CHAPTER 2

Stationary Vector Autoregressive Time Series

2.1 INTRODUCTION

The most commonly used multivariate time series model is the vector autoregressive (VAR) model, particularly so in the econometric literature for good reasons. First, the model is relatively easy to estimate. One can use the least-squares (LS) method, the maximum likelihood (ML) method, or Bayesian method. All three estimation methods have closed-form solutions. For a VAR model, the least-squares estimates are asymptotically equivalent to the ML estimates and the ordinary least-squares (OLS) estimates are the same as the generalized least-squares (GLS) estimates. Second, the properties of VAR models have been studied extensively in the literature. Finally, VAR models are similar to the multivariate multiple linear regressions widely used in multivariate statistical analysis. Many methods for making inference in multivariate multiple linear regression apply to the VAR model.

The multivariate time series $\boldsymbol{z}_t$ follows a VAR model of order p, VAR(p), if

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \boldsymbol{a}_t, \tag{2.1}$$

where $\boldsymbol{\phi}_0$ is a k-dimensional constant vector and $\boldsymbol{\phi}_i$ are $k \times k$ matrices for $i > 0$, $\boldsymbol{\phi}_p \neq \mathbf{0}$, and $\boldsymbol{a}_t$ is a sequence of independent and identically distributed (iid) random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}_a$, which is positive-definite. This is a special case of the VARMA(p, q) model of Chapter 1 with $q = 0$. With the back-shift operator, the model becomes $\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$, where $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$ is a matrix polynomial of degree p. See Equation (1.21). We shall refer to $\boldsymbol{\phi}_\ell = [\phi_{\ell,ij}]$ as the lag ℓ AR coefficient matrix.

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

To study the properties of VAR(p) models, we start with the simple VAR(1) and VAR(2) models. In many cases, we use bivariate time series in our discussion, but the results continue to hold for the k-dimensional series.

2.2 VAR(1) MODELS

To begin, consider the bivariate VAR(1) model

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t.$$

This model can be written explicitly as

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \tag{2.2}$$

or equivalently,

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

Thus, the (1,2)th element of $\boldsymbol{\phi}_1$, that is, $\phi_{1,12}$, shows the linear dependence of z_{1t} on $z_{2,t-1}$ in the presence of $z_{1,t-1}$. The (2,1)th element of $\boldsymbol{\phi}_1$, $\phi_{1,21}$, measures the linear relationship between z_{2t} and $z_{1,t-1}$ in the presence of $z_{2,t-1}$. Other parameters in $\boldsymbol{\phi}_1$ can be interpreted in a similar manner.

2.2.1 Model Structure and Granger Causality

If the off-diagonal elements of $\boldsymbol{\phi}_1$ are 0, that is, $\phi_{1,12} = \phi_{1,21} = 0$, then z_{1t} and z_{2t} are not dynamically correlated. In this particular case, each series follows a univariate AR(1) model and can be handled accordingly. We say that the two series are uncoupled.

If $\phi_{1,12} = 0$, but $\phi_{1,21} \neq 0$, then we have

$$z_{1t} = \phi_{10} + \phi_{1,11} z_{1,t-1} + a_{1t}, \tag{2.3}$$

$$z_{2t} = \phi_{20} + \phi_{1,21} z_{1,t-1} + \phi_{1,22} z_{2,t-1} + a_{2t}. \tag{2.4}$$

This particular model shows that z_{1t} does not depend on the past value of z_{2t}, but z_{2t} depends on the past value of z_{1t}. Consequently, we have a unidirectional relationship with z_{1t} acting as the input variable and z_{2t} as the output variable. In the statistical literature, the two series z_{1t} and z_{2t} are said to have a *transfer function* relationship. Transfer function models, which can be regarded as a special case of the VARMA model, are useful in control engineering as one can adjust the value of z_{1t} to influence the future value of z_{2t}. In the econometric literature, the model implies the existence of Granger causality between the two series with z_{1t} causing z_{2t}, but not being caused by z_{2t}.

Granger (1969) introduces the concept of causality, which is easy to deal with for a VAR model. Consider a bivariate series and the h step ahead forecast. In this case, we can use the VAR model and univariate models for individual components to produce forecasts. We say that z_{1t} causes z_{2t} if the bivariate forecast for z_{2t} is more accurate than its univariate forecast. Here, the accuracy of a forecast is measured by the variance of its forecast error. In other words, under Granger's framework, we say that z_{1t} causes z_{2t} if the past information of z_{1t} improves the forecast of z_{2t}.

We can make the statement more precisely. Let F_t denote the available information at time t (inclusive). Let $F_{-i,t}$ be F_t with all information concerning the ith component z_{it} removed. Consider the bivariate VAR(1) model in Equation (2.2). F_t consists of $\{\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \ldots\}$, whereas $F_{-2,t}$ consists of the past values $\{z_{1t}, z_{1,t-1}, \ldots\}$ and $F_{-1,t}$ contains $\{z_{2t}, z_{2,t-1}, \ldots\}$. Now, consider the h step ahead forecast $\boldsymbol{z}_t(h)$ based on F_t and the associated forecast error $\boldsymbol{e}_t(h)$. See Section 1.6. Let $z_{j,t+h}|F_{-i,t}$ be the h step ahead prediction of $z_{j,t+h}$ based on $F_{-i,t}$ and $e_{j,t+h}|F_{-i,t}$ be the associated forecast error, where $i \neq j$. Then, z_{1t} causes z_{2t} if Var$[e_{2t}(h)] <$ Var$[e_{2,t+h}|F_{-1,t}]$.

Return to the bivariate VAR(1) model with $\phi_{1,12} = 0$, but $\phi_{1,21} \neq 0$. We see that $z_{2,t+1}$ depends on z_{1t} so that knowing z_{1t} is helpful in forecasting $z_{2,t+1}$. On the other hand, $z_{1,t+1}$ does not depend on any past value of z_{2t} so that knowing the past values of z_{2t} will not help in predicting $z_{1,t+1}$. Thus, z_{1t} causes z_{2t}, but z_{2t} does not cause z_{1t}.

Similarly, if $\phi_{1,21} = 0$, but $\phi_{1,12} \neq 0$, then z_{2t} causes z_{1t}, but z_{1t} does not cause z_{2t}.

Remark: For the bivariate VAR(1) model in Equation (2.2), if $\boldsymbol{\Sigma}_a$ is not a diagonal matrix, then z_{1t} and z_{2t} are instantaneously correlated (or contemporaneously correlated). In this case, z_{1t} and z_{2t} have instantaneous Granger causality. The instantaneous causality goes in both ways. □

Remark: The statement that $\phi_{1,12} = 0$ and $\phi_{1,21} \neq 0$ in a VAR(1) model implies the existence of Granger's causality depends critically on the unique VAR parameterization we use, namely, the matrix polynomial $\boldsymbol{\phi}(B)$ starts with the identity matrix $\boldsymbol{I}_k$. To see this, consider the model in Equations (2.3) and (2.4). For simplicity, assume the constant term being 0, that is, $\phi_{10} = \phi_{20} = 0$. If we multiply Equation (2.4) by a nonzero parameter β and add the result to Equation (2.3), then we have

$$z_{1t} + \beta z_{2t} = (\phi_{1,11} + \beta\phi_{1,21}) z_{1,t-1} + \beta\phi_{1,22} z_{2,t-1} + a_{1t} + \beta a_{2t}.$$

Combining this equation with Equation (2.4), we have

$$\begin{bmatrix} 1 & \beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{1,11} + \beta\phi_{1,21} & \beta\phi_{1,22} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix}, \qquad (2.5)$$

where $b_{1t} = a_{1t} + \beta a_{2t}$ and $b_{2t} = a_{2t}$. Equation (2.5) remains a VAR(1) model. It has the same underlying structure as that of Equations (2.3) and (2.4). In particular, z_{1t} does not depend on $z_{2,t-1}$. Yet the parameter in the (1,2)th position of the AR coefficient matrix in Equation (2.5) is not zero. This nonzero parameter $\beta\phi_{1,22}$ is induced by the nonzero parameter β in the left side of Equation (2.5). □

The flexibility in the structure of a VAR(1) model increases with the dimension k. For example, consider the three-dimensional VAR(1) model $(\boldsymbol{I}_3 - \boldsymbol{\phi}_1 B)\boldsymbol{z}_t = \boldsymbol{a}_t$ with $\boldsymbol{\phi}_1 = [\phi_{1,ij}]_{3\times 3}$. If $\boldsymbol{\phi}_1$ is a lower triangular matrix, then z_{1t} does not depend on the past values of z_{2t} or z_{3t}, z_{2t} may depend on the past value of z_{1t}, but not on the past value of z_{3t}. In this case, we have a unidirectional relationship from z_{1t} to z_{2t} to z_{3t}. On the other hand, if $\phi_{1,13} = \phi_{1,23} = 0$ and $\phi_{1,ij} \neq 0$, otherwise, then we have a unidirectional relationship from both z_{1t} and z_{2t} to z_{3t}, whereas z_{1t} and z_{2t} are dynamically correlated. There are indeed many other possibilities.

2.2.2 Relation to Transfer Function Model

However, the model representation in Equations (2.3) and (2.4) is in general not a transfer function model because the two innovations a_{1t} and a_{2t} might be correlated. In a transfer function model, which is also known as a distributed-lag model, the input variable should be independent of the disturbance term of the output variable. To obtain a transfer function model, we perform orthogonalization of the two innovations in $\boldsymbol{a}_t$. Specifically, consider the simple linear regression

$$a_{2t} = \beta a_{1t} + \epsilon_t,$$

where $\beta = \text{cov}(a_{1t}, a_{2t})/\text{var}(a_{1t})$ and a_{1t} and ϵ_t are uncorrelated. By plugging a_{2t} into Equation (2.4), we obtain

$$(1 - \phi_{1,22}B)z_{2t} = (\phi_{20} - \beta\phi_{10}) + [\beta + (\phi_{1,21} - \beta\phi_{1,11})B]z_{1t} + \epsilon_t.$$

The prior equation further simplifies to

$$z_{2t} = \frac{\phi_{20} - \beta\phi_{10}}{1 - \phi_{1,22}} + \frac{\beta + (\phi_{1,21} - \beta\phi_{1,11})B}{1 - \phi_{1,22}B} z_{1t} + \frac{1}{1 - \phi_{1,22}B}\epsilon_t,$$

which is a transfer function model. The exogenous variable z_{1t} does not depend on the innovation ϵ_t.

2.2.3 Stationarity Condition

As defined in Chapter 1, a (weakly) stationary time series $\boldsymbol{z}_t$ has time invariant mean and covariance matrix. For these two conditions to hold, the mean of $\boldsymbol{z}_t$ should not depend on when the series started or what was its starting value. A simple way to investigate the stationarity condition for the VAR(1) model is then to exploit this feature of the series. For simplicity in discussion, we assume that the constant term $\boldsymbol{\phi}_0$ is 0, and the model reduces to $\boldsymbol{z}_t = \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$.

Suppose that the time series started at time $t = v$ with initial value $\boldsymbol{z}_v$, where v is a fixed time point. As time advances, the series $\boldsymbol{z}_t$ evolves. Specifically, by repeated substitutions, we have

$$\begin{aligned}
\boldsymbol{z}_t &= \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t \\
&= \boldsymbol{\phi}_1(\boldsymbol{\phi}_1 \boldsymbol{z}_{t-2} + \boldsymbol{a}_{t-1}) + \boldsymbol{a}_t \\
&= \boldsymbol{\phi}_1^2 \boldsymbol{z}_{t-2} + \boldsymbol{\phi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{a}_t \\
&= \boldsymbol{\phi}_1^3 \boldsymbol{z}_{t-3} + \boldsymbol{\phi}_1^2 \boldsymbol{a}_{t-2} + \boldsymbol{\phi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{a}_t \\
&= \vdots \\
&= \boldsymbol{\phi}_1^{t-v} \boldsymbol{z}_v + \sum_{i=0}^{t-1} \boldsymbol{\phi}_1^i \boldsymbol{a}_{t-i}.
\end{aligned}$$

Consequently, for $\boldsymbol{z}_t$ to be independent of $\boldsymbol{z}_v$, we need $\boldsymbol{\phi}_1^{t-v}$ goes to 0 as $v \to -\infty$. Here, $v \to -\infty$ means that the series started a long time ago. Recall that if $\{\lambda_1, \ldots, \lambda_k\}$ are the eigenvalues of $\boldsymbol{\phi}_1$, then $\{\lambda_1^n, \ldots, \lambda_k^n\}$ are the eigenvalues of $\boldsymbol{\phi}_1^n$; see Appendix A. Also, if all eigenvalues of a matrix are 0, then the matrix must be 0. Consequently, the condition for $\boldsymbol{\phi}_1^{t-v} \to \mathbf{0}$ as $v \to -\infty$ is that all eigenvalues λ_j of $\boldsymbol{\phi}_1$ must satisfy $\lambda_j^{t-v} \to 0$ as $v \to -\infty$. This implies that the absolute values of all eigenvalues λ_j of $\boldsymbol{\phi}_1$ must be less than 1.

Consequently, a necessary condition for the VAR(1) series $\boldsymbol{z}_t$ to be stationary is that all eigenvalues of $\boldsymbol{\phi}_1$ must be less than 1 in absolute value. It can also be shown that if all eigenvalues of $\boldsymbol{\phi}_1$ are less than 1 in absolute value, then the VAR(1) series $\boldsymbol{z}_t$ is stationary.

The eigenvalues of $\boldsymbol{\phi}_1$ are solutions of the determinant equation

$$|\lambda \boldsymbol{I}_k - \boldsymbol{\phi}_1| = 0,$$

which can be written as

$$\lambda^k \left| \boldsymbol{I}_k - \boldsymbol{\phi}_b \frac{1}{\lambda} \right| = 0.$$

Therefore, we can consider the determinant equation $|\boldsymbol{I}_k - \boldsymbol{\phi}_1 x| = 0$, where $x = 1/\lambda$. Eigenvalues of $\boldsymbol{\phi}_1$ are the inverses of the solutions of this new equation.

Consequently, the necessary and sufficient condition for the stationarity of a VAR(1) model is that the solutions of the determinant equation $|\boldsymbol{I}_k - \boldsymbol{\phi}_1 B| = 0$ are greater than 1 in absolute value. That is, the solutions of the determinant equation $|\boldsymbol{\phi}(B)| = 0$ are outside the unit circle.

Example 2.1 Consider the bivariate VAR(1) model

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \tag{2.6}$$

where the covariance matrix of $\boldsymbol{a}_t$ is

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 2.0 \end{bmatrix}.$$

Simple calculation shows that the eigenvalues of $\boldsymbol{\phi}_1$ are 0.5 and 0.8, which are less than 1. Therefore, the VAR(1) model in Equation (2.6) is stationary. Note that $\phi_{1,22} = 1.1$, which is greater than 1, but the series is stationary. This simple example demonstrates that the eigenvalues of $\boldsymbol{\phi}_1$ determine the stationarity of $\boldsymbol{z}_t$, not the individual elements of $\boldsymbol{\phi}_1$. □

2.2.4 Invertibility

By definition, a VAR(p) time series is a linear combination of its lagged values. Therefore, the VAR(1) model is always invertible; see the definition of invertibility in Chapter 1.

2.2.5 Moment Equations

Assume that the VAR(1) series in Equation (2.2) is stationary. Taking expectation on both sides of the equation, we have

$$\boldsymbol{\mu} = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{\mu},$$

where $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$. Consequently, we have $(\boldsymbol{I}_k - \boldsymbol{\phi}_1)\boldsymbol{\mu} = \boldsymbol{\phi}_0$, or equivalently, $\boldsymbol{\mu} = (\boldsymbol{I}_k - \boldsymbol{\phi}_1)^{-1}\boldsymbol{\phi}_0$. We can write this equation in a compact form as $\boldsymbol{\mu} = [\boldsymbol{\phi}(1)]^{-1}\boldsymbol{\phi}_0$. Plugging $\boldsymbol{\phi}_0 = (\boldsymbol{I}_k - \boldsymbol{\phi}_1)\boldsymbol{\mu}$ into the VAR(1) model, we obtain the mean-adjusted model

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{\phi}_1 \tilde{\boldsymbol{z}}_{t-1} + \boldsymbol{a}_t, \tag{2.7}$$

where $\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t - \boldsymbol{\mu}$. The covariance matrix of $\boldsymbol{z}_t$ is then

$$\begin{aligned}\boldsymbol{\Gamma}_0 &= E\left(\tilde{\boldsymbol{z}}_t\tilde{\boldsymbol{z}}_t'\right) = E\left(\boldsymbol{\phi}_1\tilde{\boldsymbol{z}}_{t-1}\tilde{\boldsymbol{z}}_{t-1}'\boldsymbol{\phi}_1'\right) + E\left(\boldsymbol{a}_t\boldsymbol{a}_t'\right)\\ &= \boldsymbol{\phi}_1\boldsymbol{\Gamma}_0\boldsymbol{\phi}_1' + \boldsymbol{\Sigma}_a,\end{aligned}$$

where we use the fact that $\boldsymbol{a}_t$ is uncorrelated with $\tilde{\boldsymbol{z}}_{t-1}$. This equation can be rewritten as

$$\text{vec}(\boldsymbol{\Gamma}_0) = (\boldsymbol{\phi}_1 \otimes \boldsymbol{\phi}_1)\text{vec}(\boldsymbol{\Gamma}_0) + \text{vec}(\boldsymbol{\Sigma}_a).$$

Consequently, we have

$$(\boldsymbol{I}_{k^2} - \boldsymbol{\phi}_1 \otimes \boldsymbol{\phi}_1)\text{vec}(\boldsymbol{\Gamma}_0) = \text{vec}(\boldsymbol{\Sigma}_a). \tag{2.8}$$

The prior equation can be used to obtain $\boldsymbol{\Gamma}_0$ from the VAR(1) model for which $\boldsymbol{\phi}_1$ and $\boldsymbol{\Sigma}_a$ are known.

Next, for any positive integer ℓ, postmultiplying $\boldsymbol{z}_{t-\ell}'$ to Equation (2.7) and taking expectation, we have

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\phi}_1\boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0, \tag{2.9}$$

where we use the property that $\boldsymbol{z}_{t-\ell}$ is uncorrelated with $\boldsymbol{a}_t$. The prior equation is referred to as the multivariate Yule–Walker equation for the VAR(1) model. It can be used in two ways. First, in conjunction with Equation (2.8), the equation can be used recursively to obtain the cross-covariance matrices of $\boldsymbol{z}_t$ and, hence, the cross-correlation matrices. Second, it can be used to obtain $\boldsymbol{\phi}_1$ from the cross-covariance matrices. For instance, $\boldsymbol{\phi}_1 = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0^{-1}$.

To demonstrate, consider the stationary VAR(1) model in Equation (2.6). The mean of the series is $\boldsymbol{\mu} = (4, -6)'$. Using Equations (2.8) and (2.9), we obtain

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} 2.29 & 3.51 \\ 3.51 & 8.62 \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 = \begin{bmatrix} 1.51 & 3.29 \\ 2.49 & 7.38 \end{bmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{bmatrix} 1.05 & 2.87 \\ 1.83 & 6.14 \end{bmatrix}.$$

The corresponding cross-correlation matrices of $\boldsymbol{z}_t$ are

$$\boldsymbol{\rho}_0 = \begin{bmatrix} 1.00 & 0.79 \\ 0.79 & 1.0 \end{bmatrix}, \quad \boldsymbol{\rho}_1 = \begin{bmatrix} 0.66 & 0.74 \\ 0.56 & 0.86 \end{bmatrix}, \quad \boldsymbol{\rho}_2 = \begin{bmatrix} 0.46 & 0.65 \\ 0.41 & 0.71 \end{bmatrix}.$$

R Demonstration

```
> phi1=matrix(c(.2,-.6,.3,1.1),2,2) % Input phi_1
> phi1
     [,1] [,2]
[1,]  0.2  0.3
[2,] -0.6  1.1
```

```
> sig=matrix(c(1,0.8,0.8,2),2,2) % Input sigma_a
> sig
     [,1] [,2]
[1,]  1.0  0.8
[2,]  0.8  2.0
> m1=eigen(phi1) % Obtain eigenvalues & vectors
> m1
$values
[1] 0.8 0.5
$vectors
           [,1]       [,2]
[1,] -0.4472136 -0.7071068
[2,] -0.8944272 -0.7071068
> I4=diag(4) ## Create the 4-by-4 identity matrix
> pp=kronecker(phi1,phi1) # Kronecker product
> pp
      [,1]  [,2]  [,3] [,4]
[1,]  0.04  0.06  0.06 0.09
[2,] -0.12  0.22 -0.18 0.33
[3,] -0.12 -0.18  0.22 0.33
[4,]  0.36 -0.66 -0.66 1.21
> c1=c(sig)
> c1
[1] 1.0 0.8 0.8 2.0
> dd=I4-pp
> ddinv=solve(dd)   ## Obtain inverse
> gam0=ddinv%*%matrix(c1,4,1)  # Obtain Gamma_0
> gam0
         [,1]
[1,] 2.288889
[2,] 3.511111
[3,] 3.511111
[4,] 8.622222
> g0=matrix(gam0,2,2)
> g1=phi1%*%g0     ## Obtain Gamma_1
> g1
         [,1]     [,2]
[1,] 1.511111 3.288889
[2,] 2.488889 7.377778
> g2=phi1%*%g1
> g2
         [,1]     [,2]
[1,] 1.048889 2.871111
[2,] 1.831111 6.142222
> D=diag(sqrt(diag(g0))) # To compute cross-correlation
matrices
> D
         [,1]     [,2]
[1,] 1.512907 0.000000
```

```
[2,] 0.000000 2.936362
> Di=solve(D)
> Di%*%g0%*%Di
          [,1]      [,2]
[1,] 1.0000000 0.7903557
[2,] 0.7903557 1.0000000
> Di%*%g1%*%Di
          [,1]      [,2]
[1,] 0.6601942 0.7403332
[2,] 0.5602522 0.8556701
> Di%*%g2%*%Di
          [,1]      [,2]
[1,] 0.4582524 0.6462909
[2,] 0.4121855 0.7123711
```

2.2.6 Implied Models for the Components

In this section, we discuss the implied marginal univariate models for each component z_{it} of a VAR(1) model. Again, for simplicity, we shall employ the mean-adjusted VAR(1) model in Equation (2.7). The AR matrix polynomial of the model is $\boldsymbol{I}_k - \boldsymbol{\phi}_1 B$. This is a $k \times k$ matrix. As such, we can consider its *adjoint* matrix; see Appendix A. For instance, consider the bivariate VAR(1) model of Example 2.1. In this case, we have

$$\boldsymbol{\phi}(B) = \begin{bmatrix} 1 - 0.2B & -0.3B \\ 0.6B & 1 - 1.1B \end{bmatrix}.$$

The adjoint matrix of $\boldsymbol{\phi}(B)$ is

$$\text{adj}[\boldsymbol{\phi}(B)] = \begin{bmatrix} 1 - 1.1B & 0.3B \\ -0.6B & 1 - 0.2B \end{bmatrix}.$$

The product of these two matrices gives

$$\text{adj}[\boldsymbol{\phi}(B)]\boldsymbol{\phi}(B) = |\boldsymbol{\phi}(B)|\boldsymbol{I}_2,$$

where $|\boldsymbol{\phi}(B)| = (1 - 0.2B)(1 - 1.1B) + 0.18B^2 = 1 - 1.3B + 0.4B^2$. The key feature of the product matrix is that it is a diagonal matrix with the determinant being its diagonal elements. This property continues to hold for a general VAR(1) model. Consequently, if we premultiply the VAR(1) model in Equation (2.7) by the adjoint matrix of $\boldsymbol{\phi}(B)$, then we have

$$|\boldsymbol{\phi}(B)|\boldsymbol{z}_t = \text{adj}[\boldsymbol{\phi}(B)]\boldsymbol{a}_t. \tag{2.10}$$

For a k-dimensional VAR(1) model, $|\boldsymbol{\phi}(B)|$ is a polynomial of degree k, and elements of adj$[\boldsymbol{\phi}(B)]$ are polynomials of degree $k-1$, because they are the determinant of a $(k-1) \times (k-1)$ matrix polynomial of order 1. Next, we use the result that any nonzero linear combination of $\{\boldsymbol{a}_t, \boldsymbol{a}_{t-1}, \ldots, \boldsymbol{a}_{t-k+1}\}$ is a univariate MA$(k-1)$ series. Consequently, Equation (2.10) shows that each component z_{it} follows a univariate ARMA$(k, k-1)$ model. The orders k and $k-1$ are the maximum orders. The actual ARMA order for the individual z_{it} can be smaller.

To demonstrate, consider again the bivariate VAR(1) model in Example 2.1. For this particular instance, Equation (2.10) becomes

$$(1 - 1.3B + 0.4B^2)\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 1-1.1B & 0.3B \\ -0.6B & 1-0.2B \end{bmatrix}\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The marginal model for z_{1t} is then

$$\begin{aligned}\left(1 - 1.3B + 0.4B^2\right) z_{1t} &= (1-1.1B)a_{1t} + 0.3Ba_{2t} \\ &= a_{1t} - 1.1a_{1,t-1} + 0.3a_{2,t-1},\end{aligned}$$

which is an ARMA(2,1) model because the MA part has serial correlation at lag 1 only and can be rewritten as $e_t - \theta e_{t-1}$. The parameters θ and Var(e_t) can be obtained from those of the VAR(1) model. The same result holds for z_{2t}.

2.2.7 Moving-Average Representation

In some applications, for example, computing variances of forecast errors, it is more convenient to consider the moving-average (MA) representation than the AR representation of a VAR model. For the VAR(1) model, one can easily obtain the MA representation using Equation (2.7). Specifically, by repeated substitutions, we have

$$\begin{aligned}\tilde{\boldsymbol{z}}_t &= \boldsymbol{a}_t + \boldsymbol{\phi}_1 \tilde{\boldsymbol{z}}_{t-1} = \boldsymbol{a}_t + \boldsymbol{\phi}_1(\boldsymbol{a}_{t-1} + \boldsymbol{\phi}_1 \tilde{\boldsymbol{z}}_{t-2}) \\ &= \boldsymbol{a}_t + \boldsymbol{\phi}_1 \boldsymbol{a}_{t-2} + \boldsymbol{\phi}_1^2(\boldsymbol{a}_{t-2} + \boldsymbol{\phi} \tilde{\boldsymbol{z}}_{t-3}) \\ &= \cdots \\ &= \boldsymbol{a}_t + \boldsymbol{\phi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\phi}_1^2 \boldsymbol{a}_{t-2} + \boldsymbol{\phi}_1^3 \boldsymbol{a}_{t-3} + \cdots\end{aligned}$$

Consequently, we have

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{a}_t + \boldsymbol{\psi}_1 \boldsymbol{a}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{a}_{t-2} + \cdots, \tag{2.11}$$

where $\boldsymbol{\psi}_i = \boldsymbol{\phi}_1^i$ for $i \geq 0$. For a stationary VAR(1) model, all eigenvalues of $\boldsymbol{\phi}_1$ are less than 1 in absolute value, thus $\boldsymbol{\psi}_i \to \mathbf{0}$ as $i \to \infty$. This implies that, as expected, the effect of the remote innovation $\boldsymbol{a}_{t-i}$ on $\boldsymbol{z}_t$ is diminishing as i increases. Eventually the effect vanishes, confirming that the initial condition of a stationary VAR(1) model has no impact on the series as time increases.

2.3 VAR(2) MODELS

A VAR(2) model assumes the form

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{\phi}_2 \boldsymbol{z}_{t-2} + \boldsymbol{a}_t. \tag{2.12}$$

It says that each component z_{it} depends on the lagged values $\boldsymbol{z}_{t-1}$ and $\boldsymbol{z}_{t-2}$. The AR coefficient can be interpreted in a way similar to that of a VAR(1) model. For instance, the $(1,2)$th element of $\boldsymbol{\phi}_1$, that is, $\phi_{1,12}$, denotes the linear dependence of z_{1t} on $z_{2,t-1}$ in the presence of other lagged values in $\boldsymbol{z}_{t-1}$ and $\boldsymbol{z}_{t-2}$.

