CHAPTER 8

Multivariate Time Series Analysis and Its Applications

Economic globalization and Internet communication have accelerated the integration of world financial markets in recent years. Price movements in one market can spread easily and instantly to another market. For this reason, financial markets are more dependent on each other than ever before, and one must consider them jointly to better understand the dynamic structure of the global finance. One market may lead the other market under some circumstances, yet the relationship may be reversed under other circumstances. Consequently, knowing how the markets are interrelated is of great importance in finance. Similarly, for an investor or a financial institution holding multiple assets, the dynamic relationships between returns of the assets play an important role in decision making. In this and the next two chapters, we introduce econometric models and methods useful for studying jointly multiple return series. In the statistical literature, these models and methods belong to vector or multivariate time series analysis.

A multivariate time series consists of multiple single series referred to as *components*. As such, concepts of vector and matrix are useful in understanding multivariate time series analysis. We use boldface notation to indicate vectors and matrices. If necessary, readers may consult Appendix A of this chapter for some basic operations and properties of vectors and matrices. Appendix B provides some results of multivariate normal distribution, which is widely used in multivariate statistical analysis (e.g., Johnson and Wichern, 1998).

Let $\boldsymbol{r}_t = (r_{1t}, r_{2t}, \ldots, r_{kt})'$ be the log returns of k assets at time t, where $\boldsymbol{a}'$ denotes the transpose of $\boldsymbol{a}$. For example, an investor holding stocks of IBM, Microsoft, Exxon Mobil, General Motors, and Wal-Mart may consider the five-dimensional daily log returns of these companies. Here r_{1t} denotes the daily log return of IBM stock, r_{2t} is that of Microsoft, and so on. As a second example, an investor who is interested in global investment may consider the return series of the S&P 500 index of the United States, the FTSE 100 index of the United

Analysis of Financial Time Series, Third Edition, By Ruey S. Tsay

Kingdom, and the Nikkei 225 index of Japan. Here the series is three-dimensional, with r_{1t} denoting the return of the S&P 500 index, r_{2t} the return of the Financial Times Stock Exchange (FTSE) 100 index, and r_{3t} the return of the Nikkei 225. The goals of this chapter are (a) to explore the basic properties of $\boldsymbol{r}_t$ and (b) to study econometric models for analyzing the multivariate data $\{\boldsymbol{r}_t | t = 1, \ldots, T\}$.

Many of the models and methods discussed in previous chapters can be generalized directly to the multivariate case. But there are situations in which the generalization requires some attention. In some situations, one needs new models and methods to handle the complicated relationships between multiple series. In this chapter, we discuss these issues with emphasis on intuition and applications. For statistical theory of multivariate time series analysis, readers are referred to Lütkepohl (2005) and Reinsel (1993).

8.1 WEAK STATIONARITY AND CROSS-CORRELATION MATRICES

Consider a k-dimensional time series $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{kt})'$. The series $\boldsymbol{r}_t$ is *weakly stationary* if its first and second moments are time invariant. In particular, the mean vector and covariance matrix of a weakly stationary series are constant over time. Unless stated explicitly to the contrary, we assume that the return series of financial assets are weakly stationary.

For a weakly stationary time series $\boldsymbol{r}_t$, we define its mean vector and covariance matrix as

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t), \quad \boldsymbol{\Gamma}_0 = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_t - \boldsymbol{\mu})'], \tag{8.1}$$

where the expectation is taken element by element over the joint distribution of $\boldsymbol{r}_t$. The mean $\boldsymbol{\mu}$ is a k-dimensional vector consisting of the unconditional expectations of the components of $\boldsymbol{r}_t$. The covariance matrix $\boldsymbol{\Gamma}_0$ is a $k \times k$ matrix. The ith diagonal element of $\boldsymbol{\Gamma}_0$ is the variance of r_{it}, whereas the (i, j)th element of $\boldsymbol{\Gamma}_0$ is the covariance between r_{it} and r_{jt}. We write $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and $\boldsymbol{\Gamma}_0 = [\Gamma_{ij}(0)]$ when the elements are needed.

8.1.1 Cross-Correlation Matrices

Let $\boldsymbol{D}$ be a $k \times k$ diagonal matrix consisting of the standard deviations of r_{it} for $i = 1, \ldots, k$. In other words, $\boldsymbol{D} = \text{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{kk}(0)}\}$. The concurrent, or lag-zero, cross-correlation matrix of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\rho}_0 \equiv [\rho_{ij}(0)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1}.$$

More specifically, the (i, j)th element of $\boldsymbol{\rho}_0$ is

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(r_{it}, r_{jt})}{\text{std}(r_{it})\text{std}(r_{jt})},$$

which is the correlation coefficient between r_{it} and r_{jt}. In time series analysis, such a correlation coefficient is referred to as a concurrent, or contemporaneous, correlation coefficient because it is the correlation of the two series at time t. It is easy to see that $\rho_{ij}(0) = \rho_{ji}(0)$, $-1 \le \rho_{ij}(0) \le 1$, and $\rho_{ii}(0) = 1$ for $1 \le i, j \le k$. Thus, $\boldsymbol{\rho}(0)$ is a symmetric matrix with unit diagonal elements.

An important topic in multivariate time series analysis is the lead–lag relationships between component series. To this end, the cross-correlation matrices are used to measure the strength of linear dependence between time series. The lag-ℓ cross-covariance matrix of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\Gamma}_\ell \equiv [\Gamma_{ij}(\ell)] = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_{t-\ell} - \boldsymbol{\mu})'], \tag{8.2}$$

where $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{r}_t$. Therefore, the (i, j)th element of $\boldsymbol{\Gamma}_\ell$ is the covariance between r_{it} and $r_{j,t-\ell}$. For a weakly stationary series, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ is a function of ℓ, not the time index t.

The lag-ℓ cross-correlation matrix (CCM) of $\boldsymbol{r}_t$ is defined as

$$\boldsymbol{\rho}_\ell \equiv [\rho_{ij}(\ell)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell\boldsymbol{D}^{-1}, \tag{8.3}$$

where, as before, $\boldsymbol{D}$ is the diagonal matrix of standard deviations of the individual series r_{it}. From the definition,

$$\rho_{ij}(\ell) = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\mathrm{Cov}(r_{it}, r_{j,t-\ell})}{\mathrm{std}(r_{it})\mathrm{std}(r_{jt})}, \tag{8.4}$$

which is the correlation coefficient between r_{it} and $r_{j,t-\ell}$. When $\ell > 0$, this correlation coefficient measures the linear dependence of r_{it} on $r_{j,t-\ell}$, which occurred prior to time t. Consequently, if $\rho_{ij}(\ell) \ne 0$ and $\ell > 0$, we say that the series r_{jt} *leads* the series r_{it} at lag ℓ. Similarly, $\rho_{ji}(\ell)$ measures the linear dependence of r_{jt} and $r_{i,t-\ell}$, and we say that the series r_{it} *leads* the series r_{jt} at lag ℓ if $\rho_{ji}(\ell) \ne 0$ and $\ell > 0$. Equation (8.4) also shows that the diagonal element $\rho_{ii}(\ell)$ is simply the lag-ℓ autocorrelation coefficient of r_{it}.

Based on this discussion, we obtain some important properties of the cross correlations when $\ell > 0$. First, in general, $\rho_{ij}(\ell) \ne \rho_{ji}(\ell)$ for $i \ne j$ because the two correlation coefficients measure different linear relationships between $\{r_{it}\}$ and $\{r_{jt}\}$. Therefore, $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$ are in general not symmetric. Second, using $\mathrm{Cov}(x, y) = \mathrm{Cov}(y, x)$ and the weak stationarity assumption, we have

$$\mathrm{Cov}(r_{it}, r_{j,t-\ell}) = \mathrm{Cov}(r_{j,t-\ell}, r_{it}) = \mathrm{Cov}(r_{jt}, r_{i,t+\ell}) = \mathrm{Cov}(r_{jt}, r_{i,t-(-\ell)}),$$

so that $\Gamma_{ij}(\ell) = \Gamma_{ji}(-\ell)$. Because $\Gamma_{ji}(-\ell)$ is the (j, i)th element of the matrix $\boldsymbol{\Gamma}_{-\ell}$ and the equality holds for $1 \le i, j \le k$, we have $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{-\ell}$ and $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$. Consequently, unlike the univariate case, $\boldsymbol{\rho}_\ell \ne \boldsymbol{\rho}_{-\ell}$ for a general vector time series when $\ell > 0$. Because $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$, it suffices in practice to consider the cross-correlation matrices $\boldsymbol{\rho}_\ell$ for $\ell \ge 0$.

8.1.2 Linear Dependence

Considered jointly, the cross-correlation matrices $\{\boldsymbol{\rho}_\ell | \ell = 0, 1, \ldots\}$ of a weakly stationary vector time series contain the following information:

1. The diagonal elements $\{\rho_{ii}(\ell) | \ell = 0, 1, \ldots\}$ are the autocorrelation function of r_{it}.
2. The off-diagonal element $\rho_{ij}(0)$ measures the concurrent linear relationship between r_{it} and r_{jt}.
3. For $\ell > 0$, the off-diagonal element $\rho_{ij}(\ell)$ measures the linear dependence of r_{it} on the past value $r_{j,t-\ell}$.

Therefore, if $\rho_{ij}(\ell) = 0$ for all $\ell > 0$, then r_{it} does not depend linearly on any past value $r_{j,t-\ell}$ of the r_{jt} series.

In general, the linear relationship between two time series $\{r_{it}\}$ and $\{r_{jt}\}$ can be summarized as follows:

1. r_{it} and r_{jt} have no linear relationship if $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ for all $\ell \geq 0$.
2. r_{it} and r_{jt} are concurrently correlated if $\rho_{ij}(0) \neq 0$.
3. r_{it} and r_{jt} have no lead–lag relationship if $\rho_{ij}(\ell) = 0$ and $\rho_{ji}(\ell) = 0$ for all $\ell > 0$. In this case, we say the two series are uncoupled.
4. There is a *unidirectional relationship* from r_{it} to r_{jt} if $\rho_{ij}(\ell) = 0$ for all $\ell > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$. In this case, r_{it} does not depend on any past value of r_{jt}, but r_{jt} depends on some past values of r_{it}.
5. There is a *feedback relationship* between r_{it} and r_{jt} if $\rho_{ij}(\ell) \neq 0$ for some $\ell > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

The conditions stated earlier are sufficient conditions. A more informative approach to study the relationship between time series is to build a multivariate model for the series because a properly specified model considers simultaneously the serial and cross correlations among the series.

8.1.3 Sample Cross-Correlation Matrices

Given the data $\{\boldsymbol{r}_t | t = 1, \ldots, T\}$, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ can be estimated by

$$\widehat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (\boldsymbol{r}_t - \bar{\boldsymbol{r}})(\boldsymbol{r}_{t-\ell} - \bar{\boldsymbol{r}})', \qquad \ell \geq 0, \tag{8.5}$$

where $\bar{\boldsymbol{r}} = (\sum_{t=1}^{T} \boldsymbol{r}_t)/T$ is the vector of sample means. The cross-correlation matrix $\boldsymbol{\rho}_\ell$ is estimated by

$$\widehat{\boldsymbol{\rho}}_\ell = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{D}}^{-1}, \qquad \ell \geq 0, \tag{8.6}$$

where $\widehat{\boldsymbol{D}}$ is the $k \times k$ diagonal matrix of the sample standard deviations of the component series.

Similar to the univariate case, asymptotic properties of the sample cross-correlation matrix $\widehat{\boldsymbol{\rho}}_\ell$ have been investigated under various assumptions; see, for instance, Fuller (1976, Chapter 6). The estimate is consistent but is biased in a finite sample. For asset return series, the finite sample distribution of $\widehat{\boldsymbol{\rho}}_\ell$ is rather complicated partly because of the presence of conditional heteroscedasticity and high kurtosis. If the finite-sample distribution of cross correlations is needed, we recommend that proper bootstrap resampling methods be used to obtain an approximate estimate of the distribution. For many applications, a crude approximation of the variance of $\hat{\rho}_{ij}(\ell)$ is sufficient.

Example 8.1. Consider the monthly log returns of IBM stock and the S&P 500 index from January 1926 to December 2008 with 996 observations. The returns include dividend payments and are in percentages. Denote the returns of IBM stock and the S&P 500 index by r_{1t} and r_{2t}, respectively. These two returns form a bivariate time series $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$. Figure 8.1 shows the time plots of $\boldsymbol{r}_t$. Figure 8.2 shows some scatterplots of the two series. The plots show that the two return series are concurrently correlated. Indeed, the sample concurrent correlation coefficient between the two returns is 0.65, which is statistically significant at the 5% level. However, the cross correlations at lag 1 are weak if any.

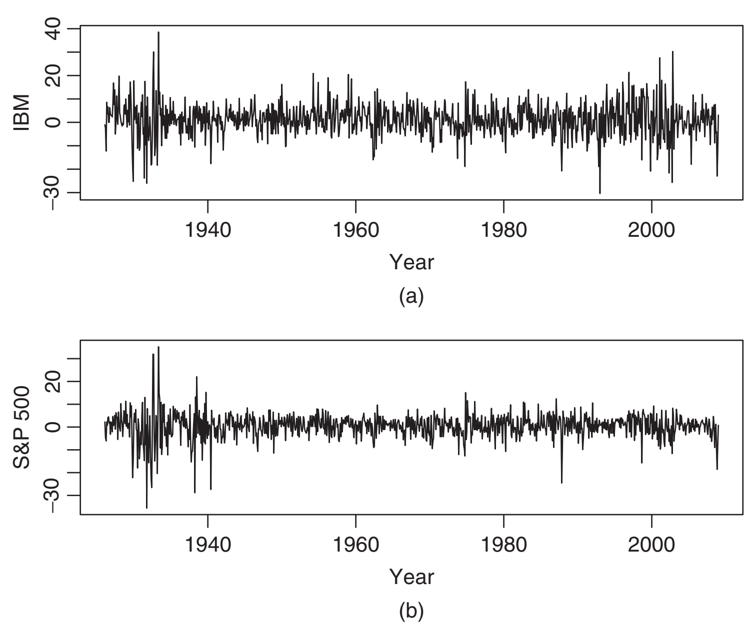


Figure 8.1 Time plots of monthly log returns, in percentages, for (a) IBM stock and (b) the S&P 500 index from January 1926 to December 2008.

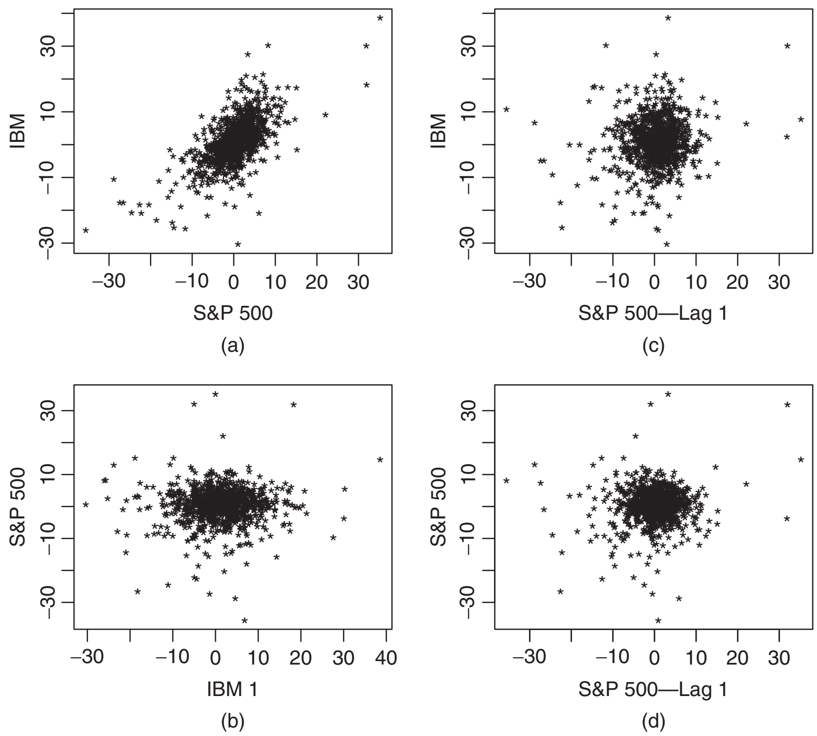


Figure 8.2 Some scatterplots for monthly log returns of IBM stock and S&P 500 index: (a) concurrent plot of IBM vs. S&P 500, (b) S&P 500 vs. lag-1 IBM, (c) IBM vs. lag-1 S&P 500, and (d) S&P 500 vs. lag-1 S&P 500.

Table 8.1 provides some summary statistics and cross-correlation matrices of the two series. For a bivariate series, each CCM is a 2×2 matrix with four correlations. Empirical experience indicates that it is rather hard to absorb simultaneously many cross-correlation matrices, especially when the dimension k is greater than 3. To overcome this difficulty, we use the simplifying notation of Tiao and Box (1981) and define a simplified cross-correlation matrix consisting of three symbols "+," "−," and "." where they have the following meaning:

1. Plus sign (+) means that the corresponding correlation coefficient is greater than or equal to $2/\sqrt{T}$.
2. Minus sign (−) means that the corresponding correlation coefficient is less than or equal to $-2/\sqrt{T}$.
3. Period (.) means that the corresponding correlation coefficient is between $-2/\sqrt{T}$ and $2/\sqrt{T}$.

And $1/\sqrt{T}$ is the asymptotic 5% critical value of the sample correlation under the assumption that $\boldsymbol{r}_t$ is a white noise series.

TABLE 8.1 Summary Statistics and Cross-Correlation Matrices of Monthly Log Returns of IBM Stock and S&P 500 Index: January 1926 to December 2008

(a) Summary Statistics

Ticker	Mean	Standard Error	Skewness	Excess Kurtosis	Minimum	Maximum
IBM	1.089	7.033	−0.068	2.622	−30.37	38.57
SP5	0.430	5.537	−0.521	7.927	−35.59	35.22

(b) Cross-Correlation Matrices

Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
0.04	0.10	0.00	−0.08	−0.01	−0.06	−0.03	−0.03	0.02	0.08
0.04	0.08	0.02	−0.02	−0.06	−0.10	0.04	0.03	0.00	0.09

(c) Simplified notation

$$\begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix} \quad \begin{bmatrix} \cdot & - \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & - \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix}$$

Table 8.1(c) shows the simplified CCM for the monthly log returns of IBM stock and the S&P 500 index. It is easily seen that significant cross correlations at the approximate 5% level appear mainly at lags 1 and 3. An examination of the sample CCMs at these two lags indicates that (a) S&P 500 index returns have some marginal autocorrelations at lags 1, 2, 3, and 5 and (b) IBM stock returns depend weakly on the previous returns of the S&P 500 index. The latter observation is based on the significance of cross correlations at the (1, 2)th element of lag-1, lag-2 and lag-5 CCMs.

Figure 8.3 shows the sample autocorrelations and cross correlations of the two series. The upper-left plot is the sample ACF of IBM stock returns and the upper-right plot shows the dependence of IBM stock returns on the lagged S&P 500 index returns. The dashed lines in the plots are the asymptotic two standard error limits of the sample auto- and cross-correlation coefficients. From the plots, the dynamic relationship is weak between the two return series, but their contemporaneous correlation is statistically significant.

Example 8.2. Consider the simple returns of monthly indexes of U.S. government bonds with maturities in 30 years, 20 years, 10 years, 5 years, and 1 year. The data obtained from the CRSP database have 696 observations starting from January 1942 to December 1999. Let $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{5t})'$ be the return series with decreasing time to maturity. Figure 8.4 shows the time plots of $\boldsymbol{r}_t$ on the same scale. The variability of the 1-year bond returns is much smaller than that of returns with longer maturities. The sample means and standard deviations of the data are $\widehat{\boldsymbol{\mu}} = 10^{-2}(0.43, 0.45, 0.45, 0.46, 0.44)'$ and $\widehat{\boldsymbol{\sigma}} = 10^{-2}(2.53, 2.43, 1.97, 1.39, 0.53)'$.

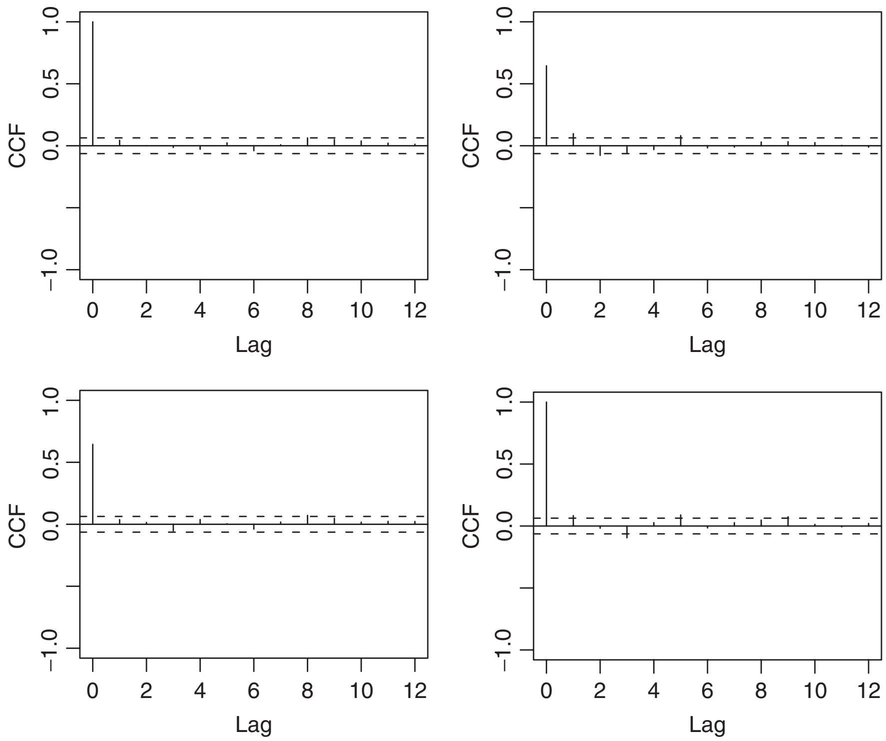


Figure 8.3 Sample auto- and cross-correlation functions (CCF) of two monthly log return series: (a) sample ACF of IBM stock returns, (b) cross-correlations between S&P 500 index and lagged IBM stock returns (lower left), (c) cross correlations between IBM stock and lagged S&P 500 index returns, and (d) sample ACF of S&P 500 index returns. Dashed lines denote 95% limits.

The concurrent correlation matrix of the series is

$$\widehat{\boldsymbol{\rho}}_0 = \begin{bmatrix} 1.00 & 0.98 & 0.92 & 0.85 & 0.63 \\ 0.98 & 1.00 & 0.91 & 0.86 & 0.64 \\ 0.92 & 0.91 & 1.00 & 0.90 & 0.68 \\ 0.85 & 0.86 & 0.90 & 1.00 & 0.82 \\ 0.63 & 0.64 & 0.68 & 0.82 & 1.00 \end{bmatrix}.$$

It is not surprising that (a) the series have high concurrent correlations, and (b) the correlations between long-term bonds are higher than those between short-term bonds.

Table 8.2 gives the lag-1 and lag-2 cross-correlation matrices of $\boldsymbol{r}_t$ and the corresponding simplified matrices. Most of the significant cross correlations are at lag 1, and the five return series appear to be intercorrelated. In addition, lag-1 and lag-2 sample ACFs of the 1-year bond returns are substantially higher than those of other series with longer maturities.

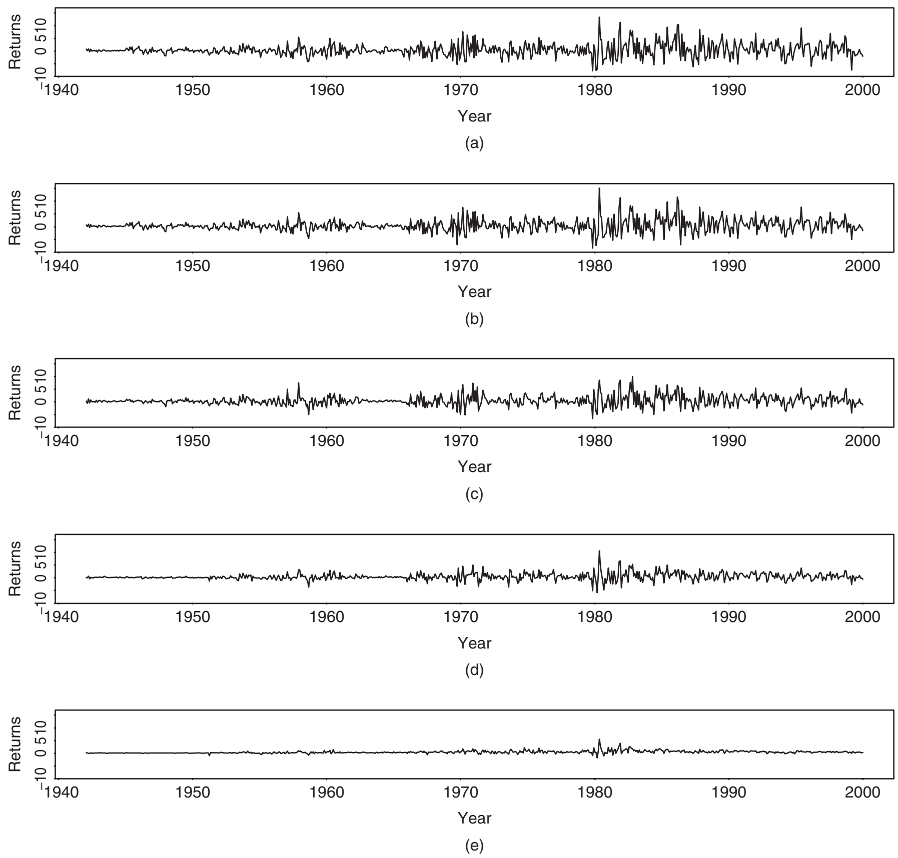


Figure 8.4 Time plots of monthly simple returns of five indexes of U.S. government bonds with maturities in (a) 30 years, (b) 20 years, (c) 10 years, (d) 5 years, and (e) 1 year. Sample period is from January 1942 to December 1999.

