ is the second entry on the diagonal of $\Sigma_{t+1|t}$ and $\sigma^2 = 1$. ▲▲

REFERENCES

Sections 4.1, 4.3–4.5. Anderson (1971), Beran (1992), Blackman and Tukey (1959), Bloomfield (1976), Brillinger (1975), Brockwell and Davis (1991), Granger and Hatanaka (1964), Hannan (1960, 1970), Harris (1967), Jenkins and Watts (1968), Kendall and Stuart (1966), Koopmans (1974), Whittle (1963), Yaglom (1962).

Section 4.2. Amemiya and Fuller (1967), Bellman (1960), Wahba (1968).

Section 4.6. Akaike (1974), Anderson and Moore (1979), Diderrich (1985), Duncan and Horn (1972), Jones (1980), Kalman (1960, 1963), Kalman and Bucy (1961), Meinhold and Singpurwalla (1983), Sage (1968), Sallas and Harville (1981), Sorenson (1966).

EXERCISES

1. Which of the following functions is the spectral density of a stationary time series? Explain why or why not.
   (a) $f(\omega) = 1 - \frac{1}{2}\omega^2$, $-\pi \leq \omega \leq \pi$.
   (b) $f(\omega) = 1 + \frac{1}{2}\omega$, $-\pi \leq \omega \leq \pi$.
   (c) $f(\omega) = 476 + \cos 13\omega$, $-\pi \leq \omega \leq \pi$.

2. Let $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ be defined by $X_t = e_t + 0.4e_{t-1}$. Compute the autocovariance function $\gamma(h)$ and the spectral density $f(\omega)$, given that the e_t are independently and identically distributed $(0, \sigma^2)$ random variables.

3. Give the spectral density for the time series defined by

$$X_t - \beta X_{t-1} = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}\,, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

   where $|\beta| < 1$, the roots of $m^2 + \alpha_1 m + \alpha_2 = 0$ are less than one in absolute value, and the e_t are uncorrelated $(0, \sigma^2)$ random variables.

4. Let

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = Z_t$$

and

$$Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} = e_t,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables, and the roots of both $m^2 + \alpha_1 m + \alpha_2 = 0$ and $r^2 + \beta_1 r + \beta_2 = 0$ are less than one in absolute value. Give an expression for the spectral density of X_t. How do you describe the time series X_t?

5. Find the covariance function and spectral distribution function for the time series

$$X_t = u_1 \cos t + u_2 \sin t + Y_t,$$

where $(u_1, u_2)'$ is distributed as a bivariate normal random variable with zero mean and diagonal covariance matrix diag(2, 2), $Y_t = e_t - e_{t-1}$, and the e_t are independent (0, 3) random variables, independent of $(u_1, u_2)'$.

6. Given the following spectral distribution function:

$$F_X(\omega) = \begin{cases} \pi + \omega, & -\pi \leq \omega < -\pi/2, \\ 5\pi + \omega, & -\pi/2 \leq \omega < \pi/2, \\ 9\pi + \omega, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

What is the variance of X_t? What is the spectral distribution function of $X_t - X_{t-1}$? Is there a k such that $X_t - X_{t-k}$ will have a continuous spectral distribution function?

7. Prove Corollary 4.3.1.3.

8. Let $\{e_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ be a time series of uncorrelated $(0, \sigma^2)$ random variables. Let $X_t = e_{t-2} + 0.5e_{t-3}$. Give the covariance matrix $\Gamma(h)$ and the spectral matrix $\mathbf{f}(\omega)$ for $(X_t, e_t)'$.

9. Let $\{e_{1t}\}$ and $\{e_{2t}\}$ be two independent sequences of uncorrelated random variables with variances 0.34 and 0.50 respectively. Let

$$X_{1t} = 0.8X_{1t-1} + e_{1t},$$
$$X_{2t} = X_{1t} + e_{2t}.$$

Find $\gamma_{22}(h)$, $\gamma_{12}(h)$, $f_{12}(\omega)$, and $f_{22}(\omega)$.