The structure and relationship to Granger causality of a VAR(2) model can be generalized straightforwardly from that of the VAR(1) model. For instance, consider a bivariate VAR(2) model. If both $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are diagonal matrices, then z_{1t} and z_{2t} follow a univariate AR(2) model and can be handled accordingly. If $\phi_{1,12} = \phi_{2,12} = 0$, but at least one of $\phi_{1,21}$ and $\phi_{2,21}$ is not zero, then we have a unidirectional relationship from z_{1t} to z_{2t}, because in this case z_{1t} does not depend on any past value of z_{2t}, but z_{2t} depends on some past value of z_{1t}. We can also derive a transfer function model for z_{2t} with z_{1t} as the input variable. In general, for the existence of a unidirectional relationship from component z_{it} to component z_{jt} of a VAR model, the parameters at the (i,j)th position of each coefficient matrix $\boldsymbol{\phi}_v$ must be 0 simultaneously.

2.3.1 Stationarity Condition

To study the stationarity condition for the VAR(2) model in Equation (2.12), we can make use of the result obtained for VAR(1) models. To see this, we consider an expanded $(2k)$-dimensional time series $\boldsymbol{Z}_t = (\boldsymbol{z}_t', \boldsymbol{z}_{t-1}')'$. Using Equation (2.12) and the identity $\boldsymbol{z}_{t-1} = \boldsymbol{z}_{t-1}$, we obtain a model for $\boldsymbol{Z}_t$. Specifically, we have

$$\begin{bmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_0 \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 \\ \boldsymbol{I}_k & \boldsymbol{0}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_{t-1} \\ \boldsymbol{z}_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_t \\ \boldsymbol{0} \end{bmatrix}, \tag{2.13}$$

where $\boldsymbol{0}$ is a k-dimensional vector of zero and $\boldsymbol{0}_k$ is a $k \times k$ zero matrix. Consequently, the expanded time series $\boldsymbol{Z}_t$ follows a VAR(1) model, say,

$$\boldsymbol{Z}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{Z}_{t-1} + \boldsymbol{b}_t, \tag{2.14}$$

where the vector $\boldsymbol{\Phi}_0$, the AR coefficient matrix $\boldsymbol{\Phi}_1$, and $\boldsymbol{b}_t$ are defined in Equation (2.13). Using the result of VAR(1) models derived in the previous section, the necessary and sufficient condition for $\boldsymbol{Z}_t$ to be stationary is that all solutions of the determinant equation $|\boldsymbol{I}_{2k} - \boldsymbol{\Phi}_1 B| = 0$ must be greater than 1 in absolute value.

Next, the determinant can be rewritten as

$$\begin{aligned}|\boldsymbol{I}_{2k} - \boldsymbol{\Phi}_1 B| &= \begin{vmatrix} \boldsymbol{I}_k - \boldsymbol{\phi}_1 B & -\boldsymbol{\phi}_2 B \\ -\boldsymbol{I}_k B & \boldsymbol{I}_k \end{vmatrix} \\ &= \begin{vmatrix} \boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2 & -\boldsymbol{\phi}_2 B \\ \boldsymbol{0}_k & \boldsymbol{I}_k \end{vmatrix} \\ &= |\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2| \\ &= |\boldsymbol{\phi}(B)|,\end{aligned}$$

where the second equality is obtained by multiplying the second column block matrix by B and adding the result to the first column block. Such an operation is valid in calculating the determinant of a matrix. In summary, the necessary and sufficient condition for the stationarity of $\boldsymbol{Z}_t$ (and, hence $\boldsymbol{z}_t$) is that all solutions of the determinant equation $|\boldsymbol{\phi}(B)| = 0$ are greater than 1 in absolute value.

2.3.2 Moment Equations

Assuming that the VAR(2) model in Equation (2.12) is stationary, we derive the moment equations for $\boldsymbol{z}_t$. First, taking expectation of Equation (2.12), we get

$$\boldsymbol{\mu} = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{\mu} + \boldsymbol{\phi}_2 \boldsymbol{\mu}.$$

Therefore, we have $(\boldsymbol{I}_k - \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)\boldsymbol{\mu} = \boldsymbol{\phi}_0$. The mean of the series is then $\boldsymbol{\mu} = [\boldsymbol{\phi}(1)]^{-1}\boldsymbol{\phi}_0$. Next, plugging $\boldsymbol{\phi}_0$ into Equation (2.12), we obtain the mean-adjusted model as

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{\phi}_1 \tilde{\boldsymbol{z}}_{t-1} + \boldsymbol{\phi}_2 \tilde{\boldsymbol{z}}_{t-2} + \boldsymbol{a}_t. \tag{2.15}$$

Postmultiplying Equation (2.15) by $\boldsymbol{a}_t'$ and using the zero correlation between $\boldsymbol{a}_t$ and the past values of $\boldsymbol{z}_t$, we obtain

$$E(\tilde{\boldsymbol{z}}_t \boldsymbol{a}_t') = E\left(\boldsymbol{a}_t \boldsymbol{a}_t'\right) = \boldsymbol{\Sigma}_a.$$

Postmultiplying Equation (2.15) by $\tilde{\boldsymbol{z}}_{t-\ell}'$ and taking expectation, we have

$$\begin{aligned}\boldsymbol{\Gamma}_0 &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{-1} + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}_{-2} + \boldsymbol{\Sigma}_a, \\ \boldsymbol{\Gamma}_\ell &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{\ell-1} + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}_{\ell-2}, \quad \text{if } \ell > 0.\end{aligned} \tag{2.16}$$

Using $\ell = 0, 1, 2$, we have a set of matrix equations that relate $\{\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2\}$ to $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\Sigma}_a\}$. In particular, using $\ell = 1$ and 2, we have

$$[\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2] = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2] \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Gamma}_1' & \boldsymbol{\Gamma}_0 \end{bmatrix}, \tag{2.17}$$

where we use $\boldsymbol{\Gamma}_{-1} = \boldsymbol{\Gamma}'_1$. This system of matrix equations is called the multivariate Yule–Walker equation for the VAR(2) model in Equation (2.12). For a stationary series $\boldsymbol{z}_t$, the $2k \times 2k$ matrix on the right side of Equation (2.17) is invertible so that we have

$$[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2] = [\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2] \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Gamma}'_1 & \boldsymbol{\Gamma}_0 \end{bmatrix}^{-1}.$$

This equation can be used to obtain the AR coefficients from the cross-covariance matrices.

In practice, one may use the expanded series $\boldsymbol{Z}_t$ in Equation (2.14) and the result of VAR(1) model in the previous section to obtain the cross-covariance matrices $\boldsymbol{\Gamma}_\ell$ of $\boldsymbol{z}_t$ from $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, and $\boldsymbol{\Sigma}_a$. Specifically, for Equation (2.14), we have

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 \\ \boldsymbol{I}_k & \boldsymbol{0}_k \end{bmatrix}, \quad \boldsymbol{\Sigma}_b = \begin{bmatrix} \boldsymbol{\Sigma}_a & \boldsymbol{0}_k \\ \boldsymbol{0}_k & \boldsymbol{0}_k \end{bmatrix}.$$

In addition, we have

$$\mathrm{Cov}(\boldsymbol{Z}_t) = \boldsymbol{\Gamma}^*_0 = \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Gamma}'_1 & \boldsymbol{\Gamma}_0 \end{bmatrix}.$$

For the expanded series $\boldsymbol{Z}_t$, Equation (2.8) becomes

$$[\boldsymbol{I}_{(2k)^2} - \boldsymbol{\Phi}_1 \otimes \boldsymbol{\Phi}_1]\mathrm{vec}(\boldsymbol{\Gamma}^*_0) = \mathrm{vec}(\boldsymbol{\Sigma}_b). \tag{2.18}$$

Consequently, we can obtain $\boldsymbol{\Gamma}^*_0$, which contains $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1$. Higher-order cross-covariances $\boldsymbol{\Gamma}_\ell$ are then obtained recursively using the moment equation in Equation (2.16).

Example 2.2 Consider a three-dimensional VAR(2) model in Equation (2.12) with parameters $\boldsymbol{\phi}_0 = \boldsymbol{0}$ and

$$\boldsymbol{\phi}_1 = \begin{bmatrix} 0.47 & 0.21 & 0 \\ 0.35 & 0.34 & 0.47 \\ 0.47 & 0.23 & 0.23 \end{bmatrix}, \quad \boldsymbol{\phi}_2 = \begin{bmatrix} 0 & 0 & 0 \\ -0.19 & -0.18 & 0 \\ -0.30 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} 0.285 & 0.026 & 0.069 \\ 0.026 & 0.287 & 0.137 \\ 0.069 & 0.137 & 0.357 \end{bmatrix}.$$

This VAR(2) model was employed later in analyzing the quarterly growth rates of gross domestic product of United Kingdom, Canada, and United States from 1980.II to 2011.II. For this particular VAR(2) model, we can calculate its cross-covariance matrices via Equation (2.18). They are

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} 0.46 & 0.22 & 0.24 \\ 0.22 & 0.61 & 0.38 \\ 0.24 & 0.38 & 0.56 \end{bmatrix}, \quad \boldsymbol{\Gamma}_1 = \begin{bmatrix} 0.26 & 0.23 & 0.19 \\ 0.25 & 0.35 & 0.38 \\ 0.24 & 0.25 & 0.25 \end{bmatrix},$$

$$\boldsymbol{\Gamma}_2 = \begin{bmatrix} 0.18 & 0.18 & 0.17 \\ 0.16 & 0.17 & 0.20 \\ 0.10 & 0.18 & 0.16 \end{bmatrix}. \qquad \square$$

2.3.3 Implied Marginal Component Models

For the VAR(2) model in Equation (2.12), we can use the same technique as that for VAR(1) models to obtain the univariate ARMA model for the component series z_{it}. The general solution is that z_{it} follows an ARMA$(2k, 2(k-1))$ model. Again, the order $(2k, 2(k-1))$ is the maximum order for each component z_{it}.

2.3.4 Moving-Average Representation

The MA representation of a VAR(2) model can be obtained in several ways. One can use repeated substitutions similar to that of the VAR(1) model. Here, we adopt an alternative approach. Consider the mean-adjusted VAR(2) model in Equation (2.15). The model can be written as

$$\left(\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2\right) \tilde{\boldsymbol{z}}_t = \boldsymbol{a}_t,$$

which is equivalent to

$$\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2\right)^{-1} \boldsymbol{a}_t.$$

On the other hand, the MA representation of the series is $\tilde{\boldsymbol{z}}_t = \boldsymbol{\psi}(B)\boldsymbol{a}_t$. Consequently, we have $(\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2)^{-1} = \boldsymbol{\psi}(B)$, which is

$$\boldsymbol{I}_k = \left(\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \boldsymbol{\phi}_2 B^2\right)\left(\boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 B + \boldsymbol{\psi}_2 B^2 + \boldsymbol{\psi}_3 B^3 + \cdots\right), \tag{2.19}$$

where $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$. Since the left-hand side of Equation (2.19) is a constant matrix, all coefficient matrices of B^i for $i > 0$ on the right-hand side of the equation must be 0. Therefore, we obtain

$$\begin{aligned} \mathbf{0} &= \boldsymbol{\psi}_1 - \boldsymbol{\phi}_1 \boldsymbol{\psi}_0, && \text{(coefficient of } B\text{)} \\ \mathbf{0} &= \boldsymbol{\psi}_2 - \boldsymbol{\phi}_1 \boldsymbol{\psi}_1 - \boldsymbol{\phi}_2 \boldsymbol{\psi}_0, && \text{(coefficient of } B^2\text{)} \\ \mathbf{0} &= \boldsymbol{\psi}_v - \boldsymbol{\phi}_1 \boldsymbol{\psi}_{v-1} - \boldsymbol{\phi}_2 \boldsymbol{\psi}_{v-2}, && \text{for } v \geq 3. \end{aligned}$$

Consequently, we have

$$\begin{aligned} \boldsymbol{\psi}_1 &= \boldsymbol{\phi}_1 \\ \boldsymbol{\psi}_v &= \boldsymbol{\phi}_1 \boldsymbol{\psi}_{v-1} + \boldsymbol{\phi}_2 \boldsymbol{\psi}_{v-2}, \quad \text{for } v \geq 2, \end{aligned} \tag{2.20}$$

where $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$. In other words, we can compute the coefficient matrix $\boldsymbol{\psi}_i$ recursively starting with $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$ and $\boldsymbol{\psi}_1 = \boldsymbol{\phi}_1$.

2.4 VAR(p) MODELS

Consider next the general k-dimensional VAR(p) model, which assumes the form

$$\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t, \tag{2.21}$$

where $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$ with $\boldsymbol{\phi}_p \neq \boldsymbol{0}$. The results for VAR(1) and VAR(2) models discussed in the previous sections continue to hold for the VAR(p) model. For example, a VAR(p) model is invertible and its structure is sufficiently flexible to encompass the transfer function model. In this section, we consider the generalizations of other properties of simple VAR models to the general VAR(p) model.

Assume that the series in Equation (2.21) is stationary. Taking the expectation, we have

$$(\boldsymbol{I}_k - \boldsymbol{\phi}_1 - \cdots - \boldsymbol{\phi}_p)\boldsymbol{\mu} = [\boldsymbol{\phi}(1)]\boldsymbol{\mu} = \boldsymbol{\phi}_0,$$

where, as before, $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$. Therefore, $\boldsymbol{\mu} = [\boldsymbol{\phi}(1)]^{-1}\boldsymbol{\phi}_0$, and the model can be rewritten as

$$\boldsymbol{\phi}(B)\tilde{\boldsymbol{z}}_t = \boldsymbol{a}_t. \tag{2.22}$$

We use this mean-adjusted representation to derive other properties of a stationary VAR(p) model. The mean has no impact on those properties and can be assumed to be 0.

2.4.1 A VAR(1) Representation

Similar to the VAR(2) model, we can express a VAR(p) model in a VAR(1) form by using an expanded series. Define $\boldsymbol{Z}_t = (\tilde{\boldsymbol{z}}_t', \tilde{\boldsymbol{z}}_{t-1}', \ldots, \tilde{\boldsymbol{z}}_{t-p+1}')'$, which is a pk-dimensional time series. The VAR(p) model in Equation (2.22) can be written as

$$\boldsymbol{Z}_t = \boldsymbol{\Phi}\boldsymbol{Z}_{t-1} + \boldsymbol{b}_t, \tag{2.23}$$

where $\boldsymbol{b}_t = (\boldsymbol{a}_t', \boldsymbol{0}')'$ with $\boldsymbol{0}$ being a $k(p-1)$-dimensional zero vector, and

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_{p-1} & \boldsymbol{\phi}_p \\ \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} \end{bmatrix},$$

where it is understood that $\boldsymbol{I}$ and $\boldsymbol{0}$ are the $k \times k$ identity and zero matrix, respectively. The matrix $\boldsymbol{\Phi}$ is called the companion matrix of the matrix polynomial $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \cdots - \boldsymbol{\phi}_p B^p$. The covariance matrix of $\boldsymbol{b}_t$ has a special structure; all of its elements are zero except those in the upper-left corner that is $\boldsymbol{\Sigma}_a$.

2.4.2 Stationarity Condition

The sufficient and necessary condition for the weak stationarity of the VAR(p) series $\boldsymbol{z}_t$ can be easily obtained using the VAR(1) representation in Equation (2.23). Since $\boldsymbol{Z}_t$ follows a VAR(1) model, the condition for its stationarity is that all solutions of the determinant equation $|\boldsymbol{I}_{kp} - \boldsymbol{\Phi} B| = 0$ must be greater than 1 in absolute value. Sometimes, we say that the solutions must be greater than 1 in modulus or they are outside the unit circle. By Lemma 2.1, we have $|\boldsymbol{I}_{kp} - \boldsymbol{\Phi} B| = |\boldsymbol{\phi}(B)|$ for the VAR(p) time series. Therefore, the necessary and sufficient condition for the weak stationarity of the VAR(p) series is that all solutions of the determinant equation $|\boldsymbol{\phi}(B)| = 0$ must be greater than 1 in modulus.

Lemma 2.1 For the $k \times k$ matrix polynomial $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$, $|\boldsymbol{I}_{kp} - \boldsymbol{\Phi} B| = |\boldsymbol{I}_k - \boldsymbol{\phi}_1 B - \cdots - \boldsymbol{\phi}_p B^p|$ holds, where $\boldsymbol{\Phi}$ is defined in Equation (2.23).

A proof of Lemma 2.1 is given in Section 2.12.

2.4.3 Moment Equations

Postmultiplying Equation (2.22) by $\tilde{\boldsymbol{z}}_{t-\ell}$ and taking expectation, we obtain

$$\boldsymbol{\Gamma}_\ell - \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{\ell-1} - \cdots - \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{\ell-p} = \begin{cases} \boldsymbol{\Sigma}_a & \text{if } \ell = 0, \\ \boldsymbol{0} & \text{if } \ell > 0. \end{cases} \tag{2.24}$$

Consider jointly the matrix equations for $\ell = 1, \ldots, p$. We have a system of matrix equations

$$[\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_p] = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_p] \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{p-1} \\ \boldsymbol{\Gamma}_1' & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_{p-1}' & \boldsymbol{\Gamma}_{p-2}' & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix}, \tag{2.25}$$

where we use $\boldsymbol{\Gamma}_{-\ell} = \boldsymbol{\Gamma}'_\ell$. This system of matrix equations is called the multivariate Yule–Walker equation for VAR(p) models. It can be used to obtain the AR coefficient matrices $\boldsymbol{\phi}_j$ from the cross-covariance matrices $\boldsymbol{\Gamma}_\ell$ for $\ell = 0, \ldots, p$. For a stationary VAR(p) model, the square matrix of Equation (2.25) is nonsingular. On the other hand, to obtain the cross-covariance matrices, hence the cross-correlation matrices, of a stationary VAR(p) model, it is convenient to use the expanded VAR(1) representation in Equation (2.23). For the expanded (kp)-dimensional series $\boldsymbol{Z}_t$, we have

$$\text{Cov}(\boldsymbol{Z}_t) = \boldsymbol{\Gamma}_0^* = \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{p-1} \\ \boldsymbol{\Gamma}'_1 & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}'_{p-1} & \boldsymbol{\Gamma}'_{p-2} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix}, \tag{2.26}$$

which is precisely the square matrix in Equation (2.25). Consequently, similar to the VAR(2) case, we can apply Equation (2.8) to obtain

$$[\boldsymbol{I}_{(kp)^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}]\text{vec}(\boldsymbol{\Gamma}_0^*) = \text{vec}(\boldsymbol{\Sigma}_b).$$

Thus, given the AR coefficient matrices $\boldsymbol{\phi}_i$ and the covariance matrix $\boldsymbol{\Sigma}_a$, we can obtain $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}_b$. The prior equation can be then used to obtain $\boldsymbol{\Gamma}_0^*$, which contains $\boldsymbol{\Gamma}_\ell$ for $\ell = 0, \ldots, p-1$. Other higher-order cross-covariance matrices, $\boldsymbol{\Gamma}_\ell$, can use computed recursively via the moment equation in Equation (2.24).

2.4.4 Implied Component Models

Using the same technique as those used in the VAR(2) model, we see that the component series z_{it} of a VAR(p) model follows a univariate ARMA($kp, (k-1)p$) model. This ARMA order $(kp, (k-1)p)$ is high for a large k or p, but it denotes the maximum order allowed. The actual order of the marginal ARMA model for z_{it} can be substantially lower. Except for seasonal time series, our experience of analyzing univariate time series shows that the order used for real-world series is typically low. Since a VAR(p) model encompasses a wide-rage of component models, one would expect that the VAR orders used for most real-world multivariate time series will not be high.

2.4.5 Moving-Average Representation

Using the same techniques as that of the VAR(2) model, we can obtain the MA representation for a VAR(p) model via a recursive method. The coefficient matrices of the MA representation are

$$\boldsymbol{\psi}_i = \sum_{j=1}^{\min(i,p)} \boldsymbol{\phi}_j \boldsymbol{\psi}_{i-j}, \quad i = 1, 2, \cdots, \tag{2.27}$$

where $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$. The matrices $\boldsymbol{\psi}_i$ are referred to as the ψ-weights of the VAR(p) model. We discuss the meanings of these $\boldsymbol{\psi}_i$ matrices later. Here, it suffices to say that, based on the MA representation, we can easily show the following result.

Lemma 2.2 For a VAR(p) model in Equation (2.21) with $\boldsymbol{a}_t$ being a serially uncorrelated innovation process with mean zero and positive-definite covariance $\boldsymbol{\Sigma}_a$, $\text{Cov}(\boldsymbol{z}_t, \boldsymbol{a}_{t-j}) = \boldsymbol{\psi}_j \boldsymbol{\Sigma}_a$ for $j \geq 0$, where $\boldsymbol{\psi}_j$ denotes the ψ-weight matrix.

2.5 ESTIMATION

A VAR(p) model can be estimated by the LS or ML method or Bayesian method. For the LS methods, we show that the GLS and the OLS methods produce the same estimates; see Zellner (1962). Under the multivariate normality assumption, that is, $\boldsymbol{a}_t$ follows a k-dimensional normal distribution, the ML estimates of a VAR(p) model are asymptotically equivalent to the LS estimates. We also briefly discuss the Bayesian estimation of a VAR(p) model.

Suppose that the sample $\{\boldsymbol{z}_t | t = 1, \ldots, T\}$ is available from a VAR(p) model. The parameters of interest are $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p\}$ and $\boldsymbol{\Sigma}_a$. In what follows, we discuss various methods for estimating these parameters and properties of the estimates.

2.5.1 Least-Squares Methods

For LS estimation, the available data enable us to consider

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \phi_1 \boldsymbol{z}_{t-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{z}_{t-p} + \boldsymbol{a}_t, \quad t = p+1, \cdots, T,$$

where the covariance matrix of $\boldsymbol{a}_t$ is $\boldsymbol{\Sigma}_a$. Here, we have $T-p$ data points for effective estimation. To facilitate the estimation, we rewrite the VAR(p) model as

$$\boldsymbol{z}_t' = \boldsymbol{x}_t' \boldsymbol{\beta} + \boldsymbol{a}_t',$$

where $\boldsymbol{x}_t = (1, \boldsymbol{z}_{t-1}', \cdots, \boldsymbol{z}_{t-p}')'$ is a $(kp + 1)$-dimensional vector and $\boldsymbol{\beta}' = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p]$ is a $k \times (kp + 1)$ matrix. With this new format, we can write the data as

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}, \tag{2.28}$$

where $\boldsymbol{Z}$ is a $(T-p) \times k$ matrix with ith row being $\boldsymbol{z}_{p+i}'$, $\boldsymbol{X}$ is a $(T-p) \times (kp+1)$ design matrix with ith row being $\boldsymbol{x}_{p+i}'$, and $\boldsymbol{A}$ is a $(T-p) \times k$ matrix with ith row being $\boldsymbol{a}'_{p+i}$. The matrix representation in Equation (2.28) is particularly convenient for the VAR(p) model. For example, column j of $\boldsymbol{\beta}$ contains parameters associated with z_{jt}. Taking the vectorization of Equation (2.28), and using the properties of Kronecker product given in Appendix A, we obtain

$$\text{vec}(\boldsymbol{Z}) = (\boldsymbol{I}_k \otimes \boldsymbol{X})\text{vec}(\boldsymbol{\beta}) + \text{vec}(\boldsymbol{A}). \tag{2.29}$$

Note that the covariance matrix of vec($\boldsymbol{A}$) is $\boldsymbol{\Sigma}_a \otimes \boldsymbol{I}_{T-p}$.

2.5.1.1 Generalized Least Squares Estimate

The GLS estimate of $\boldsymbol{\beta}$ is obtained by minimizing

$$\begin{aligned} S(\boldsymbol{\beta}) &= [\text{vec}(\boldsymbol{A})]'(\boldsymbol{\Sigma}_a \otimes \boldsymbol{I}_{T-p})^{-1}\text{vec}(\boldsymbol{A}) \\ &= [\text{vec}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta})]' \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{T-p}\right) \text{vec}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}) & (2.30) \\ &= tr[(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}_a^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta})']. & (2.31) \end{aligned}$$

The last equality holds because $\boldsymbol{\Sigma}_a$ is a symmetric matrix and we use $tr(\boldsymbol{DBC}) = \text{vec}(\boldsymbol{C}')'(\boldsymbol{B}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{D})$. From Equation (2.30), we have

$$\begin{aligned} S(\boldsymbol{\beta}) &= [\text{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\text{vec}(\boldsymbol{\beta})]' \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{T-p}\right) [\text{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\text{vec}(\boldsymbol{\beta})] \\ &= \left[\text{vec}(\boldsymbol{Z})' - \text{vec}(\boldsymbol{\beta})'(\boldsymbol{I}_k \otimes \boldsymbol{X}')\right] \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{T-p}\right) \\ &\quad \times [\text{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\text{vec}(\boldsymbol{\beta})] \\ &= \text{vec}(\boldsymbol{Z})' \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{T-p}\right) \text{vec}(\boldsymbol{Z}) - 2\text{vec}(\boldsymbol{\beta})' \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\right) \text{vec}(\boldsymbol{Z}) \\ &\quad + \text{vec}(\boldsymbol{\beta})' \left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right) \text{vec}(\boldsymbol{\beta}). \end{aligned} \tag{2.32}$$

Taking partial derivatives of $S(\boldsymbol{\beta})$ with respect to vec($\boldsymbol{\beta}$), we obtain

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \text{vec}(\boldsymbol{\beta})} = -2\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\right) \text{vec}(\boldsymbol{Z}) + 2(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\text{vec}(\boldsymbol{\beta}). \tag{2.33}$$

Equating to 0 gives the normal equations

$$(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\text{vec}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{Z}).$$

Consequently, the GLS estimate of a VAR(p) model is

$$\begin{aligned} \text{vec}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{Z}) \\ &= [\boldsymbol{\Sigma}_a \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}](\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{Z}) \\ &= [\boldsymbol{I}_k \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\text{vec}(\boldsymbol{Z}) \\ &= \text{vec}[(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z})], \end{aligned} \tag{2.34}$$

where the last equality holds because $\text{vec}(\boldsymbol{DB}) = (\boldsymbol{I} \otimes \boldsymbol{D})\text{vec}(\boldsymbol{B})$. In other words, we obtain

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}) = \left[\sum_{t=p+1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1} \sum_{t=p+1}^{T} \boldsymbol{x}_t\boldsymbol{z}_t', \tag{2.35}$$

which interestingly does not depend on $\boldsymbol{\Sigma}_a$.

Remark: The result in Equation (2.35) shows that one can obtain the GLS estimate of a VAR(p) model equation-by-equation. That is, one can consider the k multiple linear regressions of z_{it} on $\boldsymbol{x}_t$ separately, where $i = 1, \ldots, k$. This estimation method is convenient when one considers parameter constraints in a VAR(p) model. □

2.5.1.2 Ordinary Least-Squares Estimate

Readers may notice that the GLS estimate of VAR(p) model in Equation (2.35) is identical to that of the OLS estimate of the multivariate multiple linear regression in Equation (2.28). Replacing $\boldsymbol{\Sigma}_a$ in Equation (2.31) by $\boldsymbol{I}_k$, we have the objective function of the OLS estimation

$$S_o(\boldsymbol{\beta}) = tr[(\boldsymbol{Z} - \boldsymbol{X\beta})(\boldsymbol{Z} - \boldsymbol{X\beta})']. \tag{2.36}$$

The derivations discussed earlier continue to hold step-by-step with $\boldsymbol{\Sigma}_a$ replaced by $\boldsymbol{I}_k$. One thus obtains the same estimate given in Equation (2.35) for $\boldsymbol{\beta}$. The fact that the GLS estimate is the same as the OLS estimate for a VAR(p) model was first shown in Zellner (1962). In what follows, we refer to the estimate in Equation (2.35) simply as the LS estimate.

The residual of the LS estimate is

$$\hat{\boldsymbol{a}}_t = \boldsymbol{z}_t - \hat{\boldsymbol{\phi}}_0 - \sum_{i=1}^{p} \hat{\boldsymbol{\phi}}_i \boldsymbol{z}_{t-i}, \quad t = p+1, \cdots, T$$

and let $\hat{\boldsymbol{A}}$ be the residual matrix, that is, $\hat{\boldsymbol{A}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = [\boldsymbol{I}_{T-p} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y}$. The LS estimate of the innovational covariance matrix $\boldsymbol{\Sigma}_a$ is

$$\tilde{\boldsymbol{\Sigma}}_a = \frac{1}{T-(k+1)p-1} \sum_{t=p+1}^{T} \hat{\boldsymbol{a}}_t\hat{\boldsymbol{a}}_t' = \frac{1}{T-(k+1)p-1} \hat{\boldsymbol{A}}'\hat{\boldsymbol{A}},$$

where the denominator is $[T-p-(kp+1)]$, which is the effective sample size less the number of parameters in the equation for each component z_{it}. By Equation (2.28), we see that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{A}. \tag{2.37}$$

Since $E(\boldsymbol{A}) = \boldsymbol{0}$, we see that the LS estimate is an unbiased estimator. The LS estimate of a VAR(p) model has the following properties.

