

8

Multivariate Time Series

- 8.1 Examples
- 8.2 Second-Order Properties of Multivariate Time Series
- 8.3 Estimation of the Mean and Covariance Function
- 8.4 Multivariate ARMA Processes
- 8.5 Best Linear Predictors of Second-Order Random Vectors
- 8.6 Modeling and Forecasting with Multivariate AR Processes
- 8.7 Cointegration

Many time series arising in practice are best considered as components of some vector-valued (multivariate) time series $\{\mathbf{X}_t\}$ having not only serial dependence within each component series $\{X_{ti}\}$ but also interdependence between the different component series $\{X_{ti}\}$ and $\{X_{tj}\}$, $i \neq j$. Much of the theory of univariate time series extends in a natural way to the multivariate case; however, new problems arise. In this chapter we introduce the basic properties of multivariate series and consider the multivariate extensions of some of the techniques developed earlier. In Section 8.1 we introduce two sets of bivariate time series data for which we develop multivariate models later in the chapter. In Section 8.2 we discuss the basic properties of stationary multivariate time series, namely, the mean vector $\boldsymbol{\mu} = E\mathbf{X}_t$ and the covariance matrices $\Gamma(h) = E(\mathbf{X}_{t+h}\mathbf{X}_t') - \boldsymbol{\mu}\boldsymbol{\mu}'$, $h = 0, \pm 1, \pm 2, \dots$, with reference to some simple examples, including multivariate white noise. Section 8.3 deals with estimation of $\boldsymbol{\mu}$ and $\Gamma(\cdot)$ and the question of testing for serial independence on the basis of observations of $\mathbf{X}_1, \dots, \mathbf{X}_n$. In Section 8.4 we introduce multivariate ARMA processes and illustrate the problem of multivariate model identification with an example of a multivariate AR(1) process that also has an MA(1) representation. (Such examples do not exist in the univariate case.) The identification problem can be avoided by confining attention to multivariate autoregressive (or VAR) models. Forecasting multivariate time series with known second-order properties is discussed in Section 8.5, and in Section 8.6 we consider the modeling and forecasting of multivariate time series using the multivariate Yule–Walker equations and Whittle’s generalization of the

Durbin–Levinson algorithm. Section 8.7 contains a brief introduction to the notion of cointegrated time series.

8.1 Examples

In this section we introduce two examples of bivariate time series. A bivariate time series is a series of two-dimensional vectors $(X_{t1}, X_{t2})'$ observed at times t (usually $t = 1, 2, 3, \dots$). The two component series $\{X_{t1}\}$ and $\{X_{t2}\}$ could be studied independently as univariate time series, each characterized, from a second-order point of view, by its own mean and autocovariance function. Such an approach, however, fails to take into account possible dependence *between* the two component series, and such cross-dependence may be of great importance, for example in predicting future values of the two component series.

We therefore consider the series of random vectors $\mathbf{X}_t = (X_{t1}, X_{t2})'$ and define the mean vector

$$\boldsymbol{\mu}_t := E\mathbf{X}_t = \begin{bmatrix} EX_{t1} \\ EX_{t2} \end{bmatrix}$$

and covariance matrices

$$\Gamma(t+h, t) := \text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t) = \begin{bmatrix} \text{cov}(X_{t+h,1}, X_{t1}) & \text{cov}(X_{t+h,1}, X_{t2}) \\ \text{cov}(X_{t+h,2}, X_{t1}) & \text{cov}(X_{t+h,2}, X_{t2}) \end{bmatrix}.$$

The bivariate series $\{\mathbf{X}_t\}$ is said to be **(weakly) stationary** if the moments $\boldsymbol{\mu}_t$ and $\Gamma(t+h, t)$ are both independent of t , in which case we use the notation

$$\boldsymbol{\mu} = E\mathbf{X}_t = \begin{bmatrix} EX_{t1} \\ EX_{t2} \end{bmatrix}$$

and

$$\Gamma(h) = \text{Cov}(\mathbf{X}_{t+h}, \mathbf{X}_t) = \begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{bmatrix}.$$

The diagonal elements are the autocovariance functions of the univariate series $\{X_{t1}\}$ and $\{X_{t2}\}$ as defined in Chapter 2, while the off-diagonal elements are the covariances between $X_{t+h,i}$ and X_{tj} , $i \neq j$. Notice that $\gamma_{12}(h) = \gamma_{21}(-h)$.

A natural estimator of the mean vector $\boldsymbol{\mu}$ in terms of the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ is the vector of sample means

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t,$$

and a natural estimator of $\Gamma(h)$ is

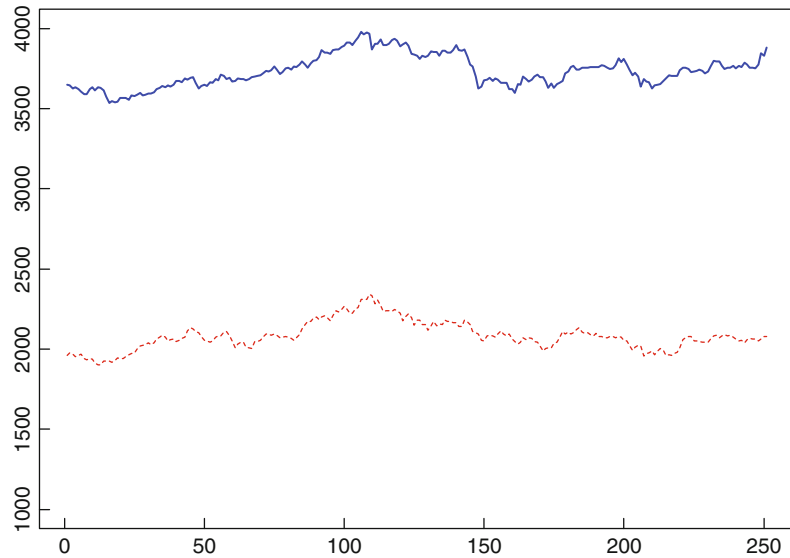
$$\hat{\Gamma}(h) = \begin{cases} n^{-1} \sum_{t=1}^{n-h} (\mathbf{X}_{t+h} - \bar{\mathbf{X}}_n) (\mathbf{X}_t - \bar{\mathbf{X}}_n)' & \text{for } 0 \leq h \leq n-1, \\ \hat{\Gamma}(-h)' & \text{for } -n+1 \leq h < 0. \end{cases}$$

The correlation $\rho_{ij}(h)$ between $X_{t+h,i}$ and X_{tj} is estimated by

$$\hat{\rho}_{ij}(h) = \hat{\gamma}_{ij}(h) (\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0))^{-1/2}.$$

If $i = j$, then $\hat{\rho}_{ij}$ reduces to the sample autocorrelation function of the i th series. These estimators will be discussed in more detail in Section 8.2.

Figure 8-1
The Dow Jones Index
(top) and Australian
All Ordinaries Index
(bottom) at closing on
251 trading days ending
August 26th, 1994



Example 8.1.1 Dow Jones and All Ordinaries Indices; DJAO2.TSM

Figure 8-1 shows the closing values D_0, \dots, D_{250} of the Dow Jones Index of stocks on the New York Stock Exchange and the closing values A_0, \dots, A_{250} of the Australian All Ordinaries Index of Share Prices, recorded at the termination of trading on 251 successive trading days up to August 26th, 1994. (Because of the time difference between Sydney and New York, the markets do not close simultaneously in both places; however, in Sydney the closing price of the Dow Jones index for the previous day is known before the opening of the market on any trading day.) The efficient market hypothesis suggests that these processes should resemble random walks with uncorrelated increments. In order to model the data as a stationary bivariate time series we first reexpress them as percentage relative price changes or **percentage returns** (filed as DJAOPC2.TSM)

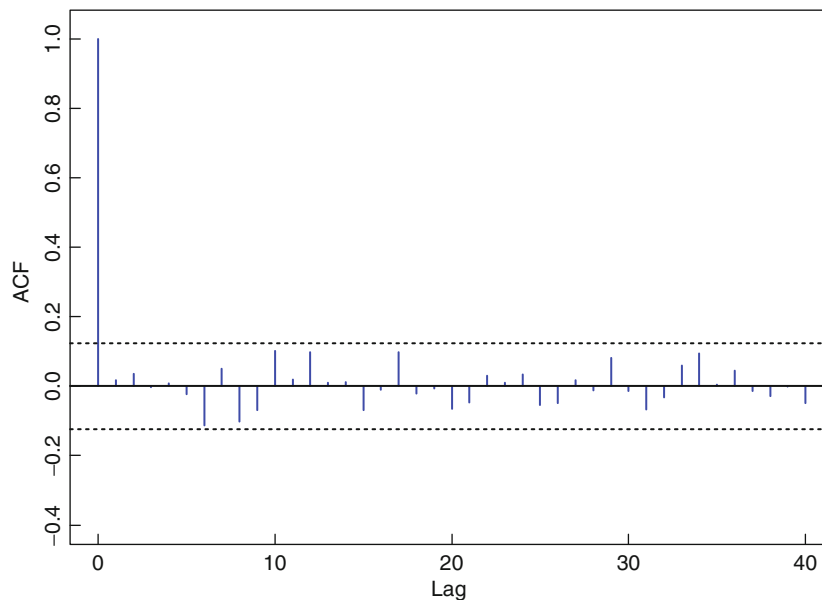
$$X_{t1} = 100 \frac{(D_t - D_{t-1})}{D_{t-1}}, \quad t = 1, \dots, 250,$$

and

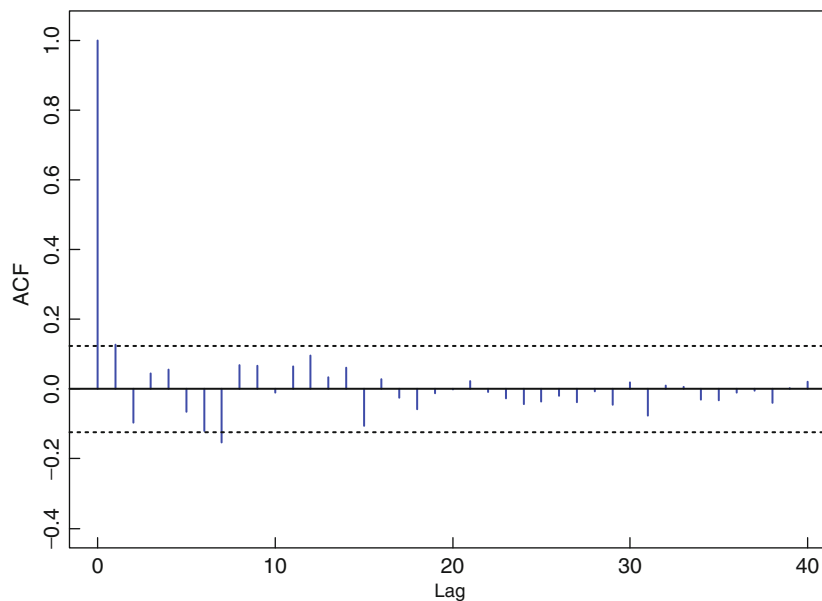
$$X_{t2} = 100 \frac{(A_t - A_{t-1})}{A_{t-1}}, \quad t = 1, \dots, 250.$$

The estimators $\hat{\rho}_{11}(h)$ and $\hat{\rho}_{22}(h)$ of the autocorrelations of the two univariate series are shown in Figures 8-2 and 8-3. They are not significantly different from zero.

To compute the sample cross-correlations $\hat{\rho}_{12}(h)$ and $\hat{\rho}_{21}(h)$ using ITSM, select File>Project>Open>Multivariate. Then click OK and double-click on the file name DJAOPC2.TSM. You will see a dialog box in which Number of columns should be set to 2 (the number of components of the observation vectors). Then click OK, and the graphs of the two component series will appear. To see the correlations, press the middle yellow button at the top of the ITSM window. The correlation functions are plotted as a 2×2 array of graphs with $\hat{\rho}_{11}(h)$, $\hat{\rho}_{12}(h)$ in the top row and $\hat{\rho}_{21}(h)$, $\hat{\rho}_{22}(h)$ in the second row. We see from these graphs (shown in Figure 8-4) that although the autocorrelations $\hat{\rho}_{ii}(h)$, $i = 1, 2$, are all small, there is a much larger correlation between $X_{t-1,1}$ and $X_{t,2}$. This indicates the importance of considering the two series jointly as components of a bivariate time series. It also suggests that the value of $X_{t-1,1}$, i.e., the Dow Jones return on day $t - 1$, may be of assistance in predicting the value of $X_{t,2}$, the All Ordinaries return on day t . This last

**Figure 8-2**

The sample ACF $\hat{\rho}_{11}$ of the observed values of $\{X_{t1}\}$ in Example 8.1.1, showing the bounds $\pm 1.96n^{-1/2}$

**Figure 8-3**

The sample ACF $\hat{\rho}_{22}$ of the observed values of $\{X_{t2}\}$ in Example 8.1.1, showing the bounds $\pm 1.96n^{-1/2}$

observation is supported by the scatterplot of the points $(x_{t-1,1}, x_{t,2})$, $t = 2, \dots, 250$, shown in Figure 8-5.