10. The complex vector random variable $\mathbf{X}$ of dimension q is distributed as a complex multivariate normal if the real vector $[(\mathrm{Re}\,\mathbf{X})', (\mathrm{Im}\,\mathbf{X})']'$ is distributed as a multivariate normal with mean $[(\mathrm{Re}\,\boldsymbol{\mu})', (\mathrm{Im}\,\boldsymbol{\mu})']'$ and covariance

matrix

$$\frac{1}{2}\begin{pmatrix} \operatorname{Re}\Sigma & -\operatorname{Im}\Sigma \\ \operatorname{Im}\Sigma & \operatorname{Re}\Sigma \end{pmatrix}.$$

where $\mu = E\{\mathbf{X}\}$ and Σ is a positive semidefinite Hermitian matrix. Let $(A_1, B_1, A_2, B_2)'$ be the multivariate normal random variable defined following (4.4.9). Show that $\mathbf{X} = (A_1 + iB_1, A_2 + iB_2)'$ is distributed as a complex bivariate normal random variable. Give the Hermitian matrix Σ.

11. Let $\{a_j\}_{j=-\infty}^{\infty}$ and $\{b_s\}_{s=-\infty}^{\infty}$ be absolutely summable, and let $\{X_t: t \in (0, \pm 1, \pm 2, \ldots)\}$ be a stationary time series with absolutely summable covariance function. Consider the following filtering operation: (1) apply the filter $\{a_j\}$ to the time series X_t to obtain a time series Z_t and then (2) apply the filter $\{b_j\}$ to Z_t to obtain a time series Y_t. What is the transfer function of this filtering operation? Express the spectral density of Y_t as a function of the spectral density of X_t and of the transfer function.

12. Let $\{a_j\}$, $\{b_j\}$, and $\{X_t\}$ be as defined in Exercise 11, and define

$$Y_t = \sum_{j=-\infty}^{\infty} (a_j + b_j) X_{t-j}.$$

Express the spectral density of Y_t as a function of the spectral density of X_t and the transfer functions of $\{a_j\}$ and $\{b_j\}$.

13. Let X_t and Y_t be defined by

$$X_t = 0.9X_{t-1} + e_t,$$
$$Y_t = X_t + u_t,$$

where $\{e_t\}$ is a sequence of normal independent $(0, 1)$ random variables independent of the sequence $\{u_t\}$. Let u_t satisfy the difference equation

$$u_t + 0.5u_{t-1} = v_t,$$

where $\{v_t\}$ is a sequence of normal independent $(0, 0.3)$ random variables. Assuming that only Y_t is observed, construct the filter $\{a_j: j = -2, -1, 0, 1, 2\}$ so that

$$\sum_{j=-2}^{2} a_j Y_{t-j}$$

is the minimum mean square error estimator of X_t. Construct the one-sided filter $\{b_j: j = 0, 1, \ldots, 5\}$ to estimate X_t. How does the mean square error of these filters compare with the lower bound for linear filters?

14. Let $f(\omega)$ be an even nonnegative continuous periodic function of period 2π. Let

$$a(h) = \int_{-\pi}^{\pi} f(\omega) e^{-i\omega h}\, d\omega .$$

Show that, for q a positive integer,

$$\gamma(h) = \begin{cases} \dfrac{q - |h|}{q} a(h), & h = 0, \pm 1, \pm 2, \ldots, \pm q, \\ 0 & \text{otherwise} \end{cases}$$

is the covariance function of a stationary time series. (*Hint:* See Exercise 3.15 and Theorem 3.1.10.)

15. Let X_t be a time series with continuous spectral density $f_X(\omega)$. Let Y_t be a time series satisfying

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

$$|f_Y(\omega) - f_X(\omega)| < \epsilon ,$$

where the e_t are uncorrelated $(0, \sigma^2)$ random variables, $\alpha_0 = 1$, and Y_t is defined by Theorem 4.3.4.

(a) Show that $|\gamma_X(h) - \gamma_Y(h)| < 2\pi\epsilon$ for all h.

(b) Let $f_X(\omega)$ be strictly positive. Prove that given $\epsilon > 0$ there is a p and a set $\{\alpha_j : j = 0, 1, \ldots, p\}$ with $\alpha_0 = 1$ such that the time series Z_t defined by

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

satisfies

$$\left| f_Z(\omega) - \frac{\sigma^2}{2\pi} \right| < \epsilon$$

for all ω, where $f_Z(\omega)$ is the spectral density of Z_t and $\gamma_Z(0) = \sigma^2$. Show that

$$\sum_{h=1}^{\infty} \gamma_Z^2(h) \leq 2\pi\epsilon^2 .$$

(c) Let $f_X(\omega)$ be strictly positive. Show that, given $\epsilon > 0$, one may define two autoregressive time series Y_{1t} and Y_{2t} with spectral densities

$$f_{Y_1}(\omega) = \frac{\sigma_1^2}{2\pi \left| \sum_{j=0}^{p_1} \alpha_{1j} e^{-i\omega j} \right|^2} ,$$

$$f_{Y_2}(\omega) = \frac{\sigma_2^2}{2\pi \left| \sum_{j=0}^{p_2} \alpha_{2j} e^{-i\omega j} \right|^2} ,$$

such that

$$f_X(\omega) - \epsilon \leq f_{Y_1}(\omega) \leq f_X(\omega) \leq f_{Y_2}(\omega) \leq f_X(\omega) + \epsilon .$$

