unconditional likelihood. For brevity, we will hereafter refer to [11.1.2] simply as the "likelihood function" and the value of $\boldsymbol{\theta}$ that maximizes [11.1.2] as the "maximum likelihood estimate."

The likelihood function is calculated in the same way as for a scalar autoregression. Conditional on the values of $\mathbf{y}$ observed through date $t-1$, the value of $\mathbf{y}$ for date t is equal to a constant,

$$\mathbf{c} + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p}, \qquad [11.1.3]$$

plus a $N(\mathbf{0}, \boldsymbol{\Omega})$ variable. Thus,

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1} \sim N\Big((\mathbf{c} + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p}), \boldsymbol{\Omega}\Big). \qquad [11.1.4]$$

It will be convenient to use a more compact expression for the conditional mean [11.1.3]. Let $\mathbf{x}_t$ denote a vector containing a constant term and p lags of each of the elements of $\mathbf{y}$:

$$\mathbf{x}_t \equiv \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}. \qquad [11.1.5]$$

Thus, $\mathbf{x}_t$ is an $[(np + 1) \times 1]$ vector. Let $\boldsymbol{\Pi}'$ denote the following $[n \times (np + 1)]$ matrix:

$$\boldsymbol{\Pi}' \equiv [\mathbf{c} \quad \boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \cdots \quad \boldsymbol{\Phi}_p]. \qquad [11.1.6]$$

Then the conditional mean [11.1.3] is equal to $\boldsymbol{\Pi}'\mathbf{x}_t$. The jth row of $\boldsymbol{\Pi}'$ contains the parameters of the jth equation in the *VAR*. Using this notation, [11.1.4] can be written more compactly as

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1} \sim N(\boldsymbol{\Pi}'\mathbf{x}_t, \boldsymbol{\Omega}). \qquad [11.1.7]$$

Thus, the conditional density of the tth observation is

$$\begin{aligned} &f_{\mathbf{Y}_t|\mathbf{Y}_{t-1},\mathbf{Y}_{t-2},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &\quad = (2\pi)^{-n/2}|\boldsymbol{\Omega}^{-1}|^{1/2} \exp[(-1/2)(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)]. \end{aligned} \qquad [11.1.8]$$

The joint density of observations 1 through t conditioned on $\mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}$ satisfies

$$\begin{aligned} &f_{\mathbf{Y}_t,\mathbf{Y}_{t-1},\ldots,\mathbf{Y}_1|\mathbf{Y}_0,\mathbf{Y}_{-1},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &\quad = f_{\mathbf{Y}_{t-1},\ldots,\mathbf{Y}_1|\mathbf{Y}_0,\mathbf{Y}_{-1},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &\quad\quad \times f_{\mathbf{Y}_t|\mathbf{Y}_{t-1},\mathbf{Y}_{t-2},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}). \end{aligned}$$

Applying this formula recursively, the likelihood for the full sample $\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1$ conditioned on $\mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}$ is the product of the individual conditional densities:

$$\begin{aligned} &f_{\mathbf{Y}_T,\mathbf{Y}_{T-1},\ldots,\mathbf{Y}_1|\mathbf{Y}_0,\mathbf{Y}_{-1},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1 | \mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}) \\ &\quad = \prod_{t=1}^{T} f_{\mathbf{Y}_t|\mathbf{Y}_{t-1},\mathbf{Y}_{t-2},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta}). \end{aligned} \qquad [11.1.9]$$

The sample log likelihood is found by substituting [11.1.8] into [11.1.9] and taking logs:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f_{\mathbf{Y}_t|\mathbf{Y}_{t-1},\mathbf{Y}_{t-2},\ldots,\mathbf{Y}_{-p+1}}(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1}; \boldsymbol{\theta})$$

$$= -(Tn/2)\log(2\pi) + (T/2)\log|\boldsymbol{\Omega}^{-1}| - (1/2)\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right]. \qquad [11.1.10]$$

Maximum Likelihood Estimate of $\boldsymbol{\Pi}$

Consider first the *MLE* of $\boldsymbol{\Pi}$, which contains the constant term $\mathbf{c}$ and autoregressive coefficients $\boldsymbol{\Phi}_j$. This turns out to be given by

$$\underset{[n\times(np+1)]}{\hat{\boldsymbol{\Pi}}'} = \left[\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_t'\right]\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}, \qquad [11.1.11]$$

which can be viewed as the sample analog of the population linear projection of $\mathbf{y}_t$ on a constant and $\mathbf{x}_t$ (equation [4.1.23]). The jth row of $\hat{\boldsymbol{\Pi}}'$ is

$$\underset{[1\times(np+1)]}{\hat{\boldsymbol{\pi}}_j'} = \left[\sum_{t=1}^{T}y_{jt}\mathbf{x}_t'\right]\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}, \qquad [11.1.12]$$

which is just the estimated coefficient vector from an *OLS* regression of y_{jt} on $\mathbf{x}_t$. Thus, maximum likelihood estimates of the coefficients for the jth equation of a *VAR* are found by an *OLS* regression of y_{jt} on a constant term and p lags of all of the variables in the system.

To verify [11.1.11], write the sum appearing in the last term in [11.1.10] as

$$\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right]$$

$$= \sum_{t=1}^{T}\left[(\mathbf{y}_t - \hat{\boldsymbol{\Pi}}'\mathbf{x}_t + \hat{\boldsymbol{\Pi}}'\mathbf{x}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\Pi}}'\mathbf{x}_t + \hat{\boldsymbol{\Pi}}'\mathbf{x}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right] \qquad [11.1.13]$$

$$= \sum_{t=1}^{T}\left[[\hat{\boldsymbol{\varepsilon}}_t + (\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t]'\boldsymbol{\Omega}^{-1}[\hat{\boldsymbol{\varepsilon}}_t + (\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t]\right],$$

where the jth element of the $(n \times 1)$ vector $\hat{\boldsymbol{\varepsilon}}_t$ is the sample residual for observation t from an *OLS* regression of y_{jt} on $\mathbf{x}_t$:

$$\hat{\boldsymbol{\varepsilon}}_t \equiv \mathbf{y}_t - \hat{\boldsymbol{\Pi}}'\mathbf{x}_t. \qquad [11.1.14]$$

Expression [11.1.13] can be expanded as

$$\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right]$$

$$= \sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_t + 2\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t + \sum_{t=1}^{T}\mathbf{x}_t'(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t. \qquad [11.1.15]$$

Consider the middle term in [11.1.15]. Since this is a scalar, it is unchanged by applying the "trace" operator:

$$\begin{aligned}\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t &= \operatorname{trace}\left[\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t\right] \\ &= \operatorname{trace}\left[\sum_{t=1}^{T} \boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t\hat{\boldsymbol{\varepsilon}}_t'\right] \\ &= \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\sum_{t=1}^{T} \mathbf{x}_t\hat{\boldsymbol{\varepsilon}}_t'\right].\end{aligned} \qquad [11.1.16]$$

But the sample residuals from an *OLS* regression are by construction orthogonal to the explanatory variables, meaning that $\Sigma_{t=1}^{T}\mathbf{x}_t\hat{\varepsilon}_{jt} = \mathbf{0}$ for all j and so $\Sigma_{t=1}^{T}\mathbf{x}_t\hat{\boldsymbol{\varepsilon}}_t' = \mathbf{0}$. Hence, [11.1.16] is identically zero, and [11.1.15] simplifies to

$$\begin{aligned}&\sum_{t=1}^{T}\left[(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right] \\ &\quad = \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_t + \sum_{t=1}^{T} \mathbf{x}_t'(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t.\end{aligned} \qquad [11.1.17]$$

Since $\boldsymbol{\Omega}$ is a positive definite matrix, $\boldsymbol{\Omega}^{-1}$ is as well.[1] Thus, defining the $(n \times 1)$ vector $\mathbf{x}_t^*$ as

$$\mathbf{x}_t^* \equiv (\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t,$$

the last term in [11.1.17] takes the form

$$\sum_{t=1}^{T} \mathbf{x}_t'(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})'\mathbf{x}_t = \sum_{t=1}^{T} [\mathbf{x}_t^*]'\boldsymbol{\Omega}^{-1}\mathbf{x}_t^*.$$

This is positive for any sequence $\{\mathbf{x}_t^*\}_{t=1}^{T}$ other than $\mathbf{x}_t^* = \mathbf{0}$ for all t. Thus, the smallest value that [11.1.17] can take on is achieved when $\mathbf{x}_t^* = \mathbf{0}$, or when $\boldsymbol{\Pi} = \hat{\boldsymbol{\Pi}}$. Since [11.1.17] is minimized by setting $\boldsymbol{\Pi} = \hat{\boldsymbol{\Pi}}$, it follows that [11.1.10] is maximized by setting $\boldsymbol{\Pi} = \hat{\boldsymbol{\Pi}}$, establishing the claim that *OLS* regressions provide the maximum likelihood estimates of the coefficients of a vector autoregression.

Some Useful Results on Matrix Derivatives

The next task is to calculate the maximum likelihood estimate of $\boldsymbol{\Omega}$. Here two results from matrix calculus will prove helpful. The first result concerns the derivative of a quadratic form in a matrix. Let a_{ij} denote the row i, column j element of an $(n \times n)$ matrix $\mathbf{A}$. Suppose that the matrix $\mathbf{A}$ is nonsymmetric and unrestricted (that is, the value of a_{ij} is unrelated to the value of a_{kl} when either $i \neq k$ or $j \neq l$). Consider a quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ for $\mathbf{x}$ an $(n \times 1)$ vector. The quadratic form can be written out explicitly as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j, \qquad [11.1.18]$$

from which

$$\frac{\partial\, \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial a_{ij}} = x_i x_j. \qquad [11.1.19]$$

[1]This follows immediately from the fact that $\boldsymbol{\Omega}^{-1}$ can be written as $\mathbf{L}'\mathbf{L}$ for $\mathbf{L}$ a nonsingular matrix as in [8.3.1].

Collecting these n^2 different derivatives into an $(n \times n)$ matrix, equation [11.1.19] can conveniently be expressed in matrix form as

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{A}} = \mathbf{x}\mathbf{x}'. \qquad [11.1.20]$$

The second result concerns the derivative of the determinant of a matrix. Let $\mathbf{A}$ be a nonsymmetric unrestricted $(n \times n)$ matrix with positive determinant. Then

$$\frac{\partial \log|\mathbf{A}|}{\partial a_{ij}} = a^{ji}, \qquad [11.1.21]$$

where a^{ji} denotes the row j, column i element of $\mathbf{A}^{-1}$. In matrix form,

$$\frac{\partial \log|\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}')^{-1}. \qquad [11.1.22]$$

To derive [11.1.22], recall the formula for the determinant of $\mathbf{A}$ (equation [A.4.10] in the Mathematical Review, Appendix A, at the end of the book):

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|, \qquad [11.1.23]$$

where $\mathbf{A}_{ij}$ denotes the $(n - 1) \times (n - 1)$ matrix formed by deleting row i and column j from $\mathbf{A}$. The derivative of [11.1.23] with respect to a_{ij} is

$$\frac{\partial|\mathbf{A}|}{\partial a_{ij}} = (-1)^{i+j}|\mathbf{A}_{ij}|, \qquad [11.1.24]$$

since the parameter a_{ij} does not appear in the matrix $\mathbf{A}_{ij}$. It follows that

$$\frac{\partial \log|\mathbf{A}|}{\partial a_{ij}} = (1/|\mathbf{A}|)\cdot(-1)^{i+j}|\mathbf{A}_{ij}|,$$

which will be recognized from equation [A.4.12] as the row j, column i element of $\mathbf{A}^{-1}$, as claimed in equation [11.1.22].

The Maximum Likelihood Estimate of $\boldsymbol{\Omega}$

We now apply these results to find the *MLE* of $\boldsymbol{\Omega}$. When evaluated at the *MLE* $\hat{\boldsymbol{\Pi}}$, the log likelihood [11.1.10] is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Pi}}) = &-(Tn/2)\log(2\pi) + (T/2)\log|\boldsymbol{\Omega}^{-1}| \\ &- (1/2)\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_t.\end{aligned} \qquad [11.1.25]$$

Our objective is to find a symmetric positive definite matrix $\boldsymbol{\Omega}$ for which this is as large as possible. It is instructive to consider first maximizing [11.1.25] by choosing $\boldsymbol{\Omega}$ to be any unrestricted $(n \times n)$ matrix. For that purpose we can just differentiate [11.1.25] with respect to the elements of $\boldsymbol{\Omega}^{-1}$ using formulas [11.1.20] and [11.1.22]:

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{\Omega}, \hat{\boldsymbol{\Pi}})}{\partial\boldsymbol{\Omega}^{-1}} &= (T/2)\frac{\partial \log|\boldsymbol{\Omega}^{-1}|}{\partial\boldsymbol{\Omega}^{-1}} - (1/2)\sum_{t=1}^{T}\frac{\partial \hat{\boldsymbol{\varepsilon}}_t'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_t}{\partial\boldsymbol{\Omega}^{-1}} \\ &= (T/2)\boldsymbol{\Omega}' - (1/2)\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'.\end{aligned} \qquad [11.1.26]$$

The likelihood is maximized when this derivative is set to zero, or when

$$\boldsymbol{\Omega}' = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'. \qquad [11.1.27]$$

The matrix $\mathbf{\Omega}$ that satisfies [11.1.27] maximizes the likelihood among the class of all unrestricted $(n \times n)$ matrices. Note, however, that the optimal unrestricted value for $\mathbf{\Omega}$ that is specified by [11.1.27] turns out to be symmetric and positive definite. The *MLE*, or the value of $\mathbf{\Omega}$ that maximizes the likelihood among the class of all symmetric positive definite matrices, is thus also given by [11.1.27]:

$$\hat{\mathbf{\Omega}} = (1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'. \qquad [11.1.28]$$

The row i, column i element of $\hat{\mathbf{\Omega}}$ is given by

$$\hat{\sigma}_i^2 = (1/T) \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2, \qquad [11.1.29]$$

which is just the average squared residual from a regression of the ith variable in the *VAR* on a constant term and p lags of all the variables. The row i, column j element of $\hat{\mathbf{\Omega}}$ is

$$\hat{\sigma}_{ij} = (1/T) \sum_{t=1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}, \qquad [11.1.30]$$

which is the average product of the *OLS* residual for variable i and the *OLS* residual for variable j.

Likelihood Ratio Tests

To perform a likelihood ratio test, we need to calculate the maximum value achieved for [11.1.25]. Thus, consider

$$\begin{aligned} \mathscr{L}(\hat{\mathbf{\Omega}}, \hat{\mathbf{\Pi}}) = & -(Tn/2) \log(2\pi) + (T/2) \log|\hat{\mathbf{\Omega}}^{-1}| \\ & - (1/2) \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t' \hat{\mathbf{\Omega}}^{-1} \hat{\boldsymbol{\varepsilon}}_t \end{aligned} \qquad [11.1.31]$$

for $\hat{\mathbf{\Omega}}$ given by [11.1.28]. The last term in [11.1.31] is

$$\begin{aligned} (1/2) \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t' \hat{\mathbf{\Omega}}^{-1} \hat{\boldsymbol{\varepsilon}}_t &= (1/2) \operatorname{trace}\left[\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t' \hat{\mathbf{\Omega}}^{-1} \hat{\boldsymbol{\varepsilon}}_t\right] \\ &= (1/2) \operatorname{trace}\left[\sum_{t=1}^{T} \hat{\mathbf{\Omega}}^{-1} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'\right] \\ &= (1/2) \operatorname{trace}[\hat{\mathbf{\Omega}}^{-1}(T\hat{\mathbf{\Omega}})] \\ &= (1/2) \operatorname{trace}(T \cdot \mathbf{I}_n) \\ &= Tn/2. \end{aligned}$$

Substituting this into [11.1.31] produces

$$\mathscr{L}(\hat{\mathbf{\Omega}}, \hat{\mathbf{\Pi}}) = -(Tn/2) \log(2\pi) + (T/2) \log|\hat{\mathbf{\Omega}}^{-1}| - (Tn/2). \qquad [11.1.32]$$

This makes likelihood ratio tests particularly simple to perform. Suppose we want to test the null hypothesis that a set of variables was generated from a Gaussian *VAR* with p_0 lags against the alternative specification of $p_1 > p_0$ lags. To estimate the system under the null hypothesis, we perform a set of n *OLS* regressions of each variable in the system on a constant term and on p_0 lags of all the variables in the system. Let $\hat{\mathbf{\Omega}}_0 = (1/T)\Sigma_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t(p_0)[\hat{\boldsymbol{\varepsilon}}_t(p_0)]'$ be the variance-covariance matrix of the residuals from these regressions. The maximum value for the log likelihood under H_0 is then

$$\mathcal{L}_0^* = -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\mathbf{\Omega}}_0^{-1}| - (Tn/2).$$

Similarly, the system is estimated under the alternative hypothesis by OLS regressions that include p_1 lags of all the variables. The maximized log likelihood under the alternative is

$$\mathcal{L}_1^* = -(Tn/2)\log(2\pi) + (T/2)\log|\hat{\mathbf{\Omega}}_1^{-1}| - (Tn/2),$$

where $\hat{\mathbf{\Omega}}_1$ is the variance-covariance matrix of the residuals from this second set of regressions. Twice the log likelihood ratio is then

$$\begin{aligned} 2(\mathcal{L}_1^* - \mathcal{L}_0^*) &= 2\{(T/2)\log|\hat{\mathbf{\Omega}}_1^{-1}| - (T/2)\log|\hat{\mathbf{\Omega}}_0^{-1}|\} \\ &= T\log(1/|\hat{\mathbf{\Omega}}_1|) - T\log(1/|\hat{\mathbf{\Omega}}_0|) \\ &= -T\log|\hat{\mathbf{\Omega}}_1| + T\log|\hat{\mathbf{\Omega}}_0| \\ &= T\{\log|\hat{\mathbf{\Omega}}_0| - \log|\hat{\mathbf{\Omega}}_1|\}. \end{aligned} \qquad [11.1.33]$$

Under the null hypothesis, this asymptotically has a χ^2 distribution with degrees of freedom equal to the number of restrictions imposed under H_0. Each equation in the specification restricted by H_0 has $(p_1 - p_0)$ fewer lags on each of n variables compared with H_1; thus, H_0 imposes $n(p_1 - p_0)$ restrictions on each equation. Since there are n such equations, H_0 imposes $n^2(p_1 - p_0)$ restrictions. Thus, the magnitude calculated in [11.1.33] is asymptotically χ^2 with $n^2(p_1 - p_0)$ degrees of freedom.

For example, suppose a bivariate VAR is estimated with three and four lags ($n = 2$, $p_0 = 3$, $p_1 = 4$). Say that the original sample contains 50 observations on each variable (denoted $\mathbf{y}_{-3}, \mathbf{y}_{-2}, \ldots, \mathbf{y}_{46}$) and that observations 1 through 46 were used to estimate both the three- and four-lag specifications so that $T = 46$. Let $\hat{\varepsilon}_{it}(p_0)$ be the sample residual for observation t from an OLS regression of y_{it} on a constant, three lags of y_{1t}, and three lags of y_{2t}. Suppose that $(1/T)\sum_{t=1}^{T}[\hat{\varepsilon}_{1t}(p_0)]^2 = 2.0$, $(1/T)\sum_{t=1}^{T}[\hat{\varepsilon}_{2t}(p_0)]^2 = 2.5$, and $(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_{1t}(p_0)\hat{\varepsilon}_{2t}(p_0) = 1.0$. Then

$$\hat{\mathbf{\Omega}}_0 = \begin{bmatrix} 2.0 & 1.0 \\ 1.0 & 2.5 \end{bmatrix}$$

and $\log|\hat{\mathbf{\Omega}}_0| = \log 4 = 1.386$. Suppose that when a fourth lag is added to each regression, the residual covariance matrix is reduced to

$$\hat{\mathbf{\Omega}}_1 = \begin{bmatrix} 1.8 & 0.9 \\ 0.9 & 2.2 \end{bmatrix},$$

for which $\log|\hat{\mathbf{\Omega}}_1| = 1.147$. Then

$$2(\mathcal{L}_1^* - \mathcal{L}_0^*) = 46(1.386 - 1.147) = 10.99.$$

The degrees of freedom for this test are $2^2(4 - 3) = 4$. Since $10.99 > 9.49$ (the 5% critical value for a $\chi^2(4)$ variable), the null hypothesis is rejected. The dynamics are not completely captured by a three-lag VAR, and a four-lag specification seems preferable.

Sims (1980, p. 17) suggested a modification to the likelihood ratio test to take into account small-sample bias. He recommended replacing [11.1.33] by

$$(T - k)\{\log|\hat{\mathbf{\Omega}}_0| - \log|\hat{\mathbf{\Omega}}_1|\}, \qquad [11.1.34]$$

where $k = 1 + np_1$ is the number of parameters estimated per equation. The adjusted test has the same asymptotic distribution as [11.1.33] but is less likely to reject the null hypothesis in small samples. For the present example, this test statistic would be

$$(46 - 9)(1.386 - 1.147) = 8.84,$$

and the earlier conclusion would be reversed (H_0 would be accepted).

Asymptotic Distribution of $\hat{\mathbf{\Pi}}$

The maximum likelihood estimates $\hat{\mathbf{\Pi}}$ and $\hat{\mathbf{\Omega}}$ will give consistent estimates of the population parameters even if the true innovations are non-Gaussian. Standard errors for $\hat{\mathbf{\Pi}}$ can be based on the usual *OLS* formulas, as the following proposition demonstrates.

Proposition 11.1: *Let*

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t$ is independent and identically distributed with mean $\mathbf{0}$, variance $\mathbf{\Omega}$, and $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) < \infty$ for all i, j, l, and m and where roots of

$$|\mathbf{I}_n - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots - \mathbf{\Phi}_p z^p| = 0 \qquad [11.1.35]$$

lie outside the unit circle. Let $k \equiv np + 1$, and let $\mathbf{x}_t'$ be the $(1 \times k)$ vector

$$\mathbf{x}_t' \equiv [1 \quad \mathbf{y}_{t-1}' \quad \mathbf{y}_{t-2}' \quad \cdots \quad \mathbf{y}_{t-p}'].$$

Let $\hat{\boldsymbol{\pi}}_T = \text{vec}(\hat{\mathbf{\Pi}}_T)$ denote the $(nk \times 1)$ vector of coefficients resulting from OLS regressions of each of the elements of $\mathbf{y}_t$ on $\mathbf{x}_t$ for a sample of size T:

$$\hat{\boldsymbol{\pi}}_T = \begin{bmatrix} \hat{\boldsymbol{\pi}}_{1.T} \\ \hat{\boldsymbol{\pi}}_{2.T} \\ \vdots \\ \hat{\boldsymbol{\pi}}_{n.T} \end{bmatrix},$$

where

$$\hat{\boldsymbol{\pi}}_{i.T} = \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t y_{it}\right];$$

and let $\boldsymbol{\pi}$ denote the $(nk \times 1)$ vector of corresponding population coefficients. Finally, let

$$\hat{\mathbf{\Omega}}_T = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t',$$

where

$$\hat{\boldsymbol{\varepsilon}}_t' = [\hat{\varepsilon}_{1t} \quad \hat{\varepsilon}_{2t} \quad \cdots \quad \hat{\varepsilon}_{nt}]$$
$$\hat{\varepsilon}_{it} = y_{it} - \mathbf{x}_t'\hat{\boldsymbol{\pi}}_{i.T}.$$

Then

(*a*) $(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{Q}$ *where* $\mathbf{Q} = E(\mathbf{x}_t\mathbf{x}_t')$;

(*b*) $\hat{\boldsymbol{\pi}}_T \xrightarrow{p} \boldsymbol{\pi}$;

(c) $\hat{\mathbf{\Omega}}_T \xrightarrow{p} \mathbf{\Omega}$;

(d) $\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{L} N(\mathbf{0}, (\mathbf{\Omega} \otimes \mathbf{Q}^{-1}))$, *where* $\otimes$ *denotes the Kronecker product.*

A proof of this proposition is provided in Appendix 11.A to this chapter.

If we are interested only in $\hat{\boldsymbol{\pi}}_{i,T}$, the coefficients of the ith regression in the *VAR*, result (d) implies that

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_{i,T} - \boldsymbol{\pi}_i) \xrightarrow{L} N(\mathbf{0}, \sigma_i^2\mathbf{Q}^{-1}), \qquad [11.1.36]$$

where $\sigma_i^2 = E(\varepsilon_{it}^2)$ is the variance of the innovation of the ith equation in the *VAR*. But σ_i^2 is estimated consistently by $\hat{\sigma}_i^2 = (1/T)\Sigma_{t=1}^T \hat{\varepsilon}_{it}^2$, the average squared residual from *OLS* estimation of this equation. Similarly, $\mathbf{Q}^{-1}$ is estimated consistently by $[(1/T)\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t']^{-1}$. Hence, [11.1.36] invites us to treat $\hat{\boldsymbol{\pi}}_i$ approximately as

$$\hat{\boldsymbol{\pi}}_i \approx N\left(\boldsymbol{\pi}_i, \hat{\sigma}_i^2\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\right). \qquad [11.1.37]$$

But this is the standard *OLS* formula for coefficient variances with $s_i^2 = [1/(T-k)]\Sigma_{t=1}^T \hat{\varepsilon}_{it}^2$ in the standard formula replaced by the maximum likelihood estimate $\hat{\sigma}_i^2$ in [11.1.37]. Clearly, s_i^2 and $\hat{\sigma}_i^2$ are asymptotically equivalent, though following Sims's argument in [11.1.34], the larger (and thus more conservative) standard errors resulting from the *OLS* formulas might be preferred. Hence, Proposition 11.1 establishes that the standard *OLS* t and F statistics applied to the coefficients of any single equation in the *VAR* are asymptotically valid and can be evaluated in the usual way.

A more general hypothesis of the form $\mathbf{R}\boldsymbol{\pi} = \mathbf{r}$ involving coefficients across different equations of the *VAR* can be tested using a generalization of the Wald form of the *OLS* χ^2 test (expression [8.2.23]). Result (d) of Proposition 11.1 establishes that

$$\sqrt{T}(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r}) \xrightarrow{L} N\left(\mathbf{0}, \mathbf{R}(\mathbf{\Omega} \otimes \mathbf{Q}^{-1})\mathbf{R}'\right).$$

In the light of results (a) and (c), the asymptotic distribution could equivalently be described as

$$\sqrt{T}(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r}) \xrightarrow{p} N\left(\mathbf{0}, \mathbf{R}(\hat{\mathbf{\Omega}}_T \otimes \mathbf{Q}_T^{-1})\mathbf{R}'\right),$$

where $\hat{\mathbf{\Omega}}_T = (1/T)\,\Sigma_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$ and $\mathbf{Q}_T = (1/T)\,\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t'$. Hence, the following statistic has an asymptotic χ^2 distribution:

$$\begin{aligned}
\chi^2(m) &= T(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r})'\left(\mathbf{R}(\hat{\mathbf{\Omega}}_T \otimes \mathbf{Q}_T^{-1})\mathbf{R}'\right)^{-1}(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r}) \\
&= (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r})'\left(\mathbf{R}[\hat{\mathbf{\Omega}}_T \otimes (T\mathbf{Q}_T)^{-1}]\mathbf{R}'\right)^{-1}(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r}) \\
&= (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r})'\left\{\mathbf{R}\left[\hat{\mathbf{\Omega}}_T \otimes \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right]\mathbf{R}'\right\}^{-1}(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{r}).
\end{aligned} \qquad [11.1.38]$$

The degrees of freedom for this statistic are given by the number of rows of $\mathbf{R}$, or the number of restrictions tested.

For example, suppose we wanted to test the hypothesis that the constant term in the first equation in the *VAR* (c_1) is equal to the constant term in the second equation (c_2). Then $\mathbf{R}$ is a $(1 \times nk)$ vector with unity in the first position, -1 in the $(k + 1)$th position, and zeros elsewhere:

$$\mathbf{R} = [1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad -1 \quad 0 \quad 0 \quad \cdots \quad 0].$$

To apply result [11.1.38], it is convenient to write $\mathbf{R}$ in Kronecker product form as

$$\mathbf{R} = \mathbf{R}_n \otimes \mathbf{R}_k, \qquad [11.1.39]$$

where $\mathbf{R}_n$ selects the equations that are involved and $\mathbf{R}_k$ selects the coefficients. For this example,

$$\underset{(1\times n)}{\mathbf{R}_n} = [1 \quad -1 \quad 0 \quad 0 \quad \cdots \quad 0]$$
$$\underset{(1\times k)}{\mathbf{R}_k} = [1 \quad 0 \quad 0 \quad 0 \quad \cdots \quad 0].$$

We then calculate

$$\begin{aligned}\mathbf{R}\left[\hat{\mathbf{\Omega}} \otimes \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right]\mathbf{R}' &= (\mathbf{R}_n \otimes \mathbf{R}_k)\left[\hat{\mathbf{\Omega}} \otimes \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right](\mathbf{R}_n' \otimes \mathbf{R}_k') \\ &= (\mathbf{R}_n\hat{\mathbf{\Omega}}\mathbf{R}_n') \otimes \left[\mathbf{R}_k \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1} \mathbf{R}_k'\right] \\ &= (\hat{\sigma}_1^2 - 2\hat{\sigma}_{12} + \hat{\sigma}_2^2) \otimes \xi^{11},\end{aligned}$$

where $\hat{\sigma}_{12}$ is the covariance between $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$ and ξ^{11} is the (1, 1) element of $(\Sigma_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t')^{-1}$. Since ξ^{11} is a scalar, the foregoing Kronecker product is a simple multiplication. The test statistic [11.1.38] is then

$$\chi^2(1) = \frac{(\hat{c}_1 - \hat{c}_2)^2}{(\hat{\sigma}_1^2 - 2\hat{\sigma}_{12} + \hat{\sigma}_2^2)\xi^{11}}.$$

Asymptotic Distribution of $\hat{\mathbf{\Omega}}$

In considering the asymptotic distribution of the estimates of variances and covariances, notice that since $\mathbf{\Omega}$ is symmetric, some of its elements are redundant. Recall that the "vec" operator transforms an $(n \times n)$ matrix into an $(n^2 \times 1)$ vector by stacking the columns. For example,

$$\operatorname{vec}\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{12} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{bmatrix}. \qquad [11.1.40]$$

An analogous "vech" operator transforms an $(n \times n)$ matrix into an $([n(n + 1)/2] \times 1)$ vector by vertically stacking those elements on or below the principal diagonal. For example,

$$\text{vech}\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix} = \begin{bmatrix}\sigma_{11}\\ \sigma_{21}\\ \sigma_{31}\\ \sigma_{22}\\ \sigma_{32}\\ \sigma_{33}\end{bmatrix}. \qquad [11.1.41]$$

Proposition 11.2: *Let*

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim$ i.i.d. $N(\mathbf{0}, \mathbf{\Omega})$ and where roots of

$$|\mathbf{I}_n - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots - \mathbf{\Phi}_p z^p| = 0$$

lie outside the unit circle. Let $\hat{\boldsymbol{\pi}}_T$, $\hat{\mathbf{\Omega}}_T$, and $\mathbf{Q}$ be as defined in Proposition 11.1. Then

$$\begin{bmatrix}\sqrt{T}[\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}]\\ \sqrt{T}[\text{vech}(\hat{\mathbf{\Omega}}_T) - \text{vech}(\mathbf{\Omega})]\end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}(\mathbf{\Omega}\otimes\mathbf{Q}^{-1}) & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{22}\end{bmatrix}\right).$$

Let σ_{ij} denote the row i, column j element of $\mathbf{\Omega}$; for example, σ_{11} is the variance of ε_{1t}. Then the element of $\mathbf{\Sigma}_{22}$ corresponding to the covariance between $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{lm}$ is given by $(\sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl})$ for all $i, j, l, m = 1, 2, \ldots, n$, including $i = j = l = m$.

