CHAPTER 1

Multivariate Linear Time Series

1.1 INTRODUCTION

Multivariate time series analysis considers simultaneously multiple time series. It is a branch of multivariate statistical analysis but deals specifically with dependent data. It is, in general, much more complicated than the univariate time series analysis, especially when the number of series considered is large. We study this more complicated statistical analysis in this book because in real life decisions often involve multiple inter-related factors or variables. Understanding the relationships between those factors and providing accurate predictions of those variables are valuable in decision making. The objectives of multivariate time series analysis thus include

1. To study the dynamic relationships between variables
2. To improve the accuracy of prediction

Let $\boldsymbol{z}_t = (z_{1t}, \cdots, z_{kt})'$ be a k-dimensional time series observed at equally spaced time points. For example, let z_{1t} be the quarterly U.S. real gross domestic product (GDP) and z_{2t} the quarterly U.S. civilian unemployment rate. By studying z_{1t} and z_{2t} jointly, we can assess the temporal and contemporaneous dependence between GDP and unemployment rate. In this particular case, $k = 2$ and the two variables are known to be instantaneously negatively correlated. Figure 1.1 shows the time plots of quarterly U.S. real GDP (in logarithm of billions of chained 2005 dollars) and unemployment rate, obtained via monthly data with averaging, from 1948 to 2011. Both series are seasonally adjusted. Figure 1.2 shows the time plots of the real GDP growth rate and the changes in unemployment rate from the second quarter of 1948 to the fourth quarter of 2011. Figure 1.3 shows the scatter plot of the two time series given in Figure 1.2. From these figures, we can see that the GDP

Multivariate Time Series Analysis: With R and Financial Applications,
First Edition. Ruey S. Tsay.

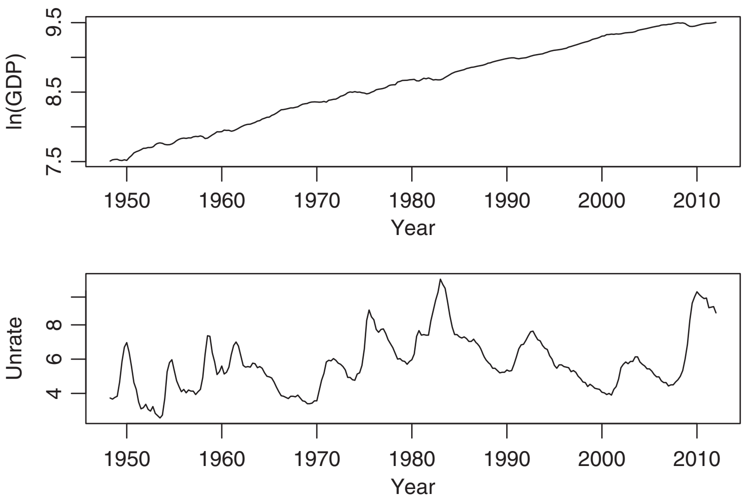

FIGURE 1.1 Time plots of U.S. quarterly real GDP (in logarithm) and unemployment rate from 1948 to 2011. The data are seasonally adjusted.

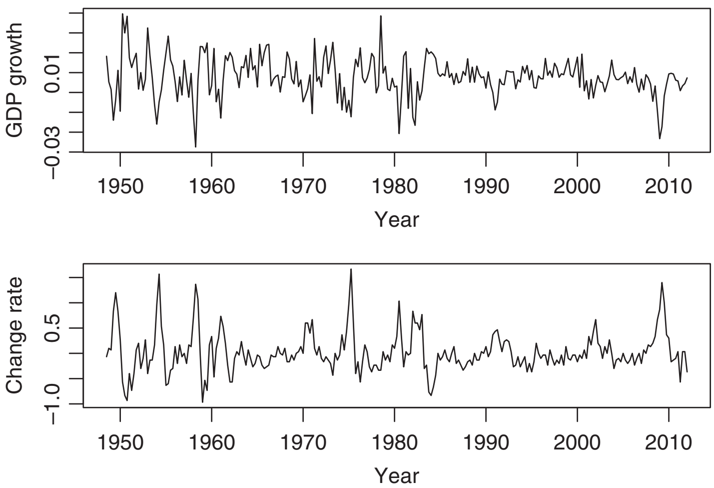

FIGURE 1.2 Time plots of the growth rate of U.S. quarterly real GDP (in logarithm) and the change series of unemployment rate from 1948 to 2011. The data are seasonally adjusted.

and unemployment rate indeed have negative instantaneous correlation. The sample correlation is -0.71.

As another example, consider $k = 3$. Let z_{1t} be the monthly housing starts of the New England division in the United States, and z_{2t} and z_{3t} be the monthly housing starts of the Middle Atlantic division and the Pacific division, respectively. By considering the three series jointly, we can investigate the relationships between the housing markets of the three geographical divisions in the United States. Figure 1.4

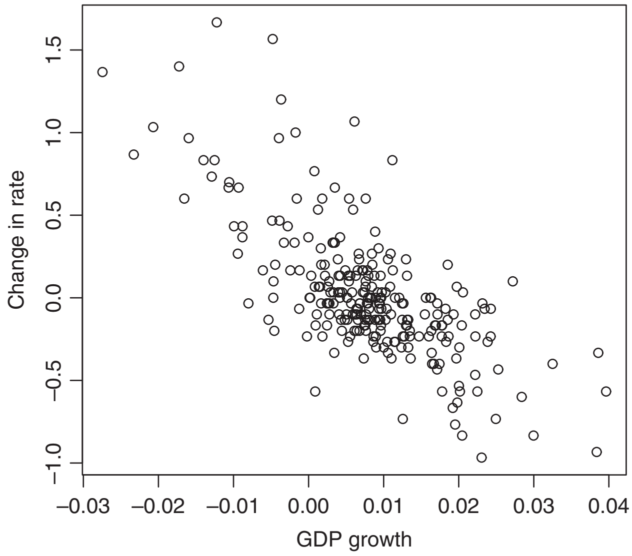

FIGURE 1.3 Scatter plot of the changes in quarterly U.S. unemployment rate versus the growth rate of quarterly real GDP (in logarithm) from the second quarter of 1948 to the last quarter of 2011. The data are seasonally adjusted.

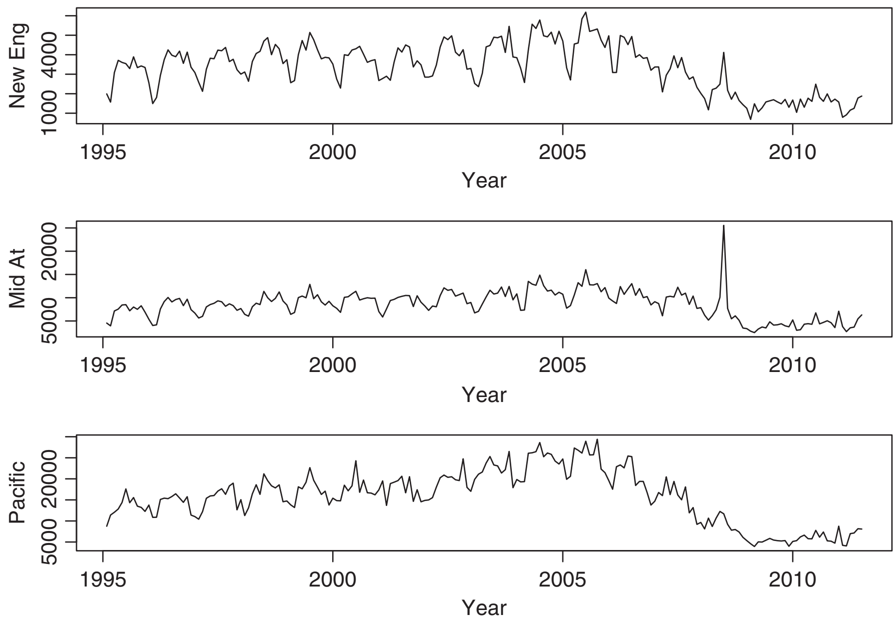

FIGURE 1.4 Time plots of the monthly housing starts for the New England, Middle Atlantic, and Pacific divisions of the United States from January 1995 to June 2011. The data are not seasonally adjusted.

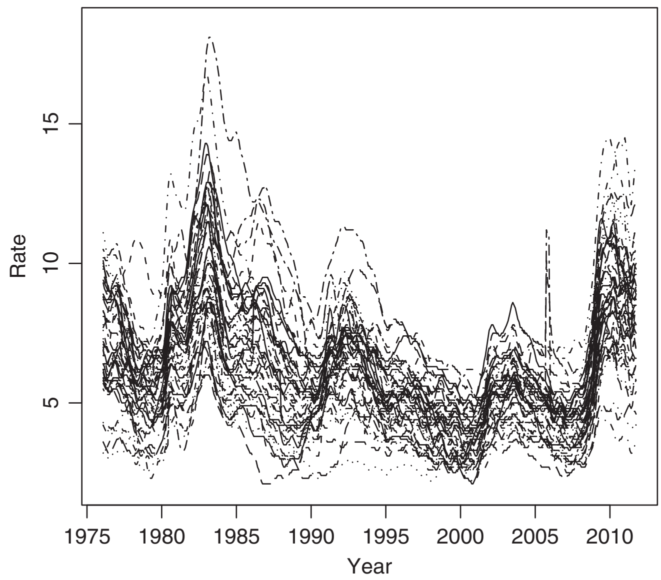

FIGURE 1.5 Time plots of the monthly unemployment rates of the 50 states in the United States from January 1976 to September 2011. The data are seasonally adjusted.

shows the time plots of the three monthly housing starts from January 1995 to June 2011. The data are not seasonally adjusted so that there exists a clear seasonal cycle in the series. From the plots, the three series show certain similarities as well as some marked differences. In some applications, we consider large k. For instance, let $\boldsymbol{z}_t$ be the monthly unemployment rates of the 50 states in the United States. Figure 1.5 shows the time plots of the monthly unemployment rates of the 50 states from January 1976 to September 2011. The data are seasonally adjusted. Here, $k = 50$ and plots are not particularly informative except that the series have certain common behavior. The objective of considering these series simultaneously may be to obtain predictions for the state unemployment rates. Such forecasts are important to state and local governments. In this particular instance, pooling information across states may be helpful in prediction because states may have similar social and economic characteristics.