□

Example 8.1.2 Sales with a leading indicator; LS2.TSM

In this example we consider the sales data $\{Y_{t2}, t = 1, \dots, 150\}$ with leading indicator $\{Y_{t1}, t = 1, \dots, 150\}$ given by Box and Jenkins (1976, p. 537). The two series are stored in the ITSM data files SALES.TSM and LEAD.TSM, respectively, and in bivariate format as LS2.TSM. The graphs of the two series and their sample autocorrelation functions strongly suggest that both series are nonstationary. Application of the operator $(1 - B)$ yields the two differenced series $\{D_{t1}\}$ and $\{D_{t2}\}$, whose properties are compatible with those of low-order ARMA processes. Using ITSM, we find that the models

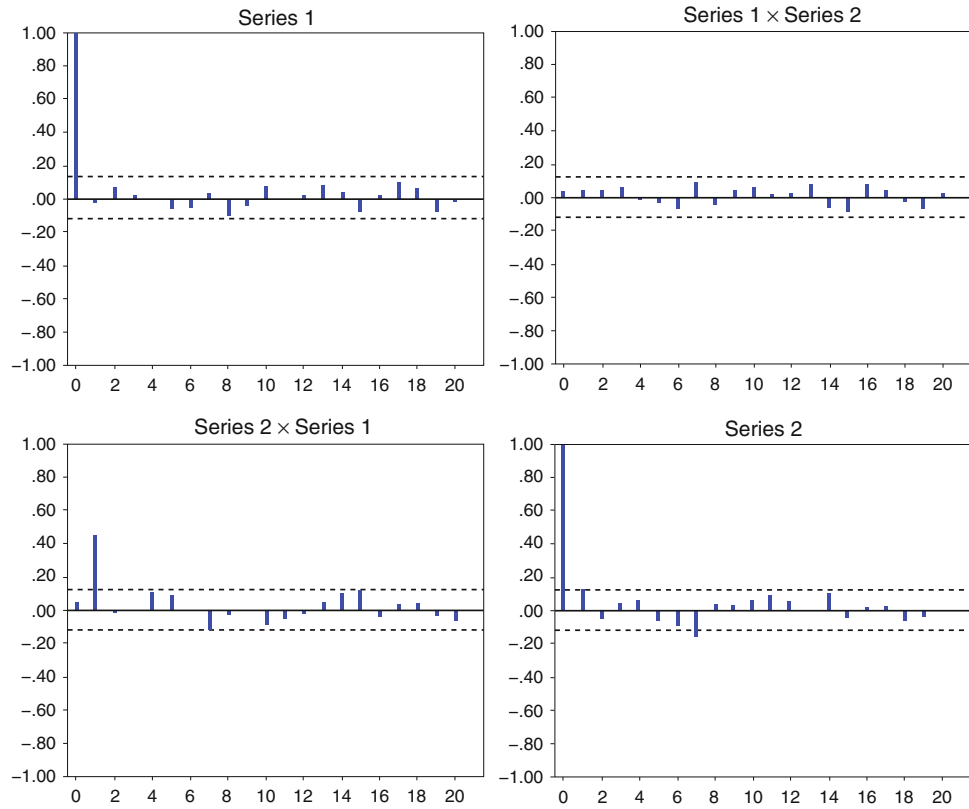


Figure 8-4
The sample correlations $\hat{\rho}_{ij}(h)$ between $X_{t+h,i}$ and $X_{t,j}$ for Example 8.1.1. ($\rho_{ij}(h)$ is plotted as the j th graph in the i th row, $i, j = 1, 2$. Series 1 and 2 consist of the daily Dow Jones and All Ordinaries percentage returns, respectively.)

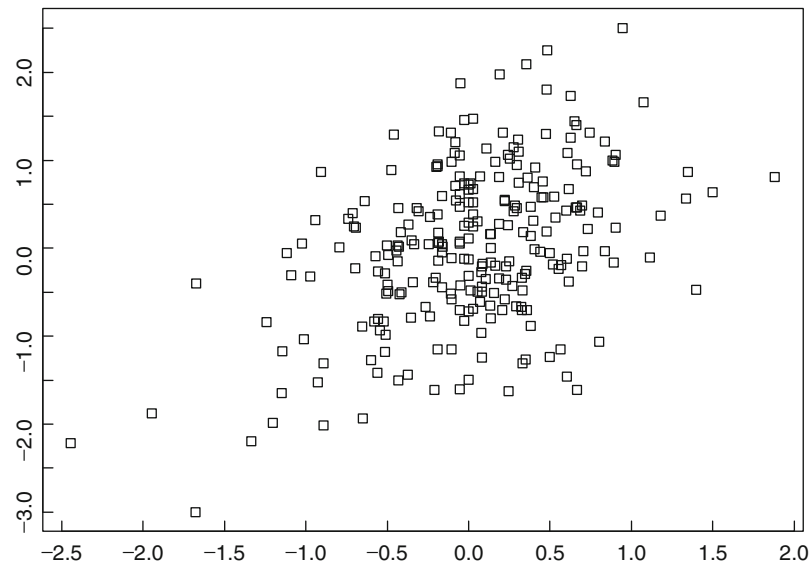


Figure 8-5
Scatterplot of $(x_{t-1,1}, x_{t,2})$, $t = 2, \dots, 250$, for the data in Example 8.1.1

$$D_{t1} - 0.0228 = Z_{t1} - 0.474Z_{t-1,1}, \quad \{Z_{t1}\} \sim \text{WN}(0, 0.0779), \quad (8.1.1)$$

$$D_{t2} - 0.838D_{t-1,2} - 0.0676 = Z_{t2} - 0.610Z_{t-1,2},$$

$$\{Z_{t2}\} \sim \text{WN}(0, 1.754), \quad (8.1.2)$$

provide good fits to the series $\{D_{t1}\}$ and $\{D_{t2}\}$.

The sample autocorrelations and cross-correlations of $\{D_{t1}\}$ and $\{D_{t2}\}$, are computed by opening the bivariate ITSM file LS2.TSM (as described in Example 8.1.1). The option `Transform>Difference`, with differencing lag equal to 1, generates

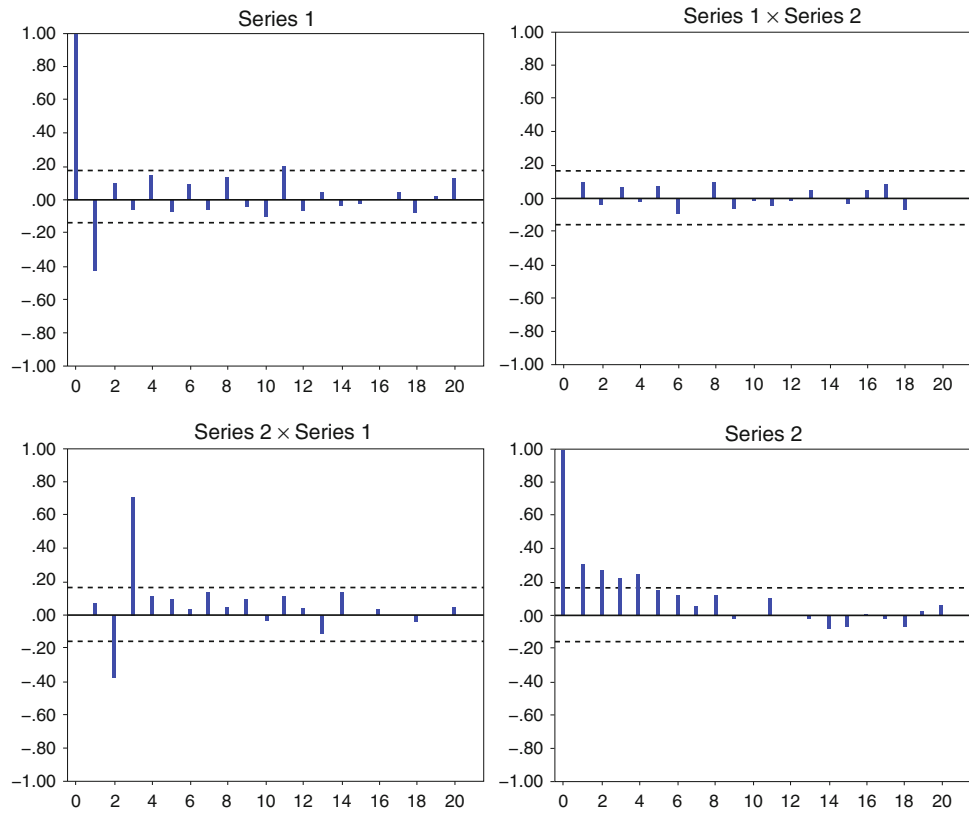


Figure 8-6
The sample correlations $\hat{\rho}_{ij}(h)$ of the series $\{D_{t1}\}$ and $\{D_{t2}\}$ of Example 8.1.2, showing the bounds $\pm 1.96n^{-1/2}$. ($\hat{\rho}_{ij}(h)$ is plotted as the j th graph in the i th row, $i, j = 1, 2$.)

the bivariate differenced series $\{(D_{t1}, D_{t2})\}$, and the correlation functions are then obtained as in Example 8.1.1 by clicking on the middle yellow button at the top of the ITSM screen. The sample auto- and cross-correlations $\hat{\rho}_{ij}(h)$, $i, j = 1, 2$, are shown in Figure 8-6. As we shall see in Section 8.3, care must be taken in interpreting the cross-correlations without first taking into account the autocorrelations of $\{D_{t1}\}$ and $\{D_{t2}\}$. □

8.2 Second-Order Properties of Multivariate Time Series

Consider m time series $\{X_{ti}, t = 0, \pm 1, \dots\}$, $i = 1, \dots, m$, with $EX_{ti}^2 < \infty$ for all t and i . If all the finite-dimensional distributions of the random variables $\{X_{ti}\}$ were multivariate normal, then the distributional properties of $\{X_{ti}\}$ would be completely determined by the means

$$\mu_{ti} := EX_{ti} \tag{8.2.1}$$

and the covariances

$$\gamma_{ij}(t+h, t) := E[(X_{t+h,i} - \mu_{ti})(X_{t,j} - \mu_{tj})]. \tag{8.2.2}$$

Even when the observations $\{X_{ti}\}$ do not have joint normal distributions, the quantities μ_{ti} and $\gamma_{ij}(t+h, t)$ specify the second-order properties, the covariances providing us with a measure of the dependence, not only between observations in the same series, but also between the observations in different series.

It is more convenient in dealing with m interrelated series to use vector notation. Thus we define

$$\mathbf{X}_t := \begin{bmatrix} X_{t1} \\ \vdots \\ X_{tm} \end{bmatrix}, \quad t = 0, \pm 1, \dots \quad (8.2.3)$$

The second-order properties of the multivariate time series $\{\mathbf{X}_t\}$ are then specified by the mean vectors

$$\boldsymbol{\mu}_t := E\mathbf{X}_t = \begin{bmatrix} \mu_{t1} \\ \vdots \\ \mu_{tm} \end{bmatrix} \quad (8.2.4)$$

and covariance matrices

$$\Gamma(t+h, t) := \begin{bmatrix} \gamma_{11}(t+h, t) & \cdots & \gamma_{1m}(t+h, t) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(t+h, t) & \cdots & \gamma_{mm}(t+h, t) \end{bmatrix}, \quad (8.2.5)$$

where

$$\gamma_{ij}(t+h, t) := \text{Cov}(X_{t+h,i}, X_{t,j}).$$

Remark 1. The matrix $\Gamma(t+h, t)$ can also be expressed as

$$\Gamma(t+h, t) := E[(\mathbf{X}_{t+h} - \boldsymbol{\mu}_{t+h})(\mathbf{X}_t - \boldsymbol{\mu}_t)'],$$

where as usual, the expected value of a random matrix A is the matrix whose components are the expected values of the components of A . \square

As in the univariate case, a particularly important role is played by the class of **multivariate stationary time series**, defined as follows.

Definition 8.2.1

The m -variate series $\{\mathbf{X}_t\}$ is (**weakly**) **stationary** if

(i) $\boldsymbol{\mu}_X(t)$ is independent of t

and

(ii) $\Gamma_X(t+h, t)$ is independent of t for each h .

For a stationary time series we shall use the notation

$$\boldsymbol{\mu} := E\mathbf{X}_t = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix} \quad (8.2.6)$$

and

$$\Gamma(h) := E[(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'] = \begin{bmatrix} \gamma_{11}(h) & \cdots & \gamma_{1m}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(h) & \cdots & \gamma_{mm}(h) \end{bmatrix}. \quad (8.2.7)$$

We shall refer to $\boldsymbol{\mu}$ as the mean of the series and to $\Gamma(h)$ as the covariance matrix at lag h . Notice that if $\{\mathbf{X}_t\}$ is stationary with covariance matrix function $\Gamma(\cdot)$, then for each i , $\{X_{ti}\}$ is stationary with covariance function $\gamma_{ii}(\cdot)$. The function $\gamma_{ij}(\cdot)$, $i \neq j$, is called the cross-covariance function of the two series $\{X_{ti}\}$ and $\{X_{tj}\}$. It should be noted

that $\gamma_{ij}(\cdot)$ is not in general the same as $\gamma_{ji}(\cdot)$. The correlation matrix function $R(\cdot)$ is defined by

$$R(h) := \begin{bmatrix} \rho_{11}(h) & \cdots & \rho_{1m}(h) \\ \vdots & \ddots & \vdots \\ \rho_{m1}(h) & \cdots & \rho_{mm}(h) \end{bmatrix}, \quad (8.2.8)$$

where $\rho_{ij}(h) = \gamma_{ij}(h)/[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$. The function $R(\cdot)$ is the covariance matrix function of the normalized series obtained by subtracting $\boldsymbol{\mu}$ from \mathbf{X}_t and then dividing each component by its standard deviation.