For example, for $n = 2$, Proposition 11.2 implies that

$$\sqrt{T}\begin{bmatrix}\hat{\sigma}_{11,T} - \sigma_{11}\\ \hat{\sigma}_{12,T} - \sigma_{12}\\ \hat{\sigma}_{22,T} - \sigma_{22}\end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2\\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22}\\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2\end{bmatrix}\right). \qquad [11.1.42]$$

Thus, a Wald test of the null hypothesis that there is no covariance between ε_{1t} and ε_{2t} is given by

$$\frac{\sqrt{T}\hat{\sigma}_{12}}{(\hat{\sigma}_{11}\hat{\sigma}_{22} + \hat{\sigma}_{12}^2)^{1/2}} \approx N(0, 1).$$

A Wald test of the null hypothesis that ε_{1t} and ε_{2t} have the same variance is given by

$$\frac{T(\hat{\sigma}_{11} - \hat{\sigma}_{22})^2}{2\hat{\sigma}_{11}^2 - 4\hat{\sigma}_{12}^2 + 2\hat{\sigma}_{22}^2} \approx \chi^2(1),$$

where $\hat{\sigma}_{11}^2$ denotes the square of the estimated variance of the innovation for the first equation.

The matrix $\mathbf{\Sigma}_{22}$ in Proposition 11.2 can be expressed more compactly using the *duplication matrix*. Notice that since $\mathbf{\Omega}$ is symmetric, the n^2 elements of vec($\mathbf{\Omega}$) in [11.1.40] are simple duplications of the $n(n + 1)/2$ elements of vech($\mathbf{\Omega}$) in [11.1.41]. There exists a unique $[n^2 \times n(n + 1)/2]$ matrix $\mathbf{D}_n$ that transforms vech($\mathbf{\Omega}$) into vec($\mathbf{\Omega}$), that is, a unique matrix satisfying

$$\mathbf{D}_n\,\text{vech}(\mathbf{\Omega}) = \text{vec}(\mathbf{\Omega}). \qquad [11.1.43]$$

For example, for $n = 2$, equation [11.1.43] is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}. \qquad [11.1.44]$$

Further, define $\mathbf{D}_n^+$ to be the following $[n(n+1)/2 \times n^2]$ matrix:[2]

$$\mathbf{D}_n^+ \equiv (\mathbf{D}_n'\mathbf{D}_n)^{-1}\mathbf{D}_n'. \qquad [11.1.45]$$

Notice that $\mathbf{D}_n^+\mathbf{D}_n = \mathbf{I}_{n(n+1)/2}$. Thus, premultiplying both sides of [11.1.43] by $\mathbf{D}_n^+$ reveals $\mathbf{D}_n^+$ to be a matrix that transforms $\text{vec}(\mathbf{\Omega})$ into $\text{vech}(\mathbf{\Omega})$ for symmetric $\mathbf{\Omega}$:

$$\text{vech}(\mathbf{\Omega}) = \mathbf{D}_n^+ \,\text{vec}(\mathbf{\Omega}). \qquad [11.1.46]$$

For example, for $n = 2$, equation [11.1.46] is

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}. \qquad [11.1.47]$$

It turns out that the matrix $\mathbf{\Sigma}_{22}$ described in Proposition 11.2 can be written as[3]

$$\mathbf{\Sigma}_{22} = 2\mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)'. \qquad [11.1.48]$$

For example, for $n = 2$, expression [11.1.48] becomes

$$\begin{aligned} 2\mathbf{D}_2^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_2^+)' &= 2\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \sigma_{11}\sigma_{11} & \sigma_{11}\sigma_{12} & \sigma_{12}\sigma_{11} & \sigma_{12}\sigma_{12} \\ \sigma_{11}\sigma_{21} & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{21} & \sigma_{12}\sigma_{22} \\ \sigma_{21}\sigma_{11} & \sigma_{21}\sigma_{12} & \sigma_{22}\sigma_{11} & \sigma_{22}\sigma_{12} \\ \sigma_{21}\sigma_{21} & \sigma_{21}\sigma_{22} & \sigma_{22}\sigma_{21} & \sigma_{22}\sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2 \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2 \end{bmatrix}, \end{aligned}$$

which reproduces [11.1.42].

11.2. *Bivariate Granger Causality Tests*

One of the key questions that can be addressed with vector autoregressions is how useful some variables are for forecasting others. This section discusses a particular summary of the forecasting relation between two variables proposed by Granger (1969) and popularized by Sims (1972). A more general discussion of a related question in larger vector systems is provided in the following section.

[2]It can be shown that $(\mathbf{D}_n'\mathbf{D}_n)$ is nonsingular. For more details, see Magnus and Neudecker (1988, pp. 48–49).

[3]Magnus and Neudecker (1988, p. 318) derived this expression directly from the information matrix.

Definition of Bivariate Granger Causality

The question investigated in this section is whether a scalar y can help forecast another scalar x. If it cannot, then we say that y does not Granger-cause x. More formally, *y fails to Granger-cause x* if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, \ldots)$ is the same as the MSE of a forecast of x_{t+s} that uses both $(x_t, x_{t-1}, \ldots)$ and $(y_t, y_{t-1}, \ldots)$. If we restrict ourselves to linear functions, y fails to Granger-cause x if

$$MSE[\hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots)] = MSE[\hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots)]. \qquad [11.2.1]$$

Equivalently, we say that *x is exogenous in the time series sense with respect to y* if [11.2.1] holds. Yet a third expression meaning the same thing is that *y is not linearly informative about future x*.

Granger's reason for proposing this definition was that if an event Y is the cause of another event X, then the event Y should precede the event X. Although one might agree with this position philosophically, there can be serious obstacles to practical implementation of this idea using aggregate time series data, as will be seen in the examples considered later in this section. First, however, we explore the mechanical implications of Granger causality for the time series representation of a bivariate system.

Alternative Implications of Granger Causality

In a bivariate VAR describing x and y, y does not Granger-cause x if the coefficient matrices $\mathbf{\Phi}_j$ are lower triangular for all j:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(2)} & 0 \\ \phi_{21}^{(2)} & \phi_{22}^{(2)} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_{11}^{(p)} & 0 \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \qquad [11.2.2]$$

From the first row of this system, the optimal one-period-ahead forecast of x depends only on its own lagged values and not on lagged y:

$$\hat{E}(x_{t+1}|x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots) = c_1 + \phi_{11}^{(1)}x_t + \phi_{11}^{(2)}x_{t-1} + \cdots + \phi_{11}^{(p)}x_{t-p+1}. \qquad [11.2.3]$$

Furthermore, the value of x_{t+2} from [11.2.2] is given by

$$x_{t+2} = c_1 + \phi_{11}^{(1)}x_{t+1} + \phi_{11}^{(2)}x_t + \cdots + \phi_{11}^{(p)}x_{t-p+2} + \varepsilon_{1,t+2}.$$

Recalling [11.2.3] and the law of iterated projections, it is clear that the date t forecast of this magnitude on the basis of $(x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots)$ also depends only on $(x_t, x_{t-1}, \ldots, x_{t-p+1})$. By induction, the same is true of an s-period-ahead forecast. Thus, for the bivariate VAR, y does not Granger-cause x if $\mathbf{\Phi}_j$ is lower triangular for all j, as claimed.

Recall from equation [10.1.19] that

$$\mathbf{\Psi}_s = \mathbf{\Phi}_1\mathbf{\Psi}_{s-1} + \mathbf{\Phi}_2\mathbf{\Psi}_{s-2} + \cdots + \mathbf{\Phi}_p\mathbf{\Psi}_{s-p} \qquad \text{for } s = 1, 2, \ldots,$$

with $\mathbf{\Psi}_0$ the identity matrix and $\mathbf{\Psi}_s = \mathbf{0}$ for $s < 0$. This expression implies that if

Φ_j is lower triangular for all j, then the moving average matrices Ψ_s for the fundamental representation will be lower triangular for all s. Thus, if y fails to Granger-cause x, then the $MA(\infty)$ representation can be written

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \psi_{11}(L) & 0 \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \qquad [11.2.4]$$

where

$$\psi_{ij}(L) = \psi_{ij}^{(0)} + \psi_{ij}^{(1)}L^1 + \psi_{ij}^{(2)}L^2 + \psi_{ij}^{(3)}L^3 + \cdots$$

with $\psi_{11}^{(0)} = \psi_{22}^{(0)} = 1$ and $\psi_{21}^{(0)} = 0$.

Another implication of Granger causality was stressed by Sims (1972).

Proposition 11.3: *Consider a linear projection of y_t on past, present, and future x's,*

$$y_t = c + \sum_{j=0}^{\infty} b_j x_{t-j} + \sum_{j=1}^{\infty} d_j x_{t+j} + \eta_t, \qquad [11.2.5]$$

where b_j and d_j are defined as population projection coefficients, that is, the values for which

$$E(\eta_t x_\tau) = 0 \qquad \textit{for all } t \textit{ and } \tau.$$

Then y fails to Granger-cause x if and only if $d_j = 0$ for $j = 1, 2, \ldots$.

Econometric Tests for Granger Causality

Econometric tests of whether a particular observed series y Granger-causes x can be based on any of the three implications [11.2.2], [11.2.4], or [11.2.5]. The simplest and probably best approach uses the autoregressive specification [11.2.2]. To implement this test, we assume a particular autoregressive lag length p and estimate

$$\begin{aligned} x_t = c_1 &+ \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \beta_1 y_{t-1} \\ &+ \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_t \end{aligned} \qquad [11.2.6]$$

by OLS. We then conduct an F test of the null hypothesis

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_p = 0. \qquad [11.2.7]$$

Recalling Proposition 8.2, one way to implement this test is to calculate the sum of squared residuals from [11.2.6],[4]

$$RSS_1 = \sum_{t=1}^{T} \hat{u}_t^2,$$

and compare this with the sum of squared residuals of a univariate autoregression for x_t,

$$RSS_0 = \sum_{t=1}^{T} \hat{e}_t^2,$$

[4]Note that in order for t to run from 1 to T as indicated, we actually need $T + p$ observations on x and y, namely, $x_{-p+1}, x_{-p+2}, \ldots, x_T$ and $y_{-p+1}, y_{-p+2}, \ldots, y_T$.

where

$$x_t = c_0 + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \cdots + \gamma_p x_{t-p} + e_t \qquad [11.2.8]$$

is also estimated by OLS. If

$$S_1 \equiv \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)} \qquad [11.2.9]$$

is greater than the 5% critical value for an $F(p, T - 2p - 1)$ distribution, then we reject the null hypothesis that y does not Granger-cause x; that is, if S_1 is sufficiently large, we conclude that y does Granger-cause x.

The test statistic [11.2.9] would have an exact F distribution for a regression with fixed regressors and Gaussian disturbances. With lagged dependent variables as in the Granger-causality regressions, however, the test is valid only asymptotically. An asymptotically equivalent test is given by

$$S_2 \equiv \frac{T(RSS_0 - RSS_1)}{RSS_1}. \qquad [11.2.10]$$

We would reject the null hypothesis that y does not Granger-cause x if S_2 is greater than the 5% critical values for a $\chi^2(p)$ variable.

An alternative approach is to base the test on the Sims form [11.2.5] instead of the Granger form [11.2.2]. A problem with the Sims form is that the error term η_t is in general autocorrelated. Thus, a standard F test of the hypothesis that $d_j = 0$ for all j in [11.2.5] will not give the correct answer. One option is to use autocorrelation-consistent standard errors for the OLS estimates as described in Section 10.5. A second option is to use a generalized least squares transformation. A third option, suggested by Geweke, Meese, and Dent (1983), is as follows. Suppose the error term η_t in [11.2.5] has Wold representation $\eta_t = \psi_{22}(L)v_{2t}$. Multiplying both sides of [11.2.5] by $h(L) \equiv [\psi_{22}(L)]^{-1}$ produces

$$y_t = c_2 - \sum_{j=1}^{\infty} h_j y_{t-j} + \sum_{j=0}^{\infty} b_j^* x_{t-j} + \sum_{j=1}^{\infty} d_j^* x_{t+j} + v_{2t}. \qquad [11.2.11]$$

The error term in [11.2.11] is white noise and uncorrelated with any of the explanatory variables. Moreover, $d_j^* = 0$ for all j if and only if $d_j = 0$ for all j. Thus, by truncating the infinite sums in [11.2.11] at some finite value, we can test the null hypothesis that y does not Granger-cause x with an F test of $d_1^* = d_2^* = \cdots = d_p^* = 0$.

A variety of other Granger-causality tests have been proposed; see Pierce and Haugh (1977) and Geweke, Meese, and Dent (1983) for selective surveys. Bouissou, Laffont, and Vuong (1986) discussed tests using discrete-valued panel data. The Monte Carlo simulations of Geweke, Meese, and Dent suggest that the simplest and most straightforward test—namely, that based on [11.2.10]—may well be the best.

The results of any empirical test for Granger causality can be surprisingly sensitive to the choice of lag length (p) or the methods used to deal with potential nonstationarity of the series. For demonstrations of the practical relevance of such issues, see Feige and Pearce (1979), Christiano and Ljungqvist (1988), and Stock and Watson (1989).

Interpreting Granger-Causality Tests

How is "Granger causality" related to the standard meaning of "causality"? We explore this question with several examples.

Example 11.1—Granger-Causality Tests and Forward-Looking Behavior

The first example uses a modification of the model of stock prices described in Chapter 2. If an investor buys one share of a stock for the price P_t at date t, then at $t + 1$ the investor will receive D_{t+1} in dividends and be able to sell the stock for P_{t+1}. The ex post rate of return from the stock (denoted r_{t+1}) is defined by

$$(1 + r_{t+1})P_t \equiv P_{t+1} + D_{t+1}. \qquad [11.2.12]$$

A simple model of stock prices holds that the expected rate of return for the stock is a constant r at all dates:[5]

$$(1 + r)P_t = E_t[P_{t+1} + D_{t+1}]. \qquad [11.2.13]$$

Here E_t denotes an expectation conditional on all information available to stock market participants at time t. The logic behind [11.2.13] is that if investors had information at time t leading them to anticipate a higher-than-normal return to stocks, they would want to buy more stocks at date t. Such purchases would drive P_t up until [11.2.13] was satisfied. This view is sometimes called the *efficient markets hypothesis.*

As noted in the discussion of equation [2.5.15] in Chapter 2, equation [11.2.13] along with a boundedness condition implies

$$P_t = E_t \sum_{j=1}^{\infty} \left[\frac{1}{1 + r}\right]^j D_{t+j}. \qquad [11.2.14]$$

Thus, according to the theory, the stock price incorporates the market's best forecast of the present value of future dividends. If this forecast is based on more information than past dividends alone, then stock prices will Granger-cause dividends as investors try to anticipate movements in dividends.

For a simple illustration of this point, suppose that

$$D_t = d + u_t + \delta u_{t-1} + v_t, \qquad [11.2.15]$$

where u_t and v_t are independent Gaussian white noise series and d is the mean dividend. Suppose that investors at time t know the values of $\{u_t, u_{t-1}, \ldots\}$ and $\{v_t, v_{t-1}, \ldots\}$. The forecast of D_{t+j} based on this information is given by

$$E_t(D_{t+j}) = \begin{cases} d + \delta u_t & \text{for } j = 1 \\ d & \text{for } j = 2, 3, \ldots. \end{cases} \qquad [11.2.16]$$

Substituting [11.2.16] into [11.2.14], the stock price would be given by

$$P_t = d/r + \delta u_t/(1 + r). \qquad [11.2.17]$$

[5]A related model was proposed by Lucas (1978):

$$U'(C_t)P_t = E_t\{\beta U'(C_{t+1})(P_{t+1} + D_{t+1})\},$$

with $U'(C_t)$ the marginal utility of consumption at date t. If we define $\tilde{P}_t$ to be the marginal-utility-weighted stock price $\tilde{P}_t \equiv U'(C_t)P_t$ and $\tilde{D}_t$ the marginal-utility-weighted dividend, then this becomes

$$\beta^{-1}\tilde{P}_t = E_t\{\tilde{P}_{t+1} + \tilde{D}_{t+1}\},$$

which is the same basic form as [11.2.13]. With risk-neutral investors, $U'(C_t)$ is a constant and the two formulations are identical. The risk-neutral version gained early support from the empirical evidence in Fama (1965).

Thus, for this example, the stock price is white noise and could not be forecast on the basis of lagged stock prices or dividends.[6] No series should Granger-cause stock prices.

On the other hand, notice from [11.2.17] that the value of u_{t-1} can be uncovered from the lagged stock price:

$$\delta u_{t-1} = (1 + r)P_{t-1} - (1 + r)d/r.$$

Recall from Section 4.7 that u_{t-1} contains additional information about D_t beyond that contained in $\{D_{t-1}, D_{t-2}, \ldots\}$. Thus, stock prices Granger-cause dividends, though dividends fail to Granger-cause stock prices. The bivariate *VAR* takes the form

$$\begin{bmatrix} P_t \\ D_t \end{bmatrix} = \begin{bmatrix} d/r \\ -d/r \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 + r & 0 \end{bmatrix} \begin{bmatrix} P_{t-1} \\ D_{t-1} \end{bmatrix} + \begin{bmatrix} \delta u_t/(1 + r) \\ u_t + v_t \end{bmatrix}.$$

Hence, in this model, Granger causation runs in the opposite direction from the true causation. Dividends fail to "Granger-cause" prices, even though investors' perceptions of dividends are the sole determinant of stock prices. On the other hand, prices do "Granger-cause" dividends, even though the market's evaluation of the stock in reality has no effect on the dividend process.

In general, time series that reflect forward-looking behavior, such as stock prices and interest rates, are often found to be excellent predictors of many key economic time series. This clearly does not mean that these series *cause* GNP or inflation to move up or down. Instead, the values of these series reflect the market's best information as to where GNP or inflation might be headed. Granger-causality tests for such series may be useful for assessing the efficient markets view or investigating whether markets are concerned with or are able to forecast GNP or inflation, but should not be used to infer a direction of causation.

There nevertheless are circumstances in which Granger causality may offer useful evidence about the direction of true causation. As an illustration of this theme, consider trying to measure the effects of oil price increases on the economy.

Example 11.2—Testing for Strict Econometric Exogeneity[7]

All but one of the economic recessions in the United States since World War II have been preceded by a sharp increase in the price of crude petroleum. Does this mean that oil shocks are a cause of recessions?

One possibility is that the correlation is a fluke—it happened just by chance that oil shocks and recessions appeared at similar times, even though the actual processes that generated the two series are unrelated. We can investigate this possibility by testing the null hypothesis that oil prices do not Granger-cause GNP. This hypothesis is rejected by the data—oil prices help predict the value of GNP, and their contribution to prediction is statistically significant. This argues against viewing the correlation as simply a coincidence.

To place a causal interpretation on this correlation, one must establish that oil price increases were not reflecting some other macroeconomic influence that was the true cause of the recessions. The major oil price increases have

[6] This result is due to the particular specification of the time series properties assumed for dividends. A completely general result is that the excess return series defined by $P_{t+1} + D_{t+1} - (1 + r)P_t$ (which for this example would equal $\delta u_{t+1}/(1 + r) + u_{t+1} + v_{t+1}$) should be unforecastable. The example in the text provides a simpler illustration of the general issues.

[7] This discussion is based on Hamilton (1983, 1985).

been associated with clear historical events such as the Suez crisis of 1956–57, the Arab-Israeli war of 1973–74, the Iranian revolution of 1978–79, the start of the Iran-Iraq war in 1980, and Iraq's invasion of Kuwait in 1990. One could take the view that these events were caused by forces entirely outside the U.S. economy and were essentially unpredictable. If this view is correct, then the historical correlation between oil prices and GNP could be given a causal interpretation. The view has the refutable implication that no series should Granger-cause oil prices. Empirically, one indeed finds very few macroeconomic series that help predict the timing of these oil shocks.

The theme of these two examples is that Granger-causality tests can be a useful tool for testing hypotheses that can be framed as statements about the predictability of a particular series. On the other hand, one may be skeptical about their utility as a general diagnostic for establishing the direction of causation between two arbitrary series. For this reason, it seems best to describe these as tests of whether y helps forecast x rather than tests of whether y causes x. The tests may have implications for the latter question, but only in conjunction with other assumptions.

Up to this point we have been discussing two variables, x and y, in isolation from any others. Suppose there are other variables that interact with x or y as well. How does this affect the forecasting relationship between x and y?

Example 11.3—Role of Omitted Information

Consider the following three-variable system:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 1 + \delta L & L & 0 \\ 0 & 1 & 0 \\ 0 & L & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix},$$

with

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \begin{cases} \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} & \text{for } t = s \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Thus, y_3 can offer no improvement in a forecast of either y_1 or y_2 beyond that achieved using lagged y_1 and y_2.

Let us now examine the bivariate Granger-causality relation between y_1 and y_3. First, consider the process for y_1:

$$y_{1t} = \varepsilon_{1t} + \delta\varepsilon_{1,t-1} + \varepsilon_{2,t-1}.$$

Notice that y_1 is the sum of an $MA(1)$ process $(\varepsilon_{1t} + \delta\varepsilon_{1,t-1})$ and an uncorrelated white noise process $(\varepsilon_{2,t-1})$. We know from equation [4.7.15] that the univariate representation for y_1 is an $MA(1)$ process:

$$y_{1t} = u_t + \theta u_{t-1}.$$

From [4.7.16], the univariate forecast error u_t can be expressed as

$$\begin{aligned} u_t = {} & (\varepsilon_{1t} - \theta\varepsilon_{1,t-1} + \theta^2\varepsilon_{1,t-2} - \theta^3\varepsilon_{1,t-3} + \cdots) \\ & + \delta(\varepsilon_{1,t-1} - \theta\varepsilon_{1,t-2} + \theta^2\varepsilon_{1,t-3} - \theta^3\varepsilon_{1,t-4} + \cdots) \\ & + (\varepsilon_{2,t-1} - \theta\varepsilon_{2,t-2} + \theta^2\varepsilon_{2,t-3} - \theta^3\varepsilon_{2,t-4} + \cdots). \end{aligned}$$

The univariate forecast error u_t is, of course, uncorrelated with its own lagged values. Notice, however, that it is correlated with $y_{3,t-1}$:

$$E(u_t)(y_{3,t-1}) = E(u_t)(\varepsilon_{3,t-1} + \varepsilon_{2,t-2}) = -\theta\sigma_2^2.$$

Thus, lagged y_3 could help improve a forecast of y_1 that had been based on lagged values of y_1 alone, meaning that y_3 Granger-causes y_1 in a bivariate system. The reason is that lagged y_3 is correlated with the omitted variable y_2, which is also helpful in forecasting y_1.[8]

11.3. *Maximum Likelihood Estimation of Restricted Vector Autoregressions*

Section 11.1 discussed maximum likelihood estimation and hypothesis testing on unrestricted vector autoregressions. In these systems each equation in the *VAR* had the same explanatory variables, namely, a constant term and lags of all the variables in the system. We showed how to calculate a Wald test of linear constraints but did not discuss estimation of the system subject to the constraints. This section examines estimation of a restricted *VAR*.

Granger Causality in a Multivariate Context

As an example of a restricted system that we might be interested in estimating, consider a vector generalization of the issues explored in the previous section. Suppose that the variables of a *VAR* are categorized into two groups, as represented by the $(n_1 \times 1)$ vector $\mathbf{y}_{1t}$ and the $(n_2 \times 1)$ vector $\mathbf{y}_{2t}$. The *VAR* may then be written

$$\mathbf{y}_{1t} = \mathbf{c}_1 + \mathbf{A}_1'\mathbf{x}_{1t} + \mathbf{A}_2'\mathbf{x}_{2t} + \boldsymbol{\varepsilon}_{1t} \tag{11.3.1}$$

$$\mathbf{y}_{2t} = \mathbf{c}_2 + \mathbf{B}_1'\mathbf{x}_{1t} + \mathbf{B}_2'\mathbf{x}_{2t} + \boldsymbol{\varepsilon}_{2t}. \tag{11.3.2}$$

Here $\mathbf{x}_{1t}$ is an $(n_1 p \times 1)$ vector containing lags of $\mathbf{y}_{1t}$, and the $(n_2 p \times 1)$ vector $\mathbf{x}_{2t}$ contains lags of $\mathbf{y}_{2t}$:

$$\mathbf{x}_{1t} \equiv \begin{bmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{1,t-2} \\ \vdots \\ \mathbf{y}_{1,t-p} \end{bmatrix} \qquad \mathbf{x}_{2t} \equiv \begin{bmatrix} \mathbf{y}_{2,t-1} \\ \mathbf{y}_{2,t-2} \\ \vdots \\ \mathbf{y}_{2,t-p} \end{bmatrix}.$$

The $(n_1 \times 1)$ and $(n_2 \times 1)$ vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ contain the constant terms of the *VAR*, while the matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{B}_1$, and $\mathbf{B}_2$ contain the autoregressive coefficients.

The group of variables represented by $\mathbf{y}_1$ is said to be *block-exogenous in the time series sense* with respect to the variables in $\mathbf{y}_2$ if the elements in $\mathbf{y}_2$ are of no help in improving a forecast of any variable contained in $\mathbf{y}_1$ that is based on lagged values of all the elements of $\mathbf{y}_1$ alone. In the system of [11.3.1] and [11.3.2], $\mathbf{y}_1$ is block-exogenous when $\mathbf{A}_2 = \mathbf{0}$. To discuss estimation of the system subject to this constraint, we first note an alternative form in which the unrestricted likelihood can be calculated and maximized.

[8]The reader may note that for this example the correlation between y_{1t} and $y_{3,t-1}$ is zero. However, there are nonzero correlations between (1) y_{1t} and $y_{1,t-1}$ and (2) $y_{1,t-1}$ and $y_{3,t-1}$, and these account for the contribution of $y_{3,t-1}$ to a forecast of y_{1t} that already includes $y_{1,t-1}$.