8.1.4 Multivariate Portmanteau Tests

The univariate Ljung–Box statistic $Q(m)$ has been generalized to the multivariate case by Hosking (1980, 1981) and Li and McLeod (1981). For a multivariate series, the null hypothesis of the test statistic is $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}$, and the alternative hypothesis $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, \ldots, m\}$. Thus, the statistic is used to test that there are no auto- and cross correlations in the vector series $\boldsymbol{r}_t$. The test statistic assumes the form

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \mathrm{tr}(\widehat{\boldsymbol{\Gamma}}'_\ell \widehat{\boldsymbol{\Gamma}}_0^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{\Gamma}}_0^{-1}), \tag{8.7}$$

where T is the sample size, k is the dimension of $\boldsymbol{r}_t$, and $\mathrm{tr}(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$, which is the sum of the diagonal elements of $\boldsymbol{A}$. Under the null

TABLE 8.2 Sample Cross-Correlation Matrices of Monthly Simple Returns of Five Indexes of U.S. Government Bonds: January 1942 to December 1999

Lag 1					Lag 2				
Cross-Correlations									
0.10	0.08	0.11	0.12	0.16	−0.01	0.00	0.00	−0.03	0.03
0.10	0.08	0.12	0.14	0.17	−0.01	0.00	0.00	−0.04	0.02
0.09	0.08	0.09	0.13	0.18	0.01	0.01	0.01	−0.02	0.07
0.14	0.12	0.15	0.14	0.22	−0.02	−0.01	0.00	−0.04	0.07
0.17	0.15	0.21	0.22	0.40	−0.02	0.00	0.02	0.02	0.22
Simplified Cross-Correlation Matrices									
$\begin{bmatrix} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$					$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + \end{bmatrix}$				

hypothesis and some regularity conditions, $Q_k(m)$ follows asymptotically a chi-squared distribution with k^2m degrees of freedom.

Remark. The $Q_k(m)$ statistics can be rewritten in terms of the sample cross-correlation matrices $\widehat{\boldsymbol{\rho}}_\ell$. Using the Kronecker product $\otimes$ and vectorization of matrices discussed in Appendix A of this chapter, we have

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

where $\boldsymbol{b}_\ell = \text{vec}(\widehat{\boldsymbol{\rho}}'_\ell)$. The test statistic proposed by Li and McLeod (1981) is

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

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

Applying the $Q_k(m)$ statistics to the bivariate monthly log returns of IBM stock and the S&P 500 index of Example 8.1, we have $Q_2(1) = 9.81$, $Q_2(5) = 47.06$, and $Q_2(10) = 71.65$. Based on asymptotic chi-squared distributions with degrees of freedom 4, 20, and 40, the p values of these $Q_2(m)$ statistics are 0.044, 0.001, and 0.002, respectively. The portmanteau tests thus confirm the existence of serial dependence in the bivariate return series at the 5% significance level. For the five-dimensional monthly simple returns of bond indexes in Example 8.2, we have $Q_5(5) = 1065.63$, which is highly significant compared with a chi-squared distribution with 125 degrees of freedom.

The $Q_k(m)$ statistic is a joint test for checking the first m cross-correlation matrices of $\boldsymbol{r}_t$ being zero. If it rejects the null hypothesis, then we build a multivariate model for the series to study the lead–lag relationships between the component series. In what follows, we discuss some simple vector models useful for modeling the linear dynamic structure of a multivariate financial time series.

8.2 VECTOR AUTOREGRESSIVE MODELS

A simple vector model useful in modeling asset returns is the vector autoregressive (VAR) model. A multivariate time series $\boldsymbol{r}_t$ is a VAR process of order 1, or VAR(1) for short, if it follows the model

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t, \tag{8.8}$$

where $\boldsymbol{\phi}_0$ is a k-dimensional vector, $\boldsymbol{\Phi}$ is a $k \times k$ matrix, and $\{\boldsymbol{a}_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. In application, the covariance matrix $\boldsymbol{\Sigma}$ is required to be positive definite; otherwise, the dimension of $\boldsymbol{r}_t$ can be reduced. In the literature, it is often assumed that $\boldsymbol{a}_t$ is multivariate normal.

Consider the bivariate case [i.e., $k = 2$, $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$, and $\boldsymbol{a}_t = (a_{1t}, a_{2t})'$]. The VAR(1) model consists of the following two equations:

$$\begin{aligned} r_{1t} &= \phi_{10} + \Phi_{11}r_{1,t-1} + \Phi_{12}r_{2,t-1} + a_{1t}, \\ r_{2t} &= \phi_{20} + \Phi_{21}r_{1,t-1} + \Phi_{22}r_{2,t-1} + a_{2t}, \end{aligned}$$

where Φ_{ij} is the (i, j)th element of $\boldsymbol{\Phi}$ and ϕ_{i0} is the ith element of $\boldsymbol{\phi}_0$. Based on the first equation, Φ_{12} denotes the linear dependence of r_{1t} on $r_{2,t-1}$ in the presence of $r_{1,t-1}$. Therefore, Φ_{12} is the conditional effect of $r_{2,t-1}$ on r_{1t} given $r_{1,t-1}$. If $\Phi_{12} = 0$, then r_{1t} does not depend on $r_{2,t-1}$, and the model shows that r_{1t} only depends on its own past. Similarly, if $\Phi_{21} = 0$, then the second equation shows that r_{2t} does not depend on $r_{1,t-1}$ when $r_{2,t-1}$ is given.

Consider the two equations jointly. If $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$, then there is a unidirectional relationship from r_{1t} to r_{2t}. If $\Phi_{12} = \Phi_{21} = 0$, then r_{1t} and r_{2t} are uncoupled. If $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$, then there is a feedback relationship between the two series.

8.2.1 Reduced and Structural Forms

In general, the coefficient matrix $\boldsymbol{\Phi}$ of Eq. (8.8) measures the dynamic dependence of $\boldsymbol{r}_t$. The concurrent relationship between r_{1t} and r_{2t} is shown by the off-diagonal element σ_{12} of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$. If $\sigma_{12} = 0$, then there is no concurrent linear relationship between the two component series. In the econometric literature, the VAR(1) model in Eq. (8.8) is called a *reduced-form* model because it does not show explicitly the concurrent dependence between the component series. If necessary, an explicit expression involving the concurrent relationship can be deduced from the reduced-form model by a simple linear transformation. Because $\boldsymbol{\Sigma}$ is positive definite, there exists a lower triangular matrix $\boldsymbol{L}$ with unit diagonal elements and a diagonal matrix $\boldsymbol{G}$ such that $\boldsymbol{\Sigma} = \boldsymbol{LGL}'$; see Appendix A on Cholesky decomposition. Therefore, $\boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{G}$.

Define $\boldsymbol{b}_t = (b_{1t}, \ldots, b_{kt})' = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then

$$E(\boldsymbol{b}_t) = \boldsymbol{L}^{-1}E(\boldsymbol{a}_t) = \boldsymbol{0}, \qquad \text{Cov}(\boldsymbol{b}_t) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^{-1})' = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{G}.$$

Since $\boldsymbol{G}$ is a diagonal matrix, the components of $\boldsymbol{b}_t$ are uncorrelated. Multiplying $\boldsymbol{L}^{-1}$ from the left to model (8.8), we obtain

$$\boldsymbol{L}^{-1}\boldsymbol{r}_t = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0 + \boldsymbol{L}^{-1}\boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{L}^{-1}\boldsymbol{a}_t = \boldsymbol{\phi}_0^* + \boldsymbol{\Phi}^*\boldsymbol{r}_{t-1} + \boldsymbol{b}_t, \tag{8.9}$$

where $\boldsymbol{\phi}_0^* = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0$ is a k-dimensional vector and $\boldsymbol{\Phi}^* = \boldsymbol{L}^{-1}\boldsymbol{\Phi}$ is a $k \times k$ matrix. Because of the special matrix structure, the kth row of $\boldsymbol{L}^{-1}$ is in the form $(w_{k1}, w_{k2}, \ldots, w_{k,k-1}, 1)$. Consequently, the kth equation of model (8.9) is

$$r_{kt} + \sum_{i=1}^{k-1} w_{ki} r_{it} = \phi_{k,0}^* + \sum_{i=1}^{k} \Phi_{ki}^* r_{i,t-1} + b_{kt}, \tag{8.10}$$

where $\phi_{k,0}^*$ is the kth element of $\boldsymbol{\phi}_0^*$ and Φ_{ki}^* is the (k, i)th element of $\boldsymbol{\Phi}^*$. Because b_{kt} is uncorrelated with b_{it} for $1 \le i < k$, Eq. (8.10) shows explicitly the concurrent linear dependence of r_{kt} on r_{it}, where $1 \le i \le k-1$. This equation is referred to as a *structural equation* for r_{kt} in the econometric literature.

For any other component r_{it} of $\boldsymbol{r}_t$, we can rearrange the VAR(1) model so that r_{it} becomes the last component of $\boldsymbol{r}_t$. The prior transformation method can then be applied to obtain a structural equation for r_{it}. Therefore, the reduced-form model (8.8) is equivalent to the structural form used in the econometric literature. In time series analysis, the reduced-form model is commonly used for two reasons. The first reason is ease in estimation. The second and main reason is that the concurrent correlations cannot be used in forecasting.

Example 8.3. To illustrate the transformation from a reduced-form model to structural equations, consider the bivariate AR(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

For this particular covariance matrix $\boldsymbol{\Sigma}$, the lower triangular matrix

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix}$$

provides a Cholesky decomposition [i.e., $\boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1}$ is a diagonal matrix]. Premultiplying $\boldsymbol{L}^{-1}$ to the previous bivariate AR(1) model, we obtain

$$\begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix}\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.7 & 0.95 \end{bmatrix}\begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix},$$

$$\boldsymbol{G} = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix},$$

where $\boldsymbol{G} = \text{Cov}(\boldsymbol{b}_t)$. The second equation of this transformed model gives

$$r_{2t} = 0.3 + 0.5r_{1t} - 0.7r_{1,t-1} + 0.95r_{2,t-1} + b_{2t},$$

which shows explicitly the linear dependence of r_{2t} on r_{1t}.

Rearranging the order of elements in $\boldsymbol{r}_t$, the bivariate AR(1) model becomes

$$\begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix}\begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} a_{2t} \\ a_{1t} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

The lower triangular matrix needed in the Cholesky decomposition of $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix}.$$

Premultiplying $\boldsymbol{L}^{-1}$ to the earlier rearranged VAR(1) model, we obtain

$$\begin{bmatrix} 1.0 & 0.0 \\ -1.0 & 1.0 \end{bmatrix}\begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ -0.8 & 0.8 \end{bmatrix}\begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix},$$

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where $\boldsymbol{G} = \text{Cov}(\boldsymbol{c}_t)$. The second equation now gives

$$r_{1t} = -0.2 + 1.0r_{2t} - 0.8r_{2,t-1} + 0.8r_{1,t-1} + c_{2t}.$$

Again this equation shows explicitly the concurrent linear dependence of r_{1t} on r_{2t}.

8.2.2 Stationarity Condition and Moments of a VAR(1) Model

Assume that the VAR(1) model in Eq. (8.8) is weakly stationary. Taking expectation of the model and using $E(\boldsymbol{a}_t) = \boldsymbol{0}$, we obtain

$$E(\boldsymbol{r}_t) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}E(\boldsymbol{r}_{t-1}).$$

Since $E(\boldsymbol{r}_t)$ is time invariant, we have

$$\boldsymbol{\mu} \equiv E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_0$$

provided that the matrix $\boldsymbol{I} - \boldsymbol{\Phi}$ is nonsingular, where $\boldsymbol{I}$ is the $k \times k$ identity matrix.

Using $\boldsymbol{\phi}_0 = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}$, the VAR(1) model in Eq. (8.8) can be written as

$$(\boldsymbol{r}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t.$$

Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\mu}$ be the mean-corrected time series. Then the VAR(1) model becomes

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t. \tag{8.11}$$

This model can be used to derive properties of a VAR(1) model. By repeated substitutions, we can rewrite Eq. (8.11) as

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \boldsymbol{\Phi}\boldsymbol{a}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{a}_{t-2} + \boldsymbol{\Phi}^3\boldsymbol{a}_{t-3} + \cdots.$$

This expression shows several characteristics of a VAR(1) process. First, since $\boldsymbol{a}_t$ is serially uncorrelated, it follows that $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-1}) = \boldsymbol{0}$. In fact, $\boldsymbol{a}_t$ is not correlated with $\boldsymbol{r}_{t-\ell}$ for all $\ell > 0$. For this reason, $\boldsymbol{a}_t$ is referred to as the *shock* or *innovation* of the series at time t. It turns out that, similar to the univariate case, $\boldsymbol{a}_t$ is uncorrelated with the past value $\boldsymbol{r}_{t-j}$ $(j > 0)$ for all time series models. Second, postmultiplying the expression by $\boldsymbol{a}_t'$, taking expectation, and using the fact of no serial correlations in the $\boldsymbol{a}_t$ process, we obtain $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$. Third, for a VAR(1) model, $\boldsymbol{r}_t$ depends on the past innovation $\boldsymbol{a}_{t-j}$ with coefficient matrix $\boldsymbol{\Phi}^j$. For such dependence to be meaningful, $\boldsymbol{\Phi}^j$ must converge to zero as $j \to \infty$. This means that the k eigenvalues of $\boldsymbol{\Phi}$ must be less than 1 in modulus; otherwise, $\boldsymbol{\Phi}^j$ will either explode or converge to a nonzero matrix as $j \to \infty$. As a matter of fact, the requirement that all eigenvalues of $\boldsymbol{\Phi}$ are less than 1 in modulus is the necessary and sufficient condition for weak stationarity of $\boldsymbol{r}_t$ provided that the covariance matrix of $\boldsymbol{a}_t$ exists. Notice that this stationarity condition reduces to that of the univariate AR(1) case in which the condition is $|\phi| < 1$. Furthermore, because

$$|\lambda \boldsymbol{I} - \boldsymbol{\Phi}| = \lambda^k \mid \boldsymbol{I} - \boldsymbol{\Phi}\frac{1}{\lambda} \mid,$$

the eigenvalues of $\boldsymbol{\Phi}$ are the inverses of the zeros of the determinant $|\boldsymbol{I} - \boldsymbol{\Phi}B|$. Thus, an equivalent sufficient and necessary condition for stationarity of $\boldsymbol{r}_t$ is that all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ are greater than one in modulus; that is, all zeros are outside the unit circle in the complex plane. Fourth, using the expression, we have

$$\text{Cov}(\boldsymbol{r}_t) = \boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}' + \boldsymbol{\Phi}^2\boldsymbol{\Sigma}(\boldsymbol{\Phi}^2)' + \cdots = \sum_{i=0}^{\infty} \boldsymbol{\Phi}^i\boldsymbol{\Sigma}(\boldsymbol{\Phi}^i)',$$

where it is understood that $\boldsymbol{\Phi}^0 = \boldsymbol{I}$, the $k \times k$ identity matrix.

Postmultiplying $\tilde{\boldsymbol{r}}'_{t-\ell}$ to Eq. (8.11), taking expectation, and using the result $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-j}) = E(\boldsymbol{a}_t\tilde{\boldsymbol{r}}'_{t-j}) = \boldsymbol{0}$ for $j > 0$, we obtain

$$E(\tilde{\boldsymbol{r}}_t\tilde{\boldsymbol{r}}'_{t-\ell}) = \boldsymbol{\Phi}E(\tilde{\boldsymbol{r}}_{t-1}\tilde{\boldsymbol{r}}_{t-\ell})', \qquad \ell > 0.$$

Therefore,

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}, \qquad \ell > 0, \tag{8.12}$$

where $\boldsymbol{\Gamma}_j$ is the lag-j cross-covariance matrix of $\boldsymbol{r}_t$. Again this result is a generalization of that of a univariate AR(1) process. By repeated substitutions, Eq. (8.12) shows that

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}^\ell\boldsymbol{\Gamma}_0, \quad \text{for} \quad \ell>0.$$

Pre- and postmultiplying Eq. (8.12) by $\boldsymbol{D}^{-1/2}$, we obtain

$$\boldsymbol{\rho}_\ell = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{D}^{1/2}\boldsymbol{D}^{-1/2}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1/2} = \boldsymbol{\Upsilon}\boldsymbol{\rho}_{\ell-1},$$

where $\boldsymbol{\Upsilon} = \boldsymbol{D}^{-1/2}\boldsymbol{\Phi}\boldsymbol{D}^{1/2}$. Consequently, the CCM of a VAR(1) model satisfies

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}^\ell\boldsymbol{\rho}_0, \qquad \text{for} \qquad \ell>0.$$

8.2.3 Vector AR(p) Models

The generalization of VAR(1) to VAR(p) models is straightforward. The time series $\boldsymbol{r}_t$ follows a VAR(p) model if it satisfies

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \qquad p>0, \tag{8.13}$$

where $\boldsymbol{\phi}_0$ and $\boldsymbol{a}_t$ are defined as before, and $\boldsymbol{\Phi}_j$ are $k \times k$ matrices. Using the back-shift operator B, the VAR(p) model can be written as

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{I}$ is the $k \times k$ identity matrix. This representation can be written in a compact form as

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ is a matrix polynomial. If $\boldsymbol{r}_t$ is weakly stationary, then we have

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$$

provided that the inverse exists. Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\mu}$. The VAR($p$) model becomes

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}_1 \tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Phi}_p \tilde{\boldsymbol{r}}_{t-p} + \boldsymbol{a}_t. \tag{8.14}$$

Using this equation and the same techniques as those for VAR(1) models, we obtain that

- $\mathrm{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{a}_t$.
- $\mathrm{Cov}(\boldsymbol{r}_{t-\ell}, \boldsymbol{a}_t) = \boldsymbol{0}$ for $\ell > 0$.
- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{\ell-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}_{\ell-p}$ for $\ell > 0$.

The last property is called the moment equations of a VAR(p) model. It is a multivariate version of the Yule–Walker equation of a univariate AR(p) model. In terms of CCM, the moment equations become

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}_1 \boldsymbol{\rho}_{\ell-1} + \cdots + \boldsymbol{\Upsilon}_p \boldsymbol{\rho}_{\ell-p} \qquad \text{for } \ell>0,$$

where $\boldsymbol{\Upsilon}_i = \boldsymbol{D}^{-1/2} \boldsymbol{\Phi}_i \boldsymbol{D}^{1/2}$.

A simple approach to understanding properties of the VAR(p) model in Eq. (8.13) is to make use of the results of the VAR(1) model in Eq. (8.8). This can be achieved by transforming the VAR(p) model of $\boldsymbol{r}_t$ into a kp-dimensional VAR(1) model. Specifically, let $\boldsymbol{x}_t = (\tilde{\boldsymbol{r}}'_{t-p+1}, \tilde{\boldsymbol{r}}'_{t-p+2}, \ldots, \tilde{\boldsymbol{r}}'_t)'$ and $\boldsymbol{b}_t = (0, \ldots, 0, \boldsymbol{a}'_t)'$ be two kp-dimensional processes. The mean of $\boldsymbol{b}_t$ is zero and the covariance matrix of $\boldsymbol{b}_t$ is a $kp \times kp$ matrix with zero everywhere except for the lower right corner, which is $\boldsymbol{\Sigma}$. The VAR(p) model for $\boldsymbol{r}_t$ can then be written in the form

$$\boldsymbol{x}_t = \boldsymbol{\Phi}^* \boldsymbol{x}_{t-1} + \boldsymbol{b}_t, \tag{8.15}$$

where $\boldsymbol{\Phi}^*$ is a $kp \times kp$ matrix given by

$$\boldsymbol{\Phi}^* = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} \\ \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_{p-2} & \boldsymbol{\Phi}_{p-3} & \cdots & \boldsymbol{\Phi}_1 \end{bmatrix},$$

where $\boldsymbol{0}$ and $\boldsymbol{I}$ are the $k \times k$ zero matrix and identity matrix, respectively. In the literature, $\boldsymbol{\Phi}^*$ is called the *companion* matrix of the matrix polynomial $\boldsymbol{\Phi}(B)$.

Equation (8.15) is a VAR(1) model for $\boldsymbol{x}_t$, which contains $\boldsymbol{r}_t$ as its last k components. The results of a VAR(1) model shown in the previous section can now be used to derive properties of the VAR(p) model via Eq. (8.15). For example, from the definition, $\boldsymbol{x}_t$ is weakly stationary if and only if $\boldsymbol{r}_t$ is weakly stationary. Therefore, the necessary and sufficient condition of weak stationarity for the VAR(p)

model in Eq. (8.13) is that all eigenvalues of $\boldsymbol{\Phi}^*$ in Eq. (8.15) are less than 1 in modulus. It is easy to show that $|\boldsymbol{I} - \boldsymbol{\Phi}^* B| = |\boldsymbol{\Phi}(B)|$. Therefore, similar to the VAR(1) case, the necessary and sufficient condition is equivalent to all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ being outside the unit circle.

Of particular relevance to financial time series analysis is the structure of the coefficient matrices $\boldsymbol{\Phi}_\ell$ of a VAR(p) model. For instance, if the (i, j)th element $\Phi_{ij}(\ell)$ of $\boldsymbol{\Phi}_\ell$ is zero for all ℓ, then r_{it} does not depend on the past values of r_{jt}. The structure of the coefficient matrices $\boldsymbol{\Phi}_\ell$ thus provides information on the lead–lag relationship between the components of $\boldsymbol{r}_t$.

8.2.4 Building a VAR(p) Model

We continue to use the iterative procedure of order specification, estimation, and model checking to build a vector AR model for a given time series. The concept of partial autocorrelation function of a univariate series can be generalized to specify the order p of a vector series. Consider the following consecutive VAR models:

$$\begin{aligned}
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{a}_t \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{r}_{t-2} + \boldsymbol{a}_t \\
&\;\;\vdots \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_i \boldsymbol{r}_{t-i} + \boldsymbol{a}_t \qquad (8.16) \\
&\;\;\vdots
\end{aligned}$$

Parameters of these models can be estimated by the ordinary least-squares (OLS) method. This is called the multivariate linear regression estimation in multivariate statistical analysis; see Johnson and Wichern (1998).

For the ith equation in Eq. (8.16), let $\widehat{\boldsymbol{\Phi}}_j^{(i)}$ be the OLS estimate of $\boldsymbol{\Phi}_j$ and $\widehat{\boldsymbol{\phi}}_0^{(i)}$ be the estimate of $\boldsymbol{\phi}_0$, where the superscript (i) is used to denote that the estimates are for a VAR(i) model. Then the residual is

$$\widehat{\boldsymbol{a}}_t^{(i)} = \boldsymbol{r}_t - \widehat{\boldsymbol{\phi}}_0^{(i)} - \widehat{\boldsymbol{\Phi}}_1^{(i)} \boldsymbol{r}_{t-1} - \cdots - \widehat{\boldsymbol{\Phi}}_i^{(i)} \boldsymbol{r}_{t-i}.$$

For $i = 0$, the residual is defined as $\widehat{\boldsymbol{r}}_t^{(0)} = \boldsymbol{r}_t - \bar{\boldsymbol{r}}$, where $\bar{\boldsymbol{r}}$ is the sample mean of $\boldsymbol{r}_t$. The residual covariance matrix is defined as

$$\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{T - 2i - 1} \sum_{t=i+1}^{T} \widehat{\boldsymbol{a}}_t^{(i)} (\widehat{\boldsymbol{a}}_t^{(i)})', \qquad i \ge 0. \qquad (8.17)$$

To specify the order p, one can test the hypothesis $H_0 : \boldsymbol{\Phi}_\ell = \boldsymbol{0}$ versus the alternative hypothesis $H_a : \boldsymbol{\Phi}_\ell \neq \boldsymbol{0}$ sequentially for $\ell = 1, 2, \ldots$. For example, using the first equation in Eq. (8.16), we can test the hypothesis $H_0 : \mathbf{\Phi}_1 = \mathbf{0}$ versus the alternative hypothesis $H_a : \mathbf{\Phi}_1 \neq \mathbf{0}$. The test statistic is

$$M(1) = -\left(T - k - \frac{5}{2}\right) \ln\left(\frac{|\widehat{\mathbf{\Sigma}}_1|}{|\widehat{\mathbf{\Sigma}}_0|}\right),$$

where $\widehat{\mathbf{\Sigma}}_i$ is defined in Eq. (8.17) and $|\boldsymbol{A}|$ denotes the determinant of the matrix $\boldsymbol{A}$. Under some regularity conditions, the test statistic $M(1)$ is asymptotically a chi-squared distribution with k^2 degrees of freedom; see Tiao and Box (1981).

In general, we use the ith and $(i-1)$th equations in Eq. (8.16) to test $H_0 : \mathbf{\Phi}_i = \mathbf{0}$ versus $H_a : \mathbf{\Phi}_i \neq \mathbf{0}$; that is, testing a VAR(i) model versus a VAR($i-1$) model. The test statistic is

$$M(i) = -\left(T - k - i - \frac{3}{2}\right) \ln\left(\frac{|\widehat{\mathbf{\Sigma}}_i|}{|\widehat{\mathbf{\Sigma}}_{i-1}|}\right). \tag{8.18}$$

Asymptotically, $M(i)$ is distributed as a chi-squared distribution with k^2 degrees of freedom.

Alternatively, one can use the Akaike information criterion (AIC) or its variants to select the order p. Assume that $\boldsymbol{a}_t$ is multivariate normal and consider the ith equation in Eq. (8.16). One can estimate the model by the maximum-likelihood (ML) method. For AR models, the OLS estimates $\widehat{\boldsymbol{\phi}}_0$ and $\widehat{\mathbf{\Phi}}_j$ are equivalent to the (conditional) ML estimates. However, there are differences between the estimates of $\mathbf{\Sigma}$. The ML estimate of $\mathbf{\Sigma}$ is

$$\tilde{\mathbf{\Sigma}}_i = \frac{1}{T} \sum_{t=i+1}^{T} \widehat{\boldsymbol{a}}_t^{(i)} [\widehat{\boldsymbol{a}}_t^{(i)}]'. \tag{8.19}$$

The AIC of a VAR(i) model under the normality assumption is defined as

$$\text{AIC}(i) = \ln(|\tilde{\mathbf{\Sigma}}_i|) + \frac{2k^2 i}{T}.$$

For a given vector time series, one selects the AR order p such that $\text{AIC}(p) = \min_{0 \leq i \leq p_0} \text{AIC}(i)$, where p_0 is a prespecified positive integer.