Example 8.2.1 Consider the bivariate stationary process $\{\mathbf{X}_t\}$ defined by

$$\begin{aligned} X_{t1} &= Z_t, \\ X_{t2} &= Z_t + 0.75Z_{t-10}, \end{aligned}$$

where $\{Z_t\} \sim \text{WN}(0, 1)$. Elementary calculations yield $\boldsymbol{\mu} = \mathbf{0}$,

$$\Gamma(-10) = \begin{bmatrix} 0 & 0.75 \\ 0 & 0.75 \end{bmatrix}, \quad \Gamma(0) = \begin{bmatrix} 1 & 1 \\ 1 & 1.5625 \end{bmatrix}, \quad \Gamma(10) = \begin{bmatrix} 0 & 0 \\ 0.75 & 0.75 \end{bmatrix},$$

and $\Gamma(j) = 0$ otherwise. The correlation matrix function is given by

$$R(-10) = \begin{bmatrix} 0 & 0.60 \\ 0 & 0.48 \end{bmatrix}, \quad R(0) = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}, \quad R(10) = \begin{bmatrix} 0 & 0 \\ 0.60 & 0.48 \end{bmatrix},$$

and $R(j) = 0$ otherwise. □

Basic Properties of $\Gamma(\cdot)$:

1. $\Gamma(h) = \Gamma'(-h)$,
2. $|\gamma_{ij}(h)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$, $i, j = 1, \dots, m$,
3. $\gamma_{ii}(\cdot)$ is an autocovariance function, $i = 1, \dots, m$, and
4. $\sum_{j,k=1}^n \mathbf{a}'_j \Gamma(j-k) \mathbf{a}_k \geq 0$ for all $n \in \{1, 2, \dots\}$ and $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$.

Proof The first property follows at once from the definition, the second from the fact that correlations cannot be greater than one in absolute value, and the third from the observation that $\gamma_{ii}(\cdot)$ is the autocovariance function of the stationary series $\{X_{it}, t = 0, \pm 1, \dots\}$. Property 4 is a statement of the obvious fact that

$$E\left(\sum_{j=1}^n \mathbf{a}'_j (\mathbf{X}_j - \boldsymbol{\mu})\right)^2 \geq 0. \quad \square$$

Remark 2. The basic properties of the matrices $\Gamma(h)$ are shared also by the corresponding matrices of correlations $R(h) = [\rho_{ij}(h)]_{i,j=1}^m$, which have the additional property

$$\rho_{ii}(0) = 1 \quad \text{for all } i.$$

The correlation $\rho_{ij}(0)$ is the correlation between X_{it} and X_{jt} , which is generally not equal to 1 if $i \neq j$ (see Example 8.2.1). It is also possible that $|\gamma_{ij}(h)| > |\gamma_{ij}(0)|$ if $i \neq j$ (see Problem 7.1). □

The simplest multivariate time series is multivariate white noise, the definition of which is quite analogous to that of univariate white noise.

Definition 8.2.2.

The m -variate series $\{\mathbf{Z}_t\}$ is called **white noise with mean $\mathbf{0}$ and covariance matrix \mathbb{Z}** , written

$$\{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \mathbb{Z}), \quad (8.2.9)$$

if $\{\mathbf{Z}_t\}$ is stationary with mean vector $\mathbf{0}$ and covariance matrix function

$$\Gamma(h) = \begin{cases} \mathbb{Z}, & \text{if } h = 0, \\ 0, & \text{otherwise.} \end{cases} \quad (8.2.10)$$

Definition 8.2.3.

The m -variate series $\{\mathbf{Z}_t\}$ is called **iid noise with mean $\mathbf{0}$ and covariance matrix \mathbb{Z}** , written

$$\{\mathbf{Z}_t\} \sim \text{iid}(\mathbf{0}, \mathbb{Z}), \quad (8.2.11)$$

if the random vectors $\{\mathbf{Z}_t\}$ are independent and identically distributed with mean $\mathbf{0}$ and covariance matrix \mathbb{Z} .

Multivariate white noise $\{\mathbf{Z}_t\}$ is used as a building block from which can be constructed an enormous variety of multivariate time series. The linear processes are generated as follows.

Definition 8.2.4.

The m -variate series $\{\mathbf{X}_t\}$ is a **linear process** if it has the representation

$$\mathbf{X}_t = \sum_{j=-\infty}^{\infty} C_j \mathbf{Z}_{t-j}, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \mathbb{Z}), \quad (8.2.12)$$

where $\{C_j\}$ is a sequence of $m \times m$ matrices whose components are absolutely summable.

The linear process (8.2.12) is stationary (Problem 7.2) with mean $\mathbf{0}$ and covariance function

$$\Gamma(h) = \sum_{j=-\infty}^{\infty} C_{j+h} \mathbb{Z} C_j', \quad h = 0, \pm 1, \dots \quad (8.2.13)$$

An **MA(∞) process** is a linear process with $C_j = 0$ for $j < 0$. Thus $\{\mathbf{X}_t\}$ is an MA(∞) process if and only if there exists a white noise sequence $\{\mathbf{Z}_t\}$ and a sequence of matrices C_j with absolutely summable components such that

$$\mathbf{X}_t = \sum_{j=0}^{\infty} C_j \mathbf{Z}_{t-j}.$$

Multivariate ARMA processes will be discussed in Section 8.4, where it will be shown in particular that any causal ARMA(p, q) process can be expressed as an MA(∞)

process, while any invertible ARMA(p, q) process can be expressed as an **AR(∞) process**, i.e., a process satisfying equations of the form

$$\mathbf{X}_t + \sum_{j=1}^{\infty} A_j \mathbf{X}_{t-j} = \mathbf{Z}_t,$$

in which the matrices A_j have absolutely summable components.

8.2.1 Second-Order Properties in the Frequency Domain

Provided that the components of the covariance matrix function $\Gamma(\cdot)$ have the property $\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty$, $i, j = 1, \dots, m$, then Γ has a **matrix-valued spectral density function**

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\lambda h} \Gamma(h), \quad -\pi \leq \lambda \leq \pi,$$

and Γ can be expressed in terms of f as

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda) d\lambda.$$

The second-order properties of the stationary process $\{\mathbf{X}_t\}$ can therefore be described equivalently in terms of $f(\cdot)$ rather than $\Gamma(\cdot)$. Similarly, $\{\mathbf{X}_t\}$ has a **spectral representation**

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{i\lambda t} d\mathbf{Z}(\lambda),$$

where $\{\mathbf{Z}(\lambda), -\pi \leq \lambda \leq \pi\}$ is a process whose components are complex-valued processes satisfying

$$E(dZ_j(\lambda)d\bar{Z}_k(\mu)) = \begin{cases} f_{jk}(\lambda)d\lambda & \text{if } \lambda = \mu, \\ 0 & \text{if } \lambda \neq \mu, \end{cases}$$

and \bar{Z}_k denotes the complex conjugate of Z_k . We shall not go into the spectral representation in this book. For details see Brockwell and Davis (1991).

8.3 Estimation of the Mean and Covariance Function

As in the univariate case, the estimation of the mean vector and covariances of a stationary multivariate time series plays an important role in describing and modeling the dependence structure of the component series. In this section we introduce estimators, for a stationary m -variate time series $\{\mathbf{X}_t\}$, of the components μ_j , $\gamma_{ij}(h)$, and $\rho_{ij}(h)$ of $\boldsymbol{\mu}$, $\Gamma(h)$, and $R(h)$, respectively. We also examine the large-sample properties of these estimators.

8.3.1 Estimation of $\boldsymbol{\mu}$

A natural unbiased estimator of the mean vector $\boldsymbol{\mu}$ based on the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ is the vector of sample means

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t.$$

The resulting estimate of the mean of the j th time series is then the univariate sample mean $(1/n) \sum_{t=1}^n X_{tj}$. If each of the univariate autocovariance functions $\gamma_{ii}(\cdot)$, $i = 1, \dots, m$, satisfies the conditions of Proposition 2.4.1, then the consistency of the estimator $\bar{\mathbf{X}}_n$ can be established by applying the proposition to each of the component time series $\{X_{ti}\}$. This immediately gives the following result.

Proposition 8.3.1. *If $\{\mathbf{X}_t\}$ is a stationary multivariate time series with mean $\boldsymbol{\mu}$ and covariance function $\Gamma(\cdot)$, then as $n \rightarrow \infty$,*

$$E(\bar{\mathbf{X}}_n - \boldsymbol{\mu})'(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \rightarrow 0 \quad \text{if } \gamma_{ii}(n) \rightarrow 0, \quad 1 \leq i \leq m,$$

and

$$nE(\bar{\mathbf{X}}_n - \boldsymbol{\mu})'(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \rightarrow \sum_{i=1}^m \sum_{h=-\infty}^{\infty} \gamma_{ii}(h) \quad \text{if } \sum_{h=-\infty}^{\infty} |\gamma_{ii}(h)| < \infty, \quad 1 \leq i \leq m.$$

Under more restrictive assumptions on the process $\{\mathbf{X}_t\}$ it can also be shown that $\bar{\mathbf{X}}_n$ is approximately normally distributed for large n . Determination of the covariance matrix of this distribution would allow us to obtain confidence regions for $\boldsymbol{\mu}$. However, this is quite complicated, and the following simple approximation is useful in practice.

For each i we construct a confidence interval for μ_i based on the sample mean \bar{X}_i of the univariate series X_{1i}, \dots, X_{ni} and combine these to form a confidence region for $\boldsymbol{\mu}$. If $f_i(\omega)$ is the spectral density of the i th process $\{X_{ti}\}$ and if the sample size n is large, then we know, under the same conditions as in Section 2.4, that $\sqrt{n}(\bar{X}_i - \mu_i)$ is approximately normally distributed with mean zero and variance

$$2\pi f_i(0) = \sum_{k=-\infty}^{\infty} \gamma_{ii}(k).$$

It can also be shown (see, e.g., Anderson 1971) that

$$2\pi \hat{f}_i(0) := \sum_{|h| \leq r} \left(1 - \frac{|h|}{r}\right) \hat{\gamma}_{ii}(h)$$

is a consistent estimator of $2\pi f_i(0)$, provided that $r = r_n$ is a sequence of numbers depending on n in such a way that $r_n \rightarrow \infty$ and $r_n/n \rightarrow 0$ as $n \rightarrow \infty$. Thus if \bar{X}_i denotes the sample mean of the i th process and Φ_α is the α -quantile of the standard normal distribution, then the bounds

$$\bar{X}_i \pm \Phi_{1-\alpha/2} \left(2\pi \hat{f}_i(0)/n\right)^{1/2}$$

are asymptotic $(1 - \alpha)$ confidence bounds for μ_i . Hence

$$\begin{aligned} P\left(|\mu_i - \bar{X}_i| \leq \Phi_{1-\alpha/2} \left(2\pi \hat{f}_i(0)/n\right)^{1/2}, i = 1, \dots, m\right) \\ \geq 1 - \sum_{i=1}^m P\left(|\mu_i - \bar{X}_i| > \Phi_{1-\alpha/2} \left(2\pi \hat{f}_i(0)/n\right)^{1/2}\right), \end{aligned}$$

where the right-hand side converges to $1 - m\alpha$ as $n \rightarrow \infty$. Consequently, as $n \rightarrow \infty$, the set of m -dimensional vectors bounded by

$$\left\{ x_i = \bar{X}_i \pm \Phi_{1-(\alpha/(2m))} \left(2\pi \hat{f}_i(0)/n \right)^{1/2}, i = 1, \dots, m \right\} \quad (8.3.1)$$

has a confidence coefficient that converges to a value *greater than or equal to* $1 - \alpha$ (and substantially greater if m is large). Nevertheless, the region defined by (8.3.1) is easy to determine and is of reasonable size, provided that m is not too large.

8.3.2 Estimation of $\Gamma(h)$

As in the univariate case, a natural estimator of the covariance $\Gamma(h) = E[(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})']$ is

$$\hat{\Gamma}(h) = \begin{cases} n^{-1} \sum_{t=1}^{n-h} (\mathbf{X}_{t+h} - \bar{\mathbf{X}}_n) (\mathbf{X}_t - \bar{\mathbf{X}}_n)' & \text{for } 0 \leq h \leq -1, \\ \hat{\Gamma}'(-h) & \text{for } -n+1 \leq h < 0. \end{cases}$$

Writing $\hat{\gamma}_{ij}(h)$ for the (i, j) -component of $\hat{\Gamma}(h)$, $i, j = 1, 2, \dots$, we estimate the cross-correlations by

$$\hat{\rho}_{ij}(h) = \hat{\gamma}_{ij}(h) (\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0))^{-1/2}.$$

If $i = j$, then $\hat{\rho}_{ij}$ reduces to the sample autocorrelation function of the i th series.

Derivation of the large-sample properties of $\hat{\gamma}_{ij}$ and $\hat{\rho}_{ij}$ is quite complicated in general. Here we shall simply note one result that is of particular importance for testing the independence of two component series. For details of the proof of this and related results, see Brockwell and Davis (1991).

Theorem 8.3.1. *Let $\{\mathbf{X}_t\}$ be the bivariate time series whose components are defined by*

$$X_{t1} = \sum_{k=-\infty}^{\infty} \alpha_k Z_{t-k,1}, \quad \{Z_{t1}\} \sim \text{IID}(0, \sigma_1^2),$$

and

$$X_{t2} = \sum_{k=-\infty}^{\infty} \beta_k Z_{t-k,2}, \quad \{Z_{t2}\} \sim \text{IID}(0, \sigma_2^2),$$

where the two sequences $\{Z_{t1}\}$ and $\{Z_{t2}\}$ are independent, $\sum_k |\alpha_k| < \infty$, and $\sum_k |\beta_k| < \infty$.

Then for all integers h and k with $h \neq k$, the random variables $n^{1/2} \hat{\rho}_{12}(h)$ and $n^{1/2} \hat{\rho}_{12}(k)$ are approximately bivariate normal with mean $\mathbf{0}$, variance $\sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j)$, and covariance $\sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j+k-h)$, for n large.

[For a related result that does not require the independence of the two series $\{X_{t1}\}$ and $\{X_{t2}\}$ see Bartlett's Formula, Section 8.3.4 below.]