An Alternative Expression for the Likelihood Function

Section 11.1 calculated the log likelihood function for a *VAR* using the prediction-error decomposition

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\theta}), \qquad [11.3.3]$$

where $\mathbf{y}_t' = (\mathbf{y}_{1t}', \mathbf{y}_{2t}')$, $\mathbf{x}_t' = (\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \ldots, \mathbf{y}_{t-p}')$, and

$$\begin{aligned}
&\log f_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\theta}) \\
&= -\frac{n_1 + n_2}{2}\log(2\pi) - \tfrac{1}{2}\log\begin{vmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{vmatrix} \\
&\quad - \tfrac{1}{2}[(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t})' \quad (\mathbf{y}_{2t} - \mathbf{c}_2 - \mathbf{B}_1'\mathbf{x}_{1t} - \mathbf{B}_2'\mathbf{x}_{2t})'] \\
&\quad \times \begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t} \\ \mathbf{y}_{2t} - \mathbf{c}_2 - \mathbf{B}_1'\mathbf{x}_{1t} - \mathbf{B}_2'\mathbf{x}_{2t}\end{bmatrix}.
\end{aligned} \qquad [11.3.4]$$

Alternatively, the joint density in [11.3.4] could be written as the product of a marginal density of $\mathbf{y}_{1t}$ with the conditional density of $\mathbf{y}_{2t}$ given $\mathbf{y}_{1t}$:

$$f_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\theta}) = f_{\mathbf{Y}_{1t}|\mathbf{X}_t}(\mathbf{y}_{1t}|\mathbf{x}_t; \boldsymbol{\theta}) \cdot f_{\mathbf{Y}_{2t}|\mathbf{Y}_{1t},\mathbf{X}_t}(\mathbf{y}_{2t}|\mathbf{y}_{1t}, \mathbf{x}_t; \boldsymbol{\theta}). \qquad [11.3.5]$$

Conditional on $\mathbf{x}_t$, the density of $\mathbf{y}_{1t}$ is

$$\begin{aligned}
f_{\mathbf{Y}_{1t}|\mathbf{X}_t}(\mathbf{y}_{1t}|\mathbf{x}_t; \boldsymbol{\theta}) &= (2\pi)^{-n_1/2}|\boldsymbol{\Omega}_{11}|^{-1/2} \\
&\quad \times \exp[-\tfrac{1}{2}(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t})'\boldsymbol{\Omega}_{11}^{-1} \\
&\quad \times (\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t})],
\end{aligned} \qquad [11.3.6]$$

while the conditional density of $\mathbf{y}_{2t}$ given $\mathbf{y}_{1t}$ and $\mathbf{x}_t$ is also Gaussian:

$$\begin{aligned}
f_{\mathbf{Y}_{2t}|\mathbf{Y}_{1t},\mathbf{X}_t}(\mathbf{y}_{2t}|\mathbf{y}_{1t}, \mathbf{x}_t; \boldsymbol{\theta}) &= (2\pi)^{-n_2/2}|\mathbf{H}|^{-1/2} \\
&\quad \times \exp[-\tfrac{1}{2}(\mathbf{y}_{2t} - \mathbf{m}_{2t})'\mathbf{H}^{-1}(\mathbf{y}_{2t} - \mathbf{m}_{2t})].
\end{aligned} \qquad [11.3.7]$$

The parameters of this conditional distribution can be calculated using the results from Section 4.6. The conditional variance is given by equation [4.6.6]:

$$\mathbf{H} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12};$$

while the conditional mean ($\mathbf{m}_{2t}$) can be calculated from [4.6.5]:

$$\mathbf{m}_{2t} = E(\mathbf{y}_{2t}|\mathbf{x}_t) + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}[\mathbf{y}_{1t} - E(\mathbf{y}_{1t}|\mathbf{x}_t)]. \qquad [11.3.8]$$

Notice from [11.3.1] that

$$E(\mathbf{y}_{1t}|\mathbf{x}_t) = \mathbf{c}_1 + \mathbf{A}_1'\mathbf{x}_{1t} + \mathbf{A}_2'\mathbf{x}_{2t},$$

while from [11.3.2],

$$E(\mathbf{y}_{2t}|\mathbf{x}_t) = \mathbf{c}_2 + \mathbf{B}_1'\mathbf{x}_{1t} + \mathbf{B}_2'\mathbf{x}_{2t}.$$

Substituting these expressions into [11.3.8],

$$\begin{aligned}
\mathbf{m}_{2t} &= (\mathbf{c}_2 + \mathbf{B}_1'\mathbf{x}_{1t} + \mathbf{B}_2'\mathbf{x}_{2t}) + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}[\mathbf{y}_{1t} - (\mathbf{c}_1 + \mathbf{A}_1'\mathbf{x}_{1t} + \mathbf{A}_2'\mathbf{x}_{2t})] \\
&= \mathbf{d} + \mathbf{D}_0'\mathbf{y}_{1t} + \mathbf{D}_1'\mathbf{x}_{1t} + \mathbf{D}_2'\mathbf{x}_{2t},
\end{aligned}$$

where

$$\mathbf{d} = \mathbf{c}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\mathbf{c}_1 \qquad [11.3.9]$$

$$\mathbf{D}_0' = \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1} \qquad [11.3.10]$$

$$\mathbf{D}_1' = \mathbf{B}_1' - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\mathbf{A}_1' \qquad [11.3.11]$$

$$\mathbf{D}_2' = \mathbf{B}_2' - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\mathbf{A}_2'. \qquad [11.3.12]$$

The log of the joint density in [11.3.4] can thus equivalently be calculated as the sum of the logs of the marginal density [11.3.6] and the conditional density [11.3.7]:

$$\log f_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}_t|\mathbf{x}_t;\boldsymbol{\theta}) = \ell_{1t} + \ell_{2t}, \qquad [11.3.13]$$

where

$$\begin{aligned}\ell_{1t} = {} & (-n_1/2)\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Omega}_{11}| \\ & - \tfrac{1}{2}[(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t})'\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t} - \mathbf{A}_2'\mathbf{x}_{2t})]\end{aligned} \qquad [11.3.14]$$

$$\begin{aligned}\ell_{2t} = {} & (-n_2/2)\log(2\pi) - \tfrac{1}{2}\log|\mathbf{H}| \\ & - \tfrac{1}{2}[(\mathbf{y}_{2t} - \mathbf{d} - \mathbf{D}_0'\mathbf{y}_{1t} - \mathbf{D}_1'\mathbf{x}_{1t} - \mathbf{D}_2'\mathbf{x}_{2t})'\mathbf{H}^{-1} \\ & \times (\mathbf{y}_{2t} - \mathbf{d} - \mathbf{D}_0'\mathbf{y}_{1t} - \mathbf{D}_1'\mathbf{x}_{1t} - \mathbf{D}_2'\mathbf{x}_{2t})].\end{aligned} \qquad [11.3.15]$$

The sample log likelihood would then be expressed as

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\ell_{1t} + \sum_{t=1}^{T}\ell_{2t}. \qquad [11.3.16]$$

Equations [11.3.4] and [11.3.13] are two different expressions for the same magnitude. As long as the parameters in the second representation are related to those of the first as in [11.3.9] through [11.3.12], either calculation would produce the identical value for the likelihood. If [11.3.3] is maximized by choice of ($\mathbf{c}_1$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{c}_2$, $\mathbf{B}_1$, $\mathbf{B}_2$, $\boldsymbol{\Omega}_{11}$, $\boldsymbol{\Omega}_{12}$, $\boldsymbol{\Omega}_{22}$), the same value for the likelihood will be achieved as by maximizing [11.3.16] by choice of ($\mathbf{c}_1$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\boldsymbol{\Omega}_{11}$, $\mathbf{H}$).

The second maximization is as easy to achieve as the first. Since the parameters ($\mathbf{c}_1$, $\mathbf{A}_1$, $\mathbf{A}_2$) appear in [11.3.16] only through $\Sigma_{t=1}^T \ell_{1t}$, the *MLE*s of these parameters can be found by *OLS* regressions of the elements of $\mathbf{y}_{1t}$ on a constant and lagged values of $\mathbf{y}_1$ and $\mathbf{y}_2$, that is, by *OLS* estimation of

$$\mathbf{y}_{1t} = \mathbf{c}_1 + \mathbf{A}_1'\mathbf{x}_{1t} + \mathbf{A}_2'\mathbf{x}_{2t} + \boldsymbol{\varepsilon}_{1t}. \qquad [11.3.17]$$

The *MLE* of $\boldsymbol{\Omega}_{11}$ is the sample variance-covariance matrix of the residuals from these regressions, $\hat{\boldsymbol{\Omega}}_{11} = (1/T)\,\Sigma_{t=1}^T \hat{\boldsymbol{\varepsilon}}_{1t}\hat{\boldsymbol{\varepsilon}}_{1t}'$. Similarly, the parameters ($\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$) appear in [11.3.16] only through $\Sigma_{t=1}^T \ell_{2t}$, and so their *MLE*s are obtained from *OLS* regressions of the elements of $\mathbf{y}_{2t}$ on a constant, current and lagged values of $\mathbf{y}_1$, and lagged values of $\mathbf{y}_2$:

$$\mathbf{y}_{2t} = \mathbf{d} + \mathbf{D}_0'\mathbf{y}_{1t} + \mathbf{D}_1'\mathbf{x}_{1t} + \mathbf{D}_2'\mathbf{x}_{2t} + \mathbf{v}_{2t}. \qquad [11.3.18]$$

The *MLE* of $\mathbf{H}$ is the sample variance-covariance matrix of the residuals from this second set of regressions, $\hat{\mathbf{H}} = (1/T)\,\Sigma_{t=1}^T \hat{\mathbf{v}}_{2t}\hat{\mathbf{v}}_{2t}'$.

Note that the population residuals associated with the second set of regressions, $\mathbf{v}_{2t}$, are uncorrelated with the population residuals of the first regressions. This is because $\mathbf{v}_{2t} = \mathbf{y}_{2t} - E(\mathbf{y}_{2t}|\mathbf{y}_{1t}, \mathbf{x}_t)$ is uncorrelated by construction with $\mathbf{y}_{1t}$ and $\mathbf{x}_t$, whereas $\boldsymbol{\varepsilon}_{1t}$ is a linear function of $\mathbf{y}_{1t}$ and $\mathbf{x}_t$. Similarly, the *OLS* sample residuals associated with the second regressions,

$$\hat{\mathbf{v}}_{2t} \equiv \mathbf{y}_{2t} - \hat{\mathbf{d}} - \hat{\mathbf{D}}_0'\mathbf{y}_{1t} - \hat{\mathbf{D}}_1'\mathbf{x}_{1t} - \hat{\mathbf{D}}_2'\mathbf{x}_{2t},$$

are orthogonal by construction to $\mathbf{y}_{1t}$, a constant term, and $\mathbf{x}_t$. Since the *OLS* sample residuals associated with the first regressions, $\hat{\boldsymbol{\varepsilon}}_{1t}$, are linear functions of these same elements, $\hat{\mathbf{v}}_{2t}$ is orthogonal by construction to $\hat{\boldsymbol{\varepsilon}}_{1t}$.

Maximum Likelihood Estimation of a VAR *Characterized by Block Exogeneity*

Now consider maximum likelihood estimation of the system subject to the constraint that $\mathbf{A}_2 = \mathbf{0}$. Suppose we view ($\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{H}$) rather than ($\mathbf{c}_2$, $\mathbf{B}_1$,

$\mathbf{B}_2$, $\boldsymbol{\Omega}_{21}$, $\boldsymbol{\Omega}_{22}$) as the parameters of interest for the second equation and take our objective to be to choose values for ($\mathbf{c}_1$, $\mathbf{A}_1$, $\boldsymbol{\Omega}_{11}$, $\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{H}$) so as to maximize the likelihood function. For this parameterization, the value of $\mathbf{A}_2$ does not affect the value of ℓ_{2t} in [11.3.15]. Thus, the full-information maximum likelihood estimates of $\mathbf{c}_1$, $\mathbf{A}_1$, and $\boldsymbol{\Omega}_{11}$ can be based solely on a restricted version of the regressions in [11.3.17],

$$\mathbf{y}_{1t} = \mathbf{c}_1 + \mathbf{A}_1'\mathbf{x}_{1t} + \boldsymbol{\varepsilon}_{1t}. \tag{11.3.19}$$

Let $\hat{\mathbf{c}}_1(0)$, $\hat{\mathbf{A}}_1(0)$, $\hat{\boldsymbol{\Omega}}_{11}(0)$ denote the estimates from these restricted regressions. The maximum likelihood estimates of the other parameters of the system ($\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{H}$) continue to be given by unrestricted *OLS* estimation of [11.3.18], with estimates denoted ($\hat{\mathbf{d}}$, $\hat{\mathbf{D}}_0$, $\hat{\mathbf{D}}_1$, $\hat{\mathbf{D}}_2$, $\hat{\mathbf{H}}$).

The maximum value achieved for the log likelihood function can be found by applying [11.1.32] to [11.3.13]:

$$\begin{aligned}\mathscr{L}[\hat{\boldsymbol{\theta}}(0)] &= \sum_{t=1}^{T} \ell_{1t}[\hat{\mathbf{c}}_1(0), \hat{\mathbf{A}}_1(0), \hat{\boldsymbol{\Omega}}_{11}(0)] + \sum_{t=1}^{T} \ell_{2t}[\hat{\mathbf{d}}, \hat{\mathbf{D}}_0, \hat{\mathbf{D}}_1, \hat{\mathbf{D}}_2, \hat{\mathbf{H}}] \\ &= [-(Tn_1/2)\log(2\pi) + (T/2)\log|\hat{\boldsymbol{\Omega}}_{11}^{-1}(0)| - (Tn_1/2)] \\ &\quad + [-(Tn_2/2)\log(2\pi) + (T/2)\log|\hat{\mathbf{H}}^{-1}| - (Tn_2/2)].\end{aligned} \tag{11.3.20}$$

By contrast, when the system is estimated with no constraints on $\mathbf{A}_2$, the value achieved for the log likelihood is

$$\begin{aligned}\mathscr{L}(\hat{\boldsymbol{\theta}}) &= \sum_{t=1}^{T} \ell_{1t}[\hat{\mathbf{c}}_1, \hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \hat{\boldsymbol{\Omega}}_{11}] + \sum_{t=1}^{T} \ell_{2t}[\hat{\mathbf{d}}, \hat{\mathbf{D}}_0, \hat{\mathbf{D}}_1, \hat{\mathbf{D}}_2, \hat{\mathbf{H}}] \\ &= [-(Tn_1/2)\log(2\pi) + (T/2)\log|\hat{\boldsymbol{\Omega}}_{11}^{-1}| - (Tn_1/2)] \\ &\quad + [-(Tn_2/2)\log(2\pi) + (T/2)\log|\hat{\mathbf{H}}^{-1}| - (Tn_2/2)],\end{aligned} \tag{11.3.21}$$

where ($\hat{\mathbf{c}}_1$, $\hat{\mathbf{A}}_1$, $\hat{\mathbf{A}}_2$, $\hat{\boldsymbol{\Omega}}_{11}$) denote estimates based on *OLS* estimation of [11.3.17]. A likelihood ratio test of the null hypothesis that $\mathbf{A}_2 = \mathbf{0}$ can thus be based on

$$\begin{aligned}2\{\mathscr{L}[\hat{\boldsymbol{\theta}}] - \mathscr{L}[\hat{\boldsymbol{\theta}}(0)]\} &= T\{\log|\hat{\boldsymbol{\Omega}}_{11}^{-1}| - \log|\hat{\boldsymbol{\Omega}}_{11}^{-1}(0)|\} \\ &= T\{\log|\hat{\boldsymbol{\Omega}}_{11}(0)| - \log|\hat{\boldsymbol{\Omega}}_{11}|\}.\end{aligned} \tag{11.3.22}$$

This will have an asymptotic χ^2 distribution with degrees of freedom equal to the number of restrictions. Since $\mathbf{A}_2$ is an $(n_1 \times n_2 p)$ matrix, the number of restrictions is $n_1 n_2 p$.

Thus, to test the null hypothesis that the n_1 variables represented by $\mathbf{y}_1$ are block-exogenous with respect to the n_2 variables represented by $\mathbf{y}_2$, perform *OLS* regressions of each of the elements of $\mathbf{y}_1$ on a constant, p lags of all of the elements of $\mathbf{y}_1$, and p lags of all of the elements of $\mathbf{y}_2$. Let $\hat{\boldsymbol{\varepsilon}}_{1t}$ denote the $(n_1 \times 1)$ vector of sample residuals for date t from these regressions and $\hat{\boldsymbol{\Omega}}_{11}$ their variance-covariance matrix ($\hat{\boldsymbol{\Omega}}_{11} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_{1t}\hat{\boldsymbol{\varepsilon}}_{1t}'$). Next perform *OLS* regressions of each of the elements of $\mathbf{y}_1$ on a constant and p lags of all the elements of $\mathbf{y}_1$. Let $\hat{\boldsymbol{\varepsilon}}_{1t}(0)$ denote the $(n_1 \times 1)$ vector of sample residuals from this second set of regressions and $\hat{\boldsymbol{\Omega}}_{11}(0)$ their variance-covariance matrix ($\hat{\boldsymbol{\Omega}}_{11}(0) = (1/T)\sum_{t=1}^{T} [\hat{\boldsymbol{\varepsilon}}_{1t}(0)][\hat{\boldsymbol{\varepsilon}}_{1t}(0)]'$). If

$$T\{\log|\hat{\boldsymbol{\Omega}}_{11}(0)| - \log|\hat{\boldsymbol{\Omega}}_{11}|\}$$

is greater than the 5% critical value for a $\chi^2(n_1 n_2 p)$ variable, then the null hypothesis is rejected, and the conclusion is that some of the elements of $\mathbf{y}_2$ are helpful in forecasting $\mathbf{y}_1$.

Thus, if our interest is in estimation of the parameters ($\mathbf{c}_1$, $\mathbf{A}_1$, $\boldsymbol{\Omega}_{11}$, $\mathbf{d}$, $\mathbf{D}_0$, $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{H}$) or testing a hypothesis about block exogeneity, all that is necessary is OLS regression on the affected equations. Suppose, however, that we wanted full-information maximum likelihood estimates of the parameters of the likelihood as originally parameterized ($\mathbf{c}_1$, $\mathbf{A}_1$, $\mathbf{\Omega}_{11}$, $\mathbf{c}_2$, $\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{\Omega}_{21}$, $\mathbf{\Omega}_{22}$). For the parameters of the first block of equations ($\mathbf{c}_1$, $\mathbf{A}_1$, $\mathbf{\Omega}_{11}$), the MLEs continue to be given by OLS estimation of [11.3.19]. The parameters of the second block can be found from the OLS estimates by inverting equations [11.3.9] through [11.3.12]:[9]

$$\begin{aligned}
\hat{\mathbf{\Omega}}_{21}(0) &= \hat{\mathbf{D}}_0'[\hat{\mathbf{\Omega}}_{11}(0)] \\
\hat{\mathbf{c}}_2(0) &= \hat{\mathbf{d}} + [\hat{\mathbf{\Omega}}_{21}(0)][\hat{\mathbf{\Omega}}_{11}(0)]^{-1}[\hat{\mathbf{c}}_1(0)] \\
[\hat{\mathbf{B}}_1(0)]' &= \hat{\mathbf{D}}_1' + [\hat{\mathbf{\Omega}}_{21}(0)][\hat{\mathbf{\Omega}}_{11}(0)]^{-1}[\hat{\mathbf{A}}_1(0)]' \\
[\hat{\mathbf{B}}_2(0)]' &= \hat{\mathbf{D}}_2' \\
\hat{\mathbf{\Omega}}_{22}(0) &= \hat{\mathbf{H}} + [\hat{\mathbf{\Omega}}_{21}(0)][\hat{\mathbf{\Omega}}_{11}(0)]^{-1}[\hat{\mathbf{\Omega}}_{12}(0)].
\end{aligned}$$

Thus, the maximum likelihood estimates for the original parameterization of [11.3.2] are found from these equations by combining the OLS estimates from [11.3.19] and [11.3.18].

Geweke's Measure of Linear Dependence

The previous subsection modeled the relation between an $(n_1 \times 1)$ vector $\mathbf{y}_{1t}$ and an $(n_2 \times 1)$ vector $\mathbf{y}_{2t}$ in terms of the pth-order VAR [11.3.1] and [11.3.2], where the innovations have a variance-covariance matrix given by

$$E\begin{bmatrix} \boldsymbol{\varepsilon}_{1t}\boldsymbol{\varepsilon}_{1t}' & \boldsymbol{\varepsilon}_{1t}\boldsymbol{\varepsilon}_{2t}' \\ \boldsymbol{\varepsilon}_{2t}\boldsymbol{\varepsilon}_{1t}' & \boldsymbol{\varepsilon}_{2t}\boldsymbol{\varepsilon}_{2t}' \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix}.$$

To test the null hypothesis that $\mathbf{y}_1$ is block exogenous with respect to $\mathbf{y}_2$, we proposed calculating the statistic in [11.3.22],

$$T\{\log|\hat{\mathbf{\Omega}}_{11}(0)| - \log|\hat{\mathbf{\Omega}}_{11}|\} \approx \chi^2(n_1 n_2 p), \qquad [11.3.23]$$

where $\hat{\mathbf{\Omega}}_{11}$ is the variance-covariance matrix of the residuals from OLS estimation of [11.3.1] and $\hat{\mathbf{\Omega}}_{11}(0)$ is the variance-covariance matrix of the residuals from OLS estimation of [11.3.1] when lagged values of $\mathbf{y}_2$ are omitted from the regression (that is, when $\mathbf{A}_2 = \mathbf{0}$ in [11.3.1]).

Clearly, to test the parallel null hypothesis that $\mathbf{y}_2$ is block-exogenous with respect to $\mathbf{y}_1$, we would calculate

$$T\{\log|\hat{\mathbf{\Omega}}_{22}(0)| - \log|\hat{\mathbf{\Omega}}_{22}|\} \approx \chi^2(n_2 n_1 p), \qquad [11.3.24]$$

where $\hat{\mathbf{\Omega}}_{22}$ is the variance-covariance matrix of the residuals from OLS estimation of [11.3.2] and $\hat{\mathbf{\Omega}}_{22}(0)$ is the variance-covariance matrix of the residuals from OLS estimation of [11.3.2] when lagged values of $\mathbf{y}_1$ are omitted from the regression (that is, when $\mathbf{B}_1 = \mathbf{0}$ in [11.3.2]).

Finally, consider maximum likelihood estimation of the VAR subject to the restriction that there is no relation whatsoever between $\mathbf{y}_1$ and $\mathbf{y}_2$, that is, subject

[9]To confirm that the resulting estimate $\hat{\mathbf{\Omega}}(0)$ is symmeteric and positive definite, notice that

$$\hat{\mathbf{\Omega}}_{22}(0) = \hat{\mathbf{H}} + \hat{\mathbf{D}}_0'[\hat{\mathbf{\Omega}}_{11}(0)]\hat{\mathbf{D}}_0$$

and so

$$\begin{bmatrix} \hat{\mathbf{\Omega}}_{11}(0) & \hat{\mathbf{\Omega}}_{12}(0) \\ \hat{\mathbf{\Omega}}_{21}(0) & \hat{\mathbf{\Omega}}_{22}(0) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \hat{\mathbf{D}}_0' & \mathbf{I}_{n_2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{\Omega}}_{11}(0) & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{H}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_1} & \hat{\mathbf{D}}_0 \\ \mathbf{0} & \mathbf{I}_{n_2} \end{bmatrix}.$$

to the restrictions that $\mathbf{A}_2 = \mathbf{0}$, $\mathbf{B}_1 = \mathbf{0}$, and $\mathbf{\Omega}_{21} = \mathbf{0}$. For this most restricted specification, the log likelihood becomes

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \bigg\{ &-(n_1/2)\log(2\pi) - (1/2)\log|\mathbf{\Omega}_{11}| \\ &- (1/2)(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t})'\mathbf{\Omega}_{11}^{-1}(\mathbf{y}_{1t} - \mathbf{c}_1 - \mathbf{A}_1'\mathbf{x}_{1t})\bigg\} \\ + \sum_{t=1}^{T} \bigg\{ &-(n_2/2)\log(2\pi) - (1/2)\log|\mathbf{\Omega}_{22}| \\ &- (1/2)(\mathbf{y}_{2t} - \mathbf{c}_2 - \mathbf{B}_2'\mathbf{x}_{2t})'\mathbf{\Omega}_{22}^{-1}(\mathbf{y}_{2t} - \mathbf{c}_2 - \mathbf{B}_2'\mathbf{x}_{2t})\bigg\}\end{aligned}$$

and the maximized value is

$$\begin{aligned}\mathcal{L}(\hat{\boldsymbol{\theta}}(0)) = &\{-(Tn_1/2)\log(2\pi) - (T/2)\log|\hat{\mathbf{\Omega}}_{11}(0)| - (Tn_1/2)\} \\ &+ \{-(Tn_2/2)\log(2\pi) - (T/2)\log|\hat{\mathbf{\Omega}}_{22}(0)| - (Tn_2/2)\}.\end{aligned}$$

A likelihood ratio test of the null hypothesis of no relation at all between $\mathbf{y}_1$ and $\mathbf{y}_2$ is thus given by

$$\begin{aligned}&2\{\mathcal{L}(\hat{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}}(0))\} \\ &\quad = T\left\{\log|\hat{\mathbf{\Omega}}_{11}(0)| + \log|\hat{\mathbf{\Omega}}_{22}(0)| - \log\begin{vmatrix}\hat{\mathbf{\Omega}}_{11} & \hat{\mathbf{\Omega}}_{12} \\ \hat{\mathbf{\Omega}}_{21} & \hat{\mathbf{\Omega}}_{22}\end{vmatrix}\right\},\end{aligned} \tag{11.3.25}$$

where $\hat{\mathbf{\Omega}}_{12}$ is the covariance matrix between the residuals from unrestricted *OLS* estimation of [11.3.1] and [11.3.2]. This null hypothesis imposed the (n_1n_2p) restrictions that $\mathbf{A}_2 = \mathbf{0}$, the (n_2n_1p) restrictions that $\mathbf{B}_1 = \mathbf{0}$, and the (n_2n_1) restrictions that $\mathbf{\Omega}_{21} = \mathbf{0}$. Hence, the statistic in [11.3.25] has a χ^2 distribution with $(n_1n_2) \times (2p + 1)$ degrees of freedom.

Geweke (1982) proposed $(1/T)$ times the magnitude in [11.3.25] as a measure of the degree of linear dependence between $\mathbf{y}_1$ and $\mathbf{y}_2$. Note that [11.3.25] can be expressed as the sum of three terms:

$$\begin{aligned}&T\left\{\log|\hat{\mathbf{\Omega}}_{11}(0)| + \log|\hat{\mathbf{\Omega}}_{22}(0)| - \log\begin{vmatrix}\hat{\mathbf{\Omega}}_{11} & \hat{\mathbf{\Omega}}_{12} \\ \hat{\mathbf{\Omega}}_{21} & \hat{\mathbf{\Omega}}_{22}\end{vmatrix}\right\} \\ &\quad = T\{\log|\hat{\mathbf{\Omega}}_{11}(0)| - \log|\hat{\mathbf{\Omega}}_{11}|\} + T\{\log|\hat{\mathbf{\Omega}}_{22}(0)| - \log|\hat{\mathbf{\Omega}}_{22}|\} \\ &\qquad + T\left\{\log|\hat{\mathbf{\Omega}}_{11}| + \log|\hat{\mathbf{\Omega}}_{22}| - \log\begin{vmatrix}\hat{\mathbf{\Omega}}_{11} & \hat{\mathbf{\Omega}}_{12} \\ \hat{\mathbf{\Omega}}_{21} & \hat{\mathbf{\Omega}}_{22}\end{vmatrix}\right\}.\end{aligned} \tag{11.3.26}$$

The first of these three terms, $T\{\log|\hat{\mathbf{\Omega}}_{11}(0)| - \log|\hat{\mathbf{\Omega}}_{11}|\}$, is a measure of the strength of the linear feedback from $\mathbf{y}_2$ to $\mathbf{y}_1$ and is the $\chi^2(n_1n_2p)$ statistic calculated in [11.3.23]. The second term, $T\{\log|\hat{\mathbf{\Omega}}_{22}(0)| - \log|\hat{\mathbf{\Omega}}_{22}|\}$, is an analogous measure of the strength of linear feedback from $\mathbf{y}_1$ to $\mathbf{y}_2$ and is the $\chi^2(n_2n_1p)$ statistic in [11.3.24]. The third term,

$$T\left\{\log|\hat{\mathbf{\Omega}}_{11}| + \log|\hat{\mathbf{\Omega}}_{22}| - \log\begin{vmatrix}\hat{\mathbf{\Omega}}_{11} & \hat{\mathbf{\Omega}}_{12} \\ \hat{\mathbf{\Omega}}_{21} & \hat{\mathbf{\Omega}}_{22}\end{vmatrix}\right\},$$

is a measure of instantaneous feedback. This corresponds to a likelihood ratio test of the null hypothesis that $\mathbf{\Omega}_{21} = \mathbf{0}$ with $\mathbf{A}_2$ and $\mathbf{B}_1$ unrestricted and has a $\chi^2(n_1n_2)$ distribution under the null.

Thus, [11.3.26] can be used to summarize the strength of any linear relation between $\mathbf{y}_1$ and $\mathbf{y}_2$ and identify the source of that relation. Geweke showed how these measures can be further decomposed by frequency.

Maximum Likelihood Estimation Under General Coefficient Constraints

We now discuss maximum likelihood estimation of a vector autoregression in which there are constraints that cannot be expressed in a block-recursive form as in the previous example. A *VAR* subject to general exclusion restrictions can be viewed as a system of "seemingly unrelated regressions" as originally analyzed by Zellner (1962).

Let $\mathbf{x}_{1t}$ be a $(k_1 \times 1)$ vector containing a constant term and lags of the variables that appear in the first equation of the *VAR*:

$$y_{1t} = \mathbf{x}_{1t}'\boldsymbol{\beta}_1 + \varepsilon_{1t}.$$

Similarly, let $\mathbf{x}_{2t}$ denote a $(k_2 \times 1)$ vector containing the explanatory variables for the second equation and $\mathbf{x}_{nt}$ a $(k_n \times 1)$ vector containing the variables for the last equation. Hence, the *VAR* consists of the system of equations

$$\begin{aligned} y_{1t} &= \mathbf{x}_{1t}'\boldsymbol{\beta}_1 + \varepsilon_{1t} \\ y_{2t} &= \mathbf{x}_{2t}'\boldsymbol{\beta}_2 + \varepsilon_{2t} \\ &\vdots \\ y_{nt} &= \mathbf{x}_{nt}'\boldsymbol{\beta}_n + \varepsilon_{nt}. \end{aligned} \qquad [11.3.27]$$

Let $k = k_1 + k_2 + \cdots + k_n$ denote the total number of coefficients to be estimated, and collect these in a $(k \times 1)$ vector:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix}.$$

Then the system of equations in [11.3.27] can be written in vector form as

$$\mathbf{y}_t = \mathcal{X}_t'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad [11.3.28]$$

where $\mathcal{X}_t'$ is the following $(n \times k)$ matrix:

$$\mathcal{X}_t' \equiv \begin{bmatrix} \boldsymbol{x}_{1t}' \\ \boldsymbol{x}_{2t}' \\ \vdots \\ \boldsymbol{x}_{nt}' \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}_{1t}' & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{x}_{2t}' & \cdots & \mathbf{0}' \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{x}_{nt}' \end{bmatrix}.$$

Thus, $\boldsymbol{x}_{it}'$ is defined as a $(1 \times k)$ vector containing the k_i explanatory variables for equation i, with zeros added so as to be conformable with the $(k \times 1)$ vector $\boldsymbol{\beta}$.