Other information criteria available for VAR(i) models are

$$\text{BIC}(i) = \ln(|\tilde{\mathbf{\Sigma}}_i|) + \frac{k^2 i \ \ln(T)}{T},$$

$$\text{HQ}(i) = \ln(|\tilde{\mathbf{\Sigma}}_i|) + \frac{2k^2 i \ \ln[\ln(T)]}{T}.$$

The HQ criterion is proposed by Hannan and Quinn (1979).

Example 8.4. Assuming that the bivariate series of monthly log returns of IBM stock and the S&P 500 index discussed in Example 8.1 follows a VAR model, we apply the $M(i)$ statistics and AIC to the data. Table 8.3 shows the results of these statistics. Both statistics indicate that a VAR(5) model might be adequate for the data. The $M(i)$ statistics are marginally significant at lags 1, 3, and 5 at the 5% level. The minimum of AIC occurs at order 5. For this particular instance, the $M(i)$ statistic is only marginally significant at the 1% level when $i = 2$, confirming the previous observation that the dynamic linear dependence between the two return series is weak.

Estimation and Model Checking

For a specified VAR model, one can estimate the parameters using either the OLS method or the ML method. The two methods are asymptotically equivalent. Under some regularity conditions, the estimates are asymptotically normal; see Reinsel (1993). A fitted model should then be checked carefully for any possible inadequacy. The $Q_k(m)$ statistic can be applied to the residual series to check the assumption that there are no serial or cross correlations in the residuals. For a fitted VAR(p) model, the $Q_k(m)$ statistic of the residuals is asymptotically a chi-squared distribution with $k^2m - g$ degrees of freedom, where g is the number of estimated parameters in the AR coefficient matrices; see Lütkepohl (2005).

Example 8.4 (Continued). Table 8.4(a) shows the estimation results of a VAR(5) model for the bivariate series of monthly log returns of IBM stock and the S&P 500 index. The specified model is in the form

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{r}_{t-2} + \boldsymbol{\Phi}_3\boldsymbol{r}_{t-3} + \boldsymbol{\Phi}_5\boldsymbol{r}_{t-5} + \boldsymbol{a}_t, \tag{8.20}$$

where the first component of $\boldsymbol{r}_t$ denotes IBM stock returns. For this particular instance, we do not use AR coefficient matrix at lag 4 because of the weak serial dependence of the data. In general, when the $M(i)$ statistics and the AIC criterion specify a VAR(5) model, all five AR lags should be used. Table 8.4(b) shows the estimation results after some statistically insignificant parameters are set to zero. The $Q_k(m)$ statistics of the residual series for the fitted model in Table 8.4(b) give $Q_2(4) = 16.64$ and $Q_2(8) = 31.55$. Since the fitted VAR(5) model has six parameters in the AR coefficient matrices, these two $Q_k(m)$ statistics are distributed asymptotically as a chi-squared distribution with degrees of freedom 10 and 26,

TABLE 8.3 Order Specification Statistics for Monthly Log Returns of IBM Stock and S&P 500 Index from January 1926 to December 2008[a]

Order	1	2	3	4	5	6
$M(i)$	10.76	13.41	10.34	7.78	12.07	1.93
AIC	6.795	6.789	6.786	6.786	6.782	6.788

[a] The 5% and 1% critical values of a chi-squared distribution with 4 degrees of freedom are 9.5 and 13.3.

respectively. The p-values of the test statistics are 0.083 and 0.208, and hence the fitted model is adequate at the 5% significance level. As shown by the univariate analysis, the return series are likely to have conditional heteroscedasticity. We discuss multivariate volatility in Chapter 10.

From the fitted model in Table 8.4(b), we make the following observations: (a) The concurrent correlation coefficient between the two innovational series is $24/\sqrt{48 \times 30} = 0.63$, which, as expected, is close to the sample correlation coefficient between r_{1t} and r_{2t}. (b) The two log return series have positive and significant means, implying that the log prices of the two series had an upward trend over the data span. (c) The model shows that

$$\begin{aligned}
\text{IBM}_t &= 1.0 + 0.13\text{SP5}_{t-1} - 0.09\text{SP5}_{t-2} + 0.09\text{SP5}_{t-5} + a_{1t}, \\
\text{SP5}_t &= 0.4 + 0.08\text{SP5}_{t-1} - 0.06\text{SP5}_{t-3} + 0.09\text{SP5}_{t-5} + a_{2t}.
\end{aligned}$$

Consequently, at the 5% significance level, there is a unidirectional dynamic relationship from the monthly S&P 500 index return to the IBM return. If the S&P 500 index represents the U.S. stock market, then IBM return is affected by the past movements of the market. However, past movements of IBM stock returns do not significantly affect the U.S. market, even though the two returns have substantial concurrent correlation. Finally, the fitted model can be written as

$$\begin{aligned}
\begin{bmatrix} \text{IBM}_t \\ \text{SP5}_t \end{bmatrix} &= \begin{bmatrix} 1.0 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.13 \\ 0.08 \end{bmatrix} \text{SP5}_{t-1} - \begin{bmatrix} 0.09 \\ 0 \end{bmatrix} \text{SP5}_{t-2} - \begin{bmatrix} 0 \\ 0.06 \end{bmatrix} \text{SP5}_{t-3} \\
&\quad + \begin{bmatrix} 0.09 \\ 0.09 \end{bmatrix} \text{SP5}_{t-5} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},
\end{aligned}$$

indicating that SP5_t is the driving factor of the bivariate series.

TABLE 8.4 Estimation Results of a VAR(5) Model for the Monthly Log Returns, in Percentages, of IBM Stock and S&P 500 Index from January 1926 to December 2008

Parameter	$\boldsymbol{\phi}_0$	$\boldsymbol{\Phi}_1$		$\boldsymbol{\Phi}_2$		$\boldsymbol{\Phi}_3$		$\boldsymbol{\Phi}_5$		$\boldsymbol{\Sigma}$	
					(a) Full Model						
Estimate	1.0	−0.03	0.15	0.10	−0.17	0.05	−0.11	−0.06	0.14	48	24
	0.4	−0.03	0.11	0.04	−0.04	0.02	−0.11	−0.07	0.15	24	30
Standard	0.23	0.04	0.05	0.04	0.05	0.04	0.05	0.04	0.05		
error	0.18	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04		
					(b) Simplified Model						
Estimate	1.0	0	0.13	0	−0.09	0	0	0	0.09	48	24
	0.4	0	0.08	0	0	0	−0.06	0	0.09	24	30
Standard	0.22	–	0.04	–	0.03	–	–	–	0.04		
error	0.18	–	0.03	–	–	–	−0.06	–	0.03		

Forecasting

Treating a properly built model as the true model, one can apply the same techniques as those in the univariate analysis to produce forecasts and standard deviations of the associated forecast errors. For a VAR(p) model, the 1-step-ahead forecast at the time origin h is $\boldsymbol{r}_h(1) = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+1-i}$, and the associated forecast error is $\boldsymbol{e}_h(1) = \boldsymbol{a}_{h+1}$. The covariance matrix of the forecast error is $\boldsymbol{\Sigma}$. For 2-step-ahead forecasts, we substitute $\boldsymbol{r}_{h+1}$ by its forecast to obtain

$$\boldsymbol{r}_h(2) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_h(1) + \sum_{i=2}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+2-i},$$

and the associated forecast error is

$$\boldsymbol{e}_h(2) = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1[\boldsymbol{r}_t - \boldsymbol{r}_h(1)] = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1 \boldsymbol{a}_{h+1}.$$

The covariance matrix of the forecast error is $\boldsymbol{\Sigma} + \boldsymbol{\Phi}_1 \boldsymbol{\Sigma} \boldsymbol{\Phi}_1'$. If $\boldsymbol{r}_t$ is weakly stationary, then the ℓ-step-ahead forecast $\boldsymbol{r}_h(\ell)$ converges to its mean vector $\boldsymbol{\mu}$ as the forecast horizon ℓ increases and the covariance matrix of its forecast error converges to the covariance matrix of $\boldsymbol{r}_t$.

Table 8.5 provides 1-step- to 6-step-ahead forecasts of the monthly log returns, in percentages, of IBM stock and the S&P 500 index at the forecast origin $h =$ 996. These forecasts are obtained by the refined VAR(5) model in Table 8.4(b). As expected, the standard errors of the forecasts converge to the sample standard errors 7.03 and 5.53, respectively, for the two log return series.

In summary, building a VAR model involves three steps: (a) Use the test statistic $M(i)$ or some information criterion to identify the order, (b) estimate the specified model by using the least-squares method and, if necessary, reestimate the model by removing statistically insignificant parameters, and (c) use the $Q_k(m)$ statistic of the residuals to check the adequacy of a fitted model. Other characteristics of the residual series, such as conditional heteroscedasticity and outliers, can also be checked. If the fitted model is adequate, then it can be used to obtain forecasts and make inference concerning the dynamic relationship between the variables.

We used SCA to perform the analysis in this section. The commands used include `miden`, `mtsm`, `mest`, and `mfore`, where the prefix `m` stands for multivariate. Details of the commands and output are shown below.

TABLE 8.5 Forecasts of a VAR(5) Model for Monthly Log Returns, in Percentages, of IBM Stock and S&P 500 Index: Forecast Origin Is December 2008

Step	1	2	3	4	5	6
IBM forecast	1.95	0.30	−0.82	0.14	1.16	1.29
Standard error	6.95	6.99	7.00	7.00	7.00	7.00
SP forecast	1.70	0.17	−1.26	−0.49	0.41	0.65
Standard error	5.48	5.50	5.50	5.51	5.51	5.53

SCA Demonstration

Output has been edited and % denotes explanation in the following:

```
input date, ibm, sp5. file 'm-ibmsp2608.txt'.
 --% compute percentage log returns.
ibm=ln(ibm+1)*100
 --
sp5=ln(sp5+1)*100
--% model identification
miden ibm,sp5. arfits 1 to 12.

TIME PERIOD ANALYZED . . . . . . . . . . . .  1  TO   996
EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          996

SERIES   NAME           MEAN       STD. ERROR
  1     IBM             1.0891      7.0298
  2     SP5             0.4301      5.5346

NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELA-
TIONS BELOW
      IS (1/NOBE**.5) =    0.03169

SAMPLE CORRELATION MATRIX OF THE SERIES
  1.00
  0.65  1.00

SUMMARIES OF CROSS CORRELATION MATRICES USING +,-,., WHERE
 + DENOTES A VALUE GREATER THAN 2/SQRT(NOBE)
     - DENOTES A VALUE LESS THAN -2/SQRT(NOBE)
     . DENOTES A NON-SIGNIFICANT VALUE BASED ON THE ABOVE
       CRITERION

CROSS CORRELATION MATRICES IN TERMS OF +,-,.
LAGS  1 THROUGH  6
     . +   . -     . .      . .      . +      . .
     . +   . .     . -      . .      . +      . .
LAGS  7 THROUGH 12
     . .   + .     . .      . .      . .      . .
     . .   + .     . +      . .      . .      . .

 ========  STEPWISE AUTOREGRESSION SUMMARY  ========
--------------------------------------------------------------
    I RESIDUAL I EIGENVAL.I CHI-SQ  I          I SIGN.
LAG I VARIANCESI OF SIGMA I  TEST   I   AIC    I PAR. AR
----+----------+----------+---------+----------+--------------
  1 I .492E+02 I .133E+02 I   10.76 I    6.795 I . +
    I .306E+02 I .665E+02 I         I          I . +
----+----------+----------+---------+----------+--------------
  2 I .486E+02 I .133E+02 I   13.41 I    6.789 I + -
```

```
    I .306E+02 I .659E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
  3 I .484E+02 I .132E+02 I   10.34 I    6.786 I . .
    I .303E+02 I .655E+02 I         I          I . -
----+----------+----------+---------+----------+--------------
  4 I .484E+02 I .131E+02 I    7.78 I    6.786 I . .
    I .302E+02 I .655E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
  5 I .480E+02 I .131E+02 I   12.07 I    6.782 I . +
    I .299E+02 I .648E+02 I         I          I - +
----+----------+----------+---------+----------+--------------
  6 I .479E+02 I .131E+02 I    1.93 I    6.788 I . .
    I .298E+02 I .647E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
  7 I .479E+02 I .130E+02 I    2.68 I    6.793 I . .
    I .298E+02 I .647E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
  8 I .477E+02 I .130E+02 I    7.09 I    6.794 I . .
    I .296E+02 I .643E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
  9 I .476E+02 I .130E+02 I    5.23 I    6.797 I . .
    I .295E+02 I .642E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
 10 I .476E+02 I .130E+02 I    1.43 I    6.803 I . .
    I .295E+02 I .641E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
 11 I .475E+02 I .130E+02 I    1.81 I    6.809 I . .
    I .294E+02 I .640E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
 12 I .475E+02 I .129E+02 I    1.88 I    6.815 I . .
    I .294E+02 I .640E+02 I         I          I . .
----+----------+----------+---------+----------+--------------
NOTE:CHI-SQUARED CRITICAL VALUES WITH 4 DEGREES OF FREEDOM ARE
          5 PERCENT:   9.5     1 PERCENT:  13.3
 -- % model specification of a VAR(5) model without lag 4.
mtsm m1. series ibm, sp5. model @
(i-p2*b-p2*b**2-p3*b**3-p5*b**5)series=c+noise.
 -- % estimation
 mestim m1. hold resi(r1,r2).
 -- % demonstration of setting zero constraint
p2(2,2)=0
 --
cp2(2,2)=1
 --
p3(1,2)=0
 --
cp3(1,2)=1
 --
mestim m1. hold resi(r1,r2)
```

```
FINAL MODEL SUMMARY WITH CONDITIONAL LIKELIHOOD PAR. EST.
----- CONSTANT VECTOR (STD ERROR) -----
     1.039  (    0.223  )
     0.390  (    0.176  )
----- PHI MATRICES -----
ESTIMATES OF    PHI(  1 ) MATRIX AND SIGNIFICANCE
      .000     .129      . +
      .000     .080      . +
STANDARD ERRORS
       --      .040
       --      .031
ESTIMATES OF    PHI(  2 ) MATRIX AND SIGNIFICANCE
      .000    -.090      . -
      .000     .000      . .
STANDARD ERRORS
       --      .031
       --       --
ESTIMATES OF    PHI(  3 ) MATRIX AND SIGNIFICANCE
      .000     .000      . .
      .000    -.061      . -
STANDARD ERRORS
       --       --
       --      .024

ESTIMATES OF    PHI(  5 ) MATRIX AND SIGNIFICANCE
      .000     .093      . +
      .000     .087      . +
STANDARD ERRORS
       --      .040
       --      .032

------------------------
ERROR COVARIANCE MATRIX
------------------------
               1           2
   1    48.328570
   2    24.361464   30.027406
--------------------------------------------------------------
-- % compute residual cross-correlation matrices
miden r1,r2. maxl 12.
 -- % prediction
mfore m1. nofs 6.
--------------------------------------------------------------
    6 FORECASTS, BEGINNING AT ORIGIN = 996
--------------------------------------------------------------
 SERIES:       IBM                 SP5
 TIME    FORECAST  STD ERR   FORECAST  STD ERR
  997      1.954    6.952      1.698    5.480
```

```
 998      0.304    6.988      0.173    5.497
 999     -0.815    7.001     -1.263    5.497
1000      0.138    7.001     -0.494    5.507
1001      1.162    7.002      0.408    5.508
1002      1.294    7.022      0.649    5.528
```

8.2.5 Impulse Response Function

Similar to the univariate case, a VAR(p) model can be written as a linear function of the past innovations, that is,

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{a}_{t-2} + \cdots, \tag{8.21}$$

where $\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$ provided that the inverse exists, and the coefficient matrices $\boldsymbol{\psi}_i$ can be obtained by equating the coefficients of B^i in the equation

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)(\boldsymbol{I} + \boldsymbol{\Psi}_1 B + \boldsymbol{\Psi}_2 B^2 + \cdots) = \boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity matrix. This is a moving-average representation of $\boldsymbol{r}_t$ with the coefficient matrix $\boldsymbol{\Psi}_i$ being the impact of the past innovation $\boldsymbol{a}_{t-i}$ on $\boldsymbol{r}_t$. Equivalently, $\boldsymbol{\Psi}_i$ is the effect of $\boldsymbol{a}_t$ on the future observation $\boldsymbol{r}_{t+i}$. Therefore, $\boldsymbol{\Psi}_i$ is often referred to as the *impulse response function* of $\boldsymbol{r}_t$. However, since the components of $\boldsymbol{a}_t$ are often correlated, the interpretation of elements in $\boldsymbol{\Psi}_i$ of Eq. (8.21) is not straightforward. To aid interpretation, one can use the Cholesky decomposition mentioned earlier to transform the innovations so that the resulting components are uncorrelated. Specifically, there exists a lower triangular matrix $\boldsymbol{L}$ such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$, where $\boldsymbol{G}$ is a diagonal matrix and the diagonal elements of $\boldsymbol{L}$ are unity. See Eq. (8.9). Let $\boldsymbol{b}_t = \boldsymbol{L}^{-1}\boldsymbol{a}_t$. Then, $\text{Cov}(\boldsymbol{b}_t) = \boldsymbol{G}$ so that the elements b_{jt} are uncorrelated. Rewrite Eq. (8.21) as

$$\begin{aligned}
\boldsymbol{r}_t &= \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{a}_{t-2} + \cdots \\
&= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_t + \boldsymbol{\Psi}_1 \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{a}_{t-2} + \cdots \\
&= \boldsymbol{\mu} + \boldsymbol{\Psi}_0^* \boldsymbol{b}_t + \boldsymbol{\Psi}_1^* \boldsymbol{b}_{t-1} + \boldsymbol{\Psi}_2^* \boldsymbol{b}_{t-2} + \cdots,
\end{aligned} \tag{8.22}$$

where $\boldsymbol{\Psi}_0^* = \boldsymbol{L}$ and $\boldsymbol{\Psi}_i^* = \boldsymbol{\Psi}_i \boldsymbol{L}$. The coefficient matrices $\boldsymbol{\Psi}_i^*$ are called the *impulse response function* of $\boldsymbol{r}_t$ with respect to the orthogonal innovations $\boldsymbol{b}_t$. Specifically, the (i, j)th element of $\boldsymbol{\Psi}_\ell^*$; that is, $\Psi_{ij}^*(\ell)$, is the impact of $b_{j,t}$ on the future observation $r_{i,t+\ell}$. In practice, one can further normalize the orthogonal innovation $\boldsymbol{b}_t$ such that the variance of b_{it} is one. A weakness of the above orthogonalization is that the result depends on the ordering of the components of $\boldsymbol{r}_t$. In particular, $b_{1t} = a_{1t}$ so that a_{1t} is not transformed. Different orderings of the components of $\boldsymbol{r}_t$ may lead to different impulse response functions. Interpretation of the impulse response function is, therefore, associated with the innovation series $\boldsymbol{b}_t$.

Both SCA and S-Plus enable one to obtain the impulse response function of a fitted VAR model. To demonstrate analysis of VAR models in S-Plus, we again use the monthly log return series of IBM stock and the S&P 500 index of Example 8.1. For details of S-Plus commands, see Zivot and Wang (2003).

S-Plus Demonstration

The following output has been edited and % denotes explanation:

```
> module(finmetrics)
> da=read.table("m-ibmsp2608.txt",header=T) % Load data
> ibm=log(da[,2]+1)*100 % Compute percentage log returns
> sp5=log(da[,3]+1)*100
> y=cbind(ibm,sp5) % Create a vector series
> y1=data.frame(y) % Crate a data frame
> ord.choice=VAR(y1,max.ar=10) % Order selection using BIC
> names(ord.choice)
 [1] "R"    "coef" "fitted"   "residuals" "Sigma" "df.resid"
 [7] "rank" "call" "ar.order" "n.na"      "terms" "Y0"
[13] "info"
> ord.choice$ar.order % selected order
[1] 1
> ord.choice$info
       ar(1)    ar(2)    ar(3)    ar(4)    ar(5)   ar(6)
BIC 12325.41 12339.42 12356.58 12376.28 12391.57 12417.2
        ar(7)    ar(8)   ar(9)   ar(10)
BIC  12442.03 12462.5212484.78 12510.91
> ord=VAR(y1,max.ar=10,criterion='AIC') % Using AIC
> ord$ar.order
[1] 5
> ord$info
       ar(1)    ar(2)     ar(3)   ar(4)    ar(5)    ar(6)
AIC 12296.04 12290.48  12288.07 12288.2 12283.91 12289.96
       ar(7)    ar(8)     ar(9)   ar(10)
AIC 12295.22 12296.13  12298.82 12305.37
```

The AIC selects a VAR(5) model as before, but BIC selects a VAR(1) model. For simplicity, we shall use VAR(1) specification in the demonstration. Note that different normalizations are used between the two packages so that the values of information criteria appear to be different; see the AIC in Table 8.3. This is not important because normalization does not affect order selection. Turn to estimation.

```
> var1.fit=VAR(y~ar(1)) % Estimate a VAR(1) model
> summary(var1.fit)
Call:
VAR(formula = y ~ ar(1))
Coefficients:
                 ibm     sp5
(Intercept)   1.0614  0.4087
  (std.err)   0.2249  0.1773
```

```
   (t.stat)   4.7198  2.3053

    ibm.lag1  -0.0320 -0.0223
   (std.err)   0.0413   0.0326
    (t.stat)  -0.7728 -0.6855

    sp5.lag1  0.1503  0.1020
   (std.err)  0.0525  0.0414
    (t.stat)  2.8612  2.4637

Regression Diagnostics:
                  ibm    sp5
     R-squared 0.0101 0.0075
Adj. R-squared 0.0081 0.0055
  Resid. Scale 7.0078 5.5247

Information Criteria:
       logL        AIC        BIC         HQ
  -6193.988  12399.977  12429.393  12411.159

                  total residual
Degree of freedom:   995      992

> plot(var1.fit)
Make a plot selection (or 0 to exit):
1: plot: All
2: plot: Response and Fitted Values
3: plot: Residuals
 ....
8: plot: PACF of Squared Residuals
Selection: 3
```

The fitted model is

$$\text{IBM}_t = 1.06 - 0.03\text{IBM}_{t-1} + 0.15\text{SP5}_{t-1} + a_{1t},$$
$$\text{SP5}_t = 0.41 - 0.02\text{IBM}_{t-1} + 0.10\text{SP5}_{t-1} + a_{2t}.$$

Based on t statistics of the coefficient estimates, only the lagged variable SP5_{t-1} is informative in both equations. Figure 8.5 shows the time plots of the two residual series, where the two horizontal lines indicate the two standard error limits. As expected, there exist clusters of outlying observations.

Next, we compute 1-step- to 6-step-ahead forecasts and the impulse response function of the fitted VAR(1) model when the IBM stock return is the first component of $\boldsymbol{r}_t$. Compared with those of a VAR(5) model in Table 8.5, the forecasts of the VAR(1) model converge faster to the sample mean of the series.

```
> var1.pred=predict(var1.fit,n.predict=6) % Compute forecasts
> summary(var1.pred)
```

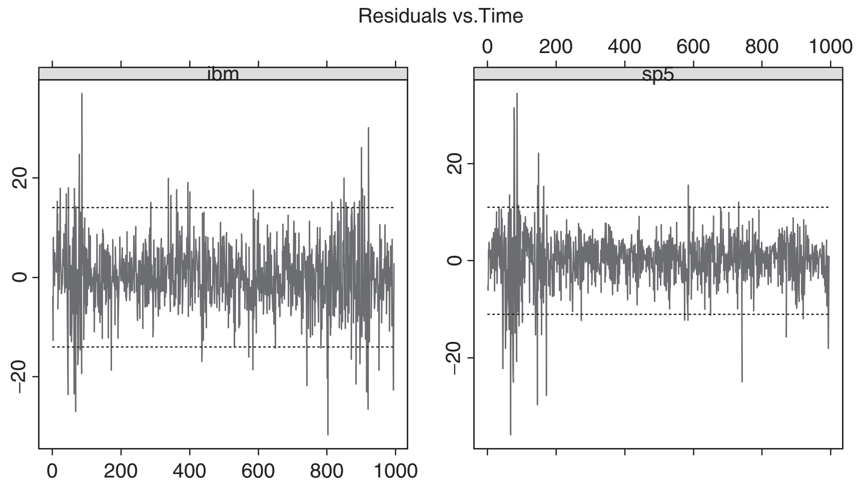


Figure 8.5 Residual plots of fitting a VAR(1) model to the monthly log returns, in percentages, of IBM stock and S&P 500 index. Sample period is from January 1926 to December 2008.

```
Predicted Values with Standard Errors:
                ibm    sp5
1-step-ahead 1.0798 0.4192
   (std.err) 7.0078 5.5247
2-step-ahead 1.0899 0.4274
   (std.err) 7.0434 5.5453
3-step-ahead 1.0908 0.4280
   (std.err) 7.0436 5.5454
 ...
6-step-ahead 1.0909 0.4280
   (std.err) 7.0436 5.5454
> plot(var1.pred,y,n.old=12) % Obtain forecast plot
% Below is to compute the impulse response function
> var1.irf=impRes(var1.fit,period=6,std.err='asymptotic')
> summary(var1.irf)
Impulse Response Function:
(with responses in rows, and innovations in columns)
, , lag.0
             ibm    sp5
      ibm 6.9973 0.0000
(std.err) 0.1569 0.0000
      sp5 3.5432 4.2280
(std.err) 0.1558 0.0948

, , lag.1
             ibm    sp5
      ibm 0.3088 0.6353
(std.err) 0.2217 0.2221
```

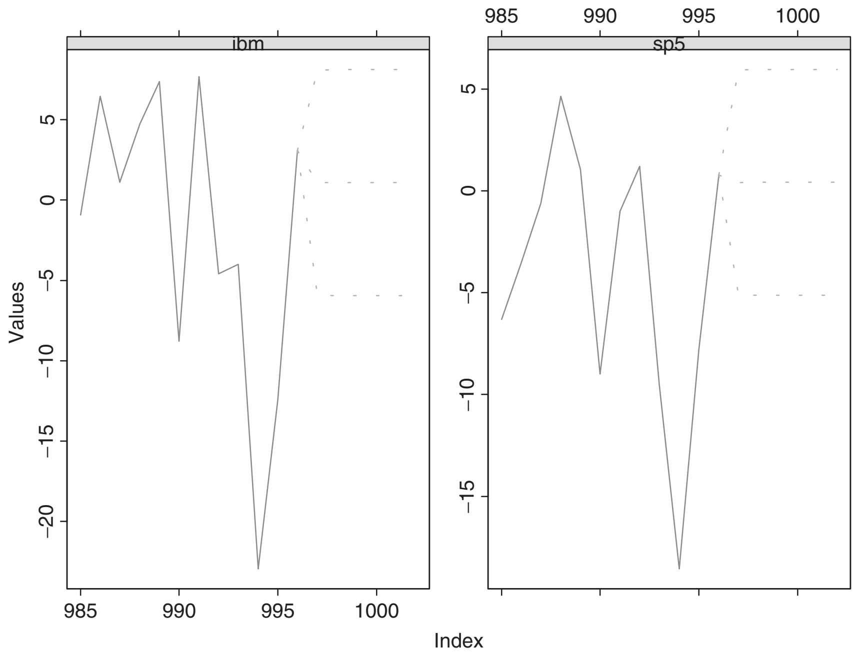


Figure 8.6 Forecasting plots of fitted VAR(1) model to monthly log returns, in percentages, of IBM stock and S&P 500 index. Sample period is from January 1926 to December 2008.

```
      sp5 0.2050 0.4312
(std.err) 0.1746 0.1750
 .....
> plot(var1.irf)
```

Figure 8.6 shows the forecasts and their pointwise 95% confidence intervals along with the last 12 data points of the series. Figure 8.7 shows the impulse response functions of the fitted VAR(1) model where the IBM stock return is the first component of $\boldsymbol{r}_t$. Since the dynamic dependence of the returns is weak, the impulse response functions exhibit simple patterns and decay quickly.