Theorem 8.3.1 is useful in testing for correlation between two time series. If one of the two processes in the theorem is white noise, then it follows at once from the theorem that $\hat{\rho}_{12}(h)$ is approximately normally distributed with mean 0 and variance

$1/n$, in which case it is straightforward to test the hypothesis that $\rho_{12}(h) = 0$. However, if neither process is white noise, then a value of $\hat{\rho}_{12}(h)$ that is large relative to $n^{-1/2}$ does not necessarily indicate that $\rho_{12}(h)$ is different from zero. For example, suppose that $\{X_{t1}\}$ and $\{X_{t2}\}$ are two independent AR(1) processes with $\rho_{11}(h) = \rho_{22}(h) = 0.8^{|h|}$. Then the large-sample variance of $\hat{\rho}_{12}(h)$ is $n^{-1} (1 + 2 \sum_{k=1}^{\infty} (0.64)^k) = 4.556n^{-1}$. It would therefore not be surprising to observe a value of $\hat{\rho}_{12}(h)$ as large as $3n^{-1/2}$ even though $\{X_{t1}\}$ and $\{X_{t2}\}$ are independent. If on the other hand, $\rho_{11}(h) = 0.8^{|h|}$ and $\rho_{22}(h) = (-0.8)^{|h|}$, then the large-sample variance of $\hat{\rho}_{12}(h)$ is $0.2195n^{-1}$, and an observed value of $3n^{-1/2}$ for $\hat{\rho}_{12}(h)$ would be very unlikely.

8.3.3 Testing for Independence of Two Stationary Time Series

Since by Theorem 8.3.1 the large-sample distribution of $\hat{\rho}_{12}(h)$ depends on both $\rho_{11}(\cdot)$ and $\rho_{22}(\cdot)$, any test for independence of the two component series cannot be based solely on estimated values of $\rho_{12}(h)$, $h = 0, \pm 1, \dots$, without taking into account the nature of the two component series.

This difficulty can be circumvented by “prewhitening” the two series before computing the cross-correlations $\hat{\rho}_{12}(h)$, i.e., by transforming the two series to white noise by application of suitable filters. If $\{X_{t1}\}$ and $\{X_{t2}\}$ are invertible ARMA (p, q) processes, this can be achieved by the transformations

$$Z_{ti} = \sum_{j=0}^{\infty} \pi_j^{(i)} X_{t-j,i},$$

where $\sum_{j=0}^{\infty} \pi_j^{(i)} z^j = \phi^{(i)}(z)/\theta^{(i)}(z)$ and $\phi^{(i)}, \theta^{(i)}$ are the autoregressive and moving-average polynomials of the i th series, $i = 1, 2$.

Since in practice the true model is nearly always unknown and since the data X_{tj} , $t \leq 0$, are not available, it is convenient to replace the sequences $\{Z_{ti}\}$ by the residuals $\{\hat{W}_{ti}\}$ after fitting a maximum likelihood ARMA model to each of the component series (see (5.3.1)). If the fitted ARMA models were in fact the true models, the series $\{\hat{W}_{ti}\}$ would be white noise sequences for $i = 1, 2$.

To test the hypothesis H_0 that $\{X_{t1}\}$ and $\{X_{t2}\}$ are independent series, we observe that under H_0 , the corresponding two prewhitened series $\{Z_{t1}\}$ and $\{Z_{t2}\}$ are also independent. Theorem 8.3.1 then implies that the sample cross-correlations $\hat{\rho}_{12}(h), \hat{\rho}_{12}(k)$, $h \neq k$, of $\{Z_{t1}\}$ and $\{Z_{t2}\}$ are for large n approximately independent and normally distributed with means 0 and variances n^{-1} . An approximate test for independence can therefore be obtained by comparing the values of $|\hat{\rho}_{12}(h)|$ with $1.96n^{-1/2}$, exactly as in Section 5.3.2. If we prewhiten only one of the two original series, say $\{X_{t1}\}$, then under H_0 Theorem 8.3.1 implies that the sample cross-correlations $\tilde{\rho}_{12}(h), \tilde{\rho}_{12}(k)$, $h \neq k$, of $\{Z_{t1}\}$ and $\{X_{t2}\}$ are for large n approximately normal with means 0, variances n^{-1} and covariance $n^{-1}\rho_{22}(k-h)$, where $\rho_{22}(\cdot)$ is the autocorrelation function of $\{X_{t2}\}$. Hence, for any fixed h , $\tilde{\rho}_{12}(h)$ also falls (under H_0) between the bounds $\pm 1.96n^{-1/2}$ with a probability of approximately 0.95.

Example 8.3.1. The sample correlation functions $\hat{\rho}_{ij}(\cdot)$, $i, j = 1, 2$, of the bivariate time series E731A.TSM (of length $n = 200$) are shown in Figure 8-7. Without taking into account the autocorrelations $\hat{\rho}_{ii}(\cdot)$, $i = 1, 2$, it is impossible to decide on the basis of the cross-correlations whether or not the two component processes are independent

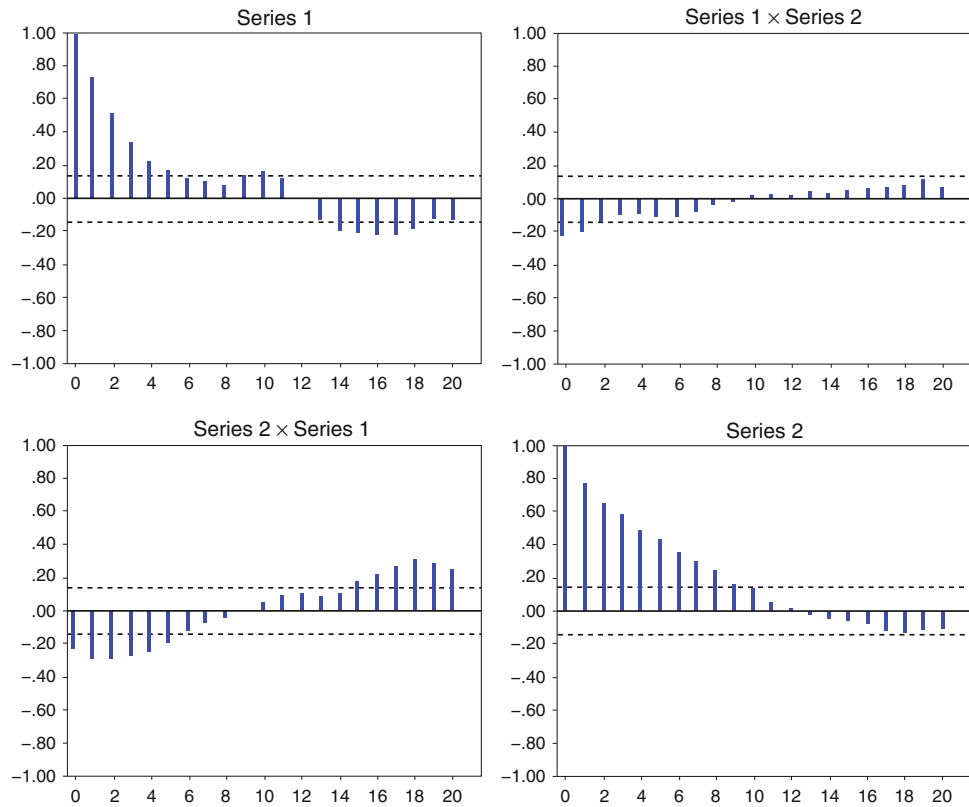


Figure 8-7
The sample correlations
of the bivariate series
E731A.TSM of
Example 8.3.1, showing the
bounds $\pm 1.96n^{-1/2}$

of each other. Notice that many of the sample cross-correlations $\hat{\rho}_{ij}(h)$, $i \neq j$, lie outside the bounds $\pm 1.96n^{-1/2} = \pm 0.139$. However, these bounds are relevant only if at least one of the component series is white noise. Since this is clearly not the case, a whitening transformation must be applied to at least one of the two component series. Analysis using ITSM leads to AR(1) models for each. The residuals from these maximum likelihood models are stored as a bivariate series in the file E731B.TSM, and their sample correlations, obtained from ITSM, are shown in Figure 8-8. All but two of the cross-correlations are between the bounds ± 0.139 , suggesting by Theorem 8.3.1 that the two residual series (and hence the two original series) are uncorrelated. The data for this example were in fact generated as two independent AR(1) series with $\phi = 0.8$ and $\sigma^2 = 1$.

□

8.3.4 Bartlett's Formula

In Section 2.4 we gave Bartlett's formula for the large-sample distribution of the sample autocorrelation vector $\hat{\rho} = (\hat{\rho}(1), \dots, \hat{\rho}(k))'$ of a univariate time series. The following theorem gives a large-sample approximation to the covariances of the sample cross-correlations $\hat{\rho}_{12}(h)$ and $\hat{\rho}_{12}(k)$ of the bivariate time series $\{\mathbf{X}_t\}$ under the assumption that $\{\mathbf{X}_t\}$ is Gaussian. However, it is not assumed (as in Theorem 8.3.1) that $\{X_{t1}\}$ is independent of $\{X_{t2}\}$.

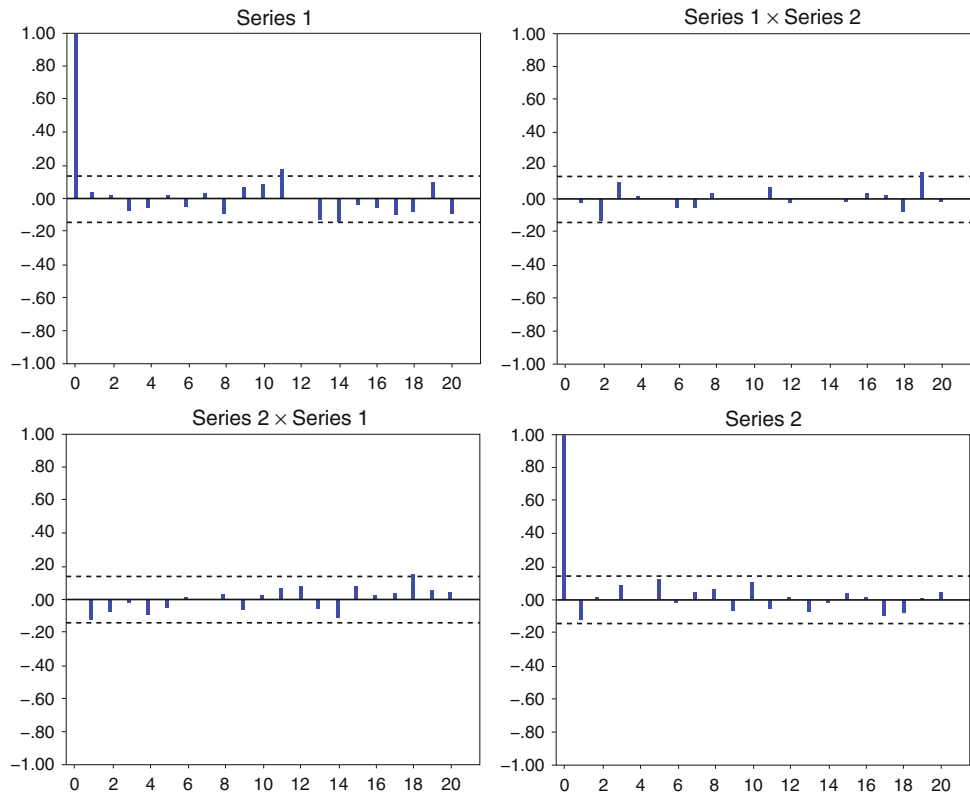


Figure 8-8
 The sample correlations of the bivariate series of residuals E731B.TSM, whose components are the residuals from the AR(1) models fitted to each of the component series in E731A.TSM

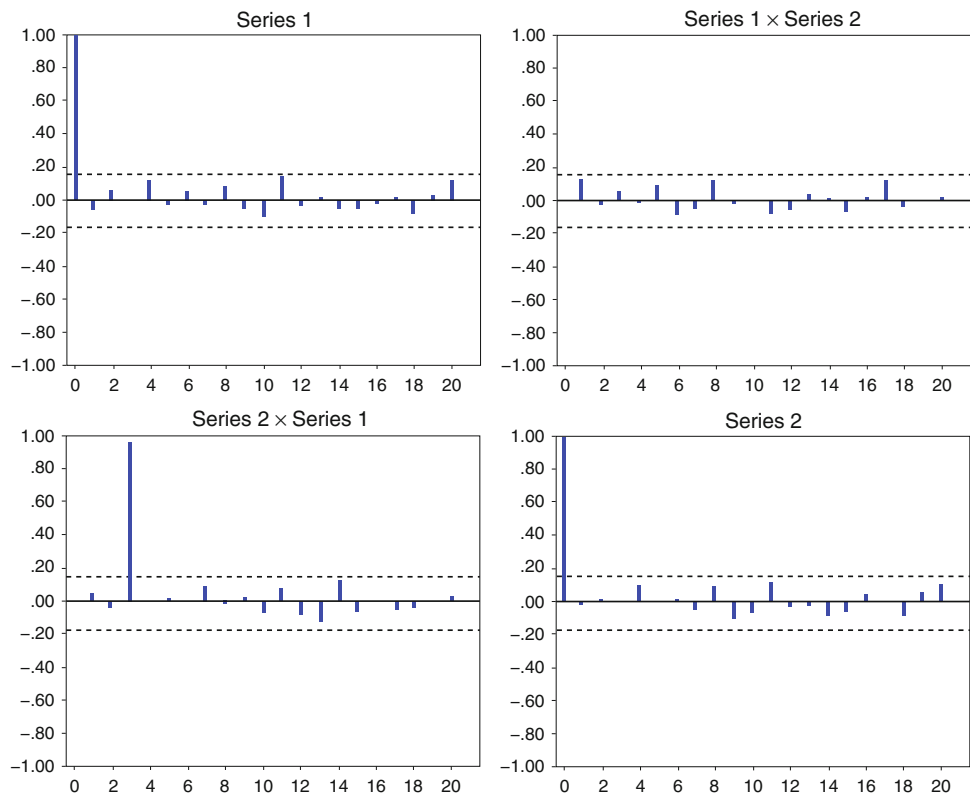


Figure 8-9
 The sample correlations of the whitened series $\hat{W}_{t+h,1}$ and $\hat{W}_{t,2}$ of Example 8.3.2, showing the bounds $\pm 1.96n^{-1/2}$