The goal is to choose $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ so as to maximize the log likelihood function

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Omega}) = &-(Tn/2)\log(2\pi) + (T/2)\log|\boldsymbol{\Omega}^{-1}| \\ &- (1/2)\sum_{t=1}^{T}(\mathbf{y}_t - \mathcal{X}_t'\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathcal{X}_t'\boldsymbol{\beta}). \end{aligned} \qquad [11.3.29]$$

This calls for choosing $\boldsymbol{\beta}$ so as to minimize

$$\sum_{t=1}^{T}(\mathbf{y}_t - \mathcal{X}_t'\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathcal{X}_t'\boldsymbol{\beta}). \qquad [11.3.30]$$

If $\boldsymbol{\Omega}^{-1}$ is written as $\mathbf{L}'\mathbf{L}$, this becomes

$$\sum_{t=1}^{T} (\mathbf{y}_t - \mathscr{X}_t'\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathscr{X}_t'\boldsymbol{\beta}) = \sum_{t=1}^{T} (\mathbf{L}\mathbf{y}_t - \mathbf{L}\mathscr{X}_t'\boldsymbol{\beta})'(\mathbf{L}\mathbf{y}_t - \mathbf{L}\mathscr{X}_t'\boldsymbol{\beta}) \\ = \sum_{t=1}^{T} (\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta})'(\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta}), \tag{11.3.31}$$

where $\tilde{\mathbf{y}}_t \equiv \mathbf{L}\mathbf{y}_t$ and

$$\tilde{\mathscr{X}}_t' \equiv \mathbf{L}\mathscr{X}_t' = \begin{bmatrix} \tilde{\boldsymbol{x}}_{1t}' \\ \tilde{\boldsymbol{x}}_{2t}' \\ \vdots \\ \tilde{\boldsymbol{x}}_{nt}' \end{bmatrix}.$$

But [11.3.31] is simply

$$\sum_{t=1}^{T} (\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta})'(\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta})$$
$$= \sum_{t=1}^{T} \begin{bmatrix} \tilde{y}_{1t} - \tilde{\boldsymbol{x}}_{1t}'\boldsymbol{\beta} \\ \tilde{y}_{2t} - \tilde{\boldsymbol{x}}_{2t}'\boldsymbol{\beta} \\ \vdots \\ \tilde{y}_{nt} - \tilde{\boldsymbol{x}}_{nt}'\boldsymbol{\beta} \end{bmatrix}' \begin{bmatrix} \tilde{y}_{1t} - \tilde{\boldsymbol{x}}_{1t}'\boldsymbol{\beta} \\ \tilde{y}_{2t} - \tilde{\boldsymbol{x}}_{2t}'\boldsymbol{\beta} \\ \vdots \\ \tilde{y}_{nt} - \tilde{\boldsymbol{x}}_{nt}'\boldsymbol{\beta} \end{bmatrix}$$
$$= \sum_{t=1}^{T} [(\tilde{y}_{1t} - \tilde{\boldsymbol{x}}_{1t}'\boldsymbol{\beta})^2 + (\tilde{y}_{2t} - \tilde{\boldsymbol{x}}_{2t}'\boldsymbol{\beta})^2 + \cdots + (\tilde{y}_{nt} - \tilde{\boldsymbol{x}}_{nt}'\boldsymbol{\beta})^2],$$

which is minimized by an *OLS* regression of $\tilde{y}_{it}$ on $\tilde{\boldsymbol{x}}_{it}$, pooling all the equations ($i = 1, 2, \ldots, n$) into one big regression. Thus, the maximum likelihood estimate is given by

$$\hat{\boldsymbol{\beta}} = \left\{\sum_{t=1}^{T} [(\tilde{\boldsymbol{x}}_{1t}\tilde{\boldsymbol{x}}_{1t}') + (\tilde{\boldsymbol{x}}_{2t}\tilde{\boldsymbol{x}}_{2t}') + \cdots + (\tilde{\boldsymbol{x}}_{nt}\tilde{\boldsymbol{x}}_{nt}')]\right\}^{-1} \\ \times \left\{\sum_{t=1}^{T} [(\tilde{\boldsymbol{x}}_{1t}\tilde{y}_{1t}) + (\tilde{\boldsymbol{x}}_{2t}\tilde{y}_{2t}) + \cdots + (\tilde{\boldsymbol{x}}_{nt}\tilde{y}_{nt})]\right\}. \tag{11.3.32}$$

Noting that the variance of the residual of this pooled regression is unity by construction,[10] the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ can be calculated from

$$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \cong \left\{\sum_{t=1}^{T} [(\tilde{\boldsymbol{x}}_{1t}\tilde{\boldsymbol{x}}_{1t}') + (\tilde{\boldsymbol{x}}_{2t}\tilde{\boldsymbol{x}}_{2t}') + \cdots + (\tilde{\boldsymbol{x}}_{nt}\tilde{\boldsymbol{x}}_{nt}')]\right\}^{-1}.$$

Construction of the variables $\tilde{y}_{it}$ and $\tilde{\boldsymbol{x}}_{it}$ to use in this pooled *OLS* regression requires knowledge of $\mathbf{L}$ and hence $\boldsymbol{\Omega}$. The parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ can be estimated jointly by maximum likelihood through the following iterative procedure. From n *OLS* regressions of y_{it} on $\mathbf{x}_{it}$, form an initial estimate of the coefficient vector

[10]That is,

$$E(\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta})(\tilde{\mathbf{y}}_t - \tilde{\mathscr{X}}_t'\boldsymbol{\beta})' = \mathbf{L}\boldsymbol{\Omega}\mathbf{L}' = \mathbf{L}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}' = \mathbf{I}_n.$$

$\hat{\boldsymbol\beta}(0) = (\mathbf{b}_1' \quad \mathbf{b}_2' \quad \cdots \quad \mathbf{b}_n')'$. Use this to form an initial estimate of the variance matrix,

$$\hat{\boldsymbol\Omega}(0) = (1/T) \sum_{t=1}^{T} [\mathbf{y}_t - \mathcal{X}_t'\hat{\boldsymbol\beta}(0)][\mathbf{y}_t - \mathcal{X}_t'\hat{\boldsymbol\beta}(0)]'.$$

Find a matrix $\hat{\mathbf{L}}(0)$ such that $[\hat{\mathbf{L}}(0)']\hat{\mathbf{L}}(0) = [\hat{\boldsymbol\Omega}(0)]^{-1}$, say, by Cholesky factorization, and form $\tilde{\mathbf{y}}_t(0) = \hat{\mathbf{L}}(0)\mathbf{y}_t$ and $\tilde{\mathcal{X}}_t'(0) = \hat{\mathbf{L}}(0)\mathcal{X}_t'$. A pooled *OLS* regression of $\tilde{y}_{it}(0)$ on $\tilde{\boldsymbol{x}}_{it}(0)$ combining $i = 1, 2, \ldots, n$ then yields the new estimate $\hat{\boldsymbol\beta}(1)$, from which $\hat{\boldsymbol\Omega}(1) = (1/T)\Sigma_{t=1}^{T}[\mathbf{y}_t - \mathcal{X}_t'\hat{\boldsymbol\beta}(1)][\mathbf{y}_t - \mathcal{X}_t'\hat{\boldsymbol\beta}(1)]'$. Iterating in this manner will produce the maximum likelihood estimates $(\hat{\boldsymbol\beta}, \hat{\boldsymbol\Omega})$, though the estimate after just one iteration has the same asymptotic distribution as the final *MLE* (see Magnus, 1978).

An alternative expression for the *MLE* in [11.3.32] is sometimes used. Notice that

$$\begin{aligned}
&[(\tilde{\boldsymbol{x}}_{1t}\tilde{\boldsymbol{x}}_{1t}') + (\tilde{\boldsymbol{x}}_{2t}\tilde{\boldsymbol{x}}_{2t}') + \cdots + (\tilde{\boldsymbol{x}}_{nt}\tilde{\boldsymbol{x}}_{nt}')] \\
&= [\tilde{\boldsymbol{x}}_{1t} \quad \tilde{\boldsymbol{x}}_{2t} \quad \cdots \quad \tilde{\boldsymbol{x}}_{nt}]\begin{bmatrix} \tilde{\boldsymbol{x}}_{1t}' \\ \tilde{\boldsymbol{x}}_{2t}' \\ \vdots \\ \tilde{\boldsymbol{x}}_{nt}' \end{bmatrix} \\
&= \tilde{\mathcal{X}}_t\tilde{\mathcal{X}}_t' \\
&= \mathcal{X}_t\mathbf{L}'\mathbf{L}\mathcal{X}_t' \\
&= \begin{bmatrix} \mathbf{x}_{1t} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2t} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{nt} \end{bmatrix}\begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1n} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma^{n1} & \sigma^{n2} & \cdots & \sigma^{nn} \end{bmatrix}\begin{bmatrix} \mathbf{x}_{1t}' & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{x}_{2t}' & \cdots & \mathbf{0}' \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{x}_{nt}' \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11}\mathbf{x}_{1t}\mathbf{x}_{1t}' & \sigma^{12}\mathbf{x}_{1t}\mathbf{x}_{2t}' & \cdots & \sigma^{1n}\mathbf{x}_{1t}\mathbf{x}_{nt}' \\ \sigma^{21}\mathbf{x}_{2t}\mathbf{x}_{1t}' & \sigma^{22}\mathbf{x}_{2t}\mathbf{x}_{2t}' & \cdots & \sigma^{2n}\mathbf{x}_{2t}\mathbf{x}_{nt}' \\ \vdots & \vdots & \cdots & \vdots \\ \sigma^{n1}\mathbf{x}_{nt}\mathbf{x}_{1t}' & \sigma^{n2}\mathbf{x}_{nt}\mathbf{x}_{2t}' & \cdots & \sigma^{nn}\mathbf{x}_{nt}\mathbf{x}_{nt}' \end{bmatrix},
\end{aligned} \qquad [11.3.33]$$

where σ^{ij} denotes the row i, column j element of $\boldsymbol\Omega^{-1}$. Similarly,

$$\begin{aligned}
&[(\tilde{\boldsymbol{x}}_{1t}\tilde{y}_{1t}) + (\tilde{\boldsymbol{x}}_{2t}\tilde{y}_{2t}) + \cdots + (\tilde{\boldsymbol{x}}_{nt}\tilde{y}_{nt})] \\
&= [\tilde{\boldsymbol{x}}_{1t} \quad \tilde{\boldsymbol{x}}_{2t} \quad \cdots \quad \tilde{\boldsymbol{x}}_{nt}]\begin{bmatrix} \tilde{y}_{1t} \\ \tilde{y}_{2t} \\ \vdots \\ \tilde{y}_{nt} \end{bmatrix} \\
&= \mathcal{X}_t\mathbf{L}'\mathbf{L}\mathbf{y}_t \\
&= \begin{bmatrix} \mathbf{x}_{1t} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2t} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{nt} \end{bmatrix}\begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1n} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma^{n1} & \sigma^{n2} & \cdots & \sigma^{nn} \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11}\mathbf{x}_{1t}y_{1t} + \sigma^{12}\mathbf{x}_{1t}y_{2t} + \cdots + \sigma^{1n}\mathbf{x}_{1t}y_{nt} \\ \sigma^{21}\mathbf{x}_{2t}y_{1t} + \sigma^{22}\mathbf{x}_{2t}y_{2t} + \cdots + \sigma^{2n}\mathbf{x}_{2t}y_{nt} \\ \vdots \\ \sigma^{n1}\mathbf{x}_{nt}y_{1t} + \sigma^{n2}\mathbf{x}_{nt}y_{2t} + \cdots + \sigma^{nn}\mathbf{x}_{nt}y_{nt} \end{bmatrix}.
\end{aligned} \qquad [11.3.34]$$

Substituting [11.3.33] and [11.3.34] into [11.3.32], the *MLE* satisfies

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sigma^{11}\Sigma\mathbf{x}_{1t}\mathbf{x}'_{1t} & \sigma^{12}\Sigma\mathbf{x}_{1t}\mathbf{x}'_{2t} & \cdots & \sigma^{1n}\Sigma\mathbf{x}_{1t}\mathbf{x}'_{nt} \\ \sigma^{21}\Sigma\mathbf{x}_{2t}\mathbf{x}'_{1t} & \sigma^{22}\Sigma\mathbf{x}_{2t}\mathbf{x}'_{2t} & \cdots & \sigma^{2n}\Sigma\mathbf{x}_{2t}\mathbf{x}'_{nt} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma^{n1}\Sigma\mathbf{x}_{nt}\mathbf{x}'_{1t} & \sigma^{n2}\Sigma\mathbf{x}_{nt}\mathbf{x}'_{2t} & \cdots & \sigma^{nn}\Sigma\mathbf{x}_{nt}\mathbf{x}'_{nt} \end{bmatrix}^{-1} \times \begin{bmatrix} \Sigma(\sigma^{11}\mathbf{x}_{1t}y_{1t} + \sigma^{12}\mathbf{x}_{1t}y_{2t} + \cdots + \sigma^{1n}\mathbf{x}_{1t}y_{nt}) \\ \Sigma(\sigma^{21}\mathbf{x}_{2t}y_{1t} + \sigma^{22}\mathbf{x}_{2t}y_{2t} + \cdots + \sigma^{2n}\mathbf{x}_{2t}y_{nt}) \\ \vdots \\ \Sigma(\sigma^{n1}\mathbf{x}_{nt}y_{1t} + \sigma^{n2}\mathbf{x}_{nt}y_{2t} + \cdots + \sigma^{nn}\mathbf{x}_{nt}y_{nt}) \end{bmatrix}, \qquad [11.3.35]$$

where Σ denotes summation over $t = 1, 2, \ldots, T$.

The result from Section 11.1 was that when there are no restrictions on the *VAR*, maximum likelihood estimation is achieved by *OLS* equation by equation. This result can be seen as a special case of [11.3.35] by setting $\mathbf{x}_{1t} = \mathbf{x}_{2t} = \cdots = \mathbf{x}_{nt}$, for then [11.3.35] becomes

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\boldsymbol{\Omega}^{-1} \otimes (\Sigma\mathbf{x}_t\mathbf{x}'_t)]^{-1}\Sigma[(\boldsymbol{\Omega}^{-1}\mathbf{y}_t) \otimes \mathbf{x}_t] \\ &= [\boldsymbol{\Omega} \otimes (\Sigma\mathbf{x}_t\mathbf{x}'_t)^{-1}]\Sigma[(\boldsymbol{\Omega}^{-1}\mathbf{y}_t) \otimes \mathbf{x}_t] \\ &= [\mathbf{I}_n \otimes (\Sigma\mathbf{x}_t\mathbf{x}'_t)^{-1}]\Sigma[\mathbf{y}_t \otimes \mathbf{x}_t] \\ &= \begin{bmatrix} (\Sigma\mathbf{x}_t\mathbf{x}'_t)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\Sigma\mathbf{x}_t\mathbf{x}'_t)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\Sigma\mathbf{x}_t\mathbf{x}'_t)^{-1} \end{bmatrix} \begin{bmatrix} \Sigma y_{1t}\mathbf{x}_t \\ \Sigma y_{2t}\mathbf{x}_t \\ \vdots \\ \Sigma y_{nt}\mathbf{x}_t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \end{aligned}$$

as shown directly in Section 11.1.

Maximum likelihood estimation with constraints on both the coefficients and the variance-covariance matrix was discussed by Magnus (1978).

11.4. *The Impulse-Response Function*

In equation [10.1.15] a *VAR* was written in vector $MA(\infty)$ form as

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots. \qquad [11.4.1]$$

Thus, the matrix $\boldsymbol{\Psi}_s$ has the interpretation

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}'_t} = \boldsymbol{\Psi}_s; \qquad [11.4.2]$$

that is, the row i, column j element of $\boldsymbol{\Psi}_s$ identifies the consequences of a one-unit increase in the jth variable's innovation at date t (ε_{jt}) for the value of the ith variable at time $t + s$ ($y_{i,t+s}$), holding all other innovations at all dates constant.

If we were told that the first element of $\boldsymbol{\varepsilon}_t$ changed by δ_1 at the same time that the second element changed by $\delta_2, \ldots,$ and the nth element by δ_n, then the combined effect of these changes on the value of the vector $\mathbf{y}_{t+s}$ would be given by

$$\Delta \mathbf{y}_{t+s} = \frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon_{1t}} \delta_1 + \frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon_{2t}} \delta_2 + \cdots + \frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon_{nt}} \delta_n = \mathbf{\Psi}_s \boldsymbol{\delta}, \qquad [11.4.3]$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)'$.

Several analytic characterizations of $\mathbf{\Psi}_s$ were given in Section 10.1. A simple way to find these dynamic multipliers numerically is by simulation. To implement the simulation, set $\mathbf{y}_{t-1} = \mathbf{y}_{t-2} = \cdots = \mathbf{y}_{t-p} = \mathbf{0}$. Set $\varepsilon_{jt} = 1$ and all other elements of $\boldsymbol{\varepsilon}_t$ to zero, and simulate the system [11.1.1] for dates $t, t+1, t+2, \ldots$, with $\mathbf{c}$ and $\boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_{t+2}, \ldots$ all zero. The value of the vector $\mathbf{y}_{t+s}$ at date $t+s$ of this simulation corresponds to the jth column of the matrix $\mathbf{\Psi}_s$. By doing a separate simulation for impulses to each of the innovations ($j = 1, 2, \ldots, n$), all of the columns of $\mathbf{\Psi}_s$ can be calculated.

A plot of the row i, column j element of $\mathbf{\Psi}_s$,

$$\frac{\partial y_{i,t+s}}{\partial \varepsilon_{jt}}, \qquad [11.4.4]$$

as a function of s is called the *impulse-response function*. It describes the response of $y_{i,t+s}$ to a one-time impulse in y_{jt} with all other variables dated t or earlier held constant.

Is there a sense in which this multiplier can be viewed as measuring the causal effect of y_j on y_i? The discussion of Granger-causality tests suggests that we should be wary of such a claim. We are on surer ground with an atheoretical *VAR* if we confine ourselves to statements about forecasts. Consider, therefore, the following question. Let

$$\mathbf{x}'_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p})$$

denote the information received about the system as of date $t-1$. Suppose we are then told that the date t value of the first variable in the autoregression, y_{1t}, was higher than expected, so that ε_{1t} is positive. How does this cause us to revise our forecast of $y_{i,t+s}$? In other words, what is

$$\frac{\partial \hat{E}(y_{i,t+s} \mid y_{1t}, \mathbf{x}_{t-1})}{\partial y_{1t}}? \qquad [11.4.5]$$

The answer to this question is given by [11.4.4] with $j = 1$ only in the special case when $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) = \mathbf{\Omega}$ is a diagonal matrix. In the more general case when the elements of $\boldsymbol{\varepsilon}_t$ are contemporaneously correlated with one another, the fact that ε_{1t} is positive gives us some useful new information about the values of $\varepsilon_{2t}, \ldots, \varepsilon_{nt}$. This information has further implications for the value of $y_{i,t+s}$. To summarize these implications, we need to calculate the vector

$$\frac{\partial \hat{E}(\boldsymbol{\varepsilon}_t \mid y_{1t}, \mathbf{x}_{t-1})}{\partial y_{1t}}$$

and then use [11.4.3] to calculate the effect of this change in all the elements of $\boldsymbol{\varepsilon}_t$ on the value of $y_{i,t+s}$.

Yet another magnitude we might propose to measure is the forecast revision resulting from new information about, say, the second variable, y_{2t}, beyond that contained in the first variable, y_{1t}. Thus, we might calculate

$$\frac{\partial \hat{E}(y_{i,t+s} \mid y_{2t}, y_{1t}, \mathbf{x}_{t-1})}{\partial y_{2t}}. \qquad [11.4.6]$$

Similarly, for the variable designated number 3, we might seek

$$\frac{\partial \hat{E}(y_{i,t+s} | y_{3t}, y_{2t}, y_{1t}, \mathbf{x}_{t-1})}{\partial y_{3t}}, \tag{11.4.7}$$

and for variable n,

$$\frac{\partial \hat{E}(y_{i,t+s} | y_{nt}, y_{n-1,t}, \ldots, y_{1t}, \mathbf{x}_{t-1})}{\partial y_{nt}}. \tag{11.4.8}$$

This last magnitude corresponds to the effect of ε_{nt} with $\varepsilon_{1t}, \ldots, \varepsilon_{n-1,t}$ constant and is given simply by the row i, column n element of $\boldsymbol{\Psi}_s$.

The recursive information ordering in [11.4.5] through [11.4.8] is quite commonly used. For this ordering, the indicated multipliers can be calculated from the moving average coefficients ($\boldsymbol{\Psi}_s$) and the variance-covariance matrix of $\boldsymbol{\varepsilon}_t$ ($\boldsymbol{\Omega}$) by a simple algorithm. Recall from Section 4.4 that for any real symmetric positive definite matrix $\boldsymbol{\Omega}$, there exists a unique lower triangular matrix $\mathbf{A}$ with 1s along the principal diagonal and a unique diagonal matrix $\mathbf{D}$ with positive entries along the principal diagonal such that

$$\boldsymbol{\Omega} = \mathbf{ADA}'. \tag{11.4.9}$$

Using this matrix $\mathbf{A}$ we can construct an $(n \times 1)$ vector $\mathbf{u}_t$ from

$$\mathbf{u}_t \equiv \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t. \tag{11.4.10}$$

Notice that since $\boldsymbol{\varepsilon}_t$ is uncorrelated with its own lags or with lagged values of $\mathbf{y}$, it follows that $\mathbf{u}_t$ is also uncorrelated with its own lags or with lagged values of $\mathbf{y}$. The elements of $\mathbf{u}_t$ are furthermore uncorrelated with each other:

$$\begin{aligned} E(\mathbf{u}_t\mathbf{u}_t') &= [\mathbf{A}^{-1}]E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')[\mathbf{A}^{-1}]' \\ &= [\mathbf{A}^{-1}]\boldsymbol{\Omega}[\mathbf{A}']^{-1} \\ &= [\mathbf{A}^{-1}]\mathbf{ADA}'[\mathbf{A}']^{-1} \\ &= \mathbf{D}. \end{aligned} \tag{11.4.11}$$

But $\mathbf{D}$ is a diagonal matrix, verifying that the elements of $\mathbf{u}_t$ are mutually uncorrelated. The (j, j) element of $\mathbf{D}$ gives the variance of u_{jt}.

If both sides of [11.4.10] are premultiplied by $\mathbf{A}$, the result is

$$\mathbf{Au}_t = \boldsymbol{\varepsilon}_t. \tag{11.4.12}$$

Writing out the equations represented by [11.4.12] explicitly,

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \vdots \\ u_{nt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}. \tag{11.4.13}$$

Thus, u_{1t} is simply ε_{1t}. The jth row of [11.4.13] states that

$$u_{jt} = \varepsilon_{jt} - a_{j1}u_{1t} - a_{j2}u_{2t} - \cdots - a_{j,j-1}u_{j-1,t}.$$

But since u_{jt} is uncorrelated with $u_{1t}, u_{2t}, \ldots, u_{j-1,t}$, it follows that u_{jt} has the interpretation as the residual from a projection of ε_{jt} on $u_{1t}, u_{2t}, \ldots, u_{j-1,t}$:

$$\hat{E}(\varepsilon_{jt} | u_{1t}, u_{2t}, \ldots, u_{j-1,t}) = a_{j1}u_{1t} + a_{j2}u_{2t} + \cdots + a_{j,j-1}u_{j-1,t}. \tag{11.4.14}$$

The fact that the u_{jt} are uncorrelated further implies that the coefficient on u_{1t} in a projection of ε_{jt} on $(u_{1t}, u_{2t}, \ldots, u_{j-1,t})$ is the same as the coefficient on

u_{1t} in a projection of ε_{jt} on u_{1t} alone:

$$\hat{E}(\varepsilon_{jt}|u_{1t}) = a_{j1}u_{1t}. \qquad [11.4.15]$$

Recalling from [11.4.13] that $\varepsilon_{1t} = u_{1t}$, we see that new information about the value of ε_{1t} would cause us to revise our forecast of ε_{jt} by the amount

$$\frac{\partial \hat{E}(\varepsilon_{jt}|\varepsilon_{1t})}{\partial \varepsilon_{1t}} = \frac{\partial \hat{E}(\varepsilon_{jt}|u_{1t})}{\partial u_{1t}} = a_{j1}. \qquad [11.4.16]$$

Now ε_{1t} has the interpretation as $y_{1t} - \hat{E}(y_{1t}|\mathbf{x}_{t-1})$ and ε_{jt} has the interpretation as $y_{jt} - \hat{E}(y_{jt}|\mathbf{x}_{t-1})$. From the formula for updating a linear projection [4.5.14], the coefficient on y_{1t} in a linear projection of y_{jt} on y_{1t} and $\mathbf{x}_{t-1}$ is the same as the coefficient on ε_{1t} in a linear projection of ε_{jt} on ε_{1t}.[11] Hence,

$$\frac{\partial \hat{E}(\varepsilon_{jt}|y_{1t}, \mathbf{x}_{t-1})}{\partial y_{1t}} = a_{j1}. \qquad [11.4.17]$$

Combining these equations for $j = 1, 2, \ldots, n$ into a vector,

$$\frac{\partial \hat{E}(\boldsymbol{\varepsilon}_t|y_{1t}, \mathbf{x}_{t-1})}{\partial y_{1t}} = \mathbf{a}_1, \qquad [11.4.18]$$

where $\mathbf{a}_1$ denotes the first column of $\mathbf{A}$:

$$\mathbf{a}_1 \equiv \begin{bmatrix} 1 \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix}.$$

Substituting [11.4.18] into [11.4.3], the consequences for $\mathbf{y}_{t+s}$ of new information about y_{1t} beyond that contained in $\mathbf{x}_{t-1}$ are given by

$$\frac{\partial \hat{E}(\mathbf{y}_{t+s}|y_{1t}, \mathbf{x}_{t-1})}{\partial y_{1t}} = \mathbf{\Psi}_s\mathbf{a}_1.$$

Similarly, the variable u_{2t} represents the new information in y_{2t} beyond that contained in $(y_{1t}, \mathbf{x}_{t-1})$. This information would, of course, not cause us to change our assessment of ε_{1t} (which we know with certainty from y_{1t} and $\mathbf{x}_{t-1}$), but from [11.4.14] would cause us to revise our estimate of ε_{jt} for $j = 2, 3, \ldots, n$ by

$$\frac{\partial \hat{E}(\varepsilon_{jt}|u_{2t}, u_{1t})}{\partial u_{2t}} = a_{j2}.$$

Substituting this into [11.4.3], we conclude that

$$\frac{\partial \hat{E}(\mathbf{y}_{t+s}|y_{2t}, y_{1t}, \mathbf{x}_{t-1})}{\partial y_{2t}} = \mathbf{\Psi}_s\mathbf{a}_2,$$

[11]That is

$$\begin{aligned}\hat{E}(y_{jt}|y_{1t}, \mathbf{x}_{t-1}) &= \hat{E}(y_{jt}|\mathbf{x}_{t-1}) \\ &\quad + \operatorname{Cov}\{[y_{jt} - \hat{E}(y_{jt}|\mathbf{x}_{t-1})], [y_{1t} - \hat{E}(y_{1t}|\mathbf{x}_{t-1})]\} \\ &\quad \times \{\operatorname{Var}[y_{1t} - \hat{E}(y_{1t}|\mathbf{x}_{t-1})]\}^{-1}[y_{1t} - \hat{E}(y_{1t}|\mathbf{x}_{t-1})] \\ &= \hat{E}(y_{jt}|\mathbf{x}_{t-1}) + \operatorname{Cov}(\varepsilon_{jt}, \varepsilon_{1t})\cdot\{\operatorname{Var}(\varepsilon_{1t})\}^{-1}\cdot\varepsilon_{1t}.\end{aligned}$$

where

$$\mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ a_{32} \\ a_{42} \\ \vdots \\ a_{n2} \end{bmatrix}.$$

In general,

$$\frac{\partial \hat{E}(\mathbf{y}_{t+s} | y_{jt}, y_{j-1,t}, \ldots, y_{1t}, \mathbf{x}_{t-1})}{\partial y_{jt}} = \boldsymbol{\Psi}_s \mathbf{a}_j, \qquad [11.4.19]$$

where $\mathbf{a}_j$ denotes the jth column of the matrix $\mathbf{A}$ defined in [11.4.9].

The magnitude in [11.4.19] is a population moment, constructed from the population parameters $\boldsymbol{\Psi}_s$ and $\boldsymbol{\Omega}$ using [11.4.9]. For a given observed sample of size T, we would estimate the autoregressive coefficients $\hat{\boldsymbol{\Phi}}_1, \ldots, \hat{\boldsymbol{\Phi}}_p$ by *OLS* and construct $\hat{\boldsymbol{\Psi}}_s$ by simulating the estimated system. *OLS* estimation would also provide the estimate $\hat{\boldsymbol{\Omega}} = (1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$, where the ith element of $\hat{\boldsymbol{\varepsilon}}_t$ is the *OLS* sample residual for the ith equation in the *VAR* for date t. Matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{D}}$ satisfying $\hat{\boldsymbol{\Omega}} = \hat{\mathbf{A}}\hat{\mathbf{D}}\hat{\mathbf{A}}'$ could then be constructed from $\hat{\boldsymbol{\Omega}}$ using the algorithm described in Section 4.4. Notice that the elements of the vector $\hat{\mathbf{u}}_t = \hat{\mathbf{A}}^{-1}\hat{\boldsymbol{\varepsilon}}_t$ are then mutually orthogonal by construction:

$$(1/T) \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' = (1/T) \sum_{t=1}^{T} \hat{\mathbf{A}}^{-1} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' (\hat{\mathbf{A}}^{-1})' = \hat{\mathbf{A}}^{-1} \hat{\boldsymbol{\Omega}} (\hat{\mathbf{A}}^{-1})' = \hat{\mathbf{D}}.$$

The sample estimate of [11.4.19] is then

$$\hat{\boldsymbol{\Psi}}_s \hat{\mathbf{a}}_j, \qquad [11.4.20]$$

where $\hat{\mathbf{a}}_j$ denotes the jth column of the matrix $\hat{\mathbf{A}}$.

A plot of [11.4.20] as a function of s is known as an *orthogonalized impulse-response function*. It is based on decomposing the original *VAR* innovations $(\varepsilon_{1t}, \ldots, \varepsilon_{nt})$ into a set of uncorrelated components $(u_{1t}, \ldots, u_{nt})$ and calculating the consequences for $\mathbf{y}_{t+s}$ of a unit impulse in u_{jt}. These multipliers describe how new information about y_{jt} causes us to revise our forecast of $\mathbf{y}_{t+s}$, though the implicit definition of "new" information is different for each variable j.