Theorem 2.1 For the stationary VAR(p) model in Equation (2.21), assume that $\boldsymbol{a}_t$ are independent and identically distributed with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_a$. Then, (i) $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is defined in Equation (2.28), (ii) $E(\tilde{\boldsymbol{\Sigma}}_a) = \boldsymbol{\Sigma}_a$, (iii) the residual $\hat{\boldsymbol{A}}$ and the LS estimate $\hat{\boldsymbol{\beta}}$ are uncorrelated, and (iv) the covariance of the parameter estimates is

$$\text{Cov}[\text{vec}(\hat{\boldsymbol{\beta}})] = \tilde{\boldsymbol{\Sigma}}_a \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1} = \tilde{\boldsymbol{\Sigma}}_a \otimes \left(\sum_{t=p+1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' \right)^{-1}.$$

Theorem 2.1 follows the LS theory for multivariate multiple linear regression. A proof can be found in Lütkepohl (2005) or in Johnson and Wichern (2007, Chapter 7).

2.5.2 Maximum Likelihood Estimate

Assume further that $\boldsymbol{a}_t$ of the VAR(p) model follows a multivariate normal distribution. Let $\boldsymbol{z}_{h:q}$ denote the observations from $t = h$ to $t = q$ (inclusive). The conditional likelihood function of the data can be written as

$$\begin{aligned}
L(\boldsymbol{z}_{(p+1):T} | \boldsymbol{z}_{1:p}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a) &= \prod_{t=p+1}^{T} p(\boldsymbol{z}_t | \boldsymbol{z}_{1:(t-1)}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a) \\
&= \prod_{t=p+1}^{T} p(\boldsymbol{a}_t | \boldsymbol{z}_{1:(t-1)}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_a) \\
&= \prod_{t=p+1}^{T} p(\boldsymbol{a}_t | \boldsymbol{\beta}, \boldsymbol{\Sigma}_a) \\
&= \prod_{t=p+1}^{T} \frac{1}{(2\pi)^{k/2} |\boldsymbol{\Sigma}_a|^{1/2}} \exp\left[\frac{-1}{2} \boldsymbol{a}_t' \boldsymbol{\Sigma}_a^{-1} \boldsymbol{a}_t \right] \\
&\propto |\boldsymbol{\Sigma}_a|^{-(T-p)/2} \exp\left[\frac{-1}{2} \sum_{t=p+1}^{T} tr(\boldsymbol{a}_t' \boldsymbol{\Sigma}_a^{-1} \boldsymbol{a}_t) \right].
\end{aligned}$$

The log-likelihood function then becomes

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a) = c - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_a|) - \frac{1}{2}\sum_{t=p+1}^{T} tr\left(\boldsymbol{a}_t'\boldsymbol{\Sigma}_a^{-1}\boldsymbol{a}_t\right)$$
$$= c - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_a|) - \frac{1}{2}tr\left(\boldsymbol{\Sigma}_a^{-1}\sum_{t=p+1}^{T}\boldsymbol{a}_t\boldsymbol{a}_t'\right),$$

where c is a constant, and we use the properties that $tr(\boldsymbol{CD}) = tr(\boldsymbol{DC})$ and $tr(\boldsymbol{C} + \boldsymbol{D}) = tr(\boldsymbol{C}) + tr(\boldsymbol{D})$. Noting that $\sum_{t=p+1}^{T}\boldsymbol{a}_t\boldsymbol{a}_t' = \boldsymbol{A}'\boldsymbol{A}$, where $\boldsymbol{A} = \boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}$ is the error matrix in Equation (2.28), we can rewrite the log-likelihood function as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a) = c - \frac{T-p}{2}\log(|\boldsymbol{\Sigma}_a|) - \frac{1}{2}S(\boldsymbol{\beta}), \tag{2.38}$$

where $S(\boldsymbol{\beta})$ is given in Equation (2.31).

Since the parameter matrix $\boldsymbol{\beta}$ only appears in the last term of $\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a)$, maximizing the log-likelihood function over $\boldsymbol{\beta}$ is equivalent to minimizing $S(\boldsymbol{\beta})$. Consequently, the ML estimate of $\boldsymbol{\beta}$ is the same as its LS estimate. Next, taking the partial derivative of the log-likelihood function with respective to $\boldsymbol{\Sigma}_a$ and using properties (i) and (j) of **Result 3** of Appendix A, we obtain

$$\frac{\partial\ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_a)}{\partial\boldsymbol{\Sigma}_a} = -\frac{T-p}{2}\boldsymbol{\Sigma}_a^{-1} + \frac{1}{2}\boldsymbol{\Sigma}_a^{-1}\hat{\boldsymbol{A}}'\hat{\boldsymbol{A}}\boldsymbol{\Sigma}_a^{-1}. \tag{2.39}$$

Equating the prior normal equation to 0, we obtain the ML estimate of $\boldsymbol{\Sigma}_a$ as

$$\hat{\boldsymbol{\Sigma}}_a = \frac{1}{T-p}\hat{\boldsymbol{A}}'\hat{\boldsymbol{A}} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{\boldsymbol{a}}_t\hat{\boldsymbol{a}}_t'. \tag{2.40}$$

This result is the same as that for the multiple linear regression. The ML estimate of $\boldsymbol{\Sigma}_a$ is only asymptotically unbiased. Finally, the Hessian matrix of $\boldsymbol{\beta}$ can be obtained by taking the partial derivative of Equation (2.33), namely,

$$-\frac{\partial^2\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a)}{\partial\text{vec}(\boldsymbol{\beta})\partial\text{vec}(\boldsymbol{\beta})'} = \frac{1}{2}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial\text{vec}(\boldsymbol{\beta})\partial\text{vec}(\boldsymbol{\beta})'} = \boldsymbol{\Sigma}_a^{-1}\otimes\boldsymbol{X}'\boldsymbol{X}.$$

Inverse of the Hessian matrix provides the asymptotic covariance matrix of the ML estimate of vec($\boldsymbol{\beta}$). Next, using property (e) of **Result 3** and the product rule of Appendix A, and taking derivative of Equation (2.39), we obtain

$$\frac{\partial^2 \ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_a)}{\partial \text{vec}(\boldsymbol{\Sigma}_a)\partial \text{vec}(\boldsymbol{\Sigma}_a)'} = \frac{T-p}{2}\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{\Sigma}_a^{-1}\right) - \frac{1}{2}\left[\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{\Sigma}_a^{-1}\right)\hat{\boldsymbol{A}}'\hat{\boldsymbol{A}}\boldsymbol{\Sigma}_a^{-1}\right]$$
$$- \frac{1}{2}\left[\boldsymbol{\Sigma}_a^{-1}\hat{\boldsymbol{A}}'\hat{\boldsymbol{A}}\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{\Sigma}_a^{-1}\right)\right].$$

Consequently, we have

$$-E\left(\frac{\partial^2 \ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_a)}{\partial \text{vec}(\boldsymbol{\Sigma}_a)\partial \text{vec}(\boldsymbol{\Sigma}_a)'}\right) = \frac{T-p}{2}\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{\Sigma}_a^{-1}\right).$$

This result provides asymptotic covariance matrix for the ML estimates of elements of $\boldsymbol{\Sigma}_a$.

Theorem 2.2 Suppose that the innovation $\boldsymbol{a}_t$ of a stationary VAR(p) model follows a multivariate normal distribution with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_a$. Then, the ML estimates are $\text{vec}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ and $\hat{\boldsymbol{\Sigma}}_a = (1/T-p)\sum_{t=p+1}^{T} \hat{\boldsymbol{a}}_t\hat{\boldsymbol{a}}_t'$. Also, (i) $(T-p)\hat{\boldsymbol{\Sigma}}_a$ is distributed as $W_{k,T-(k+1)p-1}(\boldsymbol{\Sigma}_a)$, a Wishart distribution, and (ii) $\text{vec}(\hat{\boldsymbol{\beta}})$ is normally distributed with mean $\text{vec}(\boldsymbol{\beta})$ and covariance matrix $\boldsymbol{\Sigma}_a \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}$, and (iii) $\text{vec}(\hat{\boldsymbol{\beta}})$ is independent of $\hat{\boldsymbol{\Sigma}}_a$, where $\boldsymbol{Z}$ and $\boldsymbol{X}$ are defined in Equation (2.28). Furthermore, $\sqrt{T}[\text{vec}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}]$ and $\sqrt{T}[\text{vec}(\hat{\boldsymbol{\Sigma}}_a) - \text{vec}(\boldsymbol{\Sigma}_a)]$ are asymptotically normally distributed with mean zero and covariance matrices $\boldsymbol{\Sigma}_a \otimes \boldsymbol{G}^{-1}$ and $2\boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a$, respectively, where $\boldsymbol{G} = E(\boldsymbol{x}_t\boldsymbol{x}_t')$ with $\boldsymbol{x}_t$ defined in Equation (2.28).

Finally, given the data set $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T\}$, the maximized likelihood of a VAR(p) model is

$$L\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_a | \boldsymbol{z}_{1:p}\right) = (2\pi)^{-k(T-p)/2} |\hat{\boldsymbol{\Sigma}}_a|^{-(T-p)/2} \exp\left[-\frac{k(T-p)}{2}\right]. \quad (2.41)$$

This value is useful in likelihood ratio tests to be discussed later.

2.5.3 Limiting Properties of LS Estimate

Consider a k-dimensional stationary VAR(p) model in Equation (2.21). To study the asymptotic properties of the LS estimate $\hat{\boldsymbol{\beta}}$ in Equation (2.35), we need the assumption that the innovation series $\{\boldsymbol{a}_t\}$ is a sequence of independent and identically distributed random vector with mean zero and positive-definite covariance $\boldsymbol{\Sigma}_a$. Also, $\boldsymbol{a}_t = (a_{1t}, \ldots, a_{kt})'$ is continuous and satisfies

$$E|a_{it}a_{jt}a_{ut}a_{vt}| < \infty, \quad \text{for } i, j, u, v = 1, \ldots, k \text{ and all } t. \quad (2.42)$$

In other words, the fourth moment of $\boldsymbol{a}_t$ is finite. Under this assumption, we have the following results.

Lemma 2.3 If the VAR(p) process $\boldsymbol{z}_t$ of Equation (2.21) is stationary and satisfies the condition in Equation (2.42), then, as $T \to \infty$, we have

(i) $\boldsymbol{X}'\boldsymbol{X}/(T-p) \to_p \boldsymbol{G}$,
(ii) $(1/\sqrt{T-p})\text{vec}(\boldsymbol{X}'\boldsymbol{A}) = (1/\sqrt{T-p})(\boldsymbol{I}_k \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{A}) \to_d N(\boldsymbol{0}, \boldsymbol{\Sigma}_a \otimes \boldsymbol{G})$,

where $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively, $\boldsymbol{X}$ and $\boldsymbol{A}$ are defined in Equation (2.28), and $\boldsymbol{G}$ is a nonsingular matrix given by

$$\boldsymbol{G} = \begin{bmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{\Gamma}_0^* \end{bmatrix} + \begin{bmatrix} 0 \\ \boldsymbol{u} \end{bmatrix} [0, \boldsymbol{u}'],$$

where $\boldsymbol{0}$ is a kp-dimensional vector of zeros, $\boldsymbol{\Gamma}_0^*$ is defined in Equation (2.26) and $\boldsymbol{u} = \boldsymbol{1}_p \otimes \boldsymbol{\mu}$ with $\boldsymbol{1}_p$ being a p-dimensional vector of 1.

A proof of Lemma 2.3 can be found in Theorem 8.2.3 of Fuller (1976, p. 340) or in Lemma 3.1 of Lütkepohl (2005, p. 73). Using Lemma 2.3, one can establish the asymptotic distribution of the LS estimate $\hat{\boldsymbol{\beta}}$.

Theorem 2.3 Suppose that the VAR(p) time series $\boldsymbol{z}_t$ in Equation (2.21) is stationary and its innovation $\boldsymbol{a}_t$ satisfies the assumption in Equation (2.42). Then, as $T \to \infty$,

(i) $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$,
(ii) $\sqrt{T-p}[\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\beta})] = \sqrt{T-p}[\text{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \to_d N(\boldsymbol{0}, \boldsymbol{\Sigma}_a \otimes \boldsymbol{G}^{-1})$,

where $\boldsymbol{G}$ is defined in Lemma 2.3.

Proof. By Equation (2.37), we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T-p}\right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{A}}{T-p}\right) \to_p \boldsymbol{0},$$

because the last term approaches $\boldsymbol{0}$. This establishes the consistency of $\hat{\boldsymbol{\beta}}$. For result (ii), we can use Equation (2.34) to obtain

$$\begin{aligned} \sqrt{T-p}\left[\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\beta})\right] &= \sqrt{T-p}\left[\boldsymbol{I}_k \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right] \text{vec}(\boldsymbol{A}) \\ &= \sqrt{T-p}\left[\boldsymbol{I}_k \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right]\left[\boldsymbol{I}_k \otimes \boldsymbol{X}'\right] \text{vec}(\boldsymbol{A}) \\ &= \left[\boldsymbol{I}_k \otimes \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{T-p}\right)^{-1}\right] \frac{1}{\sqrt{T-p}} \left[\boldsymbol{I}_k \otimes \boldsymbol{X}'\right] \text{vec}(\boldsymbol{A}). \end{aligned}$$

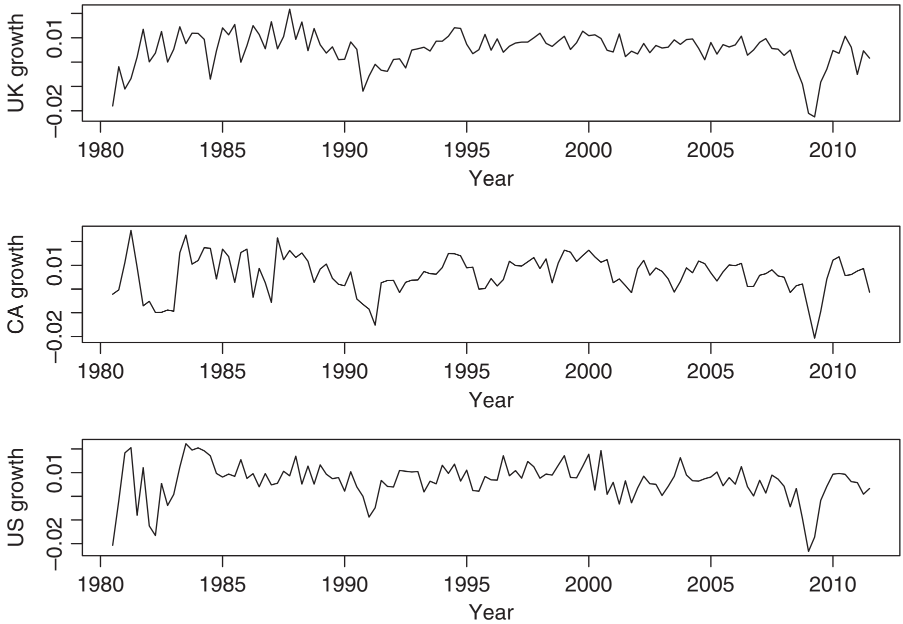

FIGURE 2.1 Time plots of the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011.

Therefore, the limiting distribution of $\sqrt{T-p}[\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\beta})]$ is the same as that of

$$\left(\boldsymbol{I}_k \otimes \boldsymbol{G}^{-1}\right) \frac{1}{\sqrt{T-p}} (\boldsymbol{I}_k \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{A}).$$

Hence, by Lemma 2.3, the limiting distribution of $\sqrt{T-p}[\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\beta})]$ is normal and the covariance matrix is

$$\left(\boldsymbol{I}_k \otimes \boldsymbol{G}^{-1}\right) (\boldsymbol{\Sigma}_a \otimes \boldsymbol{G}) \left(\boldsymbol{I}_k \otimes \boldsymbol{G}^{-1}\right) = \boldsymbol{\Sigma}_a \otimes \boldsymbol{G}^{-1}.$$

The proof is complete. □

Example 2.3 Consider the quarterly growth rates, in percentages, of real gross domestic product (GDP) of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The data were seasonally adjusted and downloaded from the database of Federal Reserve Bank at St. Louis. The GDP were in millions of local currency, and the growth rate denotes the differenced series of log GDP. Figure 2.1 shows the time plots of the three GDP growth rates. In our demonstration, we employ a VAR(2) model. In this particular instance, we have $k = 3$, $p = 2$, and $T = 125$. Using the notation defined in Section 2.5, we have

$$\tilde{\mathbf{\Sigma}}_a = \begin{bmatrix} 0.299 & 0.028 & 0.079 \\ 0.028 & 0.309 & 0.148 \\ 0.079 & 0.148 & 0.379 \end{bmatrix}, \quad \hat{\mathbf{\Sigma}}_a = \begin{bmatrix} 0.282 & 0.027 & 0.074 \\ 0.027 & 0.292 & 0.139 \\ 0.074 & 0.139 & 0.357 \end{bmatrix},$$

and the LS estimates vec($\hat{\boldsymbol{\beta}}$) and their standard errors and t-ratios are given in the following R demonstration. The two estimates of the covariance matrix differ by a factor of $116/123 = 0.943$. From the output, the t-ratios indicate that some of the LS estimates are not statistically significant at the usual 5% level. We shall discuss model checking and refinement later. □

R Demonstration

```
> da=read.table("q-gdp-ukcaus.txt",header=T)
> gdp=log(da[,3:5])
> dim(gdp)
[1] 126   3
> z=gdp[2:126,]-gdp[1:125,]  ## Growth rate
> z=z*100   ## Percentage growth rates
> dim(z)
[1] 125   3
> Z=z[3:125,]
> X=cbind(rep(1,123),z[2:124,],z[1:123,])
> X=as.matrix(X)
> XPX=t(X)%*%X
> XPXinv=solve(XPX)
> Z=as.matrix(Z)
> XPZ=t(X)%*%Z
> bhat=XPXinv%*%XPZ
> bhat
                    uk           ca          us
rep(1, 123) 0.12581630  0.123158083  0.28955814
uk          0.39306691  0.351313628  0.49069776
ca          0.10310572  0.338141505  0.24000097
us          0.05213660  0.469093555  0.23564221
uk          0.05660120 -0.191350134 -0.31195550
ca          0.10552241 -0.174833458 -0.13117863
us          0.01889462 -0.008677767  0.08531363
> A=Z-X%*%bhat
> Sig=t(A)%*%A/(125-(3+1)*2-1)
> Sig
           uk         ca         us
uk 0.29948825 0.02814252 0.07883967
ca 0.02814252 0.30917711 0.14790523
us 0.07883967 0.14790523 0.37850674
```

```
> COV=kronecker(Sig,XPXinv)
> se=sqrt(diag(COV))
> para=cbind(beta,se,beta/se)
> para
              beta         se    t-ratio
 [1,]  0.125816304 0.07266338  1.7314953
 [2,]  0.393066914 0.09341839  4.2075968
 [3,]  0.103105720 0.09838425  1.0479901
 [4,]  0.052136600 0.09112636  0.5721353
 [5,]  0.056601196 0.09237356  0.6127424
 [6,]  0.105522415 0.08755896  1.2051584
 [7,]  0.018894618 0.09382091  0.2013903
 [8,]  0.123158083 0.07382941  1.6681440
 [9,]  0.351313628 0.09491747  3.7012536
[10,]  0.338141505 0.09996302  3.3826660
[11,]  0.469093555 0.09258865  5.0664259
[12,] -0.191350134 0.09385587 -2.0387658
[13,] -0.174833458 0.08896401 -1.9652155
[14,] -0.008677767 0.09532645 -0.0910321
[15,]  0.289558145 0.08168880  3.5446492
[16,]  0.490697759 0.10502176  4.6723437
[17,]  0.240000969 0.11060443  2.1699038
[18,]  0.235642214 0.10244504  2.3001819
[19,] -0.311955500 0.10384715 -3.0039871
[20,] -0.131178630 0.09843454 -1.3326484
[21,]  0.085313633 0.10547428  0.8088572
> Sig1=t(A)%*%A/(125-2)  ## MLE of Sigma_a
> Sig1
           uk         ca         us
uk 0.28244420 0.02654091 0.07435286
ca 0.02654091 0.29158166 0.13948786
us 0.07435286 0.13948786 0.35696571
```

The aforementioned demonstration is to provide details about LS and ML estimation of a VAR model. In practice, we use some available packages in R to perform estimation. For example, we can use the command `VAR` in the `MTS` package to estimate a VAR model. The command and the associated output, which is in the matrix form, are shown next:

R Demonstration: Estimation of VAR models.

```
> da=read.table("q-gdp-ukcaus.txt",header=T)
> gdp=log(da[,3:5])
> z=gdp[2:126,]-gdp[1:125,]
> z=z*100
> m1=VAR(z,2)
Constant term:
Estimates:  0.1258163 0.1231581 0.2895581
```

```
Std.Error:  0.07266338 0.07382941 0.0816888
AR coefficient matrix
AR( 1 )-matrix
      [,1]  [,2]   [,3]
[1,] 0.393 0.103 0.0521
[2,] 0.351 0.338 0.4691
[3,] 0.491 0.240 0.2356
standard error
       [,1]   [,2]   [,3]
[1,] 0.0934 0.0984 0.0911
[2,] 0.0949 0.1000 0.0926
[3,] 0.1050 0.1106 0.1024
AR( 2 )-matrix
        [,1]   [,2]     [,3]
[1,]  0.0566  0.106  0.01889
[2,] -0.1914 -0.175 -0.00868
[3,] -0.3120 -0.131  0.08531
standard error
       [,1]   [,2]   [,3]
[1,] 0.0924 0.0876 0.0938
[2,] 0.0939 0.0890 0.0953
[3,] 0.1038 0.0984 0.1055

Residuals cov-mtx:
           [,1]       [,2]       [,3]
[1,] 0.28244420 0.02654091 0.07435286
[2,] 0.02654091 0.29158166 0.13948786
[3,] 0.07435286 0.13948786 0.35696571

det(SSE) =  0.02258974
AIC =  -3.502259
BIC =  -3.094982
HQ  =  -3.336804
```

From the output, the fitted VAR(2) model for the percentage growth rates of quarterly GDP of United Kingdom, Canada, and United States is

$$\boldsymbol{z}_t = \begin{bmatrix} 0.13 \\ 0.12 \\ 0.29 \end{bmatrix} + \begin{bmatrix} 0.38 & 0.10 & 0.05 \\ 0.35 & 0.34 & 0.47 \\ 0.49 & 0.24 & 0.24 \end{bmatrix} \boldsymbol{z}_{t-1}$$

$$+ \begin{bmatrix} 0.06 & 0.11 & 0.02 \\ -0.19 & -0.18 & -0.01 \\ -0.31 & -0.13 & 0.09 \end{bmatrix} \boldsymbol{z}_{t-2} + \boldsymbol{a}_t,$$

where the residual covariance matrix is

$$\hat{\mathbf{\Sigma}}_a = \begin{bmatrix} 0.28 & 0.03 & 0.07 \\ 0.03 & 0.29 & 0.14 \\ 0.07 & 0.14 & 0.36 \end{bmatrix}.$$

Standard errors of the coefficient estimates are given in the output. Again, some of the estimates are not statistically significant at the usual 5% level.

Remark: In our derivation, we used Equation (2.28). An alternative approach is to use

$$\boldsymbol{Y} = \boldsymbol{\varpi}\boldsymbol{W} + \boldsymbol{U},$$

where $\boldsymbol{Y} = [\boldsymbol{z}_{p+1}, \boldsymbol{z}_{p+2}, \ldots, \boldsymbol{z}_T]$, a $k \times (T-p)$ matrix, $\boldsymbol{\varpi} = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p]$, a $k \times (kp+1)$ parameter matrix, $\boldsymbol{W} = [\boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_T]$ with $\boldsymbol{x}_t$ defined in Equation (2.28), and $\boldsymbol{U} = [\boldsymbol{a}_{p+1}, \ldots, \boldsymbol{a}_T]$. Obviously, we have $\boldsymbol{Y} = \boldsymbol{Z}'$, $\boldsymbol{\varpi} = \boldsymbol{\beta}'$, $\boldsymbol{W} = \boldsymbol{X}'$, and $\boldsymbol{U} = \boldsymbol{A}'$. One can then derive the limiting distribution of the LS estimate $\hat{\boldsymbol{\varpi}}$ of $\boldsymbol{\varpi}$ via using exactly the same argument as we used. Under this approach, Lemma 2.3 (ii) becomes

$$\frac{1}{\sqrt{T_p}}\text{vec}(\boldsymbol{U}\boldsymbol{W}') = \frac{1}{\sqrt{T_p}}(\boldsymbol{W} \otimes \boldsymbol{I}_k)\text{vec}(\boldsymbol{U}) \to_d N(\boldsymbol{0}, \boldsymbol{G} \otimes \mathbf{\Sigma}_a),$$

where $T_p = T - p$, and $\boldsymbol{G} = \lim(\boldsymbol{W}\boldsymbol{W}')/T_p$, which is the same as defined in Lemma 2.3. Furthermore, Theorem 2.3 becomes

$$\sqrt{T_p}\text{vec}\left(\hat{\boldsymbol{\varpi}} - \boldsymbol{\varpi}\right) = \sqrt{T_p}\text{vec}\left(\hat{\boldsymbol{\beta}}' - \boldsymbol{\beta}'\right) \to_d N\left(\boldsymbol{0}, \boldsymbol{G}^{-1} \otimes \mathbf{\Sigma}_a\right). \tag{2.43}$$

□

2.5.4 Bayesian Estimation

We consider Bayesian estimation of a stationary VAR(p) model in this section. The basic framework used is the multivariate multiple linear regression in Equation (2.28). We begin with a brief review of Bayesian inference.

2.5.4.1 Review of Bayesian Paradigm

Consider a statistical inference problem. Denote the set of unknown parameters by $\mathbf{\Theta}$. Our prior beliefs of $\mathbf{\Theta}$ are often expressed via a probability distribution with density function $f(\mathbf{\Theta})$. Let $\boldsymbol{D}$ denote the observed data. The information provided by $\boldsymbol{D}$ is the likelihood function $f(\boldsymbol{D}|\mathbf{\Theta})$. By Bayes' theorem, we can combine the prior and likelihood to produce the distribution of the parameters conditional on the data and prior:

$$f(\boldsymbol{\Theta}|\boldsymbol{D}) = \frac{f(\boldsymbol{D}, \boldsymbol{\Theta})}{f(\boldsymbol{D})} = \frac{f(\boldsymbol{D}|\boldsymbol{\Theta})f(\boldsymbol{\Theta})}{f(\boldsymbol{D})}, \tag{2.44}$$

where $f(\boldsymbol{D}) = \int f(\boldsymbol{D}, \boldsymbol{\Theta})d\boldsymbol{\Theta} = \int f(\boldsymbol{D}|\boldsymbol{\Theta})f(\boldsymbol{\Theta})d\boldsymbol{\Theta}$ is the marginal distribution of $\boldsymbol{D}$, obtained by integrating out $\boldsymbol{\Theta}$. The density function $f(\boldsymbol{\Theta}|\boldsymbol{D})$ is called the *posterior* distribution. Bayesian inference on $\boldsymbol{\Theta}$ is drawn from this posterior distribution.

The marginal distribution $f(\boldsymbol{D})$ in Equation (2.44) serves as the normalization constant, referred to as the constant of proportionality, and its actual value is not critical in many applications provided that it exists. Therefore, we may rewrite the equation as

$$f(\boldsymbol{\Theta}|\boldsymbol{D}) \propto f(\boldsymbol{D}|\boldsymbol{\Theta})f(\boldsymbol{\Theta}). \tag{2.45}$$

If the prior distribution $f(\boldsymbol{\Theta})$ and the posterior distribution $f(\boldsymbol{\Theta}|\boldsymbol{D})$ belong to the same family of distributions, then the prior is called a *conjugate prior*. Conjugate priors are often used in Bayesian inference because they enable us to obtain analytic expressions for the posterior distribution.