Bartlett's Formula:

If $\{\mathbf{X}_t\}$ is a bivariate Gaussian time series with covariances satisfying $\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty$, $i, j = 1, 2$, then

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}(\hat{\rho}_{12}(h), \hat{\rho}_{12}(k)) = & \sum_{j=-\infty}^{\infty} \left[\rho_{11}(j) \rho_{22}(j+k-h) + \rho_{12}(j+k) \rho_{21}(j-h) \right. \\ & - \rho_{12}(h) \{ \rho_{11}(j) \rho_{12}(j+k) + \rho_{22}(j) \rho_{21}(j-k) \} \\ & - \rho_{12}(k) \{ \rho_{11}(j) \rho_{12}(j+h) + \rho_{22}(j) \rho_{21}(j-h) \} \\ & \left. + \rho_{12}(h) \rho_{12}(k) \left\{ \frac{1}{2} \rho_{11}^2(j) + \rho_{12}^2(j) + \frac{1}{2} \rho_{22}^2(j) \right\} \right] \end{aligned}$$

Corollary 8.3.1. *If $\{\mathbf{X}_t\}$ satisfies the conditions for Bartlett's formula, if either $\{X_{t1}\}$ or $\{X_{t2}\}$ is white noise, and if*

$$\rho_{12}(h) = 0, \quad h \notin [a, b],$$

then

$$\lim_{n \rightarrow \infty} n \text{Var}(\hat{\rho}_{12}(h)) = 1, \quad h \notin [a, b].$$

Example 8.3.2. Sales with a leading indicator

We consider again the differenced series $\{D_{t1}\}$ and $\{D_{t2}\}$ of Example 8.1.2, for which we found the maximum likelihood models (8.1.1) and (8.1.2) using ITSM. The residuals from the two models (which can be fitted by ITSM) are the two “whitened” series $\{\hat{W}_{t1}\}$ and $\{\hat{W}_{t2}\}$ with sample variances 0.0779 and 1.754, respectively. This bivariate series is contained in the file E732.TSM.

The sample auto- and cross-correlations of $\{D_{t1}\}$ and $\{D_{t2}\}$ were shown in Figure 8-6. Without taking into account the autocorrelations, it is not possible to draw any conclusions about the dependence between the two component series from the cross-correlations.

Examination of the sample cross-correlation function of the whitened series $\{\hat{W}_{t1}\}$ and $\{\hat{W}_{t2}\}$, on the other hand, is much more informative. From Figure 8-9 it is apparent that there is one large-sample cross-correlation (between $\hat{W}_{t+3,2}$ and $\hat{W}_{t,1}$), while the others are all between $\pm 1.96n^{-1/2}$. □

If $\{\hat{W}_{t1}\}$ and $\{\hat{W}_{t2}\}$ are assumed to be jointly Gaussian, Corollary 8.3.1 indicates the compatibility of the cross-correlations with a model for which

$$\rho_{12}(-3) \neq 0$$

and

$$\rho_{12}(h) = 0, \quad h \neq -3.$$

The value $\hat{\rho}_{12}(-3) = 0.969$ suggests the model

$$\hat{W}_{t2} = 4.74 \hat{W}_{t-3,1} + N_t, \tag{8.3.2}$$

where the stationary noise $\{N_t\}$ has small variance compared with $\{\hat{W}_{t2}\}$ and $\{\hat{W}_{t1}\}$, and the coefficient 4.74 is the square root of the ratio of sample variances of $\{\hat{W}_{t2}\}$ and

$\{\hat{W}_{t1}\}$. A study of the sample values of $\{\hat{W}_{t2} - 4.74\hat{W}_{t-3,1}\}$ suggests the model

$$(1 + 0.345B)N_t = U_t, \quad \{U_t\} \sim \text{WN}(0, 0.0782) \quad (8.3.3)$$

for $\{N_t\}$. Finally, replacing \hat{W}_{t2} and $\hat{W}_{t-3,1}$ in (8.3.2) by Z_{t2} and $Z_{t-3,1}$, respectively, and then using (8.1.1) and (8.1.2) to express Z_{t2} and $Z_{t-3,1}$ in terms of $\{D_{t2}\}$ and $\{D_{t1}\}$, we obtain a model relating $\{D_{t1}\}$, $\{D_{t2}\}$, and $\{U_{t1}\}$, namely,

$$D_{t2} + 0.0773 = (1 - 0.610B)(1 - 0.838B)^{-1}[4.74(1 - 0.474B)^{-1}D_{t-3,1} + (1 + 0.345B)^{-1}U_{t1}].$$

This model should be compared with the one derived later in Section 11.1 by the more systematic technique of transfer function modeling.

8.4 Multivariate ARMA Processes

As in the univariate case, we can define an extremely useful class of multivariate stationary processes $\{\mathbf{X}_t\}$ by requiring that $\{\mathbf{X}_t\}$ should satisfy a set of linear difference equations with constant coefficients. Multivariate white noise $\{\mathbf{Z}_t\}$ (see Definition 8.2.2) is a fundamental building block from which these ARMA processes are constructed.

Definition 8.4.1.

$\{\mathbf{X}_t\}$ is an **ARMA(p, q) process** if $\{\mathbf{X}_t\}$ is stationary and if for every t ,

$$\mathbf{X}_t - \Phi_1\mathbf{X}_{t-1} - \cdots - \Phi_p\mathbf{X}_{t-p} = \mathbf{Z}_t + \Theta_1\mathbf{Z}_{t-1} + \cdots + \Theta_q\mathbf{Z}_{t-q}, \quad (8.4.1)$$

where $\{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma)$. ($\{\mathbf{X}_t\}$ is an **ARMA(p, q) process with mean μ** if $\{\mathbf{X}_t - \mu\}$ is an ARMA(p, q) process.)

Equations (8.4.1) can be written in the more compact form

$$\Phi(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma), \quad (8.4.2)$$

where $\Phi(z) := I - \Phi_1z - \cdots - \Phi_pz^p$ and $\Theta(z) := I + \Theta_1z + \cdots + \Theta_qz^q$ are matrix-valued polynomials, I is the $m \times m$ identity matrix, and B as usual denotes the backward shift operator. (Each component of the matrices $\Phi(z)$, $\Theta(z)$ is a polynomial with real coefficients and degree less than or equal to p, q , respectively.)

Example 8.4.1.

The multivariate AR(1) process

Setting $p = 1$ and $q = 0$ in (8.4.1) gives the defining equations

$$\mathbf{X}_t = \Phi\mathbf{X}_{t-1} + \mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma), \quad (8.4.3)$$

for the multivariate AR(1) series $\{\mathbf{X}_t\}$. By exactly the same argument as used in Example 2.2.1, we can express \mathbf{X}_t as

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Phi^j \mathbf{Z}_{t-j}, \quad (8.4.4)$$

provided that all the eigenvalues of Φ are less than 1 in absolute value, i.e., provided that

$$\det(I - z\Phi) \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ such that } |z| \leq 1. \quad (8.4.5)$$

If this condition is satisfied, then the coefficients Φ^j are absolutely summable, and hence the series in (8.4.4) converges; i.e., each component of the matrix $\sum_{j=0}^n \Phi^j \mathbf{Z}_{t-j}$ converges (see Remark 1 of Section 2.2). The same argument as in Example 2.2.1 also shows that (8.4.4) is the unique stationary solution of (8.4.3). The condition that all the eigenvalues of Φ should be less than 1 in absolute value (or equivalently (8.4.5)) is just the multivariate analogue of the condition $|\phi| < 1$ required for the existence of a causal stationary solution of the univariate AR(1) equations (2.2.8). \square

Causality and invertibility of a multivariate ARMA(p, q) process are defined precisely as in Section 3.1, except that the coefficients ψ_j, π_j in the representations $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ and $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ are replaced by $m \times m$ matrices Ψ_j and Π_j whose components are required to be absolutely summable. The following two theorems (proofs of which can be found in Brockwell and Davis (1991)) provide us with criteria for causality and invertibility analogous to those of Section 3.1.

Causality:

An ARMA(p, q) process $\{\mathbf{X}_t\}$ is **causal**, or a **causal function of $\{\mathbf{Z}_t\}$** , if there exist matrices $\{\Psi_j\}$ with absolutely summable components such that

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{Z}_{t-j} \quad \text{for all } t. \quad (8.4.6)$$

Causality is equivalent to the condition

$$\det \Phi(z) \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ such that } |z| \leq 1. \quad (8.4.7)$$

The matrices Ψ_j are found recursively from the equations

$$\Psi_j = \Theta_j + \sum_{k=1}^{\infty} \Phi_k \Psi_{j-k}, \quad j = 0, 1, \dots, \quad (8.4.8)$$

where we define $\Theta_0 = I$, $\Theta_j = 0$ for $j > q$, $\Phi_j = 0$ for $j > p$, and $\Psi_j = 0$ for $j < 0$.

Invertibility:

An ARMA(p, q) process $\{\mathbf{X}_t\}$ is **invertible** if there exist matrices $\{\Pi_j\}$ with absolutely summable components such that

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \Pi_j \mathbf{X}_{t-j} \quad \text{for all } t. \quad (8.4.9)$$

Invertibility is equivalent to the condition

$$\det \Theta(z) \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ such that } |z| \leq 1. \quad (8.4.10)$$

The matrices Π_j are found recursively from the equations

$$\Pi_j = -\Phi_j - \sum_{k=1}^{\infty} \Theta_k \Pi_{j-k}, \quad j = 0, 1, \dots, \quad (8.4.11)$$

where we define $\Phi_0 = -I$, $\Phi_j = 0$ for $j > p$, $\Theta_j = 0$ for $j > q$, and $\Pi_j = 0$ for $j < 0$.

Example 8.4.2. For the multivariate AR(1) process defined by (8.4.3), the recursions (8.4.8) give

$$\begin{aligned}\Psi_0 &= I, \\ \Psi_1 &= \Phi \Psi_0 = \Phi, \\ \Psi_2 &= \Phi \Psi_1 = \Phi^2, \\ &\vdots \\ \Psi_j &= \Phi \Psi_{j-1} = \Phi^j, \quad j \geq 3,\end{aligned}$$

as already found in Example 8.4.1. □

Remark 3. For the bivariate AR(1) process (8.4.3) with

$$\Phi = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \end{bmatrix}$$

it is easy to check that $\Psi_j = \Phi^j = 0$ for $j > 1$ and hence that $\{\mathbf{X}_t\}$ has the alternative representation

$$\mathbf{X}_t = \mathbf{Z}_t + \Phi \mathbf{Z}_{t-1}$$

as an MA(1) process. This example shows that it is not always possible to distinguish between multivariate ARMA models of different orders without imposing further restrictions. If, for example, attention is restricted to pure AR processes, the problem does not arise. For detailed accounts of the identification problem for general ARMA(p, q) models see Hannan and Deistler (1988) and Lütkepohl (1993). □

8.4.1 The Covariance Matrix Function of a Causal ARMA Process

From (8.2.13) we can express the covariance matrix $\Gamma(h) = E(\mathbf{X}_{t+h}\mathbf{X}'_t)$ of the causal process (8.4.6) as

$$\Gamma(h) = \sum_{j=0}^{\infty} \Psi_{h+j} \Sigma \Psi'_j, \quad h = 0, \pm 1, \dots, \quad (8.4.12)$$

where the matrices Ψ_j are found from (8.4.8) and $\Psi_j := 0$ for $j < 0$.

The covariance matrices $\Gamma(h)$, $h = 0, \pm 1, \dots$, can also be found by solving the Yule–Walker equations

$$\Gamma(j) - \sum_{r=1}^p \Phi_r \Gamma(j-r) = \sum_{j \leq r \leq q} \Theta_r \Sigma \Psi'_{r-j}, \quad j = 0, 1, 2, \dots, \quad (8.4.13)$$

obtained by postmultiplying (8.4.1) by \mathbf{X}'_{t-j} and taking expectations. The first $p+1$ of the equation (8.4.13) can be solved for the components of $\Gamma(0), \dots, \Gamma(p)$ using the fact that $\Gamma(-h) = \Gamma'(h)$. The remaining equations then give $\Gamma(p+1), \Gamma(p+2), \dots$ recursively. An explicit form of the solution of these equations can be written down by making use of Kronecker products and the vec operator (see e.g., Lütkepohl 1993).

Remark 4. If z_0 is the root of $\det \Phi(z) = 0$ with smallest absolute value, then it can be shown from the recursions (8.4.8) that $\Psi_j/r^j \rightarrow 0$ as $j \rightarrow \infty$ for all r such that $|z_0|^{-1} < r < 1$. Hence, there is a constant C such that each component of Ψ_j is smaller

in absolute value than Cr^j . This implies in turn that there is a constant K such that each component of the matrix $\Psi_{h+j}\Sigma\Psi_j'$ on the right of (8.4.12) is bounded in absolute value by Kr^{2j} . Provided that $|z_0|$ is not very close to 1, this means that the series (8.4.12) converges rapidly, and the error incurred in each component by truncating the series after the term with $j = k - 1$ is smaller in absolute value than $\sum_{j=k}^{\infty} Kr^{2j} = Kr^{2k} / (1 - r^2)$.