What is the rationale for treating each variable differently? Clearly, if the *VAR* is being used as a purely atheoretical summary of the dynamics of a group of variables, there can be none—we could just as easily have labeled the second variable y_{1t} and the first variable y_{2t}, in which case we would have obtained different dynamic multipliers. By choosing a particular recursive ordering of the variables, the researcher is implicitly asking a set of questions about forecasting of the form of [11.4.5] through [11.4.8]. Whether we should orthogonalize in this way and how the variables should be ordered would seem to depend on why we want to ask such questions about forecasting in the first place. We will explore this issue in more depth in Section 11.6.

Before leaving the recursive orthogonalization, we note another popular form in which it is implemented and reported. Recall that $\mathbf{D}$ is a diagonal matrix whose (j, j) element is the variance of u_{jt}. Let $\mathbf{D}^{1/2}$ denote the diagonal matrix whose (j, j) element is the standard deviation of u_{jt}. Note that [11.4.9] could be written as

$$\boldsymbol{\Omega} = \mathbf{A}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{A}' = \mathbf{P}\mathbf{P}', \qquad [11.4.21]$$

where

$$\mathbf{P} \equiv \mathbf{A}\mathbf{D}^{1/2}.$$

Expression [11.4.21] is the *Cholesky decomposition* of the matrix $\mathbf{\Omega}$. Note that, like $\mathbf{A}$, the $(n \times n)$ matrix $\mathbf{P}$ is lower triangular, though whereas $\mathbf{A}$ has 1s along its principal diagonal, $\mathbf{P}$ has the standard deviation of $\mathbf{u}_t$ along its principal diagonal.

In place of $\mathbf{u}_t$ defined in [11.4.10], some researchers use

$$\mathbf{v}_t \equiv \mathbf{P}^{-1}\boldsymbol{\varepsilon}_t = \mathbf{D}^{-1/2}\mathbf{A}^{-1}\boldsymbol{\varepsilon}_t = \mathbf{D}^{-1/2}\mathbf{u}_t.$$

Thus, v_{jt} is just u_{jt} divided by its standard deviation $\sqrt{d_{jj}}$. A one-unit increase in v_{jt} is the same as a one-standard-deviation increase in u_{jt}.

In place of the dynamic multiplier $\partial y_{i,t+s}/\partial u_{jt}$, these researchers then report $\partial y_{i,t+s}/\partial v_{jt}$. The relation between these multipliers is clearly

$$\frac{\partial \mathbf{y}_{t+s}}{\partial v_{jt}} = \frac{\partial \mathbf{y}_{t+s}}{\partial u_{jt}}\sqrt{d_{jj}} = \mathbf{\Psi}_s\mathbf{a}_j\sqrt{d_{jj}}.$$

But $\mathbf{a}_j\sqrt{d_{jj}}$ is just the jth column of $\mathbf{A}\mathbf{D}^{1/2}$, which is the jth column of the Cholesky factor matrix $\mathbf{P}$. Denoting the jth column of $\mathbf{P}$ by $\mathbf{p}_j$, we have

$$\frac{\partial \mathbf{y}_{t+s}}{\partial v_{jt}} = \mathbf{\Psi}_s\mathbf{p}_j. \qquad [11.4.22]$$

Expression [11.4.22] is just [11.4.19] multiplied by the constant $\sqrt{\text{Var}(u_{jt})}$. Expression [11.4.19] gives the consequences of a one-unit increase in y_{jt}, where the units are those in which y_{jt} itself is measured. Expression [11.4.22] gives the consequences if y_{jt} were to increase by $\sqrt{\text{Var}(u_{jt})}$ units.

11.5. *Variance Decomposition*

Equations [10.1.14] and [10.1.16] identify the error in forecasting a *VAR* s periods into the future as

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\varepsilon}_{t+s} + \mathbf{\Psi}_1\boldsymbol{\varepsilon}_{t+s-1} + \mathbf{\Psi}_2\boldsymbol{\varepsilon}_{t+s-2} + \cdots + \mathbf{\Psi}_{s-1}\boldsymbol{\varepsilon}_{t+1}. \qquad [11.5.1]$$

The mean squared error of this s-period-ahead forecast is thus

$$\begin{aligned} MSE(\hat{\mathbf{y}}_{t+s|t}) &= E[(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t})'] \\ &= \mathbf{\Omega} + \mathbf{\Psi}_1\mathbf{\Omega}\mathbf{\Psi}_1' + \mathbf{\Psi}_2\mathbf{\Omega}\mathbf{\Psi}_2' + \cdots + \mathbf{\Psi}_{s-1}\mathbf{\Omega}\mathbf{\Psi}_{s-1}', \end{aligned} \qquad [11.5.2]$$

where

$$\mathbf{\Omega} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'). \qquad [11.5.3]$$

Let us now consider how each of the orthogonalized disturbances $(u_{1t}, \ldots, u_{nt})$ contributes to this *MSE*. Write [11.4.12] as

$$\boldsymbol{\varepsilon}_t = \mathbf{A}\mathbf{u}_t = \mathbf{a}_1u_{1t} + \mathbf{a}_2u_{2t} + \cdots + \mathbf{a}_nu_{nt}, \qquad [11.5.4]$$

where, as before, $\mathbf{a}_j$ denotes the jth column of the matrix $\mathbf{A}$ given in [11.4.9]. Recalling that the u_{jt}'s are uncorrelated, postmultiplying equation [11.5.4] by its transpose and taking expectations produces

$$\begin{aligned} \mathbf{\Omega} &= E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') \\ &= \mathbf{a}_1\mathbf{a}_1'\cdot\text{Var}(u_{1t}) + \mathbf{a}_2\mathbf{a}_2'\cdot\text{Var}(u_{2t}) + \cdots + \mathbf{a}_n\mathbf{a}_n'\cdot\text{Var}(u_{nt}), \end{aligned} \qquad [11.5.5]$$

where $\text{Var}(u_{jt})$ is the row j, column j element of the matrix $\mathbf{D}$ in [11.4.9]. Substituting [11.5.5] into [11.5.2], the *MSE* of the s-period-ahead forecast can be written as the sum of n terms, one arising from each of the disturbances u_{jt}:

$$\begin{aligned} MSE(\hat{\mathbf{y}}_{t+s|t}) = \sum_{j=1}^{n} \{\text{Var}(u_{jt})\cdot[\mathbf{a}_j\mathbf{a}_j' + \mathbf{\Psi}_1\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_1' \\ + \mathbf{\Psi}_2\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_2' + \cdots + \mathbf{\Psi}_{s-1}\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_{s-1}']\}. \end{aligned} \qquad [11.5.6]$$

With this expression, we can calculate the contribution of the jth orthogonalized innovation to the MSE of the s-period-ahead forecast:

$$\text{Var}(u_{jt})\cdot[\mathbf{a}_j\mathbf{a}_j' + \mathbf{\Psi}_1\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_1' + \mathbf{\Psi}_2\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_2' + \cdots + \mathbf{\Psi}_{s-1}\mathbf{a}_j\mathbf{a}_j'\mathbf{\Psi}_{s-1}'].$$

Again, this magnitude in general depends on the ordering of the variables.

As $s \to \infty$ for a covariance-stationary VAR, $MSE(\hat{\mathbf{y}}_{t+s|t}) \to \mathbf{\Gamma}_0$, the unconditional variance of the vector $\mathbf{y}_t$. Thus, [11.5.6] permits calculation of the portion of the total variance of y_i that is due to the disturbance u_j by letting s become suitably large.

Alternatively, recalling that $\mathbf{a}_j\cdot\sqrt{\text{Var}(u_{jt})}$ is equal to $\mathbf{p}_j$, the jth column of the Cholesky factor $\mathbf{P}$, result [11.5.6] can equivalently be written as

$$\begin{aligned} MSE(\hat{\mathbf{y}}_{t+s|t}) = \sum_{j=1}^{n} [&\mathbf{p}_j\mathbf{p}_j' + \mathbf{\Psi}_1\mathbf{p}_j\mathbf{p}_j'\mathbf{\Psi}_1' + \mathbf{\Psi}_2\mathbf{p}_j\mathbf{p}_j'\mathbf{\Psi}_2' \\ &+ \cdots + \mathbf{\Psi}_{s-1}\mathbf{p}_j\mathbf{p}_j'\mathbf{\Psi}_{s-1}']. \end{aligned} \qquad [11.5.7]$$

11.6. *Vector Autoregressions and Structural Econometric Models*

Pitfalls in Estimating Dynamic Structural Models

The vector autoregression was introduced in Section 10.1 as a statistical description of the dynamic interrelations between n different variables contained in the vector $\mathbf{y}_t$. This description made no use of prior theoretical ideas about how these variables are expected to be related, and therefore cannot be used to test our theories or interpret the data in terms of economic principles. This section explores the relation between VARs and structural econometric models.

Suppose that we would like to estimate a money demand function that expresses the public's willingness to hold cash as a function of the level of income and interest rates. The following specification was used by some early researchers:

$$M_t - P_t = \beta_0 + \beta_1 Y_t + \beta_2 I_t + \beta_3(M_{t-1} - P_{t-1}) + v_t^D. \qquad [11.6.1]$$

Here, M_t is the log of the nominal money balances held by the public at date t, P_t is the log of the aggregate price level, Y_t is the log of real GNP, and I_t is a nominal interest rate. The parameters β_1 and β_2 represent the effect of income and interest rates on desired cash holdings. Part of the adjustment in money balances to a change in income is thought to take place immediately, with further adjustments coming in subsequent periods. The parameter β_3 characterizes this partial adjustment. The disturbance v_t^D represents factors other than income and interest rates that influence money demand.

It was once common practice to estimate such a money demand equation with Cochrane-Orcutt adjustment for first-order serial correlation. The implicit assumption behind this procedure is that

$$v_t^D = \rho v_{t-1}^D + u_t^D, \qquad [11.6.2]$$

where u_t^D is white noise. Write equation [11.6.2] as $(1 - \rho L)v_t^D = u_t^D$ and multiply both sides of [11.6.1] by $(1 - \rho L)$:

$$\begin{aligned} M_t - P_t = {} & (1 - \rho)\beta_0 + \beta_1 Y_t - \beta_1\rho Y_{t-1} + \beta_2 I_t - \beta_2\rho I_{t-1} \\ & + (\beta_3 + \rho)(M_{t-1} - P_{t-1}) - \beta_3\rho(M_{t-2} - P_{t-2}) + u_t^D. \end{aligned} \qquad [11.6.3]$$

Equation [11.6.3] is a restricted version of

$$\begin{aligned} M_t - P_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \alpha_3 I_t + \alpha_4 I_{t-1} \\ + \alpha_5(M_{t-1} - P_{t-1}) + \alpha_6(M_{t-2} - P_{t-2}) + u_t^D, \end{aligned} \qquad [11.6.4]$$

where the seven parameters $(\alpha_0, \alpha_1, \ldots, \alpha_6)$ are restricted in [11.6.3] to be nonlinear functions of the underlying five parameters $(\rho, \beta_0, \beta_1, \beta_2, \beta_3)$. The assumption of [11.6.2] can thus be tested by comparing the fit of [11.6.3] with that from unconstrained estimation of [11.6.4].

By definition, v_t^D represents factors influencing money demand for which the researcher has no explicit theory. It therefore seems odd to place great confidence in a detailed specification of its dynamics such as [11.6.2] without testing this assumption against the data. For example, there do not seem to be clear theoretical grounds for ruling out a specification such as

$$v_t^D = \rho_1 v_{t-1}^D + \rho_2 v_{t-2}^D + u_t^D,$$

or, for that matter, a specification in which v_t^D is correlated with lagged values of Y or I.

Equation [11.6.1] further assumes that the dynamic multiplier relating money demand to income is proportional to that relating money demand to the interest rate:

$$\frac{\partial(M_{t+s} - P_{t+s})}{\partial Y_t} = \beta_1 \beta_3^s$$

$$\frac{\partial(M_{t+s} - P_{t+s})}{\partial I_t} = \beta_2 \beta_3^s.$$

Again, it seems a good idea to test this assumption before imposing it, by comparing the fit of [11.6.1] with that of a more general dynamic model. Finally, inflation may have effects on money demand that are not captured by nominal interest rates. The specification in [11.6.1] incorporates very strong assumptions about the way nominal money demand responds to the price level.

To summarize, a specification such as [11.6.1] and [11.6.2] implicitly imposes many restrictions on dynamics for which there is little or no justification on the basis of economic theory. Before relying on the inferences of [11.6.1] and [11.6.2], it seems a good idea to test that model against a more general specification such as

$$\begin{aligned} M_t = k_1 &+ \beta_{12}^{(0)} P_t + \beta_{13}^{(0)} Y_t + \beta_{14}^{(0)} I_t \\ &+ \beta_{11}^{(1)} M_{t-1} + \beta_{12}^{(1)} P_{t-1} + \beta_{13}^{(1)} Y_{t-1} + \beta_{14}^{(1)} I_{t-1} \\ &+ \beta_{11}^{(2)} M_{t-2} + \beta_{12}^{(2)} P_{t-2} + \beta_{13}^{(2)} Y_{t-2} + \beta_{14}^{(2)} I_{t-2} + \cdots \\ &+ \beta_{11}^{(p)} M_{t-p} + \beta_{12}^{(p)} P_{t-p} + \beta_{13}^{(p)} Y_{t-p} + \beta_{14}^{(p)} I_{t-p} + u_t^D. \end{aligned} \qquad [11.6.5]$$

Like equation [11.6.1], the specification in [11.6.5] is regarded as a structural money demand equation; $\beta_{13}^{(0)}$ and $\beta_{14}^{(0)}$ are interpreted as the effects of current income and the interest rate on desired money holdings, and u_t^D represents factors influencing money demand other than inflation, income, and interest rates. Compared with [11.6.1], the specification in [11.6.5] generalizes the dynamic behavior for the error term v_t^D, the partial adjustment process, and the influence of the price level on desired money holdings.

Although [11.6.5] relaxes many of the dubious restrictions on the dynamics implied by [11.6.1], it is still not possible to estimate [11.6.5] by *OLS*, because of simultaneous equations bias. *OLS* estimation of [11.6.5] will summarize the correlation between money, the price level, income, and the interest rate. The public's money demand adjustments are one reason these variables will be correlated, but not the only one. For example, each period, the central bank may be adjusting the interest rate I_t to a level consistent with its policy objectives, which may depend on current and lagged values of income, the interest rate, the price level, and the money supply:

$$\begin{aligned} I_t = k_4 &+ \beta_{41}^{(0)} M_t + \beta_{42}^{(0)} P_t + \beta_{43}^{(0)} Y_t \\ &+ \beta_{41}^{(1)} M_{t-1} + \beta_{42}^{(1)} P_{t-1} + \beta_{43}^{(1)} Y_{t-1} + \beta_{44}^{(1)} I_{t-1} \\ &+ \beta_{41}^{(2)} M_{t-2} + \beta_{42}^{(2)} P_{t-2} + \beta_{43}^{(2)} Y_{t-2} + \beta_{44}^{(2)} I_{t-2} + \cdots \\ &+ \beta_{41}^{(p)} M_{t-p} + \beta_{42}^{(p)} P_{t-p} + \beta_{43}^{(p)} Y_{t-p} + \beta_{44}^{(p)} I_{t-p} + u_t^C. \end{aligned} \qquad [11.6.6]$$

Here, for example, $\beta_{42}^{(0)}$ captures the effect of the current price level on the interest rate that the central bank tries to achieve. The disturbance u_t^C captures changes in policy that cannot be described as a deterministic function of current and lagged money, the price level, income, and the interest rate. If the money demand disturbance u_t^D is unusually large, this will make M_t unusually large. If $\beta_{41}^{(0)} > 0$, this would cause I_t to be unusually large as well, in which case u_t^D would be positively correlated with the explanatory variable I_t in equation [11.6.5]. Thus, [11.6.5] cannot be estimated by *OLS*.

Nor is central bank policy and endogeneity of I_t the only reason to be concerned about simultaneous equations bias. Money demand disturbances and changes in central bank policy also have effects on aggregate output and the price level, so that Y_t and P_t in [11.6.5] are endogenous as well. An aggregate demand equation, for example, might be postulated that relates the level of output to the money supply, price level, and interest rate:

$$\begin{aligned} Y_t = k_3 &+ \beta_{31}^{(0)} M_t + \beta_{32}^{(0)} P_t + \beta_{34}^{(0)} I_t \\ &+ \beta_{31}^{(1)} M_{t-1} + \beta_{32}^{(1)} P_{t-1} + \beta_{33}^{(1)} Y_{t-1} + \beta_{34}^{(1)} I_{t-1} \\ &+ \beta_{31}^{(2)} M_{t-2} + \beta_{32}^{(2)} P_{t-2} + \beta_{33}^{(2)} Y_{t-2} + \beta_{34}^{(2)} I_{t-2} + \cdots \\ &+ \beta_{31}^{(p)} M_{t-p} + \beta_{32}^{(p)} P_{t-p} + \beta_{33}^{(p)} Y_{t-p} + \beta_{34}^{(p)} I_{t-p} + u_t^A, \end{aligned} \qquad [11.6.7]$$

with u_t^A representing other factors influencing aggregate demand. Similarly, an aggregate supply curve might relate the aggregate price level to the other variables being studied. The logical conclusion of such reasoning is that *all* of the date t explanatory variables in [11.6.5] should be treated as endogenous.

Relation Between Dynamic Structural Models and Vector Autoregressions

The system of equations [11.6.5] through [11.6.7] (along with an analogous aggregate supply equation describing P_t) can be collected and written in vector form as

$$\mathbf{B}_0 \mathbf{y}_t = \mathbf{k} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{B}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{B}_p \mathbf{y}_{t-p} + \mathbf{u}_t, \qquad [11.6.8]$$

where

$$\mathbf{y}_t = (M_t, P_t, Y_t, I_t)'$$
$$\mathbf{u}_t = (u_t^D, u_t^S, u_t^A, u_t^C)'$$
$$\mathbf{B}_0 = \begin{bmatrix} 1 & -\beta_{12}^{(0)} & -\beta_{13}^{(0)} & -\beta_{14}^{(0)} \\ -\beta_{21}^{(0)} & 1 & -\beta_{23}^{(0)} & -\beta_{24}^{(0)} \\ -\beta_{31}^{(0)} & -\beta_{32}^{(0)} & 1 & -\beta_{34}^{(0)} \\ -\beta_{41}^{(0)} & -\beta_{42}^{(0)} & -\beta_{43}^{(0)} & 1 \end{bmatrix}$$
$$\mathbf{k} = (k_1, k_2, k_3, k_4)'$$

and $\mathbf{B}_s$ is a (4×4) matrix whose row i, column j element is given by $\beta_{ij}^{(s)}$ for $s = 1, 2, \ldots, p$. A large class of structural models for an $(n \times 1)$ vector $\mathbf{y}_t$ can be written in the form of [11.6.8].

Generalizing the argument in [11.6.3], it is assumed that a sufficient number of lags of p are included and the matrices $\mathbf{B}_s$ are defined so that $\mathbf{u}_t$ is vector white noise. If instead, say, $\mathbf{u}_t$ followed an rth-order VAR, with

$$\mathbf{u}_t = \mathbf{F}_1\mathbf{u}_{t-1} + \mathbf{F}_2\mathbf{u}_{t-2} + \cdots + \mathbf{F}_r\mathbf{u}_{t-r} + \boldsymbol{\varepsilon}_t,$$

then we could premultiply [11.6.8] by $(\mathbf{I}_n - \mathbf{F}_1 L^1 - \mathbf{F}_2 L^2 - \cdots - \mathbf{F}_r L^r)$ to arrive at a system of the same basic form as [11.6.8] with p replaced by $(p + r)$ and with $\mathbf{u}_t$ replaced by the white noise disturbance $\boldsymbol{\varepsilon}_t$.

If each side of [11.6.8] is premultiplied by $\mathbf{B}_0^{-1}$, the result is

$$\mathbf{y}_t = \mathbf{c} + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad [11.6.9]$$

where

$$\mathbf{c} = \mathbf{B}_0^{-1}\mathbf{k} \qquad [11.6.10]$$
$$\boldsymbol{\Phi}_s = \mathbf{B}_0^{-1}\mathbf{B}_s \qquad \text{for } s = 1, 2, \ldots, p \qquad [11.6.11]$$
$$\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t. \qquad [11.6.12]$$

Assuming that [11.6.8] is parameterized sufficiently richly that $\mathbf{u}_t$ is vector white noise, then $\boldsymbol{\varepsilon}_t$ will also be vector white noise and [11.6.9] will be recognized as the vector autoregressive representation for the dynamic structural system [11.6.8]. Thus, a VAR can be viewed as the reduced form of a general dynamic structural model.

Interpreting Impulse-Response Functions

In Section 11.4 we calculated the impulse-response function

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon_{jt}}. \qquad [11.6.13]$$

This magnitude describes the effect of an innovation in the jth variable on future values of each of the variables in the system. According to [11.6.12], the VAR innovation ε_{jt} is a linear combination of the structural disturbances $\mathbf{u}_t$. For example, it might turn out that

$$\varepsilon_{1t} = 0.3u_t^D - 0.6u_t^S + 0.1u_t^A - 0.5u_t^C.$$

In this case, if the cash held by the public is larger than would have been forecast using the VAR (ε_{1t} is positive), this might be because the public's demand for cash is higher than is normally associated with the current level of income and interest rate (that is, u_t^D is positive). Alternatively, ε_{1t} might be positive because the central bank has chosen to ease credit (u_t^C is negative), or a variety of other factors. In general, ε_{1t} represents a combination of all the different influences that matter for any variables in the economy. Viewed this way, it is not clear why the magnitude [11.6.13] is of particular interest.

By contrast, if we were able to calculate

$$\frac{\partial \mathbf{y}_{t+s}}{\partial u_t^C}, \qquad [11.6.14]$$

this would be of considerable interest. Expression [11.6.14] identifies the dynamic consequences for the economy if the central bank were to tighten credit more than usual and is a key magnitude for describing the effects of monetary policy on the economy.

Section 11.4 also discussed calculation of an orthogonalized impulse-response function. For $\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$, we found a lower triangular matrix $\mathbf{A}$ and a diagonal matrix $\mathbf{D}$ such that $\boldsymbol{\Omega} = \mathbf{ADA}'$. We then constructed the vector $\mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$ and calculated the consequences of changes in each element of this vector for future values of $\mathbf{y}$.

Recall from [11.6.12] that the structural disturbances $\mathbf{u}_t$ are related to the VAR innovations $\boldsymbol{\varepsilon}_t$ by

$$\mathbf{u}_t = \mathbf{B}_0\boldsymbol{\varepsilon}_t. \qquad [11.6.15]$$

Suppose that it happened to be the case that the matrix of structural parameters $\mathbf{B}_0$ was exactly equal to the matrix $\mathbf{A}^{-1}$. Then the orthogonalized innovations would coincide with the true structural disturbances:

$$\mathbf{u}_t = \mathbf{B}_0\boldsymbol{\varepsilon}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t. \qquad [11.6.16]$$

In this case, the method described in Section 11.4 could be used to find the answers to important questions such as [11.6.14].

Is there any reason to hope that $\mathbf{B}_0$ and $\mathbf{A}^{-1}$ would be the same matrix? Since $\mathbf{A}$ is lower triangular, this clearly requires $\mathbf{B}_0$ to be lower triangular. In the example [11.6.8], this would require that the current values of P, Y, and I do not influence money demand, that the current value of M but not that of Y or I enters into the aggregate supply curve, and so on. Such assumptions are rather unusual, though there may be another way to order the variables such that a recursive structure is more palatable. For example, a Keynesian might argue that prices respond to other economic variables only with a lag, so that the coefficients on current variables in the aggregate supply equation are all zero. Perhaps money and interest rates influence aggregate demand only with a lag, so that their current values are excluded from the aggregate demand equation. One might try to argue further that the interest rate affects desired money holdings only with a lag as well. Because most central banks monitor current economic conditions quite carefully, perhaps all the current values should be included in the equation for I_t. These assumptions suggest ordering the variables as $\mathbf{y}_t = (P_t, Y_t, M_t, I_t)'$, for which the structural model would be

$$\begin{bmatrix} P_t \\ Y_t \\ M_t \\ I_t \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{21}^{(0)} & 0 & 0 & 0 \\ \beta_{31}^{(0)} & \beta_{32}^{(0)} & 0 & 0 \\ \beta_{41}^{(0)} & \beta_{42}^{(0)} & \beta_{43}^{(0)} & 0 \end{bmatrix} \begin{bmatrix} P_t \\ Y_t \\ M_t \\ I_t \end{bmatrix}$$

$$+ \begin{bmatrix} \beta_{11}^{(1)} & \beta_{12}^{(1)} & \beta_{13}^{(1)} & \beta_{14}^{(1)} \\ \beta_{21}^{(1)} & \beta_{22}^{(1)} & \beta_{23}^{(1)} & \beta_{24}^{(1)} \\ \beta_{31}^{(1)} & \beta_{32}^{(1)} & \beta_{33}^{(1)} & \beta_{34}^{(1)} \\ \beta_{41}^{(1)} & \beta_{42}^{(1)} & \beta_{43}^{(1)} & \beta_{44}^{(1)} \end{bmatrix} \begin{bmatrix} P_{t-1} \\ Y_{t-1} \\ M_{t-1} \\ I_{t-1} \end{bmatrix} + \cdots \qquad [11.6.17]$$

$$+ \begin{bmatrix} \beta_{11}^{(p)} & \beta_{12}^{(p)} & \beta_{13}^{(p)} & \beta_{14}^{(p)} \\ \beta_{21}^{(p)} & \beta_{22}^{(p)} & \beta_{23}^{(p)} & \beta_{24}^{(p)} \\ \beta_{31}^{(p)} & \beta_{32}^{(p)} & \beta_{33}^{(p)} & \beta_{34}^{(p)} \\ \beta_{41}^{(p)} & \beta_{42}^{(p)} & \beta_{43}^{(p)} & \beta_{44}^{(p)} \end{bmatrix} \begin{bmatrix} P_{t-p} \\ Y_{t-p} \\ M_{t-p} \\ I_{t-p} \end{bmatrix} + \begin{bmatrix} u_t^S \\ u_t^A \\ u_t^D \\ u_t^C \end{bmatrix}.$$

Suppose there exists such an ordering of the variables for which $\mathbf{B}_0$ is lower triangular. Write the dynamic structural model [11.6.8] as

$$\mathbf{B}_0\mathbf{y}_t = -\mathbf{\Gamma}\mathbf{x}_t + \mathbf{u}_t, \qquad [11.6.18]$$

where

$$\underset{[n\times(np+1)]}{-\mathbf{\Gamma}} \equiv [\mathbf{k} \quad \mathbf{B}_1 \quad \mathbf{B}_2 \quad \cdots \quad \mathbf{B}_p]$$

$$\underset{[(np+1)\times 1]}{\mathbf{x}_t} \equiv \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}.$$

Suppose, furthermore, that the disturbances in the structural equations are serially uncorrelated and uncorrelated with each other:

$$E(\mathbf{u}_t\mathbf{u}_\tau') = \begin{cases} \mathbf{D} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise,} \end{cases} \qquad [11.6.19]$$

where $\mathbf{D}$ is a diagonal matrix. The VAR is the reduced form of the dynamic structural model [11.6.18] and can be written as

$$\mathbf{y}_t = \mathbf{\Pi}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \qquad [11.6.20]$$

where

$$\mathbf{\Pi}' = -\mathbf{B}_0^{-1}\mathbf{\Gamma} \qquad [11.6.21]$$

$$\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t. \qquad [11.6.22]$$

Letting $\mathbf{\Omega}$ denote the variance-covariance matrix of $\boldsymbol{\varepsilon}_t$, [11.6.22] implies

$$\mathbf{\Omega} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{B}_0^{-1}E(\mathbf{u}_t\mathbf{u}_t')(\mathbf{B}_0^{-1})' = \mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})'. \qquad [11.6.23]$$

Note that if the only restrictions on the dynamic structural model are that $\mathbf{B}_0$ is lower triangular with unit coefficients along the principal diagonal and that $\mathbf{D}$ is diagonal, then the structural model is just identified. To see this, note that these restrictions imply that $\mathbf{B}_0^{-1}$ must also be lower triangular with unit coefficients along the principal diagonal. Recall from Section 4.4 that given any positive definite symmetric matrix $\mathbf{\Omega}$, there exist a unique lower triangular matrix $\mathbf{A}$ with 1s along the principal diagonal and a diagonal matrix $\mathbf{D}$ with positive entries along the principal diagonal such that $\mathbf{\Omega} = \mathbf{ADA}'$. Thus, unique values $\mathbf{B}_0^{-1}$ and $\mathbf{D}$ of the required form can always be found that satisfy [11.6.23]. Moreover, any $\mathbf{B}_0$ matrix of this form is nonsingular, so that $\mathbf{\Gamma}$ in [11.6.21] can be calculated uniquely from $\mathbf{B}_0$ and $\mathbf{\Pi}$ as $\mathbf{\Gamma} = -\mathbf{B}_0\mathbf{\Pi}'$. Thus, given any allowable values for the reduced-form parameters ($\mathbf{\Pi}$ and $\mathbf{\Omega}$), there exist unique values for the structural parameters ($\mathbf{B}_0$, $\mathbf{\Gamma}$, and $\mathbf{D}$) of the specified form, establishing that the structural model is just identified.