(d) For the three time series defined in part (c) prove that

$$\mathrm{Var}\left\{\sum_{t=1}^{n} a_t Y_{1t}\right\} \leq \mathrm{Var}\left\{\sum_{t=1}^{n} a_t X_t\right\} \leq \mathrm{Var}\left\{\sum_{t=1}^{n} a_t Y_{2t}\right\}$$

for any fixed real numbers $\{a_t\colon t = 1, 2, \ldots, n\}$.

16. Let

$$X_{1t} = e_t - 0.8e_{t-1} ,$$
$$X_{2t} = u_t - 0.9u_{t-4} ,$$

where $\{e_t\}$ is a sequence of independent $(0, 1)$ random variables independent of $\{u_t\}$, a sequence of independent $(0, 6)$ random variables. Express

$$Y_t = X_{1t} + X_{2t}$$

as a moving average process.

17. Let Y_t be defined by

$$Y_t = S_t + Z_t ,$$
$$S_t = 0.9\, S_{t-4} + u_t ,$$
$$Z_t = 0.8\, Z_{t-1} + e_t ,$$

where the sequence $\{(e_t, u_t)'\}$ is a sequence of normal independent $(\mathbf{0}, \boldsymbol{\Sigma})$ random variables with $\boldsymbol{\Sigma} = \mathrm{diag}(0.1, 0.6)$. Construct the optimum filter $\{a_j\colon j = -9, -8, \ldots, 8, 9\}$ to estimate S_t, where the estimator is defined by $\Sigma_{j=-9}^{9} a_j Y_{t-j}$. Construct the best one-sided filter $\{b_j\colon j = 0, 1, \ldots, 8, 9\}$ to estimate S_t. Compare the mean square error of these filters with the lower bound for linear filters.

18. Let $\mathbf{C}$ and $\mathbf{D}$ be $k \times k$ nonsingular matrices. Show that

$$(\mathbf{C}^{-1} + \mathbf{D}^{-1})^{-1}\mathbf{D}^{-1} = \mathbf{C}(\mathbf{C} + \mathbf{D})^{-1} .$$

19. Let $X \sim N(0, \sigma_X^2)$, and let $Y = X + u$, where $u \sim N(0, \sigma_u^2)$ independent of X. Give the covariance matrix of (Y, X) and the conditional expected value of X given Y. Consider the regression problem

$$(Y, 0)' = (1, 1)'X + (u, \epsilon)' ,$$

where $(u, \epsilon)' \sim N[(0, 0)', \mathrm{diag}\{\sigma_u^2, \sigma_X^2\}]$. Show that the generalized regression

estimator of X constructed with known (σ_X^2, σ_u^2) is the conditional expected value of X given Y.

20. Let $\mathbf{\Sigma}_{uu}$ be an $r \times r$ nonsingular matrix, $\mathbf{\Sigma}_{ww}$ a $p \times p$ nonsingular matrix, and $\mathbf{H}$ an $r \times p$ matrix. Show that

$$[\mathbf{H}'\mathbf{\Sigma}_{uu}^{-1}\mathbf{H} + \mathbf{\Sigma}_{ww}^{-1}]^{-1} = \mathbf{\Sigma}_{ww} - \mathbf{\Sigma}_{ww}\mathbf{H}'(\mathbf{\Sigma}_{uu} + \mathbf{H}\mathbf{\Sigma}_{ww}\mathbf{H}')^{-1}\mathbf{H}\mathbf{\Sigma}_{ww},$$

and hence verify (4.6.21). [See Duncan and Horn (1972).]

21. Using Exercise 20, verify equation (4.6.23).

22. Let the model (4.6.1)–(4.6.2) hold, and assume it is desired to construct a recursive estimator of X_{t+s} given the observations $Y_1, Y_2, \ldots, Y_t$. Construct such an estimator. Consider both $s > 0$ and $s < 0$.