8.5 Best Linear Predictors of Second-Order Random Vectors

Let $\{\mathbf{X}_t = (X_{t1}, \dots, X_{tm})'\}$ be an m -variate time series with means $E\mathbf{X}_t = \boldsymbol{\mu}_t$ and covariance function given by the $m \times m$ matrices

$$K(i, j) = E(\mathbf{X}_i\mathbf{X}_j') - \boldsymbol{\mu}_i\boldsymbol{\mu}_j'.$$

If $\mathbf{Y} = (Y_1, \dots, Y_m)'$ is a random vector with finite second moments and $E\mathbf{Y} = \boldsymbol{\mu}$, we define

$$P_n(\mathbf{Y}) = (P_n Y_1, \dots, P_n Y_m)', \quad (8.5.1)$$

where $P_n Y_j$ is the best linear predictor of the component Y_j of \mathbf{Y} in terms of all of the components of the vectors \mathbf{X}_t , $t = 1, \dots, n$, and the constant 1. It follows immediately from the properties of the prediction operator (Section 2.5) that

$$P_n(\mathbf{Y}) = \boldsymbol{\mu} + A_1(\mathbf{X}_n - \boldsymbol{\mu}_n) + \dots + A_n(\mathbf{X}_1 - \boldsymbol{\mu}_1) \quad (8.5.2)$$

for some matrices A_1, \dots, A_n , and that

$$\mathbf{Y} - P_n(\mathbf{Y}) \perp \mathbf{X}_{n+1-i}, \quad i = 1, \dots, n, \quad (8.5.3)$$

where we say that two m -dimensional random vectors \mathbf{X} and \mathbf{Y} are orthogonal (written $\mathbf{X} \perp \mathbf{Y}$) if $E(\mathbf{X}\mathbf{Y}')$ is a matrix of zeros. The vector of best predictors (8.5.1) is uniquely determined by (8.5.2) and (8.5.3), although it is possible that there may be more than one possible choice for A_1, \dots, A_n .

As a special case of the above, if $\{\mathbf{X}_t\}$ is a zero-mean time series, the best linear predictor $\hat{\mathbf{X}}_{n+1}$ of \mathbf{X}_{n+1} in terms of $\mathbf{X}_1, \dots, \mathbf{X}_n$ is obtained on replacing \mathbf{Y} by \mathbf{X}_{n+1} in (8.5.1). Thus

$$\hat{\mathbf{X}}_{n+1} = \begin{cases} \mathbf{0}, & \text{if } n = 0, \\ P_n(\mathbf{X}_{n+1}), & \text{if } n \geq 1. \end{cases}$$

Hence, we can write

$$\hat{\mathbf{X}}_{n+1} = \Phi_{n1}\mathbf{X}_n + \dots + \Phi_{nn}\mathbf{X}_1, \quad n = 1, 2, \dots, \quad (8.5.4)$$

where, from (8.5.3), the coefficients Φ_{nj} , $j = 1, \dots, n$, are such that

$$E(\hat{\mathbf{X}}_{n+1}\mathbf{X}'_{n+1-i}) = E(\mathbf{X}_{n+1}\mathbf{X}'_{n+1-i}), \quad i = 1, \dots, n, \quad (8.5.5)$$

i.e.,

$$\sum_{j=1}^n \Phi_{nj}K(n+1-j, n+1-i) = K(n+1, n+1-i), \quad i = 1, \dots, n.$$

In the case where $\{\mathbf{X}_t\}$ is stationary with $K(i, j) = \Gamma(i - j)$, the prediction equations simplify to the m -dimensional analogues of (2.5.7), i.e.,

$$\sum_{j=1}^n \Phi_{nj}\Gamma(i-j) = \Gamma(i), \quad i = 1, \dots, n. \quad (8.5.6)$$

Provided that the covariance matrix of the nm components of $\mathbf{X}_1, \dots, \mathbf{X}_n$ is nonsingular for every $n \geq 1$, the coefficients $\{\Phi_{nj}\}$ can be determined recursively using a multivariate version of the Durbin–Levinson algorithm given by Whittle (1963) (for details see Brockwell and Davis (1991), Proposition 11.4.1). Whittle’s recursions also determine the covariance matrices of the one-step prediction errors, namely, $V_0 = \Gamma(0)$ and, for $n \geq 1$,

$$\begin{aligned} V_n &= E(\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1})(\mathbf{X}_{n+1} - \hat{\mathbf{X}}_{n+1})' \\ &= \Gamma(0) - \Phi_{n1}\Gamma(-1) - \dots - \Phi_{nn}\Gamma(-n). \end{aligned} \quad (8.5.7)$$

Remark 5. The innovations algorithm also has a multivariate version that can be used for prediction in much the same way as the univariate version described in Section 2.5.4 (for details see Brockwell and Davis (1991), Proposition 11.4.2). \square

8.6 Modeling and Forecasting with Multivariate AR Processes

If $\{\mathbf{X}_t\}$ is any zero-mean second-order multivariate time series, it is easy to show from the results of Section 8.5 (Problem 8.4) that the one-step prediction errors $\mathbf{X}_j - \hat{\mathbf{X}}_j$, $j = 1, \dots, n$, have the property

$$E(\mathbf{X}_j - \hat{\mathbf{X}}_j)(\mathbf{X}_k - \hat{\mathbf{X}}_k)' = 0 \text{ for } j \neq k. \quad (8.6.1)$$

Moreover, the matrix M such that

$$\begin{bmatrix} \mathbf{X}_1 - \hat{\mathbf{X}}_1 \\ \mathbf{X}_2 - \hat{\mathbf{X}}_2 \\ \mathbf{X}_3 - \hat{\mathbf{X}}_3 \\ \vdots \\ \mathbf{X}_n - \hat{\mathbf{X}}_n \end{bmatrix} = M \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \quad (8.6.2)$$

is lower triangular with ones on the diagonal and therefore has determinant equal to 1.

If the series $\{\mathbf{X}_t\}$ is also Gaussian, then (8.6.1) implies that the prediction errors $\mathbf{U}_j = \mathbf{X}_j - \hat{\mathbf{X}}_j$, $j = 1, \dots, n$, are independent with covariance matrices V_0, \dots, V_{n-1} , respectively (as specified in (8.5.7)). Consequently, the joint density of the prediction errors is the product

$$f(\mathbf{u}_1, \dots, \mathbf{u}_n) = (2\pi)^{-nm/2} \left(\prod_{j=1}^n \det V_{j-1} \right)^{-1/2} \exp \left[-\frac{1}{2} \sum_{j=1}^n \mathbf{u}_j' V_{j-1}^{-1} \mathbf{u}_j \right].$$

Since the determinant of the matrix M in (8.6.2) is equal to 1, the joint density of the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ at $\mathbf{x}_1, \dots, \mathbf{x}_n$ is obtained on replacing $\mathbf{u}_1, \dots, \mathbf{u}_n$ in the last expression by the values of $\mathbf{X}_j - \hat{\mathbf{X}}_j$ corresponding to the observations $\mathbf{x}_1, \dots, \mathbf{x}_n$.

If we suppose that $\{\mathbf{X}_t\}$ is a zero-mean m -variate AR(p) process with coefficient matrices $\Phi = \{\Phi_1, \dots, \Phi_p\}$ and white noise covariance matrix Σ , we can therefore express the likelihood of the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ as

$$L(\Phi, \Sigma) = (2\pi)^{-nm/2} \left(\prod_{j=1}^n \det V_{j-1} \right)^{-1/2} \exp \left[-\frac{1}{2} \sum_{j=1}^n \mathbf{U}_j' V_{j-1}^{-1} \mathbf{U}_j \right],$$

where $\mathbf{U}_j = \mathbf{X}_j - \hat{\mathbf{X}}_j$, $j = 1, \dots, n$, and $\hat{\mathbf{X}}_j$ and V_j are found from (8.5.4), (8.5.6), and (8.5.7).

Maximization of the Gaussian likelihood is much more difficult in the multivariate than in the univariate case because of the potentially large number of parameters involved and the fact that it is not possible to compute the maximum likelihood estimator of Φ independently of Σ as in the univariate case. In principle, maximum likelihood estimators can be computed with the aid of efficient nonlinear optimization algorithms, but it is important to begin the search with preliminary estimates that are reasonably close to the maximum. For pure AR processes good preliminary estimates can be obtained using Whittle's algorithm or a multivariate version of Burg's algorithm given by Jones (1978). We shall restrict our discussion here to the use of Whittle's algorithm (the multivariate option `AR-Model>Estimation>Yule-Walker` in ITSM), but Jones's multivariate version of Burg's algorithm is also available (`AR-Model>Estimation>Burg`). Other useful algorithms can be found in Lütkepohl (1993), in particular the method of conditional least squares and the method of Hannan and Rissanen (1982), the latter being useful also for preliminary estimation in the more difficult problem of fitting ARMA(p, q) models with $q > 0$. Spectral methods of estimation for multivariate ARMA processes are also frequently used. A discussion of these (as well as some time-domain methods) is given in Anderson (1980).

Order selection for multivariate autoregressive models can be made by minimizing a multivariate analogue of the univariate AICC statistic

$$\text{AICC} = -2 \ln L(\Phi_1, \dots, \Phi_p, \Sigma) + \frac{2(pm^2 + 1)nm}{nm - pm^2 - 2}. \quad (8.6.3)$$

8.6.1 Estimation for Autoregressive Processes Using Whittle's Algorithm

If $\{\mathbf{X}_t\}$ is the (causal) multivariate AR(p) process defined by the difference equations

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma), \quad (8.6.4)$$

then postmultiplying by \mathbf{X}'_{t-j} , $j = 0, \dots, p$, and taking expectations gives the equations

$$\Sigma = \Gamma(0) - \sum_{j=1}^p \Phi_j \Gamma(-j) \quad (8.6.5)$$

and

$$\Gamma(i) = \sum_{j=1}^n \Phi_j \Gamma(i-j), \quad i = 1, \dots, p. \quad (8.6.6)$$

Given the matrices $\Gamma(0), \dots, \Gamma(p)$, equation (8.6.6) can be used to determine the coefficient matrices Φ_1, \dots, Φ_p . The white noise covariance matrix Σ can then be found from (8.6.5). The solution of these equations for Φ_1, \dots, Φ_p , and Σ is identical to the solution of (8.5.6) and (8.5.7) for the prediction coefficient matrices $\Phi_{p1}, \dots, \Phi_{pp}$ and the corresponding prediction error covariance matrix V_p . Consequently, Whittle's algorithm can be used to carry out the algebra.

The Yule-Walker estimators $\hat{\Phi}_1, \dots, \hat{\Phi}_p$, and $\hat{\Sigma}$ for the model (8.6.4) fitted to the data $\mathbf{X}_1, \dots, \mathbf{X}_n$ are obtained by replacing $\Gamma(j)$ in (8.6.5) and (8.6.6) by $\hat{\Gamma}(j)$, $j = 0, \dots, p$, and solving the resulting equations for Φ_1, \dots, Φ_p , and Σ . The solution of these equations is obtained from ITSM by selecting the multivariate option `AR-Model>Estimation>Yule-Walker`. The mean vector of the fitted model is the sample mean of the data, and Whittle's algorithm is used to solve the

equations (8.6.5) and (8.6.6) for the coefficient matrices and the white noise covariance matrix. The fitted model is displayed by ITSM in the form

$$\mathbf{X}_t = \phi_0 + \Phi_1 \mathbf{X}_{t-1} + \cdots + \Phi_p \mathbf{X}_{t-p} + \mathbf{Z}_t, \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma).$$

Note that the mean $\boldsymbol{\mu}$ of this model is not the vector ϕ_0 , but

$$\boldsymbol{\mu} = (I - \Phi_1 - \cdots - \Phi_p)^{-1} \phi_0.$$

In fitting multivariate autoregressive models using ITSM, check the box Find minimum AICC model to find the AR(p) model with $0 \leq p \leq 20$ that minimizes the AICC value as defined in (8.6.3).

Analogous calculations using Jones's multivariate version of Burg's algorithm can be carried out by selecting AR-Model>Estimation>Burg.