Since the model is just identified, full-information maximum likelihood (*FIML*) estimates of ($\mathbf{B}_0$, $\mathbf{\Gamma}$, and $\mathbf{D}$) can be obtained by first maximizing the likelihood function with respect to the reduced-form parameters ($\mathbf{\Pi}$ and $\mathbf{\Omega}$) and then using the unique mapping from reduced-form parameters to find the structural parameters. The maximum likelihood estimates of $\mathbf{\Pi}$ are found from *OLS* regressions of the elements of $\mathbf{y}_t$ on $\mathbf{x}_t$, and the *MLE* of $\mathbf{\Omega}$ is obtained from the variance-covariance matrix of the residuals from these regressions. The estimates $\hat{\mathbf{B}}_0^{-1}$ and $\hat{\mathbf{D}}$ are then found from the triangular factorization of $\hat{\mathbf{\Omega}}$. This, however, is precisely the procedure described in calculating the orthogonalized innovations in Section 11.4. The estimate $\hat{\mathbf{A}}$ described there is thus the same as the *FIML* estimate of $\mathbf{B}_0^{-1}$. The vector of orthogonalized residuals $\mathbf{u}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$ would correspond to the vector of structural disturbances, and the orthogonalized impulse-response coefficients would give the dynamic consequences of the structural events represented by $\mathbf{u}_t$, provided that the structural model is lower triangular as in [11.6.17].

Nonrecursive Structural VAR*s*

Even if the structural model cannot be written in lower triangular form, it may be possible to give a structural interpretation to a *VAR* using a similar idea to that in equation [11.6.23]. Specifically, a structural model specifies a set of restrictions on $\mathbf{B}_0$ and $\mathbf{D}$, and we can try to find values satisfying these restrictions such that $\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})' = \mathbf{\Omega}$. This point was developed by Bernanke (1986), Blanchard and Watson (1986), and Sims (1986).

For illustration, consider again the model of supply and demand discussed in equations [9.3.2] and [9.3.3]. In that specification, quantity (q_t) and price (p_t) were endogeneous variables and weather (w_t) was exogenous, and it was assumed that both disturbances were i.i.d. The structural *VAR* approach to this model would allow quite general dynamics by adding p lags of all three variables to equations [9.3.2] and [9.3.3], as well as adding a third equation to describe the dynamic behavior of weather. Weather presumably does not depend on the behavior of the market, so the third equation would for this example just be a univariate autoregression. The model would then be

$$\begin{aligned} q_t = \beta p_t &+ \beta_{11}^{(1)}q_{t-1} + \beta_{12}^{(1)}p_{t-1} + \beta_{13}^{(1)}w_{t-1} \\ &+ \beta_{11}^{(2)}q_{t-2} + \beta_{12}^{(2)}p_{t-2} + \beta_{13}^{(2)}w_{t-2} + \cdots \\ &+ \beta_{11}^{(p)}q_{t-p} + \beta_{12}^{(p)}p_{t-p} + \beta_{13}^{(p)}w_{t-p} + u_t^d \end{aligned} \qquad [11.6.24]$$

$$\begin{aligned} q_t = {} & \gamma p_t + h w_t + \beta_{21}^{(1)} q_{t-1} + \beta_{22}^{(1)} p_{t-1} + \beta_{23}^{(1)} w_{t-1} \\ & + \beta_{21}^{(2)} q_{t-2} + \beta_{22}^{(2)} p_{t-2} + \beta_{23}^{(2)} w_{t-2} + \cdots \\ & + \beta_{21}^{(p)} q_{t-p} + \beta_{22}^{(p)} p_{t-p} + \beta_{23}^{(p)} w_{t-p} + u_t^s \end{aligned} \qquad [11.6.25]$$

$$w_t = \beta_{33}^{(1)} w_{t-1} + \beta_{33}^{(2)} w_{t-2} + \cdots + \beta_{33}^{(p)} w_{t-p} + u_t^w. \qquad [11.6.26]$$

We could then take $(u_t^d, u_t^s, u_t^w)'$ to be a white noise vector with diagonal variance-covariance matrix given by $\mathbf{D}$. This is an example of a structural model [11.6.18] in which

$$\mathbf{B}_0 = \begin{bmatrix} 1 & -\beta & 0 \\ 1 & -\gamma & -h \\ 0 & 0 & 1 \end{bmatrix}. \qquad [11.6.27]$$

There is no way to order the variables so as to make the matrix $\mathbf{B}_0$ lower triangular. However, equation [11.6.22] indicates that the structural disturbances $\mathbf{u}_t$ are related to the *VAR* residuals $\boldsymbol{\varepsilon}_t$ by $\boldsymbol{\varepsilon}_t = \mathbf{B}_0^{-1}\mathbf{u}_t$. Thus, if $\mathbf{B}_0$ is estimated by maximum likelihood, then the impulse-response functions could be calculated as in Section 11.4 with $\mathbf{A}$ replaced by $\mathbf{B}_0^{-1}$, and the results would give the effects of each of the structural disturbances on subsequent values of variables of the system. Specifically,

$$\frac{\partial \boldsymbol{\varepsilon}_t}{\partial \mathbf{u}_t'} = \mathbf{B}_0^{-1},$$

so that the effect on $\boldsymbol{\varepsilon}_t$ of the jth structural disturbance u_{jt} is given by $\mathbf{b}^j$, the jth column of $\mathbf{B}_0^{-1}$. Thus, we would calculate

$$\frac{\partial \mathbf{y}_{t+s}}{\partial u_{jt}} = \frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} \frac{\partial \boldsymbol{\varepsilon}_t}{\partial u_{jt}} = \boldsymbol{\Psi}_s \mathbf{b}^j$$

for $\boldsymbol{\Psi}_s$ the $(n \times n)$ matrix of coefficients for the sth lag of the $MA(\infty)$ representation [11.4.1].

FIML *Estimation of a Structural* VAR *with Unrestricted Dynamics*

FIML estimation is particularly simple if there are no restrictions on the coefficients $\boldsymbol{\Gamma}$ on lagged variables in [11.6.18]. For example, this would require including lagged values of p_{t-j} and q_{t-j} in the weather equation [11.6.26]. Using [11.6.23], the log likelihood function for the system [11.6.18] can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{B}_0, \mathbf{D}, \boldsymbol{\Pi}) = {} & -(Tn/2)\log(2\pi) - (T/2)\log|\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})'| \\ & - (1/2)\sum_{t=1}^{T} [\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t]'[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}[\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t]. \end{aligned} \qquad [11.6.28]$$

If there are no restrictions on lagged dynamics, this is maximized with respect to $\boldsymbol{\Pi}$ by *OLS* regression of $\mathbf{y}_t$ on $\mathbf{x}_t$. Substituting this estimate into [11.6.28] as in

[11.1.25] produces

$$\mathscr{L}(\mathbf{B}_0, \mathbf{D}, \hat{\mathbf{\Pi}}) = -(Tn/2)\log(2\pi) - (T/2)\log|\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})'| - (1/2)\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}\hat{\boldsymbol{\varepsilon}}_t. \qquad [11.6.29]$$

But

$$\begin{aligned}\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t'[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}\hat{\boldsymbol{\varepsilon}}_t &= \sum_{t=1}^{T} \operatorname{trace}\{\hat{\boldsymbol{\varepsilon}}_t'[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}\hat{\boldsymbol{\varepsilon}}_t\} \\ &= \sum_{t=1}^{T} \operatorname{trace}\{[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'\} \\ &= \operatorname{trace}\{[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}T\cdot\hat{\mathbf{\Omega}}\} \\ &= T \times \operatorname{trace}\{[\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})']^{-1}\hat{\mathbf{\Omega}}\} \\ &= T \times \operatorname{trace}\{(\mathbf{B}_0'\mathbf{D}^{-1}\mathbf{B}_0)\hat{\mathbf{\Omega}}\}.\end{aligned} \qquad [11.6.30]$$

Furthermore,

$$\log|\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})'| = \log\{|\mathbf{B}_0^{-1}|\cdot|\mathbf{D}|\cdot|\mathbf{B}_0^{-1}|\} = -\log|\mathbf{B}_0|^2 + \log|\mathbf{D}|. \qquad [11.6.31]$$

Substituting [11.6.31] and [11.6.30] into [11.6.29], *FIML* estimates of the structural parameters are found by choosing $\mathbf{B}_0$ and $\mathbf{D}$ so as to maximize

$$\mathscr{L}(\mathbf{B}_0, \mathbf{D}, \hat{\mathbf{\Pi}}) = -(Tn/2)\log(2\pi) + (T/2)\log|\mathbf{B}_0|^2 - (T/2)\log|\mathbf{D}| - (T/2)\operatorname{trace}\{(\mathbf{B}_0'\mathbf{D}^{-1}\mathbf{B}_0)\hat{\mathbf{\Omega}}\}. \qquad [11.6.32]$$

Using calculations similar to those used to analyze [11.1.25], one can show that if there exist unique matrices $\mathbf{B}_0$ and $\mathbf{D}$ of the required form satisfying $\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})' = \mathbf{\Omega}$, then maximization of [11.6.32] will produce estimates $\hat{\mathbf{B}}_0$ and $\hat{\mathbf{D}}$ satisfying

$$\hat{\mathbf{B}}_0^{-1}\hat{\mathbf{D}}(\hat{\mathbf{B}}_0^{-1})' = \hat{\mathbf{\Omega}}. \qquad [11.6.33]$$

This is a nonlinear system of equations, and numerical maximization of [11.6.32] offers a convenient general approach to finding a solution to this system of equations.

Identification of Structural VARs

The existence of a unique maximum of [11.6.32] requires both an order condition and a rank condition for identification. The order condition is that $\mathbf{B}_0$ and $\mathbf{D}$ have no more unknown parameters than $\mathbf{\Omega}$. Since $\mathbf{\Omega}$ is symmetric, it can be summarized by $n(n+1)/2$ distinct values. If $\mathbf{D}$ is diagonal, it requires n parameters, meaning that $\mathbf{B}_0$ can have no more than $n(n-1)/2$ free parameters. For the supply-and-demand example of [11.6.24] through [11.6.26], $n = 3$, and the matrix $\mathbf{B}_0$ in [11.6.27] has $3(3-1)/2 = 3$ free parameters (β, γ, and h). Thus, that example satisfies the order condition for identification.

Even if the order condition is satisfied, the model may still not be identified. For example, suppose that

$$\mathbf{B}_0 = \begin{bmatrix} 1 & -\beta & 0 \\ 1 & -\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Even though this specification satisfies the order condition, it fails the rank condition, since the value of the likelihood function will be unchanged if β and γ are switched along with σ_d^2 and σ_s^2.

To characterize the rank condition, suppose that there are n_B elements of $\mathbf{B}_0$ that must be estimated; collect these in an $(n_B \times 1)$ vector $\boldsymbol{\theta}_B$. The identifying assumptions can be represented as a known $(n^2 \times n_B)$ matrix $\mathbf{S}_B$ and a known $(n^2 \times 1)$ vector $\mathbf{s}_B$ for which

$$\text{vec}(\mathbf{B}_0) = \mathbf{S}_B\boldsymbol{\theta}_B + \mathbf{s}_B. \qquad [11.6.34]$$

For example, for the dynamic model of supply and demand represented by [11.6.27],

$$\text{vec}(\mathbf{B}_0) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -\beta \\ -\gamma \\ 0 \\ 0 \\ -h \\ 1 \end{bmatrix} \qquad \boldsymbol{\theta}_B = \begin{bmatrix} \beta \\ \gamma \\ h \end{bmatrix}$$

$$\mathbf{S}_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{s}_B = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Similarly, collect the unknown elements of $\mathbf{D}$ in an $(n_D \times 1)$ vector $\boldsymbol{\theta}_D$, with

$$\text{vec}(\mathbf{D}) = \mathbf{S}_D\boldsymbol{\theta}_D + \mathbf{s}_D \qquad [11.6.35]$$

for $\mathbf{S}_D$ an $(n^2 \times n_D)$ matrix and $\mathbf{s}_D$ an $(n^2 \times 1)$ vector. For the supply-and-demand example,

$$\text{vec}(\mathbf{D}) = \begin{bmatrix} \sigma_d^2 \\ 0 \\ 0 \\ 0 \\ \sigma_s^2 \\ 0 \\ 0 \\ 0 \\ \sigma_w^2 \end{bmatrix} \qquad \boldsymbol{\theta}_D = \begin{bmatrix} \sigma_d^2 \\ \sigma_s^2 \\ \sigma_w^2 \end{bmatrix}$$

$$\mathbf{S}_D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{s}_D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since [11.6.33] is an equation relating two symmetric matrices, there are $n^* \equiv n(n+1)/2$ separate conditions, represented by

$$\text{vech}(\boldsymbol{\Omega}) = \text{vech}\Big([\mathbf{B}_0(\boldsymbol{\theta}_B)]^{-1}[\mathbf{D}(\boldsymbol{\theta}_D)]\{[\mathbf{B}_0(\boldsymbol{\theta}_B)]^{-1}\}'\Big). \qquad [11.6.36]$$

Denote the right side of [11.6.36] by $\mathbf{f}(\boldsymbol{\theta}_B, \boldsymbol{\theta}_D)$, where $\mathbf{f}: (\mathbb{R}^{n_B} \times \mathbb{R}^{n_D}) \to \mathbb{R}^{n^*}$:

$$\text{vech}(\boldsymbol{\Omega}) = \mathbf{f}(\boldsymbol{\theta}_B, \boldsymbol{\theta}_D). \qquad [11.6.37]$$

Appendix 11.B shows that the $[n^* \times (n_B + n_D)]$ matrix of derivatives of this function is given by

$$\begin{aligned}\mathbf{J} &= \begin{bmatrix} \dfrac{\partial\, \text{vech}(\boldsymbol{\Omega})}{\partial \boldsymbol{\theta}_B'} & \dfrac{\partial\, \text{vech}(\boldsymbol{\Omega})}{\partial \boldsymbol{\theta}_D'} \end{bmatrix} \\ &= \Big[[-2\mathbf{D}_n^+(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1})\mathbf{S}_B] \qquad \mathbf{D}_n^+[(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})]\mathbf{S}_D\Big],\end{aligned} \qquad [11.6.38]$$

where $\mathbf{D}_n^+$ is the $(n^* \times n^2)$ matrix defined in [11.1.45].

Suppose that the columns of the matrix in [11.6.38] were linearly dependent; that is, suppose there exists a nonzero $[(n_B + n_D) \times 1]$ vector $\boldsymbol{\lambda}$ such that $\mathbf{J}\boldsymbol{\lambda} = \mathbf{0}$. This would mean that if a small multiple of $\boldsymbol{\lambda}$ were added to $(\boldsymbol{\theta}_B', \boldsymbol{\theta}_D')'$, the model would imply the same probability distribution for the data. We would have no basis for distinguishing between these alternative values for $(\boldsymbol{\theta}_B', \boldsymbol{\theta}_D')$, meaning that the model would be unidentified.

Thus, the rank condition for identification of a structural *VAR* requires that the $(n_B + n_D)$ columns of the matrix $\mathbf{J}$ in [11.6.38] be linearly independent.[12] The order condition is that the number of rows of $\mathbf{J}$ $(n^* = n(n+1)/2)$ be at least as great as the number of columns.

To check this condition in practice, the simplest approach is usually to make a guess as to the values of the structural parameters and check $\mathbf{J}$ numerically. Giannini (1992) derived an alternative expression for the rank condition and provided computer software for checking it numerically.

Structural VAR *with Restrictions on* $\boldsymbol{\Pi}$

The supply-and-demand example of [11.6.24] to [11.6.26] did not satisfy the assumptions behind the derivation of [11.6.32], because [11.6.26] imposed the restriction that lagged values of p and q did not belong in the weather equation. Where such restrictions are imposed, it is no longer that case that the *FIML* estimates of $\boldsymbol{\Pi}$ are obtained by *OLS*, and system parameters would have to be estimated as described in Section 11.3. As an alternative, *OLS* estimation of [11.6.24] through [11.6.26] would still give consistent estimates of $\boldsymbol{\Pi}$, and the variance-covariance matrix of the residuals from these regressions would provide a consistent estimate $\hat{\boldsymbol{\Omega}}$. One could still use this estimate in [11.6.32], and the resulting maximization problem would give reasonable estimates of $\mathbf{B}_0$ and $\mathbf{D}$.

Structural VAR*s and Forward-Looking Behavior*

The supply-and-demand example assumed that lagged values of price and quantity did not appear in the equation for weather. The spirit of *VAR*s is that

[12] This condition characterizes *local* identification; it may be that even if a model satisfies both the rank and the order condition, there are two noncontiguous values of $(\boldsymbol{\theta}_B', \boldsymbol{\theta}_D')$ for which the likelihood has the same value for all realizations of the data. See Rothenberg (1971, Theorem 6, p. 585).

such assumptions ought to be tested before being imposed. What should we conclude if, contrary to our prior expectations, the price of oranges turned out to Granger-cause the weather in Florida? It certainly cannot be that the price is a cause of the weather. Instead, such a finding would suggest forward-looking behavior on the part of buyers or sellers of oranges; for example, it may be that if buyers anticipate bad weather in the future, they bid up the price of oranges today. If this should prove to be the case, the identifying assumption in [11.6.24] that demand depends on the weather only through its effect on the current price needs to be reexamined. Proper modeling of forward-looking behavior can provide an alternative way to identify *VAR*s, as explored by Flavin (1981), Hansen and Sargent (1981), and Keating (1990), among others.

Other Approaches to Identifying Structural VAR*s*

Identification was discussed in previous subsections primarily in terms of exclusion restrictions on the matrix of structural coefficients $\mathbf{B}_0$. Blanchard and Diamond (1989, 1990) used a priori assumptions about the signs of structural parameters to identify a range of values of $\mathbf{B}_0$ consistent with the data. Shapiro and Watson (1988) and Blanchard and Quah (1989) used assumptions about long-run multipliers to achieve identification.

A Critique of Structural VAR*s*

Structural *VAR*s have appeal for two different kinds of inquiry. The first potential user is someone who is primarily interested in estimating a structural equation such as the money demand function in [11.6.1]. If a model imposes restrictions on the dynamics of the relationship, it seems good practice to test these restrictions against a more general specification such as [11.6.5] before relying on the restricted model for inference. Furthermore, in order to estimate the dynamic consequences of, say, income on money demand, we have to take into account the fact that, historically, when income goes up, this has typically been associated with future changes in income and interest rates. What time path for these explanatory variables should be assumed in order to assess the consequences for money demand at time $t + s$ of a change in income at time t? A *VAR* offers a framework for posing this question — we use the time path that would historically be predicted for those variables following an unanticipated change in income.

A second potential user is someone who is interested in summarizing the dynamics of a vector $\mathbf{y}_t$ while imposing as few restrictions as possible. Insofar as this summary includes calculation of impulse-response functions, we need some motivation for what the statictics mean. Suppose we find that there is a temporary rise in income following an innovation in money. One is tempted to interpret this finding as suggesting that expansionary monetary policy has a positive but temporary effect on output. However, such an interpretation implicitly assumes that the orthogonalized "money innovation" is the same as the disturbance term in a description of central bank policy. Insofar as impulse-response functions are used to make statements that are structural in nature, it seems reasonable to try to use an orthogonalization that represents our understanding of these relationships as well as possible. This point has been forcefully argued by Cooley and LeRoy (1985), Leamer (1985), Bernanke (1986), and Blanchard (1989), among others.

Even so, it must be recognized that convincing identifying assumptions are hard to come by. For example, the ordering in [11.6.17] is clearly somewhat arbitrary, and the exclusion restrictions are difficult to defend. Indeed, if there were compelling identifying assumptions for such a system, the fierce debates among macroeconomists would have been settled long ago! Simultaneous equations bias is very pervasive in the social sciences, and drawing structural inferences from observed correlations must always proceed with great care. We surely cannot always expect to find credible identifying assumptions to enable us to identify the causal relations among any arbitrary set of n variables on which we have data.

11.7. *Standard Errors for Impulse-Response Functions*

Standard Errors for Nonorthogonalized Impulse-Response Function Based on Analytical Derivatives

Section 11.4 discussed how $\boldsymbol{\Psi}_s$, the matrix of impulse-response coefficients at lag s, would be constructed from knowledge of the autoregressive coefficients. In practice, the autoregressive coefficients are not known with certainty but must be estimated by *OLS* regressions. When the estimated values of the autoregressive coefficients are used to calculate $\boldsymbol{\Psi}_s$, it is useful to report the implied standard errors for the estimates $\hat{\boldsymbol{\Psi}}_s$.[13]

Adopting the notation from Proposition 11.1, let $k = np + 1$ denote the number of coefficients in each equation of the *VAR* and let $\boldsymbol{\pi} \equiv \text{vec}(\boldsymbol{\Pi})$ denote the $(nk \times 1)$ vector of parameters for all the equations; the first k elements of $\boldsymbol{\pi}$ give the constant term and autoregressive coefficients for the first equation, the next k elements of $\boldsymbol{\pi}$ give the parameters for the second equation, and so on. Let $\boldsymbol{\psi}_s \equiv \text{vec}(\boldsymbol{\Psi}_s')$ denote the $(n^2 \times 1)$ vector of moving average coefficients associated with lag s. The first n elements of $\boldsymbol{\psi}_s$ are given by the first row of $\boldsymbol{\Psi}_s$ and identify the response of $y_{1,t+s}$ to $\boldsymbol{\varepsilon}_t$. The next n elements of $\boldsymbol{\psi}_s$ are given by the second row of $\boldsymbol{\Psi}_s$ and identify the response of $y_{2,t+s}$ to $\boldsymbol{\varepsilon}_t$, and so on. Given the values of the autoregressive coefficients in $\boldsymbol{\pi}$, the *VAR* can be simulated to calculate $\boldsymbol{\psi}_s$. Thus, $\boldsymbol{\psi}_s$ could be regarded as a nonlinear function of $\boldsymbol{\pi}$, represented by the function $\boldsymbol{\psi}_s(\boldsymbol{\pi})$, $\boldsymbol{\psi}_s\colon \mathbb{R}^{nk} \to \mathbb{R}^{n^2}$.

The impulse-response coefficients are estimated by replacing $\boldsymbol{\pi}$ with the *OLS* estimates $\hat{\boldsymbol{\pi}}_T$, generating the estimate $\hat{\boldsymbol{\psi}}_{s,T} = \boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}}_T)$. Recall that under the conditions of Proposition 11.1, $\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{L} \mathbf{X}$, where

$$\mathbf{X} \sim N\left(\mathbf{0}, (\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1})\right). \qquad [11.7.1]$$

Standard errors for $\hat{\boldsymbol{\psi}}_s$ can then be calculated by applying Proposition 7.4:

$$\sqrt{T}(\hat{\boldsymbol{\psi}}_{s,T} - \boldsymbol{\psi}_s) \xrightarrow{L} \mathbf{G}_s\mathbf{X},$$

where

$$\underset{(n^2 \times nk)}{\mathbf{G}_s} = \frac{\partial \boldsymbol{\psi}_s(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}'}. \qquad [11.7.2]$$

That is,

$$\sqrt{T}(\hat{\boldsymbol{\psi}}_{s,T} - \boldsymbol{\psi}_s) \xrightarrow{L} N\left(\mathbf{0}, \mathbf{G}_s(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1})\mathbf{G}_s'\right). \qquad [11.7.3]$$

Standard errors for an estimated impulse-response coefficient are given by the

[13]Calculations related to those developed in this section appeared in Baillie (1987), Lütkepohl (1989, 1990), and Giannini (1992). Giannini provided computer software for calculating some of these magnitudes.

square root of the associated diagonal element of $(1/T)\hat{\mathbf{G}}_{s,T}(\hat{\mathbf{\Omega}}_T \otimes \hat{\mathbf{Q}}_T^{-1})\hat{\mathbf{G}}_{s,T}'$, where

$$\hat{\mathbf{G}}_{s,T} = \left.\frac{\partial \boldsymbol{\psi}_s(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}'}\right|_{\boldsymbol{\pi}=\hat{\boldsymbol{\pi}}_T}$$

$$\hat{\mathbf{Q}}_T = (1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t',$$

with $\mathbf{x}_t$ and $\hat{\mathbf{\Omega}}_T$ as defined in Proposition 11.1.

To apply this result, we need an expression for the matrix $\mathbf{G}_s$ in [11.7.2]. Appendix 11.B to this chapter establishes that the sequence $\{\mathbf{G}_s\}_{s=1}^{m}$ can be calculated by iterating on

$$\begin{aligned}\mathbf{G}_s = {} & [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \mathbf{\Psi}_{s-1}' \quad \mathbf{\Psi}_{s-2}' \quad \cdots \quad \mathbf{\Psi}_{s-p}')] + (\mathbf{\Phi}_1 \otimes \mathbf{I}_n)\mathbf{G}_{s-1} \\ & + (\mathbf{\Phi}_2 \otimes \mathbf{I}_n)\mathbf{G}_{s-2} + \cdots + (\mathbf{\Phi}_p \otimes \mathbf{I}_n)\mathbf{G}_{s-p}.\end{aligned} \qquad [11.7.4]$$

Here $\mathbf{0}_{n1}$ denotes an $(n \times 1)$ vector of zeros. The iteration is initialized by setting $\mathbf{G}_0 = \mathbf{G}_{-1} = \cdots = \mathbf{G}_{-p+1} = \mathbf{0}_{n^2,nk}$. It is also understood that $\mathbf{\Psi}_0 = \mathbf{I}_n$ and $\mathbf{\Psi}_s = \mathbf{0}_{nn}$ for $s < 0$. Thus, for example,

$$\mathbf{G}_1 = [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \mathbf{I}_n \quad \mathbf{0}_{nn} \quad \cdots \quad \mathbf{0}_{nn})]$$
$$\mathbf{G}_2 = [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \mathbf{\Psi}_1' \quad \mathbf{I}_n \quad \cdots \quad \mathbf{0}_{nn})] + (\mathbf{\Phi}_1 \otimes \mathbf{I}_n)\mathbf{G}_1.$$

A closed-form solution for [11.7.4] is given by

$$\mathbf{G}_s = \sum_{i=1}^{s} [\mathbf{\Psi}_{i-1} \otimes (\mathbf{0}_{n1} \quad \mathbf{\Psi}_{s-i}' \quad \mathbf{\Psi}_{s-i-1}' \quad \cdots \quad \mathbf{\Psi}_{s-i-p+1}')]. \qquad [11.7.5]$$

Alternative Approaches to Calculating Standard Errors for Nonorthogonalized Impulse-Response Function

The matrix of derivatives $\mathbf{G}_s$ can alternatively be calculated numerically as follows. First we use the *OLS* estimates $\hat{\boldsymbol{\pi}}$ to calculate $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}})$ for $s = 1, 2, \ldots, m$. We then increase the value of the ith element of $\boldsymbol{\pi}$ by some small amount Δ, holding all other elements constant, and evaluate $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}} + \mathbf{e}_i\Delta)$ for $s = 1, 2, \ldots, m$, where $\mathbf{e}_i$ denotes the ith column of $\mathbf{I}_{nk}$. Then the $(n^2 \times 1)$ vector

$$\frac{\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}} + \mathbf{e}_i\Delta) - \boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}})}{\Delta}$$

gives an estimate of the ith column of $\mathbf{G}_s$. By conducting separate evaluations of the sequence $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}} + \mathbf{e}_i\Delta)$ for each $i = 1, 2, \ldots, nk$, all of the columns of $\mathbf{G}_s$ can be filled in.

Monte Carlo methods can also be used to infer the distribution of $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}})$. Here we would randomly generate an $(nk \times 1)$ vector drawn from a $N(\hat{\boldsymbol{\pi}}, (1/T)(\hat{\mathbf{\Omega}} \otimes \hat{\mathbf{Q}}^{-1}))$ distribution. Denote this vector by $\boldsymbol{\pi}^{(1)}$, and calculate $\boldsymbol{\psi}_s(\boldsymbol{\pi}^{(1)})$. Draw a second vector $\boldsymbol{\pi}^{(2)}$ from the same distribution and calculate $\boldsymbol{\psi}_s(\boldsymbol{\pi}^{(2)})$. Repeat this for, say, 10,000 separate simulations. If 9500 of these simulations result in a value of the first element of $\boldsymbol{\psi}_s$ that is between $\underline{\psi}_{s1}$ and $\overline{\psi}_{s1}$, then $(\underline{\psi}_{s1}, \overline{\psi}_{s1})$ can be used as a 95% confidence interval for the first element of $\hat{\boldsymbol{\psi}}_s$.

Runkle (1987) employed a related approach based on *bootstrapping*. The idea behind bootstrapping is to obtain an estimate of the small-sample distribution of $\hat{\boldsymbol{\pi}}$ without assuming that the innovations $\boldsymbol{\varepsilon}_t$ are Gaussian. To implement this procedure, first estimate the *VAR* and save the coefficient estimates $\hat{\boldsymbol{\pi}}$ and the fitted residuals $\{\hat{\boldsymbol{\varepsilon}}_1, \hat{\boldsymbol{\varepsilon}}_2, \ldots, \hat{\boldsymbol{\varepsilon}}_T\}$. Then consider an artificial random variable $\mathbf{u}$ that has probability $(1/T)$ of taking on each of the particular values $\{\hat{\boldsymbol{\varepsilon}}_1, \hat{\boldsymbol{\varepsilon}}_2, \ldots, \hat{\boldsymbol{\varepsilon}}_T\}$. The hope is that the distribution of $\mathbf{u}$ is similar to the distribution of the true population $\boldsymbol{\varepsilon}$'s. Then take a random draw from this distribution (denoted $\mathbf{u}_1^{(1)}$), and use this to construct the first innovation in an artificial sample; that is, set

$$\mathbf{y}_1^{(1)} = \hat{\mathbf{c}} + \hat{\mathbf{\Phi}}_1 \mathbf{y}_0 + \hat{\mathbf{\Phi}}_2 \mathbf{y}_{-1} + \cdots + \hat{\mathbf{\Phi}}_p \mathbf{y}_{-p+1} + \mathbf{u}_1^{(1)},$$

where $\mathbf{y}_0, \mathbf{y}_{-1}, \ldots,$ and $\mathbf{y}_{-p+1}$ denote the presample values of $\mathbf{y}$ that were actually observed in the historical data. Taking a second draw $\mathbf{u}_2^{(1)}$, generate

$$\mathbf{y}_2^{(1)} = \hat{\mathbf{c}} + \hat{\mathbf{\Phi}}_1 \mathbf{y}_1^{(1)} + \hat{\mathbf{\Phi}}_2 \mathbf{y}_0 + \cdots + \hat{\mathbf{\Phi}}_p \mathbf{y}_{-p+2} + \mathbf{u}_2^{(1)}.$$

Note that this second draw is with replacement; that is, there is a $(1/T)$ chance that $\mathbf{u}_1^{(1)}$ is exactly the same as $\mathbf{u}_2^{(1)}$. Proceeding in this fashion, a full sample $\{\mathbf{y}_1^{(1)}, \mathbf{y}_2^{(1)}, \ldots, \mathbf{y}_T^{(1)}\}$ can be generated. A *VAR* can be fitted by *OLS* to these simulated data (again taking presample values of $\mathbf{y}$ as their historical values), producing an estimate $\hat{\boldsymbol{\pi}}^{(1)}$. From this estimate, the magnitude $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}}^{(1)})$ can be calculated. Next, generate a second set of T draws from the distribution of $\mathbf{u}$, denoted $\{\mathbf{u}_1^{(2)}, \mathbf{u}_2^{(2)}, \ldots, \mathbf{u}_T^{(2)}\}$, fit $\hat{\boldsymbol{\pi}}^{(2)}$ to these data by *OLS*, and calculate $\boldsymbol{\psi}_s(\hat{\boldsymbol{\pi}}^{(2)})$. A series of 10,000 such simulations could be undertaken, and a 95% confidence interval for $\psi_{s1}(\hat{\boldsymbol{\pi}})$ is then inferred from the range that includes 95% of the values for $\psi_{s1}(\hat{\boldsymbol{\pi}}^{(i)})$.

