5. INTRODUCTION TO VAR ANALYSIS

When we are not confident that a variable is actually exogeneous, a natural extension of transfer function analysis is to treat each variable symmetrically. In the two-variable case, we can let the time path of $\{y_t\}$ be affected by current and past realizations of the $\{z_t\}$ sequence *and* let the time path of the $\{z_t\}$ sequence be affected by current and past realizations of the $\{y_t\}$ sequence. Consider the simple bivariate system:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} \tag{5.17}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} \tag{5.18}$$

where it is assumed that (i) both y_t and z_t are stationary; (ii) ε_{yt} and ε_{zt} are white-noise disturbances with standard deviations of σ_y and σ_z, respectively; and (iii) $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are uncorrelated white-noise disturbances.

Equations (5.17) and (5.18) constitute a *first-order* vector autoregression (VAR) because the longest lag length is unity. This simple two-variable first-order VAR is useful for illustrating the multivariate higher order systems that are introduced in Section 8. The structure of the system incorporates feedback because y_t and z_t are allowed to affect each other. For example, $-b_{12}$ is the contemporaneous effect of a unit change of z_t on y_t and γ_{12} is the effect of a unit change in z_{t-1} on y_t. Note that the terms ε_{yt} and ε_{zt} are pure innovations (or shocks) in y_t and z_t, respectively. Of course, if b_{21} is not equal to zero, ε_{yt} has an indirect contemporaneous effect on z_t, and if b_{12} is not equal to zero, ε_{zt} has an indirect contemporaneous effect on y_t. Such a system could be used to capture the feedback effects in our temperature-thermostat example. The first equation allows current and past values of the thermostat setting to affect the time path of the temperature; the second allows for feedback between current and past values of the temperature and the thermostat setting.

Equations (5.17) and (5.18) cannot be estimated by OLS since y_t has a contemporaneous effect on z_t and z_t has a contemporaneous effect on y_t. The OLS estimates would suffer from simultaneous equation bias since the regressors and the error terms would be correlated. Fortunately, it is possible to transform the system of equations into a more usable form. Using matrix algebra, we can write the system in the compact form:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

or

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

where

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix},$$

$$\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

2134904 - Wiley US ©

Premultiplication by B^{-1} allows us to obtain the VAR model in *standard* form:

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{5.19}$$

where $A_0 = B^{-1}\Gamma_0, A_1 = B^{-1}\Gamma_1$, and $e_t = B^{-1}\varepsilon_t$.

For notational purposes, we can define a_{i0} as element i of the vector A_0, a_{ij} as the element in row i and column j of the matrix A_1, and e_{it} as the element i of the vector e_t. Using this new notation, we can rewrite (5.19) in the equivalent form:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \tag{5.20}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \tag{5.21}$$

To distinguish between the systems represented by (5.17) and (5.18) versus (5.20) and (5.21), the first is called a structural VAR or the primitive system and the second is called a VAR in standard form. It is important to note that the error terms (i.e., e_{1t} and e_{2t}) are composites of the two shocks ε_{yt} and ε_{zt}. Since $e_t = B^{-1}\varepsilon_t$, we can compute e_{1t} and e_{2t} as

$$e_{1t} = (\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12}b_{21}) \tag{5.22}$$

$$e_{2t} = (\varepsilon_{zt} - b_{21}\varepsilon_{yt})/(1 - b_{12}b_{21}) \tag{5.23}$$

Since ε_{yt} and ε_{zt} are white-noise processes, it follows that both e_{1t} and e_{2t} have zero means and constant variances and are individually serially uncorrelated. To find the properties of $\{e_{1t}\}$, first take the expected value of (5.22):

$$Ee_{1t} = E(\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12}b_{21}) = 0$$

The variance of e_{1t} is given by

$$\begin{aligned} Ee_{1t}^2 &= E[(\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12}b_{21})]^2 \\ &= (\sigma_y^2 + b_{12}^2\sigma_z^2)/(1 - b_{12}b_{21})^2 \end{aligned} \tag{5.24}$$

Thus, the variance of e_{1t} is time independent. The autocorrelations of e_{1t} and e_{1t-i} are

$$Ee_{1t}e_{1t-i} = E[(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{yt-i} - b_{12}\varepsilon_{zt-i})]/(1 - b_{12}b_{21})^2 = 0 \quad \text{for } i \neq 0$$