Example 8.6.1 The Dow Jones and All Ordinaries Indices

To find the minimum AICC Yule–Walker model (of order less than or equal to 20) for the bivariate series $\{(X_{t1}, X_{t2})', t = 1, \dots, 250\}$ of Example 8.1.1, proceed as follows. Select File>Project>Open>Multivariate, click OK, and then double-click on the file name, DJAOPC2.TSM. Check that Number of columns is set to 2, the dimension of the observation vectors, and click OK again to see graphs of the two component time series. No differencing is required (recalling from Example 8.1.1 that $\{X_{t1}\}$ and $\{X_{t2}\}$ are the daily percentage price changes of the original Dow Jones and All Ordinaries Indices). Select AR-Model>Estimation>Yule-Walker, check the box Find minimum AICC Model, click OK, and you will obtain the model

$$\begin{bmatrix} X_{t1} \\ X_{t2} \end{bmatrix} = \begin{bmatrix} 0.0288 \\ 0.00836 \end{bmatrix} + \begin{bmatrix} -0.0148 & 0.0357 \\ 0.6589 & 0.0998 \end{bmatrix} \begin{bmatrix} X_{t-1,1} \\ X_{t-1,2} \end{bmatrix} + \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix},$$

where

$$\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} \sim \text{WN} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.3653 & 0.0224 \\ 0.0224 & 0.6016 \end{bmatrix} \right).$$

□

Example 8.6.2 Sales with a leading indicator

The series $\{Y_{t1}\}$ (leading indicator) and $\{Y_{t2}\}$ (sales) are stored in bivariate form (Y_{t1} in column 1 and Y_{t2} in column 2) in the file LS2.TSM. On opening this file in ITSM you will see the graphs of the two component time series. Inspection of the graphs immediately suggests, as in Example 8.2.2, that the differencing operator $\nabla = 1 - B$ should be applied to the data before a stationary AR model is fitted. Select Transform>Difference and specify 1 for the differencing lag. Click OK and you will see the graphs of the two differenced series. Inspection of the series and their correlation functions (obtained by pressing the second yellow button at the top of the ITSM window) suggests that no further differencing is necessary. The next step is to select AR-model>Estimation>Yule-Walker with the option Find minimum AICC model. The resulting model has order $p = 5$ and parameters $\phi_0 = (0.0328 \ 0.0156)'$,

$$\hat{\phi}_1 = \begin{bmatrix} -0.517 & 0.024 \\ -0.019 & -0.051 \end{bmatrix}, \hat{\phi}_2 = \begin{bmatrix} -0.192 & -0.018 \\ 0.047 & 0.250 \end{bmatrix}, \hat{\phi}_3 = \begin{bmatrix} -0.073 & 0.010 \\ 4.678 & 0.207 \end{bmatrix},$$

$$\hat{\Phi}_4 = \begin{bmatrix} -0.032 & -0.009 \\ 3.664 & 0.004 \end{bmatrix}, \hat{\Phi}_5 = \begin{bmatrix} 0.022 & 0.011 \\ 1.300 & 0.029 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 0.076 & -0.003 \\ -0.003 & 0.095 \end{bmatrix},$$

with AICC=109.49. (Analogous calculations using Burg's algorithm give an AR(8) model for the differenced series.) The sample cross-correlations of the residual vectors $\hat{\mathbf{Z}}_t$ can be plotted by clicking on the last blue button at the top of the ITSM window. These are nearly all within the bounds $\pm 1.96/\sqrt{n}$, suggesting that the model is a good fit. The components of the residual vectors themselves are plotted by selecting AR Model>Residual Analysis>Plot Residuals. Simulated observations from the fitted model can be generated using the option AR Model>Simulate. The fitted model has the interesting property that the upper right component of each of the coefficient matrices is close to zero. This suggests that $\{X_{t1}\}$ can be effectively modeled independently of $\{X_{t2}\}$. In fact, the MA(1) model

$$X_{t1} = (1 - 0.474B)U_t, \quad \{U_t\} \sim \text{WN}(0, 0.0779), \quad (8.6.7)$$

provides an adequate fit to the univariate series $\{X_{t1}\}$. Inspecting the bottom rows of the coefficient matrices and deleting small entries, we find that the relation between $\{X_{t1}\}$ and $\{X_{t2}\}$ can be expressed approximately as

$$X_{t2} = 0.250X_{t-2,2} + 0.207X_{t-3,2} + 4.678X_{t-3,1} + 3.664X_{t-4,1} + 1.300X_{t-5,1} + W_t,$$

or equivalently,

$$X_{t2} = \frac{4.678B^3(1 + 0.783B + 0.278B^2)}{1 - 0.250B^2 - 0.207B^3}X_{t1} + \frac{W_t}{1 - 0.250B^2 - 0.207B^3}, \quad (8.6.8)$$

where $\{W_t\} \sim \text{WN}(0, 0.095)$. Moreover, since the estimated noise covariance matrix is essentially diagonal, it follows that the two sequences $\{X_{t1}\}$ and $\{W_t\}$ are uncorrelated. This reduced model defined by (8.6.7) and (8.6.8) is an example of a transfer function model that expresses the “output” series $\{X_{t2}\}$ as the output of a linear filter with “input” $\{X_{t1}\}$ plus added noise. A more direct approach to the fitting of transfer function models is given in Section 11.1 and applied to this same data set. \square

8.6.2 Forecasting Multivariate Autoregressive Processes

The technique developed in Section 8.5 allows us to compute the minimum mean squared error one-step linear predictors $\hat{\mathbf{X}}_{n+1}$ for any multivariate stationary time series from the mean $\boldsymbol{\mu}$ and autocovariance matrices $\Gamma(h)$ by recursively determining the coefficients Φ_{ni} , $i = 1, \dots, n$, and evaluating

$$\hat{\mathbf{X}}_{n+1} = \boldsymbol{\mu} + \Phi_{n1}(\mathbf{X}_n - \boldsymbol{\mu}) + \dots + \Phi_{nm}(\mathbf{X}_1 - \boldsymbol{\mu}). \quad (8.6.9)$$

The situation is simplified when $\{\mathbf{X}_t\}$ is the causal AR(p) process defined by (8.6.4), since for $n \geq p$ (as is almost always the case in practice)

$$\hat{\mathbf{X}}_{n+1} = \Phi_1\mathbf{X}_n + \dots + \Phi_p\mathbf{X}_{n+1-p}. \quad (8.6.10)$$

To verify (8.6.10) it suffices to observe that the right-hand side has the required form (8.5.2) and that the prediction error

$$\mathbf{X}_{n+1} - \Phi_1\mathbf{X}_n - \dots - \Phi_p\mathbf{X}_{n+1-p} = \mathbf{Z}_{n+1}$$

is orthogonal to $\mathbf{X}_1, \dots, \mathbf{X}_n$ in the sense of (8.5.3). (In fact, the prediction error is orthogonal to *all* \mathbf{X}_j , $-\infty < j \leq n$, showing that if $n \geq p$, then (8.6.10) is also the best linear predictor of \mathbf{X}_{n+1} in terms of all components of \mathbf{X}_j , $-\infty < j \leq n$.) The covariance matrix of the one-step prediction error is clearly $E(\mathbf{Z}_{n+1}\mathbf{Z}'_{n+1}) = \mathbb{A}$.

To compute the best h -step linear predictor $P_n\mathbf{X}_{n+h}$ based on all the components of $\mathbf{X}_1, \dots, \mathbf{X}_n$ we apply the linear operator P_n to (8.6.4) to obtain the recursions

$$P_n\mathbf{X}_{n+h} = \Phi_1 P_n\mathbf{X}_{n+h-1} + \dots + \Phi_p P_n\mathbf{X}_{n+h-p}. \quad (8.6.11)$$

These equations are easily solved recursively, first for $P_n\mathbf{X}_{n+1}$, then for $P_n\mathbf{X}_{n+2}$, $P_n\mathbf{X}_{n+3}$, \dots , etc. If $n \geq p$, then the h -step predictors based on all components of \mathbf{X}_j , $-\infty < j \leq n$, also satisfy (8.6.11) and are therefore the same as the h -step predictors based on $\mathbf{X}_1, \dots, \mathbf{X}_n$.

To compute the h -step error covariance matrices, recall from (8.4.6) that

$$\mathbf{X}_{n+h} = \sum_{j=0}^{\infty} \Psi_j \mathbf{Z}_{n+h-j}, \quad (8.6.12)$$

where the coefficient matrices Ψ_j are found from the recursions (8.4.8) with $q = 0$. From (8.6.12) we find that for $n \geq p$,

$$P_n\mathbf{X}_{n+h} = \sum_{j=h}^{\infty} \Psi_j \mathbf{Z}_{n+h-j}. \quad (8.6.13)$$

Subtracting (8.6.13) from (8.6.12) gives the h -step prediction error

$$\mathbf{X}_{n+h} - P_n\mathbf{X}_{n+h} = \sum_{j=0}^{h-1} \Psi_j \mathbf{Z}_{n+h-j}, \quad (8.6.14)$$

with covariance matrix

$$E[(\mathbf{X}_{n+h} - P_n\mathbf{X}_{n+h})(\mathbf{X}_{n+h} - P_n\mathbf{X}_{n+h})'] = \sum_{j=0}^{h-1} \Psi_j \mathbb{A} \Psi_j', \quad n \geq p. \quad (8.6.15)$$

For the (not necessarily zero-mean) causal AR(p) process defined by

$$\mathbf{X}_t = \phi_0 + \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \mathbb{A}),$$

Equations (8.6.10) and (8.6.11) remain valid, provided that ϕ_0 is added to each of their right-hand sides. The error covariance matrices are the same as in the case $\phi_0 = \mathbf{0}$.

The above calculations are all based on the assumption that the AR(p) model for the series is known. However, in practice, the parameters of the model are usually estimated from the data, and the uncertainty in the predicted values of the series will be larger than indicated by (8.6.15) because of parameter estimation errors. See Lütkepohl (1993).

Example 8.6.3 The Dow Jones and All Ordinaries Indices

The VAR(1) model fitted to the series $\{\mathbf{X}_t, t = 1, \dots, 250\}$ in Example 8.6.1 was

$$\begin{bmatrix} X_{t1} \\ X_{t2} \end{bmatrix} = \begin{bmatrix} 0.0288 \\ 0.00836 \end{bmatrix} + \begin{bmatrix} -0.0148 & 0.0357 \\ 0.6589 & 0.0998 \end{bmatrix} \begin{bmatrix} X_{t-1,1} \\ X_{t-1,2} \end{bmatrix} + \begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix},$$

where

$$\begin{bmatrix} Z_{t1} \\ Z_{t2} \end{bmatrix} \sim \text{WN} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.3653 & 0.0224 \\ 0.0224 & 0.6016 \end{bmatrix} \right).$$

The one-step mean squared error for prediction of X_{t2} , assuming the validity of this model, is thus 0.6016. This is a substantial reduction from the estimated mean squared error $\hat{\gamma}_{22}(0) = 0.7712$ when the sample mean $\hat{\mu}_2 = 0.0309$ is used as the one-step predictor.

If we fit a univariate model to the series $\{X_{t2}\}$ using ITSM, we find that the autoregression with minimum AICC value (645.0) is

$$X_{t2} = 0.0273 + 0.1180X_{t-1,2} + Z_t, \quad \{Z_t\} \sim \text{WN}(0, 0.7604).$$

Assuming the validity of this model, we thus obtain a mean squared error for one-step prediction of 0.7604, which is slightly less than the estimated mean squared error (0.7712) incurred when the sample mean is used for one-step prediction.

The preceding calculations suggest that there is little to be gained from the point of view of one-step prediction by fitting a univariate model to $\{X_{t2}\}$, while there is a substantial reduction achieved by the bivariate AR(1) model for $\{\mathbf{X}_t = (X_{t1}, X_{t2})'\}$.

To test the models fitted above, we consider the next forty values $\{\mathbf{X}_t, t = 251, \dots, 290\}$, which are stored in the file DJAOPCF.TSM. We can use these values, in conjunction with the bivariate and univariate models fitted to the data for $t = 1, \dots, 250$, to compute one-step predictors of X_{t2} , $t = 251, \dots, 290$. The results are as follows:

Predictor	Average Squared Error
$\hat{\mu} = 0.0309$	0.4706
AR(1)	0.4591
VAR(1)	0.3962

It is clear from these results that the sample variance of the series $\{X_{t2}, t = 251, \dots, 290\}$ is rather less than that of the series $\{X_{t2}, t = 1, \dots, 250\}$, and consequently, the average squared errors of all three predictors are substantially less than expected from the models fitted to the latter series. Both the AR(1) and VAR(1) models show an improvement in one-step average squared error over the sample mean $\hat{\mu}$, but the improvement shown by the bivariate model is much more pronounced. \square

The calculation of predictors and their error covariance matrices for multivariate ARIMA and SARIMA processes is analogous to the corresponding univariate calculation, so we shall simply state the pertinent results. Suppose that $\{\mathbf{Y}_t\}$ is a nonstationary process satisfying $D(B)\mathbf{Y}_t = \mathbf{U}_t$ where $D(z) = 1 - d_1z - \dots - d_rz^r$ is a polynomial with $D(1) = 0$ and $\{\mathbf{U}_t\}$ is a causal invertible ARMA process with mean $\boldsymbol{\mu}$. Then $\mathbf{X}_t = \mathbf{U}_t - \boldsymbol{\mu}$ satisfies

$$\Phi(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \Sigma). \quad (8.6.16)$$

Under the assumption that the random vectors $\mathbf{Y}_{-r+1}, \dots, \mathbf{Y}_0$ are uncorrelated with the sequence $\{\mathbf{Z}_t\}$, the best linear predictors $\tilde{P}_n\mathbf{Y}_j$ of $\mathbf{Y}_j, j > n > 0$, based on 1 and the components of $\mathbf{Y}_j, -r+1, \leq j \leq n$, are found as follows. Compute the observed values of $\mathbf{U}_t = D(B)\mathbf{Y}_t, t = 1, \dots, n$, and use the ARMA model for $\mathbf{X}_t = \mathbf{U}_t - \boldsymbol{\mu}$ to compute predictors $P_n\mathbf{U}_{n+h}$. Then use the recursions

$$\tilde{P}_n\mathbf{Y}_{n+h} = P_n\mathbf{U}_{n+h} + \sum_{j=1}^r d_j\tilde{P}_n\mathbf{Y}_{n+h-j} \quad (8.6.17)$$

to compute successively $\tilde{P}_n \mathbf{Y}_{n+1}$, $\tilde{P}_n \mathbf{Y}_{n+2}$, $\tilde{P}_n \mathbf{Y}_{n+3}$, etc. The error covariance matrices are approximately (for large n)

$$E \left[(\mathbf{Y}_{n+h} - \tilde{P}_n \mathbf{Y}_{n+h})(\mathbf{Y}_{n+h} - \tilde{P}_n \mathbf{Y}_{n+h})' \right] = \sum_{j=0}^{h-1} \Psi_j^* \Sigma \Psi_j^{*'}, \quad (8.6.18)$$

where Ψ_j^* is the coefficient of z^j in the power series expansion

$$\sum_{j=0}^{\infty} \Psi_j^* z^j = D(z)^{-1} \Phi^{-1}(z) \Theta(z), \quad |z| < 1.$$

The matrices Ψ_j^* are most readily found from the recursions (8.4.8) after replacing Φ_j , $j = 1, \dots, p$, by Φ_j^* , $j = 1, \dots, p + r$, where Φ_j^* is the coefficient of z^j in $D(z)\Phi(z)$.