In this book, we refer to $\{z_{it}\}$ as the ith component of the multivariate time series $\boldsymbol{z}_t$. The objectives of the analysis discussed in this book include (a) to investigate the dynamic relationships between the components of $\boldsymbol{z}_t$ and (b) to improve the prediction of z_{it} using information in all components of $\boldsymbol{z}_t$.

Suppose we are interested in predicting $\boldsymbol{z}_{T+1}$ based on the data $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T\}$. To this end, we may entertain the model

$$\hat{\boldsymbol{z}}_{T+1} = g(\boldsymbol{z}_T, \boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_1),$$

where $\hat{\boldsymbol{z}}_{T+1}$ denotes a prediction of $\boldsymbol{z}_{T+1}$ and $g(.)$ is some suitable function. The goal of multivariate time series analysis is to specify the function $g(.)$ based on the available data. In many applications, $g(.)$ is a smooth, differentiable function and can be well approximated by a linear function, say,

$$\hat{\boldsymbol{z}}_{T+1} \approx \boldsymbol{\pi}_0 + \boldsymbol{\pi}_1 \boldsymbol{z}_T + \boldsymbol{\pi}_2 \boldsymbol{z}_{T-1} + \cdots + \boldsymbol{\pi}_T \boldsymbol{z}_1,$$

where $\boldsymbol{\pi}_0$ is a k-dimensional vector, and $\boldsymbol{\pi}_i$ are $k \times k$ constant real-valued matrices (for $i = 1, \ldots, T$). Let $\boldsymbol{a}_{T+1} = \boldsymbol{z}_{T+1} - \hat{\boldsymbol{z}}_{T+1}$ be the forecast error. The prior equation states that

$$\boldsymbol{z}_{T+1} = \boldsymbol{\pi}_0 + \boldsymbol{\pi}_1 \boldsymbol{z}_T + \boldsymbol{\pi}_2 \boldsymbol{z}_{T-1} + \cdots + \boldsymbol{\pi}_T \boldsymbol{z}_1 + \boldsymbol{a}_{T+1}$$

under the linearity assumption.

To build a solid foundation for making prediction described in the previous paragraph, we need sound statistical theories and methods. The goal of this book is to provide some useful statistical models and methods for analyzing multivariate time series. To begin with, we start with some basic concepts of multivariate time series.

1.2 SOME BASIC CONCEPTS

Statistically speaking, a k-dimensional time series $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{kt})'$ is a random vector consisting of k random variables. As such, there exists an underlying probability space on which the random variables are defined. What we observe in practice is a *realization* of this random vector. For simplicity, we use the same notation $\boldsymbol{z}_t$ for the random vector and its realization. When we discuss properties of $\boldsymbol{z}_t$, we treat it as a random vector. On the other hand, when we consider an application, we treat $\boldsymbol{z}_t$ as a realization. In this book, we assume that $\boldsymbol{z}_t$ follows a continuous multivariate probability distribution. In other words, the discrete-valued (or categorical) multivariate time series are not considered. Because we are dealing with random vectors, vector and matrix are used extensively in the book. If necessary, readers can consult Appendix A for a brief review of mathematics and statistics.

1.2.1 Stationarity

A k-dimensional time series $\boldsymbol{z}_t$ is said to be weakly stationary if (a) $E(\boldsymbol{z}_t) = \boldsymbol{\mu}$, a k-dimensional constant vector, and (b) $\text{Cov}(\boldsymbol{z}_t) = E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_t - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}_z$, a constant $k \times k$ positive-definite matrix. Here, $E(\boldsymbol{z})$ and $\text{Cov}(\boldsymbol{z})$ denote the expectation and covariance matrices of the random vector $\boldsymbol{z}$, respectively. Thus, the mean and covariance matrices of a weakly stationary time series $\boldsymbol{z}_t$ do not depend on time, that is, the first two moments of $\boldsymbol{z}_t$ are time invariant. Implicit in the definition, we require that the mean and covariance matrices of a weakly stationary time series exist.

A k-dimensional time series $\boldsymbol{z}_t$ is strictly stationary if the joint distribution of the m collection, $(\boldsymbol{z}_{t_1}, \ldots, \boldsymbol{z}_{t_m})$, is the same as that of $(\boldsymbol{z}_{t_1+j}, \ldots, \boldsymbol{z}_{t_m+j})'$, where m, j, and $(t_1, \ldots, t_m)$ are arbitrary positive integers. In statistical terms, strict stationarity requires that the probability distribution of an arbitrary collection of $\boldsymbol{z}_t$ is time invariant. An example of strictly stationary time series is the sequence of independent and identically distributed random vectors of standard multivariate normal distribution. From the definitions, a strictly stationary time series $\boldsymbol{z}_t$ is weakly stationary provided that its first two moments exist.

In this chapter, we focus mainly on the weakly stationary series because strict stationarity is hard to verify in practice. We shall consider nonstationary time series later. In what follows, stationarity means weak stationarity.

1.2.2 Linearity

We focus on multivariate linear time series in this book. Strictly speaking, real multivariate time series are nonlinear, but linear models can often provide accurate approximations for making inference. A k-dimensional time series $\boldsymbol{z}_t$ is linear if

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i}, \tag{1.1}$$

where $\boldsymbol{\mu}$ is a k-dimensional constant vector, $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$, the $k \times k$ identity matrix, $\boldsymbol{\psi}_i$ $(i > 0)$ are $k \times k$ constant matrices, and $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed random vectors with mean zero and a positive-definite covariance matrix $\boldsymbol{\Sigma}_a$.

We require $\boldsymbol{\Sigma}_a$ to be positive-definite; otherwise, the dimension k can be reduced—see the principal component analysis discussed in Chapter 2. The condition that $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$ is satisfied because we allow $\boldsymbol{\Sigma}_a$ to be a general positive-definite matrix. An alternative approach to express a linear time series is to require $\boldsymbol{\psi}_0$ to be a lower triangular matrix with diagonal elements being 1 and $\boldsymbol{\Sigma}_a$ a diagonal matrix. This is achieved by using the Cholesky decomposition of $\boldsymbol{\Sigma}_a$; see Appendix A. Specifically, decomposite the covariance matrix as $\boldsymbol{\Sigma}_a = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$, where $\boldsymbol{G}$ is a diagonal matrix and $\boldsymbol{L}$ is a $k \times k$ lower triangular matrix with 1 being its diagonal elements. Let $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then, $\boldsymbol{a}_t = \boldsymbol{L}\boldsymbol{b}_t$, and

$$\text{Cov}(\boldsymbol{b}_t) = \text{Cov}(\boldsymbol{L}^{-1}\boldsymbol{a}_t) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}_a(\boldsymbol{L}^{-1})' = \boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{G}\boldsymbol{L}')(\boldsymbol{L}')^{-1} = \boldsymbol{G}.$$

With the sequence $\{\boldsymbol{b}_t\}$, Equation (1.1) can be written as

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} (\boldsymbol{\psi}_i \boldsymbol{L}) \boldsymbol{b}_{t-i} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i^* \boldsymbol{b}_{t-i}, \tag{1.2}$$

where $\boldsymbol{\psi}_0^* = \boldsymbol{L}$, which is a lower triangular matrix, $\boldsymbol{\psi}_i^* = \boldsymbol{\psi}_i \boldsymbol{L}$ for $i > 0$, and the covariance matrix of $\boldsymbol{b}_t$ is a diagonal matrix.

For a stationary, purely stochastic process $\boldsymbol{z}_t$, Wold decomposition states that it can be written as a linear combination of a sequence of serially uncorrelated process $\boldsymbol{e}_t$. This is close, but not identical, to Equation (1.1) because $\{\boldsymbol{e}_t\}$ do not necessarily have the same distribution. An example of $\boldsymbol{z}_t$ that satisfies the Wold decomposition, but not a linear time series, is the multivariate autoregressive conditional heteroscedastic process. We discuss multivariate volatility modeling in Chapter 7. The Wold decomposition, however, shows that the conditional mean of $\boldsymbol{z}_t$ can be written as a linear combination of the lagged values $\boldsymbol{z}_{t-i}$ for $i > 0$ if $\boldsymbol{z}_t$ is stationary and purely stochastic. This provides a justification for starting with linear time series because the conditional mean of $\boldsymbol{z}_t$ plays an important role in forecasting.