2.5.4.2 VAR Estimation

To derive the Bayesian estimation of a stationary VAR(p) model, we employ the model in Equation (2.28), namely,

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}, \tag{2.46}$$

where $\boldsymbol{Z}$ and $\boldsymbol{A}$ are $(T-p) \times k$ matrices, $\boldsymbol{\beta}' = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p]$ is a $k \times (kp+1)$ matrix of coefficient parameters, and the ith rows of $\boldsymbol{Z}$ and $\boldsymbol{A}$ are $\boldsymbol{z}'_{p+i}$ and $\boldsymbol{a}'_{p+i}$, respectively. The matrix $\boldsymbol{X}$ is a $(T-p) \times (kp+1)$ design matrix with ith row being $(1, \boldsymbol{z}'_{p+i-1}, \ldots, \boldsymbol{z}'_i)$. The unknown parameters of the VAR(p) model are $\boldsymbol{\Theta} = [\boldsymbol{\beta}', \boldsymbol{\Sigma}_a]$. Rossi, Allenby, and McCulloch (2005) provide a good treatment of Bayesian estimation for the model in Equation (2.46). We adopt their approach.

For ease in notation, let $n = T - p$ be the effective sample size for estimation. Also, we shall omit the condition $\boldsymbol{z}_{1:p}$ from all equations. As stated in Section 2.5.2, the likelihood function of the data is

$$f(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_a) \propto |\boldsymbol{\Sigma}_a|^{-n/2} \exp\left[-\tfrac{1}{2}tr\left\{(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}_a^{-1}\right\}\right].$$

Using the LS properties in Appendix A, we have

$$(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta}) = \hat{\boldsymbol{A}}'\hat{\boldsymbol{A}} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}).$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ is the LS estimate of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{A}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is the residual matrix. The likelihood function can be written as

$$\begin{aligned} f(\boldsymbol{Z}|\boldsymbol{\beta},\boldsymbol{\Sigma}_a) \propto\ & |\boldsymbol{\Sigma}_a|^{-(n-k)/2}\exp\left[-\tfrac{1}{2}tr(\boldsymbol{S}\boldsymbol{\Sigma}_a^{-1})\right] \\ & \times |\boldsymbol{\Sigma}_a|^{-k/2}\exp\left[-\tfrac{1}{2}tr\left\{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}_a^{-1}\right\}\right], \end{aligned} \tag{2.47}$$

where $\boldsymbol{S} = \hat{\boldsymbol{A}}'\hat{\boldsymbol{A}}$. The first term of Equation (2.47) does not depend on $\boldsymbol{\beta}$. This suggests that the natural conjugate prior for $\boldsymbol{\Sigma}_a$ is the inverted Wishart distribution and the prior for $\boldsymbol{\beta}$ can be conditioned on $\boldsymbol{\Sigma}_a$. Let $\boldsymbol{K} = (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}_a^{-1}$ be the matrix in the exponent of Equation (2.47). The exponent of the second term in Equation (2.47) can be rewritten as

$$\begin{aligned} tr(\boldsymbol{K}) &= \left[\mathrm{vec}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right]'\mathrm{vec}\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}_a^{-1}\right] \\ &= \left[\mathrm{vec}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right]'\left(\boldsymbol{\Sigma}_a^{-1}\otimes\boldsymbol{X}'\boldsymbol{X}\right)\mathrm{vec}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) \\ &= \left[\mathrm{vec}(\boldsymbol{\beta})-\mathrm{vec}(\hat{\boldsymbol{\beta}})\right]'\left(\boldsymbol{\Sigma}_a^{-1}\otimes\boldsymbol{X}'\boldsymbol{X}\right)\left[\mathrm{vec}(\boldsymbol{\beta})-\mathrm{vec}(\hat{\boldsymbol{\beta}})\right]. \end{aligned}$$

Consequently, the second term in Equation (2.47) is a multivariate normal kernel. This means that the natural conjugate prior for vec($\boldsymbol{\beta}$) is multivariate normal conditional on $\boldsymbol{\Sigma}_a$.

The conjugate priors for the VAR(p) model in Equation (2.46) are of the form

$$\begin{aligned} f(\boldsymbol{\beta},\boldsymbol{\Sigma}_a) &= f(\boldsymbol{\Sigma}_a)f(\boldsymbol{\beta}|\boldsymbol{\Sigma}_a) \\ \boldsymbol{\Sigma}_a &\sim W^{-1}(\boldsymbol{V}_o, n_o) \\ \mathrm{vec}(\boldsymbol{\beta}) &\sim N\left[\mathrm{vec}(\boldsymbol{\beta}_o), \boldsymbol{\Sigma}_a\otimes\boldsymbol{C}^{-1}\right], \end{aligned} \tag{2.48}$$

where $\boldsymbol{V}_o$ is $k\times k$ and $\boldsymbol{C}$ is $(kp+1)\times(kp+1)$, both are positive-definite, $\boldsymbol{\beta}_o$ is a $k\times(kp+1)$ matrix and n_o is a real number. These quantities are referred to as the hyper parameters in Bayesian inference and their values are assumed to be known in this section. For information on Wishart and inverted Wishart distributions, see Appendix A.

Using the pdf of inverted Wishart distribution in Equation (A.12), the posterior distribution is

$$\begin{aligned} f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a | \boldsymbol{Z}, \boldsymbol{X}) \propto & |\boldsymbol{\Sigma}_a|^{-(v_o+k+1)/2} \exp\left[-\tfrac{1}{2} tr(\boldsymbol{V}_o \boldsymbol{\Sigma}_a^{-1})\right] \\ & \times |\boldsymbol{\Sigma}_a|^{-k/2} \exp\left[-\tfrac{1}{2} tr\left\{(\boldsymbol{\beta}-\boldsymbol{\beta}_o)' \boldsymbol{C} (\boldsymbol{\beta}-\boldsymbol{\beta}_o) \boldsymbol{\Sigma}_a^{-1}\right\}\right] \\ & \times |\boldsymbol{\Sigma}_a|^{-n/2} \exp\left[-\tfrac{1}{2} tr\left\{(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta}) \boldsymbol{\Sigma}_a^{-1}\right\}\right]. \end{aligned} \tag{2.49}$$

To simplify, we can combine the two terms in Equation (2.49) involving $\boldsymbol{\beta}$ via the LS properties in Appendix A. Specifically, denote the Cholesky decomposition of $\boldsymbol{C}$ as $\boldsymbol{C} = \boldsymbol{U}'\boldsymbol{U}$, where $\boldsymbol{U}$ is an upper triangular matrix. Define

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{U} \end{bmatrix}, \quad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{U}\boldsymbol{\beta}_o \end{bmatrix}.$$

Then, we have

$$(\boldsymbol{\beta}-\boldsymbol{\beta}_o)' \boldsymbol{C} (\boldsymbol{\beta}-\boldsymbol{\beta}_o) + (\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{\beta}).$$

Applying Property (ii) of the LS estimation in Appendix A, we have

$$\begin{aligned} (\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{\beta}) &= (\boldsymbol{Y}-\boldsymbol{W}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{W}\tilde{\boldsymbol{\beta}}) + (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\boldsymbol{W}'\boldsymbol{W}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}), \\ &= \tilde{\boldsymbol{S}} + (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\boldsymbol{W}'\boldsymbol{W}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}), \end{aligned} \tag{2.50}$$

where

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{C})^{-1}(\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{C}\boldsymbol{\beta}_o),$$

and

$$\tilde{\boldsymbol{S}} = (\boldsymbol{Y}-\boldsymbol{W}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{W}\tilde{\boldsymbol{\beta}}) = (\boldsymbol{Z}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Z}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_o)'\boldsymbol{C}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_o).$$

Using Equation(2.50), we can write the posterior distribution as

$$\begin{aligned} f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a | \boldsymbol{Z}, \boldsymbol{X}) \propto & |\boldsymbol{\Sigma}_a|^{-k/2} \exp\left[-\tfrac{1}{2} tr\left\{(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\boldsymbol{W}'\boldsymbol{W}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right)\boldsymbol{\Sigma}_a^{-1}\right\}\right] \\ & \times |\boldsymbol{\Sigma}_a|^{-(n_o+n+k+1)/2} \exp\left[-\tfrac{1}{2} tr\left\{\left(\boldsymbol{V}_o + \tilde{\boldsymbol{S}}\right)\boldsymbol{\Sigma}_a^{-1}\right\}\right]. \end{aligned} \tag{2.51}$$

The first term of Equation (2.51) is a multivariate normal kernel, whereas the second term is an inverted Wishart kernel. Consequently, the posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_a$ are

$$\begin{aligned} \boldsymbol{\Sigma}_a | \boldsymbol{Z}, \boldsymbol{X} &\sim W^{-1}(\boldsymbol{V}_o + \tilde{\boldsymbol{S}}, n_o + n) \\ \text{vec}(\boldsymbol{\beta}) | \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{\Sigma}_a &\sim N[\text{vec}(\tilde{\boldsymbol{\beta}}), \boldsymbol{\Sigma}_a \otimes (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{C})^{-1}], \end{aligned} \tag{2.52}$$

where $n = T - p$ and

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{C})^{-1}(\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{C}\boldsymbol{\beta}_o), \\ \tilde{\boldsymbol{S}} &= (\boldsymbol{Z} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Z} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o)'\boldsymbol{C}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_o). \end{aligned}$$

The inverse of the covariance matrix of a multivariate normal distribution is called the *precision* matrix. We can interpret $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{C}$ as the precision matrices of the LS estimate $\hat{\boldsymbol{\beta}}$ and the prior distribution of $\boldsymbol{\beta}$, respectively. (They are relative precision matrices as the covariance matrices involve $\boldsymbol{\Sigma}_a$.) The posterior mean $\tilde{\boldsymbol{\beta}}$ is a weighted average between the LS estimates and the prior mean. The precision of the posterior distribution of $\boldsymbol{\beta}$ is simply the sum of the two precision matrices. These results are generalizations of those of the multiple linear regression; see, for instance, Tsay (2010, Chapter 12).

We can use posterior means as point estimates of the VAR(p) parameters. Therefore, the Bayesian estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_a$ are

$$\breve{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} \quad \text{and} \quad \breve{\boldsymbol{\Sigma}}_a = \frac{\boldsymbol{V}_o + \tilde{\boldsymbol{S}}}{n_o + T - p - k - 1}.$$

The covariance matrix of vec($\breve{\boldsymbol{\beta}}$) is $\breve{\boldsymbol{\Sigma}}_a \otimes (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{C})^{-1}$.

In practice, we may not have concrete prior information about $\boldsymbol{\beta}$ for a stationary VAR(p) model, especially the correlations between the parameters. In this case, we may choose $\boldsymbol{\beta}_o = \boldsymbol{0}$ and a large covariance matrix $\boldsymbol{C}^{-1}$, say $\boldsymbol{C}^{-1} = \delta \boldsymbol{I}_{kp+1}$ with a large δ. These particular prior choices are referred to as *vague priors* and would result in a small $\boldsymbol{C} = \delta^{-1}\boldsymbol{I}_{kp+1}$, and Equation (2.52) shows that the Bayesian estimate $\tilde{\boldsymbol{\beta}}$ is close to the LS estimate $\hat{\boldsymbol{\beta}}$. A small $\boldsymbol{C}$ also leads to $\tilde{\boldsymbol{S}}$ being close to $\boldsymbol{S}$ of the LS estimate given in Equation (2.47). Consequently, if we also choose a small $\boldsymbol{V}_o$ for the prior of $\boldsymbol{\Sigma}_a$, then the Bayesian estimate of $\boldsymbol{\Sigma}_a$ is also close to the LS estimate.

The choices of prior distributions are subjective in a real application. Our use of conjugate priors and $\boldsymbol{C} = \lambda \boldsymbol{I}_{kp+1}$ with a small λ are for convenience. One can apply several priors to study the sensitivity of the analysis to prior specification. For instance, one can specify $\boldsymbol{C}$ as a diagonal matrix with different diagonal elements to reflect the common prior belief that higher-order AR lags are of decreasing importance. For stationary VAR(p) models, Bayesian estimates are typically not too sensitive to any reasonable prior specification, especially when the sample size is large. In the literature, Litterman (1986) and Doan, Litterman, and Sims (1984) describe a specific prior for stationary VAR(p) models. The prior is known as the *Minnesota prior*. For $\boldsymbol{\beta}$, the prior is multivariate normal with mean zero and a diagonal covariance matrix. That is, Minnesota prior replaces $\boldsymbol{\Sigma}_a \otimes \boldsymbol{C}^{-1}$ by a diagonal matrix $\boldsymbol{V}$. For the AR coefficient $\boldsymbol{\phi}_{\ell,ij}$, the prior variance is given by

$$\text{Var}(\phi_{\ell,ij}) = \begin{cases} (\lambda/\ell)^2 & \text{if } i = j \\ (\lambda\theta/\ell)^2 \times (\sigma_{ii}/\sigma_{jj}) & \text{if } i \neq j, \end{cases}$$

where λ is a real number, $0 < \theta < 1$, σ_{ii} is the (i,i)th element of $\boldsymbol{\Sigma}_a$, and $\ell = 1, \ldots, p$. From the specification, the prior states that $\phi_{\ell,ij}$ is close to 0 as ℓ increases. The choices of λ and θ for the Minnesota prior are subjective in an application.

Example 2.4 Consider, again, the percentage growth rates of quarterly real GDP of United Kingdom, Canada, and United States employed in Example 2.3. We specify a VAR(2) model and use the noninformative conjugate priors with

$$\boldsymbol{C} = 0.1 \times \boldsymbol{I}_7, \quad \boldsymbol{V}_o = \boldsymbol{I}_3, \quad n_o = 5, \quad \boldsymbol{\beta}_o = \boldsymbol{0}.$$

The estimation results are shown in the following R demonstration. As expected, the Bayesian estimates of $\boldsymbol{\phi}_i$ and $\boldsymbol{\Sigma}_a$ are close to the LS estimates. The command for Bayesian VAR estimation in the `MTS` package is `BVAR`. □

R Demonstration: Bayesian estimation.

```
> da=read.table("q-gdp-ukcaus.txt",header=T)
> x=log(da[,3:5])
> dim(x)
[1] 126   3
> dx=x[2:126,]-x[1:125,]
> dx=dx*100
> C=0.1*diag(7)  ### lambda = 0.1
> V0=diag(3) ### Vo = I_3
> mm=BVAR(dx,p=2,C,V0)
Bayesian estimate:
                 Est       s.e.     t-ratio
 [1,]  0.125805143 0.07123059  1.76616742
 [2,]  0.392103983 0.09150764  4.28493158
 [3,]  0.102894946 0.09633822  1.06805941
 [4,]  0.052438976 0.08925487  0.58751947
 [5,]  0.056937547 0.09048722  0.62923303
 [6,]  0.105553695 0.08578002  1.23051603
 [7,]  0.019147973 0.09188759  0.20838475
 [8,]  0.123256168 0.07237470  1.70302833
 [9,]  0.350253306 0.09297745  3.76707803
[10,]  0.337525508 0.09788562  3.44816232
```

```
[11,]  0.468440207 0.09068850  5.16537628
[12,] -0.190144541 0.09194064 -2.06812294
[13,] -0.173964344 0.08715783 -1.99596908
[14,] -0.008627966 0.09336351 -0.09241262
[15,]  0.289317667 0.07987129  3.62229886
[16,]  0.489072359 0.10260807  4.76641231
[17,]  0.239456311 0.10802463  2.21668257
[18,]  0.235601116 0.10008202  2.35408023
[19,] -0.310286945 0.10146386 -3.05810301
[20,] -0.130271750 0.09618566 -1.35437813
[21,]  0.085039470 0.10303411  0.82535258
Covariance matrix:
           uk         ca         us
uk 0.28839063 0.02647455 0.07394349
ca 0.02647455 0.29772937 0.13875034
us 0.07394349 0.13875034 0.36260138
```

2.6 ORDER SELECTION

Turn to model building. We follow the iterated procedure of Box and Jenkins consisting of model specification, estimation, and diagnostic checking. See Box, Jenkins, and Reinsel (2008). For VAR models, model specification is to select the order p. Several methods have been proposed in the literature to select the VAR order. We discuss two approaches. The first approach adopts the framework of multivariate multiple linear regression and uses sequential likelihood ratio tests. The second approach employs information criteria.

2.6.1 Sequential Likelihood Ratio Tests

This approach to selecting the VAR order was recommended by Tiao and Box (1981). The basic idea of the approach is to compare a VAR(ℓ) model with a VAR($\ell - 1$) model. In statistical terms, it amounts to consider the hypothesis testing

$$H_0 : \boldsymbol{\phi}_\ell = \mathbf{0} \quad \text{versus} \quad H_a : \boldsymbol{\phi}_\ell \neq \mathbf{0}. \tag{2.53}$$

This is a problem of nested hypotheses, and a natural test statistic to use is the likelihood ratio statistic. As shown in Section 2.5, we can employ the multivariate linear regression framework for a VAR model. Let $\boldsymbol{\beta}'_\ell = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_\ell]$ be the matrix of coefficient parameters of a VAR(ℓ) model and $\boldsymbol{\Sigma}_{a,\ell}$ be the corresponding innovation covariance matrix. Under the normality assumption, the likelihood ratio for the testing problem in Equation (2.53) is

$$\Lambda = \frac{\max L(\boldsymbol{\beta}_{\ell-1}, \boldsymbol{\Sigma}_a)}{\max L(\boldsymbol{\beta}_\ell, \boldsymbol{\Sigma}_a)} = \left(\frac{|\hat{\boldsymbol{\Sigma}}_{a,\ell}|}{|\hat{\boldsymbol{\Sigma}}_{a,\ell-1}|} \right)^{(T-\ell)/2}.$$

This equation follows the maximized likelihood function of a VAR model in Equation (2.41). Note that here we estimate the VAR($\ell - 1$) model using the regression setup of the VAR(ℓ) model. In other words, the $\boldsymbol{Z}$ matrix of Equation (2.28) consists of $\boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_T$. The likelihood ratio test of H_0 is equivalent to rejecting H_0 for large values of

$$-2\ln(\Lambda) = -(T-\ell)\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}_{a,\ell}|}{|\hat{\boldsymbol{\Sigma}}_{a,\ell-1}|}\right).$$

A commonly used test statistic is then

$$M(\ell) = -(T-\ell-1.5-k\ell)\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}_{a,\ell}|}{|\hat{\boldsymbol{\Sigma}}_{a,\ell-1}|}\right),$$

which follows asymptotically a chi-square distribution with k^2 degrees of freedom. This test statistic is widely used in the multivariate statistical analysis. See, for instance, Result 7.11 of Johnson and Wichern (2007).

To simplify the computation, Tiao and Box (1981) suggest the following procedure to compute the $M(\ell)$ statistic and to select the VAR order:

1. Select a positive integer P, which is the maximum VAR order entertained.
2. Setup the multivariate multiple linear regression framework of Equation (2.28) for the VAR(P) model. That is, there are $T - P$ observations in the $\boldsymbol{Z}$ data matrix for estimation.
3. For $\ell = 0, \ldots, P$, compute the LS estimate of the AR coefficient matrix, that is, compute $\hat{\boldsymbol{\beta}}_\ell$. For $\ell = 0$, $\boldsymbol{\beta}'$ is simply the constant vector $\boldsymbol{\phi}_0$. Then, compute the ML estimate of $\boldsymbol{\Sigma}_a$, that is, compute $\hat{\boldsymbol{\Sigma}}_{a,\ell} = (1/T-P)\hat{\boldsymbol{A}}_\ell'\hat{\boldsymbol{A}}_\ell$, where $\hat{\boldsymbol{A}}_\ell = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\ell$ is the residual matrix of the fitted VAR(ℓ) model.
4. For $\ell = 1, \ldots, P$, compute the modified likelihood ratio test statistic

$$M(\ell) = -(T-P-1.5-k\ell)\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}_{a,\ell}|}{|\hat{\boldsymbol{\Sigma}}_{a,\ell-1}|}\right), \tag{2.54}$$

   and its p-value, which is based on the asymptotic $\chi^2_{k^2}$ distribution.
5. Examine the test statistics sequentially starting with $\ell = 1$. If all p-values of the $M(\ell)$ test statistics are greater than the specified type I error for $\ell > p$, then a VAR(p) model is specified. This is so because the test rejects the null hypothesis $\boldsymbol{\phi}_p = \boldsymbol{0}$, but fails to reject $\boldsymbol{\phi}_\ell = \boldsymbol{0}$ for $\ell > p$.

In practice, a simpler model is preferred. Thus, we often start with a small p. This is particularly so when the dimension k is high.

2.6.2 Information Criteria

Information criteria have been shown to be effective in selecting a statistical model. In the time series literature, several criteria have been proposed. All criteria are likelihood based and consist of two components. The first component is concerned with the goodness of fit of the model to the data, whereas the second component penalizes more heavily complicated models. The goodness of fit of a model is often measured by the maximized likelihood. For normal distribution, the maximized likelihood is equivalent to the determinant of the covariance matrix of the innovations; see Equation (2.41). This determinant is known as the *generalized variance* in multivariate analysis. The selection of the penalty, on the other hand, is relatively subjective. Different penalties result in different information criteria.

Three criteria functions are commonly used to determine VAR order. Under the normality assumption, these three criteria for a VAR(ℓ) model are

$$
\begin{aligned}
\text{AIC}(\ell) &= \ln|\hat{\mathbf{\Sigma}}_{a,\ell}| + \frac{2}{T}\ell k^2, \\
\text{BIC}(\ell) &= \ln|\hat{\mathbf{\Sigma}}_{a,\ell}| + \frac{\ln(T)}{T}\ell k^2, \\
\text{HQ}(\ell) &= \ln|\hat{\mathbf{\Sigma}}_{a,\ell}| + \frac{2\ln[\ln(T)]}{T}\ell k^2,
\end{aligned}
$$

where T is the sample size, $\hat{\mathbf{\Sigma}}_{a,\ell}$ is the ML estimate of $\mathbf{\Sigma}_a$ discussed in Section 2.5.2, AIC is the Akaike information criterion proposed in Akaike (1973). BIC stands for Bayesian information criterion (see Schwarz 1978), and HQ(ℓ) is proposed by Hannan and Quinn (1979) and Quinn (1980). The AIC penalizes each parameter by a factor of 2. BIC and HQ, on the other hand, employ penalties that depend on the sample size. For large T, BIC penalizes complicated models more heavily; for example, when $\ln(T) > 2$. HQ penalizes each parameter by $2\ln(\ln(T))$, which is greater than 2 when $T > 15$.

If $\boldsymbol{z}_t$ is indeed a Gaussian VAR(p) time series with $p < \infty$, then both BIC and HQ are consistent in the sense that they will select the true VAR(p) model with probability 1 as $T \to \infty$. The AIC, on the other hand, is not consistent as it has positive probabilities to select VAR(ℓ) models for $\ell > p$. The criterion, however, does not select a VAR(ℓ) model with $\ell < p$ when $T \to \infty$. See Quinn (1980). There is discussion in the literature about the validity of using consistency to compare information criteria because consistency requires the existence of a true model, yet there are no true models in real applications. Shibata (1980) derived asymptotic optimality properties of AIC for univariate time series.

Example 2.5 Consider, again, the growth rates of quarterly real GDP of United Kingdom, Canada, and United States from 1980.II to 2011.II. We apply the sequential likelihood ratio tests and all three information criteria to the data. The maximum order entertained is 13. Table 2.1 summarizes these statistics. From the table, we see that the orders selected by AIC, BIC, and HQ are 2, 1, and 1, respectively. The

TABLE 2.1 Order Selection Statistics for the Quarterly Growth Rates of the Real GDP of United Kingdom, Canada, and United States from the Second Quarter of 1980 to the Second Quarter of 2011

p	AIC	BIC	HQ	$M(p)$	p-Value
0	−30.956	−30.956	−30.956	0.000	0.000
1	−31.883	−31.679	−31.800	115.13	0.000
2	−31.964	−31.557	−31.799	23.539	0.005
3	−31.924	−31.313	−31.675	10.486	0.313
4	−31.897	−31.083	−31.566	11.577	0.238
5	−31.782	−30.764	−31.368	2.741	0.974
6	−31.711	−30.489	−31.215	6.782	0.660
7	−31.618	−30.192	−31.039	4.547	0.872
8	−31.757	−30.128	−31.095	24.483	0.004
9	−31.690	−29.857	−30.945	6.401	0.669
10	−31.599	−29.563	−30.772	4.322	0.889
11	−31.604	−29.364	−30.694	11.492	0.243
12	−31.618	−29.175	−30.626	11.817	0.224
13	−31.673	−29.025	−30.596	14.127	0.118

The information criteria used are AIC, BIC, and HQ. The M-statistics are given in Equation (2.54).

sequential M-statistic of Equation (2.54) selects the order $p = 2$, except for a violation at $p = 8$. This example demonstrates that different criteria may select different orders for a multivariate time series. Keep in mind, however, that these statistics are estimates. As such, one cannot take the values too seriously. Figure 2.2 shows the time plots of the three information criteria. The AIC shows a relatively close values for $p \in \{1, 2, 3, 4\}$, BIC shows a clear minimum at $p = 1$, whereas HQ shows a minimum at $p = 1$ with $p = 2$ as a close second. All three criteria show a drop at $p = 8$. In summary, a VAR(1) or VAR(2) model may serve as a starting model for the three-dimensional GDP series. □

R Demonstration: Order selection.

```
> z1=z/100  ### Original growth rates
> m2=VARorder(z1)
selected order: aic =  2
selected order: bic =  1
selected order: hq =  1
M statistic and its p-value
        Mstat       pv
 [1,] 115.133 0.000000
 [2,]  23.539 0.005093
 [3,]  10.486 0.312559
```

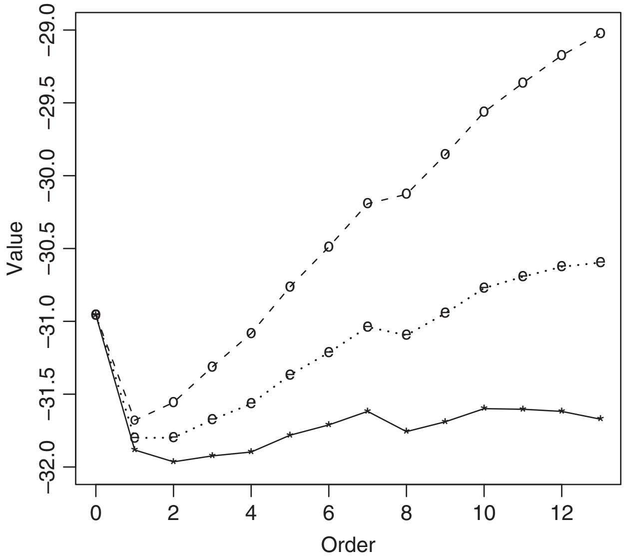

FIGURE 2.2 Information criteria for the quarterly growth rates, in percentages, of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The solid, dashed, and dotted lines are for AIC, BIC, and HQ, respectively.

```
 [4,]  11.577 0.238240
 [5,]   2.741 0.973698
 [6,]   6.782 0.659787
 [7,]   4.547 0.871886
 [8,]  24.483 0.003599
 [9,]   6.401 0.699242
[10,]   4.323 0.888926
[11,]  11.492 0.243470
[12,]  11.817 0.223834
[13,]  14.127 0.117891
Summary table:
       p     AIC     BIC      HQ     M(p)   p-value
 [1,]  0 -30.956 -30.956 -30.956   0.0000 0.0000000
 [2,]  1 -31.883 -31.679 -31.800 115.1329 0.0000000
 [3,]  2 -31.964 -31.557 -31.799  23.5389 0.0050930
 [4,]  3 -31.924 -31.313 -31.675  10.4864 0.3125594
 [5,]  4 -31.897 -31.083 -31.566  11.5767 0.2382403
 [6,]  5 -31.782 -30.764 -31.368   2.7406 0.9736977
 [7,]  6 -31.711 -30.489 -31.215   6.7822 0.6597867
 [8,]  7 -31.618 -30.192 -31.039   4.5469 0.8718856
 [9,]  8 -31.757 -30.128 -31.095  24.4833 0.0035992
[10,]  9 -31.690 -29.857 -30.945   6.4007 0.6992417
[11,] 10 -31.599 -29.563 -30.772   4.3226 0.8889256
[12,] 11 -31.604 -29.364 -30.694  11.4922 0.2434698
```

```
[13,] 12 -31.618 -29.175 -30.626  11.8168 0.2238337
[14,] 13 -31.672 -29.025 -30.596  14.1266 0.1178914

> names(m2)
[1] "aic"  "aicor" "bic"  "bicor" "hq"   "hqor"  "Mstat" "Mpv"
```

Remark: There are different ways to compute the information criteria for a given time series realization $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T\}$. The first approach is to use the same number of observations as discussed in the calculation of the $M(\ell)$ statistics in Equation (2.54). Here, one uses the data from $t = P + 1$ to T to evaluate the likelihood functions, where P is the maximum AR order. In the `MTS` package, the command `VARorder` uses this approach. The second approach is to use the data from $t = \ell + 1$ to T to fit a VAR(ℓ) model. In this case, different VAR models use different numbers of observations in estimation. For a large T, the two approaches should give similar results. However, when the sample size T is moderate compared with the dimension k, the two approaches may give different order selections even for the same criterion function. In the `MTS` package, the command `VARorderI` uses the second approach. □

2.7 MODEL CHECKING

Model checking, also known as diagnostic check or residual analysis, plays an important role in model building. Its main objectives include (i) to ensure that the fitted model is adequate and (ii) to suggest directions for further improvements if needed. The adequacy of a fitted model is judged according to some selected criteria, which may depend on the objective of the analysis. Typically a fitted model is said to be adequate if (a) all fitted parameters are statistically significant (at a specified level), (b) the residuals have no significant serial or cross-sectional correlations, (c) there exist no structural changes or outlying observations, and (d) the residuals do not violate the distributional assumption, for example, multivariate normality. We discuss some methods for model checking in this section.