23. Let the following linear model hold

$$\mathbf{Y}_1 = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{u}_1,$$
$$\mathbf{Y}_2 = \mathbf{X}_2\boldsymbol{\beta} + \mathbf{u}_2,$$

where $\mathbf{Y}_1$ is $n_1 \times 1$, $\mathbf{X}_1$ is $n_1 \times k$, $\boldsymbol{\beta}$ is $k \times 1$, $\mathbf{Y}_2$ is $n_2 \times 1$, $\mathbf{X}_2$ is $n_2 \times k$, and $(\mathbf{u}_1', \mathbf{u}_2')' \sim N[\mathbf{0}, \mathbf{I}\sigma^2]$. Let $\mathbf{X}_1'\mathbf{X}_1$ be nonsingular. Let $\mathbf{X}_1'\mathbf{X}_1$ and $\hat{\boldsymbol{\beta}}_1$, where

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1,$$

be given. Construct the best linear unbiased estimator of $\boldsymbol{\beta}$ as a function of $\hat{\boldsymbol{\beta}}_1$, $\mathbf{X}_1'\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{Y}_2$.

24. Prove Corollary 4.2.2.

25. Prove the following lemma.

Lemma. Let $\mathbf{\Sigma}$ be the nonsingular covariance matrix of the vector $\mathbf{Y}_t' = (\mathbf{Y}_{t1}', \mathbf{Y}_{t2}', \mathbf{Y}_{t3}')$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} & \mathbf{\Sigma}_{33} \end{pmatrix}$$

and $E\{\mathbf{Y}_t\} = \mathbf{0}$. Let

$$(\hat{\mathbf{Y}}_{t3|1}', \hat{\mathbf{Y}}_{t2|1}') = (\mathbf{Y}_{t1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{13}, \mathbf{Y}_{t1}'\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}]$$

be the best linear estimator of $(\mathbf{Y}_{t3}', \mathbf{Y}_{t2}')$ based upon $\mathbf{Y}_{t1}$. Let

$$(\mathbf{e}_{t2}', \mathbf{e}_{t3}') = [(\mathbf{Y}_{t2} - \hat{\mathbf{Y}}_{t2|1})', (\mathbf{Y}_{t3} - \hat{\mathbf{Y}}_{t3|1})'],$$

and let $\mathbf{V}$ be the covariance matrix of $(\mathbf{e}_{t2}', \mathbf{e}_{t3}')'$, where

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{32} & \mathbf{V}_{33} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{23} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{32} - \boldsymbol{\Sigma}_{31}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{33} - \boldsymbol{\Sigma}_{31}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{13} \end{pmatrix}.$$

Then the best linear estimator of $\mathbf{Y}_{t3}$ given $(\mathbf{Y}_{t1}', \mathbf{Y}_{t2}')$ is the best linear estimator of $\mathbf{Y}_{t3}$ given $\mathbf{Y}_{t1}$, plus the best linear estimator of $\mathbf{e}_{t3|1}$ given $\mathbf{e}_{t2|1}$. That is,

$$\hat{\mathbf{Y}}_{t3|(1,2)} = \hat{\mathbf{Y}}_{t3|1} + \hat{\mathbf{e}}_{t3|2},$$

where $\hat{\mathbf{e}}_{t3|2}' = \mathbf{e}_{t2}'\mathbf{V}_{22}^{-1}\mathbf{V}_{23}$, and

$$\hat{\mathbf{Y}}_{t3|(1,2)} = (\boldsymbol{\Sigma}_{31}, \boldsymbol{\Sigma}_{32}) \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_{t1} \\ \mathbf{Y}_{t2} \end{pmatrix}.$$

26. Show that the predictor of Y_{t+2} given $(Y_1, \ldots, Y_t)$ of Example 4.6.4 is the first element of $\mathbf{A}^2\hat{\mathbf{X}}_t$. Show that the variance of the prediction error is the upper left element of

$$\mathbf{A}^2\boldsymbol{\Sigma}_{vv|t}\mathbf{A}^{2\prime} + \mathbf{A}\boldsymbol{\Sigma}_{\epsilon\epsilon}\mathbf{A}' + \boldsymbol{\Sigma}_{\epsilon\epsilon}.$$

27. Assume that the time series Y_t has a spectral density $f_Y(\omega)$ such that $f_Y(\omega) > 0$ for all ω and such that the Fourier coefficients of $[f_Y(\omega)]^{-1}$ are absolutely summable. Cleveland (1972) defines the *inverse autocorrelation function* by $\rho i(h) = [\gamma i(0)]^{-1}\gamma i(h)$, where

$$\gamma i(h) = \int_{-\pi}^{\pi} (4\pi^2)^{-1}[f_Y(\omega)]^{-1} e^{e\omega h}\, d\omega.$$