Remark 6. In the special case where $\Theta(z) = I$ (i.e., in the purely autoregressive case) the expression (8.6.18) for the h -step error covariance matrix is exact for all $n \geq p$ (i.e., if there are at least $p + r$ observed vectors). The program ITSM allows differencing transformations and subtraction of the mean before fitting a multivariate autoregression. Predicted values for the original series and the standard deviations of the prediction errors can be determined using the multivariate option `Forecasting>AR Model`. \square

Remark 7. In the multivariate case, simple differencing of the type discussed in this section where the same operator $D(B)$ is applied to all components of the random vectors is rather restrictive. It is useful to consider more general linear transformations of the data for the purpose of generating a stationary series. Such considerations lead to the class of **cointegrated models** discussed briefly in Section 8.7 below. \square

Example 8.6.4 Sales with a leading indicator

Assume that the model fitted to the bivariate series $\{\mathbf{Y}_t, t = 0, \dots, 149\}$ in Example 8.6.2 is correct, i.e., that

$$\Phi(B)\mathbf{X}_t = \mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \hat{\Sigma}),$$

where

$$\mathbf{X}_t = (1 - B)\mathbf{Y}_t - (0.0228, 0.420)', \quad t = 1, \dots, 149,$$

$\Phi(B) = I - \hat{\Phi}_1 B - \dots - \hat{\Phi}_5 B^5$, and $\hat{\Phi}_1, \dots, \hat{\Phi}_5, \hat{\Sigma}$ are the matrices found in Example 8.6.2. Then the one- and two-step predictors of \mathbf{X}_{150} and \mathbf{X}_{151} are obtained from (8.6.11) as

$$P_{149}\mathbf{X}_{150} = \hat{\Phi}_1 \mathbf{X}_{149} + \dots + \hat{\Phi}_5 \mathbf{X}_{145} = \begin{bmatrix} 0.163 \\ -0.217 \end{bmatrix}$$

and

$$P_{149}\mathbf{X}_{151} = \hat{\Phi}_1 P_{149}\mathbf{X}_{150} + \hat{\Phi}_2 \mathbf{X}_{149} + \dots + \hat{\Phi}_5 \mathbf{X}_{146} = \begin{bmatrix} -0.027 \\ 0.816 \end{bmatrix}$$

with error covariance matrices, from (8.6.15),

$$\Sigma = \begin{bmatrix} 0.076 & -0.003 \\ -0.003 & 0.095 \end{bmatrix}$$

and

$$\Sigma + \hat{\phi}_1 \Sigma \hat{\phi}_1' = \begin{bmatrix} 0.096 & -0.002 \\ -0.002 & 0.095 \end{bmatrix},$$

respectively.

Similarly, the one- and two-step predictors of \mathbf{Y}_{150} and \mathbf{Y}_{151} are obtained from (8.6.17) as

$$\tilde{P}_{149} \mathbf{Y}_{150} = \begin{bmatrix} 0.0228 \\ 0.420 \end{bmatrix} + P_{149} \mathbf{X}_{150} + \mathbf{Y}_{149} = \begin{bmatrix} 13.59 \\ 262.90 \end{bmatrix}$$

and

$$\tilde{P}_{149} \mathbf{Y}_{151} = \begin{bmatrix} 0.0228 \\ 0.420 \end{bmatrix} + P_{149} \mathbf{X}_{151} + \tilde{P}_{149} \mathbf{Y}_{150} = \begin{bmatrix} 13.59 \\ 264.14 \end{bmatrix}$$

with error covariance matrices, from (8.6.18),

$$\Sigma = \begin{bmatrix} 0.076 & -0.003 \\ -0.003 & 0.095 \end{bmatrix}$$

and

$$\Sigma + (I + \hat{\phi}_1) \Sigma (I + \hat{\phi}_1)' = \begin{bmatrix} 0.094 & -0.003 \\ -0.003 & 0.181 \end{bmatrix},$$

respectively. The predicted values and the standard deviations of the predictors can easily be verified with the aid of the program ITSM. It is also of interest to compare the results with those obtained by fitting a transfer function model to the data as described in Section 11.1 below. □

8.7 Cointegration

We have seen that nonstationary univariate time series can frequently be made stationary by applying the differencing operator $\nabla = 1 - B$ repeatedly. If $\{\nabla^d X_t\}$ is stationary for some positive integer d but $\{\nabla^{d-1} X_t\}$ is nonstationary, we say that $\{X_t\}$ is **integrated of order d** , or more concisely, $\{X_t\} \sim I(d)$. Many macroeconomic time series are found to be integrated of order 1.

If $\{\mathbf{X}_t\}$ is a k -variate time series, we define $\{\nabla^d \mathbf{X}_t\}$ to be the series whose j th component is obtained by applying the operator $(1 - B)^d$ to the j th component of $\{\mathbf{X}_t\}$, $j = 1, \dots, k$. The idea of a cointegrated multivariate time series was introduced by Granger (1981) and developed by Engle and Granger (1987). Here we use the slightly different definition of Lütkepohl (1993). We say that the k -dimensional time series $\{\mathbf{X}_t\}$ is integrated of order d (or $\{\mathbf{X}_t\} \sim I(d)$) if d is a positive integer, $\{\nabla^d \mathbf{X}_t\}$ is stationary, and $\{\nabla^{d-1} \mathbf{X}_t\}$ is nonstationary. The $I(d)$ process $\{\mathbf{X}_t\}$ is said to be **cointegrated with cointegration vector α** if α is a $k \times 1$ vector such that $\{\alpha' \mathbf{X}_t\}$ is of order less than d .

Example 8.7.1 A simple example is provided by the bivariate process whose first component is the random walk

$$X_t = \sum_{j=1}^t Z_j, \quad t = 1, 2, \dots, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2),$$

and whose second component consists of noisy observations of the same random walk,

$$Y_t = X_t + W_t, \quad t = 1, 2, \dots, \quad \{W_t\} \sim \text{IID}(0, \tau^2),$$

where $\{W_t\}$ is independent of $\{Z_t\}$. Then $\{(X_t, Y_t)'\}$ is integrated of order 1 and cointegrated with cointegration vector $\alpha = (1, -1)'$.

The notion of cointegration captures the idea of univariate nonstationary time series “moving together.” Thus, even though $\{X_t\}$ and $\{Y_t\}$ in Example 8.7.1 are both nonstationary, they are linked in the sense that they differ only by the stationary sequence $\{W_t\}$. Series that behave in a cointegrated manner are often encountered in economics. Engle and Granger (1991) give as an illustrative example the prices of tomatoes U_t and V_t in Northern and Southern California. These are linked by the fact that if one were to increase sufficiently relative to the other, the profitability of buying in one market and selling for a profit in the other would tend to push the prices $(U_t, V_t)'$ toward the straight line $v = u$ in \mathbb{R}^2 . This line is said to be an **attractor** for $(U_t, V_t)'$, since although U_t and V_t may both vary in a nonstationary manner as t increases, the points $(U_t, V_t)'$ will exhibit relatively small random deviations from the line $v = u$. \square

Example 8.7.2 If we apply the operator $\nabla = 1 - B$ to the bivariate process defined in Example 8.7.1 in order to render it stationary, we obtain the series $(U_t, V_t)'$, where

$$U_t = Z_t$$

and

$$V_t = Z_t + W_t - W_{t-1}.$$

The series $\{(U_t, V_t)'\}$ is clearly a stationary multivariate MA(1) process

$$\begin{bmatrix} U_t \\ V_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Z_t \\ Z_t + W_t \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-1} + W_{t-1} \end{bmatrix}.$$

However, the process $\{(U_t, V_t)'\}$ cannot be represented as an AR(∞) process, since the matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - z \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix}$ has zero determinant when $z = 1$, thus violating condition (8.4.10). Care is therefore needed in the estimation of parameters for such models (and the closely related error-correction models). We shall not go into the details here but refer the reader to Engle and Granger (1987) and Lütkepohl (1993). \square

Problems

8.1 Let $\{Y_t\}$ be a stationary process and define the bivariate process $X_{t1} = Y_t$, $X_{t2} = Y_{t-d}$, where $d \neq 0$. Show that $\{(X_{t1}, X_{t2})'\}$ is stationary and express its cross-correlation function in terms of the autocorrelation function of $\{Y_t\}$. If $\rho_Y(h) \rightarrow 0$ as $h \rightarrow \infty$, show that there exists a lag k for which $\rho_{12}(k) > \rho_{12}(0)$.

8.2 Show that the covariance matrix function of the multivariate linear process defined by (8.2.12) is as specified in (8.2.13).

8.3 Let $\{X_t\}$ be the bivariate time series whose components are the MA(1) processes defined by

$$X_{t1} = Z_{t,1} + 0.8Z_{t-1,1}, \quad \{Z_{t1}\} \sim \text{IID}(0, \sigma_1^2),$$

and

$$X_{t2} = Z_{t,2} - 0.6Z_{t-1,2}, \quad \{Z_{t2}\} \sim \text{IID}(0, \sigma_2^2),$$

where the two sequences $\{Z_{t1}\}$ and $\{Z_{t2}\}$ are independent.

- Find a large-sample approximation to the variance of $n^{1/2}\hat{\rho}_{12}(h)$.
- Find a large-sample approximation to the covariance of $n^{1/2}\hat{\rho}_{12}(h)$ and $n^{1/2}\hat{\rho}_{12}(k)$ for $h \neq k$.

8.4 Use the characterization (8.5.3) of the multivariate best linear predictor of \mathbf{Y} in terms of $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ to establish the orthogonality of the one-step prediction errors $\mathbf{X}_j - \hat{\mathbf{X}}_j$ and $\mathbf{X}_k - \hat{\mathbf{X}}_k$, $j \neq k$, as asserted in (8.6.1).

8.5 Determine the covariance matrix function of the ARMA(1,1) process satisfying

$$\mathbf{X}_t - \Phi\mathbf{X}_{t-1} = \mathbf{Z}_t + \Theta\mathbf{Z}_{t-1}, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, I_2),$$

where I_2 is the 2×2 identity matrix and $\Phi = \Theta' = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{bmatrix}$.

8.6 a. Let $\{\mathbf{X}_t\}$ be a causal AR(p) process satisfying the recursions

$$\mathbf{X}_t = \Phi_1\mathbf{X}_{t-1} + \dots + \Phi_p\mathbf{X}_{t-p} + \mathbf{Z}_t, \quad \{\mathbf{Z}_t\} \sim \text{WN}(\mathbf{0}, \mathbb{Z}).$$

For $n \geq p$ write down recursions for the predictors $P_n\mathbf{X}_{n+h}$, $h \geq 0$, and find explicit expressions for the error covariance matrices in terms of the AR coefficients and \mathbb{Z} when $h = 1, 2$, and 3.

- Suppose now that $\{\mathbf{Y}_t\}$ is the multivariate ARIMA($p, 1, 0$) process satisfying $\nabla\mathbf{Y}_t = \mathbf{X}_t$, where $\{\mathbf{X}_t\}$ is the AR process in (a). Assuming that $E(\mathbf{Y}_0\mathbf{X}_t') = 0$, for $t \geq 1$, show (using (8.6.17) with $r = 1$ and $d = 1$) that

$$\tilde{P}_n(\mathbf{Y}_{n+h}) = \mathbf{Y}_n + \sum_{j=1}^h P_n\mathbf{X}_{n+j},$$

and derive the error covariance matrices when $h = 1, 2$, and 3. Compare these results with those obtained in Example 8.6.4.

8.7 Use the program ITSM to find the minimum AICC AR model of order less than or equal to 20 for the bivariate series $\{(X_{t1}, X_{t2})', t = 1, \dots, 200\}$ with components filed as APPJK2.TSM. Use the fitted model to predict $(X_{t1}, X_{t2})'$, $t = 201, 202, 203$ and estimate the error covariance matrices of the predictors (assuming that the fitted model is appropriate for the data).

8.8 Let $\{X_{t1}, t = 1, \dots, 63\}$ and $\{X_{t2}, t = 1, \dots, 63\}$ denote the differenced series $\{\nabla \ln Y_{t1}\}$ and $\{\nabla \ln Y_{t2}\}$, where $\{Y_{t1}\}$ and $\{Y_{t2}\}$ are the annual mink and muskrat trappings filed as APPH.TSM and APPL.TSM, respectively).

- Use ITSM to construct and save the series $\{X_{t1}\}$ and $\{X_{t2}\}$ as univariate data files X1.TSM and X2.TSM, respectively. (After making the required transformations press the red EXP button and save each transformed series to a file with the appropriate name.) To enter X1 and X2 as a bivariate series in ITSM, open X1 as a multivariate series with Number of columns equal to 1. Then open X2 as a univariate series. Click the project editor button (at the top left of the ITSM window), click on the plus signs next to the projects X1.TSM and X2.TSM, then click on the series that appears just below X2.TSM and drag it to the first line of the project X1.TSM. It will then be added as a second component, making X1.TSM a bivariate project consisting of the two component series X1 and X2. Click OK to close the project editor and close

the ITSM window labeled X2.TSM. You will then see the graphs of X_1 and X_2 . Press the second yellow button to see the correlation functions of $\{X_{t1}\}$ and $\{X_{t2}\}$. For more information on the project editor in ITSM consult the Project Editor section of the PDF file ITSM_HELP.

b. Conduct a test for independence of the two series $\{X_{t1}\}$ and $\{X_{t1}\}$.

- 8.9** Use ITSM to open the data file STOCK7.TSM, which contains the daily returns on seven different stock market indices from April 27th, 1998, through April 9th, 1999. (Consult the Data Sets section of the PDF file ITSM_HELP for more information.) Fit a multivariate autoregression to the trivariate series consisting of the returns on the Dow Jones Industrials, All Ordinaries, and Nikkei indices. Check the model for goodness of fit and interpret the results.