2.7.1 Residual Cross-Correlations

The residuals of an adequate model should behave like a white noise series. Checking the serial and cross-correlations of the residuals thus becomes an integral part of model checking. Let $\hat{\boldsymbol{A}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ be the residual matrix of a fitted VAR(p) model, using the notation in Equation (2.28). The ith row of $\hat{\boldsymbol{A}}$ contains $\hat{\boldsymbol{a}}_{p+i} = \boldsymbol{z}_{p+i} - \hat{\boldsymbol{\phi}}_0 - \sum_{i=1}^{p} \hat{\boldsymbol{\phi}}_i \boldsymbol{z}_{t-i}$. The lag ℓ cross-covariance matrix of the residual series is defined as

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

In particular, we have $\hat{\boldsymbol{C}}_0 = \hat{\boldsymbol{\Sigma}}_a$ is the residual covariance matrix. In matrix notation, we can rewrite the lag ℓ residual cross-covariance matrix $\hat{\boldsymbol{C}}_\ell$ as

$$\hat{\boldsymbol{C}}_\ell = \frac{1}{T-p}\hat{\boldsymbol{A}}'\boldsymbol{B}^\ell\hat{\boldsymbol{A}}, \quad \ell \geq 0 \tag{2.55}$$

where $\boldsymbol{B}$ is a $(T-p) \times (T-p)$ back-shift matrix defined as

$$\boldsymbol{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0}'_{T-p-1} \\ \boldsymbol{I}_{T-p-1} & \mathbf{0}_{T-p-1} \end{bmatrix},$$

where $\mathbf{0}_h$ is the h-dimensional vector of zero. The lag ℓ residual cross-correlation matrix is defined as

$$\hat{\boldsymbol{R}}_\ell = \hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{C}}_\ell\hat{\boldsymbol{D}}^{-1}, \tag{2.56}$$

where $\hat{\boldsymbol{D}}$ is the diagonal matrix of the standard errors of the residual series, that is, $\hat{\boldsymbol{D}} = \sqrt{\text{diag}(\hat{\boldsymbol{C}}_0)}$. In particular, $\hat{\boldsymbol{R}}_0$ is the residual correlation matrix.

Residual cross-covariance matrices are useful tools for model checking, so we study next their limiting properties. To this end, we consider the asymptotic joint distribution of the residual cross-covariance matrices $\hat{\boldsymbol{\Xi}}_m = [\hat{\boldsymbol{C}}_1, \ldots, \hat{\boldsymbol{C}}_m]$. Using the notation in Equation (2.55), we have

$$\hat{\boldsymbol{\Xi}}_m = \frac{1}{T-p}\hat{\boldsymbol{A}}'\left[\boldsymbol{B}\hat{\boldsymbol{A}}, \boldsymbol{B}^2\hat{\boldsymbol{A}}, \ldots, \boldsymbol{B}^m\hat{\boldsymbol{A}}\right] = \frac{1}{T-p}\hat{\boldsymbol{A}}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \hat{\boldsymbol{A}}), \tag{2.57}$$

where $\boldsymbol{B}_m = [\boldsymbol{B}, \boldsymbol{B}^2, \ldots, \boldsymbol{B}^m]$ is a $(T-p) \times m(T-p)$ matrix. Next, from Equation (2.28), we have

$$\hat{\boldsymbol{A}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{A} - \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Therefore, via Equation (2.57) and letting $T_p = T - p$, we have

$$\begin{aligned} T_p\hat{\boldsymbol{\Xi}}_m &= \boldsymbol{A}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \boldsymbol{A}) - \boldsymbol{A}'\boldsymbol{B}_m\left[\boldsymbol{I}_m \otimes \boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] \\ &\quad - \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{X}'\boldsymbol{B}_m\left(\boldsymbol{I}_m \otimes \boldsymbol{A}\right) + \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{X}'\boldsymbol{B}_m\left[\boldsymbol{I}_m \otimes \boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right]. \end{aligned} \tag{2.58}$$

We can use Equation (2.58) to study the limiting distribution of $\hat{\boldsymbol{\Xi}}_m$. Adopting an approach similar to that of Lütkepohl (2005), we divide the derivations into several steps.

Lemma 2.4 Suppose that $\boldsymbol{z}_t$ follows a stationary VAR(p) model of Equation (2.21) with $\boldsymbol{a}_t$ being a white noise process with mean zero and positive covariance matrix $\boldsymbol{\Sigma}_a$. Also, assume that the assumption in Equation (2.42) holds and the parameter matrix $\boldsymbol{\beta}$ of the model in Equation (2.28) is consistently estimated by a method discussed in Section 2.5 and the residual cross-covariance matrix is defined in Equation (2.55). Then, $\sqrt{T_p}\text{vec}(\hat{\boldsymbol{\Xi}}_m)$ has the same limiting distribution as $\sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m) - \sqrt{T_p}\boldsymbol{H}\text{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']$, where $T_p = T - p$, $\boldsymbol{\Xi}_m$ is the theoretical counterpart of $\hat{\boldsymbol{\Xi}}_m$ obtained by dividing the first term of Equation (2.58) by T_p, and $\boldsymbol{H} = \boldsymbol{H}_*' \otimes \boldsymbol{I}_k$ with

$$\boldsymbol{H}_* = \begin{bmatrix} \boldsymbol{0}' & \boldsymbol{0}' & \cdots & \boldsymbol{0}' \\ \boldsymbol{\Sigma}_a & \boldsymbol{\psi}_1\boldsymbol{\Sigma}_a & \cdots & \boldsymbol{\psi}_{m-1}\boldsymbol{\Sigma}_a \\ \boldsymbol{0}_k & \boldsymbol{\Sigma}_a & \cdots & \boldsymbol{\psi}_{m-2}\boldsymbol{\Sigma}_a \\ \vdots & \vdots & & \vdots \\ \boldsymbol{0}_k & \boldsymbol{0}_k & \cdots & \boldsymbol{\psi}_{m-p}\boldsymbol{\Sigma}_a \end{bmatrix}_{(kp+1)\times km},$$

where $\boldsymbol{0}$ is a k-dimensional vector of zero, $\boldsymbol{0}_k$ is a $k \times k$ matrix of zero, and $\boldsymbol{\psi}_i$ are the coefficient matrices of the MA representation of the VAR(p) model in Equation (2.27).

Proof of Lemma 2.4 is given in Section 2.12. The first term of Equation (2.58) is

$$\boldsymbol{A}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \boldsymbol{A}) = \left[\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A}, \boldsymbol{A}'\boldsymbol{B}^2\boldsymbol{A}, \ldots, \boldsymbol{A}'\boldsymbol{B}^m\boldsymbol{A}\right] \equiv T_p\boldsymbol{\Xi}_m.$$

It is then easy to see that

$$\sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m) \to_d N(\boldsymbol{0}, \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a).$$

In fact, by using part (ii) of Lemma 2.3 and direct calculation, we have the following result. Details can be found in Ahn (1988).

Lemma 2.5 Assume that $\boldsymbol{z}_t$ is a stationary VAR(p) series satisfying the conditions of Lemma 2.4, then

$$\begin{bmatrix} \frac{1}{T_p}\text{vec}(\boldsymbol{A}'\boldsymbol{X}) \\ \sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m) \end{bmatrix} \to_d N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{G} & \boldsymbol{H}_* \\ \boldsymbol{H}_*' & \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a \end{bmatrix} \otimes \boldsymbol{\Sigma}_a\right),$$

where $\boldsymbol{G}$ is defined in Lemma 2.3 and $\boldsymbol{H}_*$ is defined in Lemma 2.4.

Using Lemmas 2.4 and 2.5, we can obtain the limiting distribution of the cross-covariance matrices of a stationary VAR(p) time series.

Theorem 2.4 Suppose that $\boldsymbol{z}_t$ follows a stationary VAR(p) model of Equation (2.21) with $\boldsymbol{a}_t$ being a white noise process with mean zero and positive covariance matrix $\boldsymbol{\Sigma}_a$. Also, assume that the assumption in Equation (2.42) holds and the parameter matrix $\boldsymbol{\beta}$ of the model in Equation (2.28) is consistently estimated by a method discussed in Section 2.5 and the residual cross-covariance matrix is defined in Equation (2.55). Then,

$$\sqrt{T_p}\text{vec}(\hat{\boldsymbol{\Xi}}_m) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{c,m}),$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{c,m} &= (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a - \boldsymbol{H}_*'\boldsymbol{G}^{-1}\boldsymbol{H}_*) \otimes \boldsymbol{\Sigma}_a \\ &= \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a - \tilde{\boldsymbol{H}}[(\boldsymbol{\Gamma}_0^*)^{-1} \otimes \boldsymbol{\Sigma}_a]\tilde{\boldsymbol{H}}',\end{aligned}$$

where $\boldsymbol{H}_*$ and $\boldsymbol{G}$ are defined in Lemma 2.5, $\boldsymbol{\Gamma}_0^*$ is the expanded covariance matrix defined in Equation (2.26), and $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{H}}_* \otimes \boldsymbol{I}_k$ with $\tilde{\boldsymbol{H}}_*$ being a submatrix of $\boldsymbol{H}_*$ with the first row of zeros removed.

Proof. Using Lemma 2.4, the limiting distribution of $\sqrt{T_p}\text{vec}(\hat{\boldsymbol{\Xi}}_m)$ can be obtained by considering

$$\begin{aligned}&\sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m) - \sqrt{T_p}\text{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\ &= [-\boldsymbol{H}, \boldsymbol{I}_{mk^2}]\begin{bmatrix}\sqrt{T_p}\text{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\ \sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m)\end{bmatrix} \\ &= [-\boldsymbol{H}_*' \otimes \boldsymbol{I}_k, \boldsymbol{I}_{mk^2}]\begin{bmatrix}\left(\dfrac{\boldsymbol{X}'\boldsymbol{X}}{\boldsymbol{T_p}}\right)^{-1} \otimes \boldsymbol{I}_k & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{I}_{mk^2}\end{bmatrix}\begin{bmatrix}\dfrac{1}{T_p}\text{vec}(\boldsymbol{A}'\boldsymbol{X}) \\ \sqrt{T_p}\text{vec}(\boldsymbol{\Xi}_m)\end{bmatrix},\end{aligned}$$

where $\mathbf{0}$ is a $k(kp + 1) \times mk^2$ matrix of zero, and we have used $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{A}$ and properties of vec operator. Since $\boldsymbol{X}'\boldsymbol{X}/T_p$ converges to the nonsingular matrix $\boldsymbol{G}$ defined in Lemma 2.3, we can apply Lemma 2.5 and properties of multivariate normal distribution to complete the proof. Specifically, the first two factors of the prior equation converge to

$$[-\boldsymbol{H}_*' \otimes \boldsymbol{I}_k, \boldsymbol{I}_{mk^2}]\begin{bmatrix}\boldsymbol{G}^{-1} \otimes \boldsymbol{I}_k & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{I}_{mk^2}\end{bmatrix} = [-\boldsymbol{H}_*'\boldsymbol{G}^{-1} \otimes \boldsymbol{I}_k, \boldsymbol{I}_{mk^2}],$$

and we have

$$\begin{aligned}&[-\boldsymbol{H}_*'\boldsymbol{G}^{-1} \otimes \boldsymbol{I}_k, \boldsymbol{I}_{mk^2}]\left\{\begin{bmatrix}\boldsymbol{G} & \boldsymbol{H}_* \\ \boldsymbol{H}_*' & \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a\end{bmatrix} \otimes \boldsymbol{\Sigma}_a\right\}\begin{bmatrix}-\boldsymbol{G}^{-1}\boldsymbol{H}_* \otimes \boldsymbol{I}_k \\ \boldsymbol{I}_{mk^2}\end{bmatrix} \\ &= (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a - \boldsymbol{H}_*'\boldsymbol{G}^{-1}\boldsymbol{H}_*) \otimes \boldsymbol{\Sigma}_a \\ &= \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a - (\boldsymbol{H}_*' \otimes \boldsymbol{I}_k)(\boldsymbol{G}^{-1} \otimes \boldsymbol{\Sigma}_a)(\boldsymbol{H}_* \otimes \boldsymbol{I}_k) \\ &= \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_a \otimes \boldsymbol{\Sigma}_a - \tilde{\boldsymbol{H}}[(\boldsymbol{\Gamma}_0^*)^{-1} \otimes \boldsymbol{\Sigma}_a]\tilde{\boldsymbol{H}}',\end{aligned}$$

where the last equality holds because the first row of $\boldsymbol{H}_*$ is 0. □

Comparing results of Lemma 2.5 and Theorem 2.4, we see that the asymptotic variances of elements of the cross-covariance matrices of residuals $\hat{\boldsymbol{a}}_t$ are less than or equal to those of elements of the cross-covariance matrices of the white noise series $\boldsymbol{a}_t$. This seems counter-intuitive, but the residuals are, strictly speaking, not independent.

Let $\boldsymbol{D}$ be the diagonal matrix of the standard errors of the components of $\boldsymbol{a}_t$, that is, $\boldsymbol{D}=\text{diag}\{\sqrt{\sigma_{11,a}},\ldots,\sqrt{\sigma_{kk,a}}\}$, where $\boldsymbol{\Sigma}_a=[\sigma_{ij,a}]$. We can apply Theorem 2.4 to obtain the limiting distribution of the cross-correlation matrices $\hat{\boldsymbol{\xi}}_m=[\hat{\boldsymbol{R}}_1,\ldots,\hat{\boldsymbol{R}}_m]$, where the cross-correlation matrix $\hat{\boldsymbol{R}}_j$ is defined in Equation (2.56).

Theorem 2.5 Assume that the conditions of Theorem 2.4 hold. Then,

$$\sqrt{T_p}\text{vec}(\hat{\boldsymbol{\xi}}_m) \to_d N(\boldsymbol{0},\boldsymbol{\Sigma}_{r,m}),$$

where $\boldsymbol{\Sigma}_{r,m}=[(\boldsymbol{I}_m\otimes\boldsymbol{R}_0)-\boldsymbol{H}_0'\boldsymbol{G}^{-1}\boldsymbol{H}_0]\otimes\boldsymbol{R}_0$, where $\boldsymbol{R}_0$ is the lag-0 cross-correlation matrix of $\boldsymbol{a}_t$, $\boldsymbol{H}_0=\boldsymbol{H}_*(\boldsymbol{I}_m\otimes\boldsymbol{D}^{-1})$, and $\boldsymbol{G}$, as before, is defined in Lemma 2.3.

Proof. Let $\hat{\boldsymbol{D}}^{-1}$ be the diagonal matrix defined in Equation (2.56). It is easy to see that $\hat{\boldsymbol{D}}^{-1}$ is a consistent estimate of $\boldsymbol{D}^{-1}$. We can express the statistic of interest as

$$\begin{aligned}\text{vec}(\hat{\boldsymbol{\xi}}_m) &= \text{vec}\left[\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\Xi}}_m\left(\boldsymbol{I}_m\otimes\hat{\boldsymbol{D}}^{-1}\right)\right]\\ &= \left(\boldsymbol{I}_m\otimes\hat{\boldsymbol{D}}^{-1}\otimes\hat{\boldsymbol{D}}^{-1}\right)\text{vec}(\hat{\boldsymbol{\Xi}}_m).\end{aligned}$$

Applying the result of Theorem 2.4, we obtain that $\sqrt{T_p}\text{vec}(\hat{\boldsymbol{\xi}}_m)$ follows asymptotically a multivariate normal distribution with mean zero and covariance matrix

$$\begin{aligned}&\left(\boldsymbol{I}_m\otimes\boldsymbol{D}^{-1}\otimes\boldsymbol{D}^{-1}\right)\left[\left(\boldsymbol{I}_m\otimes\boldsymbol{\Sigma}_a-\boldsymbol{H}_*'\boldsymbol{G}^{-1}\boldsymbol{H}_*\right)\otimes\boldsymbol{\Sigma}_a\right]\left(\boldsymbol{I}_m\otimes\boldsymbol{D}^{-1}\otimes\boldsymbol{D}^{-1}\right)\\ &\quad=\left[(\boldsymbol{I}_m\otimes\boldsymbol{R}_0)-\boldsymbol{H}_0'\boldsymbol{G}^{-1}\boldsymbol{H}_0\right]\otimes\boldsymbol{R}_0,\end{aligned}$$

where we have used $\boldsymbol{R}_0=\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_a\boldsymbol{D}^{-1}$. □

Based on Theorem 2.5, we can obtain the limiting distribution of the lag j cross-correlation matrix $\hat{\boldsymbol{R}}_j$. Specifically,

$$\sqrt{T_p}\text{vec}(\hat{\boldsymbol{R}}_j) \to_d N(\boldsymbol{0},\boldsymbol{\Sigma}_{r,(j)}),$$

where

$$\boldsymbol{\Sigma}_{r,(j)}=\left[\boldsymbol{R}_0-\boldsymbol{D}^{-1}\boldsymbol{\Sigma}_a\boldsymbol{\Psi}_p\boldsymbol{G}^{-1}\boldsymbol{\Psi}_p'\boldsymbol{\Sigma}_a\boldsymbol{D}^{-1}\right]\otimes\boldsymbol{R}_0,$$

where $\boldsymbol{\Psi}_p = [\mathbf{0}, \boldsymbol{\psi}'_{j-1}, \ldots, \boldsymbol{\psi}'_{j-p}]$ with $\boldsymbol{\psi}_\ell$ being the coefficient matrices of the MA representation of $\boldsymbol{z}_t$ so that $\boldsymbol{\psi}_\ell = \mathbf{0}$ for $\ell < 0$ and $\mathbf{0}$ is a k-dimensional vector of zero.

2.7.2 Multivariate Portmanteau Statistics

Let $\boldsymbol{R}_\ell$ be the theoretical lag ℓ cross-correlation matrix of innovation $\boldsymbol{a}_t$. The hypothesis of interest in model checking is

$$H_0 : \boldsymbol{R}_1 = \cdots = \boldsymbol{R}_m = \mathbf{0} \quad \text{versus} \quad H_a : \boldsymbol{R}_j \neq \mathbf{0} \quad \text{for some } 1 \leq j \leq m, \tag{2.59}$$

where m is a prespecified positive integer. The Portmanteau statistic of Equation (1.11) is often used to perform the test. For residual series, the statistic becomes

$$\begin{aligned} Q_k(m) &= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{R}}'_\ell \hat{\boldsymbol{R}}_0^{-1} \hat{\boldsymbol{R}}_\ell \hat{\boldsymbol{R}}_0^{-1}\right) \\ &= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{R}}'_\ell \hat{\boldsymbol{R}}_0^{-1} \hat{\boldsymbol{R}}_\ell \hat{\boldsymbol{R}}_0^{-1} \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{D}}\right) \\ &= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{D}} \hat{\boldsymbol{R}}'_\ell \hat{\boldsymbol{D}} \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{R}}_0^{-1} \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{D}} \hat{\boldsymbol{R}}_\ell \hat{\boldsymbol{D}} \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{R}}_0^{-1} \hat{\boldsymbol{D}}^{-1}\right) \\ &= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{C}}'_\ell \hat{\boldsymbol{C}}_0^{-1} \hat{\boldsymbol{C}}_\ell \hat{\boldsymbol{C}}_0^{-1}\right). \end{aligned} \tag{2.60}$$

Theorem 2.6 Suppose that $\boldsymbol{z}_t$ follows a stationary VAR(p) model of Equation (2.21) with $\boldsymbol{a}_t$ being a white noise process with mean zero and positive covariance matrix $\boldsymbol{\Sigma}_a$. Also, assume that the assumption in Equation (2.42) holds and the parameter matrix $\boldsymbol{\beta}$ of the model in Equation (2.28) is consistently estimated by a method discussed in Section 2.5 and the residual cross-covariance matrix is defined in Equation (2.55). Then, the test statistic $Q_k(m)$ is asymptotically distributed as a chi-square distribution with $(m-p)k^2$ degrees of freedom.

The proof of Theorem 2.6 is relatively complicated. Readers may consult Li and McLeod (1981), Hosking (1981), and Lütkepohl (2005) for further information. Compared with the Portmanteau test of Chapter 1, the degrees of freedom of the chi-square distribution in Theorem 2.6 is adjusted by pk^2, which is the number of AR parameters in a VAR(p) model. In practice, some of the AR parameters in a VAR(p) model are fixed to 0. In this case, the adjustment in the degrees of freedom of the chi-square distribution is set to the number of estimated AR parameters.

In the literature, some Lagrange multiplier tests have also been suggested for checking a fitted VAR model. However, the asymptotic chi-square distribution of the Lagrange multiplier test is found to be a poor approximation. See, for instance, Edgerton and Shukur (1999). For this reason, we shall not discuss the Lagrange multiplier tests.

Example 2.6 Consider again the quarterly GDP growth rates, in percentages, of United Kingdom, Canada, and United States employed in Example 2.3. We apply the multivariate Portmanteau test statistics to the residuals of the fitted VAR(2) model. The results are given in the following R demonstration. Since we used $p = 2$, the $Q_2(m)$ statistics requires $m > 2$ to have positive degrees of freedom for the asymptotic chi-square distribution. For this reason, the p-values are set to 1 for $m = 1$ and 2. From the demonstration, it is seen that the fitted VAR(2) model has largely removed the dynamic dependence in the GDP growth rates, except for some minor violation at $m = 4$. We shall return to this point later as there exist some insignificant parameter estimates in the fitted VAR(2) model. □

R Demonstration: Multivariate Portmanteau statistics.

```
> names(m1)
 [1] "data"      "cnst"    "order"   "coef"    "aic"     "bic"
 [7] "residuals" "secoef"  "Sigma"   "Phi"     "Ph0"
> resi=m1$residuals  ### Obtain the residuals of VAR(2) fit.
> mq(resi,adj=18) ## adj is used to adjust the degrees of
freedom. Ljung-Box Statistics:
         m      Q(m)   p-value
 [1,]   1.000    0.816    1.00
 [2,]   2.000    3.978    1.00
 [3,]   3.000   16.665    0.05
 [4,]   4.000   35.122    0.01
 [5,]   5.000   38.189    0.07
 [6,]   6.000   41.239    0.25
 [7,]   7.000   47.621    0.37
 [8,]   8.000   61.677    0.22
 [9,]   9.000   67.366    0.33
[10,]  10.000   76.930    0.32
[11,]  11.000   81.567    0.46
[12,]  12.000   93.112    0.39
```

2.7.3 Model Simplification

Multivariate time series models may contain many parameters if the dimension k is moderate or large. In practice, we often observe that some of the parameters are not statistically significant at a given significant level. It is then advantageous to simplify the model by removing the insignificant parameters. This is particularly so when no prior knowledge is available to support those parameters. However, there exists no optimal method to simplify a fitted model. We discuss some methods commonly used in practice.

2.7.3.1 Testing Zero Parameters

An obvious approach to simplify a fitted VAR(p) model is to remove insignificant parameters. Given a specified significant level, for example, $\alpha = 0.05$, we can identify the *target parameters* for removal. By target parameters we meant those parameters whose individual t-ratio is less than the critical value of the normal distribution with type I error α. They are the target for removal because parameter estimates are correlated and marginal statistics could be misleading. To confirm that those parameters can indeed be removed, we consider a test procedure using the limiting distribution of Theorem 2.3.

Let $\hat{\boldsymbol{\omega}}$ be a v-dimensional vector consisting of the target parameters. In other words, v is the number of parameters to be fixed to 0. Let $\boldsymbol{\omega}$ be the counterpart of $\hat{\boldsymbol{\omega}}$ in the parameter matrix $\boldsymbol{\beta}$ in Equation (2.28). The hypothesis of interest is

$$H_0 : \boldsymbol{\omega} = \boldsymbol{0} \quad \text{versus} \quad H_a : \boldsymbol{\omega} \neq \boldsymbol{0}.$$

Clearly, there exists a $v \times k(kp+1)$ locating matrix $\boldsymbol{K}$ such that

$$\boldsymbol{K}\text{vec}(\boldsymbol{\beta}) = \boldsymbol{\omega}, \quad \text{and} \quad \boldsymbol{K}\text{vec}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\omega}}. \tag{2.61}$$

By Theorem 2.3 and properties of multivariate normal distribution, we have

$$\sqrt{T_p}\,(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \to_d N\left[\boldsymbol{0}, \boldsymbol{K}\left(\boldsymbol{\Sigma}_a \otimes \boldsymbol{G}^{-1}\right)\boldsymbol{K}'\right], \tag{2.62}$$

where $T_p = T - p$ is the effective sample size. Consequently, under H_0, we have

$$T_p\hat{\boldsymbol{\omega}}'\left[\boldsymbol{K}\left(\boldsymbol{\Sigma}_a \otimes \boldsymbol{G}^{-1}\right)\boldsymbol{K}'\right]^{-1}\hat{\boldsymbol{\omega}} \to_d \chi^2_v, \tag{2.63}$$

where $v = \dim(\boldsymbol{\omega})$. This chi-square test can also be interpreted as a likelihood-ratio test under the normality assumption of $\boldsymbol{a}_t$. The null hypothesis H_0: $\boldsymbol{\omega} = \boldsymbol{0}$ denotes a reduced VAR(p) model. Therefore, one can use likelihood-ratio statistic, which is asymptotically equivalent to the chi-square test of Equation (2.63).

If the hypothesis of interest is

$$H_0 : \boldsymbol{\omega} = \boldsymbol{\omega}_o \quad \text{versus} \quad H_a : \boldsymbol{\omega} \neq \boldsymbol{\omega}_o,$$

where $\boldsymbol{\omega}_o$ is a prespecified v-dimensional vector, and we replace $\boldsymbol{\Sigma}_a$ and $\boldsymbol{G}$ by their estimates, then the test statistic in Equation (2.63) becomes

$$\begin{aligned}\lambda_W &= T_p(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_o)'\left[\boldsymbol{K}\left(\hat{\boldsymbol{\Sigma}}_a \otimes \hat{\boldsymbol{G}}^{-1}\right)\boldsymbol{K}'\right]^{-1}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_o) \\ &= (\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_o)'\left[\boldsymbol{K}\left\{\hat{\boldsymbol{\Sigma}}_a \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right\}\boldsymbol{K}'\right]^{-1}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_o). \end{aligned} \tag{2.64}$$

This is a *Wald statistic* and is asymptotically distributed as χ^2_v, provided that the assumptions of Theorem 2.3 hold so that $\hat{\boldsymbol{\Sigma}}_a$ and $\hat{\boldsymbol{G}} = \boldsymbol{X}'\boldsymbol{X}/T_p$ are consistent. This Wald test can be used to make inference such as testing for Granger causality.