Consider Equation (1.1). We see that $\boldsymbol{z}_{t-1}$ is a function of $\{\boldsymbol{a}_{t-1}, \boldsymbol{a}_{t-2}, \cdots\}$. Therefore, at time index $t-1$, the only *unknown* quantity of $\boldsymbol{z}_t$ is $\boldsymbol{a}_t$. For this reason, we call $\boldsymbol{a}_t$ the *innovation* of the time series $\boldsymbol{z}_t$ at time t. One can think of $\boldsymbol{a}_t$ as the *new* information about the time series obtained at time t. We shall make the concept of innovation more precisely later when we discuss forecasting. The innovation $\boldsymbol{a}_t$ is also known as the *shock* to the time series at time t.

For the linear series $\boldsymbol{z}_t$ in Equation (1.1) to be stationary, the coefficient matrices must satisfy

$$\sum_{i=1}^{\infty} \|\boldsymbol{\psi}_i\| < \infty,$$

where $\|\boldsymbol{A}\|$ denotes a norm of the matrix $\boldsymbol{A}$, for example, the Frobenius norm $\|\boldsymbol{A}\| = \sqrt{tr(\boldsymbol{A}\boldsymbol{A}')}$. Based on the properties of a convergence series, this implies that $\|\boldsymbol{\psi}_i\| \to 0$ as $i \to \infty$. Thus, for a stationary linear time series $\boldsymbol{z}_t$ in Equation (1.1), we have $\boldsymbol{\psi}_i \to \boldsymbol{0}$ as $i \to \infty$. Furthermore, we have

$$E(\boldsymbol{z}_t) = \boldsymbol{\mu}, \quad \text{and} \quad \text{Cov}(\boldsymbol{z}_t) = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_a \boldsymbol{\psi}_i'. \tag{1.3}$$

We shall discuss the stationarity conditions of $\boldsymbol{z}_t$ later for various models.

1.2.3 Invertibility

In many situations, for example, forecasting, we like to express the time series $\boldsymbol{z}_t$ as a function of its lagged values $\boldsymbol{z}_{t-i}$ for $i > 0$ plus new information at time t. A time series $\boldsymbol{z}_t$ is said to be invertible if it can be written as

$$\boldsymbol{z}_t = \boldsymbol{c} + \boldsymbol{a}_t + \sum_{j=1}^{\infty} \boldsymbol{\pi}_j \boldsymbol{z}_{t-j}, \tag{1.4}$$

where $\boldsymbol{c}$ is a k-dimensional constant vector, $\boldsymbol{a}_t$ is defined as before in Equation (1.1), and $\boldsymbol{\pi}_i$ are $k \times k$ constant matrices. An obvious example of an invertible time series is a vector autoregressive (VAR) series of order 1, namely, $\boldsymbol{z}_t = \boldsymbol{c} + \boldsymbol{\pi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$. Again, we shall discuss the invertibility conditions later. Here, it suffices to say that, for an invertible series $\boldsymbol{z}_t$, $\boldsymbol{\pi}_i \to \boldsymbol{0}$ as $i \to \infty$.

1.3 CROSS-COVARIANCE AND CORRELATION MATRICES

To measure the linear dynamic dependence of a stationary time series $\boldsymbol{z}_t$, we define its lag ℓ cross-covariance matrix as

$$\begin{aligned}\boldsymbol{\Gamma}_\ell &= \text{Cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-\ell}) = E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-\ell} - \boldsymbol{\mu})'] \\ &= \begin{bmatrix} E(\tilde{z}_{1t}\tilde{z}_{1,t-\ell}) & E(\tilde{z}_{1t}\tilde{z}_{2,t-\ell}) & \cdots & E(\tilde{z}_{1t}\tilde{z}_{k,t-\ell}) \\ \vdots & \vdots & & \vdots \\ E(\tilde{z}_{kt}\tilde{z}_{1,t-\ell}) & E(\tilde{z}_{kt}\tilde{z}_{2,t-\ell}) & \cdots & E(\tilde{z}_{kt}\tilde{z}_{k,t-\ell}) \end{bmatrix},\end{aligned} \tag{1.5}$$

where $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$ is the mean vector of $\boldsymbol{z}_t$ and $\tilde{\boldsymbol{z}}_t = (\tilde{z}_{1t}, \ldots, \tilde{z}_{kt})' \equiv \boldsymbol{z}_t - \boldsymbol{\mu}$ is the mean-adjusted time series. This cross-covariance matrix is a function of ℓ, not the time index t, because $\boldsymbol{z}_t$ is stationary. For $\ell = 0$, we have the covariance matrix $\boldsymbol{\Gamma}_0$ of $\boldsymbol{z}_t$. In some cases, we use the notation $\boldsymbol{\Sigma}_z$ to denote the covariance matrix of $\boldsymbol{z}_t$, that is, $\boldsymbol{\Sigma}_z = \boldsymbol{\Gamma}_0$.

Denote the (i,j)th element of $\boldsymbol{\Gamma}_\ell$ as $\gamma_{\ell,ij}$, that is, $\boldsymbol{\Gamma}_\ell = [\gamma_{\ell,ij}]$. From the definition in Equation (1.5), we see that $\gamma_{\ell,ij}$ is the covariance between $z_{i,t}$ and $z_{j,t-\ell}$. Therefore, for a positive lag ℓ, $\gamma_{\ell,ij}$ can be regarded as a measure of the linear dependence of the ith component z_{it} on the ℓth lagged value of the jth component z_{jt}. This interpretation is important because we use matrix in the book and one must understand the meaning of each element in a matrix.

From the definition in Equation (1.5), for negative lag ℓ, we have

$$\begin{aligned}\boldsymbol{\Gamma}_\ell &= E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-\ell} - \boldsymbol{\mu})'] \\ &= E[(\boldsymbol{z}_{t+\ell} - \boldsymbol{\mu})(\boldsymbol{z}_t - \boldsymbol{\mu})'], && \text{(because of stationarity)} \\ &= \{E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t+\ell} - \boldsymbol{\mu})']\}', && \text{(because } \boldsymbol{C} = (\boldsymbol{C}')')\text{)} \\ &= \left\{E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-(-\ell)} - \boldsymbol{\mu})']\right\}' \\ &= \{\boldsymbol{\Gamma}_{-\ell}\}', && \text{(by definition)} \\ &= \boldsymbol{\Gamma}'_{-\ell}.\end{aligned}$$

Therefore, unlike the case of univariate stationary time series for which the autocovariances of lag ℓ and lag $-\ell$ are identical, one must take the transpose of a positive-lag cross-covariance matrix to obtain the negative-lag cross-covariance matrix.

Remark: Some researchers define the cross-covariance matrix of $\boldsymbol{z}_t$ as $\boldsymbol{G}_\ell = E[(\boldsymbol{z}_{t-\ell} - \boldsymbol{\mu})(\boldsymbol{z}_t - \boldsymbol{\mu})']$, which is the transpose matrix of Equation (1.5). This is also a valid definition; see the property $\boldsymbol{\Gamma}_{-\ell} = \boldsymbol{\Gamma}'_\ell$. However, the meanings of the off-diagonal elements of $\boldsymbol{G}_\ell$ are different from those defined in Equation (1.5) for $\ell > 0$. As a matter of fact, the (i,j)th element $g_{\ell,ij}$ of $\boldsymbol{G}_\ell$ measures the linear dependence of z_{jt} on the lagged value $z_{i,t-\ell}$ of z_{it}. So long as readers understand the meanings of elements of a cross-covariance matrix, either definition can be used. □

For a stationary multivariate linear time series $\boldsymbol{z}_t$ in Equation (1.1), we have, for $\ell \geq 0$,

$$\begin{aligned}\boldsymbol{\Gamma}_\ell &= E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-\ell} - \boldsymbol{\mu})'] \\ &= E[(\boldsymbol{a}_t + \boldsymbol{\psi}_1\boldsymbol{a}_{t-1} + \cdots)(\boldsymbol{a}_{t-\ell} + \boldsymbol{\psi}_1\boldsymbol{a}_{t-\ell-1} + \cdots)'] \\ &= E[(\boldsymbol{a}_t + \boldsymbol{\psi}_1\boldsymbol{a}_{t-1} + \cdots)(\boldsymbol{a}'_{t-\ell} + \boldsymbol{a}'_{t-\ell-1}\boldsymbol{\psi}'_1 + \cdots)] \\ &= \sum_{i=\ell}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_a \boldsymbol{\psi}'_{i-\ell},\end{aligned} \tag{1.6}$$

where the last equality holds because $\boldsymbol{a}_t$ has no serial covariances and $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$.

For a stationary series $\boldsymbol{z}_t$, the lag ℓ cross-correlation matrix (CCM) $\boldsymbol{\rho}_\ell$ is defined as

$$\boldsymbol{\rho}_\ell = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell\boldsymbol{D}^{-1} = [\rho_{\ell,ij}], \tag{1.7}$$

where $\boldsymbol{D} = \text{diag}\{\sigma_1, \ldots, \sigma_k\}$ is the diagonal matrix of the standard deviations of the components of $\boldsymbol{z}_t$. Specifically, $\sigma_i^2 = \text{Var}(z_{it}) = \gamma_{0,ii}$, that is, the (i,i)th element of $\boldsymbol{\Gamma}_0$. Obviously, $\boldsymbol{\rho}_0$ is symmetric with diagonal elements being 1. The off-diagonal elements of $\boldsymbol{\rho}_0$ are the instantaneous correlations between the components of $\boldsymbol{z}_t$. For $\ell > 0$, $\boldsymbol{\rho}_\ell$ is not symmetric in general because $\rho_{\ell,ij}$ is the correlation coefficient between z_{it} and $z_{j,t-\ell}$, whereas $\rho_{\ell,ji}$ is the correlation coefficient between z_{jt} and $z_{i,t-\ell}$. Using properties of $\boldsymbol{\Gamma}_\ell$, we have $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$.