Standard Errors for Parameters of a Structural VAR

Recall from Proposition 11.2 and equation [11.1.48] that if the innovations are Gaussian,

$$\sqrt{T}[\mathrm{vech}(\hat{\mathbf{\Omega}}_T) - \mathrm{vech}(\mathbf{\Omega})] \xrightarrow{L} N\left(\mathbf{0}, 2\mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)'\right).$$

The estimates of the parameters of a structural *VAR* ($\hat{\mathbf{B}}_0$ and $\hat{\mathbf{D}}$) are determined as implicit functions of $\hat{\mathbf{\Omega}}$ from

$$\hat{\mathbf{\Omega}} = \hat{\mathbf{B}}_0^{-1}\hat{\mathbf{D}}(\hat{\mathbf{B}}_0^{-1})'. \qquad [11.7.6]$$

As in equation [11.6.34], the unknown elements of $\mathbf{B}_0$ are summarized by an $(n_B \times 1)$ vector $\boldsymbol{\theta}_B$ with $\mathrm{vec}(\mathbf{B}_0) = \mathbf{S}_B\boldsymbol{\theta}_B + \mathbf{s}_B$. Similarly, as in [11.6.35], it is assumed that $\mathrm{vec}(\mathbf{D}) = \mathbf{S}_D\boldsymbol{\theta}_D + \mathbf{s}_D$ for $\boldsymbol{\theta}_D$ an $(n_D \times 1)$ vector. It then follows from Proposition 7.4 that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{B,T} - \boldsymbol{\theta}_B) \xrightarrow{L} N\left(\mathbf{0}, 2\mathbf{G}_B\mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)'\mathbf{G}_B'\right) \qquad [11.7.7]$$

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{D,T} - \boldsymbol{\theta}_D) \xrightarrow{L} N\left(\mathbf{0}, 2\mathbf{G}_D\mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)'\mathbf{G}_D'\right), \qquad [11.7.8]$$

where

$$\underset{(n_B \times n^*)}{\mathbf{G}_B} = \frac{\partial\boldsymbol{\theta}_B}{\partial[\mathrm{vech}(\mathbf{\Omega})]'} \qquad [11.7.9]$$

$$\underset{(n_D \times n^*)}{\mathbf{G}_D} = \frac{\partial\boldsymbol{\theta}_D}{\partial[\mathrm{vech}(\mathbf{\Omega})]'} \qquad [11.7.10]$$

and $n^* \equiv n(n + 1)/2$.

Equation [11.6.38] gave an expression for the $[n^* \times (n_B + n_D)]$ matrix:

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial\, \mathrm{vech}(\mathbf{\Omega})}{\partial\boldsymbol{\theta}_B'} & \dfrac{\partial\, \mathrm{vech}(\mathbf{\Omega})}{\partial\boldsymbol{\theta}_D'} \end{bmatrix}.$$

We noted there that if the model is to be identified, the columns of this matrix must be linearly independent. In the just-identified case, $n^* = (n_B + n_D)$ and $\mathbf{J}^{-1}$

exists, from which

$$\begin{bmatrix} \mathbf{G}_B \\ \mathbf{G}_D \end{bmatrix} = \mathbf{J}^{-1}. \qquad [11.7.11]$$

Standard Errors for Orthogonalized Impulse-Response Functions

Section 11.6 described calculation of the following $(n \times n)$ matrix:

$$\mathbf{H}_s = \mathbf{\Psi}_s \mathbf{B}_0^{-1}. \qquad [11.7.12]$$

The row i, column j element of this matrix measures the effect of the jth structural disturbance (u_{jt}) on the ith variable in the system $(y_{i,t+s})$ after a lag of s periods. Collect these magnitudes in an $(n^2 \times 1)$ vector $\mathbf{h}_s \equiv \text{vec}(\mathbf{H}_s')$. Thus, the first n elements of $\mathbf{h}_s$ give the effect of $\mathbf{u}_t$ on $y_{1,t+s}$, the next n elements give the effect of $\mathbf{u}_t$ on $y_{2,t+s}$, and so on.

Since $\hat{\mathbf{\Psi}}_s$ is a function of $\hat{\boldsymbol{\pi}}$ and since $\hat{\mathbf{B}}_0$ is a function of vech$(\hat{\mathbf{\Omega}})$, the distributions of both the autoregressive coefficients and the variances affect the asymptotic distribution of $\hat{\mathbf{h}}_s$. It follows from Proposition 11.2 that with Gaussian innovations,

$$\begin{aligned} &\sqrt{T}(\hat{\mathbf{h}}_{s,T} - \mathbf{h}_s) \\ &\xrightarrow{L} N\left(\mathbf{0}, [\mathbf{\Xi}_\pi \quad \mathbf{\Xi}_\sigma] \begin{bmatrix} \mathbf{\Omega} \otimes \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)' \end{bmatrix} \begin{bmatrix} \mathbf{\Xi}_\pi' \\ \mathbf{\Xi}_\sigma' \end{bmatrix}\right) \qquad [11.7.13] \\ &\sim N\left(\mathbf{0}, [\mathbf{\Xi}_\pi(\mathbf{\Omega} \otimes \mathbf{Q}^{-1})\mathbf{\Xi}_\pi' + 2\mathbf{\Xi}_\sigma \mathbf{D}_n^+(\mathbf{\Omega} \otimes \mathbf{\Omega})(\mathbf{D}_n^+)'\mathbf{\Xi}_\sigma']\right), \end{aligned}$$

where Appendix 11.B demonstrates that

$$\mathbf{\Xi}_\pi = \partial \mathbf{h}_s / \partial \boldsymbol{\pi}' = [\mathbf{I}_n \otimes (\mathbf{B}_0')^{-1}]\mathbf{G}_s \qquad [11.7.14]$$

$$\mathbf{\Xi}_\sigma = \frac{\partial \mathbf{h}_s}{\partial[\text{vech}(\mathbf{\Omega})]'} = -[\mathbf{H}_s \otimes (\mathbf{B}_0')^{-1}]\mathbf{S}_{B'}\mathbf{G}_B. \qquad [11.7.15]$$

Here $\mathbf{G}_s$ is the matrix given in [11.7.5], $\mathbf{G}_B$ is the matrix given in [11.7.11], and $\mathbf{S}_{B'}$ is an $(n^2 \times n_B)$ matrix that takes the elements of $\boldsymbol{\theta}_B$ and puts them in the corresponding position to construct vec$(\mathbf{B}_0')$:

$$\text{vec}(\mathbf{B}_0') = \mathbf{S}_{B'}\boldsymbol{\theta}_B + \mathbf{s}_{B'}.$$

For the supply-and-demand examples of [11.6.24] to [11.6.26],

$$\mathbf{S}_{B'} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Practical Experience with Standard Errors

In practice, the standard errors for dynamic inferences based on *VAR*s often turn out to be disappointingly large (see Runkle, 1987, and Lütkepohl, 1990).

Although a VAR imposes few restrictions on the dynamics, the cost of this generality is that the inferences drawn are not too precise. To gain more precision, it is necessary to impose further restrictions. One approach is to fit the multivariate dynamics using a restricted model with far fewer parameters, provided that the data allow us to accept the restrictions. A second approach is to place greater reliance on prior expectations about the system dynamics. This second approach is explored in the next chapter.

APPENDIX 11.A. *Proofs of Chapter 11 Propositions*

■ **Proof of Proposition 11.1.** The condition on the roots of [11.1.35] ensures that the $MA(\infty)$ representation is absolutely summable. Thus $\mathbf{y}_t$ is ergodic for first moments, from Propositions 10.2(b) and 10.5(a), and is also ergodic for second moments, from Proposition 10.2(d). This establishes result 11.1(a).

The proofs of results (b) and (c) are virtually identical to those for a single OLS regression with stochastic regressors (results [8.2.5] and [8.2.12]).

To verify result (d), notice that

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_{i,T} - \boldsymbol{\pi}_i) = \left[(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_{it}\right]$$

and so

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) = \begin{bmatrix} \mathbf{Q}_T^{-1}(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_{1t} \\ \mathbf{Q}_T^{-1}(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_{2t} \\ \vdots \\ \mathbf{Q}_T^{-1}(1/\sqrt{T})\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_{nt} \end{bmatrix}, \qquad [11.A.1]$$

where

$$\mathbf{Q}_T \equiv \left[(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right].$$

Define $\boldsymbol{\xi}_t$ to be the following $(nk \times 1)$ vector:

$$\boldsymbol{\xi}_t \equiv \begin{bmatrix} \mathbf{x}_t\varepsilon_{1t} \\ \mathbf{x}_t\varepsilon_{2t} \\ \vdots \\ \mathbf{x}_t\varepsilon_{nt} \end{bmatrix}.$$

Notice that $\boldsymbol{\xi}_t$ is a martingale difference sequence with finite fourth moments and variance

$$\begin{aligned} E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') &= \begin{bmatrix} E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{1t}^2) & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{1t}\varepsilon_{2t}) & \cdots & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{1t}\varepsilon_{nt}) \\ E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{2t}\varepsilon_{1t}) & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{2t}^2) & \cdots & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{2t}\varepsilon_{nt}) \\ \vdots & \vdots & \cdots & \vdots \\ E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{nt}\varepsilon_{1t}) & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{nt}\varepsilon_{2t}) & \cdots & E(\mathbf{x}_t\mathbf{x}_t')\cdot E(\varepsilon_{nt}^2) \end{bmatrix} \\ &= \begin{bmatrix} E(\varepsilon_{1t}^2) & E(\varepsilon_{1t}\varepsilon_{2t}) & \cdots & E(\varepsilon_{1t}\varepsilon_{nt}) \\ E(\varepsilon_{2t}\varepsilon_{1t}) & E(\varepsilon_{2t}^2) & \cdots & E(\varepsilon_{2t}\varepsilon_{nt}) \\ \vdots & \vdots & \cdots & \vdots \\ E(\varepsilon_{nt}\varepsilon_{1t}) & E(\varepsilon_{nt}\varepsilon_{2t}) & \cdots & E(\varepsilon_{nt}^2) \end{bmatrix} \otimes E(\mathbf{x}_t\mathbf{x}_t') \\ &= \boldsymbol{\Omega} \otimes \mathbf{Q}. \end{aligned}$$

It can further be shown that

$$(1/T)\sum_{t=1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_t' \xrightarrow{p} \boldsymbol{\Omega} \otimes \mathbf{Q} \qquad [11.A.2]$$

(see Exercise 11.1). It follows from Proposition 7.9 that

$$(1/\sqrt{T}) \sum_{t=1}^{T} \xi_t \xrightarrow{L} N\Big(\mathbf{0}, (\mathbf{\Omega} \otimes \mathbf{Q})\Big). \qquad [11.A.3]$$

Now, expression [11.A.1] can be written

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) = \begin{bmatrix} \mathbf{Q}_T^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_T^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Q}_T^{-1} \end{bmatrix} \begin{bmatrix} (1/\sqrt{T}) \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_{1t} \\ (1/\sqrt{T}) \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_{2t} \\ \vdots \\ (1/\sqrt{T}) \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_{nt} \end{bmatrix}$$

$$= (\mathbf{I}_n \otimes \mathbf{Q}_T^{-1})(1/\sqrt{T}) \sum_{t=1}^{T} \xi_t.$$

But result (a) implies that $\mathbf{Q}_T^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$. Thus,

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{p} (\mathbf{I}_n \otimes \mathbf{Q}^{-1})(1/\sqrt{T}) \sum_{t=1}^{T} \xi_t. \qquad [11.A.4]$$

But from [11.A.3], this has a distribution that is Gaussian with mean $\mathbf{0}$ and variance

$$(\mathbf{I}_n \otimes \mathbf{Q}^{-1})(\mathbf{\Omega} \otimes \mathbf{Q})(\mathbf{I}_n \otimes \mathbf{Q}^{-1}) = (\mathbf{I}_n \mathbf{\Omega} \mathbf{I}_n) \otimes (\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}) = \mathbf{\Omega} \otimes \mathbf{Q}^{-1},$$

as claimed. ■

■ **Proof of Proposition 11.2.** Define $\hat{\mathbf{\Omega}}_T^* \equiv (1/T)\Sigma_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$ to be the estimate of $\mathbf{\Omega}$ based on the true residuals. We first note that $\hat{\mathbf{\Omega}}_T$ has the same asymptotic distribution as $\hat{\mathbf{\Omega}}_T^*$. To see this, observe that

$$\begin{aligned} \hat{\mathbf{\Omega}}_T^* &= (1/T) \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t)(\mathbf{y}_t - \mathbf{\Pi}'\mathbf{x}_t)' \\ &= (1/T) \sum_{t=1}^{T} [\mathbf{y}_t - \hat{\mathbf{\Pi}}_T'\mathbf{x}_t + (\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})'\mathbf{x}_t][\mathbf{y}_t - \hat{\mathbf{\Pi}}_T'\mathbf{x}_t + (\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})'\mathbf{x}_t]' \\ &= (1/T) \sum_{t=1}^{T} (\mathbf{y}_t - \hat{\mathbf{\Pi}}_T'\mathbf{x}_t)(\mathbf{y}_t - \hat{\mathbf{\Pi}}_T'\mathbf{x}_t)' \qquad [11.A.5] \\ &\quad + (\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})'(1/T) \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'(\hat{\mathbf{\Pi}}_T - \mathbf{\Pi}) \\ &= \hat{\mathbf{\Omega}}_T + (\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})'(1/T) \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'(\hat{\mathbf{\Pi}}_T - \mathbf{\Pi}), \end{aligned}$$

where cross-product terms were dropped in the third equality on the right in the light of the *OLS* orthogonality condition $(1/T)\Sigma_{t=1}^T(\mathbf{y}_t - \hat{\mathbf{\Pi}}_T'\mathbf{x}_t)\mathbf{x}_t' = \mathbf{0}$. Equation [11.A.5] implies that

$$\sqrt{T}(\hat{\mathbf{\Omega}}_T^* - \hat{\mathbf{\Omega}}_T) = (\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})'(1/T) \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'[\sqrt{T}(\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})].$$

But Proposition 11.1 established that $(\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})' \xrightarrow{p} \mathbf{0}$, $(1/T)\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} \mathbf{Q}$, and $\sqrt{T}(\hat{\mathbf{\Pi}}_T - \mathbf{\Pi})$ converges in distribution. Thus, from Proposition 7.3, $\sqrt{T}(\hat{\mathbf{\Omega}}_T^* - \hat{\mathbf{\Omega}}_T) \xrightarrow{p} \mathbf{0}$ meaning that $\sqrt{T}(\hat{\mathbf{\Omega}}_T^* - \mathbf{\Omega}) \xrightarrow{p} \sqrt{T}(\hat{\mathbf{\Omega}}_T - \mathbf{\Omega})$.

Recalling [11.A.4],

$$\begin{bmatrix} \sqrt{T}[\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}] \\ \sqrt{T}[\text{vech}(\hat{\mathbf{\Omega}}_T) - \text{vech}(\mathbf{\Omega})] \end{bmatrix} \xrightarrow{p} \begin{bmatrix} (\mathbf{I}_n \otimes \mathbf{Q}^{-1})(1/\sqrt{T}) \sum_{t=1}^{T} \xi_t \\ (1/\sqrt{T}) \sum_{t=1}^{T} \boldsymbol{\lambda}_t \end{bmatrix}, \qquad [11.A.6]$$

where $\xi_t = \varepsilon_t \otimes \mathbf{x}_t$ and

$$\lambda_t \equiv \operatorname{vech}\begin{bmatrix} \varepsilon_{1t}^2 - \sigma_{11} & \varepsilon_{1t}\varepsilon_{2t} - \sigma_{12} & \cdots & \varepsilon_{1t}\varepsilon_{nt} - \sigma_{1n} \\ \varepsilon_{2t}\varepsilon_{1t} - \sigma_{21} & \varepsilon_{2t}^2 - \sigma_{22} & \cdots & \varepsilon_{2t}\varepsilon_{nt} - \sigma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \varepsilon_{nt}\varepsilon_{1t} - \sigma_{n1} & \varepsilon_{nt}\varepsilon_{2t} - \sigma_{n2} & \cdots & \varepsilon_{nt}^2 - \sigma_{nn} \end{bmatrix}.$$

It is straightforward to show that $(\xi_t', \lambda_t')'$ is a martingale difference sequence that satisfies the conditions of Proposition 7.9, from which

$$\begin{bmatrix} (1/\sqrt{T}) \sum_{t=1}^{T} \xi_t \\ (1/\sqrt{T}) \sum_{t=1}^{T} \lambda_t \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right), \qquad [11.A.7]$$

where

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} E(\xi_t\xi_t') & E(\xi_t\lambda_t') \\ E(\lambda_t\xi_t') & E(\lambda_t\lambda_t') \end{bmatrix}.$$

Recall from the proof of Proposition 11.1 that

$$\Sigma_{11} = E(\xi_t\xi_t') = \Omega \otimes \mathbf{Q}.$$

A typical element of Σ_{12} is of the form

$$E(\mathbf{x}_t\varepsilon_{lt})(\varepsilon_{it}\varepsilon_{jt} - \sigma_{ij}) = E(\mathbf{x}_t)\cdot E(\varepsilon_{lt}\varepsilon_{it}\varepsilon_{jt}) - \sigma_{ij}\cdot E(\mathbf{x}_t)\cdot E(\varepsilon_{lt}),$$

which equals zero for all i, j, and l. Hence, [11.A.7] becomes

$$\begin{bmatrix} (1/\sqrt{T}) \sum_{t=1}^{T} \xi_t \\ (1/\sqrt{T}) \sum_{t=1}^{T} \lambda_t \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega \otimes \mathbf{Q} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right),$$

and so, from [11.A.6],

$$\begin{bmatrix} \sqrt{T}[\hat{\pi}_T - \pi] \\ \sqrt{T}[\operatorname{vech}(\hat{\Omega}_T) - \operatorname{vech}(\Omega)] \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega \otimes \mathbf{Q}^{-1} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right).$$

Hence, Proposition 11.2 will be established if we can show that $E(\lambda_t\lambda_t')$ is given by the matrix Σ_{22} described in the proposition; that is, we must show that

$$E(\varepsilon_{it}\varepsilon_{jt} - \sigma_{ij})(\varepsilon_{lt}\varepsilon_{mt} - \sigma_{lm}) = \sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl} \qquad [11.A.8]$$

for all i, j, l, and m.

To derive [11.A.8], let $\Omega = \mathbf{P}\mathbf{P}'$ denote the Cholesky decomposition of Ω, and define

$$\mathbf{v}_t \equiv \mathbf{P}^{-1}\varepsilon_t. \qquad [11.A.9]$$

Then $E(\mathbf{v}_t\mathbf{v}_t') = \mathbf{P}^{-1}\Omega(\mathbf{P}^{-1})' = \mathbf{I}_n$. Thus, v_{it} is Gaussian with zero mean, unit variance, and fourth moment given by $E(v_{it}^4) = 3$. Moreover, v_{it} is independent of v_{jt} for $i \neq j$.

Equation [11.A.9] implies

$$\varepsilon_t = \mathbf{P}\mathbf{v}_t. \qquad [11.A.10]$$

Let p_{ij} denote the row i, column j element of $\mathbf{P}$. Then the ith row of [11.A.10] states that

$$\varepsilon_{it} = p_{i1}v_{1t} + p_{i2}v_{2t} + \cdots + p_{in}v_{nt} \qquad [11.A.11]$$

and

$$\varepsilon_{it}\varepsilon_{jt} = (p_{i1}v_{1t} + p_{i2}v_{2t} + \cdots + p_{in}v_{nt}) \times (p_{j1}v_{1t} + p_{j2}v_{2t} + \cdots + p_{jn}v_{nt}). \qquad [11.A.12]$$

Second moments of ε_t can be found by taking expectations of [11.A.12], recalling that $E(v_{it}v_{jt}) = 1$ if $i = j$ and is zero otherwise:

$$E(\varepsilon_{it}\varepsilon_{jt}) = p_{i1}p_{j1} + p_{i2}p_{j2} + \cdots + p_{in}p_{jn}. \qquad [11.A.13]$$

Similarly, fourth moments can be found from

$$
\begin{aligned}
E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) &= E[(p_{i1}v_{1t} + p_{i2}v_{2t} + \cdots + p_{in}v_{nt})(p_{j1}v_{1t} + p_{j2}v_{2t} + \cdots + p_{jn}v_{nt}) \\
&\quad \times (p_{l1}v_{1t} + p_{l2}v_{2t} + \cdots + p_{ln}v_{nt})(p_{m1}v_{1t} + p_{m2}v_{2t} + \cdots + p_{mn}v_{nt})] \\
&= [3(p_{i1}p_{j1}p_{l1}p_{m1} + p_{i2}p_{j2}p_{l2}p_{m2} + \cdots + p_{in}p_{jn}p_{ln}p_{mn})] \\
&\quad + [(p_{i1}p_{j1})(p_{l2}p_{m2} + p_{l3}p_{m3} + \cdots + p_{ln}p_{mn}) \\
&\qquad + (p_{i2}p_{j2})(p_{l1}p_{m1} + p_{l3}p_{m3} + \cdots + p_{ln}p_{mn}) + \cdots \\
&\qquad + (p_{in}p_{jn})(p_{l1}p_{m1} + p_{l2}p_{m2} + \cdots + p_{l,n-1}p_{m,n-1})] \\
&\quad + [(p_{i1}p_{l1})(p_{j2}p_{m2} + p_{j3}p_{m3} + \cdots + p_{jn}p_{mn}) \\
&\qquad + (p_{i2}p_{l2})(p_{j1}p_{m1} + p_{j3}p_{m3} + \cdots + p_{jn}p_{mn}) + \cdots \\
&\qquad + (p_{in}p_{ln})(p_{j1}p_{m1} + p_{j2}p_{m2} + \cdots + p_{j,n-1}p_{m,n-1})] \\
&\quad + [(p_{i1}p_{m1})(p_{j2}p_{l2} + p_{j3}p_{l3} + \cdots + p_{jn}p_{ln}) \\
&\qquad + (p_{i2}p_{m2})(p_{j1}p_{l1} + p_{j3}p_{l3} + \cdots + p_{jn}p_{ln}) + \cdots \\
&\qquad + (p_{in}p_{mn})(p_{j1}p_{l1} + p_{j2}p_{l2} + \cdots + p_{j,n-1}p_{l,n-1})] \\
&= [(p_{i1}p_{j1} + p_{i2}p_{j2} + \cdots + p_{in}p_{jn})(p_{l1}p_{m1} + p_{l2}p_{m2} + \cdots + p_{ln}p_{mn})] \\
&\quad + [(p_{i1}p_{l1} + p_{i2}p_{l2} + \cdots + p_{in}p_{ln})(p_{j1}p_{m1} + p_{j2}p_{m2} + \cdots + p_{jn}p_{mn})] \\
&\quad + [(p_{i1}p_{m1} + p_{i2}p_{m2} + \cdots + p_{in}p_{mn})(p_{j1}p_{l1} + p_{j2}p_{l2} + \cdots + p_{jn}p_{ln})] \\
&= \sigma_{ij}\sigma_{lm} + \sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl},
\end{aligned}
\qquad [11.A.14]
$$

where the last line follows from [11.A.13]. Then

$$E[(\varepsilon_{it}\varepsilon_{jt} - \sigma_{ij})(\varepsilon_{lt}\varepsilon_{mt} - \sigma_{lm})] = E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) - \sigma_{ij}\sigma_{lm} = \sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl},$$

as claimed in [11.A.8]. ■

■ **Proof of Proposition 11.3.** First suppose that y fails to Granger-cause x, so that the process can be written as in [11.2.4]. Define v_{2t} to be the residual from a projection of ε_{2t} on ε_{1t}, with b_0 defined to be the projection coefficient:

$$v_{2t} \equiv \varepsilon_{2t} - b_0\varepsilon_{1t}.$$

Thus, v_{2t} and ε_{1t} are uncorrelated, and, recalling that $\boldsymbol{\varepsilon}_t$ is white noise, v_{2t} must be uncorrelated with $\varepsilon_{1\tau}$ for all $t \neq \tau$ as well. From the first row of [11.2.4], this means that v_{2t} and x_τ are uncorrelated for all t and τ. With this definition of v_{2t}, the second row of [11.2.4] can be written as

$$y_t = \mu_2 + \psi_{21}(L)\varepsilon_{1t} + \psi_{22}(L)[v_{2t} + b_0\varepsilon_{1t}]. \qquad [11.A.15]$$

Furthermore, from the first row of [11.2.4],

$$\varepsilon_{1t} = [\psi_{11}(L)]^{-1}(x_t - \mu_1). \qquad [11.A.16]$$

Substituting [11.A.16] into [11.A.15] gives

$$y_t = c + b(L)x_t + \eta_t, \qquad [11.A.17]$$

where we have defined $b(L) \equiv \{[\psi_{21}(L) + b_0\psi_{22}(L)][\psi_{11}(L)]^{-1}\}$, $c \equiv \mu_2 - b(1)\mu_1$, and $\eta_t \equiv \psi_{22}(L)v_{2t}$. But η_t, being constructed from v_{2t}, is uncorrelated with x_τ for all τ. Furthermore, only current and lagged values of x, as summarized by the operator $b(L)$, appear in equation [11.A.17]. We have thus shown that if [11.2.4] holds, then $d_j = 0$ for all j in [11.2.5].

To prove the converse, suppose that $d_j = 0$ for all j in [11.2.5]. Let

$$x_t = \mu_1 + \psi_{11}(L)\varepsilon_{1t} \qquad [11.A.18]$$

denote the univariate Wold representation for x_t; thus, $\psi_{11}^{(0)} = 1$. We will be using notation consistent with the form of [11.2.4] in anticipation of the final answer that will be derived; for now, the reader should view [11.A.18] as a new definition of $\psi_{11}(L)$ in terms of the univariate Wold representation for x. There also exists a univariate Wold representation for the error term in [11.2.5], denoted

$$\eta_t = \psi_{22}(L)\nu_{2t}, \qquad [11.A.19]$$

with $\psi_{22}^{(0)} = 1$. Notice that η_t as defined in [11.2.5] is uncorrected with x_s for all t and s. It follows that ν_{2t} is uncorrelated with x_τ or $\varepsilon_{1\tau}$ for all t and τ.

Substituting [11.A.18] and [11.A.19] into [11.2.5],

$$y_t = c + b(1)\mu_1 + b(L)\psi_{11}(L)\varepsilon_{1t} + \psi_{22}(L)\nu_{2t}. \qquad [11.A.20]$$

Define

$$\varepsilon_{2t} \equiv \nu_{2t} + b_0\varepsilon_{1t} \qquad [11.A.21]$$

for b_0 the coefficient on L^0 of $b(L)$ and

$$\mu_2 \equiv c + b(1)\mu_1. \qquad [11.A.22]$$

Observe that $(\varepsilon_{1t}, \varepsilon_{2t})'$ is vector white noise. Substituting [11.A.21] and [11.A.22] into [11.A.20] produces

$$y_t = \mu_2 + [b(L)\psi_{11}(L) - b_0\psi_{22}(L)]\varepsilon_{1t} + \psi_{22}(L)\varepsilon_{2t}. \qquad [11.A.23]$$

Finally, define

$$\psi_{21}(L) \equiv [b(L)\psi_{11}(L) - b_0\psi_{22}(L)],$$

noting that $\psi_{21}^{(0)} = 0$. Then, substituting this into [11.A.23] produces

$$y_t = \mu_2 + \psi_{21}(L)\varepsilon_{1t} + \psi_{22}(L)\varepsilon_{2t}.$$

This combined with [11.A.18] completes the demonstration that [11.2.5] implies [11.2.4]. ■

APPENDIX 11.B. *Calculation of Analytic Derivatives*

This appendix calculates the derivatives reported in Sections 11.6 and 11.7.