Example 2.7 Consider again the quarterly GDP growth rates, in percentages, of United Kingdom, Canada, and United States employed in Example 2.3. Based on the R demonstration of Example 2.3, there are insignificant parameters in a fitted VAR(2) model. To simplify the model, we apply the chi-square test of Equation (2.63). We used type I error $\alpha = 0.05$ and $\alpha = 0.1$, respectively, to identify target parameters. The corresponding critical values are 1.96 and 1.645, respectively. It turns out that there are 10 and 8 possible zero parameters for these two choices of α. The chi-square test for all 10 targeted parameters being 0 is 31.69 with p-value 0.0005. Thus, we cannot simultaneously set all 10 parameters with the smallest absolute t-ratios to 0. On the other hand, the chi-square test is 15.16 with p-value 0.056 when we test that all eight parameters with the smallest t-ratios, in absolute value, are 0. Consequently, we can simplify the fitted VAR(2) model by letting the eight parameters with smallest t-ratio (in absolute) to 0. □

Remark: To perform the chi-square test for zero parameters, we use the command `VARchi` in the `MTS` package. The subcommand `thres` of `VARchi` can be used to set type I error for selecting the target parameters. The default is `thres = 1.645`. □

R Demonstration: Testing zero parameters.

```
> m3=VARchi(z,p=2)
Number of targeted parameters:  8
Chi-square test and p-value:  15.16379 0.05603778
> m3=VARchi(z,p=2,thres=1.96)
Number of targeted parameters:  10
Chi-square test and p-value:  31.68739 0.000451394
```

2.7.3.2 Information Criteria

An alternative approach to the chi-square test is to use the information criteria discussed in Section 2.6. For instance, we can estimate the unconstrained VAR(p) model (under H_a) and the constrained VAR(p) model (under H_0). If the constrained model has a smaller value for a selected criterion, then H_0 cannot be rejected according to that criterion.

2.7.3.3 Stepwise Regression

Finally, we can make use of the fact that for VAR(p) models the estimation can be performed equation-by-equation; see the discussion in Section 2.5. As such, each equation is a multiple linear regression, and we can apply the traditional idea of stepwise regression to remove insignificant parameters. Readers may consult variable selection in multiple linear regression for further information about the stepwise regression.

Remark: We use the command `refVAR` of the `MTS` package to carry out model simplification of a fitted VAR model. The command uses a threshold to select target parameters for removal and computes information criteria of the simplified model for validation. The default threshold is 1.00. The command also allows the user to specify zero parameters with a subcommand `fixed`. □

To demonstrate, we consider again the quarterly growth rates, in percentages, of United Kingdom, Canada, and United States employed in Example 2.7. The simplified VAR(2) model has 12 parameters, instead of 21, for the unconstrained VAR(2) model. The AIC of the simplified model is -3.53, which is smaller than -3.50 of the unconstrained model. For this particular instance, all three criteria have a smaller value for the constrained model. It is possible in practice that different criteria lead to different conclusions.

R Demonstration: Model simplification.

```
> m1=VAR(zt,2)  # fit a un-constrained VAR(2) model.
> m2=refVAR(m1,thres=1.96)  # Model refinement.
Constant term:
Estimates:  0.1628247 0 0.2827525
Std.Error:  0.06814101 0 0.07972864
AR coefficient matrix
AR( 1 )-matrix
      [,1]  [,2]  [,3]
[1,] 0.467 0.207 0.000
[2,] 0.334 0.270 0.496
[3,] 0.468 0.225 0.232
standard error
       [,1]   [,2]   [,3]
[1,] 0.0790 0.0686 0.0000
[2,] 0.0921 0.0875 0.0913
[3,] 0.1027 0.0963 0.1023
AR( 2 )-matrix
       [,1] [,2] [,3]
[1,]  0.000    0    0
[2,] -0.197    0    0
[3,] -0.301    0    0
standard error
       [,1] [,2] [,3]
[1,] 0.0000    0    0
[2,] 0.0921    0    0
[3,] 0.1008    0    0

Residuals cov-mtx:
           [,1]       [,2]       [,3]
[1,] 0.29003669 0.01803456 0.07055856
[2,] 0.01803456 0.30802503 0.14598345
[3,] 0.07055856 0.14598345 0.36268779
```

```
det(SSE) =  0.02494104
AIC =  -3.531241
BIC =  -3.304976
HQ  =  -3.439321
```

From the output, the simplified VAR(2) model for the percentage growth rates of quarterly GDP of United Kingdom, Canada, and United States is

$$\boldsymbol{z}_t = \begin{bmatrix} 0.16 \\ - \\ 0.28 \end{bmatrix} + \begin{bmatrix} 0.47 & 0.21 & - \\ 0.33 & 0.27 & 0.50 \\ 0.47 & 0.23 & 0.23 \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} - & - & - \\ -0.20 & - & - \\ -0.30 & - & - \end{bmatrix} \boldsymbol{z}_{t-2} + \boldsymbol{a}_t, \tag{2.65}$$

where the residual covariance matrix is

$$\hat{\boldsymbol{\Sigma}}_a = \begin{bmatrix} 0.29 & 0.02 & 0.07 \\ 0.02 & 0.31 & 0.15 \\ 0.07 & 0.15 & 0.36 \end{bmatrix}.$$

All estimates are now significant at the usual 5% level. Limiting properties of the constrained parameter estimates are discussed in the next section.

Finally, we return to model checking for the simplified VAR(2) model in Equation (2.65). Since all estimates are statistically significant at the usual 5% level, we perform a careful residual analysis. Figure 2.3 shows the time plots of the three residual series, whereas Figure 2.4 shows the residual cross-correlation matrices. The dashed lines of the plots in Figure 2.4 indicate the approximate 2 standard-error limits of the cross-correlations, that is $\pm 2/\sqrt{T}$. Strictly speaking, these limits are only valid for higher-order lags; see Theorem 2.5. However, it is relatively complicated to compute the asymptotic standard errors of the cross-correlations so that the simple approximation is often used in model checking. Based on the plots, the residuals of the model in Equation (2.65) do not have any strong serial or cross-correlations. The only possible exception is the $(1, 3)$th position of lag-4 cross-correlation matrix.

The multivariate Portmanteau test can be used to check the residual cross-correlations. The results are given in the following R demonstration. Figure 2.5 plots the p-values of the $Q_3(m)$ statistics applied to the residuals of the simplified VAR(2) model in Equation (2.65). Since there are 12 parameters, the degrees of freedom of the chi-square distribution for $Q_3(m)$ is $9m - 12$. Therefore, the asymptotic chi-square distribution works if $m \geq 2$. From the plot and the R demonstration, the $Q_k(m)$ statistics indicate that there are no strong serial or cross-correlations in the residuals of the simplified VAR(2) model. Based on $Q_k(4)$, one may consider some further improvement, such as using a VAR(4) model. We leave it as an exercise.

Remark: Model checking of a fitted VARMA model can be carried out by the command `MTSdiag` of the `MTS` package. The default option checks 24 lags of cross-correlation matrix. □

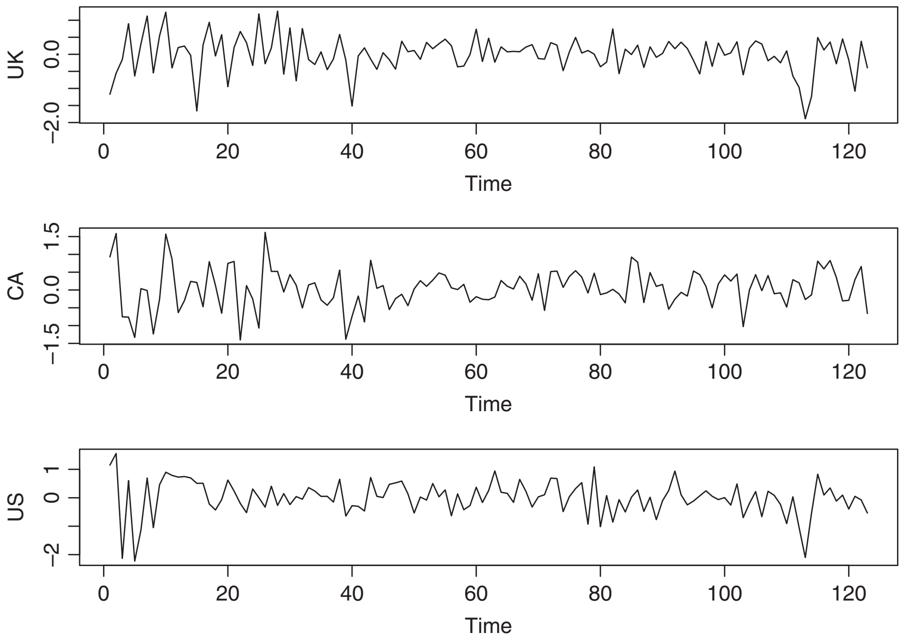

FIGURE 2.3 Residual plots of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages.

R Demonstration: Model checking.

```
> MTSdiag(m2,adj=12)
[1] "Covariance matrix:"
       [,1]   [,2]   [,3]
[1,] 0.2924 0.0182 0.0711
[2,] 0.0182 0.3084 0.1472
[3,] 0.0711 0.1472 0.3657
CCM at lag:  0
       [,1]   [,2]  [,3]
[1,] 1.0000 0.0605 0.218
[2,] 0.0605 1.0000 0.438
[3,] 0.2175 0.4382 1.000
Simplified matrix:
CCM at lag:  1
. . .
. . .
. . .
CCM at lag:  2
. . .
. . .
. . .
```

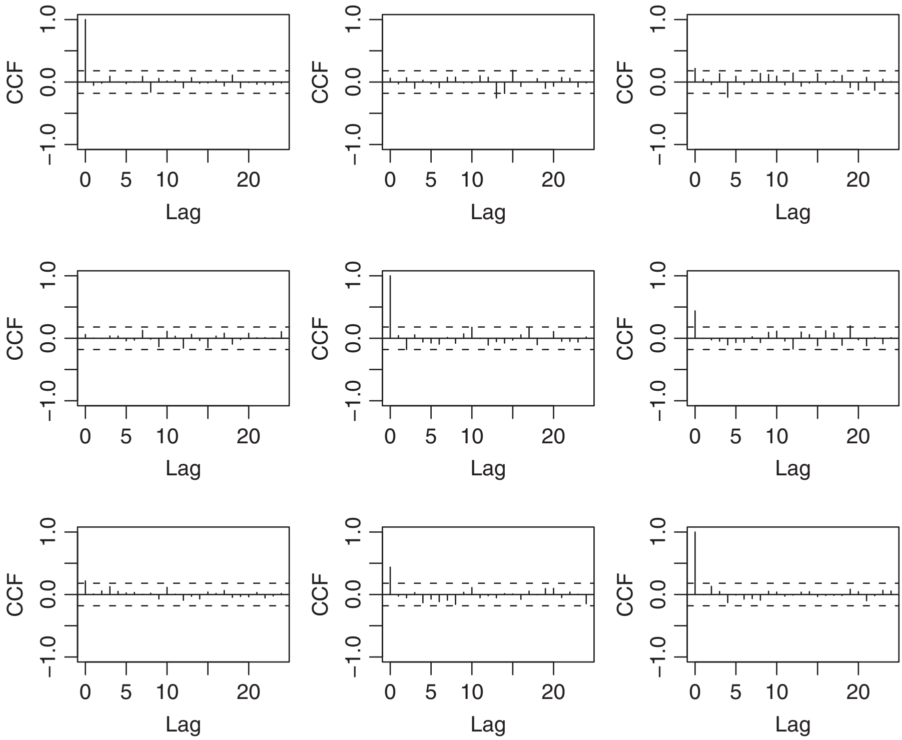

FIGURE 2.4 Residual cross-correlation matrices of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages.

```
CCM at lag:  3
. . .
. . .
. . .
CCM at lag:  4
. . -
. . .
. . .
Hit Enter to compute MQ-statistics:

Ljung-Box Statistics:
         m       Q(m)   p-value
 [1,]   1.00      1.78     1.00
 [2,]   2.00     12.41     0.05
 [3,]   3.00     22.60     0.09
 [4,]   4.00     37.71     0.04
 [5,]   5.00     41.65     0.14
 [6,]   6.00     44.95     0.35
 [7,]   7.00     51.50     0.45
```

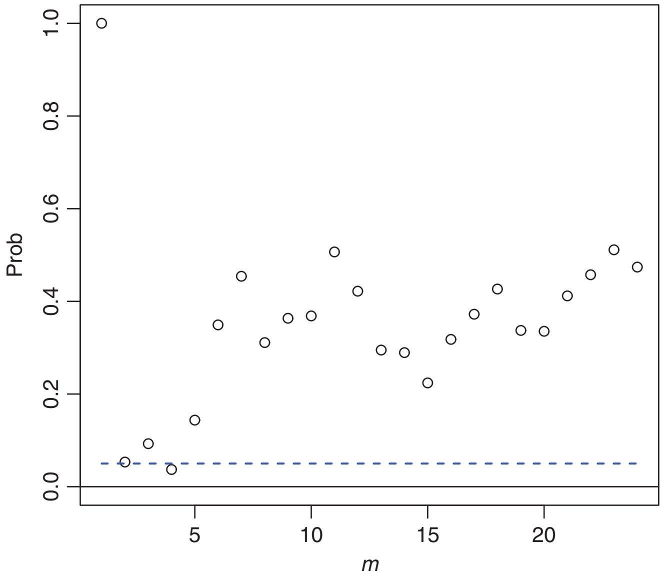

FIGURE 2.5 Plot of p-values of the $Q_k(m)$ statistics applied to the residuals of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages.

```
 [8,]   8.00     64.87     0.31
 [9,]   9.00     72.50     0.36
[10,]  10.00     81.58     0.37
[11,]  11.00     86.12     0.51
[12,]  12.00     98.08     0.42
```

In conclusion, the simplified VAR(2) model in Equation (2.65) is adequate for the GDP growth rate series. The model can be written as

$$\begin{aligned}
\text{UK}: &\quad z_{1t} = 0.16 + 0.47z_{1,t-1} + 0.21z_{2,t-1} + a_{1t},\\
\text{CA}: &\quad z_{2t} = 0.33z_{1,t-1} + 0.27z_{2,t-1} + 0.5z_{3,t-1} - 0.2z_{1,t-2} + a_{2t},\\
\text{US}: &\quad z_{3t} = 0.28 + 0.47z_{1,t-1} + 0.23z_{2,t-1} + 0.23z_{3,t-1} - 0.3z_{1,t-2} + a_{3t}.
\end{aligned}$$

The correlation matrix of the residuals is

$$\boldsymbol{R}_0 = \begin{bmatrix} 1.00 & 0.06 & 0.22 \\ 0.06 & 1.00 & 0.44 \\ 0.22 & 0.44 & 1.00 \end{bmatrix}.$$

This correlation matrix indicates that the quarterly GDP growth rates of United Kingdom and Canada are not instantaneously correlated. The fitted three-dimensional model shows that the GDP growth rate of United Kingdom does not depend on the lagged growth rates of the United States in the presence of lagged Canadian GDP growth rates, but the United Kingdom growth rate depends on the past growth rate of Canada. On the other hand, the GDP growth rate of Canada is dynamically related to the growth rates of United Kingdom and United States. Similarly, the GDP growth rate of the United States depends on the lagged growth rates of United Kingdom and Canada. If one further considers the dependence of z_{1t} on $z_{3,t-4}$, then all three GDP growth rates are directly dynamically correlated. In summary, the simplified VAR(2) model indicates that the GDP growth rate of United Kingdom is conditionally independent of the growth rate of the United States given the Canadian growth rate.

2.8 LINEAR CONSTRAINTS

Linear parameter constraints can be handled easily in estimation of a VAR(p) model. Consider the matrix representation in Equation (2.28). Any linear parameter constraint can be expressed as

$$\text{vec}(\boldsymbol{\beta}) = \boldsymbol{J}\boldsymbol{\gamma} + \boldsymbol{r}, \tag{2.66}$$

where $\boldsymbol{J}$ is $k(kp+1) \times P$ constant matrix of rank P, $\boldsymbol{r}$ is a $k(kp+1)$-dimensional constant vector, and $\boldsymbol{\gamma}$ denotes a P-dimensional vector of unknown parameters. Here $\boldsymbol{J}$ and $\boldsymbol{r}$ are known. For example, consider the two-dimensional VAR(1) model, $\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t$. Here we have

$$\boldsymbol{\beta} = \begin{bmatrix} \phi_{0,1} & \phi_{0,2} \\ \phi_{1,11} & \phi_{1,21} \\ \phi_{1,12} & \phi_{1,22} \end{bmatrix}, \quad \text{where} \quad \boldsymbol{\phi}_0 = \begin{bmatrix} \phi_{0,1} \\ \phi_{0,2} \end{bmatrix}, \quad \boldsymbol{\phi}_1 = [\phi_{1,ij}].$$

Suppose the actual model is

$$\boldsymbol{z}_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.6 & 0 \\ 0.2 & 0.8 \end{bmatrix} \boldsymbol{z}_{t-1} + \boldsymbol{a}_t,$$

which has four parameters. In this particular case, we have $\boldsymbol{r} = \boldsymbol{0}_6$ and

$$\text{vec}(\boldsymbol{\beta}) = \begin{bmatrix} \phi_{0,1} \\ \phi_{1,11} \\ \phi_{1,12} \\ \phi_{0,2} \\ \phi_{1,21} \\ \phi_{1,22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.0 \\ 0.6 \\ 0.2 \\ 0.8 \end{bmatrix} \equiv \boldsymbol{J}\boldsymbol{\gamma}.$$

Under the linear constraints in Equation (2.66), Equation (2.29) becomes

$$\begin{aligned}\text{vec}(\boldsymbol{Z}) &= (\boldsymbol{I}_k \otimes \boldsymbol{X})(\boldsymbol{J}\boldsymbol{\gamma} + \boldsymbol{r}) + \text{vec}(\boldsymbol{A})\\ &= (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{J}\boldsymbol{\gamma} + (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{r} + \text{vec}(\boldsymbol{A}).\end{aligned}$$

Since $\boldsymbol{r}$ is known, $(\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{r}$ is a known $k(T-p) \times 1$ vector. Therefore, we can rewrite the prior equation as

$$\text{vec}(\boldsymbol{Z}_*) = (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{J}\boldsymbol{\gamma} + \text{vec}(\boldsymbol{A}),$$

where $\text{vec}(\boldsymbol{Z}_*) = \text{vec}(\boldsymbol{Z}) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{r}$. Following the same argument as the VAR(p) estimation in Equation (2.30), the GLS estimate of $\boldsymbol{\gamma}$ is obtained by minimizing

$$\begin{aligned}S(\boldsymbol{\gamma}) &= [\text{vec}(\boldsymbol{A})]'(\boldsymbol{\Sigma}_a \otimes \boldsymbol{I}_{T-p})^{-1}\text{vec}(\boldsymbol{A})\\ &= [\text{vec}(\boldsymbol{Z}_*) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{J}\boldsymbol{\gamma}]'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{T-p})[\text{vec}(\boldsymbol{Z}_*) - (\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{J}\boldsymbol{\gamma}].\end{aligned} \tag{2.67}$$

Using the same method as that in Equation (2.32), we have

$$\begin{aligned}\hat{\boldsymbol{\gamma}} &= \left[\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right)\boldsymbol{J}\right]^{-1}\left[\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{Z}_*)\right.\\ &= \left[\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right)\boldsymbol{J}\right]^{-1}\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\right)\left[(\boldsymbol{I}_k \otimes \boldsymbol{X})\boldsymbol{J}\boldsymbol{\gamma} + \text{vec}(\boldsymbol{A})\right]\\ &= \boldsymbol{\gamma} + \left[\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\boldsymbol{J}\right]^{-1}\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\right)\text{vec}(\boldsymbol{A})\\ &= \boldsymbol{\gamma} + \left[\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\boldsymbol{J}\right]^{-1}\boldsymbol{J}'\text{vec}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{\Sigma}_a^{-1}\right)\\ &= \boldsymbol{\gamma} + \left[\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\boldsymbol{J}\right]^{-1}\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{kp+1}\right)\text{vec}\left(\boldsymbol{X}'\boldsymbol{A}\right).\end{aligned} \tag{2.68}$$

From Equation (2.68), we have

$$\sqrt{T_p}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \left[\boldsymbol{J}'\left(\boldsymbol{\Sigma}_a^{-1} \otimes \frac{\boldsymbol{X}'\boldsymbol{X}}{T_p}\right)\boldsymbol{J}\right]^{-1}\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{I}_{kp+1})\frac{1}{\sqrt{T_p}}\text{vec}(\boldsymbol{X}'\boldsymbol{A}), \tag{2.69}$$

where, as before, $T_p = T - p$.

Theorem 2.7 Assume that the stationary VAR(p) process $\boldsymbol{z}_t$ satisfies the conditions of Theorem 2.3. Assume further that the parameters of the model satisfy the linear constraints given in Equation (2.66), where $\boldsymbol{J}$ is of rank P, which is the number of coefficient parameters to be estimated. Then, the GLS estimate $\hat{\boldsymbol{\gamma}}$ of Equation (2.68) is a consistent estimate of $\boldsymbol{\gamma}$ and

$$\sqrt{T_p}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \to_d N\left(\boldsymbol{0}, \left[\boldsymbol{J}'(\boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{G})\boldsymbol{J}\right]^{-1}\right),$$

where $T_p = T - p$ and $\boldsymbol{G}$ is the limit of $\boldsymbol{X}'\boldsymbol{X}/T_p$ as $T \to \infty$ and is defined in Lemma 2.3.

Proof. By part (ii) of Lemma 2.3,

$$\frac{1}{\sqrt{T_p}}\text{vec}(\boldsymbol{X}'\boldsymbol{A}) \to_d N(\boldsymbol{0}, \boldsymbol{\Sigma}_a^{-1} \otimes \boldsymbol{G}).$$

The theorem then follows from Equation (2.69). □

2.9 FORECASTING

Let h be the forecast origin, $\ell > 0$ be the forecast horizon, and F_h be the information available at time h (inclusive). We discuss forecasts of a VAR(p) model in this section.

2.9.1 Forecasts of a Given Model

To begin, assume that the VAR model is known, that is, we ignore for a moment that the parameters are estimated. Following the discussion in Section 1.6, the minimum mean-squared error forecast of $\boldsymbol{z}_{h+\ell}$ is simply the conditional expectation of $\boldsymbol{z}_{h+\ell}$ given F_h. For the VAR(p) model in Equation (2.21), the one-step ahead prediction is trivial,

$$\boldsymbol{z}_h(1) = E(\boldsymbol{z}_{h+1}|F_h) = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{h+1-i}.$$

For two-step ahead prediction, we have

$$\begin{aligned}
\boldsymbol{z}_h(2) &= E(\boldsymbol{z}_{h+2}|F_h) \\
&= \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 E(\boldsymbol{z}_{h+1}|F_h) + \sum_{i=2}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{h+2-i} \\
&= \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_h(1) + \sum_{i=2}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{h+2-i}.
\end{aligned}$$

In general, for the ℓ-step ahead forecast, we have

$$\boldsymbol{z}_h(\ell) = E(\boldsymbol{z}_{h+\ell}|F_h) = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_h(\ell - i), \tag{2.70}$$

where it is understood that $\boldsymbol{z}_h(j) = \boldsymbol{z}_{h+j}$ for $j \leq 0$. Thus, the point forecasts of a VAR(p) model can be computed recursively.

Using $\boldsymbol{\phi}_0 = (\boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_1)\boldsymbol{\mu}$, Equation (2.70) can be rewritten as

$$\boldsymbol{z}_h(\ell) - \boldsymbol{\mu} = \sum_{i=1}^{p} \boldsymbol{\phi}_i[\boldsymbol{z}_h(\ell - i) - \boldsymbol{\mu}].$$

The prior equation implies that

$$[\boldsymbol{z}_h(\ell) - \boldsymbol{\mu}] - \sum_{i=1}^{p} \boldsymbol{\phi}_i[\boldsymbol{z}_h(\ell - i) - \boldsymbol{\mu}] = \boldsymbol{0},$$

which can be rewritten as

$$\left(\boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i\right)[\boldsymbol{z}_h(\ell) - \boldsymbol{\mu}] = \boldsymbol{0}, \quad \text{or simply} \quad \boldsymbol{\phi}(B)[\boldsymbol{z}_h(\ell) - \boldsymbol{\mu}] = \boldsymbol{0}, \tag{2.71}$$

where it is understood that the back-shift operator B operates on ℓ as the forecast origin h is fixed. Define the expanded forecast vector $\boldsymbol{Z}_h(\ell)$ as

$$\boldsymbol{Z}_h(\ell) = ([\boldsymbol{z}_h(\ell) - \boldsymbol{\mu}]', [\boldsymbol{z}_h(\ell - 1) - \boldsymbol{\mu}]', \ldots, [\boldsymbol{z}_h(\ell - p + 1) - \boldsymbol{\mu}]')'.$$

Then, Equation (2.71) implies that

$$\boldsymbol{Z}_h(\ell) = \boldsymbol{\Phi}\boldsymbol{Z}_h(\ell - 1), \quad \ell > 1, \tag{2.72}$$

where $\boldsymbol{\Phi}$, defined in Equation (2.23), is the companion matrix of the polynomial matrix $\boldsymbol{\phi}(B) = \boldsymbol{I}_k - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$. By repeated application of Equation (2.72), we have

$$\boldsymbol{Z}_h(\ell) = \boldsymbol{\Phi}^{\ell-1}\boldsymbol{Z}_h(1), \quad \ell > 1. \tag{2.73}$$

For a stationary VAR(p) model, all eigenvalues of $\boldsymbol{\Phi}$ are less than 1 in absolute value. Therefore, $\boldsymbol{\Phi}^j \to \boldsymbol{0}$ as $j \to \infty$. Consequently, we have

$$\boldsymbol{z}_h(\ell) - \boldsymbol{\mu} \to \boldsymbol{0}, \quad \text{as} \quad \ell \to \infty.$$

This says that the stationary VAR(p) process is mean-reverting because its point forecasts converge to the mean of the process as the forecast horizon increases. The speed of mean-reverting is determined by the magnitude of the largest eigenvalue, in modulus, of $\boldsymbol{\Phi}$.

Turn to forecast errors. For ℓ-step ahead forecast, the forecast error is

$$\boldsymbol{e}_h(\ell) = \boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell).$$

To study this forecast error, it is most convenient to use the MA representation of the VAR(p) model,

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

where $\boldsymbol{\mu} = [\boldsymbol{\phi}(1)]^{-1}\boldsymbol{\phi}_0$, $\boldsymbol{\psi}_0 = \boldsymbol{I}_k$, and $\boldsymbol{\psi}_i$ can be obtained recursively via Equation (2.27). As shown in Equation (1.14), the ℓ-step ahead forecast error is

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

Consequently, the covariance matrix of the forecast error is

$$\text{Cov}[\boldsymbol{e}_h(\ell)] = \boldsymbol{\Sigma}_a + \sum_{i=1}^{\ell-1} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_a \boldsymbol{\psi}_i'. \tag{2.74}$$

As the forecast horizon increases, we see that

$$\text{Cov}[\boldsymbol{e}_h(\ell)] \to \boldsymbol{\Sigma}_a + \sum_{i=1}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_a \boldsymbol{\psi}_i' = \text{Cov}(\boldsymbol{z}_t).$$

This is consistent with the mean-reverting of $\boldsymbol{z}_t$, as $\boldsymbol{z}_h(\ell)$ approaches $\boldsymbol{\mu}$ the uncertainty in forecasts is the same as that of $\boldsymbol{z}_t$. Also, from Equation (2.74), it is easy to see that

$$\text{Cov}[\boldsymbol{e}_h(\ell)] = \text{Cov}[\boldsymbol{e}_h(\ell-1)] + \boldsymbol{\psi}_{\ell-1} \boldsymbol{\Sigma}_a \boldsymbol{\psi}_{\ell-1}', \quad \ell > 1.$$

The covariances of forecast errors thus can also be computed recursively.