To study the linear dynamic dependence between the components of $\boldsymbol{z}_t$, it suffices to consider $\boldsymbol{\rho}_\ell$ for $\ell \geq 0$, because for negative ℓ we can use the property $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$. For a k-dimensional series $\boldsymbol{z}_t$, each matrix $\boldsymbol{\rho}_\ell$ is a $k \times k$ matrix. When k is large, it is hard to decipher $\boldsymbol{\rho}_\ell$ simultaneously for several values of ℓ. To summarize the information, one can consider k^2 plots of the elements of $\boldsymbol{\rho}_\ell$ for $\ell = 0, \ldots, m$, where m is a prespecified positive integer. Specifically, for each (i,j)th position, we plot $\rho_{\ell,ij}$ versus ℓ. This plot shows the linear dynamic dependence of z_{it} on $z_{j,t-\ell}$ for $\ell = 0, 1, \ldots, m$. We refer to these k^2 plots as the cross-correlation plots of $\boldsymbol{z}_t$.

1.4 SAMPLE CCM

Given the sample $\{\boldsymbol{z}_t\}_{t=1}^T$, we obtain the sample mean vector and covariance matrix as

$$\hat{\boldsymbol{\mu}}_z = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{z}_t, \quad \hat{\boldsymbol{\Gamma}}_0 = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)(\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)'. \tag{1.8}$$

These sample quantities are estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}_0$, respectively. The lag ℓ sample cross-covariance matrix is defined as

$$\hat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T-1} \sum_{t=\ell+1}^{T} (\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)(\boldsymbol{z}_{t-\ell} - \hat{\boldsymbol{\mu}}_z)'.$$

The lag ℓ sample CCM is then

$$\hat{\boldsymbol{\rho}}_\ell = \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{D}}^{-1},$$

where $\hat{\boldsymbol{D}} = \text{diag}\{\hat{\gamma}_{0,11}^{1/2}, \cdots, \hat{\gamma}_{0,kk}^{1/2}\}$, in which $\hat{\gamma}_{0,ii}$ is the (i,i)th element of $\hat{\boldsymbol{\Gamma}}_0$. If $\boldsymbol{z}_t$ is a stationary linear process and $\boldsymbol{a}_t$ follows a multivariate normal distribution, then $\hat{\boldsymbol{\rho}}_\ell$ is a consistent estimate of $\boldsymbol{\rho}_\ell$. The normality condition can be relaxed by assuming the existence of finite fourth-order moments of $\boldsymbol{z}_t$. The asymptotic covariance matrix between elements of $\hat{\boldsymbol{\rho}}_\ell$ is complicated in general. An approximate formula has been obtained in the literature when $\boldsymbol{z}_t$ has zero fourth-order cumulants (see Bartlett 1955, Box, Jenkins, and Reinsel 1994, Chapter 11, and Reinsel 1993, Section 4.1.2). However, the formula can be simplified for some special cases. For instance, if $\boldsymbol{z}_t$ is a white noise series with positive-definite covariance matrix $\boldsymbol{\Sigma}_z$, then we have

$$\begin{aligned}
\text{Var}(\hat{\rho}_{\ell,ij}) &\approx \frac{1}{T} \quad \text{for} \quad \ell > 0, \\
\text{Var}(\hat{\rho}_{0,ij}) &\approx \frac{(1-\rho_{0,ij}^2)^2}{T} \quad \text{for} \quad i \neq j, \\
\text{Cov}(\hat{\rho}_{\ell,ij}, \hat{\rho}_{-\ell,ij}) &\approx \frac{\rho_{0,ij}^2}{T}, \\
\text{Cov}(\hat{\rho}_{\ell,ij}, \hat{\rho}_{h,uv}) &\approx 0, \quad \ell \neq h.
\end{aligned}$$

Another special case of interest is that $\boldsymbol{z}_t$ follows a vector moving-average (VMA) model, which will be discussed in Chapter 3. For instance, if $\boldsymbol{z}_t$ is a VMA(1) process, then

$$\text{Var}(\hat{\rho}_{\ell,ii}) \approx \frac{1 - 3\rho_{1,ii}^2 + 4\rho_{1,ii}^4}{T}, \quad \text{Var}(\hat{\rho}_{\ell,ij}) \approx \frac{1 + 2\rho_{1,ii}\rho_{1,jj}}{T},$$

for $\ell = \pm 2, \pm 3, \ldots$. If $\boldsymbol{z}_t$ is a VMA(q) process with $q > 0$, then

$$\text{Var}(\hat{\rho}_{\ell,ij}) \approx \frac{1}{T}\left(1 + 2\sum_{v=1}^{q} \rho_{v,ii}\rho_{v,jj}\right), \quad \text{for} \quad |\ell| > q. \tag{1.9}$$

In data analysis, we often examine the sample CCM $\hat{\boldsymbol{\rho}}_\ell$ to study the linear dynamic dependence in the data. As mentioned before, when the dimension k is large, it is hard to comprehend the k^2 cross-correlations simultaneously. To aid our ability to decipher the dependence structure of the data, we adopt the *simplified matrix* of Tiao and Box (1981). For each sample CCM $\hat{\boldsymbol{\rho}}_\ell$, we define a simplified matrix $\boldsymbol{s}_\ell = [s_{\ell,ij}]$ as

$$s_{\ell,ij} = \begin{cases} + & \text{if } \hat{\rho}_{\ell,ij} \geq 2/\sqrt{T}, \\ - & \text{if } \hat{\rho}_{\ell,ij} \leq -2/\sqrt{T}, \\ \cdot & \text{if } |\hat{\rho}_{\ell,ij}| < 2/\sqrt{T}. \end{cases} \tag{1.10}$$

This simplified matrix provides a summary of the sample CCM $\hat{\boldsymbol{\rho}}_\ell$ by applying the approximate 5% significance test to individual elements of $\boldsymbol{\rho}_\ell$ under the white noise assumption.

Another approach to check the linear dynamic dependence of $\boldsymbol{z}_t$ is to consider the sample counterpart of the cross-correlation plot. For each (i,j)th position of the sample CCM, we plot $\hat{\rho}_{\ell,ij}$ versus ℓ for $\ell = 0, 1, \ldots, m$, where m is a positive integer. This is a generalization of the sample autocorrelation function (ACF) of the univariate time series. For a k-dimensional series $\boldsymbol{z}_t$, we have k^2 plots. To simplify further the reading of these k^2 plots, an approximate 95% pointwise confidence interval is often imposed on the plot. Here, the 95% interval is often computed using $0 \pm 2/\sqrt{T}$. In other words, we use $1/\sqrt{T}$ as the standard error for the sample cross-correlations. This is justified in the sense that we are checking whether the observed time series is a white noise series. As mentioned before, if $\boldsymbol{z}_t$ is a white noise series with a positive-definite covariance matrix, then $\boldsymbol{\rho}_\ell = \mathbf{0}$ and the asymptotic variance of the sample cross-correlation $\hat{\rho}_{\ell,ij}$ is $1/T$ for $\ell > 0$.

To demonstrate, we use the command `ccm` of the `MTS` package in R to obtain the cross-correlation plots for a dataset consisting of 300 independent and identically distributed (iid) random draws from the two-dimensional standard Gaussian distribution. In this particular case, we have $\boldsymbol{\Sigma}_z = \boldsymbol{I}_2$ and $\boldsymbol{\rho}_\ell = \mathbf{0}$ for $\ell > 0$ so that we expect $\hat{\boldsymbol{\rho}}_\ell$ to be small for $\ell > 0$ and most of the sample cross-correlations to be within the 95% confidence intervals. Figure 1.6 shows the sample cross-correlation plots. As expected, these plots confirm that $\boldsymbol{z}_t$ has zero cross-correlations for all positive lags.

R Demonstration: Output edited.

```
> sig=diag(2) % create the 2-by-2 identity matrix
> x=rmvnorm(300,rep(0,2),sig) % generate random draws
> MTSplot(x) % Obtain time series plots (output not shown)
> ccm(x)
[1] "Covariance matrix:"
       [,1]   [,2]
[1,]  1.006 -0.101
[2,] -0.101  0.994
CCM at lag:  0
       [,1]   [,2]
```