8.3 VECTOR MOVING-AVERAGE MODELS

A vector moving-average model of order q, or VMA(q), is in the form

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1 \boldsymbol{a}_{t-1} - \cdots - \boldsymbol{\Theta}_q \boldsymbol{a}_{t-q} \quad \text{or} \quad \boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{8.23}$$

where $\boldsymbol{\theta}_0$ is a k-dimensional vector, $\boldsymbol{\Theta}_i$ are $k \times k$ matrices, and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ is the MA matrix polynomial in the back-shift operator B. Similar to the univariate case, VMA(q) processes are weakly stationary provided

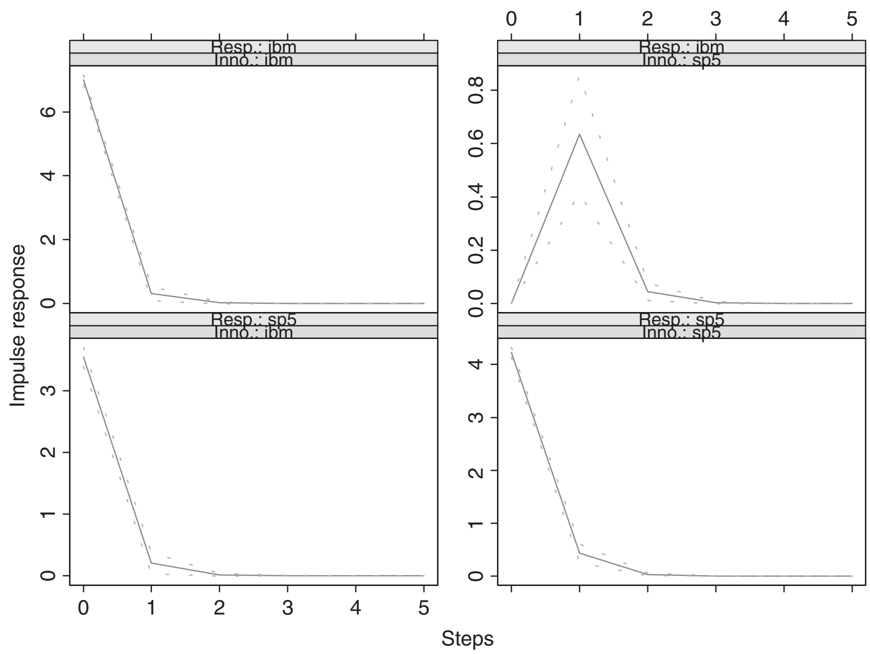


Figure 8.7 Plots of impulse response functions of orthogonal innovations for fitted VAR(1) model to monthly log returns, in percentages, of IBM stock and S&P 500 index. Sample period is from January 1926 to December 2008.

that the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$ exists. Taking expectation of Eq. (8.23), we obtain that $\boldsymbol{\mu} = E(\boldsymbol{r}_t) = \boldsymbol{\theta}_0$. Thus, the constant vector $\boldsymbol{\theta}_0$ is the mean vector of $\boldsymbol{r}_t$ for a VMA model.

Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$ be the mean-corrected VAR(q) process. Then using Eq. (8.23) and the fact that $\{\boldsymbol{a}_t\}$ has no serial correlations, we have

1. $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$.
2. $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1\boldsymbol{\Sigma}\boldsymbol{\Theta}_1' + \cdots + \boldsymbol{\Theta}_q\boldsymbol{\Sigma}\boldsymbol{\Theta}_q'$.
3. $\boldsymbol{\Gamma}_\ell = \mathbf{0}$ if $\ell > q$.
4. $\boldsymbol{\Gamma}_\ell = \sum_{j=\ell}^{q} \boldsymbol{\Theta}_j\boldsymbol{\Sigma}\boldsymbol{\Theta}_{j-\ell}'$ if $1 \le \ell \le q$, where $\boldsymbol{\Theta}_0 = -\boldsymbol{I}$.

Since $\boldsymbol{\Gamma}_\ell = \mathbf{0}$ for $\ell > q$, the cross-correlation matrices (CCMs) of a VMA(q) process $\boldsymbol{r}_t$ satisfy

$$\boldsymbol{\rho}_\ell = \mathbf{0}, \qquad \ell > q. \tag{8.24}$$

Therefore, similar to the univariate case, the sample CCMs can be used to identify the order of a VMA process.

To better understand the VMA processes, let us consider the bivariate MA(1) model

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}\boldsymbol{a}_{t-1} = \boldsymbol{\mu} + \boldsymbol{a}_t - \boldsymbol{\Theta}\boldsymbol{a}_{t-1}, \tag{8.25}$$

where, for simplicity, the subscript of $\boldsymbol{\Theta}_1$ is removed. This model can be written explicitly as

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \tag{8.26}$$

It says that the current return series $\boldsymbol{r}_t$ only depends on the current and past shocks. Therefore, the model is a finite-memory model.

Consider the equation for r_{1t} in Eq. (8.26). The parameter Θ_{12} denotes the linear dependence of r_{1t} on $a_{2,t-1}$ in the presence of $a_{1,t-1}$. If $\Theta_{12} = 0$, then r_{1t} does not depend on the lagged values of a_{2t} and, hence, the lagged values of r_{2t}. Similarly, if $\Theta_{21} = 0$, then r_{2t} does not depend on the past values of r_{1t}. The off-diagonal elements of $\boldsymbol{\Theta}$ thus show the dynamic dependence between the component series. For this simple VMA(1) model, we can classify the relationships between r_{1t} and r_{2t} as follows:

1. They are uncoupled series if $\Theta_{12} = \Theta_{21} = 0$.
2. There is a unidirectional dynamic relationship from r_{1t} to r_{2t} if $\Theta'_{12} = 0$, but $\Theta_{21} \neq 0$. The opposite unidirectional relationship holds if $\Theta_{21} = 0$, but $\Theta_{12} \neq 0$.
3. There is a feedback relationship between r_{1t} and r_{2t} if $\Theta_{12} \neq 0$ and $\Theta_{21} \neq 0$.

Finally, the concurrent correlation between r_{it} is the same as that between a_{it}. The previous classification can be generalized to a VMA(q) model.

Estimation

Unlike the VAR models, estimation of VMA models is much more involved; see Hillmer and Tiao (1979), Lütkepohl (2005), and the references therein. For the likelihood approach, there are two methods available. The first is the conditional-likelihood method that assumes that $\boldsymbol{a}_t = \boldsymbol{0}$ for $t \leq 0$. The second is the exact-likelihood method that treats $\boldsymbol{a}_t$ with $t \leq 0$ as additional parameters of the model. To gain some insight into the problem of estimation, we consider the VMA(1) model in Eq. (8.25). Suppose that the data are $\{\boldsymbol{r}_t | t = 1, \ldots, T\}$ and $\boldsymbol{a}_t$ is multivariate normal. For a VMA(1) model, the data depend on $\boldsymbol{a}_0$.

Conditional MLE

The conditional-likelihood method assumes that $\boldsymbol{a}_0 = \boldsymbol{0}$. Under such an assumption and rewriting the model as $\boldsymbol{a}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}$, we can compute the shock $\boldsymbol{a}_t$ recursively as

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0, \qquad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}_1\boldsymbol{a}_1, \quad \ldots.$$

Consequently, the likelihood function of the data becomes

$$f(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T|\boldsymbol{\theta}_0, \boldsymbol{\Theta}_1, \boldsymbol{\Sigma}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{a}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_t\right),$$

which can be evaluated to obtain the parameter estimates.

Exact MLE

For the exact-likelihood method, $\boldsymbol{a}_0$ is an unknown vector that must be estimated from the data to evaluate the likelihood function. For simplicity, let $\tilde{r}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$ be the mean-corrected series. Using $\tilde{\boldsymbol{r}}_t$ and Eq. (8.25), we have

$$\boldsymbol{a}_t = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}. \tag{8.27}$$

By repeated substitutions, $\boldsymbol{a}_0$ is related to all $\tilde{\boldsymbol{r}}_t$ as

$$\begin{aligned}
\boldsymbol{a}_1 &= \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\boldsymbol{a}_0,\\
\boldsymbol{a}_2 &= \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_2 + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^2\boldsymbol{a}_0,\\
&\vdots\\
\boldsymbol{a}_T &= \tilde{\boldsymbol{r}}_T + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{T-1} + \cdots + \boldsymbol{\Theta}^{T-1}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^T\boldsymbol{a}_0.
\end{aligned} \tag{8.28}$$

Thus, $\boldsymbol{a}_0$ is a linear function of the data if $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}$ are given. This result enables us to estimate $\boldsymbol{a}_0$ using the data and initial estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}$. More specifically, given $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and the data, we can define

$$\boldsymbol{r}_t^* = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Theta}^{t-1}\tilde{\boldsymbol{r}}_1, \qquad \text{for} \quad t = 1, 2, \ldots, T.$$

Equation (8.28) can then be rewritten as

$$\begin{aligned}
\boldsymbol{r}_1^* &= -\boldsymbol{\Theta}\boldsymbol{a}_0 + \boldsymbol{a}_1,\\
\boldsymbol{r}_2^* &= -\boldsymbol{\Theta}^2\boldsymbol{a}_0 + \boldsymbol{a}_2,\\
&\vdots\\
\boldsymbol{r}_T^* &= -\boldsymbol{\Theta}^T\boldsymbol{a}_0 + \boldsymbol{a}_T.
\end{aligned}$$

This is in the form of a multiple linear regression with parameter vector $\boldsymbol{a}_0$, even though the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{a}_t$ may not be a diagonal matrix. If initial estimate of $\boldsymbol{\Sigma}$ is also available, one can premultiply each equation of the prior system by $\boldsymbol{\Sigma}^{-1/2}$, which is the square root matrix of $\boldsymbol{\Sigma}$. The resulting system is indeed a multiple linear regression, and the ordinary least-squares method can be used to obtain an estimate of $\boldsymbol{a}_0$. Denote the estimate by $\widehat{\boldsymbol{a}}_0$.

Using the estimate $\widehat{\boldsymbol{a}}_0$, we can compute the shocks $\boldsymbol{a}_t$ recursively as

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\widehat{a}_0, \qquad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_1, \quad \ldots .$$

This recursion is a linear transformation from $(\boldsymbol{a}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ to $(\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_T)$, from which we can (a) obtain the joint distribution of $\boldsymbol{a}_0$ and the data, and (2) integrate out $\boldsymbol{a}_0$ to derive the exact-likelihood function of the data. The resulting likelihood function can then be evaluated to obtain the exact ML estimates. For details, see Hillmer and Tiao (1979).

In summary, the exact-likelihood method works as follows. Given initial estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$, one uses Eq. (8.28) to derive an estimate of $\boldsymbol{a}_0$. This estimate is in turn used to compute $\boldsymbol{a}_t$ recursively using Eq. (8.27) and starting with $\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\widehat{\boldsymbol{a}}_0$. The resulting $\{\boldsymbol{a}_t\}_{t=1}^{T}$ are then used to evaluate the exact-likelihood function of the data to update the estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Sigma}$. The whole process is then repeated until the estimates converge. This iterative method to evaluate the exact-likelihood function applies to the general VMA(q) models.

From the previous discussion, the exact-likelihood method requires more intensive computation than the conditional-likelihood approach does. But it provides more accurate parameter estimates, especially when some eigenvalues of $\boldsymbol{\Theta}$ are close to 1 in modulus. Hillmer and Tiao (1979) provide some comparison between the conditional- and exact-likelihood estimations of VMA models. In multivariate time series analysis, the exact maximum-likelihood method becomes important if one suspects that the data might have been overdifferenced. Overdifferencing may occur in many situations (e.g., differencing individual components of a cointegrated system; see discussion later on cointegration).

In summary, building a VMA model involves three steps: (a) Use the sample cross-correlation matrices to specify the order q—for a VMA(q) model, $\boldsymbol{\rho}_\ell = \boldsymbol{0}$ for $\ell > q$; (b) estimate the specified model by using either the conditional- or exact-likelihood method—the exact method is preferred when the sample size is not large; and (c) the fitted model should be checked for adequacy [e.g., applying the $Q_k(m)$ statistics to the residual series]. Finally, forecasts of a VMA model can be obtained by using the same procedure as a univariate MA model.

Example 8.5. Consider again the bivariate series of monthly log returns in percentages of IBM stock and the S&P 500 index from January 1926 to December 2008. Since significant cross correlations occur mainly at lags 1, 2, 3 and 5, we employ the VMA(5) model

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1\boldsymbol{a}_{t-1} - \boldsymbol{\Theta}_2\boldsymbol{a}_{t-2} - \boldsymbol{\Theta}_3\boldsymbol{a}_{t-3} - \boldsymbol{\Theta}_5\boldsymbol{a}_{t-5} \tag{8.29}$$

for the data. Table 8.6 shows the estimation results of the model. The $Q_k(m)$ statistics for the residuals of the simplified model give $Q_2(4) = 16.00$ and $Q_2(8) = 29.46$. Compared with chi-squared distributions with 10 and 26 degrees of freedom, the p values of these statistics are 0.10 and 0.291, respectively. Thus, the model is adequate at the 5% significance level.

TABLE 8.6 Estimation Results for Monthly Log Returns of IBM Stock and S&P 500 Index Using the Vector Moving-Average Model in Eq. (8.29)[a]

Parameter	$\boldsymbol{\theta}_0$	$\boldsymbol{\Theta}_1$		$\boldsymbol{\Theta}_2$		$\boldsymbol{\Theta}_3$		$\boldsymbol{\Theta}_5$	
(a) Full Model with Conditional-Likelihood Method									
Estimate	1.1	0.02	−0.15	−0.09	0.15	−0.05	0.11	0.05	−0.15
	0.4	0.02	−0.10	−0.04	0.04	−0.01	0.11	0.07	−0.15
Standard error	0.24	0.04	0.05	0.04	0.05	0.04	0.05	0.04	0.05
	0.19	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04
(b) Full Model with Exact-Likelihood Method									
Estimate	1.1	0.02	−0.05	−0.09	0.15	−0.05	0.11	0.05	−0.15
	0.4	0.02	−0.10	−0.04	0.04	−0.01	0.11	0.07	−0.15
Standard error	0.24	0.04	0.05	0.04	0.05	0.04	0.05	0.04	0.05
	0.19	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04
(c) Simplified Model with Exact-Likelihood Method									
Estimate	1.1	0.0	−0.13	0.0	0.08	0.0	0.0	0.0	−0.10
	0.4	0.0	−0.09	0.0	0.0	0.0	0.07	0.0	−0.09
Standard error	0.24	—	0.04	—	0.03	—	—	—	0.04
	0.19	—	0.03	—	—	—	0.02	—	0.03

[a]The sample period is from January 1926 to December 2008. The residual covariance matrix is not shown as it is similar to that in Table 8.4

From Table 8.6, we make the following observations:

1. The difference between conditional- and exact-likelihood estimates is small for this particular example. This is not surprising because the sample size is not small and, more important, the dynamic structure of the data is weak.
2. The VMA(5) model provides essentially the same dynamic relationship for the series as that of the VAR(5) model in Example 8.4. The monthly log return of IBM stock depends on the previous returns of the S&P 500 index. The market return, in contrast, does not depend on lagged returns of IBM stock. In other words, the dynamic structure of the data is driven by the market return, not by the IBM return. The concurrent correlation between the two returns remains strong, however.

8.4 VECTOR ARMA MODELS

Univariate ARMA models can also be generalized to handle vector time series. The resulting models are called VARMA models. The generalization, however, encounters some new issues that do not occur in developing VAR and VMA models. One of the issues is the *identifiability* problem. Unlike the univariate ARMA models, VARMA models may not be uniquely defined. For example, the VMA(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

is *identical* to the VAR(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The equivalence of the two models can easily be seen by examining their component models. For the VMA(1) model, we have

$$r_{1t} = a_{1t} - 2a_{2,t-1}, \qquad r_{2t} = a_{2t}.$$

For the VAR(1) model, the equations are

$$r_{1t} + 2r_{2,t-1} = a_{1t}, \qquad r_{2t} = a_{2t}.$$

From the model for r_{2t}, we have $r_{2,t-1} = a_{2,t-1}$. Therefore, the models for r_{1t} are identical. This type of identifiability problem is harmless because either model can be used in a real application.

Another type of identifiability problem is more troublesome. Consider the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

This model is identical to the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix},$$

for any nonzero ω and η. In this particular instance, the equivalence occurs because we have $r_{2t} = a_{2t}$ in both models. The effects of the parameters ω and η on the system cancel out between AR and MA parts of the second model. Such an identifiability problem is serious because, without proper constraints, the likelihood function of a vector ARMA(1,1) model for the data is not uniquely defined, resulting in a situation similar to the exact multicollinearity in a regression analysis. This type of identifiability problem can occur in a vector model even if none of the components is a white noise series.

These two simple examples highlight the new issues involved in the generalization to VARMA models. Building a VARMA model for a given data set thus requires some attention. In the time series literature, methods of *structural specification* have been proposed to overcome the identifiability problem; see Tiao and Tsay (1989), Tsay (1991), and the references therein. We do not discuss the detail of structural specification here because VAR and VMA models are sufficient in most financial applications. When VARMA models are used, only lower order models are entertained [e.g., a VARMA(1,1) or VARMA(2,1) model] especially when the time series involved are not seasonal.

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

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t,$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are two $k \times k$ matrix polynomials. We assume that the two matrix polynomials have no left common factors; otherwise, the model can be simplified. The necessary and sufficient condition of weak stationarity for $\boldsymbol{r}_t$ is the same as that for the VAR(p) model with matrix polynomial $\boldsymbol{\Phi}(B)$. For $v > 0$, the (i, j)th elements of the coefficient matrices $\boldsymbol{\Phi}_v$ and $\boldsymbol{\Theta}_v$ measure the linear dependence of r_{1t} on $r_{j,t-v}$ and $a_{j,t-v}$, respectively. If the (i, j)th element is zero for all AR and MA coefficient matrices, then r_{it} does not depend on the lagged values of r_{jt}. However, the converse proposition does not hold in a VARMA model. In other words, nonzero coefficients at the (i, j)th position of AR and MA matrices may exist even when r_{it} does not depend on any lagged value of r_{jt}.

To illustrate, consider the following bivariate model

$$\begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \Theta_{11}(B) & \Theta_{12}(B) \\ \Theta_{21}(B) & \Theta_{22}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Here the necessary and sufficient conditions for the existence of a unidirectional dynamic relationship from r_{1t} to r_{2t} are

$$\Phi_{22}(B)\Theta_{12}(B) - \Phi_{12}(B)\Theta_{22}(B) = 0,$$

but

$$\Phi_{11}(B)\Theta_{21}(B) - \Phi_{21}(B)\Theta_{11}(B) \neq 0. \tag{8.30}$$

These conditions can be obtained as follows. Letting

$$\Omega(B) = |\boldsymbol{\Phi}(B)| = \Phi_{11}(B)\Phi_{22}(B) - \Phi_{12}(B)\Phi_{21}(B)$$

be the determinant of the AR matrix polynomial and premultiplying the model by the matrix

$$\begin{bmatrix} \Phi_{22}(B) & -\Phi_{12}(B) \\ -\Phi_{21}(B) & \Phi_{11}(B) \end{bmatrix},$$

we can rewrite the bivariate model as

$$\begin{aligned} \Omega(B)\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} &= \begin{bmatrix} \Phi_{22}(B)\Theta_{11}(B) - \Phi_{12}(B)\Theta_{21}(B) & \Phi_{22}(B)\Theta_{12}(B) - \Phi_{12}(B)\Theta_{22}(B) \\ \Phi_{11}(B)\Theta_{21}(B) - \Phi_{21}(B)\Theta_{11}(B) & \Phi_{11}(B)\Theta_{22}(B) - \Phi_{21}(B)\Theta_{12}(B) \end{bmatrix} \\ &\quad \times \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}. \end{aligned}$$

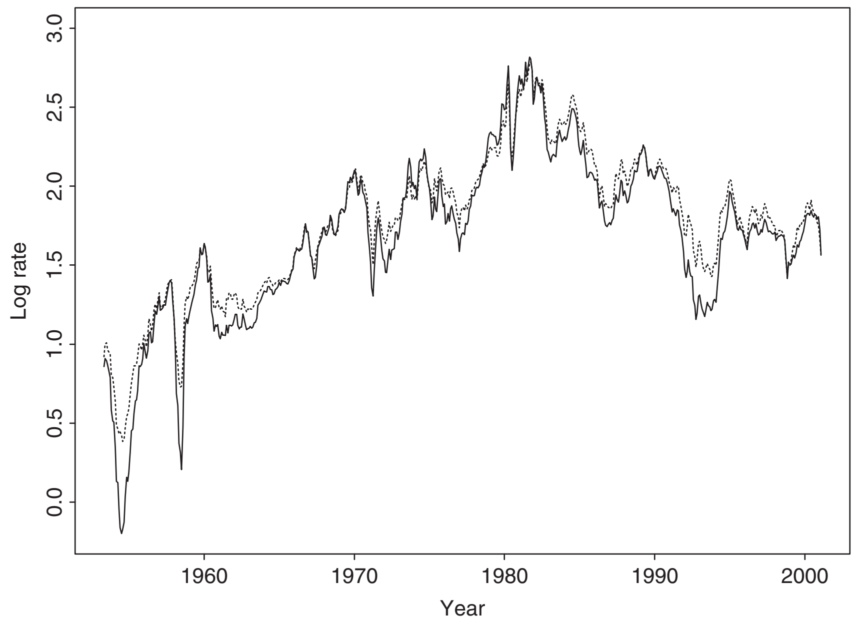


Figure 8.8 Time plots of log U.S. monthly interest rates from April 1953 to January 2001. Solid line denotes 1-year Treasury constant maturity rate and dashed line denotes 3-year rate.

Consider the equation for r_{1t}. The first condition in Eq. (8.30) shows that r_{1t} does not depend on any past value of a_{2t} or r_{2t}. From the equation for r_{2t}, the second condition in Eq. (8.30) implies that r_{2t} indeed depends on some past values of a_{1t}. Based on Eq. (8.30), $\Theta_{12}(B) = \Phi_{12}(B) = 0$ is a sufficient, but not necessary, condition for the unidirectional relationship from r_{1t} to r_{2t}.

Estimation of a VARMA model can be carried out by either the conditional or exact maximum-likelihood method. The $Q_k(m)$ statistic continues to apply to the residual series of a fitted model, but the degrees of freedom of its asymptotic chi-squared distribution are $k^2m - g$, where g is the number of estimated parameters in both the AR and MA coefficient matrices.

Example 8.6. To demonstrate VARMA modeling, we consider two U.S. monthly interest rate series. The first series is the 1-year Treasury constant maturity rate, and the second series is the 3-year Treasury constant maturity rate. The data are obtained from the Federal Reserve Bank of St. Louis, and the sampling period is from April 1953 to January 2001. There are 574 observations. To ensure the positiveness of U.S. interest rates, we analyze the log series. Figure 8.8 shows the time plots of the two log interest rate series. The solid line denotes the 1-year maturity rate. The two series moved closely in the sampling period.

The $M(i)$ statistics and AIC criterion specify a VAR(4) model for the data. However, we employ a VARMA(2,1) model because the two models provide similar

TABLE 8.7 Parameter Estimates of VARMA(2,1) Model for Two Monthly U.S. Interest Rate Series Based on Exact-Likelihood Method

Parameter	$\mathbf{\Phi}_1$		$\mathbf{\Phi}_2$		$\boldsymbol{\phi}_0$	$\mathbf{\Theta}_1$		$\mathbf{\Sigma} \times 10^3$	
Estimate	1.82	−0.97	−0.84	0.98	0.028	0.90	−1.66	3.58	2.50
	—	0.99	—	—	0.025	—	−0.47	2.50	2.19
Standard error	0.03	0.08	0.03	0.08	0.014	0.03	0.10		
	—	0.01	—	—	0.011	—	0.04		

fits. Table 8.7 shows the parameter estimates of the VARMA(2,1) model obtained by the exact-likelihood method. We removed the insignificant parameters and reestimated the simplified model. The residual series of the fitted model has some minor serial and cross correlations at lags 7 and 11. Figure 8.9 shows the residual plots and indicates the existence of some outlying data points. The model can be further improved, but it seems to capture the dynamic structure of the data reasonably well.