2.9.2 Forecasts of an Estimated Model

In practice, the parameters of a VAR(p) model are unknown, and one would like to take into account the parameter uncertainty in forecasting. For simplicity and similar to real-world applications, we assume that the parameters are estimated using the information available at the forecast origin $t = h$. That is, estimation is carried out based on the available information in F_h. Under this assumption, parameter estimates are functions of F_h and, hence, the ℓ-step ahead minimum mean squared error (MSE) forecast of $\boldsymbol{z}_{h+\ell}$ with estimated parameters is

$$\hat{\boldsymbol{z}}_h(\ell) = \hat{\boldsymbol{\phi}}_0 + \sum_{i=1}^{p} \hat{\boldsymbol{\phi}}_i \hat{\boldsymbol{z}}_h(\ell - i), \tag{2.75}$$

where, as before, $\hat{\boldsymbol{z}}_h(j) = \boldsymbol{z}_{h+j}$ for $j \leq 0$. The point forecasts using estimated parameters thus remain the same as before. This is so because the estimates are unbiased and the forecast is out-of-sample forecast. However, the associated forecast error is

$$\begin{aligned}\hat{\boldsymbol{e}}_h(\ell) &= \boldsymbol{z}_{h+\ell} - \hat{\boldsymbol{z}}_h(\ell) = \boldsymbol{z}_{h+\ell} - \boldsymbol{z}_h(\ell) + \boldsymbol{z}_h(\ell) - \hat{\boldsymbol{z}}_h(\ell) \\ &= \boldsymbol{e}_h(\ell) + [\boldsymbol{z}_h(\ell) - \hat{\boldsymbol{z}}_h(\ell)]. \end{aligned} \tag{2.76}$$

Notice that $\boldsymbol{e}_h(\ell)$ are functions of $\{\boldsymbol{a}_{h+1}, \ldots, \boldsymbol{a}_{h+\ell}\}$ and the second term in the right side of Equation (2.76) is a function of F_h. The two terms of the forecast errors $\hat{\boldsymbol{e}}_h(\ell)$ are therefore uncorrelated and we have

$$\begin{aligned}\text{Cov}[\hat{\boldsymbol{e}}_h(\ell)] &= \text{Cov}[\boldsymbol{e}_h(\ell)] + E\{[\boldsymbol{z}_h(\ell) - \hat{\boldsymbol{z}}_h(\ell)][\boldsymbol{z}_h(\ell) - \hat{\boldsymbol{z}}_h(\ell)]'\} \\ &\equiv \text{Cov}[\boldsymbol{e}_h(\ell)] + \text{MSE}[\boldsymbol{z}_h(\ell) - \hat{\boldsymbol{z}}_h(\ell)], \end{aligned} \tag{2.77}$$

where the notation $\equiv$ is used to denote equivalence. To derive the MSE in Equation (2.77), we follow the approach of Samaranayake and Hasza (1988) and Basu and Sen Roy (1986).

Using the model form in Equation (2.28) for a VAR(p) model, we denote the parameter matrix via $\boldsymbol{\beta}$. Letting $T_p = T - p$ be the effective sample size in estimation, we assume that the parameter estimates satisfy

$$\sqrt{T_p}\text{vec}\left(\hat{\boldsymbol{\beta}}' - \boldsymbol{\beta}'\right) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\beta'}).$$

As discussed in Section 2.5, several estimation methods can produce such estimates for a stationary VAR(p) model. Since $\boldsymbol{z}_h(\ell)$ is a differentiable function of vec($\boldsymbol{\beta}'$), one can show that

$$\sqrt{T_p}[\hat{\boldsymbol{z}}_h(\ell) - \boldsymbol{z}_h(\ell)|F_h] \to_d N\left(\mathbf{0}, \frac{\partial \boldsymbol{z}_h(\ell)}{\partial \text{vec}(\boldsymbol{\beta}')'} \boldsymbol{\Sigma}_\beta \frac{\partial \boldsymbol{z}_h(\ell)'}{\partial \text{vec}(\boldsymbol{\beta}')}\right).$$

This result suggests that we can approximate the MSE in Equation (2.77) by

$$\boldsymbol{\Omega}_\ell = E\left[\frac{\partial \boldsymbol{z}_h(\ell)}{\partial \text{vec}(\boldsymbol{\beta}')'} \boldsymbol{\Sigma}_\beta \frac{\partial \boldsymbol{z}_h(\ell)'}{\partial \text{vec}(\boldsymbol{\beta}')}\right].$$

If we further assume that $\boldsymbol{a}_t$ is multivariate normal, then we have

$$\sqrt{T_p}\,[\hat{\boldsymbol{z}}_h(\ell) - \boldsymbol{z}_h(\ell)] \to_d N(\mathbf{0}, \boldsymbol{\Omega}_\ell).$$

Consequently, we have

$$\text{Cov}\,[\hat{\boldsymbol{e}}_h(\ell)] = \text{Cov}[\boldsymbol{e}_h(\ell)] + \frac{1}{T_p}\boldsymbol{\Omega}_\ell. \tag{2.78}$$

It remains to derive the quantity $\boldsymbol{\Omega}_\ell$. To this end, we need to obtain the derivatives $\partial \boldsymbol{z}_h(\ell)/\partial \text{vec}(\boldsymbol{\beta}')$. (The reason for using $\boldsymbol{\beta}'$ instead of $\boldsymbol{\beta}$ is to simplify the matrix derivatives later.) As shown in Equation (2.70), $\boldsymbol{z}_h(\ell)$ can be calculated recursively. Moreover, we can further generalize the expanded series to include the constant $\boldsymbol{\phi}_0$. Specifically, let $\boldsymbol{x}_h = (1, \boldsymbol{z}'_h, \boldsymbol{z}'_{h-1}, \ldots, \boldsymbol{z}'_{h-p+1})'$ be the $(kp+1)$-dimensional vector at the forecast origin $t = h$. Then, by Equation (2.70), we have

$$\boldsymbol{z}_h(\ell) = \boldsymbol{J}\boldsymbol{P}^\ell \boldsymbol{x}_h, \quad \ell \geq 1, \tag{2.79}$$

where

$$\boldsymbol{P} = \begin{bmatrix} 1 & \boldsymbol{0}'_{kp} \\ \boldsymbol{\nu} & \boldsymbol{\Phi} \end{bmatrix}_{(kp+1)\times(kp+1)}, \qquad \boldsymbol{J} = [\boldsymbol{0}_k, \boldsymbol{I}_k, \boldsymbol{0}_{k\times k(p-1)}]_{k\times(kp+1)},$$

and $\boldsymbol{\nu} = [\boldsymbol{\phi}'_0, \boldsymbol{0}_{k(p-1)}]'$, where $\boldsymbol{\Phi}$ is the companion matrix of $\boldsymbol{\phi}(B)$ as defined in Equation (2.23), $\boldsymbol{0}_m$ is an m-dimensional vector of zero, and $\boldsymbol{0}_{m\times n}$ is an $m \times n$ matrix of zero. This is a generalized version of Equation (2.73) to include the constant vector $\boldsymbol{\phi}_0$ in the recursion and it can be shown by mathematical induction. Using Equation (2.79) and part (k) of **Result 3** in Appendix A, we have

$$\begin{aligned}
\frac{\partial \boldsymbol{z}_h(\ell)}{\partial \text{vec}(\boldsymbol{\beta}')'} &= \frac{\partial \text{vec}(\boldsymbol{J}\boldsymbol{P}^\ell \boldsymbol{x}_h)}{\partial \text{vec}(\boldsymbol{\beta}')'} = (\boldsymbol{x}'_h \otimes \boldsymbol{J})\frac{\partial \text{vec}(\boldsymbol{P}^\ell)}{\partial \text{vec}(\boldsymbol{\beta}')'} \\
&= (\boldsymbol{x}'_h \otimes \boldsymbol{J})\left[\sum_{i=0}^{\ell-1}(\boldsymbol{P}')^{\ell-1-i} \otimes \boldsymbol{P}^i\right]\frac{\partial \text{vec}(\boldsymbol{P}^\ell)}{\partial \text{vec}(\boldsymbol{\beta}')'} \\
&= (\boldsymbol{x}'_h \otimes \boldsymbol{J})\left[\sum_{i=0}^{\ell-1}(\boldsymbol{P}')^{\ell-1-i} \otimes \boldsymbol{P}^i\right](\boldsymbol{I}_{kp+1} \otimes \boldsymbol{J}') \\
&= \sum_{i=0}^{\ell-1} \boldsymbol{x}'_h(\boldsymbol{P}')^{\ell-1-i} \otimes \boldsymbol{J}\boldsymbol{P}^i\boldsymbol{J}' \\
&= \sum_{i=0}^{\ell-1} \boldsymbol{x}'_h(\boldsymbol{P}')^{\ell-1-i} \otimes \boldsymbol{\psi}_i,
\end{aligned}$$

where we have used the fact that $\boldsymbol{J}\boldsymbol{P}^i\boldsymbol{J} = \boldsymbol{\psi}_i$. Using the LS estimate $\hat{\boldsymbol{\beta}}$, we have, via Equation (2.43), $\boldsymbol{\Sigma}_{\beta'} = \boldsymbol{G}^{-1} \otimes \boldsymbol{\Sigma}_a$. Therefore,

$$\begin{aligned}
\boldsymbol{\Omega}_\ell &= E\left[\frac{\partial \boldsymbol{z}_h(\ell)}{\partial \mathrm{vec}(\boldsymbol{\beta}')'}\left(\boldsymbol{G}^{-1}\otimes \boldsymbol{\Sigma}_a\right)\frac{\partial \boldsymbol{z}_h(\ell)'}{\partial \mathrm{vec}(\boldsymbol{\beta}')}\right] \\
&= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} E\left(\boldsymbol{x}_h'(\boldsymbol{P}')^{\ell-1-i}\boldsymbol{G}^{-1}\boldsymbol{P}^{\ell-1-j}\boldsymbol{x}_h\right)\otimes \boldsymbol{\psi}_i\boldsymbol{\Sigma}_a\boldsymbol{\psi}_j' \\
&= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} E\left[tr\left(\boldsymbol{x}_h'(\boldsymbol{P}')^{\ell-1-i}\boldsymbol{G}^{-1}\boldsymbol{P}^{\ell-1-j}\boldsymbol{x}_h\right)\right]\boldsymbol{\psi}_i\boldsymbol{\Sigma}_a\boldsymbol{\psi}_j' \\
&= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} tr\left[\left(\boldsymbol{P}'\right)^{\ell-1-i}\boldsymbol{G}^{-1}\boldsymbol{P}^{\ell-1-j}E\left(\boldsymbol{x}_h\boldsymbol{x}_h'\right)\right]\boldsymbol{\psi}_i\boldsymbol{\Sigma}_a\boldsymbol{\psi}_j' \\
&= \sum_{i=0}^{\ell-1}\sum_{j=0}^{\ell-1} tr\left[\left(\boldsymbol{P}'\right)^{\ell-1-i}\boldsymbol{G}^{-1}\boldsymbol{P}^{\ell-1-j}\boldsymbol{G}\right]\boldsymbol{\psi}_i\boldsymbol{\Sigma}_a\boldsymbol{\psi}_j'. \qquad (2.80)
\end{aligned}$$

In particular, if $\ell = 1$, then

$$\boldsymbol{\Omega}_1 = tr(\boldsymbol{I}_{kp+1})\boldsymbol{\Sigma}_a = (kp+1)\boldsymbol{\Sigma}_a,$$

and

$$\mathrm{Cov}[\hat{\boldsymbol{z}}_h(1)] = \boldsymbol{\Sigma}_a + \frac{kp+1}{T_p}\boldsymbol{\Sigma}_a = \frac{T_p+kp+1}{T_p}\boldsymbol{\Sigma}_a.$$

Since $kp+1$ is the number of parameters in the model equation for z_{it}, the prior result can be interpreted as each parameter used increases the MSE of one-step ahead forecasts by a factor of $1/T_p$, where T_p is the effective sample size used in the estimation. When K or p is large, but T_p is small, the impact of using estimated parameters in forecasting could be substantial. The result, therefore, provides support for removing insignificant parameters in a VAR(p) model. In other words, it pays to employ parsimonious models.

In practice, we can replace the quantities in Equation (2.80) by their LS estimates to compute the $\mathrm{Cov}[\hat{\boldsymbol{z}}_h(\ell)]$.

Example 2.8 Consider, again, the quarterly GDP growth rates, in percentages, of United Kingdom, Canada, and United States employed in Example 2.6, where a VAR(2) model is fitted. Using this model, we consider one-step to eight-step ahead forecasts of the GDP growth rates at the forecast origin 2011.II. We also provide the standard errors and root mean-squared errors of the predictions. The root mean-squared errors include the uncertainty due to the use of estimated parameters. The results are given in Table 2.2. From the table, we make the following observations. First, the point forecasts of the three series move closer to the sample means of the data as the forecast horizon increases, showing evidence of mean reverting. Second,

TABLE 2.2 Forecasts of Quarterly GDP Growth Rates, in Percentages, for United Kingdom, Canada, and United States via a VAR(2) Model

	Forecasts			Standard Errors			Root MSE		
Step	United Kingdom	Canada	United States	United Kingdom	Canada	United States	United Kingdom	Canada	United States
1	0.31	0.05	0.17	0.53	0.54	0.60	0.55	0.55	0.61
2	0.26	0.32	0.49	0.58	0.72	0.71	0.60	0.78	0.75
3	0.31	0.48	0.52	0.62	0.77	0.73	0.64	0.79	0.75
4	0.38	0.53	0.60	0.65	0.78	0.74	0.66	0.78	0.75
5	0.44	0.57	0.63	0.66	0.78	0.75	0.67	0.78	0.75
6	0.48	0.59	0.65	0.67	0.78	0.75	0.67	0.78	0.75
7	0.51	0.61	0.66	0.67	0.78	0.75	0.67	0.78	0.75
8	0.52	0.62	0.67	0.67	0.78	0.75	0.67	0.78	0.75
Data	0.52	0.62	0.65	0.71	0.79	0.79	0.71	0.79	0.79

The forecast origin is the second quarter of 2011. The last row of the table gives the sample means and sample standard errors of the series.

as expected, the standard errors and root mean-squared errors of forecasts increase with the forecast horizon. The standard errors should converge to the standard errors of the time series as the forecast horizon increases. Third, the effect of using estimated parameters is evident when the forecast horizon is small. The effect vanishes quickly as the forecast horizon increases. This is reasonable because a stationary VAR model is mean-reverting. The standard errors and mean-squared errors of prediction should converge to the standard errors of the series. The standard errors and root mean-squared errors of Table 2.2 can be used to construct interval predictions. For instance, a two-step ahead 95% interval forecast for U.S. GDP growth rate is $0.49 \pm 1.96 \times 0.71$ and $0.49 \pm 1.96 \times 0.75$, respectively, for predictions without and with parameter uncertainty.

Finally, we provide estimates of the $\mathbf{\Omega}_\ell$ matrix in the following R demonstration. As expected, $\mathbf{\Omega}_\ell$ decreases in magnitude as the forecast horizon increases. The forecasting results of the example are obtained by using the command `VARpred` of the `MTS` package. □

R Demonstration: Prediction.

```
> VARpred(m1,8)
Forecasts at origin:  125
     uk      ca     us
 0.3129 0.05166 0.1660
 0.2647 0.31687 0.4889
 ....
 0.5068 0.60967 0.6630
 0.5247 0.61689 0.6688
```

```
Standard Errors of predictions:
       [,1]   [,2]   [,3]
[1,] 0.5315 0.5400 0.5975
[2,] 0.5804 0.7165 0.7077
 ....
[7,] 0.6719 0.7842 0.7486
[8,] 0.6729 0.7843 0.7487
Root Mean square errors of predictions:
       [,1]   [,2]   [,3]
[1,] 0.5461 0.5549 0.6140
[2,] 0.6001 0.7799 0.7499
 ....
[7,] 0.6730 0.7844 0.7487
[8,] 0.6734 0.7844 0.7487
> colMeans(z)  ## Compute sample means
       uk        ca        us
0.5223092 0.6153672 0.6473996
> sqrt(apply(z,2,var))  ## Sample standard errors
       uk        ca        us
0.7086442 0.7851955 0.7872912
>
Omega matrix at horizon:  1
            [,1]        [,2]       [,3]
[1,] 0.015816875 0.001486291 0.00416376
[2,] 0.001486291 0.016328573 0.00781132
[3,] 0.004163760 0.007811320 0.01999008
Omega matrix at horizon:  2
           [,1]       [,2]       [,3]
[1,] 0.02327855 0.03708068 0.03587541
[2,] 0.03708068 0.09490535 0.07211282
[3,] 0.03587541 0.07211282 0.06154730
Omega matrix at horizon:  3
           [,1]       [,2]       [,3]
[1,] 0.02044253 0.02417433 0.01490480
[2,] 0.02417433 0.03218999 0.02143570
[3,] 0.01490480 0.02143570 0.01652968
Omega matrix at horizon:  4
            [,1]        [,2]        [,3]
[1,] 0.015322037 0.010105520 0.009536199
[2,] 0.010105520 0.007445308 0.006700067
[3,] 0.009536199 0.006700067 0.006181433
```

2.10 IMPULSE RESPONSE FUNCTIONS

In studying the structure of a VAR model, we discussed the Granger causality and the relation to transfer function models. There is another approach to explore the relation between variables. As a matter of fact, we are often interested in knowing the effect of changes in one variable on another variable in multivariate time series analysis. For instance, suppose that the bivariate time series $\boldsymbol{z}_t$ consists of monthly income and expenditure of a household, we might be interested in knowing the effect on expenditure if the monthly income of the household is increased or decreased by a certain amount, for example, 5%. This type of study is referred to as the *impulse response function* in the statistical literature and the *multiplier analysis* in the econometric literature. In this section, we use the MA representation of a VAR(p) model to assess such effects.

In multiplier analysis, we can assume $E(\boldsymbol{z}_t) = \mathbf{0}$ because the mean does not affect the pattern of the response of $\boldsymbol{z}_t$ to any shock. To study the effects of changes in z_{1t} on $\boldsymbol{z}_{t+j}$ for $j > 0$ while holding other quantities unchanged, we can assume that $t = 0$, $\boldsymbol{z}_t = \mathbf{0}$ for $t \leq 0$, and $\boldsymbol{a}_0 = (1, 0, \ldots)'$. In other words, we would like to study the behavior of $\boldsymbol{z}_t$ for $t > 0$ while z_{10} increases by 1. To this end, we can trace out $\boldsymbol{z}_t$ for $t = 1, 2, \ldots$, assuming $\boldsymbol{a}_t = \mathbf{0}$ for $t > 0$. Using the MA representation of a VAR(p) model with coefficient matrix $\boldsymbol{\psi}_\ell = [\psi_{\ell,ij}]$ given in Equation (2.27), we have

$$\boldsymbol{z}_0 = \boldsymbol{a}_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{z}_1 = \boldsymbol{\psi}_1 \boldsymbol{a}_0 = \begin{bmatrix} \psi_{1,11} \\ \psi_{1,21} \\ \vdots \\ \psi_{1,k1} \end{bmatrix}, \quad \boldsymbol{z}_2 = \boldsymbol{\psi}_2 \boldsymbol{a}_0 = \begin{bmatrix} \psi_{2,11} \\ \psi_{2,21} \\ \vdots \\ \psi_{2,k1} \end{bmatrix}, \ldots$$

The results are simply the first columns of the coefficient matrices $\boldsymbol{\psi}_i$. Therefore, in this particular case, we have $\boldsymbol{z}_t = \boldsymbol{\psi}_{t,.1}$ where $\boldsymbol{\psi}_{\ell,.1}$ denotes the first column of $\boldsymbol{\psi}_\ell$. Similarly, to study the effect on $\boldsymbol{z}_{t+j}$ by increasing the ith series z_{it} by 1, we have $\boldsymbol{a}_0 = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the ith unit vector in R^k, the k-dimensional Euclidean space, we have

$$\boldsymbol{z}_0 = \boldsymbol{e}_i, \quad \boldsymbol{z}_1 = \boldsymbol{\psi}_{1,.i}, \quad \boldsymbol{z}_2 = \boldsymbol{\psi}_{2,.i}, \quad \ldots.$$

They are the ith columns of the coefficient matrices $\boldsymbol{\psi}_i$ of the MA representation of $\boldsymbol{z}_t$. For this reason, the coefficient matrix $\boldsymbol{\psi}_i$ of the MA representation of a VAR(p) model is referred to as the coefficients of impulse response functions. The summation

$$\underline{\boldsymbol{\psi}}_n = \sum_{i=0}^{n} \boldsymbol{\psi}_i$$

denotes the *accumulated responses* over n periods to a unit shock to $\boldsymbol{z}_t$. Elements of $\underline{\boldsymbol{\psi}}_n$ are referred to as the nth interim multipliers. The total accumulated responses for all future periods are defined as

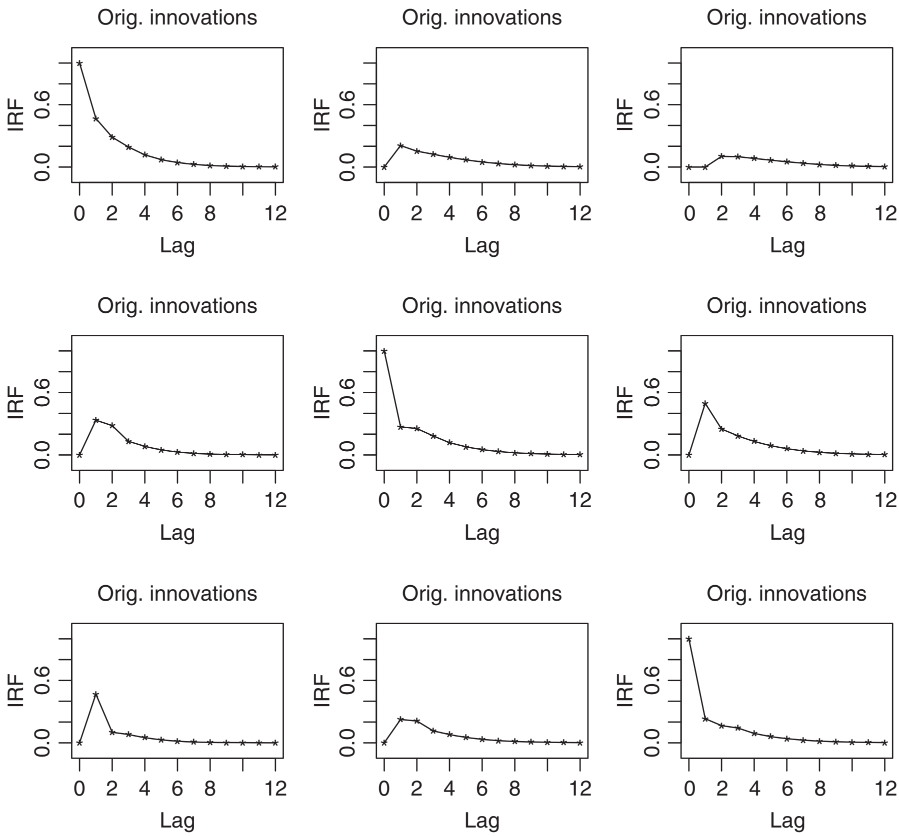

FIGURE 2.6 Impulse response functions of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages.

$$\underline{\psi}_{\infty} = \sum_{i=0}^{\infty} \psi_i.$$

Often $\underline{\psi}_{\infty}$ is called the *total multipliers* or *long-run effects*.

To demonstrate, consider again the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates, in percentages, of United Kingdom, Canada, and United States. Figure 2.6 shows the impulse response functions of the fitted three-dimensional model. From the plots, the impulse response functions decay to 0 quickly. This is expected for a stationary series. The upper-right plot shows that there is a delayed effect on the U.K. GDP growth rate if one changed the U.S. growth rate by 1. This delayed effect is due to the fact that a change in U.S. rate at time t affects the Canadian rate at time $t + 1$, which in turn affects the U.K. rate at time $t + 2$. This plot thus shows that the impulse response functions show the marginal

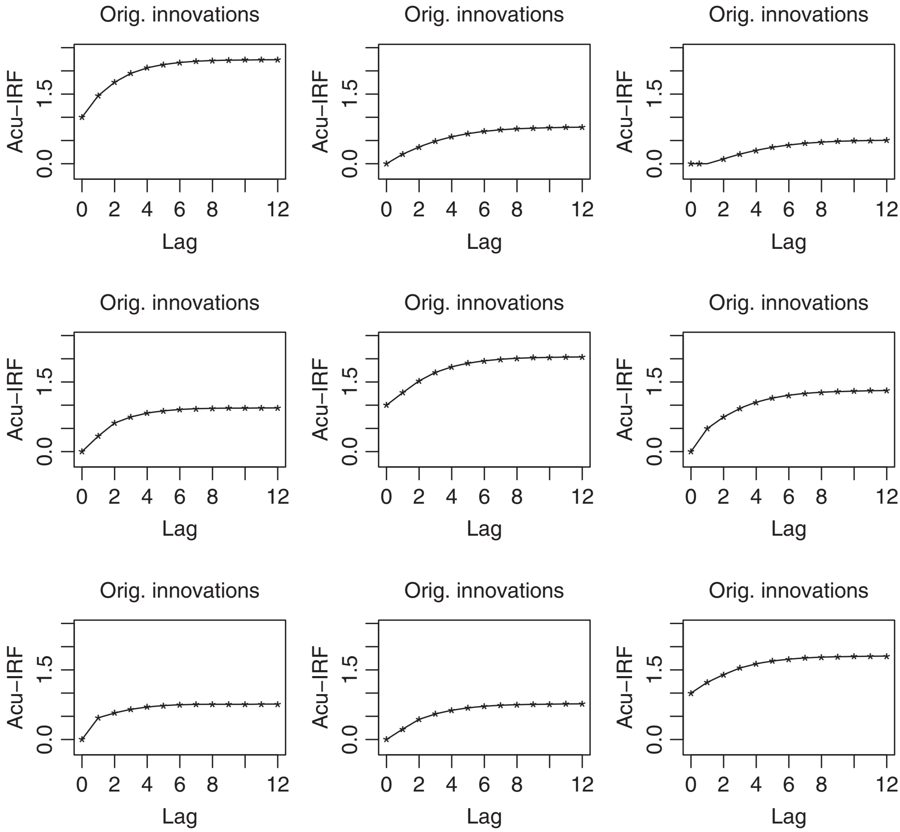

FIGURE 2.7 Accumulated responses of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages.

effects, not the conditional effects. Recall that conditional on the lagged Canadian rate, the growth rate of U.K. GDP does not depend on lagged values of U.S. growth rate. Figure 2.7 shows the accumulated responses implied by the simplified VAR(2) model. As expected, the accumulated effects converge to the total multipliers quickly.

2.10.1 Orthogonal Innovations

In practice, elements of $\boldsymbol{a}_t$ tend to be correlated, that is, $\boldsymbol{\Sigma}_a$ is not a diagonal matrix. As such, change in one component of $\boldsymbol{a}_t$ will simultaneously affect other components of $\boldsymbol{a}_t$. Consequently, the impulse response functions introduced in the previous section would encounter difficulties in a real application because one cannot arbitrarily increase z_{1t} by 1 without altering other components z_{it}. Mathematically, the problem can be shown as follows. The effect of change in a_{1t} on the future series $\boldsymbol{z}_{t+j}$ $(j \geq 0)$ can be quantified as $\partial \boldsymbol{z}_{t+j}/\partial a_{1t}$. Since $\{\boldsymbol{a}_t\}$ are serially uncorrelated, we can use the MA representation of $\boldsymbol{z}_{t+j}$ and obtain

$$\frac{\partial \boldsymbol{z}_{t+j}}{\partial a_{1t}} = \boldsymbol{\psi}_j \frac{\partial \boldsymbol{a}_t}{\partial a_{1t}} = \boldsymbol{\psi}_j \begin{bmatrix} \frac{\partial a_{1t}}{\partial a_{1t}} \\ \frac{\partial a_{2t}}{\partial a_{1t}} \\ \vdots \\ \frac{\partial a_{kt}}{\partial a_{1t}} \end{bmatrix} = \boldsymbol{\psi}_j \begin{bmatrix} 1 \\ \frac{\sigma_{a,21}}{\sigma_{a,11}} \\ \vdots \\ \frac{\sigma_{a,k1}}{\sigma_{a,11}} \end{bmatrix} = \boldsymbol{\psi}_j \boldsymbol{\Sigma}_{a,.1} \sigma_{a,11}^{-1},$$

where $\boldsymbol{\Sigma}_{a,.1}$ denotes the first column of $\boldsymbol{\Sigma}_a$, and we have used $\boldsymbol{\Sigma}_a = [\sigma_{a,ij}]$ and the simple linear regression $a_{it} = (\sigma_{a,i1}/\sigma_{a,11})a_{1t} + \epsilon_{it}$ with ϵ_{it} denoting the error term. In general, we have $\partial \boldsymbol{z}_{t+j}/\partial a_{it} = \boldsymbol{\psi}_j \boldsymbol{\Sigma}_{a,.i} \sigma_{a,ii}^{-1}$. Thus, the correlations between components of $\boldsymbol{a}_t$ cannot be ignored.