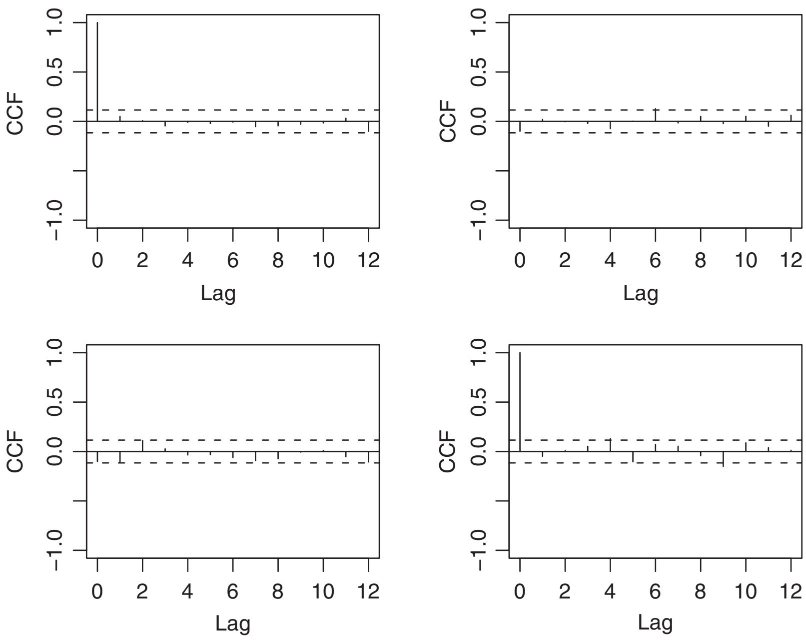

FIGURE 1.6 Sample cross-correlation plots for 300 observations drawn independently from the bivariate standard normal distribution. The dashed lines indicate pointwise 95% confidence intervals.

```
[1,]  1.000 -0.101
[2,] -0.101  1.000
Simplified matrix:
CCM at lag:  1
. .
. .
CCM at lag:  2
. .
. .
CCM at lag:  3
. .
. .
```

1.5 TESTING ZERO CROSS-CORRELATIONS

A basic test in multivariate time series analysis is to detect the existence of linear dynamic dependence in the data. This amounts to testing the null hypothesis H_0 : $\boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}$ versus the alternative hypothesis $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some i satisfying $1 \leq i \leq m$, where m is a positive integer. The *Portmanteau test* of univariate time series has been generalized to the multivariate case by several authors. See, for instance, Hosking (1980, 1981), Li and McLeod (1981), and Li (2004). In particular, the multivariate Ljung–Box test statistic is defined as

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{\Gamma}}'_\ell \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{\Gamma}}_0^{-1}\right), \tag{1.11}$$

where $tr(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$ and T is the sample size. This is referred to as the *multivariate Portmanteau test*. It can be rewritten as

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \hat{\boldsymbol{b}}'_\ell \left(\hat{\boldsymbol{\rho}}_0^{-1} \otimes \hat{\boldsymbol{\rho}}_0^{-1}\right) \hat{\boldsymbol{b}}_\ell,$$

where $\hat{\boldsymbol{b}}_\ell = \text{vec}(\hat{\boldsymbol{\rho}}'_\ell)$ and $\otimes$ is the Kronecker product of two matrices. Here, vec($\boldsymbol{A}$) denotes the column-stacking vector of matrix $\boldsymbol{A}$. Readers are referred to Appendix A for the definitions of vectors and the Kronecker product of two matrices.

Under the null hypothesis that $\boldsymbol{\Gamma}_\ell = \boldsymbol{0}$ for $\ell > 0$ and the condition that $\boldsymbol{z}_t$ is normally distributed, $Q_k(m)$ is asymptotically distributed as $\chi^2_{mk^2}$, that is, a chi-square distribution with mk^2 degrees of freedom. Roughly speaking, assume that $E(\boldsymbol{z}_t) = \boldsymbol{0}$ because covariance matrices do not depend on the mean vectors. Under the assumption $\boldsymbol{\Gamma}_\ell = \boldsymbol{0}$ for $\ell > 0$, we have $\boldsymbol{z}_t = \boldsymbol{a}_t$, a white noise series. Then, the lag ℓ sample autocovariance matrix of $\boldsymbol{a}_t$ is

$$\hat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} \boldsymbol{a}_t \boldsymbol{a}'_{t-\ell}.$$

Using vec($\boldsymbol{AB}$) $= (\boldsymbol{B}' \otimes \boldsymbol{I})$vec($\boldsymbol{A}$), and letting $\hat{\boldsymbol{\gamma}}_\ell = \text{vec}(\hat{\boldsymbol{\Gamma}}_\ell)$, we have

$$\hat{\boldsymbol{\gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (\boldsymbol{a}_{t-\ell} \otimes \boldsymbol{I}_k) \boldsymbol{a}_t.$$

Therefore, we have $E(\hat{\boldsymbol{\gamma}}_\ell) = \boldsymbol{0}$ and

$$\text{Cov}(\hat{\boldsymbol{\gamma}}_\ell) = E(\hat{\boldsymbol{\gamma}}_\ell \hat{\boldsymbol{\gamma}}'_\ell) = \frac{T-\ell}{T^2} \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a.$$

In the aforementioned equation, we have used

$$E[(\boldsymbol{a}_{t-\ell} \otimes \boldsymbol{I}_k) \boldsymbol{a}_t \boldsymbol{a}'_t (\boldsymbol{a}'_{t-\ell} \otimes \boldsymbol{I}_k)] = E(\boldsymbol{a}_{t-\ell} \boldsymbol{a}'_{t-\ell}) \otimes E(\boldsymbol{a}_t \otimes \boldsymbol{a}_t) = \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a.$$

Moreover, by iterated expectation, we have $\text{Cov}(\hat{\boldsymbol{\gamma}}_\ell, \hat{\boldsymbol{\gamma}}_v) = \boldsymbol{0}$ for $\ell \neq v$. In fact, the vectors $T^{1/2}\hat{\boldsymbol{\gamma}}_\ell$, $\ell = 1, \ldots, m$, are jointly asymptotically normal by application of the martingale central limit theorem; see Hannan (1970, p. 228). Therefore,

$$\frac{T^2}{T-\ell}\hat{\gamma}'_\ell\left(\boldsymbol{\Sigma}_a^{-1}\otimes\boldsymbol{\Sigma}_a^{-1}\right)\hat{\gamma}_\ell = \frac{T^2}{T-\ell}tr\left(\boldsymbol{\Sigma}_a^{-1}\hat{\boldsymbol{\Gamma}}'_\ell\boldsymbol{\Sigma}_a^{-1}\hat{\boldsymbol{\Gamma}}_\ell\right) \tag{1.12}$$

is asymptotically distributed as chi-square with k^2 degrees of freedom.

Remark: Strictly speaking, the test statistic of Li and McLeod (1981) is

$$Q_k^*(m) = T\sum_{\ell=1}^{m}\hat{\boldsymbol{b}}'_\ell\left(\hat{\boldsymbol{\rho}}_0^{-1}\otimes\hat{\boldsymbol{\rho}}_0^{-1}\right)\hat{\boldsymbol{b}}_\ell + \frac{k^2 m(m+1)}{2T},$$

which is asymptotically equivalent to $Q_k(m)$. □

To demonstrate the $Q_k(m)$ statistic, we consider the bivariate time series $\boldsymbol{z}_t = (z_{1t}, z_{2t})'$ of Figure 1.2, where z_{1t} is the growth rate of U.S. quarterly real GDP and z_{2t} is the change in the U.S. quarterly unemployment rate. Obviously, there exists certain linear dynamic dependence in the data so that we expect the test statistic to reject the null hypothesis of no cross-correlations. This is indeed the case. The p-values of $Q_k(m)$ are also close to 0 for $m > 0$. See the R demonstration given later, where we use the command `mq` of the `MTS` package to perform the test. We also apply the $Q_k(m)$ statistic to a random sample of 200 observations drawn from the three-dimensional standard normal distribution. In this particular case, the statistic does not reject the null hypothesis of zero cross-correlations. Figure 1.7 shows the time

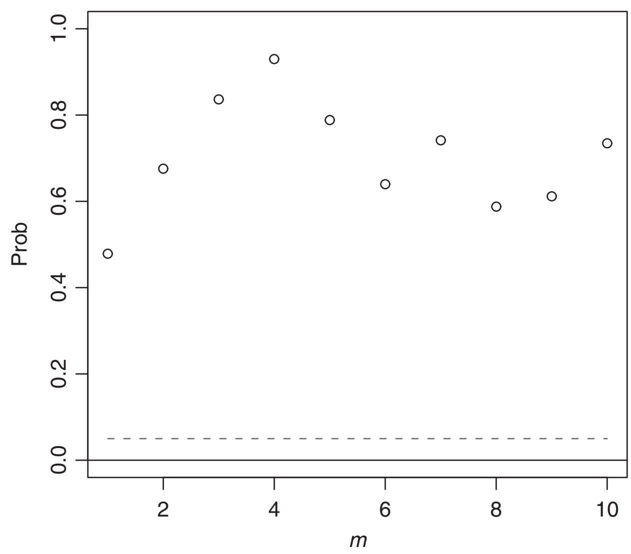

FIGURE 1.7 Plot of p-values for the $Q_k(m)$ statistics for a simulated data consisting of 200 random draws from the three-dimensional standard normal distribution. The dashed line denotes type I error of 5%.

plot of p-values of the $Q_k(m)$ statistic for the simulated three-dimensional white noise series. This is part of the output of the command mq. The dashed line of the plot denotes the type I error of 5%. For this particular simulation, as expected, all p-values are greater than 0.05, confirming that the series has no zero CCMs.