The final VARMA(2,1) model shows some interesting characteristics of the data. First, the interest rate series are highly contemporaneously correlated. The concurrent correlation coefficient is $2.5/\sqrt{3.58 \times 2.19} = 0.893$. Second, there is a

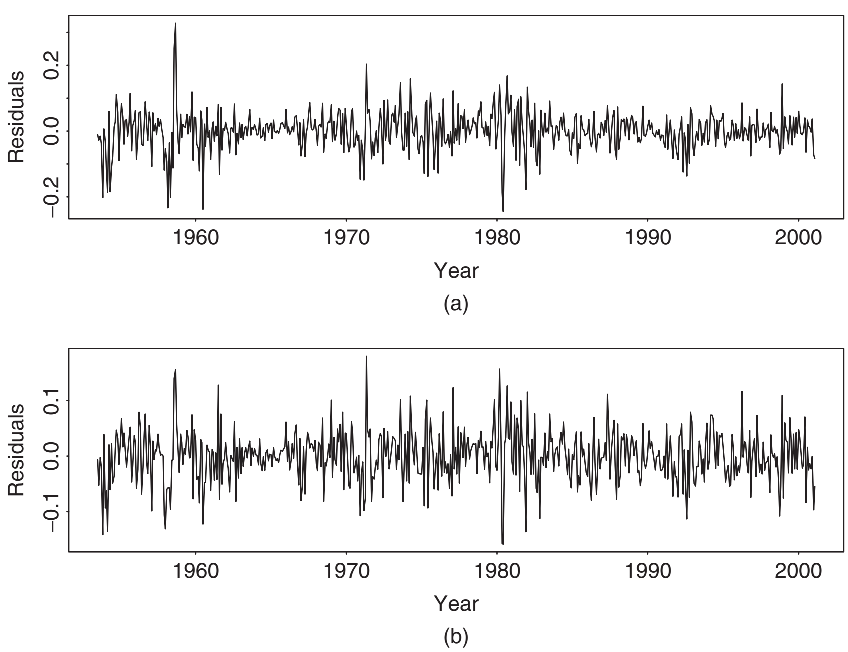


Figure 8.9 Residual plots for log U.S. monthly interest rate series of Example 8.6. Fitted model is VARMA(2,1): (a) 1-year rate and (b) 3-year rate.

unidirectional linear relationship from the 3-year rate to the 1-year rate because the (2, 1)th elements of all AR and MA matrices are zero, but some (1, 2)th element is not zero. As a matter of fact, the model in Table 8.7 shows that

$$\begin{aligned} r_{3t} &= 0.025 + 0.99 r_{3,t-1} + a_{3t} + 0.47 a_{3,t-1}, \\ r_{1t} &= 0.028 + 1.82 r_{1,t-1} - 0.84 r_{1,t-2} - 0.97 r_{3,t-1} + 0.98 r_{3,t-2} \\ &\quad + a_{1t} - 0.90 a_{1,t-1} + 1.66 a_{3,t-1}, \end{aligned}$$

where r_{it} is the log series of i-year interest rate and a_{it} is the corresponding shock series. Therefore, the 3-year interest rate does not depend on the past values of the 1-year rate, but the 1-year rate depends on the past values of the 3-year rate. Third, the two interest rate series appear to be unit-root nonstationary. Using the back-shift operator B, the model can be rewritten approximately as

$$\begin{aligned} (1-B)r_{3t} &= 0.03 + (1 + 0.47B)a_{3t}, \\ (1-B)(1-0.82B)r_{1t} &= 0.03 - 0.97B(1-B)r_{3,t} + (1-0.9B)a_{1t} + 1.66Ba_{3,t}. \end{aligned}$$

Finally, the SCA commands used in the analysis are given in Appendix C.

8.4.1 Marginal Models of Components

Given a vector model for $\boldsymbol{r}_t$, the implied univariate models for the components r_{it} are the *marginal* models. For a k-dimensional ARMA(p, q) model, the marginal models are ARMA$[kp, (k-1)p+q]$. This result can be obtained in two steps. First, the marginal model of a VMA(q) model is univariate MA(q). Assume that $\boldsymbol{r}_t$ is a VMA(q) process. Because the cross-correlation matrix of $\boldsymbol{r}_t$ vanishes after lag q (i.e., $\boldsymbol{\rho}_\ell = \boldsymbol{0}$ for $\ell > q$), the ACF of r_{it} is zero beyond lag q. Therefore, r_{it} is an MA process and its univariate model is in the form $r_{it} = \theta_{i,0} + \sum_{j=1}^{q} \theta_{i,j} b_{i,t-j}$, where $\{b_{it}\}$ is a sequence of uncorrelated random variables with mean zero and variance σ_{ib}^2. The parameters $\theta_{i,j}$ and σ_{ib} are functions of the parameters of the VMA model for $\boldsymbol{r}_t$.

The second step to obtain the result is to diagonalize the AR matrix polynomial of a VARMA(p, q) model. For illustration, consider the bivariate AR(1) model

$$\begin{bmatrix} 1-\Phi_{11}B & -\Phi_{12}B \\ -\Phi_{21}B & 1-\Phi_{22}B \end{bmatrix} \begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Premultiplying the model by the matrix polynomial

$$\begin{bmatrix} 1-\Phi_{22}B & \Phi_{12}B \\ \Phi_{21}B & 1-\Phi_{11}B \end{bmatrix},$$

we obtain

$$[(1-\Phi_{11}B)(1-\Phi_{22}B) - \Phi_{12}\Phi_{22}B^2]\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} 1-\Phi_{22}B & -\Phi_{12}B \\ -\Phi_{21}B & 1-\Phi_{11}B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The left-hand side of the prior equation shows that the univariate AR polynomials for r_{it} are of order 2. In contrast, the right-hand side of the equation is in a VMA(1) form. Using the result of VMA models in step 1, we show that the univariate model for r_{it} is ARMA(2,1). The technique generalizes easily to the k-dimensional VAR(1) model, and the marginal models are ARMA(k, $k-1$). More generally, for a k-dimensional VAR(p) model, the marginal models are ARMA[kp, $(k-1)p$]. The result for VARMA models follows directly from those of VMA and VAR models.

The order [kp, $(k-1)p+q$] is the maximum order (i.e., the upper bound) for the marginal models. The actual marginal order of r_{it} can be much lower.

8.5 UNIT-ROOT NONSTATIONARITY AND COINTEGRATION

When modeling several unit-root nonstationary time series jointly, one may encounter the case of *cointegration*. Consider the bivariate ARMA(1,1) model

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}, \qquad (8.31)$$

where the covariance matrix $\boldsymbol{\Sigma}$ of the shock $\boldsymbol{a}_t$ is positive definite. This is not a weakly stationary model because the two eigenvalues of the AR coefficient matrix are 0 and 1. Figure 8.10 shows the time plots of a simulated series of the model with 200 data points and $\boldsymbol{\Sigma} = \boldsymbol{I}$, whereas Figure 8.11 shows the sample autocorrelations of the two component series x_{it}. It is easy to see that the two series have high autocorrelations and exhibit features of unit-root nonstationarity. The two marginal models of $\boldsymbol{x}_t$ are indeed unit-root nonstationary. Rewrite the model as

$$\begin{bmatrix} 1-0.5B & B \\ 0.25B & 1-0.5B \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.2B & 0.4B \\ 0.1B & 1-0.2B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Premultiplying the above equation by

$$\begin{bmatrix} 1-0.5B & -B \\ -0.25B & 1-0.5B \end{bmatrix},$$

we obtain the result

$$\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1-0.7B & -0.6B \\ -0.15B & 1-0.7B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Therefore, each component x_{it} of the model is unit-root nonstationary and follows an ARIMA(0,1,1) model.

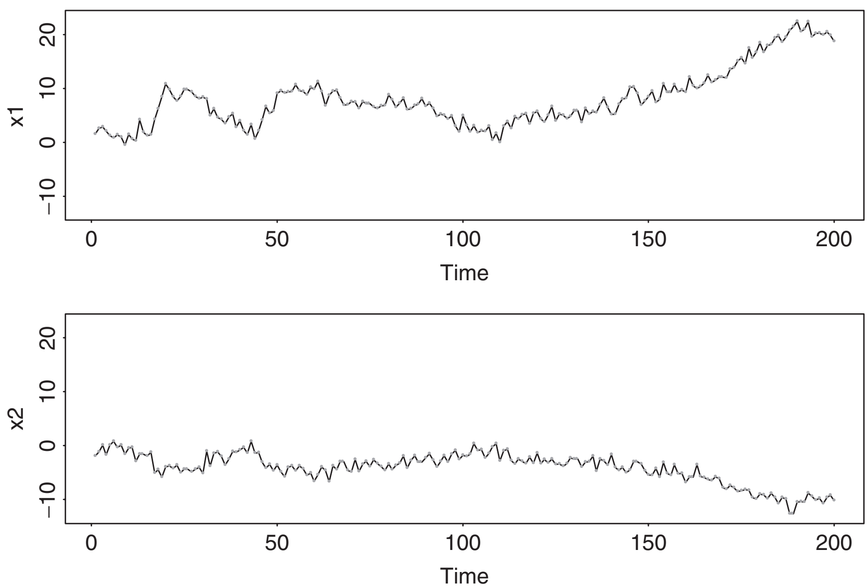


Figure 8.10 Time plots of simulated series based on model (8.31) with identity covariance matrix for shocks.

However, we can consider a linear transformation by defining

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{x}_t,$$

$$\begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{a}_t.$$

The VARMA model of the transformed series $\boldsymbol{y}_t$ can be obtained as follows:

$$\begin{aligned} \boldsymbol{L}\boldsymbol{x}_t &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{x}_{t-1}) + \boldsymbol{b}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{b}_{t-1}. \end{aligned}$$

Thus, the model for $\boldsymbol{y}_t$ is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{1,t-1} \\ b_{2,t-1} \end{bmatrix}. \quad (8.32)$$

From the prior model, we see that (a) y_{1t} and y_{2t} are uncoupled series with concurrent correlation equal to that between the shocks b_{1t} and b_{2t}, (b) y_{1t} follows a univariate ARIMA(0,1,1) model, and (c) y_{2t} is a white noise series (i.e., $y_{2t} = b_{2t}$).

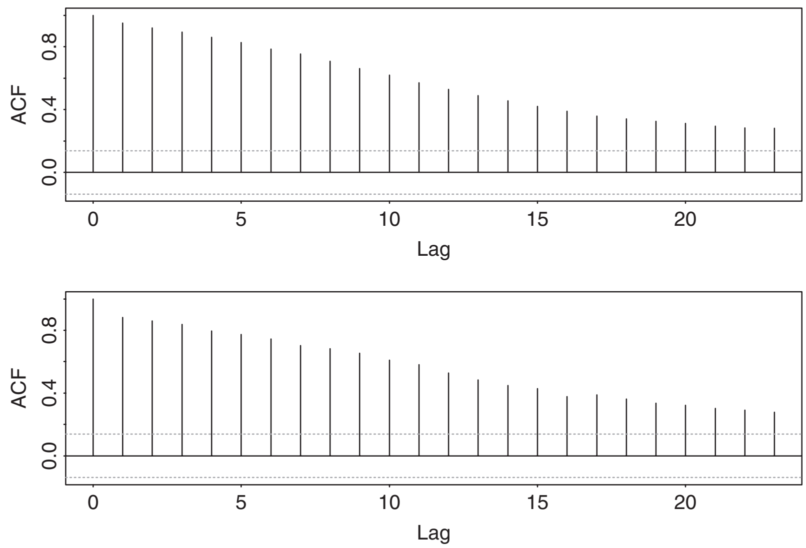


Figure 8.11 Sample autocorrelation functions of two simulated component series. There are 200 observations, and model is given by Eq. (8.31) with identity covariance matrix for shocks.

In particular, the model in Eq. (8.32) shows that there is *only* a single unit root in the system. Consequently, the unit roots of x_{1t} and x_{2t} are introduced by the unit root of y_{1t}. In the literature, y_{1t} is referred to as the *common trend* of x_{1t} and x_{2t}.

The phenomenon that both x_{1t} and x_{2t} are unit-root nonstationary, but there is only a single unit root in the vector series, is referred to as *cointegration* in the econometric and time series literature. Another way to define cointegration is to focus on linear transformations of unit-root nonstationary series. For the simulated example of model (8.31), the transformation shows that the linear combination $y_{2t} = 0.5x_{1t} + x_{2t}$ does not have a unit root. Consequently, x_{1t} and x_{2t} are cointegrated if (a) both of them are unit-root nonstationary, and (b) they have a linear combination that is unit-root stationary.

Generally speaking, for a k-dimensional unit-root nonstationary time series, cointegration exists if there are less than k unit roots in the system. Let h be the number of unit roots in the k-dimensional series $\boldsymbol{x}_t$. Cointegration exists if $0 < h < k$, and the quantity $k - h$ is called the number of cointegrating factors. Alternatively, the number of cointegrating factors is the number of different linear combinations that are unit-root stationary. The linear combinations are called the cointegrating vectors. For the prior simulated example, $y_{2t} = (0.5, 1)\boldsymbol{x}_t$ so that $(0.5, 1)'$ is a cointegrating vector for the system. For more discussions on cointegration and cointegration tests, see Box and Tiao (1977), Engle and Granger (1987), Stock and Watson (1988), and Johansen (1988). We discuss cointegrated VAR models in Section 8.6.

The concept of cointegration is interesting and has attracted a lot of attention in the literature. However, there are difficulties in testing for cointegration in a real application. The main source of difficulties is that cointegration tests overlook the scaling effects of the component series. Interested readers are referred to Cochrane (1988) and Tiao, Tsay, and Wang (1993) for further discussion.

While I have some misgivings on the practical value of cointegration tests, the idea of cointegration is highly relevant in financial study. For example, consider the stock of Finnish Nokia Corporation. Its price on the Helsinki Stock Market must move in unison with the price of its American Depositary Receipts on the New York Stock Exchange; otherwise there exists some arbitrage opportunity for investors. If the stock price has a unit root, then the two price series must be cointegrated. In practice, such a cointegration can exist after adjusting for transaction costs and exchange rate risk. We discuss issues like this later in Section 8.7.

8.5.1 An Error Correction Form

Because there are more unit-root nonstationary components than the number of unit roots in a cointegrated system, differencing individual components to achieve stationarity results in overdifferencing. Overdifferencing leads to the problem of unit roots in the MA matrix polynomial, which in turn may encounter difficulties in parameter estimation. If the MA matrix polynomial contains unit roots, the vector time series is said to be noninvertible.

Engle and Granger (1987) discuss an error correction representation for a cointegrated system that overcomes the difficulty of estimating noninvertible VARMA models. Consider the cointegrated system in Eq. (8.31). Let $\Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$ be the differenced series. Subtracting $\boldsymbol{x}_{t-1}$ from both sides of the equation, we obtain a model for $\Delta \boldsymbol{x}_t$ as

$$\begin{aligned}\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} &= \begin{bmatrix} -0.5 & -1.0 \\ -0.25 & -0.5 \end{bmatrix}\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix}\begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}[0.5, 1.0]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.2 & -0.4 \\ -0.1 & 0.2 \end{bmatrix}\begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.\end{aligned}$$

This is a stationary model because both $\Delta \boldsymbol{x}_t$ and $[0.5, 1.0]\boldsymbol{x}_t = y_{2t}$ are unit-root stationary. Because $\boldsymbol{x}_{t-1}$ is used on the right-hand side of the previous equation, the MA matrix polynomial is the same as before and, hence, the model does not encounter the problem of noninvertibility. Such a formulation is referred to as an error correction model for $\Delta \boldsymbol{x}_t$, and it can be extended to the general cointegrated VARMA model. For a cointegrated VARMA(p, q) model with m cointegrating factors ($m < k$), an error correction representation is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{x}_{t-i} + \boldsymbol{a}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{a}_{t-j}, \tag{8.33}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $k \times m$ full-rank matrices. The AR coefficient matrices $\boldsymbol{\Phi}_i^*$ are functions of the original coefficient matrices $\boldsymbol{\Phi}_j$. Specifically, we have

$$\boldsymbol{\Phi}_j^* = -\sum_{i=j+1}^{p} \boldsymbol{\Phi}_i, \qquad j = 1, \ldots, p-1,$$
$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_{p-1} + \cdots + \boldsymbol{\Phi}_1 - \boldsymbol{I} = -\boldsymbol{\Phi}(1). \tag{8.34}$$

These results can be obtained by equating coefficient matrices of the AR matrix polynomials. The time series $\boldsymbol{\beta}'\boldsymbol{x}_t$ is unit-root stationary, and the columns of $\boldsymbol{\beta}$ are the cointegrating vectors of $\boldsymbol{x}_t$.

Existence of the stationary series $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$ in the error correction representation (8.33) is natural. It can be regarded as a "compensation" term for the overdifferenced system $\Delta\boldsymbol{x}_t$. The stationarity of $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$ can be justified as follows. The theory of unit-root time series shows that the sample correlation coefficient between a unit-root nonstationary series and a stationary series converges to zero as the sample size goes to infinity; see Tsay and Tiao (1990) and the references therein. In an error correction representation, $\boldsymbol{x}_{t-1}$ is unit-root nonstationary, but $\Delta\boldsymbol{x}_t$ is stationary. Therefore, the only way that $\Delta\boldsymbol{x}_t$ can relate meaningfully to $\boldsymbol{x}_{t-1}$ is through a stationary series $\boldsymbol{\beta}'\boldsymbol{x}_{t-1}$.

Remark. Our discussion of cointegration assumes that all unit roots are of multiplicity 1, but the concept can be extended to cases in which the unit roots have different multiplicities. Also, if the number of cointegrating factors m is given, then the error correction model in Eq. (8.33) can be estimated by likelihood methods. We discuss the simple case of cointegrated VAR models in the next section. Finally, there are many ways to construct an error correction representation. In fact, one can use any $\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-v}$ for $1 \le v \le p$ in Eq. (8.33) with some modifications to the AR coefficient matrices $\boldsymbol{\Phi}_i^*$. □

8.6 COINTEGRATED VAR MODELS

To better understand cointegration, we focus on VAR models for their simplicity in estimation. Consider a k-dimensional VAR(p) time series $\boldsymbol{x}_t$ with possible time trend so that the model is

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{x}_{t-p} + \boldsymbol{a}_t, \tag{8.35}$$

where the innovation $\boldsymbol{a}_t$ is assumed to be Gaussian and $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are k-dimensional constant vectors. Write $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$. Recall that if all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ are outside the unit circle, then $\boldsymbol{x}_t$ is unit-root stationary. In the literature, a unit-root stationary series is said to be an $I(0)$ process; that is, it is not integrated. If $|\boldsymbol{\Phi}(1)| = 0$, then $\boldsymbol{x}_t$ is unit-root nonstationary. For simplicity, we assume that $\boldsymbol{x}_t$ is at most an integrated process of order 1; that is, an $I(1)$ process. This means that $(1-B)x_{it}$ is unit-root stationary if x_{it} itself is not.

An error correction model (ECM) for the VAR(p) process $\boldsymbol{x}_t$ is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi}\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^*\Delta\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta\boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t, \tag{8.36}$$

where $\boldsymbol{\Phi}_j^*$ are defined in Eq. (8.34) and $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = -\boldsymbol{\Phi}(1)$. We refer to the term $\boldsymbol{\Pi}\boldsymbol{x}_{t-1}$ of Eq. (8.36) as the *error correction term*, which plays a key role in cointegration study. Notice that $\boldsymbol{\Phi}_i$ can be recovered from the ECM representation via

$$\boldsymbol{\Phi}_1 = \boldsymbol{I} + \boldsymbol{\Pi} + \boldsymbol{\Phi}_1^*,$$
$$\boldsymbol{\Phi}_i = \boldsymbol{\Phi}_i^* - \boldsymbol{\Phi}_{i-1}^*, \qquad i = 2, \ldots, p,$$

where $\boldsymbol{\Phi}_p^* = \boldsymbol{0}$, the zero matrix. Based on the assumption that $\boldsymbol{x}_t$ is at most $I(1)$, $\Delta\boldsymbol{x}_t$ of Eq. (8.36) is an $I(0)$ process.

If $\boldsymbol{x}_t$ contains unit roots, then $|\boldsymbol{\Phi}(1)| = 0$ so that $\boldsymbol{\Pi} = -\boldsymbol{\Phi}(1)$ is singular. Therefore, three cases are of interest in considering the ECM in Eq. (8.36):

1. Rank($\boldsymbol{\Pi}$) $= 0$. This implies $\boldsymbol{\Pi} = \boldsymbol{0}$ and $\boldsymbol{x}_t$ is not cointegrated. The ECM of Eq. (8.36) reduces to

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1^*\Delta\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta\boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

   so that $\Delta\boldsymbol{x}_t$ follows a VAR($p-1$) model with deterministic trend $\boldsymbol{\mu}_t$.
2. Rank($\boldsymbol{\Pi}$) $= k$. This implies that $|\boldsymbol{\Phi}(1)| \neq 0$ and $\boldsymbol{x}_t$ contains no unit roots; that is, $\boldsymbol{x}_t$ is $I(0)$. The ECM model is not informative and one studies $\boldsymbol{x}_t$ directly.
3. $0 <$ Rank($\boldsymbol{\Pi}$) $= m < k$. In this case, one can write $\boldsymbol{\Pi}$ as

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}', \tag{8.37}$$

   where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $k \times m$ matrices with Rank($\boldsymbol{\alpha}$) = Rank($\boldsymbol{\beta}$) $= m$. The ECM of Eq. (8.36) becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^*\Delta\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta\boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t. \tag{8.38}$$

   This means that $\boldsymbol{x}_t$ is cointegrated with m linearly independent cointegrating vectors, $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$, and has $k-m$ unit roots that give $k-m$ common stochastic trends of $\boldsymbol{x}_t$.

If $\boldsymbol{x}_t$ is cointegrated with Rank($\boldsymbol{\Pi}$) $= m$, then a simple way to obtain a presentation of the $k-m$ common trends is to obtain an orthogonal complement matrix $\boldsymbol{\alpha}_\perp$ of $\boldsymbol{\alpha}$; that is, $\boldsymbol{\alpha}_\perp$ is a $k \times (k-m)$ matrix such that $\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha} = \boldsymbol{0}$, a $(k-m) \times m$ zero matrix, and use $\boldsymbol{y}_t = \boldsymbol{\alpha}_\perp'\boldsymbol{x}_t$. To see this, one can premultiply the ECM by $\boldsymbol{\alpha}_\perp'$

and use $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ to see that there would be no error correction term in the resulting equation. Consequently, the $(k-m)$-dimensional series $\boldsymbol{y}_t$ should have $k-m$ unit roots. For illustration, consider the bivariate example of Section 8.5.1. For this special series, $\boldsymbol{\alpha} = (-1, -0.5)'$ and $\boldsymbol{\alpha}_\perp = (1, -2)'$. Therefore, $y_t = (1, -2)\boldsymbol{x}_t = x_{1t} - 2x_{2t}$, which is precisely the unit-root nonstationary series y_{1t} in Eq. (8.32).

Note that the factorization in Eq. (8.37) is not unique because for any $m \times m$ orthogonal matrix $\boldsymbol{\Omega}$ satisfying $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}$, we have a

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{\Omega}\boldsymbol{\Omega}'\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{\Omega})(\boldsymbol{\beta}\boldsymbol{\Omega})' \equiv \boldsymbol{\alpha}_*\boldsymbol{\beta}_*',$$

where both $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}_*$ are also of rank m. Additional constraints are needed to uniquely identify $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. It is common to require that $\boldsymbol{\beta}' = [\boldsymbol{I}_m, \boldsymbol{\beta}_1']$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix and $\boldsymbol{\beta}_1$ is a $(k-m) \times m$ matrix. In practice, this may require reordering of the elements of $\boldsymbol{x}_t$ such that the first m components all have a unit root. The elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ must also satisfy other constraints for the process $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ to be unit-root stationary. For example, consider the case of a bivariate VAR(1) model with one cointegrating vector. Here $k = 2$, $m = 1$, and the ECM is

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_t + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [1, \beta_1]\boldsymbol{x}_{t-1} + \boldsymbol{a}_t.$$

Premultiplying the prior equation by $\boldsymbol{\beta}'$, using $w_{t-i} = \boldsymbol{\beta}'\boldsymbol{x}_{t-i}$, and moving w_{t-1} to the right-hand side of the equation, we obtain

$$w_t = \boldsymbol{\beta}'\boldsymbol{\mu}_t + (1 + \alpha_1 + \alpha_2\beta_1)w_{t-1} + b_t,$$

where $b_t = \boldsymbol{\beta}'\boldsymbol{a}_t$. This implies that w_t is a stationary AR(1) process. Consequently, α_i and β_1 must satisfy the stationarity constraint $|1 + \alpha_1 + \alpha_2\beta_1| < 1$.

The prior discussion shows that the rank of $\boldsymbol{\Pi}$ in the ECM of Eq. (8.36) is the number of cointegrating vectors. Thus, to test for cointegration, one can examine the rank of $\boldsymbol{\Pi}$. This is the approach taken by Johansen (1988, 1995) and Reinsel and Ahn (1992).

8.6.1 Specification of the Deterministic Function

Similar to the univariate case, the limiting distributions of cointegration tests depend on the deterministic function $\boldsymbol{\mu}_t$. In this section, we discuss some specifications of $\boldsymbol{\mu}_t$ that have been proposed in the literature. To understand some of the statements made below, keep in mind that $\boldsymbol{\alpha}_\perp'\boldsymbol{x}_t$ provides a presentation for the common stochastic trends of $\boldsymbol{x}_t$ if it is cointegrated.

1. $\boldsymbol{\mu}_t = \boldsymbol{0}$: In this case, all the component series of $\boldsymbol{x}_t$ are $I(1)$ without drift and the stationary series $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{x}_t$ has mean zero.

2. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{c}_0$, where $\boldsymbol{c}_0$ is an m-dimensional nonzero constant vector. The ECM becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{c}_0) + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

so that the components of $\boldsymbol{x}_t$ are $I(1)$ without drift, but $\boldsymbol{w}_t$ have a nonzero mean $-\boldsymbol{c}_0$. This is referred to as the case of restricted constant.

3. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$, which is nonzero. Here the component series of $\boldsymbol{x}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ and $\boldsymbol{w}_t$ may have a nonzero mean.
4. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{c}_1 t$, where $\boldsymbol{c}_1$ is a nonzero vector. The ECM becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{c}_1 t) + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

so that the components of $\boldsymbol{x}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ and $\boldsymbol{w}_t$ has a linear time trend related to $\boldsymbol{c}_1 t$. This is the case of restricted trend.

5. $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, where $\boldsymbol{\mu}_i$ are nonzero. Here both the constant and trend are unrestricted. The components of $\boldsymbol{x}_t$ are $I(1)$ and have a quadratic time trend and $\boldsymbol{w}_t$ have a linear trend.

Obviously, the last case is not common in empirical work. The first case is not common for economic time series but may represent the log price series of some assets. The third case is also useful in modeling asset prices.