■ **Derivation of [11.6.38].** Let the scalar ξ represent some particular element of $\boldsymbol{\theta}_B$ or $\boldsymbol{\theta}_D$, and let $\partial\boldsymbol{\Omega}/\partial\xi$ denote the $(n^2 \times n^2)$ matrix that results when each element of $\boldsymbol{\Omega}$ is differentiated with respect to ξ. Thus, differentiating [11.6.33] with respect to ξ results in

$$\partial\boldsymbol{\Omega}/\partial\xi = (\partial\mathbf{B}_0^{-1}/\partial\xi)\mathbf{D}(\mathbf{B}_0^{-1})' + \mathbf{B}_0^{-1}(\partial\mathbf{D}/\partial\xi)(\mathbf{B}_0^{-1})' + (\mathbf{B}_0^{-1})\mathbf{D}[\partial(\mathbf{B}_0^{-1})'/\partial\xi]. \qquad [11.B.1]$$

Define

$$\boldsymbol{\chi} \equiv (\partial\mathbf{B}_0^{-1}/\partial\xi)\mathbf{D}(\mathbf{B}_0^{-1})' \qquad [11.B.2]$$

and notice that

$$\boldsymbol{\chi}' = (\mathbf{B}_0^{-1})\mathbf{D}[\partial(\mathbf{B}_0^{-1})'/\partial\xi],$$

since $\mathbf{D}$ is a variance-covariance matrix and must therefore be symmetric. Thus, [11.B.1] can be written

$$\partial\boldsymbol{\Omega}/\partial\xi = \boldsymbol{\chi} + \mathbf{B}_0^{-1}(\partial\mathbf{D}/\partial\xi)(\mathbf{B}_0^{-1})' + \boldsymbol{\chi}'. \qquad [11.B.3]$$

Recall from Proposition 10.4 that

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\cdot\text{vec}(\mathbf{B}). \qquad [11.B.4]$$

Thus, if the vec operator is applied to [11.B.3], the result is

$$\frac{\partial\,\text{vec}(\boldsymbol{\Omega})}{\partial\xi} = \text{vec}(\boldsymbol{\chi} + \boldsymbol{\chi}') + [(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})]\,\text{vec}(\partial\mathbf{D}/\partial\xi). \qquad [11.B.5]$$

Let $\mathbf{D}_n$ denote the $(n^2 \times n^*)$ duplication matrix introduced in [11.1.43]. Notice that for any $(n \times n)$ matrix $\boldsymbol{\chi}$, the elements of $\mathbf{D}_n' \operatorname{vec}(\boldsymbol{\chi})$ are of the form χ_{ii} for diagonal elements of $\boldsymbol{\chi}$ and of the form $(\chi_{ij} + \chi_{ji})$ for off-diagonal elements. Hence, $\mathbf{D}_n' \operatorname{vec}(\boldsymbol{\chi}) = \mathbf{D}_n' \operatorname{vec}(\boldsymbol{\chi}')$. If [11.B.5] is premultiplied by $\mathbf{D}_n^+ = (\mathbf{D}_n'\mathbf{D}_n)^{-1}\mathbf{D}_n'$, the result is thus

$$\frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \xi} = 2\mathbf{D}_n^+ \operatorname{vec}(\boldsymbol{\chi}) + \mathbf{D}_n^+[(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})] \operatorname{vec}(\partial \mathbf{D}/\partial \xi), \qquad [11.B.6]$$

since from [11.1.46] $\mathbf{D}_n^+ \operatorname{vec}(\boldsymbol{\Omega}) = \operatorname{vech}(\boldsymbol{\Omega})$.

Differentiating the identity $\mathbf{B}_0^{-1}\mathbf{B}_0 = \mathbf{I}_n$ with respect to ξ produces

$$(\partial \mathbf{B}_0^{-1}/\partial \xi)\mathbf{B}_0 + \mathbf{B}_0^{-1}(\partial \mathbf{B}_0/\partial \xi) = \mathbf{0}_{nn}$$

or

$$\partial \mathbf{B}_0^{-1}/\partial \xi = -\mathbf{B}_0^{-1}(\partial \mathbf{B}_0/\partial \xi)\mathbf{B}_0^{-1}. \qquad [11.B.7]$$

Thus, [11.B.2] can be written

$$\boldsymbol{\chi} = -\mathbf{B}_0^{-1}(\partial \mathbf{B}_0/\partial \xi)\mathbf{B}_0^{-1}\mathbf{D}(\mathbf{B}_0^{-1})' = -\mathbf{B}_0^{-1}(\partial \mathbf{B}_0/\partial \xi)\boldsymbol{\Omega}.$$

Applying the vec operator as in [11.B.4] results in

$$\operatorname{vec}(\boldsymbol{\chi}) = -(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1}) \frac{\partial \operatorname{vec}(\mathbf{B}_0)}{\partial \xi}.$$

Substituting this expression into [11.B.6] gives

$$\begin{aligned}\frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \xi} &= -2\mathbf{D}_n^+(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1}) \frac{\partial \operatorname{vec}(\mathbf{B}_0)}{\partial \xi} + \mathbf{D}_n^+[(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})] \frac{\partial \operatorname{vec}(\mathbf{D})}{\partial \xi} \\ &= -2\mathbf{D}_n^+(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1})\mathbf{S}_B \frac{\partial \boldsymbol{\theta}_B}{\partial \xi} + \mathbf{D}_n^+[(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})]\mathbf{S}_D \frac{\partial \boldsymbol{\theta}_D}{\partial \xi}.\end{aligned} \qquad [11.B.8]$$

Expression [11.B.8] is an $(n^* \times 1)$ vector that gives the effect of a change in some element of $\boldsymbol{\theta}_B$ or $\boldsymbol{\theta}_D$ on each of the n^* elements of $\operatorname{vech}(\boldsymbol{\Omega})$. If ξ corresponds to the first element of $\boldsymbol{\theta}_B$, then $\partial \boldsymbol{\theta}_B/\partial \xi = \mathbf{e}_1$, the first column of the $(n_B \times n_B)$ identity matrix, and $\partial \boldsymbol{\theta}_D/\partial \xi = \mathbf{0}$. If ξ corresponds to the second element of $\boldsymbol{\theta}_B$, then $\partial \boldsymbol{\theta}_B/\partial \xi = \mathbf{e}_2$. If we stack the vectors in [11.B.8] associated with $\xi = \theta_{B,1}, \xi = \theta_{B,2}, \ldots, \xi = \theta_{B,n_B}$ side by side, the result is

$$\begin{aligned}&\left[\frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \theta_{B,1}} \quad \frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \theta_{B,2}} \quad \cdots \quad \frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \theta_{B,n_B}}\right] \\ &\qquad = [-2\mathbf{D}_n^+(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1})\mathbf{S}_B][\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_{n_B}].\end{aligned} \qquad [11.B.9]$$

That is,

$$\frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \boldsymbol{\theta}_B'} = [-2\mathbf{D}_n^+(\boldsymbol{\Omega} \otimes \mathbf{B}_0^{-1})\mathbf{S}_B]. \qquad [11.B.10]$$

Similarly, letting the scalar ξ in [11.B.8] correspond to each of the elements of $\boldsymbol{\theta}_D$ in succession and stacking the resulting columns horizontally results in

$$\frac{\partial \operatorname{vech}(\boldsymbol{\Omega})}{\partial \boldsymbol{\theta}_D'} = \mathbf{D}_n^+[(\mathbf{B}_0^{-1}) \otimes (\mathbf{B}_0^{-1})]\mathbf{S}_D. \qquad [11.B.11]$$

Equation [11.6.38] then follows immediately from [11.B.10] and [11.B.11]. ■

■ **Derivation of [11.7.4].** Recall from equation [10.1.19] that

$$\boldsymbol{\Psi}_s = \boldsymbol{\Phi}_1\boldsymbol{\Psi}_{s-1} + \boldsymbol{\Phi}_2\boldsymbol{\Psi}_{s-2} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{\Psi}_{s-p}. \qquad [11.B.12]$$

Taking transposes,

$$\boldsymbol{\Psi}_s' = \boldsymbol{\Psi}_{s-1}'\boldsymbol{\Phi}_1' + \boldsymbol{\Psi}_{s-2}'\boldsymbol{\Phi}_2' + \cdots + \boldsymbol{\Psi}_{s-p}'\boldsymbol{\Phi}_p'. \qquad [11.B.13]$$

Let the scalar ξ denote some particular element of $\boldsymbol{\pi}$, and differentiate [11.B.13] with respect to ξ:

$$\begin{aligned}
\frac{\partial \boldsymbol{\Psi}_s'}{\partial \xi} &= \boldsymbol{\Psi}_{s-1}' \frac{\partial \boldsymbol{\Phi}_1'}{\partial \xi} + \boldsymbol{\Psi}_{s-2}' \frac{\partial \boldsymbol{\Phi}_2'}{\partial \xi} + \cdots + \boldsymbol{\Psi}_{s-p}' \frac{\partial \boldsymbol{\Phi}_p'}{\partial \xi} \\
&\quad + \frac{\partial \boldsymbol{\Psi}_{s-1}'}{\partial \xi} \boldsymbol{\Phi}_1' + \frac{\partial \boldsymbol{\Psi}_{s-2}'}{\partial \xi} \boldsymbol{\Phi}_2' + \cdots + \frac{\partial \boldsymbol{\Psi}_{s-p}'}{\partial \xi} \boldsymbol{\Phi}_p' \\
&= [\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}'] \begin{bmatrix} \partial \mathbf{c}'/\partial \xi \\ \partial \boldsymbol{\Phi}_1'/\partial \xi \\ \partial \boldsymbol{\Phi}_2'/\partial \xi \\ \vdots \\ \partial \boldsymbol{\Phi}_p'/\partial \xi \end{bmatrix} \\
&\quad + \frac{\partial \boldsymbol{\Psi}_{s-1}'}{\partial \xi} \boldsymbol{\Phi}_1' + \frac{\partial \boldsymbol{\Psi}_{s-2}'}{\partial \xi} \boldsymbol{\Phi}_2' + \cdots + \frac{\partial \boldsymbol{\Psi}_{s-p}'}{\partial \xi} \boldsymbol{\Phi}_p' \\
&= [\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}'] \frac{\partial \boldsymbol{\Pi}}{\partial \xi} \\
&\quad + \frac{\partial \boldsymbol{\Psi}_{s-1}'}{\partial \xi} \boldsymbol{\Phi}_1' + \frac{\partial \boldsymbol{\Psi}_{s-2}'}{\partial \xi} \boldsymbol{\Phi}_2' + \cdots + \frac{\partial \boldsymbol{\Psi}_{s-p}'}{\partial \xi} \boldsymbol{\Phi}_p'.
\end{aligned} \tag{11.B.14}$$

Recall result [11.B.4], and note the special case when $\mathbf{A}$ is the $(n \times n)$ identity matrix, $\mathbf{B}$ is an $(n \times r)$ matrix, and $\mathbf{C}$ is an $(r \times q)$ matrix:

$$\text{vec}(\mathbf{BC}) = (\mathbf{C}' \otimes \mathbf{I}_n) \,\text{vec}(\mathbf{B}). \tag{11.B.15}$$

For example,

$$\text{vec}\left(\frac{\partial \boldsymbol{\Psi}_{s-1}'}{\partial \xi} \boldsymbol{\Phi}_1'\right) = (\boldsymbol{\Phi}_1 \otimes \mathbf{I}_n) \,\text{vec}\left(\frac{\partial \boldsymbol{\Psi}_{s-1}'}{\partial \xi}\right) = (\boldsymbol{\Phi}_1 \otimes \mathbf{I}_n)\left(\frac{\partial \boldsymbol{\psi}_{s-1}}{\partial \xi}\right). \tag{11.B.16}$$

Another implication of [11.B.4] can be obtained by letting $\mathbf{A}$ be an $(m \times q)$ matrix, $\mathbf{B}$ a $(q \times n)$ matrix, and $\mathbf{C}$ the $(n \times n)$ identity matrix:

$$\text{vec}(\mathbf{AB}) = (\mathbf{I}_n \otimes \mathbf{A}) \,\text{vec}(\mathbf{B}). \tag{11.B.17}$$

For example,

$$\begin{aligned}
&\text{vec}\left([\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}'] \frac{\partial \boldsymbol{\Pi}}{\partial \xi}\right) \\
&= [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}')] \left(\frac{\partial \,\text{vec}(\boldsymbol{\Pi})}{\partial \xi}\right) \\
&= [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}')] \left(\frac{\partial \boldsymbol{\pi}}{\partial \xi}\right).
\end{aligned} \tag{11.B.18}$$

Applying the vec operator to [11.B.14] and using [11.B.18] and [11.B.16] gives

$$\begin{aligned}
\frac{\partial \boldsymbol{\psi}_s}{\partial \xi} &= [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}_{s-1}' \quad \boldsymbol{\Psi}_{s-2}' \quad \cdots \quad \boldsymbol{\Psi}_{s-p}')] \left(\frac{\partial \boldsymbol{\pi}}{\partial \xi}\right) \\
&\quad + (\boldsymbol{\Phi}_1 \otimes \mathbf{I}_n) \left(\frac{\partial \boldsymbol{\psi}_{s-1}}{\partial \xi}\right) + (\boldsymbol{\Phi}_2 \otimes \mathbf{I}_n) \left(\frac{\partial \boldsymbol{\psi}_{s-2}}{\partial \xi}\right) \\
&\quad + \cdots + (\boldsymbol{\Phi}_p \otimes \mathbf{I}_n) \left(\frac{\partial \boldsymbol{\psi}_{s-p}}{\partial \xi}\right).
\end{aligned} \tag{11.B.19}$$

Letting ξ successively represent each of the elements of $\boldsymbol{\pi}$ and stacking the resulting equations horizontally as in [11.B.9] results in

$$\frac{\partial \boldsymbol{\psi}_s}{\partial \boldsymbol{\pi}'} = [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-1} \quad \boldsymbol{\Psi}'_{s-2} \quad \cdots \quad \boldsymbol{\Psi}'_{s-p})]$$
$$+ (\boldsymbol{\Phi}_1 \otimes \mathbf{I}_n)\left[\frac{\partial \boldsymbol{\psi}_{s-1}}{\partial \boldsymbol{\pi}'}\right] + \cdots + (\boldsymbol{\Phi}_p \otimes \mathbf{I}_n)\left[\frac{\partial \boldsymbol{\psi}_{s-p}}{\partial \boldsymbol{\pi}'}\right],$$

as claimed in [11.7.4]. ■

■ **Derivation of [11.7.5].** Here the task is to verify that if $\mathbf{G}_s$ is given by [11.7.5], then [11.7.4] holds:

$$\mathbf{G}_s = [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-1} \quad \boldsymbol{\Psi}'_{s-2} \quad \cdots \quad \boldsymbol{\Psi}'_{s-p})] + \sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n)\mathbf{G}_{s-k}. \qquad [11.B.20]$$

Notice that for $\mathbf{G}_s$ given by [11.7.5],

$$\begin{aligned}
&\sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n)\mathbf{G}_{s-k} \\
&= \sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n) \sum_{i=1}^{s-k} [\boldsymbol{\Psi}_{i-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-k-i} \quad \boldsymbol{\Psi}'_{s-k-i-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-k-i-p+1})] \\
&= \sum_{k=1}^{p} \sum_{i=1}^{s-k} [\boldsymbol{\Phi}_k \boldsymbol{\Psi}_{i-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-k-i} \quad \boldsymbol{\Psi}'_{s-k-i-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-k-i-p+1})].
\end{aligned}$$

For any given value for k and i, define $v \equiv k + i$. When $i = 1$, then $v = k + 1$; when $i = 2$, then $v = k + 2$; and so on:

$$\sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n)\mathbf{G}_{s-k} = \sum_{k=1}^{p} \sum_{v=k+1}^{s} [\boldsymbol{\Phi}_k \boldsymbol{\Psi}_{v-k-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})].$$

Recalling further that $\boldsymbol{\Psi}_{v-k-1} = \mathbf{0}$ for $v = 2, 3, \ldots, k$, we could equally well write

$$\begin{aligned}
&\sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n)\mathbf{G}_{s-k} \\
&= \sum_{k=1}^{p} \sum_{v=2}^{s} [\boldsymbol{\Phi}_k \boldsymbol{\Psi}_{v-k-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})] \\
&= \sum_{v=2}^{s} \sum_{k=1}^{p} [\boldsymbol{\Phi}_k \boldsymbol{\Psi}_{v-k-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})] \\
&= \sum_{v=2}^{s} \left[\left(\sum_{k=1}^{p} \boldsymbol{\Phi}_k \boldsymbol{\Psi}_{v-k-1}\right) \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})\right] \\
&= \sum_{v=2}^{s} [\boldsymbol{\Psi}_{v-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})],
\end{aligned} \qquad [11.B.21]$$

by virtue of [11.B.12]. If the first term on the right side of [11.B.20] is added to [11.B.21], the result is

$$\begin{aligned}
&[\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-1} \quad \boldsymbol{\Psi}'_{s-2} \quad \cdots \quad \boldsymbol{\Psi}'_{s-p})] + \sum_{k=1}^{p} (\boldsymbol{\Phi}_k \otimes \mathbf{I}_n)\mathbf{G}_{s-k} \\
&\quad = [\mathbf{I}_n \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-1} \quad \boldsymbol{\Psi}'_{s-2} \quad \cdots \quad \boldsymbol{\Psi}'_{s-p}) \\
&\qquad + \sum_{v=2}^{s} [\boldsymbol{\Psi}_{v-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})] \\
&\quad = \sum_{v=1}^{s} [\boldsymbol{\Psi}_{v-1} \otimes (\mathbf{0}_{n1} \quad \boldsymbol{\Psi}'_{s-v} \quad \boldsymbol{\Psi}'_{s-v-1} \quad \cdots \quad \boldsymbol{\Psi}'_{s-v-p+1})],
\end{aligned}$$

which is indeed the expression for $\mathbf{G}_s$ given in [11.7.5]. ■

■ **Derivation of [11.7.14] and [11.7.15].** Postmultiplying [11.7.12] by $\mathbf{B}_0$ and transposing results in

$$\mathbf{B}_0'\mathbf{H}_s' = \boldsymbol{\Psi}_s'. \qquad [11.B.22]$$

Let the scalar ξ denote some element of $\boldsymbol{\pi}$ or $\boldsymbol{\Omega}$, and differentiate [11.B.22] with respect to ξ:

$$(\partial \mathbf{B}_0'/\partial \xi)\mathbf{H}_s' + \mathbf{B}_0'(\partial \mathbf{H}_s'/\partial \xi) = \partial \boldsymbol{\Psi}_s'/\partial \xi. \qquad [11.B.23]$$

Applying the vec operator to [11.B.23] and using [11.B.15] and [11.B.17],

$$(\mathbf{H}_s \otimes \mathbf{I}_n)(\partial\, \text{vec}(\mathbf{B}_0')/\partial\xi) + (\mathbf{I}_n \otimes \mathbf{B}_0')(\partial\, \text{vec}(\mathbf{H}_s')/\partial\xi) = \partial\, \text{vec}(\mathbf{\Psi}_s')/\partial\xi,$$

implying that

$$\begin{aligned} \partial\mathbf{h}_s/\partial\xi &= -(\mathbf{I}_n \otimes \mathbf{B}_0')^{-1}(\mathbf{H}_s \otimes \mathbf{I}_n)(\partial\, \text{vec}(\mathbf{B}_0')/\partial\xi) + (\mathbf{I}_n \otimes \mathbf{B}_0')^{-1}\, \partial\boldsymbol{\psi}_s/\partial\xi \\ &= -[\mathbf{H}_s \otimes (\mathbf{B}_0')^{-1}](\partial\, \text{vec}(\mathbf{B}_0')/\partial\xi) + [\mathbf{I}_n \otimes (\mathbf{B}_0')^{-1}]\, \partial\boldsymbol{\psi}_s/\partial\xi. \end{aligned} \qquad [11.B.24]$$

Noticing that $\mathbf{B}_0'$ does not depend on $\boldsymbol{\pi}$, if [11.B.24] is stacked horizontally for $\xi = \pi_1, \pi_2, \ldots, \pi_{nk}$, the result is

$$\partial\mathbf{h}_s/\partial\boldsymbol{\pi}' = [\mathbf{I}_n \otimes (\mathbf{B}_0')^{-1}]\, \partial\boldsymbol{\psi}_s/\partial\boldsymbol{\pi}',$$

as claimed in [11.7.14]. Similarly, if ξ is an element of $\mathbf{\Omega}$, then ξ has no effect on $\mathbf{\Psi}_s$, and its influence on $\mathbf{B}_0'$ is given by

$$\frac{\partial\, \text{vec}(\mathbf{B}_0')}{\partial\xi} = \mathbf{S}_{B'}\frac{\partial\boldsymbol{\theta}_B}{\partial\xi}.$$

Stacking [11.B.24] horizontally with ξ representing each of the elements of vech($\mathbf{\Omega}$) thus produces

$$\frac{\partial\mathbf{h}_s}{\partial[\text{vech}(\mathbf{\Omega})]'} = -[\mathbf{H}_s \otimes (\mathbf{B}_0')^{-1}]\mathbf{S}_{B'}\frac{\partial\boldsymbol{\theta}_B}{\partial[\text{vech}(\mathbf{\Omega})]'},$$

as claimed in [11.7.15]. ■

Chapter 11 Exercises

11.1. Verify result [11.A.2].

11.2. Consider the following three-variable *VAR*:

$$\begin{aligned} y_{1t} &= \alpha y_{1,t-1} + \beta y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} &= \gamma y_{1,t-1} + \varepsilon_{2t} \\ y_{3t} &= \xi y_{1,t-1} + \zeta y_{2,t-1} + \eta y_{3,t-1} + \varepsilon_{3t}. \end{aligned}$$

(a) Is y_{1t} block-exogenous with respect to the vector $(y_{2t}, y_{3t})'$?
(b) Is the vector (y_{1t}, y_{2t}) block-exogenous with respect to y_{3t}?
(c) Is y_{3t} block-exogenous with respect to the vector $(y_{1t}, y_{2t})'$?

11.3. Consider the following bivariate *VAR*:

$$\begin{aligned} y_{1t} &= \alpha_1 y_{1,t-1} + \alpha_2 y_{1,t-2} + \cdots + \alpha_p y_{1,t-p} \\ &\quad + \beta_1 y_{2,t-1} + \beta_2 y_{2,t-2} + \cdots + \beta_p y_{2,t-p} + \varepsilon_{1t} \\ y_{2t} &= \gamma_1 y_{1,t-1} + \gamma_2 y_{1,t-2} + \cdots + \gamma_p y_{1,t-p} \\ &\quad + \delta_1 y_{2,t-1} + \delta_2 y_{2,t-2} + \cdots + \delta_p y_{2,t-p} + \varepsilon_{2t} \end{aligned}$$

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_\tau') = \begin{cases} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Use the results of Section 11.3 to write this in the form

$$\begin{aligned} y_{1t} &= \zeta_1 y_{1,t-1} + \zeta_2 y_{1,t-2} + \cdots + \zeta_p y_{1,t-p} \\ &\quad + \eta_1 y_{2,t-1} + \eta_2 y_{2,t-2} + \cdots + \eta_p y_{2,t-p} + u_{1t} \\ y_{2t} &= \lambda_0 y_{1t} + \lambda_1 y_{1,t-1} + \lambda_2 y_{1,t-2} + \cdots + \lambda_p y_{1,t-p} \\ &\quad + \xi_1 y_{2,t-1} + \xi_2 y_{2,t-2} + \cdots + \xi_p y_{2,t-p} + u_{2t}, \end{aligned}$$

where

$$E(\mathbf{u}_t\mathbf{u}_\tau') = \begin{cases} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

What is the relation between the parameters of the first representation ($\alpha_j, \beta_j, \gamma_j, \delta_j, \Omega_{il}$) and those of the second representation ($\zeta_j, \eta_j, \lambda_j, \xi_j, \sigma_i^2$)? What is the relation between $\boldsymbol{\varepsilon}_t$ and $\mathbf{u}_t$?

11.4. Write the result for Exercise 11.3 as

$$\begin{bmatrix} 1 - \zeta(L) & -\eta(L) \\ -\lambda_0 - \lambda(L) & 1 - \xi(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

or

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{u}_t.$$

Premultiply this system by the adjoint of $\mathbf{A}(L)$,

$$\mathbf{A}^*(L) \equiv \begin{bmatrix} 1 - \xi(L) & \eta(L) \\ \lambda_0 + \lambda(L) & 1 - \zeta(L) \end{bmatrix},$$

to deduce that y_{1t} and y_{2t} each admit a univariate $ARMA(2p, p)$ representation. Show how the argument generalizes to establish that if the $(n \times 1)$ vector $\mathbf{y}_t$ follows a pth-order autoregression, then each individual element y_{it} follows an $ARMA[np, (n-1)p]$ process. (See Zellner and Palm, 1974).

11.5. Consider the following bivariate VAR:

$$y_{1t} = 0.3y_{1,t-1} + 0.8y_{2,t-1} + \varepsilon_{1t}$$
$$y_{2t} = 0.9y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{2t},$$

with $E(\varepsilon_{1t}\varepsilon_{1\tau}) = 1$ for $t = \tau$ and 0 otherwise, $E(\varepsilon_{2t}\varepsilon_{2\tau}) = 2$ for $t = \tau$ and 0 otherwise, and $E(\varepsilon_{1t}\varepsilon_{2\tau}) = 0$ for all t and τ.

(a) Is this system covariance-stationary?

(b) Calculate $\mathbf{\Psi}_s = \partial\mathbf{y}_{t+s}/\partial\boldsymbol{\varepsilon}_t'$ for $s = 0$, 1, and 2. What is the limit as $s \to \infty$?

(c) Calculate the fraction of the MSE of the two-period-ahead forecast error for variable 1,

$$E[y_{1,t+2} - \hat{E}(y_{1,t+2}|y_t, y_{t-1}, \ldots)]^2,$$

that is due to $\varepsilon_{1,t+1}$ and $\varepsilon_{1,t+2}$.

Chapter 11 References

Ashley, Richard. 1988. "On the Relative Worth of Recent Macroeconomic Forecasts." *International Journal of Forecasting* 4:363–76.

Baillie, Richard T. 1987. "Inference in Dynamic Models Containing 'Surprise' Variables." *Journal of Econometrics* 35:101–17.

Bernanke, Ben. 1986. "Alternative Explanations of the Money-Income Correlation." *Carnegie-Rochester Conference Series on Public Policy* 25:49–100.

Blanchard, Olivier. 1989. "A Traditional Interpretation of Macroeconomic Fluctuations." *American Economic Review* 79:1146–64.

——— and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity* I:1989, 1–60.

——— and ———. 1990. "The Cyclical Behavior of the Gross Flows of U.S. Workers." *Brookings Papers on Economic Activity* II:1990, 85–155.

——— and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Aggregate Supply Disturbances." *American Economic Review* 79:655–73.

——— and Mark Watson. 1986. "Are Business Cycles All Alike?" in Robert J. Gordon, ed., *The American Business Cycle.* Chicago: University of Chicago Press.

Bouissou, M. B., J. J. Laffont, and Q. H. Vuong. 1986. "Tests of Noncausality under Markov Assumptions for Qualitative Panel Data." *Econometrica* 54:395–414.

Christiano, Lawrence J., and Lars Ljungqvist. 1988. "Money Does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics* 22:217–35.

Cooley, Thomas F., and Stephen F. LeRoy. 1985. "Atheoretical Macroeconometrics: A Critique." *Journal of Monetary Economics* 16:283–308.

Fama, Eugene F. 1965. "The Behavior of Stock Market Prices." *Journal of Business* 38:34–105.

Feige, Edgar L., and Douglas K. Pearce. 1979. "The Casual Causal Relationship between Money and Income: Some Caveats for Time Series Analysis." *Review of Economics and Statistics* 61:521–33.

Flavin, Marjorie A. 1981. "The Adjustment of Consumption to Changing Expectations about Future Income." *Journal of Political Economy* 89:974–1009.

Geweke, John. 1982. "Measurement of Linear Dependence and Feedback between Multiple Time Series." *Journal of the American Statistical Association* 77:304–13.

———, Richard Meese, and Warren Dent. 1983. "Comparing Alternative Tests of Causality in Temporal Systems: Analytic Results and Experimental Evidence." *Journal of Econometrics* 21:161–94.

Giannini, Carlo. 1992. *Topics in Structural VAR Econometrics*. New York: Springer-Verlag.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37:424–38.

Hamilton, James D. 1983. "Oil and the Macroeconomy since World War II." *Journal of Political Economy* 91:228–48.

———. 1985. "Historical Causes of Postwar Oil Shocks and Recessions." *Energy Journal* 6:97–116.

Hansen, Lars P., and Thomas J. Sargent. 1981. "Formulating and Estimating Dynamic Linear Rational Expectations Models," in Robert E. Lucas, Jr., and Thomas J. Sargent, eds., *Rational Expectations and Econometric Practice*, Vol. I. Minneapolis: University of Minnesota Press.

Keating, John W. 1990. "Identifying VAR Models under Rational Expectations." *Journal of Monetary Economics* 25:453–76.

Leamer, Edward. 1985. "Vector Autoregressions for Causal Inference?" *Carnegie-Rochester Conference Series on Public Policy* 22:255–303.

Lucas, Robert E., Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46:1429–45.

Lütkepohl, Helmut. 1989. "A Note on the Asymptotic Distribution of Impulse Response Functions of Estimated VAR Models with Orthogonal Residuals." *Journal of Econometrics* 42:371–76.

———. 1990. "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models." *Review of Economics and Statistics* 72:116–25.

Magnus, Jan R. 1978. "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix." *Journal of Econometrics* 7:281–312.

——— and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Pierce, David A., and Larry D. Haugh. 1977. "Causality in Temporal Systems: Characterization and a Survey." *Journal of Econometrics* 5:265–93.

Rothenberg, Thomas J. 1971. "Identification in Parametric Models." *Econometrica* 39:577–91.

———. 1973. *Efficient Estimation with a Priori Information*. New Haven, Conn.: Yale University Press.

Runkle, David E. 1987. "Vector Autoregressions and Reality." *Journal of Business and Economic Statistics* 5:437–42.

Shapiro, Matthew D., and Mark W. Watson. 1988. "Sources of Business Cycle Fluctuations," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1988*. Cambridge, Mass.: MIT Press.

Sims, Christopher A. 1972. "Money, Income and Causality." *American Economic Review* 62:540–52.

———. 1980. "Macroeconomics and Reality." *Econometrica* 48:1–48.

———. 1986. "Are Forecasting Models Usable for Policy Analysis?" *Quarterly Review of the Federal Reserve Bank of Minneapolis* (Winter), 2–16.

Stock, James H., and Mark W. Watson. 1989. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40:161–81.

Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57:348–68.

——— and Franz Palm. 1974. "Time Series Analysis and Simultaneous Equation Econometric Models." *Journal of Econometrics* 2:17–54.