R Demonstration

```
> da=read.table("q-gdpunemp.txt",header=T) % Load the data
> head(da)
  year mon      gdp     rate
1 1948   1 1821.809 3.733333
 ....
6 1949   4 1835.512 5.866667
> x=cbind(diff(da$gdp),diff(da$rate)) % compute differenced
                                      % series
> mq(x,lag=10)  % Compute Q(m) statistics
Ljung-Box Statistics:
        m       Q(m)   p-value
 [1,]     1       140        0
 [2,]     2       196        0
 [3,]     3       213        0
 [4,]     4       232        0
 [5,]     5       241        0
 [6,]     6       246        0
 [7,]     7       250        0
 [8,]     8       261        0
 [9,]     9       281        0
[10,]    10       290        0
>
> sig=diag(3)  %% Simulation study
> z=rmvnorm(200,rep(0,3),sig)
> mq(z,10)
Ljung-Box Statistics:
        m       Q(m)   p-value
 [1,]  1.00      8.56     0.48
 [2,]  2.00     14.80     0.68
 [3,]  3.00     19.86     0.84
 [4,]  4.00     24.36     0.93
 [5,]  5.00     37.22     0.79
 [6,]  6.00     49.73     0.64
 [7,]  7.00     55.39     0.74
 [8,]  8.00     68.72     0.59
 [9,]  9.00     76.79     0.61
[10,] 10.00     81.23     0.73
```

Remark: When the dimension k is large, it becomes cumbersome to plot the CCMs. A possible solution is to summarize the information of $\hat{\mathbf{\Gamma}}_\ell$ by the chi-squared statistic in Equation (1.12). In particular, we can compute the p-value of the chi-squared statistic for testing $H_0 : \mathbf{\Gamma}_\ell = \mathbf{0}$ versus $H_a : \mathbf{\Gamma}_\ell \neq \mathbf{0}$. By plotting the p-value against the lag, we obtain a multivariate generalization of the ACF plot. □

1.6 FORECASTING

Prediction is one of the objectives of the multivariate time series analysis. Suppose we are interested in predicting $\boldsymbol{z}_{h+\ell}$ based on information available at time $t = h$ (inclusive). Such a prediction is called the ℓ-step ahead forecast of the series at the time index h. Here, h is called the *forecast origin* and ℓ the *forecast horizon*. Let F_t denote the available information at time t, which, in a typical situation, consists of the observations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t$. In a time series analysis, the data-generating process is unknown so that we must use the information in F_h to build a statistical model for prediction. As such, the model itself is uncertain. A careful forecaster must consider such uncertainty in making predictions. In practice, it is hard to handle model uncertainty and we make the simplifying assumption that the model used in prediction is the true data-generating process. Keep in mind, therefore, that the forecasts produced by any method that assumes the fitted model as the true model are likely to underestimate the true variability of the time series. In Chapter 2, we discuss the effect of parameter estimates on the mean square of forecast errors for VAR models.

Forecasts produced by an econometric model also depend on the loss function used. In this book, we follow the tradition by using the minimum mean square error (MSE) prediction. Let $\boldsymbol{x}_h$ be an arbitrary forecast of $\boldsymbol{z}_{h+\ell}$ at the forecast origin h. The forecast error is $\boldsymbol{z}_{h+\ell} - \boldsymbol{x}_h$, and the mean square of forecast error is

$$\text{MSE}(\boldsymbol{x}_h) = E[(\boldsymbol{z}_{h+\ell} - \boldsymbol{x}_h)(\boldsymbol{z}_{h+\ell} - \boldsymbol{x}_h)'].$$

Let $\boldsymbol{z}_h(\ell) = E(\boldsymbol{z}_{h+\ell}|F_h)$ be the conditional expectation of $\boldsymbol{z}_{h+\ell}$ given the information F_h, including the model. Then, we can rewrite the MSE of $\boldsymbol{x}_h$ as

$$\begin{aligned}\text{MSE}(\boldsymbol{x}_h) &= E[\{\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell) + \boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}\{\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell) + \boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}'] \\ &= E[\{\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell)\}\{\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell)\}'] + E[\{\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}\{\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}'] \\ &= \text{MSE}[\boldsymbol{z}_h(\ell)] + E[\{\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}\{\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}'], \qquad (1.13)\end{aligned}$$

where we have used the property

$$E[\{\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell)\}\{\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h\}'] = \mathbf{0}.$$

This equation holds because $\boldsymbol{z}_h(\ell) - \boldsymbol{x}_h$ is a vector of functions of F_h, but $\boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell)$ is a vector of functions of the innovations $\{\boldsymbol{a}_{h+\ell}, \ldots, \boldsymbol{a}_{h+1}\}$. Consequently, by using the iterative expectation and $E(\boldsymbol{a}_{t+i}) = \mathbf{0}$, the result in Equation (1.13) holds.

Consider Equation (1.13). Since $E[\{\boldsymbol{z}_h(\ell)-\boldsymbol{x}_h\}\{\boldsymbol{z}_h(\ell)-\boldsymbol{x}_h\}']$ is a nonnegative-definite matrix, we conclude that

$$\text{MSE}(\boldsymbol{x}_h) \geq \text{MSE}[\boldsymbol{z}_h(\ell)],$$

and the equality holds if and only if $\boldsymbol{x}_h = \boldsymbol{z}_h(\ell)$. Consequently, the minimum MSE forecast of $\boldsymbol{z}_{h+\ell}$ at the forecast origin $t = h$ is the conditional expectation of $\boldsymbol{z}_{h+\ell}$ given F_h. For the linear model in Equation (1.1), we have

$$\boldsymbol{z}_h(\ell) = \boldsymbol{\mu} + \boldsymbol{\psi}_\ell \boldsymbol{a}_h + \boldsymbol{\psi}_{\ell+1}\boldsymbol{a}_{h-1} + \cdots .$$

Let $\boldsymbol{e}_h(\ell) = \boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell)$ be the ℓ-step ahead forecast error. Then, we have

$$\boldsymbol{e}_h(\ell) = \boldsymbol{a}_{h+\ell} + \boldsymbol{\psi}_1 \boldsymbol{a}_{h+\ell-1} + \cdots + \boldsymbol{\psi}_{\ell-1}\boldsymbol{a}_{h+1}. \tag{1.14}$$

The covariance matrix of the forecast error is then

$$\text{Cov}[\boldsymbol{e}_h(\ell)] = \boldsymbol{\Sigma}_a + \sum_{i=1}^{\ell-1} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_a \boldsymbol{\psi}_i' = [\sigma_{e,ij}]. \tag{1.15}$$

If we further assume that $\boldsymbol{a}_t$ is multivariate normal, then we can obtain *interval forecasts* for $\boldsymbol{z}_{h+\ell}$. For instance, a 95% interval forecast for the component $z_{i,h+\ell}$ is

$$z_{ih}(\ell) \pm 1.96\sqrt{\sigma_{e,ii}},$$

where $z_{ih}(\ell)$ is the ith component of $\boldsymbol{z}_h(\ell)$ and $\sigma_{e,ii}$ is the (i,i)th diagonal element of Cov$[\boldsymbol{e}_h(\ell)]$ defined in Equation (1.15). One can also construct confidence regions and simultaneous confidence intervals using the methods available in multivariate statistical analysis; see, for instance, Johnson and Wichern (2007, Section 5.4). An approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{z}_{t+h}$ is the ellipsoid determined by

$$(\boldsymbol{z}_h(\ell) - \boldsymbol{z}_{h+\ell})'\text{Cov}[\boldsymbol{e}_h(\ell)]^{-1}(\boldsymbol{z}_h(\ell) - \boldsymbol{z}_{h+\ell}) \leq \chi^2_{k,1-\alpha},$$

where $\chi^2_{k,1-\alpha}$ denotes the $100(1-\alpha)$ quantile of a chi-square distribution with k degrees of freedom and $0 < \alpha < 1$. Also, $100(1-\alpha)\%$ simultaneous confidence intervals for all components of $\boldsymbol{z}_t$ are

$$z_{ih}(\ell) \pm \sqrt{\chi^2_{k,1-\alpha} \times \sigma_{e,ii}}, \quad i = 1, \ldots, k.$$

An alternative approach to construct simultaneous confidence intervals for the k components is to use the *Bonferroni's inequality*. Consider a probability space and events $E_1, \ldots, E_k$. The inequality says that

$$Pr(\cup_{i=1}^{k} E_i) \leq \sum_{i=1}^{k} Pr(E_i).$$

Therefore,

$$Pr(\cap_{i=1}^{k} E_i) \geq 1 - \sum_{i=1}^{k} Pr(E_i^c),$$

where E_i^c denotes the complement of the event E_i. By choosing a $(100 - (\alpha/k))\%$ forecast interval for each component z_{it}, we apply the inequality to ensure that the probability that the following forecast intervals hold is at least $100(1 - \alpha)$:

$$z_{ih}(\ell) \pm Z_{1-(\alpha/k)} \sqrt{\sigma_{e,ii}},$$

where Z_{1-v} is the $100(1 - v)$ quantile of a standard normal distribution.

From Equation (1.14), we see that the one step ahead forecast error is

$$\boldsymbol{e}_h(1) = \boldsymbol{a}_{h+1}.$$

This says that $\boldsymbol{a}_{h+1}$ is the unknown quantity of $\boldsymbol{z}_{h+1}$ at time h. Therefore, $\boldsymbol{a}_{h+1}$ is called the *innovation* of the series at time index $h + 1$. This provides the justification for using the term innovation in Section 1.2.

1.7 MODEL REPRESENTATIONS

The linear model in Equation (1.1) is commonly referred to as the moving-average (MA) representation of a multivariate time series. This representation is useful in forecasting, such as computing the covariance of a forecast error shown in Equation (1.15). It is also used in studying the *impulse response functions*. Again, details are given in later chapters of the book. For an invertible series, the model in Equation (1.4) is referred to as the autoregressive (AR) representation of the model. This model is useful in understanding how $\boldsymbol{z}_t$ depends on its lag values $\boldsymbol{z}_{t-i}$ for $i > 0$.