8.6.2 Maximum-Likelihood Estimation

In this section, we briefly outline the maximum-likelihood estimation (MLE) of a cointegrated VAR(p) model. Suppose that the data are $\{\boldsymbol{x}_t | t = 1, \ldots, T\}$. Without loss of generality, we write $\boldsymbol{\mu}_t = \boldsymbol{\mu}\boldsymbol{d}_t$, where $\boldsymbol{d}_t = [1, t]'$, and it is understood that $\boldsymbol{\mu}_t$ depends on the specification of the previous section. For a given m, which is the rank of $\boldsymbol{\Pi}$, the ECM model becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t, \qquad (8.39)$$

where $t = p + 1, \ldots, T$. A key step in the estimation is to concentrate the likelihood function with respect to the deterministic term and the stationary effects. This is done by considering the following two multivariate linear regressions:

$$\Delta \boldsymbol{x}_t = \boldsymbol{\gamma}_0 \boldsymbol{d}_t + \boldsymbol{\Omega}_1 \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Omega}_{p-1} \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{u}_t, \qquad (8.40)$$

$$\boldsymbol{x}_{t-1} = \boldsymbol{\gamma}_1 \boldsymbol{d}_t + \boldsymbol{\Xi}_1 \Delta \boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Xi}_{p-1} \Delta \boldsymbol{x}_{t-p+1} + \boldsymbol{v}_t. \qquad (8.41)$$

Let $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$ be the residuals of Eqs. (8.40) and (8.41), respectively. Define the sample covariance matrices

$$\boldsymbol{S}_{00} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_t', \quad \boldsymbol{S}_{01} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{v}}_t', \quad \boldsymbol{S}_{11} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\boldsymbol{v}}_t \hat{\boldsymbol{v}}_t'.$$

Next, compute the eigenvalues and eigenvectors of $\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ with respect to $\boldsymbol{S}_{11}$. This amounts to solving the eigenvalue problem

$$|\lambda \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}| = 0.$$

Denote the eigenvalue and eigenvector pairs by $(\hat{\lambda}_i, \boldsymbol{e}_i)$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_k$. Here the eigenvectors are normalized so that $\boldsymbol{e}'\boldsymbol{S}_{11}\boldsymbol{e} = \boldsymbol{I}$, where $\boldsymbol{e} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k]$ is the matrix of eigenvectors.

The unnormalized MLE of the cointegrating vector $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$ and from which we can obtain an MLE for $\boldsymbol{\beta}$ that satisfies the identifying constraint and normalization condition. Denote the resulting estimate by $\hat{\boldsymbol{\beta}}_c$ with the subscript c signifying constraints. The MLE of other parameters can then be obtained by the multivariate linear regression

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\alpha}\hat{\boldsymbol{\beta}}_c'\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^*\Delta\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta\boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t.$$

The maximized value of the likelihood function based on m cointegrating vectors is

$$L_{\max}^{-2/T} \propto |\boldsymbol{S}_{00}| \prod_{i=1}^{m}(1 - \hat{\lambda}_i).$$

This value is used in the maximum-likelihood ratio test for testing Rank$(\boldsymbol{\Pi}) = m$. Finally, estimates of the orthogonal complements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be obtained using

$$\hat{\boldsymbol{\alpha}}_\perp = \boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{11}[\boldsymbol{e}_{m+1}, \ldots, \boldsymbol{e}_k], \qquad \hat{\boldsymbol{\beta}}_\perp = \boldsymbol{S}_{11}[\boldsymbol{e}_{m+1}, \ldots, \boldsymbol{e}_k].$$

8.6.3 Cointegration Test

For a specified deterministic term $\boldsymbol{\mu}_t$, we now discuss the maximum-likelihood test for testing the rank of the $\boldsymbol{\Pi}$ matrix in Eq. (8.36). Let $H(m)$ be the null hypothesis that the rank of $\boldsymbol{\Pi}$ is m. For example, under $H(0)$, Rank$(\boldsymbol{\Pi}) = 0$ so that $\boldsymbol{\Pi} = \boldsymbol{0}$ and there is no cointegration. The hypotheses of interest are

$$H(0) \subset \cdots \subset H(m) \subset \cdots \subset H(k).$$

For testing purpose, the ECM in Eq. (8.39) becomes

$$\Delta \boldsymbol{x}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\Pi}\boldsymbol{x}_{t-1} + \boldsymbol{\Phi}_1^*\,\Delta\boldsymbol{x}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\,\Delta\boldsymbol{x}_{t-p+1} + \boldsymbol{a}_t,$$

where $t = p+1, \ldots, T$. Our goal is to test the rank of $\boldsymbol{\Pi}$. Mathematically, the rank of $\boldsymbol{\Pi}$ is the number of nonzero eigenvalues of $\boldsymbol{\Pi}$, which can be obtained if a consistent estimate of $\boldsymbol{\Pi}$ is available. Based on the prior equation, which is in the form of a multivariate linear regression, we see that $\boldsymbol{\Pi}$ is related to the covariance matrix between $\boldsymbol{x}_{t-1}$ and $\Delta\boldsymbol{x}_t$ after adjusting for the effects of $\boldsymbol{d}_t$ and $\Delta\boldsymbol{x}_{t-i}$ for i

$= 1, \ldots, p-1$. The necessary adjustments can be achieved by the techniques of multivariate linear regression shown in the previous section. Indeed, the adjusted series of $\boldsymbol{x}_{t-1}$ and $\Delta\boldsymbol{x}_t$ are $\hat{\boldsymbol{v}}_t$ and $\hat{\boldsymbol{u}}_t$, respectively. The equation of interest for the cointegration test then becomes

$$\hat{\boldsymbol{u}}_t = \boldsymbol{\Pi}\hat{\boldsymbol{v}}_t + \boldsymbol{a}_t.$$

Under the normality assumption, the likelihood ratio test for testing the rank of $\boldsymbol{\Pi}$ in the prior equation can be done by using the canonical correlation analysis between $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$. See Johnson and Wichern (1998) for information on canonical correlation analysis. The associated canonical correlations are the *partial canonical correlations* between $\Delta\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_{t-1}$ because the effects of $\boldsymbol{d}_t$ and $\Delta\boldsymbol{x}_{t-i}$ have been adjusted. The quantities $\{\hat{\lambda}_i\}$ are the squared canonical correlations between $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$.

Consider the hypotheses

$$H_0 : \text{Rank}(\boldsymbol{\Pi}) = m \quad \text{versus} \quad H_a : \text{Rank}(\boldsymbol{\Pi}) > m.$$

Johansen (1988) proposes the likelihood ratio (LR) statistic

$$\text{LR}_{\text{tr}}(m) = -(T-p)\sum_{i=m+1}^{k} \ln(1-\hat{\lambda}_i) \tag{8.42}$$

to perform the test. If $\text{Rank}(\boldsymbol{\Pi}) = m$, then $\hat{\lambda}_i$ should be small for $i > m$ and hence $\text{LR}_{\text{tr}}(m)$ should be small. This test is referred to as the *trace* cointegration test. Due the presence of unit roots, the asymptotic distribution of $\text{LR}_{\text{tr}}(m)$ is not chi squared but a function of standard Brownian motions. Thus, critical values of $\text{LR}_{\text{tr}}(m)$ must be obtained via simulation.

Johansen (1988) also considers a sequential procedure to determine the number of cointegrating vectors. Specifically, the hypotheses of interest are

$$H_0 : \text{Rank}(\boldsymbol{\Pi}) = m \quad \text{versus} \quad H_a : \text{Rank}(\boldsymbol{\Pi}) = m+1.$$

The LR ratio test statistic, called the maximum eigenvalue statistic, is

$$\text{LR}_{\max}(m) = -(T-p)\ln(1-\hat{\lambda}_{m+1}).$$

Again, critical values of the test statistics are nonstandard and must be evaluated via simulation.

8.6.4 Forecasting of Cointegrated VAR Models

The fitted ECM can be used to produce forecasts. First, conditioned on the estimated parameters, the ECM equation can be used to produce forecasts of the differenced series $\Delta\boldsymbol{x}_t$. Such forecasts can in turn be used to obtain forecasts of $\boldsymbol{x}_t$. A difference between ECM forecasts and the traditional VAR forecasts is that the ECM approach imposes the cointegration relationships in producing the forecasts.

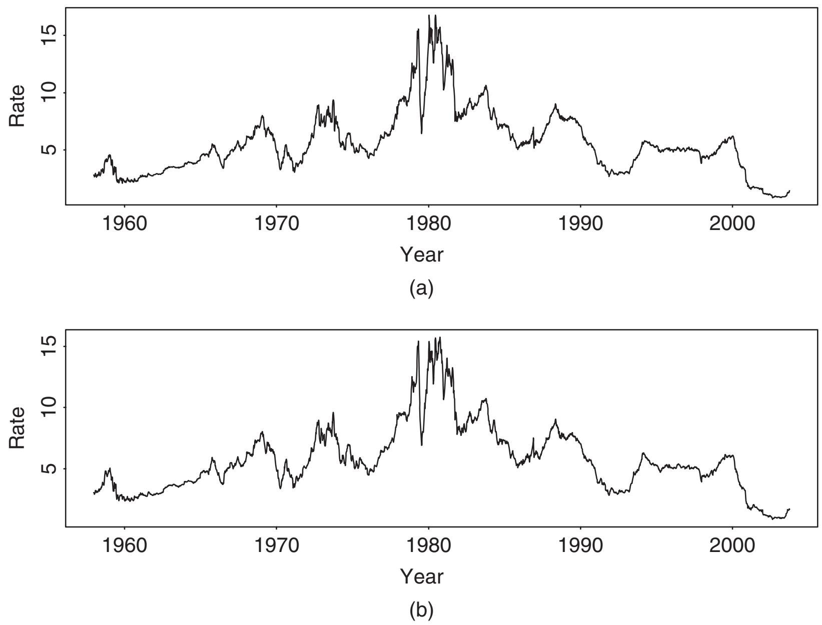


Figure 8.12 Time plots of weekly U.S. interest rate from December 12, 1958, to August 6, 2004. (a) The 3-month Treasury bill rate and (b) 6-month Treasury bill rate. Rates are from secondary market.

8.6.5 An Example

To demonstrate the analysis of cointegrated VAR models, we consider two weekly U.S. short-term interest rates. The series are the 3-month Treasury bill (TB) rate and 6-month Treasury bill rate from December 12, 1958, to August 6, 2004, for 2383 observations. The TB rates are from the secondary market and obtained from the Federal Reserve Bank of St. Loius. Figure 8.12 shows the time plots of the interest rates. As expected, the two series move closely together.

Our analysis uses the S-Plus software with commands VAR for VAR analysis, coint for cointegration test, and VECM for vector error correction estimation. Denote the two series by tb3m and tb6m and define the vector series $\boldsymbol{x}_t = (\text{tb3m}_t, \text{tb6m}_t)'$. The augmented Dickey–Fuller unit-root tests fail to reject the hypothesis of a unit root in the individual series; see Chapter 2. Indeed, the test statistics are -2.34 and -2.33 with p value about 0.16 for the 3-month and 6-month interest rate when an AR(3) model is used. Thus, we proceed to VAR modeling.

For the bivariate series $\boldsymbol{x}_t$, the BIC criterion selects a VAR(3) model:

```
> x=cbind(tb3m,tb6m)
> y=data.frame(x)
> ord.choice$ar.order
[1] 3
```

To perform a cointegration test, we choose a restricted constant for $\boldsymbol{\mu}_t$ because there is no reason a priori to believe the existence of a drift in the U.S. interest rate. Both Johansen's tests confirm that the two series are cointegrated with one cointegrating vector when a VAR(3) model is entertained.

```
> cointst.rc=coint(x,trend='rc', lags=2)  % lags = p-1.
> cointst.rc
Call:
coint(Y = x, lags = 2, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests sign. at the 5% level are flagged by ' +'.
Trace tests sign. at the 1% level are flagged by '++'.
Max Eig. tests sign. at the 5% level are flagged by ' *'.
Max Eig. tests sign. at the 1% level are flagged by '**'.

Tests for Cointegration Rank:
         Eigenvalue Trace Stat  95% CV  99% CV
H(0)++**  0.0322    83.2712     19.96   24.60
H(1)      0.0023     5.4936      9.24   12.97

         Max Stat  95% CV  99% CV
H(0)++**  77.7776  15.67   20.20
H(1)       5.4936   9.24   12.97
```

Next, we perform the maximum-likelihood estimation of the specified cointegrated VAR(3) model using an ECM presentation. The results are as follows:

```
> vecm.fit=VECM(cointst.rc)
> summary(vecm.fit)
Call:
VECM(test = cointst.rc)

Cointegrating Vectors:
             coint.1
              1.0000

      tb6m   -1.0124
 (std.err)    0.0086
  (t.stat) -118.2799

Intercept*    0.2254
 (std.err)    0.0545
  (t.stat)    4.1382

VECM Coefficients:
```

```
             tb3m    tb6m
  coint.1 -0.0949 -0.0211
(std.err)  0.0199  0.0179
 (t.stat) -4.7590 -1.1775

tb3m.lag1  0.0466 -0.0419
(std.err)  0.0480  0.0432
(t.stat)   0.9696 -0.9699

tb6m.lag1  0.2650  0.3164
(std.err)  0.0538  0.0484
 (t.stat)  4.9263  6.5385

tb3m.lag2 -0.2067 -0.0346
(std.err)  0.0481  0.0433
 (t.stat) -4.2984 -0.8005

tb6m.lag2  0.2547  0.0994
(std.err)  0.0543  0.0488
 (t.stat)  4.6936  2.0356

Regression Diagnostics:
                 tb3m   tb6m
     R-squared 0.1081 0.0913
Adj. R-squared 0.1066 0.0898
  Resid. Scale 0.2009 0.1807

> plot(vecm.fit)
Make a plot selection (or 0 to exit):

1: plot: All
2: plot: Response and Fitted Values
3: plot: Residuals
...
13: plot: PACF of Squared Cointegrating Residuals
Selection:
```

As expected, the output shows that the stationary series is $w_t \approx \text{tb3m}_t - \text{tb6m}_t$ and the mean of w_t is about -0.225. The fitted ECM is

$$\Delta \boldsymbol{x}_t = \begin{bmatrix} -0.09 \\ -0.02 \end{bmatrix} (w_{t-1} + 0.23) + \begin{bmatrix} 0.05 & 0.27 \\ -0.04 & 0.32 \end{bmatrix} \Delta \boldsymbol{x}_{t-1}$$
$$+ \begin{bmatrix} -0.21 & 0.25 \\ -0.03 & 0.10 \end{bmatrix} \Delta \boldsymbol{x}_{t-2} + \boldsymbol{a}_t,$$

and the estimated standard errors of a_{it} are 0.20 and 0.18, respectively. Adequacy of the fitted ECM can be examined via various plots. For illustration, Figure 8.13 shows the cointegrating residuals. Some large residuals are shown

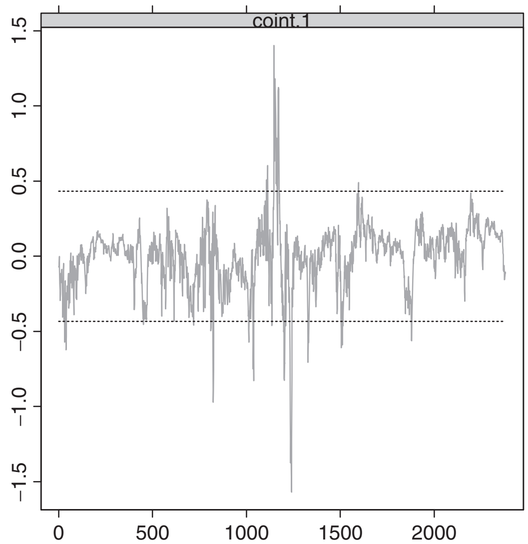


Figure 8.13 Time plot of cointegrating residuals for an ECM fit to weekly U.S. interest rate series. Data span is from December 12, 1958, to August 6, 2004.

in the plot, which occurred in the early 1980s when the interest rates were high and volatile.

Finally, we use the fitted ECM to produce 1-step- to 10-step-ahead forecasts for both $\Delta \boldsymbol{x}_t$ and $\boldsymbol{x}_t$. The forecast origin is August 6, 2004.

```
> vecm.fst=predict(vecm.fit, n.predict=10)
> summary(vecm.fst)
Predicted Values with Standard Errors:

                 tb3m     tb6m
 1-step-ahead -0.0378  -0.0642
     (std.err)  0.2009   0.1807
 2-step-ahead -0.0870  -0.0864
     (std.err)  0.3222   0.2927
 ...
10-step-ahead -0.2276 -0.1314
      (std.err)  0.8460  0.8157
> plot(vecm.fst,xold=diff(x),n.old=12)

> vecm.fit.level=VECM(cointst.rc,levels=T)
> vecm.fst.level=predict(vecm.fit.level, n.predict=10)
> summary(vecm.fst.level)

Predicted Values with Standard Errors:
               tb3m    tb6m
 1-step-ahead 1.4501  1.7057
```

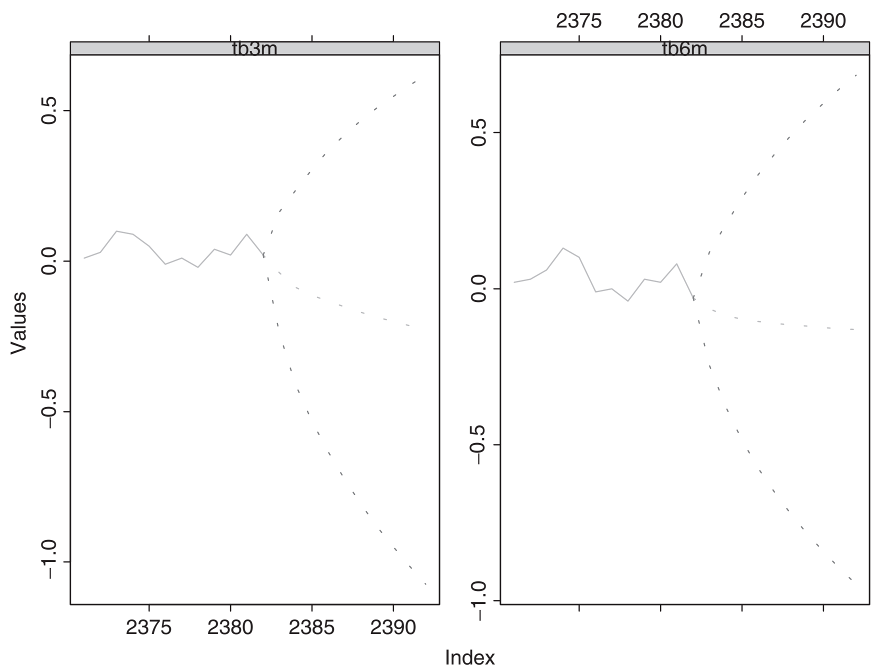


Figure 8.14 Forecasting plots of fitted ECM model for weekly U.S. interest rate series. Forecasts are for differenced series and forecast origin is August 6, 2004.

```
    (std.err) 0.2009  0.1807
 2-step-ahead 1.4420  1.7017
    (std.err) 0.3222  0.2927
  ...
10-step-ahead 1.4722 1.7078
     (std.err) 0.8460 0.8157
> plot(vecm.fst.level, xold=x, n.old=50)
```

The forecasts are shown in Figures 8.14 and 8.15 for the differenced data and the original series, respectively, along with some observed data points. The dashed lines in the plots are pointwise 95% confidence intervals. Because of unit-root nonstationarity, the intervals are wide and not informative.

Remark. The package `urca` of R can be used to perform Johansen's co-integration test. The command is `ca.jo`. It requires specification of some sub-commands. See the section of pairs trading for demonstration. □

8.7 THRESHOLD COINTEGRATION AND ARBITRAGE

In this section, we focus on detecting arbitrage opportunities in index trading by using multivariate time series methods. We also demonstrate that simple univariate

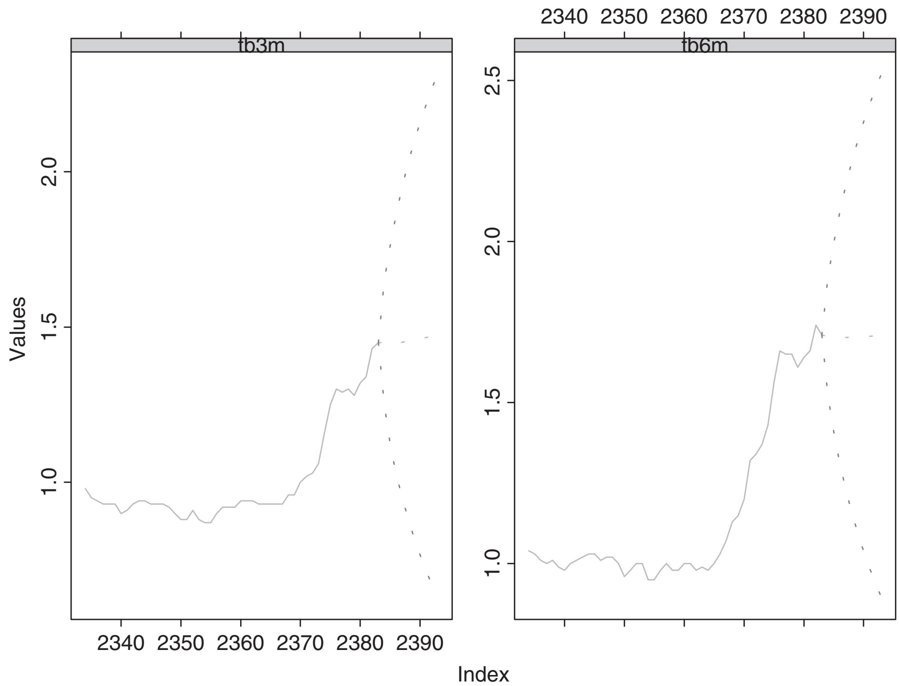


Figure 8.15 Forecasting plots of fitted ECM model for weekly U.S. interest rate series. Forecasts are for interest rates and forecast origin is August 6, 2004.

nonlinear models of Chapter 4. can be extended naturally to the multivariate case in conjunction with the idea of cointegration.

Our study considers the relationship between the price of the S&P 500 index futures and the price of the shares underlying the index on the cash market. Let $f_{t,\ell}$ be the log price of the index futures at time t with maturity ℓ, and let s_t be the log price of the shares underlying the index on the cash market at time t. A version of the *cost-of-carry model* in the finance literature states

$$f_{t,\ell} - s_t = (r_{t,\ell} - q_{t,\ell})(\ell - t) + z_t^*, \tag{8.43}$$

where $r_{t,\ell}$ is the risk-free interest rate, $q_{t,\ell}$ is the dividend yield with respect to the cash price at time t, and $(\ell - t)$ is the time to maturity of the futures contract; see Brenner and Kroner (1995), Dwyer, Locke, and Yu (1996), and the references therein.

The z_t^* process of model (8.43) must be unit-root stationary; otherwise there exist *persistent* arbitrage opportunities. Here an arbitrage trading consists of simultaneously buying (short-selling) the security index and selling (buying) the index futures whenever the log prices diverge by more than the cost of carrying the index over time until maturity of the futures contract. Under the weak stationarity of z_t^*, for arbitrage to be profitable, z_t^* must exceed a certain value in modulus determined by transaction costs and other economic and risk factors.

It is commonly believed that the $f_{t,\ell}$ and s_t series of the S&P 500 index contain a unit root, but Eq. (8.43) indicates that they are cointegrated after adjusting for the effect of interest rate and dividend yield. The cointegrating vector is $(1, -1)$ after the adjustment, and the cointegrated series is z_t^*. Therefore, one should use an error correction form to model the return series $\boldsymbol{r}_t = (\Delta f_t, \Delta s_t)'$, where $\Delta f_t = f_{t,\ell} - f_{t-1,\ell}$ and $\Delta s_t = s_t - s_{t-1}$, where for ease in notation we drop the maturity time ℓ from the subscript of Δf_t.

8.7.1 Multivariate Threshold Model

In practice, arbitrage tradings affect the dynamic of the market, and hence the model for $\boldsymbol{r}_t$ may vary over time depending on the presence or absence of arbitrage tradings. Consequently, the prior discussions lead naturally to the following model:

$$\boldsymbol{r}_t = \begin{cases} \boldsymbol{c}_1 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(1)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_1 z_{t-1} + \boldsymbol{a}_t^{(1)} & \text{if } z_{t-1} \le \gamma_1, \\ \boldsymbol{c}_2 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(2)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_2 z_{t-1} + \boldsymbol{a}_t^{(2)} & \text{if } \gamma_1 < z_{t-1} \le \gamma_2, \\ \boldsymbol{c}_3 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(3)} \boldsymbol{r}_{t-i} + \boldsymbol{\beta}_3 z_{t-1} + \boldsymbol{a}_t^{(3)} & \text{if } \gamma_2 < z_{t-1}, \end{cases} \tag{8.44}$$

where $z_t = 100z_t^*$, $\gamma_1 < 0 < \gamma_2$ are two real numbers, and $\{\boldsymbol{a}_t^{(i)}\}$ are sequences of two-dimensional white noises and are independent of each other. Here we use $z_t = 100z_t^*$ because the actual value of z_t^* is relatively small.

The model in Eq. (8.44) is referred to as a multivariate threshold model with three regimes. The two real numbers γ_1 and γ_2 are the thresholds and z_{t-1} is the threshold variable. The threshold variable z_{t-1} is supported by the data; see Tsay (1998). In general, one can select z_{t-d} as a threshold variable by considering $d \in \{1, \ldots, d_0\}$, where d_0 is a prespecified positive integer.

Model (8.44) is a generalization of the threshold autoregressive model of Chapter 4. It is also a generalization of the error correlation model of Eq. (8.33). As mentioned earlier, an arbitrage trading is profitable only when z_t^* or, equivalently, z_t is large in modulus. Therefore, arbitrage tradings only occurred in regimes 1 and 3 of model (8.44). As such, the dynamic relationship between $f_{t,\ell}$ and s_t in regime 2 is determined mainly by the normal market force, and hence the two series behave more or less like a random walk. In other words, the two log prices in the middle regime should be free from arbitrage effects and, hence, free from the cointegration constraint. From an econometric viewpoint, this means that the estimate of $\boldsymbol{\beta}_2$ in the middle regime should be insignificant.