To overcome this difficulty, one can take a proper transformation of $\boldsymbol{a}_t$ such that components of the innovation become uncorrelated, that is, diagonalize the covariance matrix $\boldsymbol{\Sigma}_a$. A simple way to achieve orthogonalization of the innovation is to consider the Cholesky decomposition of $\boldsymbol{\Sigma}_a$ as discussed in Chapter 1. Specifically, we have

$$\boldsymbol{\Sigma}_a = \boldsymbol{U}'\boldsymbol{U},$$

where $\boldsymbol{U}$ is an upper triangular matrix with positive diagonal elements. Let $\boldsymbol{\eta}_t = (\boldsymbol{U}')^{-1}\boldsymbol{a}_t$. Then,

$$\text{Cov}(\boldsymbol{\eta}_t) = (\boldsymbol{U}')^{-1}\text{Cov}(\boldsymbol{a}_t)\boldsymbol{U}^{-1} = (\boldsymbol{U}')^{-1}(\boldsymbol{U}'\boldsymbol{U})(\boldsymbol{U})^{-1} = \boldsymbol{I}_k.$$

Thus, components of $\boldsymbol{\eta}_t$ are uncorrelated and have unit variance.

From the MA representation of $\boldsymbol{z}_t$, we have

$$\begin{aligned} \boldsymbol{z}_t &= \boldsymbol{\psi}(B)\boldsymbol{a}_t = \boldsymbol{\psi}(B)\boldsymbol{U}'(\boldsymbol{U}')^{-1}\boldsymbol{a}_t, \\ &= [\boldsymbol{\psi}(B)\boldsymbol{U}']\boldsymbol{\eta}_t, \\ &= [\underline{\boldsymbol{\psi}}_0 + \underline{\boldsymbol{\psi}}_1 B + \underline{\boldsymbol{\psi}}_2 B^2 + \cdots]\boldsymbol{\eta}_t, \end{aligned} \tag{2.81}$$

where $\underline{\boldsymbol{\psi}}_\ell = \boldsymbol{\psi}_\ell \boldsymbol{U}'$ for $\ell \geq 0$.

Letting $[\underline{\psi}_{\ell,ij}] = \underline{\boldsymbol{\psi}}_\ell$, we can easily see that $\partial \boldsymbol{z}_{t+\ell}/\partial \eta_{it} = \underline{\boldsymbol{\psi}}_{\ell,.i}$ for $\ell > 0$. We call $\underline{\boldsymbol{\psi}}_\ell$ the impulse response coefficients of $\boldsymbol{z}_t$ with *orthogonal innovations*. The plot of $\underline{\psi}_{\ell,ij}$ against ℓ is called the *impulse response function* of $\boldsymbol{z}_t$ with orthogonal innovations. Specifically, $\underline{\psi}_{\ell,ij}$ denotes the impact of a shock with size being "one standard deviation" of the jth innovation at time t on the future value of $z_{i,t+\ell}$. In particular, the (i, j)th element of the transformation matrix $(\boldsymbol{U}')^{-1}$ (where $i > j$) denotes the instantaneous effect of the shock η_{jt} on z_{it}. We can define accumulated responses by summing over the coefficient matrices $\underline{\boldsymbol{\psi}}_\ell$ in a similar manner as before.

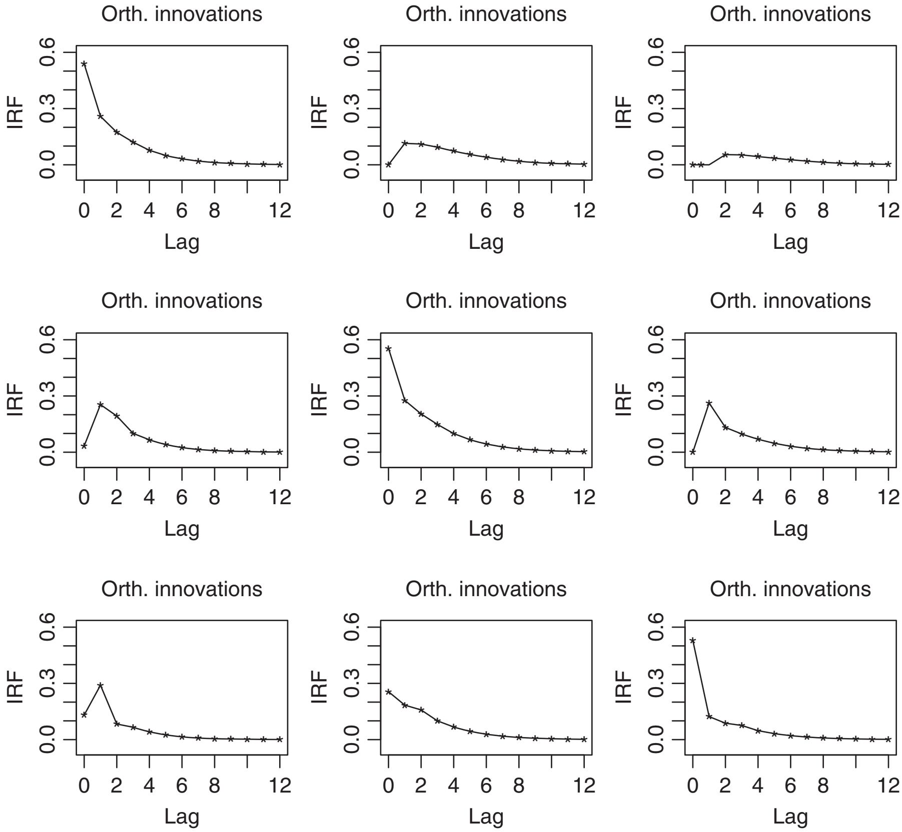

FIGURE 2.8 Impulse response functions of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages and innovations are orthogonalized.

Figure 2.8 shows the impulse response functions of the simplified VAR(2) model in Equation (2.65) when the innovations are orthogonalized. For this particular instance, the instantaneous correlations between the components of $\boldsymbol{a}_t$ are not strong so that the orthogonalization of the innovations has only a small impact. Figure 2.9 shows the accumulated responses of the GDP series with orthogonal innovations. Again, the pattern of the impulse response functions is similar to those shown in Figure 2.7.

Remark: The prior definition of impulse response function depends on the *ordering* of the components in $\boldsymbol{z}_t$. The lower triangular structure of $\boldsymbol{U}'$ indicates that the η_{1t} is a function of a_{1t}, η_{2t} is a function of a_{1t} and a_{2t}, and so on. Thus, η_{it} is not affected by a_{jt} for $j > i$. Different orderings of the components of $\boldsymbol{a}_t$ thus lead to different impulse response functions for a given VAR(p) model. However, one

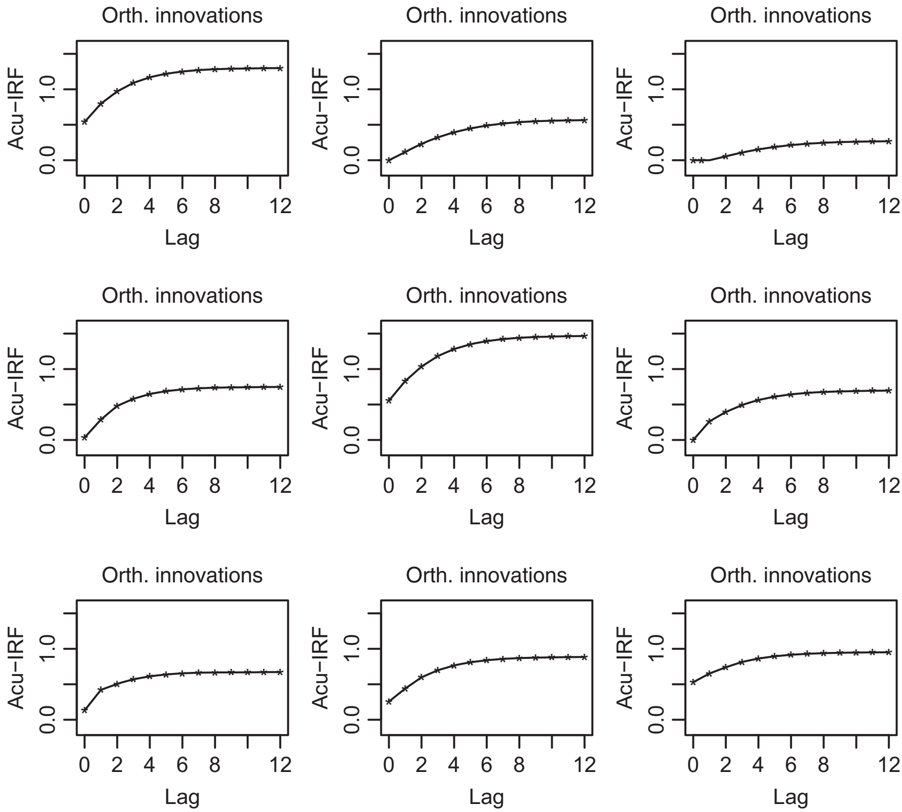

FIGURE 2.9 Accumulated responses of the simplified VAR(2) model in Equation (2.65) for the quarterly growth rates of real gross domestic products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The growth rates are in percentages and innovations are orthogonalized.

should remember that the meanings of the transformed innovations $\boldsymbol{\eta}_t$ also depend on the ordering of $\boldsymbol{a}_t$. □

Remark: The impulse response functions of a VAR(p) with or without the orthogonalization of the innovations can be obtained via the command `VARirf` in the `MTS` package. The input includes the AR coefficient matrix $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p]$ and the innovation covariance matrix $\boldsymbol{\Sigma}_a$. The default option uses orthogonal innovations. □

R Demonstration: Impulse response functions of a VAR model.

```
> Phi = m2$Phi  ### m2 is the simplified VAR(2) model
> Sig = m2$Sigma
> VARirf(Phi,Sig)  ### Orthogonal innovations
> VARirf(Phi,Sig,orth=F) ## Original innovations
```

2.11 FORECAST ERROR VARIANCE DECOMPOSITION

Using the MA representation of a VAR(p) model in Equation (2.81) and the fact that $\text{Cov}(\boldsymbol{\eta}_t) = \boldsymbol{I}_k$, we see that the ℓ-step ahead forecast error of $\boldsymbol{z}_{h+\ell}$ at the forecast origin $t = h$ can be written as

$$\boldsymbol{e}_h(\ell) = \underline{\boldsymbol{\psi}}_0 \boldsymbol{\eta}_{h+\ell} + \underline{\boldsymbol{\psi}}_1 \boldsymbol{\eta}_{h+\ell-1} + \cdots + \underline{\boldsymbol{\psi}}_{\ell-1} \boldsymbol{\eta}_{h+1},$$

and the covariance matrix of the forecast error is

$$\text{Cov}[\boldsymbol{e}_h(\ell)] = \sum_{v=0}^{\ell-1} \underline{\boldsymbol{\psi}}_v \underline{\boldsymbol{\psi}}_v'. \tag{2.82}$$

From Equation (2.82), the variance of the forecast error $e_{h,i}(\ell)$, which is the ith component of $\boldsymbol{e}_h(\ell)$, is

$$\text{Var}[e_{h,i}(\ell)] = \sum_{v=0}^{\ell-1} \sum_{j=1}^{k} \underline{\psi}_{v,ij}^2 = \sum_{j=1}^{k} \sum_{v=0}^{\ell-1} \underline{\psi}_{v,ij}^2. \tag{2.83}$$

Using Equation (2.83), we define

$$w_{ij}(\ell) = \sum_{v=0}^{\ell-1} \underline{\psi}_{v,ij}^2,$$

and obtain

$$\text{Var}[e_{h,i}(\ell)] = \sum_{j=1}^{k} w_{ij}(\ell). \tag{2.84}$$

Therefore, the quantity $w_{ij}(\ell)$ can be interpreted as the contribution of the jth shock η_{jt} to the variance of the ℓ-step ahead forecast error of z_{it}. Equation (2.84) is referred to as the forecast error decomposition. In particular, $w_{ij}(\ell)/\text{Var}[e_{h,i}(\ell)]$ is the percentage of contribution from the shock η_{jt}.

To demonstrate, consider the simplified VAR(2) model in Equation (2.65) for the quarterly percentage GDP growth rates of United Kingdom, Canada, and United States. The forecast error variance decompositions for the one-step to five-step ahead predictions at the forecast origin 2011.II are given in Table 2.3. From the table, we see that the decomposition, again, depends on the ordering of the components in $\boldsymbol{z}_t$. However, the results confirm that the three growth rate series are interrelated.

TABLE 2.3 Forecast Error Variance Decomposition for One-step to Five-step Ahead Predictions for the Quarterly GDP Growth Rates, in Percentages, of United Kingdom, Canada, and United States

Variable	Step	United Kingdom	Canada	United States
United Kingdom	1	1.0000	0.0000	0.0000
	2	0.9645	0.0355	0.0000
	3	0.9327	0.0612	0.0071
	4	0.9095	0.0775	0.0130
	5	0.8956	0.0875	0.0170
Canada	1	0.0036	0.9964	0.0000
	2	0.1267	0.7400	0.1333
	3	0.1674	0.6918	0.1407
	4	0.1722	0.6815	0.1462
	5	0.1738	0.6767	0.1495
United States	1	0.0473	0.1801	07726
	2	0.2044	0.1999	0.5956
	3	0.2022	0.2320	0.5658
	4	0.2028	0.2416	0.5556
	5	0.2028	0.2460	0.5512

The forecast origin is the second quarter of 2011. The simplified VAR(2) model in Equation (2.65) is used.

R Demonstration: Forecast error decomposition.

```
> m1=VAR(z,2)
> m2=refVAR(m1)
> names(m2)
 [1] "data"  "order"      "cnst"    "coef"    "aic"   "bic"
 [7] "hq"    "residuals"  "secoef"  "Sigma"   "Phi"   "Ph0"
> Phi=m2$Phi
> Sig=m2$Sigma
> Theta=NULL
> FEVdec(Phi,Theta,Sig,lag=5)
Order of the ARMA mdoel:
[1] 2 0
Standard deviation of forecast error:
          [,1]      [,2]      [,3]      [,4]      [,5]
# Forecat horison
[1,] 0.5385505 0.6082891 0.6444223 0.6644656 0.6745776
[2,] 0.5550000 0.7197955 0.7839243 0.8100046 0.8217975
[3,] 0.6022357 0.7040833 0.7317336 0.7453046 0.7510358
Forecast-Error-Variance Decomposition
Forecast horizon:  1
            [,1]      [,2]      [,3]
[1,] 1.000000000 0.0000000 0.0000000
[2,] 0.003640595 0.9963594 0.0000000
[3,] 0.047327504 0.1801224 0.7725501
```

```
Forecast horizon:  2
          [,1]      [,2]      [,3]
[1,] 0.9645168 0.0354832 0.0000000
[2,] 0.1266584 0.7400392 0.1333023
[3,] 0.2044415 0.1999232 0.5956353
```

2.12 PROOFS

We provide proofs for some lemmas and theorems stated in the chapter.

Proof of Lemma 2.1:

$$|\boldsymbol{I} - \boldsymbol{\Phi} B| = \begin{vmatrix} \boldsymbol{I}_k - \boldsymbol{\phi}_1 B & -\boldsymbol{\phi}_2 B & \cdots & -\boldsymbol{\phi}_{p-1} B & -\boldsymbol{\phi}_p B \\ -\boldsymbol{I}_k B & \boldsymbol{I}_k & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_k B & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\boldsymbol{I}_k B & \boldsymbol{I}_k \end{vmatrix}.$$

Multiplying the last column block by B and adding the result to the $(p-1)$th column block, we have

$$|\boldsymbol{I} - \boldsymbol{\Phi} B| = \begin{vmatrix} \boldsymbol{I}_k - \boldsymbol{\phi}_1 B & -\boldsymbol{\phi}_2 B & \cdots & -\boldsymbol{\phi}_{p-1} B - \boldsymbol{\phi}_p B^2 & -\boldsymbol{\phi}_p B \\ -\boldsymbol{I}_k B & \boldsymbol{I}_k & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_k B & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I}_k \end{vmatrix}.$$

Next, multiplying the $(p-1)$th column block by B and adding the result to the $(p-2)$th column block, we have $-\boldsymbol{\phi}_{p-2} B - \boldsymbol{\phi}_{p-1} B^2 - \boldsymbol{\phi}_p^3$ at the $(1, p-2)$ block and 0 at $(p-1, p-2)$ block. By repeating the procedure, we obtain $|\boldsymbol{I} - \boldsymbol{\Phi} B| =$

$$\begin{vmatrix} \boldsymbol{\phi}(B) & -\sum_{j=2}^{p} \boldsymbol{\phi}_j B^{j-1} & \cdots & -\boldsymbol{\phi}_{p-1} B - \boldsymbol{\phi}_p B^2 & -\boldsymbol{\phi}_p B \\ \boldsymbol{0} & \boldsymbol{I}_k & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I}_k \end{vmatrix} = |\boldsymbol{\phi}(B)|.$$

$\square$

Proof of Lemma 2.4: We make use of Equation (2.58), Lemma 2.2, and the limiting distribution of the estimate $\hat{\boldsymbol{\beta}}$. Dividing the second last term of Equation (2.58) by $\sqrt{T_p}$ and applying vec operator, we have

$$\begin{aligned}
&\text{vec}\{\boldsymbol{A}'\boldsymbol{B}_m[\boldsymbol{I}_m \otimes \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]/\sqrt{T_p}\} \\
&\quad = \sqrt{T_p}\text{vec}\left[\frac{1}{T_p}\boldsymbol{A}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \boldsymbol{X})[\boldsymbol{I}_m \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]\right] \\
&\quad = \left[\boldsymbol{I}_k \otimes \frac{1}{T_p}\boldsymbol{A}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \boldsymbol{X})\right]\sqrt{T_p}\text{vec}[\boldsymbol{I}_m \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})].
\end{aligned}$$

The first factor of the prior equation converges to $\mathbf{0}$, whereas the second factor has a limiting normal distribution. Therefore, the second term of Equation (2.58) converges to $\mathbf{0}$ in probability after it is divided by $\sqrt{T_p}$.

Dividing the last term of Equation (2.58) by $\sqrt{T_p}$ and applying the vec operator, we have

$$\begin{aligned}
&\sqrt{T_p}\text{vec}\left\{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\boldsymbol{X}'\boldsymbol{B}_m\left[\boldsymbol{I}_m \otimes \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]\right\} \\
&\quad = \left\{\left[\boldsymbol{I}_m \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right]\frac{(\boldsymbol{I}_m \otimes \boldsymbol{X}')\,\boldsymbol{B}_m'\boldsymbol{X}}{T_p} \otimes \boldsymbol{I}_k\right\}\sqrt{T_p}\text{vec}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right].
\end{aligned}$$

The last factor of the prior equation follows a limiting distribution, the second factor converges to a fixed matrix, but the first factor converges to $\mathbf{0}$. Therefore, the last term of Equation (2.58) converges to $\mathbf{0}$ in probability after it is divided by $\sqrt{T_p}$.

It remains to consider the third term of Equation (2.58). To this end, consider

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{B}_m(\boldsymbol{I}_m \otimes \boldsymbol{A}) &= \boldsymbol{X}'[\boldsymbol{B}\boldsymbol{A}, \boldsymbol{B}^2\boldsymbol{A}, \ldots, \boldsymbol{B}^m\boldsymbol{A}] \\
&= [\boldsymbol{X}'\boldsymbol{B}\boldsymbol{A}, \boldsymbol{X}'\boldsymbol{B}^2\boldsymbol{A}, \ldots, \boldsymbol{X}'\boldsymbol{B}^m\boldsymbol{A}].
\end{aligned}$$

Also, for $i = 1, \ldots, m$,

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{B}^i\boldsymbol{A} &= \sum_{t=p+1}^{n} \boldsymbol{x}_t\boldsymbol{a}_{t-i}' \\
&= \sum_{t=p+1}^{T} \begin{bmatrix} 1 \\ \boldsymbol{z}_{t-1} \\ \vdots \\ \boldsymbol{z}_{t-p} \end{bmatrix} \boldsymbol{a}_{t-i}'.
\end{aligned}$$

Using the MA representation for $\boldsymbol{z}_{t-j} (j = 1, \ldots, p)$ and Lemma 2.2, we obtain

$$\frac{1}{T_p} \boldsymbol{X}' \boldsymbol{B}^i \boldsymbol{A} \rightarrow_p \begin{bmatrix} \boldsymbol{0}' \\ \boldsymbol{\psi}_{i-1} \boldsymbol{\Sigma}_a \\ \boldsymbol{\psi}_{i-2} \boldsymbol{\Sigma}_a \\ \vdots \\ \boldsymbol{\psi}_{i-p} \boldsymbol{\Sigma}_a \end{bmatrix},$$

where it is understood that $\boldsymbol{\psi}_j = \boldsymbol{0}$ for $j < 0$. Therefore,

$$\frac{1}{T_p} \boldsymbol{X}' \boldsymbol{B}_m (\boldsymbol{I}_m \otimes \boldsymbol{A}) \rightarrow_p \boldsymbol{H}_*,$$

where $\boldsymbol{H}_*$ is defined in the statement of Lemma 2.4. Dividing Equation (2.58) by $\sqrt{T_p} = \sqrt{T-p}$, taking the vec operator, and using the properties of vec operator, we have

$$\begin{aligned} \sqrt{T_p} \mathrm{vec}(\hat{\boldsymbol{\Xi}}_m) &\approx \sqrt{T_p} \mathrm{vec}\,(\boldsymbol{\Xi}_m) - \frac{1}{\sqrt{T_p}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{X}' \boldsymbol{B}_m \, (\boldsymbol{I}_m \otimes \boldsymbol{A}) \\ &= \sqrt{T_p} \mathrm{vec}(\boldsymbol{\Xi}_m) - \sqrt{T_p} \left[\frac{(\boldsymbol{I}_m \otimes \boldsymbol{A}') \boldsymbol{B}_m' \boldsymbol{X}}{T_p} \otimes \boldsymbol{I}_k \right] \mathrm{vec}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'], \end{aligned}$$

where $\approx$ is used to denote asymptotic equivalence. This completes the proof of Lemma 2.4. □

EXERCISES

2.1 Prove Lemma 2.2.

2.2 Consider the growth rates, in percentages, of the quarterly real GDP of United Kingdom, Canada, and the United States used in the chapter. Fit a VAR(4) model to the series, simplify the model by removing insignificant parameters with type I error $\alpha = 0.05$, and perform model checking. Finally, compare the simplified VAR(4) model with the simplified VAR(2) model of Section 2.7.3.

2.3 Consider a bivariate time series $\boldsymbol{z}_t$, where z_{1t} is the change in monthly U.S. treasury bills with maturity 3 months and z_{2t} is the inflation rate, in percentage, of the U.S. monthly consumer price index (CPI). The CPI used is the consumer price index for all urban consumers: all items (CPIAUCSL). The original data are downloaded from the Federal Reserve Bank of St. Louis. The CPI rate is 100 times the first difference of the log CPI index. The sample period is from January 1947 to December 2012. The original data are in the file `m-cpitb3m.txt`.

- Construct the z_t series. Obtain the time plots of z_t.
- Select a VAR order for z_t using the BIC criterion.
- Fit the specified VAR model and simplify the fit by the command `refVAR` with threshold 1.65. Write down the fitted model.
- Is the fitted model adequate? Why?
- Compute the impulse response functions of the fitted model using orthogonal innovations. Show the plots and draw conclusion based on the plots.
- Consider the residual covariance matrix. Obtain its Cholesky decomposition and the transformed innovations. Plot the orthogonal innovations.

2.4 Consider the U.S. quarterly gross private saving (GPSAVE) and gross private domestic investment (GPDI) from first quarter of 1947 to the third quarter of 2012. The data are from the Federal Reserve Bank of St. Louis and are in billions of dollars. See the file `m-gpsavedi.txt`.

- Construct the growth series by taking the first difference of the log data. Denote the growth series by z_t. Plot the growth series.
- Build a VAR model for z_t, including simplification and model checking. Write down the fitted model.
- Perform a chi-square test to confirm that one can remove the insignificant parameters in the previous question. You may use 5% significant level.
- Obtain the impulse response functions of the fitted model. What is the relationship between the private investment and saving?
- Obtain one-step to eight-step ahead predictions of z_t at the forecast origin 2012.III (last data point).
- Obtain the forecast error variance decomposition.

2.5 Consider, again, the quarterly growth series z_t of Problem 4. Obtain Bayesian estimation of a VAR(4) model. Write down the fitted model.

2.6 Consider four components of U.S. monthly industrial production index from January 1947 to December 2012 for 792 data points. The four components are durable consumer goods (IPDCONGD), nondurable consumer goods (IPNCONGD), business equivalent (IPBUSEQ), and materials (IPMAT). The original data are from the Federal Reserve Bank of St. Louis and are seasonally adjusted. See the file `m-ip4comp.txt`.

- Construct the growth rate series z_t of the four industrial production index, that is, take the first difference of the log data. Obtain time plots of z_t. Comment on the time plot.
- Build a VAR model for z_t, including simplification and model checking. Write down the fitted model.
- Compute one-step to six-step ahead predictions of z_t at the forecast origin $h = 791$ (December 2012). Obtain 95% interval forecasts for each component series.

2.7 Consider, again, the z_t series of Problem 6. The time plots show the existence of possible aberrant observations, especially at the beginning of the series. Repeat the analyses of Problem 6, but use the subsample for t from 201 to 791.

2.8 Consider the quarterly U.S. federal government debt from the first quarter of 1970 to the third quarter of 2012. The first series is the federal debt held by foreign and international investors and the second series is federal debt held by Federal Reserve Banks. The data are from Federal Reserve Bank of St Louis, in billions of dollars, and are not seasonally adjusted. See the file `q-fdebt.txt`.

- Construct the bivariate time series z_t of the first difference of log federal debt series. Plot the data.
- Fit a VAR(6) model to the z_t series. Perform model checking. Is the model adequate? Why?
- Perform a chi-square test to verify that all coefficient estimates of the VAR(6) model with t-ratios less than 1.96 can indeed be removed based on an approximate 5% type I error.
- Based on the simplified model, is there any Granger causality between the two time series? Why?

2.9 Consider the quarterly growth rates (percentage change a year ago) of real gross domestic products of Brazil, South Korea, and Israel from 1997.I to 2012.II for 62 observations. The data are from the Federal Reserve Bank of St. Louis and in the file `q-rdgp-brkris.txt`.

- Specify a VAR model for the three-dimensional series.
- Fit the specified model, refine it if necessary, and perform model checking. Is the model adequate? Why?
- Compute the impulse response functions (using the observed innovations) of the fitted model. State the implications of the model.

2.10 Consider the monthly unemployment rates of the States of Illinois, Michigan, and Ohio of the United States from January 1976 to November 2009. The data were seasonally adjusted and are available from FRED of the Federal Reserve Bank of St. Louis. See also the file `m-3state-un.txt`.

- Build a VAR model for the three unemployment rates. Write down the fitted model.
- Use the impulse response functions of the VAR model built to study the relationship between the three unemployment rates. Describe the relationships.
- Use the fitted model to produce point and interval forecasts for the unemployment rates of December 2009 and January 2010 at the forecast origin of November 2009.

REFERENCES

Ahn, S. K. (1988). Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika*, **75**: 590–593.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (eds.). *2nd International Symposium on Information Theory*, pp. 267–281. Akademia Kiado, Budapest.

Basu, A. K. and Sen Roy, S. (1986). On some asymptotic results for multivariate autoregressive models with estimated parameters. *Calcutta Statistical Association Bulletin*, **35**: 123–132.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*. 4th Edition. John Wiley & Sons, Inc, Hoboken, NJ.

Doan, T., Litterman, R. B., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**: 1–144.

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective. *Econometric Reviews*, **18**: 343–386.

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

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**: 424–438.

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.

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

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

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

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics*, **4**: 25–38.

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

Quinn, B. G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society, Series B*, **42**: 182–185.

Rossi, P. E., Allenby, G. M., and McCulloch, R. E. (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons, Inc, Hoboken, NJ.

Samaranayake, V. A. and Hasza, D. P. (1988). Properties of predictors for multivariate autoregressive models with estimated parameters. *Journal of Time Series Analysis*, **9**: 361–383.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**: 461–464.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, **8**: 147–164.

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

Tsay, R. S. (2010). *Analysis of Financial Time Series*. 3rd Edition. John Wiley & Sons, Inc, Hoboken, NJ.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal the American Statistical Association*, **57**: 348–368.