(a) Show that if Y_t is a stationary autoregressive process satisfying

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

then the $\rho i(h)$ are the autocorrelations of the moving average process

$$X_t = e_t + \sum_{i=1}^{p} \alpha_i e_{t-i}.$$

(b) Show that if Y_t is an ARMA(p, q) process, then $\rho i(h)$ is the autocorrelation of the ARMA(q, p) process with the autoregressive and moving average coefficients interchanged.

28. Let X_t have the spectral density

$$f_X(\omega) = \begin{cases} (2\pi)^{-1}, & \pi/4 \leq \omega \leq \pi/2, \\ (2\pi)^{-1}, & -\pi/2 \leq \omega \leq -\pi/4, \\ 0 & \text{otherwise}. \end{cases}$$

The time series X_t is sometimes called "bandpassed white noise." Give the covariance function of X_t. Plot the covariance function. Using the covariance function (or otherwise), approximate X_t by a third order autoregressive process. Plot the spectral density of the third order process with the spectral density of the original process. What would be the error made in predicting X_t one period ahead using the third order autoregressive process?

Given the availability of a random number generator, explain how you would create 200 observations from the time series X_t using the "bandpass" idea. Assume you wish to create a normal time series.

29. Let

$$Y_t = S_t + u_t,$$

where S_t is a stationary time series satisfying

$$S_t = 0.9S_{t-2} + e_t,$$

where the e_t are NI(0, 1.9) random variables. Let u_t be a sequence of NI(0, 1) random variables where u_t is independent of S_j for all t, j. Construct a centered two-sided filter of length seven to estimate S_t given $(Y_{t-3}, Y_{t-2}, \ldots, Y_{t+3})$. Construct a one-sided filter of length 6 using $(Y_t, Y_{t-1}, \ldots, Y_{t-5})$ to estimate S_t. For the two-sided filter define

$$X_t = S_t - \hat{S}_t,$$

where $\hat{S}_t$ is the estimate of S_t computed using the filter. Compare the variance of X_t with the lower bound for a two-sided filter.

30. Let Y_t be a stationary autoregressive process satisfying

$$\sum_{i=0}^{p} \alpha_i Y_{t-i} = e_t, \qquad \text{or} \quad B(\mathcal{B})Y_t = e_t,$$

where $\alpha_0 = 1$, $B(\mathcal{B}) = \sum_{i=0}^{p} \alpha_i \mathcal{B}^i$, e_t are independent $(0, \sigma^2)$ random variables, and $\mathcal{B}$ is the backshift operator. The process also satisfies

$$\sum_{i=0}^{p} \alpha_i Y_{t+i} = a_t, \qquad \text{or} \quad A(\mathcal{B})Y_t = a_t,$$

where $A(\mathcal{B}) = \sum_{i=0}^{p} \alpha_i \mathcal{B}^{p-i}$ and a_t are uncorrelated $(0, \sigma^2)$ random variables.

Show that the partial correlation between Y_t and Y_{t-p} given $(Y_{t-p}, \ldots, Y_{t-1}, Y_{t+1}, \ldots, Y_{t+p})$ is zero for $j = 1, 2, \ldots$, and that the partial correlation between Y_t and Y_{t-p-j} given $(Y_{t-p}, \ldots, Y_{t-1}, Y_{t+1}, \ldots, Y_{t+p})$ is zero for $j = 1, 2, \ldots$.

31. Let

$$Y_t = X_t + u_t,$$
$$X_t = \theta X_{t-1} + e_t,$$

where $e_t \sim \text{NI}(0, \sigma_e^2)$, $u_t \sim \text{NI}(0, \sigma_u^2)$, $\{e_t\}$ is independent of $\{u_t\}$, and X_t is stationary with $|\theta| < 1$. Show that

$$Y_t = \theta Y_{t-1} + v_t + \beta v_{t-1},$$

where $\beta = -[\sigma_e^2 + (1 + \theta^2)\sigma_u^2]^{-1}\theta\sigma_u^2$, v_t are $\text{NI}(0, \sigma_v^2)$, and $(1 + \beta^2)\sigma_v^2 = \sigma_e^2 + (1 + \theta^2)\sigma_u^2$.