In summary, we expect that the cointegration effects between the log price of the futures and the log price of security index on the cash market are significant in regimes 1 and 3, but insignificant in regime 2. This phenomenon is referred to as a *threshold cointegration*; see Balke and Fomby (1997).

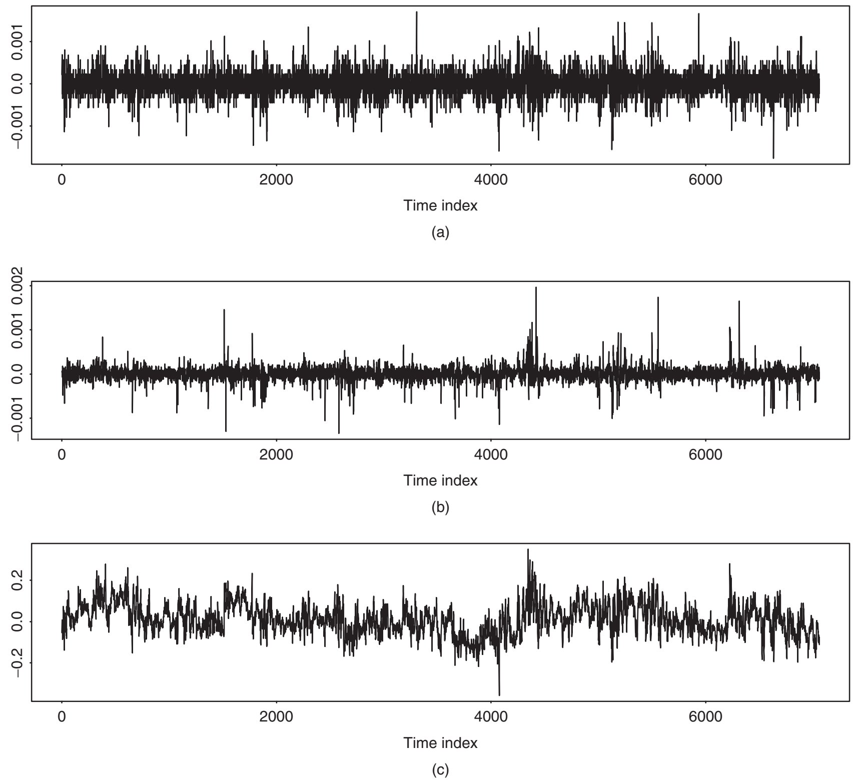


Figure 8.16 Time plots of 1-minute log returns of S&P 500 index futures and cash prices and associated threshold variable in May 1993: (a) log returns of index futures, (b) log returns of index cash prices, and (c) z_t series.

8.7.2 The Data

The data used in this case study are the intraday transaction data of the S&P 500 index in May 1993 and its June futures contract traded at the Chicago Mercantile Exchange; see Forbes, Kalb, and Kofman (1999), who used the data to construct a minute-by-minute bivariate price series with 7060 observations. To avoid the undue influence of unusual returns, I replaced 10 extreme values (5 on each side) by the simple average of their two nearest neighbors. This step does not affect the qualitative conclusion of the analysis but may affect the conditional heteroscedasticity in the data. For simplicity, we do not consider conditional heteroscedasticity in the study. Figure 8.16 shows the time plots of the log returns of the index futures and cash prices and the associated threshold variable $z_t = 100z_t^*$ of model (8.43).

8.7.3 Estimation

A formal specification of the multivariate threshold model in Eq. (8.44) includes selecting the threshold variable, determining the number of regimes, and choosing the order p for each regime. Interested readers are referred to Tsay (1998) and Forbes, Kalb, and Kofman (1999). The thresholds γ_1 and γ_2 can be estimated by using some information criteria [e.g., the Akaike information criterion (AIC) or the sum of squares of residuals]. Assuming $p = 8$, $d \in \{1, 2, 3, 4\}$, $\gamma_1 \in [-0.15, -0.02]$, and $\gamma_2 \in [0.025, 0.145]$, and using a grid search method with 300 points on each of the two intervals, the AIC selects z_{t-1} as the threshold variable with thresholds $\hat{\gamma}_1 = -0.0226$ and $\hat{\gamma}_2 = 0.0377$. Details of the parameter estimates are given in Table 8.8.

From Table 8.8, we make the following observations. First, the t ratios of $\widehat{\boldsymbol{\beta}}_2$ in the middle regime show that, as expected, the estimates are insignificant at the 5% level, confirming that there is no cointegration between the two log prices in the absence of arbitrage opportunities. Second, Δf_t depends negatively on Δf_{t-1} in all three regimes. This is in agreement with the bid–ask bounce discussed in Chapter 5. Third, past log returns of the index futures seem to be more informative than the past log returns of the cash prices because there are more significant t ratios in Δf_{t-i} than in Δs_{t-i}. This is reasonable because futures series are in general more liquid. For more information on index arbitrage, see Dwyer, Locke, and Yu (1996).

8.8 PAIRS TRADING

Pairs trading is a market-neutral trading strategy. There are several versions of pairs trading in the equity markets. In this section, we focus on the statistical arbitrage pairs trading, which makes use of the ideas of cointegration and error correction model discussed in the chapter. Our discussion will be brief. For more information concerning pairs trading and statistical arbitrage, see Vidyamurthy (2004) and Pole (2007).

The general theme for trading in the equity markets is to buy undervalued stocks and sell overvalued ones. However, the true price of a stock is hard to assess. Pairs trading attempts to resolve this difficulty using the idea of relative pricing. Based on the arbitrage pricing theory (APT) in finance, if two stocks have similar characteristics, then the prices of both stocks must be more or less the same. If the prices differ, then it is likely that one of the stocks is overpriced and the other underpriced. Pairs trading involves selling the higher priced stock and buying the lower priced stock with the hope that the mispricing will correct itself in the future. Note that the true prices of the two stocks are not important. The observed prices may be wrong. What is important is that the observed prices be the same. The gap (properly scaled) between the two observed prices is called the spread. For pairs trading, the greater the spread, the larger the magnitude of mispricing and the greater the profit potential. Before discussing a trading strategy, we first introduce the theoretical framework.

8.8.1 Theoretical Framework

Consider two stocks. Let P_{it} be the observed price of stock i at time t and $p_{it} = \ln(P_{it})$ be the corresponding log price. As mentioned in earlier chapters, it is

TABLE 8.8 Least-Squares Estimates and Their t Ratios of Multivariate Threshold Model in Eq. (8.43) for S&P 500 Index Data in May 1993[a]

	Regime 1		Regime 2		Regime 3	
	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t
ϕ_0	0.00002	0.00005	0.00000	0.00000	−0.00001	−0.00005
t	(1.47)	(7.64)	(−0.07)	(0.53)	(−0.74)	(−6.37)
Δf_{t-1}	−0.08468	0.07098	−0.03861	0.04037	−0.04102	0.02305
t	(−3.83)	(6.15)	(−1.53)	(3.98)	(−1.72)	(1.96)
Δf_{t-2}	−0.00450	0.15899	0.04478	0.08621	−0.02069	0.09898
t	(−0.20)	(13.36)	(1.85)	(8.88)	(−0.87)	(8.45)
Δf_{t-3}	0.02274	0.11911	0.07251	0.09752	0.00365	0.08455
t	(0.95)	(9.53)	(3.08)	(10.32)	(0.15)	(7.02)
Δf_{t-4}	0.02429	0.08141	0.01418	0.06827	−0.02759	0.07699
t	(0.99)	(6.35)	(0.60)	(7.24)	(−1.13)	(6.37)
Δf_{t-5}	0.00340	0.08936	0.01185	0.04831	−0.00638	0.05004
t	(0.14)	(7.10)	(0.51)	(5.13)	(−0.26)	(4.07)
Δf_{t-6}	0.00098	0.07291	0.01251	0.03580	−0.03941	0.02615
t	(0.04)	(5.64)	(0.54)	(3.84)	(−1.62)	(2.18)
Δf_{t-7}	−0.00372	0.05201	0.02989	0.04837	−0.02031	0.02293
t	(−0.15)	(4.01)	(1.34)	(5.42)	(−0.85)	(1.95)
Δf_{t-8}	0.00043	0.00954	0.01812	0.02196	−0.04422	0.00462
t	(0.02)	(0.76)	(0.85)	(2.57)	(−1.90)	(0.40)
Δs_{t-1}	−0.08419	0.00264	−0.07618	−0.05633	0.06664	0.11143
t	(−2.01)	(0.12)	(−1.70)	(−3.14)	(1.49)	(5.05)
Δs_{t-2}	−0.05103	0.00256	−0.10920	−0.01521	0.04099	−0.01179
t	(−1.18)	(0.11)	(−2.59)	(−0.90)	(0.92)	(−0.53)
Δs_{t-3}	0.07275	−0.03631	−0.00504	0.01174	−0.01948	−0.01829
t	(1.65)	(−1.58)	(−0.12)	(0.71)	(−0.44)	(−0.84)
Δs_{t-4}	0.04706	0.01438	0.02751	0.01490	0.01646	0.00367
t	(1.03)	(0.60)	(0.71)	(0.96)	(0.37)	(0.17)
Δs_{t-5}	0.08118	0.02111	0.03943	0.02330	−0.03430	−0.00462
t	(1.77)	(0.88)	(0.97)	(1.43)	(−0.83)	(−0.23)
Δs_{t-6}	0.04390	0.04569	0.01690	0.01919	0.06084	−0.00392
t	(0.96)	(1.92)	(0.44)	(1.25)	(1.45)	(−0.19)
Δs_{t-7}	−0.03033	0.02051	−0.08647	0.00270	−0.00491	0.03597
t	(−0.70)	(0.91)	(−2.09)	(0.16)	(−0.13)	(1.90)
Δs_{t-8}	−0.02920	0.03018	0.01887	−0.00213	0.00030	0.02171
t	(−0.68)	(1.34)	(0.49)	(−0.14)	(0.01)	(1.14)
z_{t-1}	0.00024	0.00097	−0.00010	0.00012	0.00025	0.00086
t	(1.34)	(10.47)	(−0.30)	(0.86)	(1.41)	(9.75)

[a]The numbers of data points for the three regimes are 2234, 2410, and 2408, respectively.

reasonable to assume that p_{it} is unit-root nonstationary and follows a random-walk model; that is, $p_{it} = p_{i,t-1} + r_{it}$, where $\{r_{it}\}$ is the return and forms a sequence of uncorrelated innovations. If the two stocks have similar risk factors, then they should have similar returns based on APT. Therefore, p_{1t} and p_{2t} are likely to be driven by a common component and are cointegrated. In other words, there exists a linear combination $w_t = p_{1t} - \gamma p_{2t}$, which is unit-root stationary and, hence, mean reverting. The two price series $\{p_{1t}\}$ and $\{p_{2t}\}$ thus assume an error correction form

$$\begin{bmatrix} p_{1t} - p_{1,t-1} \\ p_{2t} - p_{2,t-1} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} (w_{t-1} - \mu_w) + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \tag{8.45}$$

where $\mu_w = E(w_t)$ denotes the mean of w_t. The four parameters γ, μ_w, α_1, and α_2 can be estimated, for instance, by the maximum-likelihood or least-squares methods; see Section 8.6.2. We refer to the stationary series w_t as the *spread* between the two log stock prices.

The left-hand side of Eq. (8.45) consists the log returns of the two stocks. The equation says that the returns depend on w_{t-1}, which is the stationary. Specifically, $w_{t-1} - \mu_w$ denotes the deviation from the log-run equilibrium between the two stocks. Equation (8.45) shows that, for cointegrated stocks, the returns depend on the past deviation from equilibrium. The coefficients α_1 and α_2 show the effect of past deviation on the returns r_{1t} and r_{2t}, respectively. In practice, α_1 and α_2 should have opposite signs, indicating reversion to the equilibrium.

Next, consider a portfolio with long one share of stock 1 and short γ shares of stock 2. The return of the portfolio for a given time period i is

$$\begin{aligned} r_{p,t+i} &= (p_{1,t+i} - p_{1t}) - \gamma(p_{2,t+i} - p_{2t}) \\ &= (p_{1,t+i} - \gamma p_{2,t+i}) - (p_{1t} - \gamma p_{2t}) \\ &= w_{t+i} - w_t. \end{aligned}$$

Therefore, the return $r_{p,t+i}$ of the portfolio is the increment of the spread in the time period. As expected, the return of the portfolio does not depend on the mean of w_t.

8.8.2 Trading Strategy

The idea behind a pairs-trading strategy is to trade on the oscillations about the equilibrium value of the spread. The oscillations in spread occur because the spread is mean reverting. Since the equilibrium value is the mean of w_t, that is, μ_w, we can put on a trade when w_t deviates substantially from its mean and unwind the trade when the equilibrium is restored. In practice, how big the deviation needs to be in order for the trading to be profitable depends on several factors. Trading costs, marginal interest rates, and bid–ask spreads of the two stocks are three obvious factors. Mathematically, let η be the cost involved in carrying out a pairs trading. Let Δ be a target deviation of w_t from its mean μ_w for pairs trading. Then, conditioned on $2\Delta > \eta$, a simple trading strategy is as follows:

- Buy a share of stock 1 and short γ shares of stock 2 at time t if $w_t = p_{1t} - \gamma p_{2t} = \mu_w - \Delta$.
- Unwind the position at time $t+i$ $(i > 0)$ if $w_{t+i} = p_{1,i+i} - \gamma p_{2,t+i} = \mu_w + \Delta$.

One can identify the time point t so long as Δ is not too large compared with the standard deviation of w_t. The time point $t+i$ will occur because of the mean reverting of the spread series. In this particular instance, the return of the portfolio $w_{t+i} - w_t = 2\Delta$ and the net profit of the trade is $2\Delta - \eta > 0$.

Discussion. The aforementioned trading strategy is just one of many possibilities. For instance, if $\Delta > \eta$, one can unwind the position when $w_{t+i} = \mu_w$. The net profit of the pairs trading then is $\Delta - \eta > 0$. This may result in more transactions and trading costs, but it shortens the holding period of the portfolio. If Δ is negative, then one can short one share of stock 1 and buy γ shares of stock 2 to make a net profit $-2\Delta - \eta$. The quantity η is the threshold for trading and is likely to depend on several factors such as transaction fees and bid–ask spreads of the two stocks. □

8.8.3 Simple Illustration

To demonstrate pairs trading, we consider two stocks traded on the New York Stock Exchange. The two companies are the Billiton Ltd. of Australia and the Vale S.A. of Brazil with stock symbols BHP and VALE, respectively. BHP of Australia is a natural resources company with business in Australia, the Americans, and Southern Africa. Vale of Brazil is a worldwide metals and mining company. Thus, both multinational companies belong to the natural resources industry and encounter similar risk factors. The daily prices of the two stocks were downloaded from Yahoo Finance, and we employ adjusted closing prices from July 1, 2002, to March 31, 2006, in our study.

Figure 8.17 shows the time plots of the daily log prices of the two stocks (adjusted closing prices). The upper plot is for the BHP stock. From the plots, the prices of the two stocks exhibit certain characteristics of comovement. Let p_{1t} and p_{2t} be the daily log closing prices of BHP and VALE, respectively. We analyze the series using both the least-squares and maximum-likelihood methods.

Least-Squares Estimation

A simple way to verify that the two stocks are suitable for pairs trading is to check the cointegration of their log stock prices. To this end, we consider the simple linear regression $p_{1t} = \beta_0 + \beta_1 p_{2t} + w_t$, where w_t denotes the residual series. For the BHP and VALE stocks, we have

$$p_{1t} = 1.823 + 0.717 p_{2t} + \hat{w}_t, \qquad \sigma_w = 0.044.$$

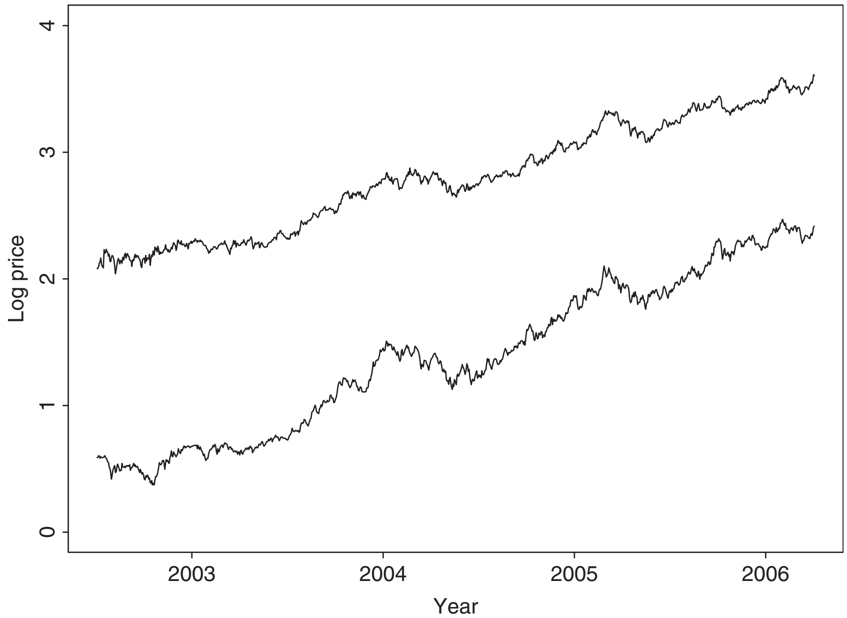


Figure 8.17 Daily log (adjusted) closing prices of BHP and VALE stocks from July 1, 2002, to March 31, 2006. Upper plot is for BHP stock.

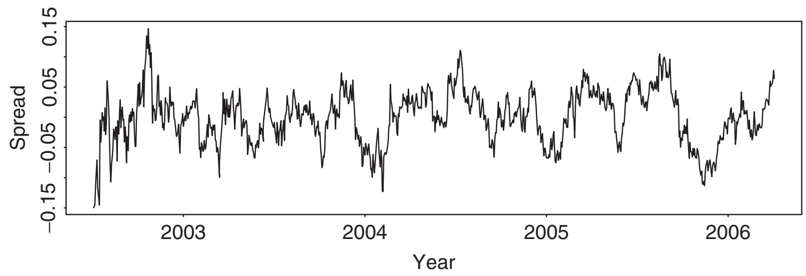
Figure 8.18(a) shows the time plot of the residual series $\hat{w}_t$. The plot shows that the residual series has certain characteristics of a stationary time series. In particular, it has mean zero and fluctuates around its mean within a fixed range. Figure 8.18(b) gives the sample ACF of $\hat{w}_t$. The ACFs decay exponentially, supporting that $\hat{w}_t$ is indeed stationary. To further confirm the stationarity assertion, we fit an AR(2) model to $\hat{w}_t$ and obtain

$$(1 - 0.805B - 0.122B^2)\hat{w}_t = a_t, \qquad \sigma_a = 0.018.$$

Following the discussion of Chapter 2, we can obtain the two characteristic roots of the fitted AR(2) model. Indeed, the model can be rewritten as $(1 - 0.935B)(1 - 0.130B)\hat{w}_t = a_t$. Hence, $\hat{w}_t$ is stationary. Finally, we conduct an augmented Dickey–Fuller unit-root test on $\hat{w}_t$ using an AR(2) model and find that the test statistic is -6.04 with a p value of 0.01. The unit-root hypothesis is clearly rejected.

Maximum-Likelihood Estimation

A formal approach to verify the cointegration of the two log stock prices is to perform a cointegration test. Let $\boldsymbol{x}_t = (p_{1t}, p_{2t})'$. Using information criteria, a VAR(1) model is specified for $\boldsymbol{x}_t$. We then conduct cointegration tests with restricted and unrestricted constant. Both tests give similar results so that we only report the results for the case of restricted constant.



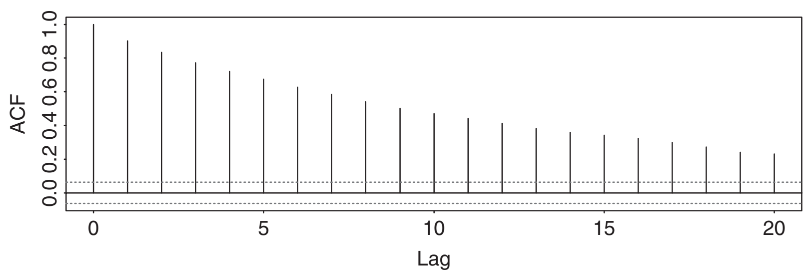


Figure 8.18 Results of least-squares estimation: (a) Time plot of the estimated spread between BHP and VALE daily log stock prices. (b) Sample autocorrelation functions of estimated spread.

```
> coint2=coint(xx,trend="rc")
> coint2
coint(Y = xt, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests signif. at the 5% level are flagged by ' +'.
Trace tests signif. at the 1% level are flagged by '++'.
Max Eigenvalue tests signif. at the 5% level are
    flagged by ' *'.
Max Eigenvalue tests signif. at the 1% level are
    flagged by '**'.

Tests for Cointegration Rank:
         Eigenvalue TraceSt  95%-CV 99%-CV Max-St 95%-CV 99%-
CV
H(0)++**   0.0415  47.7400  19.960 24.600 39.965 15.670 20.200
H(1)       0.0082   7.7748   9.240 12.970  7.774  9.240 12.970
```

The test confirms that $\boldsymbol{x}_t$ is cointegrated. Next, we perform the maximum-likelihood estimation of the error correction model. The results are given below:

```
> n3=VECM(coint2)
> summary(n3)
```

```
VECM(test = coint2)

Cointegrating Vectors:
             coint.1
              1.0000

      vale   -0.7177
 (std.err)    0.0112
  (t.stat)  -64.0913

Intercept*   -1.8144
 (std.err)    0.0169
  (t.stat) -107.0430

VECM Coefficients:
              bhp    vale
  coint.1 -0.0671  0.0263
(std.err)  0.0145  0.0168
 (t.stat) -4.6462  1.5659

 bhp.lag1 -0.1119  0.0659
(std.err)  0.0366  0.0425
 (t.stat) -3.0596  1.5516

vale.lag1  0.0732  0.0445
(std.err)  0.0320  0.0371
 (t.stat)  2.2920  1.1986

Regression Diagnostics:
                  bhp   vale
     R-squared 0.0370 0.0104
Adj. R-squared 0.0350 0.0083
  Resid. Scale 0.0193 0.0224
```

Based on the estimation result, we have the model

$$\Delta \boldsymbol{x}_t = \begin{bmatrix} -0.067 \\ 0.026 \end{bmatrix} (w_{t-1} - 1.81) + \begin{bmatrix} -0.11 & 0.07 \\ 0.07 & 0.04 \end{bmatrix} \Delta \boldsymbol{x}_{t-1} + \boldsymbol{a}_t,$$

where the estimated standard errors of a_{it} are 0.019 and 0.022, respectively. In addition, the spread series is $w_t = p_{1t} - 0.718 p_{2t}$, which is stationary with mean 1.81. Clearly, the result is very close to that of the least-squares estimation. In particular, the γ parameter for the pairs trading is $\hat{\gamma} = 0.718$. Also, as expected, α_1 is negative whereas α_2 is positive.

Trading Strategy

Since the standard error of the spread series w_t is 0.044, we can select $\Delta = 0.045$, which is slightly greater than one standard error of w_t, for pairs trading. This choice

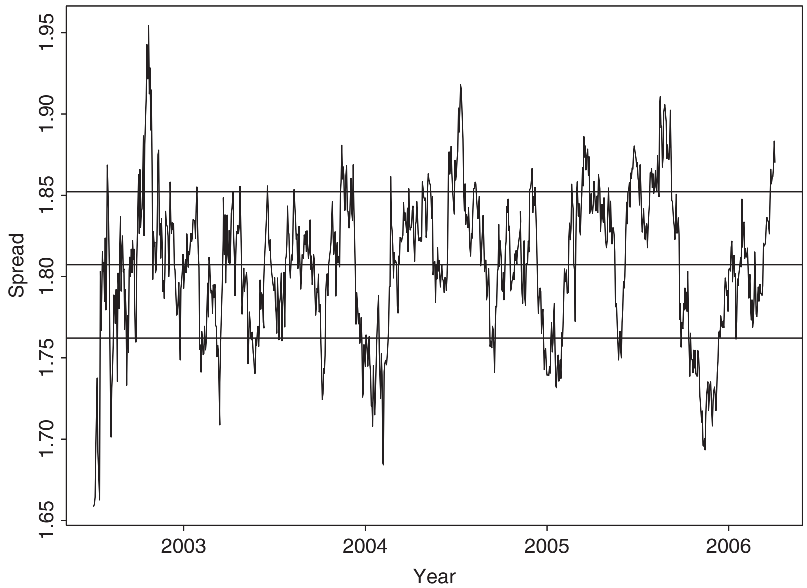


Figure 8.19 Time plot of fitted spread series between daily log prices of BHP and VALE stocks. Three horizontal lines denotes μ_w, $\mu_w + 0.045$, and $\mu_w - 0.045$ with $\mu_w = E(w_t)$.

of Δ ensures that the probability for the spread w_t to deviate Δ away from its mean is not small. In fact, under the normality assumption, the probability is about 30%. Figure 8.19 shows the time plot of the spread series w_t of the fitted error correction model. Three horizontal lines are imposed on the plot. They are μ_w, $\mu_w + 0.045$, and $\mu_w - 0.045$ with the latter two serving as boundaries for pairs trading. Since w_t varies from the lower boundary to the upper one (or from the upper boundary to the lower one) several times, there are many pairs-trading opportunities. From the discussion of Section 8.8.2, the log return of each pairs trading is $2\Delta = 0.09$, which is not small. A more realistic demonstration is to implement the trading in a out-of-sample period. However, the example shows that pairs trading is feasible.

Finally, an important question in pairs trading is to identify the cointegrated pairs of stocks. There are some procedures available in the literature. It seems reasonable to consider pairs of stocks that have similar risk factors. In other words, one should make use of finance theory to guide the selection.