If the time series is both stationary and invertible, then these two model presentations are equivalent and one can obtain one representation from the other. To see this, we first consider the mean of $\boldsymbol{z}_t$. Taking expectation on both sides of Equation (1.4), we have

$$\boldsymbol{\mu} = \boldsymbol{c} + \sum_{i=1}^{\infty} \boldsymbol{\pi}_i \boldsymbol{\mu}.$$

Letting $\boldsymbol{\pi}_0 = \boldsymbol{I}_k$, we obtain, from the prior equation,

$$\left(\sum_{i=0}^{\infty} \boldsymbol{\pi}_i\right) \boldsymbol{\mu} = \boldsymbol{c}.$$

Plugging in $\boldsymbol{c}$, we can rewrite Equation (1.4) as

$$\tilde{\boldsymbol{z}}_t = \sum_{i=1}^{\infty} \boldsymbol{\pi}_i \tilde{\boldsymbol{z}}_{t-i} + \boldsymbol{a}_t, \tag{1.16}$$

where, as before, $\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t - \boldsymbol{\mu}$ is the mean-adjusted time series.

Next, we consider the relationship between the coefficient matrices $\boldsymbol{\psi}_i$ and $\boldsymbol{\pi}_j$, using the mean-adjusted series $\tilde{\boldsymbol{z}}_t$. The MA representation is

$$\tilde{\boldsymbol{z}}_t = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i}.$$

Let B be the back-shift operator defined by $B\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$ for any time series $\boldsymbol{x}_t$. In the econometric literature, the back-shift operator is called the lag operator and the notation L is often used. Using the back-shift operator, the MA representation of $\tilde{\boldsymbol{z}}_t$ becomes

$$\tilde{\boldsymbol{z}}_t = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i} = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i B^i \boldsymbol{a}_t = \boldsymbol{\psi}(B)\boldsymbol{a}_t, \tag{1.17}$$

where $\boldsymbol{\psi}(B) = \boldsymbol{I}_k + \boldsymbol{\psi}_1 B + \boldsymbol{\psi}_2 B^2 + \boldsymbol{\psi}_3 B^3 + \cdots$. On the other hand, we can also rewrite the AR representation in Equation (1.16) using the back-shift operator as

$$\tilde{\boldsymbol{z}}_t - \sum_{i=1}^{\infty} \boldsymbol{\pi}_i \tilde{\boldsymbol{z}}_{t-i} = \boldsymbol{a}_t \quad \text{or} \quad \boldsymbol{\pi}(B)\tilde{\boldsymbol{z}}_t = \boldsymbol{a}_t, \tag{1.18}$$

where $\boldsymbol{\pi}(B) = \boldsymbol{I}_i - \boldsymbol{\pi}_1 B - \boldsymbol{\pi}_2 B^2 - \cdots$. Plugging Equation (1.17) into Equation (1.18), we obtain

$$\boldsymbol{\pi}(B)\boldsymbol{\psi}(B)\boldsymbol{a}_t = \boldsymbol{a}_t.$$

Consequently, we have $\boldsymbol{\pi}(B)\boldsymbol{\psi}(B) = \boldsymbol{I}_k$. That is,

$$(\boldsymbol{I}_k - \boldsymbol{\pi}_1 B - \boldsymbol{\pi}_2 B^2 - \boldsymbol{\pi}_3 B^3 - \cdots)(\boldsymbol{I}_i + \boldsymbol{\psi}_1 B + \boldsymbol{\psi}_2 B^2 + \cdots) = \boldsymbol{I}_k.$$

This equation implies that all coefficient matrices of B^i on the left-hand side, for $i > 0$, must be 0. Consequently, we have

$$\begin{aligned}
\boldsymbol{\psi}_1 - \boldsymbol{\pi}_1 &= \mathbf{0}, \quad \text{(coefficient matrix of } B^1\text{)}\\
\boldsymbol{\psi}_2 - \boldsymbol{\pi}_1\boldsymbol{\psi}_1 - \boldsymbol{\pi}_2 &= \mathbf{0}, \quad \text{(coefficient matrix of } B^2\text{)}\\
\boldsymbol{\psi}_3 - \boldsymbol{\pi}_1\boldsymbol{\psi}_2 - \boldsymbol{\pi}_2\boldsymbol{\psi}_1 - \boldsymbol{\pi}_3 &= \mathbf{0}, \quad \text{(coefficient matrix of } B^3\text{)}\\
\vdots &= \vdots
\end{aligned}$$

In general, we can obtain $\boldsymbol{\psi}_\ell$ recursively from $\{\boldsymbol{\pi}_i | i = 1, 2, \ldots\}$ via

$$\boldsymbol{\psi}_\ell = \sum_{i=0}^{\ell-1} \boldsymbol{\pi}_{\ell-i}\boldsymbol{\psi}_i, \quad \ell \geq 1, \tag{1.19}$$

where $\boldsymbol{\psi}_0 = \boldsymbol{\pi}_0 = \boldsymbol{I}_k$. Similarly, we can obtain $\boldsymbol{\pi}_\ell$ recursively from $\{\boldsymbol{\psi}_i | i = 1, 2, \ldots\}$ via

$$\boldsymbol{\pi}_1 = \boldsymbol{\psi}_1 \quad \text{and} \quad \boldsymbol{\pi}_\ell = \boldsymbol{\psi}_\ell - \sum_{i=1}^{\ell-1} \boldsymbol{\pi}_i \boldsymbol{\psi}_{\ell-i}, \quad \ell > 1. \tag{1.20}$$

Finally, neither the AR representation in Equation (1.4) nor the MA representation in Equation (1.1) is particularly useful in estimation if they involve too many coefficient matrices. To facilitate model estimation and to gain a deeper understanding of the models used, we postulate that the coefficient matrices $\boldsymbol{\pi}_i$ and $\boldsymbol{\psi}_j$ depend only on a finite number of parameters. This consideration leads to the use of vector autoregressive moving-average (VARMA) models, which are also known as the multivariate autoregressive moving-average (MARMA) models.

A general VARMA(p, q) model can be written as

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-1} + \boldsymbol{a}_t - \sum_{i=1}^{q} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i}, \tag{1.21}$$

where p and q are nonnegative integers, $\boldsymbol{\phi}_0$ is a k-dimensional constant vector, $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_j$ are $k \times k$ constant matrices, and $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed random vectors with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_a$. Using the back-shift operator B, we can write the VARMA model in a compact form as

$$\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{1.22}$$

where $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \cdots - \boldsymbol{\phi}_p B^p$ and $\boldsymbol{\theta}(B) = \boldsymbol{I}_k - \boldsymbol{\theta}_1 B - \cdots - \boldsymbol{\theta}_q B^q$ are matrix polynomials in B. Certain conditions on $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are needed to render the VARMA model stationary, invertible, and identifiable. We shall discuss these conditions in detail in later chapters of the book.

For a stationary series $\boldsymbol{z}_t$, by taking expectation on both sides of Equation (1.21), we have

$$\boldsymbol{\mu} = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{\mu},$$

where $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$. Consequently, we have

$$\left(\boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i \right) \boldsymbol{\mu} = \boldsymbol{\phi}_0. \tag{1.23}$$

This equation can be conveniently written as $\boldsymbol{\phi}(1)\boldsymbol{\mu} = \boldsymbol{\phi}_0$. Plugging Equation (1.23) into the VARMA model in Equation (1.22), we obtain a mean-adjusted VARMA(p, q) model as

$$\boldsymbol{\phi}(B)\tilde{\boldsymbol{z}}_t = \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{1.24}$$

where, as before, $\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t - \boldsymbol{\mu}$.

The AR and MA representations of $\boldsymbol{z}_t$ can be obtained from the VARMA model by matrix multiplication. Assuming for simplicity that the matrix inversion involved exists, we can rewrite Equation (1.24) as

$$\tilde{\boldsymbol{z}}_t = [\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B)\boldsymbol{a}_t.$$

Consequently, comparing with the MA representation in Equation (1.17), we have $\boldsymbol{\psi}(B) = [\boldsymbol{\phi}(B)]^{-1}\boldsymbol{\theta}(B)$, or equivalently

$$\boldsymbol{\phi}(B)\boldsymbol{\psi}(B) = \boldsymbol{\theta}(B).$$

By equating the coefficient matrices of B^i on both sides of the prior equation, we can obtain recursively $\boldsymbol{\psi}_i$ from $\boldsymbol{\phi}_j$ and $\boldsymbol{\theta}_v$ with $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$.

If we rewrite the VARMA model in Equation (1.24) as $[\boldsymbol{\theta}(B)]^{-1}\boldsymbol{\phi}(B)\tilde{\boldsymbol{z}}_t = \boldsymbol{a}_t$ and compare it with the AR representation in Equation (1.18), we see that $[\boldsymbol{\theta}(B)]^{-1}\boldsymbol{\phi}(B) = \boldsymbol{\pi}(B)$. Consequently,

$$\boldsymbol{\psi}(B) = \boldsymbol{\theta}(B)\boldsymbol{\pi}(B).$$

Again, by equating the coefficient matrices of B^i on both sides of the prior equation, we can obtain recursively the coefficient matrix $\boldsymbol{\pi}_i$ from $\boldsymbol{\phi}_j$ and $\boldsymbol{\theta}_v$.