R Demonstration

The following output has been edited:

```
> library(urca)
> help(ca.jo)
> da=read.table("d-bhp0206.txt",header=T)
> da1=read.table("d-vale0206.txt",header=T)
> bhp=log(da[,9])
```

```
> vale=log(da1[,9])
> m1=lm(bhp~vale)
> summary(m1)
Call:
lm(formula = bhp ~ vale)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.822648   0.003662   497.7   >2e-16 ***
vale        0.716664   0.002354   304.4   >2e-16 ***
---
Residual standard error: 0.04421 on 944 degrees of freedom
Multiple R-squared: 0.9899,     Adjusted R-squared: 0.9899
F-statistic: 9.266e+04 on 1 and 944 DF,  p-value: < 2.2e-16

> wt=m1$residuals
> m3=arima(wt,order=c(2,0,0),include.mean=F)
> m3
Call:
arima(x = wt, order = c(2, 0, 0), include.mean = F)

Coefficients:
         ar1     ar2
      0.8051  0.1219
s.e.  0.0322  0.0325

sigma^2 estimated as 0.0003326: log likelihood=2444.76
> p1=c(1,-m3$coef)
> x=polyroot(p1)
> x
[1]  1.069100+0i -7.675365-0i
> 1/Mod(x)
[1] 0.9353661 0.1302870

> xt=cbind(bhp,vale)
> mm=ar(xt)
> mm$order
[1] 2
> cot=ca.jo(xt,ecdet="const",type='trace',K=2,
spec='transitory')
> summary(cot)
######################
# Johansen-Procedure #
######################
Test type: trace statistic, without linear trend and
  constant in cointegration

Eigenvalues (lambda):
```

```
[1]  4.148282e-02  8.206470e-03 -4.610389e-18

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  7.78  7.52  9.24 12.97
r = 0  | 47.77 17.85 19.96 24.60

Eigenvectors, normalised to first column:
(These are the cointegration relations)

            bhp.l1    vale.l1  constant
bhp.l1    1.000000  1.0000000  1.000000
vale.l1  -0.717704 -0.7327542  2.047274
constant -1.828460 -1.5411890 -5.712629

Weights W:
(This is the loading matrix)

            bhp.l1     vale.l1     constant
bhp.d  -0.06731196 0.004568985 9.341093e-18
vale.d  0.02545606 0.007541565 1.015639e-18

> co1=ca.jo(xt,ecdet="const",type='eigen',K=2,
spec='transitory')
> summary(co1)
######################
# Johansen-Procedure #
######################
Test type: maximal eigenvalue statistic (lambda max), without
linear trend and constant in cointegration

Eigenvalues (lambda):
[1]  4.148282e-02  8.206470e-03 -4.610389e-18

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  7.78  7.52  9.24 12.97
r = 0  | 40.00 13.75 15.67 20.20

Eigenvectors, normalised to first column:
(These are the cointegration relations)

            bhp.l1    vale.l1  constant
bhp.l1    1.000000  1.0000000  1.000000
vale.l1  -0.717704 -0.7327542  2.047274
constant -1.828460 -1.5411890 -5.712629
```

```
Weights W:
(This is the loading matrix)

            bhp.l1     vale.l1     constant
bhp.d  -0.06731196 0.004568985 9.341093e-18
vale.d  0.02545606 0.007541565 1.015639e-18
```

APPENDIX A: REVIEW OF VECTORS AND MATRICES

In this appendix, we briefly review some algebra and properties of vectors and matrices. No proofs are given as they can be found in standard textbooks on matrices (e.g., Graybill, 1969).

An $m \times n$ real-valued matrix is an $m \times n$ array of real numbers. For example,

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 8 \\ -1 & 3 & 4 \end{bmatrix}$$

is a 2×3 matrix. This matrix has two rows and three columns. In general, an $m \times n$ matrix is written as

$$\boldsymbol{A} \equiv [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{mn} \end{bmatrix}. \tag{8.46}$$

The positive integers m and n are the *row dimension* and *column dimension* of $\boldsymbol{A}$. The real number a_{ij} is referred to as the (i, j)th element of $\boldsymbol{A}$. In particular, the elements a_{ii} are the *diagonal elements* of the matrix.

An $m \times 1$ matrix forms an m-dimensional column vector, and a $1 \times n$ matrix is an n-dimensional row vector. In the literature, a vector is often meant to be a column vector. If $m = n$, then the matrix is a square matrix. If $a_{ij} = 0$ for $i \neq j$ and $m = n$, then the matrix $\boldsymbol{A}$ is a *diagonal matrix*. If $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = 1$ for all i, then $\boldsymbol{A}$ is the $m \times m$ *identity matrix*, which is commonly denoted by $\boldsymbol{I}_m$ or simply $\boldsymbol{I}$ if the dimension is clear.

The $n \times m$ matrix

$$\boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m-1,1} & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m-1,2} & a_{m2} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{m-1,n} & a_{mn} \end{bmatrix}$$

is the *transpose* of the matrix $\boldsymbol{A}$. For example,

$$\begin{bmatrix} 2 & -1 \\ 5 & 3 \\ 8 & 4 \end{bmatrix} \quad \text{is the transpose of} \quad \begin{bmatrix} 2 & 5 & 8 \\ -1 & 3 & 4 \end{bmatrix}.$$

We use the notation $\boldsymbol{A}' = [a'_{ij}]$ to denote the transpose of $\boldsymbol{A}$. From the definition, $a'_{ij} = a_{ji}$ and $(\boldsymbol{A}')' = \boldsymbol{A}$. If $\boldsymbol{A}' = \boldsymbol{A}$, then $\boldsymbol{A}$ is a *symmetric matrix*.

Basic Operations

Suppose that $\boldsymbol{A} = [a_{ij}]_{m\times n}$ and $\boldsymbol{C} = [c_{ij}]_{p\times q}$ are two matrices with dimensions given in the subscript. Let b be a real number. Some basic matrix operations are defined next:

- Addition: $\boldsymbol{A} + \boldsymbol{C} = [a_{ij} + c_{ij}]_{m\times n}$ if $m = p$ and $n = q$.
- Subtraction: $\boldsymbol{A} - \boldsymbol{C} = [a_{ij} - c_{ij}]_{m\times n}$ if $m = p$ and $n = q$.
- Scalar multiplication: $b\boldsymbol{A} = [ba_{ij}]_{m\times n}$.
- Multiplication: $\boldsymbol{AC} = [\sum_{v=1}^{n} a_{iv}c_{vj}]_{m\times q}$ provided that $n = p$.

When the dimensions of matrices satisfy the condition for multiplication to take place, the two matrices are said to be *conformable*. An example of matrix multiplication is

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & -4 \end{bmatrix} = \begin{bmatrix} 2\cdot 1 - 1\cdot 1 & 2\cdot 2 + 1\cdot 2 & 2\cdot 3 - 1\cdot 4 \\ 1\cdot 1 - 1\cdot 1 & 1\cdot 2 + 1\cdot 2 & 1\cdot 3 - 1\cdot 4 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 6 & 2 \\ 0 & 4 & -1 \end{bmatrix}.$$

Important rules of matrix operations include (a) $(\boldsymbol{AC})' = \boldsymbol{C}'\boldsymbol{A}'$ and (b) $\boldsymbol{AC} \neq \boldsymbol{CA}$ in general.

Inverse, Trace, Eigenvalue, and Eigenvector

A square matrix $\boldsymbol{A}_{m\times m}$ is *nonsingular* or *invertible* if there exists a unique matrix $\boldsymbol{C}_{m\times m}$ such that $\boldsymbol{AC} = \boldsymbol{CA} = \boldsymbol{I}_m$, the $m \times m$ identity matrix. In this case, $\boldsymbol{C}$ is called the *inverse* matrix of $\boldsymbol{A}$ and is denoted by $\boldsymbol{C} = \boldsymbol{A}^{-1}$.

The trace of $\boldsymbol{A}_{m\times m}$ is the sum of its diagonal elements [i.e., $\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{m} a_{ii}$]. It is easy to see that (a) $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{C}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{C})$, (b) $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}')$, and (c) $\mathrm{tr}(\boldsymbol{AC}) = \mathrm{tr}(\boldsymbol{CA})$ provided that the two matrices are conformable.

A number λ and an $m \times 1$ vector $\boldsymbol{b}$, possibly complex valued, are a right *eigenvalue* and *eigenvector* pair of the matrix $\boldsymbol{A}$ if $\boldsymbol{Ab} = \lambda\boldsymbol{b}$. There are m possible eigenvalues for the matrix $\boldsymbol{A}$. For a real-valued matrix $\boldsymbol{A}$, complex eigenvalues occur in conjugated pairs. The matrix $\boldsymbol{A}$ is nonsingular if and only if all of its eigenvalues are nonzero. Denote the eigenvalues by $\{\lambda_i | i = 1, \ldots, m\}$: We have $\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{m} \lambda_i$. In addition, the *determinant* of the matrix $\boldsymbol{A}$ can be defined as $|\boldsymbol{A}| = \prod_{i=1}^{m} \lambda_i$. For a general definition of determinant of a matrix, see a standard textbook on matrices (e.g., Graybill, 1969).

Finally, the rank of the matrix $\boldsymbol{A}_{m\times n}$ is the number of nonzero eigenvalues of the symmetric matrix $\boldsymbol{AA}'$. Also, for a nonsingular matrix $\boldsymbol{A}$, $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$.

Positive-Definite Matrix

A square matrix $\boldsymbol{A}$ ($m \times m$) is a *positive-definite* matrix if (a) $\boldsymbol{A}$ is symmetric and (b) all eigenvalues of $\boldsymbol{A}$ are positive. Alternatively, $\boldsymbol{A}$ is a positive-definite matrix if for any nonzero m-dimensional vector $\boldsymbol{b}$, we have $\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b} > 0$.

Useful properties of a positive-definite matrix $\boldsymbol{A}$ include (a) all eigenvalues of $\boldsymbol{A}$ are real and positive, and (b) the matrix can be decomposed as

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}',$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of all eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an $m \times m$ matrix consisting of the m right eigenvectors of $\boldsymbol{A}$. It is common to write the eigenvalues as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and the eigenvectors as $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ such that $\boldsymbol{A}\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i$ and $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$. In addition, these eigenvectors are orthogonal to each other—namely, $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$ if $i \neq j$—if the eigenvalues are distinct. The matrix $\boldsymbol{P}$ is an *orthogonal* matrix and the decomposition is referred to as the *spectral decomposition* of the matrix $\boldsymbol{A}$. Consider, for example, the simple 2×2 matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

which is positive definite. Simple calculations show that

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Therefore, 3 and 1 are eigenvalues of $\boldsymbol{\Sigma}$ with normalized eigenvectors $(1/\sqrt{2}, 1/\sqrt{2})'$ and $(1/\sqrt{2}, -1/\sqrt{2})'$, respectively. It is easy to verify that the spectral decomposition holds—that is,

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

For a symmetric matrix $\boldsymbol{A}$, there exists a lower triangular matrix $\boldsymbol{L}$ with diagonal elements being 1 and a diagonal matrix $\boldsymbol{G}$ such that $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$; see Chapter 1 of Strang (1980). If $\boldsymbol{A}$ is positive definite, then the diagonal elements of $\boldsymbol{G}$ are positive. In this case, we have

$$\boldsymbol{A} = \boldsymbol{L}\sqrt{\boldsymbol{G}}\sqrt{\boldsymbol{G}}\boldsymbol{L}' = (\boldsymbol{L}\sqrt{\boldsymbol{G}})(\boldsymbol{L}\sqrt{\boldsymbol{G}})',$$

where $\boldsymbol{L}\sqrt{\boldsymbol{G}}$ is again a lower triangular matrix and the square root is taken element by element. Such a decomposition is called the *Cholesky decomposition* of $\boldsymbol{A}$. This decomposition shows that a positive-definite matrix $\boldsymbol{A}$ can be diagonalized as

$$\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}')^{-1} = \boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{-1})' = \boldsymbol{G}.$$

Since $\boldsymbol{L}$ is a lower triangular matrix with unit diagonal elements, $\boldsymbol{L}^{-1}$ is also lower triangular matrix with unit diagonal elements. Consider again the prior 2×2 matrix $\boldsymbol{\Sigma}$. It is easy to verify that

$$\boldsymbol{L} = \begin{bmatrix} 1.0 & 0.0 \\ 0.5 & 1.0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 1.5 \end{bmatrix}$$

satisfy $\boldsymbol{\Sigma} = \boldsymbol{LGL}'$. In addition,

$$\boldsymbol{L}^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^{-1})' = \boldsymbol{G}.$$

Vectorization and Kronecker Product

Writing an $m \times n$ matrix $\boldsymbol{A}$ in its columns as $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$, we define the stacking operation as $\text{vec}(\boldsymbol{A}) = (\boldsymbol{a}_1', \boldsymbol{a}_2', \ldots, \boldsymbol{a}_m')'$, which is an $mn \times 1$ vector. For two matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{C}_{p\times q}$, the Kronecker product between $\boldsymbol{A}$ and $\boldsymbol{C}$ is

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{bmatrix} a_{11}\boldsymbol{C} & a_{12}\boldsymbol{C} & \cdots & a_{1n}\boldsymbol{C} \\ a_{21}\boldsymbol{C} & a_{22}\boldsymbol{C} & \cdots & a_{2n}\boldsymbol{C} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{C} & a_{m2}\boldsymbol{C} & \cdots & a_{mn}\boldsymbol{C} \end{bmatrix}_{mp\times nq}.$$

For example, assume that

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \qquad \boldsymbol{C} = \begin{bmatrix} 4 & -1 & 3 \\ -2 & 5 & 2 \end{bmatrix}.$$

Then $\text{vec}(\boldsymbol{A}) = (2, -1, 1, 3)'$, $\text{vec}(\boldsymbol{C}) = (4, -2, -1, 5, 3, 2)'$, and

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{bmatrix} 8 & -2 & 6 & 4 & -1 & 3 \\ -4 & 10 & 4 & -2 & 5 & 2 \\ -4 & 1 & -3 & 12 & -3 & 9 \\ 2 & -5 & -2 & -6 & 15 & 6 \end{bmatrix}.$$

Assuming that the dimensions are appropriate, we have the following useful properties for the two operators:

1. $\boldsymbol{A} \otimes \boldsymbol{C} \neq \boldsymbol{C} \otimes \boldsymbol{A}$ in general.
2. $(\boldsymbol{A} \otimes \boldsymbol{C})' = \boldsymbol{A}' \otimes \boldsymbol{C}'$.
3. $\boldsymbol{A} \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D}$.
4. $(\boldsymbol{A} \otimes \boldsymbol{C})(\boldsymbol{F} \otimes \boldsymbol{G}) = (\boldsymbol{AF}) \otimes (\boldsymbol{CG})$.
5. If $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible, then $(\boldsymbol{A} \otimes \boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{C}^{-1}$.
6. For square matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, $\text{tr}(\boldsymbol{A} \otimes \boldsymbol{C}) = \text{tr}(\boldsymbol{A})\text{tr}(\boldsymbol{C})$.

7. $\text{vec}(\boldsymbol{A}+\boldsymbol{C}) = \text{vec}(\boldsymbol{A}) + \text{vec}(\boldsymbol{C})$.
8. $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\ \text{vec}(\boldsymbol{B})$.
9. $\text{tr}(\boldsymbol{AC}) = \text{vec}(\boldsymbol{C}')'\text{vec}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A}')'\text{vec}(\boldsymbol{C})$.
10. $$\begin{aligned}\text{tr}(\boldsymbol{ABC}) &= \text{vec}(\boldsymbol{A}')'(\boldsymbol{C}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{B}) = \text{vec}(\boldsymbol{A}')'(\boldsymbol{I} \otimes \boldsymbol{B})\text{vec}(\boldsymbol{C})\\ &= \text{vec}(\boldsymbol{B}')'(\boldsymbol{A}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{C}) = \text{vec}(\boldsymbol{B}')'(\boldsymbol{I} \otimes \boldsymbol{C})\text{vec}(\boldsymbol{A})\\ &= \text{vec}(\boldsymbol{C}')'(\boldsymbol{B}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{A}) = \text{vec}(\boldsymbol{C}')'(\boldsymbol{I} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B}).\end{aligned}$$

In multivariate statistical analysis, we often deal with symmetric matrices. It is therefore convenient to generalize the stacking operation to the *half-stacking* operation, which consists of elements on or below the main diagonal. Specifically, for a symmetric square matrix $\boldsymbol{A} = [a_{ij}]_{k\times k}$, define

$$\text{vech}(\boldsymbol{A}) = (\boldsymbol{a}'_{1.}, \boldsymbol{a}'_{2*}, \ldots, \boldsymbol{a}'_{k*})',$$

where $\boldsymbol{a}_{1.}$ is the first column of $\boldsymbol{A}$, and $\boldsymbol{a}_{i*} = (a_{ii}, a_{i+1,i}, \ldots, a_{ki})'$ is a $(k-i+1)$-dimensional vector. The dimension of vech($\boldsymbol{A}$) is $k(k+1)/2$. For example, suppose that $k = 3$. Then we have $\text{vech}(\boldsymbol{A}) = (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})'$, which is a six-dimensional vector.

APPENDIX B: MULTIVARIATE NORMAL DISTRIBUTIONS

A k-dimensional random vector $\boldsymbol{x} = (x_1, \ldots, x_k)'$ follows a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and positive-definite covariance matrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$ if its probability density function (pdf) is

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]. \tag{8.47}$$

We use the notation $\boldsymbol{x} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote that $\boldsymbol{x}$ follows such a distribution. This normal distribution plays an important role in multivariate statistical analysis and it has several nice properties. Here we consider only those properties that are relevant to our study. Interested readers are referred to Johnson and Wichern (1998) for details.

To gain insight into multivariate normal distributions, consider the bivariate case (i.e., $k = 2$). In this case, we have

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \qquad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}.$$

Using the correlation coefficient $\rho = \sigma_{12}/(\sigma_1\sigma_2)$, where $\sigma_i = \sqrt{\sigma_{ii}}$ is the standard deviation of x_i, we have $\sigma_{12} = \rho\sqrt{\sigma_{11}\sigma_{22}}$ and $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22}(1-\rho^2)$. The pdf of $\boldsymbol{x}$ then becomes

$$f(x_1, x_2|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[Q(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})]\right\},$$

where

$$Q(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right).$$

Chapter 4 of Johnson and Wichern (1998) contains some plots of this pdf function.

Let $\boldsymbol{c} = (c_1, \ldots, c_k)'$ be a nonzero k-dimensional vector. Partition the random vector as $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, where $\boldsymbol{x}_1 = (x_1, \ldots, x_p)'$ and $\boldsymbol{x}_2 = (x_{p+1}, \ldots, x_k)'$ with $1 \le p < k$. Also partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly as

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right).$$

Some properties of $\boldsymbol{x}$ are as follows:

1. $\boldsymbol{c}'\boldsymbol{x} \sim N(\boldsymbol{c}'\boldsymbol{\mu}, \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c})$. That is, any nonzero linear combination of $\boldsymbol{x}$ is univariate normal. The inverse of this property also holds. Specifically, if $\boldsymbol{c}'\boldsymbol{x}$ is univariate normal for any nonzero vector $\boldsymbol{c}$, then $\boldsymbol{x}$ is multivariate normal.
2. The marginal distribution of $\boldsymbol{x}_i$ is normal. In fact, $\boldsymbol{x}_i \sim N_{k_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$ for $i =$ 1 and 2, where $k_1 = p$ and $k_2 = k - p$.
3. $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$ if and only if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent.
4. The random variable $y = (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ follows a chi-squared distribution with m degrees of freedom.
5. The conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2 = \boldsymbol{b}$ is also normally distributed as

$$(\boldsymbol{x}_1 | \boldsymbol{x}_2 = \boldsymbol{b}) \sim N_p[\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}].$$

The last property is useful in many scientific areas. For instance, it forms the basis for time series forecasting under the normality assumption and for recursive least-squares estimation.

APPENDIX C: SOME SCA COMMANDS

The following SCA commands are used in the analysis of Example 8.6:

```
input x1,x2. file 'm-gs1n3-5301.txt'  % Load data
 --
r1=ln(x1)  % Take log transformation
 --
r2=ln(x2)
 --
miden r1,r2. no ccm. arfits 1 to 8.
 -- % Denote the model by v21.
mtsm v21. series r1,r2. model (i-p1*b-p2*b**2)series= @
```

```
c+(i-t1*b)noise.
 --
mestim v21.   % Initial estimation
 --
p1(2,1)=0     % Set zero constraints
 --
cp1(2,1)=1
 --
p2(2,1)=0
 --
cp2(2,1)=1
 --
p2(2,2)=0
 --
cp2(2,2)=1
 --
t1(2,1)=0
 --
ct1(2,1)=1
 -- % Refine estimation and store residuals
 mestim v21. method exact. hold resi(res1,res2)
 --
miden res1,res2.
```

EXERCISES

8.1. Consider the monthly log stock returns, in percentages and including dividends, of Merck & Company, Johnson & Johnson, General Electric, General Motors, Ford Motor Company, and value-weighted index from January 1960 to December 2008; see the file `m-mrk2vw.txt`.

(a) Compute the sample mean, covariance matrix, and correlation matrix of the data.

(b) Test the hypothesis $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_6 = \mathbf{0}$, where $\boldsymbol{\rho}_i$ is the lag-i cross-correlation matrix of the data. Draw conclusions based on the 5% significance level.

(c) Is there any lead–lag relationship among the six return series?

8.2. The Federal Reserve Bank of St. Louis publishes selected interest rates and U.S. financial data on its website: http://research.stlouisfed.org/fred2/.
Consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2009 for 679 observations; see the file `m-gs1n10.txt`. The rates are in percentages.

(a) Let $c_t = r_t - r_{t-1}$ be the change series of the monthly interest rate r_t. Build a bivariate autoregressive model for the two change series. Discuss the implications of the model. Transform the model into a structural form.

(b) Build a bivariate moving-average model for the two change series. Discuss the implications of the model and compare it with the bivariate AR model built earlier.

8.3. Again consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2009. Consider the log series of the data and build a VARMA model for the series. Discuss the implications of the model obtained.

8.4. Again consider the monthly 1-year and 10-year Treasury constant maturity rates from April 1953 to October 2009. Are the two interest rate series threshold cointegrated? Use the interest spread $s_t = r_{10,t} - r_{1,t}$ as the threshold variable, where r_{it} is the i-year Treasury constant maturity rate. If they are threshold cointegrated, build a multivariate threshold model for the two series.

8.5. The bivariate AR(4) model $\boldsymbol{x}_t - \boldsymbol{\Phi}_4\boldsymbol{x}_{t-4} = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$ is a special seasonal model with periodicity 4, where $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Such a seasonal model may be useful in studying quarterly earnings of a company. (a) Assume that $\boldsymbol{x}_t$ is weakly stationary. Derive the mean vector and covariance matrix of $\boldsymbol{x}_t$. (b) Derive the necessary and sufficient condition of weak stationarity for $\boldsymbol{x}_t$. (c) Show that $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_4\boldsymbol{\Gamma}_{\ell-4}$ for $\ell > 0$, where $\boldsymbol{\Gamma}_\ell$ is the lag-ℓ autocovariance matrix of $\boldsymbol{x}_t$.

8.6. The bivariate MA(4) model $\boldsymbol{x}_t = \boldsymbol{a}_t - \boldsymbol{\Theta}_4\boldsymbol{a}_{t-4}$ is another seasonal model with periodicity 4, where $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Derive the covariance matrices $\boldsymbol{\Gamma}_\ell$ of $\boldsymbol{x}_t$ for $\ell = 0, \ldots, 5$.

8.7. Consider the monthly U.S. 1-year and 3-year Treasury constant maturity rates from April 1953 to March 2004. The data can be obtained from the Federal Reserve Bank of St. Louis or from the file `m-gs1n3-5304.txt` (1-year, 3-year, dates). See also Example 8.6, which uses a shorter data span. Here we use the interest rates directly without the log transformation and define $\boldsymbol{x}_t = (x_{1t}, x_{2t})'$, where x_{1t} is the 1-year maturity rate and x_{2t} is the 3-year maturity rate.

(a) Identify a VAR model for the bivariate interest rate series. Write down the fitted model.

(b) Compute the impulse response functions of the fitted VAR model. It suffices to use the first 6 lags.

(c) Use the fitted VAR model to produce 1-step- to 12-step-ahead forecasts of the interest rates, assuming that the forecast origin is March 2004.

(d) Are the two interest rate series cointegrated, when a restricted constant term is used? Use 5% significance level to perform the test.

(e) If the series are cointegrated, build an ECM for the series. Write down the fitted model.

(f) Use the fitted ECM to produce 1-step- to 12-step-ahead forecasts of the interest rates, assuming that the forecast origin is March 2004.

(g) Compare the forecasts produced by the VAR model and the ECM.

REFERENCES

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review* **38**: 627–645.

Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**: 355–366.

Brenner, R. J. and Kroner, K. F. (1995). Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. *Journal of Financial and Quantitative Analysis* **30**: 23–42.

Cochrane, J. H. (1988). How big is the random walk in the GNP? *Journal of Political Economy* **96**: 893–920.

Dwyer, G. P. Jr., Locke, P., and Yu, W. (1996). Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash. *Review of Financial Studies* **9**: 301–332.

Engle, R. F. and Granger, C. W. J. (1987). Cointegration and error correction representation, estimation and testing. *Econometrica* **55**: 251–276.

Forbes, C. S., Kalb, G. R. J., and Kofman, P. (1999). Bayesian arbitrage threshold analysis. *Journal of Business & Economic Statistics* **17**: 364–372.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth, Belmont, CA.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society Series B* **41**: 190–195.

Hillmer, S. C. and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association* **74**: 652–660.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**: 602–608.

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

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.

Johansen, S. (1995). *Likelihood Based Inference in Cointegrated Vector Error Correction Models*. Oxford University Press, Oxford, UK.

Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th ed. Prentice Hall, Upper Saddle River, NJ.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.

Pole, A. (2007). *Statistical Arbitrage*. Wiley, Hoboken, NJ.

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

Reinsel, G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis*, **13**: 353–375.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association* **83**: 1097–1107.

Strang, G. (1980). *Linear Algebra and Its Applications*, 2nd ed. Harcourt Brace Jovanovich, Chicago.

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

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series (with discussions). *Journal of the Royal Statistical Society Series B* **51**: 157–213.

Tiao, G. C., Tsay, R. S., and Wang, T. (1993). Usefulness of linear transformations in multivariate time series analysis. *Empirical Economics* **18**: 567–593.

Tsay, R. S. (1991). Two canonical forms for vector ARMA processes. *Statistica Sinica* **1**: 247–269.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1202.

Tsay, R. S. and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics* **18**: 220–250.

Vidyamurthy, G. (2004). *Pairs Trading*. Wiley, Hoboken, NJ.

Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*. Springer, New York.