The requirement that both the $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ matrix polynomials of Equation (1.21) start with the $k \times k$ identity matrix is possible because the covariance matrix of $\boldsymbol{a}_t$ is a general positive-definite matrix. Similar to Equation (1.2), we can have alternative parameterizations for the VARMA(p, q) model. Specifically, consider the Cholesky decomposition $\boldsymbol{\Sigma}_a = \boldsymbol{L}\boldsymbol{\Omega}\boldsymbol{L}'$. Let $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. We have

$\text{Cov}(\boldsymbol{b}_t) = \boldsymbol{\Omega}$, which is a diagonal matrix, and $\boldsymbol{a}_t = \boldsymbol{L}\boldsymbol{b}_t$. Using the same method as that of Equation (1.2), we can rewrite the VARMA model in Equation (1.21) as

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \boldsymbol{L}\boldsymbol{b}_t - \sum_{j=1}^{q} \boldsymbol{\theta}_j^* \boldsymbol{b}_{t-j},$$

where $\boldsymbol{\theta}_j^* = \boldsymbol{\theta}_j \boldsymbol{L}$. In this particular formulation, we have $\theta^*(B) = \boldsymbol{L} - \sum_{j=1}^{q} \boldsymbol{\theta}_j^* B^j$. Also, because $\boldsymbol{L}$ is a lower triangular matrix with 1 being the diagonal elements, $\boldsymbol{L}^{-1}$ is also a lower triangular matrix with 1 being the diagonal elements. Premultiplying Equation (1.21) by $\boldsymbol{L}^{-1}$ and letting $\boldsymbol{\phi}_0^* = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0$, we obtain

$$\boldsymbol{L}^{-1}\boldsymbol{z}_t = \boldsymbol{\phi}_0^* + \sum_{i=1}^{p} \boldsymbol{L}^{-1}\boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \boldsymbol{b}_t - \sum_{j=1}^{q} \boldsymbol{L}^{-1}\boldsymbol{\theta}_j \boldsymbol{a}_{t-j}.$$

By inserting $\boldsymbol{L}\boldsymbol{L}^{-1}$ in front of $\boldsymbol{a}_{t-j}$, we can rewrite the prior equation as

$$\boldsymbol{L}^{-1}\boldsymbol{z}_t = \boldsymbol{\phi}_0^* + \sum_{i=1}^{p} \boldsymbol{\phi}_i^* \boldsymbol{z}_{t-i} + \boldsymbol{b}_t - \sum_{j=1}^{q} \tilde{\boldsymbol{\theta}}_j \boldsymbol{b}_{t-j},$$

where $\boldsymbol{\phi}_i^* = \boldsymbol{L}^{-1}\boldsymbol{\phi}_i$ and $\tilde{\boldsymbol{\theta}}_j = \boldsymbol{L}^{-1}\boldsymbol{\theta}_j\boldsymbol{L}$. In this particular formulation, we have $\boldsymbol{\phi}^*(B) = \boldsymbol{L}^{-1} - \sum_{i=1}^{p} \boldsymbol{\phi}_i^* B^i$. From the discussion, we see that there are several equivalent ways to write a VARMA(p, q) model. The important issue in studying a VARMA model is not how to write a VARMA model, but what is the dynamic structure embedded in a given model.

1.8 OUTLINE OF THE BOOK

The book comprises seven chapters. Chapter 2 focuses on the VAR models. It considers the properties of VAR models, starting with simple models of orders 1 and 2. It then introduces estimation and model building. Both the least-squares and Bayesian estimation methods are discussed. Estimation with linear parameter constraints is also included. It also discusses forecasting and the decomposition of the forecast-error covariances. The concepts and calculations of impulse response function are given in detail. Chapter 3 studies the stationary and invertible VARMA models. Again, it starts with the properties of simple MA models. For estimation, both the conditional and the exact likelihood methods are introduced. It then investigates the identifiability and implications of VARMA models. Various approaches to study the likelihood function of a VARMA model are given. For model building, the chapter introduces the method of extended CCMs.

Chapter 4 studies the structural specification of VARMA models. Two methods are given that can specify the simplifying structure (or skeleton) of a vector VARMA time series and, hence, overcome the difficulty of identifiability. Chapter 5 focuses on unit-root nonstationarity. The asymptotic properties of unit-root processes are discussed. It then introduces spurious regression, cointegration, and error-correction forms of VARMA models. Finally, the chapter considers cointegration tests and estimation of error-correction models. Applications of cointegration in finance are briefly discussed. Chapter 6 considers factor models and some selected topics in vector time series. Most factor models available in the literature are included and discussed. Both the orthogonal factor models and the approximate factor models are considered. For selected topics, the chapter includes seasonal vector time series, principal component analysis, missing values, regression models with vector time series errors, and model-based clustering. Finally, Chapter 7 studies multivariate volatility models. It discusses various multivariate volatility models that are relatively easy to estimate and produce positive-definite volatility matrices.

1.9 SOFTWARE

Real examples are used throughout the book to demonstrate the concepts and analyses of vector time series. These empirical analyses were carried out via the `MTS` package developed by the author for the book. Not a trained programmer, I am certain that most of the programs in the package are not as efficient as they can be. With high probability, the program may even contain bugs. My goal in preparing the package is to ensure that readers can reproduce the results shown in the book and gain experience in analyzing real-world vector time series. Interested readers and more experienced researchers can certainly improve the package. I sincerely welcome the suggestions for improvements and corrections for any bug.

EXERCISES

1.1 Simulation is helpful in learning vector time series. Define the matrices

$$\boldsymbol{C} = \begin{bmatrix} 0.8 & 0.4 \\ -0.3 & 0.6 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 2.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}.$$

Use the command

```
m1 = VARMAsim(300,arlags = c(1),phi = C,sigma = S);zt = m1$series
```

to generate 300 observations from the VAR(1) model

$$\boldsymbol{z}_t = \boldsymbol{C}\boldsymbol{z}_{t-1} + \boldsymbol{a}_t,$$

where $\boldsymbol{a}_t$ are iid bivariate normal random variates with mean zero and Cov$(\boldsymbol{a}_t) = \boldsymbol{S}$.

- Plot the time series $\boldsymbol{z}_t$.
- Obtain the first five lags of sample CCMs of $\boldsymbol{z}_t$.

- Test $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_{10} = \mathbf{0}$ versus $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some i, where $i \in \{1, \ldots, 10\}$. Draw the conclusion using the 5% significance level.

1.2 Use the matrices of Problem 1 and the following command

```
m2 = VARMAsim(200,malags = c(1),theta = C,sigma = S);zt = m2$series
```

to generate 200 observations from the VMA(1) model, $\boldsymbol{z}_t = \boldsymbol{a}_t - \boldsymbol{C}\boldsymbol{a}_{t-1}$, where $\boldsymbol{a}_t$ are iid $N(\mathbf{0}, \boldsymbol{S})$.

- Plot the time series $\boldsymbol{z}_t$.
- Obtain the first two lags of sample CCMs of $\boldsymbol{z}_t$.
- Test $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_5 = \mathbf{0}$ versus $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, \ldots, 5\}$. Draw the conclusion using the 5% significance level.

1.3 The file `q-fdebt.txt` contains the U.S. quarterly federal debts held by (a) foreign and international investors, (b) federal reserve banks, and (c) the public. The data are from the Federal Reserve Bank of St. Louis, from 1970 to 2012 for 171 observations, and not seasonally adjusted. The debts are in billions of dollars. Take the log transformation and the first difference for each time series. Let $\boldsymbol{z}_t$ be the differenced log series.

- Plot the time series $\boldsymbol{z}_t$.
- Obtain the first five lags of sample CCMs of $\boldsymbol{z}_t$.
- Test $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_{10} = \mathbf{0}$ versus $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some i, where $i \in \{1, \ldots, 10\}$. Draw the conclusion using the 5% significance level.

Hint: You may use the following commands of `MTS` to process the data:

```
da=read.table("q-fdebt.txt",header=T)
debt=log(da[,3:5]); tdx=da[,1]+da[,2]/12
MTSplot(debt,tdx); zt=diffM(debt); MTSplot(zt,tdx[-1])
```

1.4 The file `m-pgspabt.txt` consists of monthly simple returns of Procter & Gamble stock, S&P composite index, and Abbott Laboratories from January 1962 to December 2011. The data are from CRSP. Transform the simple returns into log returns. Let $\boldsymbol{z}_t$ be the monthly log returns.

- Plot the time series $\boldsymbol{z}_t$.
- Obtain the first two lags of sample CCMs of $\boldsymbol{z}_t$.
- Test $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_5 = \mathbf{0}$ versus $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, \ldots, 5\}$. Draw the conclusion using the 5% significance level.

1.5 For a VARMA time series $\boldsymbol{z}_t$, derive the result of Equation (1.20).

REFERENCES

Bartlett, M. S. (1955). *Stochastic Processes*. Cambridge University Press, Cambridge, UK.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*. 3rd Edition. Prentice-Hall, Englewood Cliffs, NJ.

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons, Inc, New York.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, **75**: 602–607.

Hosking, J. R. M. (1981). Lagrange-multiplier tests of multivariate time series model. *Journal of the Royal Statistical Society, Series B*, **43**: 219–230.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. 6th Edition. Pearson Prentice Hall, Upper Saddle River, NJ.

Li, W. K. (2004). *Diagnostic Checks in Time Series*. Chapman & Hall/CRC, Boca Raton, FL.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate time series models. *Journal of the Royal Statistical Society, Series B*, **43**: 231–239.

Reinsel, G. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, **76**: 802–816.