Similarly, (5.23) can be used to demonstrate that e_{2t} is a stationary process with zero mean, constant variance, and all autocovariances equal to zero. A critical point to note is that e_{1t} and e_{2t} are correlated. The covariance of the two terms is

$$\begin{aligned} Ee_{1t}e_{2t} &= E[(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{zt} - b_{21}\varepsilon_{yt})]/(1 - b_{12}b_{21})^2 \\ &= -(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)/(1 - b_{12}b_{21})^2 \end{aligned} \tag{5.25}$$

In general, (5.25) will not be zero so that the two shocks will be correlated. In the special case where $b_{12} = b_{21} = 0$ (i.e., if there are no contemporaneous effects of y_t on z_t and z_t on y_t), the shocks will be uncorrelated. It is useful to define the variance/covariance matrix of the e_{1t} and e_{2t} shocks as

$$\Sigma = \begin{bmatrix} \text{var}(e_{1t}) & \text{cov}(e_{1t}, e_{2t}) \\ \text{cov}(e_{1t}, e_{2t}) & \text{var}(e_{2t}) \end{bmatrix}$$

2134904 - Wiley US ©

Since all elements of Σ are time independent, we can use the more compact form

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \tag{5.26}$$

where $\operatorname{var}(e_{it}) = \sigma_i^2$ and $\operatorname{cov}(e_{1t}, e_{2t}) = \sigma_{12} = \sigma_{21}$.

Stability and Stationarity

In the first-order autoregressive model $y = a_0 + a_1 y_{t-1} + \varepsilon_t$, the stability condition is that a_1 be less than unity in absolute value. There is a direct analog between this stability condition and the matrix A_1 in the first-order VAR model of (5.19). Using the brute force method to solve the system, iterate (5.19) backward to obtain

$$\begin{aligned} x_t &= A_0 + A_1(A_0 + A_1 x_{t-2} + e_{t-1}) + e_t \\ &= (I + A_1)A_0 + A_1^2 x_{t-2} + A_1 e_{t-1} + e_t \end{aligned}$$

where $I = 2 \times 2$ identity matrix.

After n iterations,

$$x_t = (I + A_1 + \cdots + A_1^n)A_0 + \sum_{i=0}^{n} A_1^i e_{t-i} + A_1^{n+1} x_{t-n-1}$$

If we continue to iterate backward, it is clear that convergence requires that the expression A_1^n vanish as n approaches infinity. As shown below, stability requires that the roots of $(1 - a_{11}L)(1 - a_{22}L) - (a_{12}a_{21}L^2)$ lie outside the unit circle (the stability condition for higher-order systems is derived in Appendix 6.2 of Chapter 6). For the time being, if we assume that the stability condition is met, we can write the particular solution for x_t as

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} \tag{5.27}$$

where $\mu = [\bar{y}\ \ \bar{z}]'$ and

$$\begin{aligned} \bar{y} &= [a_{10}(1 - a_{22}) + a_{12}a_{20}]/\Delta; \qquad \bar{z} = [a_{20}(1 - a_{11}) + a_{21}a_{10}]/\Delta \\ \Delta &= (1 - a_{11})(1 - a_{22}) - a_{12}a_{21}. \end{aligned}$$

If we take the expected value of (5.27), the unconditional mean of x_t is μ; hence, the unconditional means of y_t and z_t are $\bar{y}$ and $\bar{z}$, respectively. The variances and covariances of y_t and z_t can be obtained as follows. First, form the variance/covariance matrix as

$$E(x_t - \mu)^2 = E\left[\sum_{i=0}^{\infty} A_1^i e_{t-i}\right]^2$$

Next, using (5.26) note that

$$\begin{aligned} Ee_t^2 &= \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \begin{bmatrix} e_{1t} & e_{2t} \end{bmatrix} \\ &= \Sigma \end{aligned}$$

2134904 - Wiley US ©

Since $Ee_te_{t-i} = 0$ for $i \neq 0$, it follows that

$$E(x_1 - \mu)^2 = (I + A_1^2 + A_1^4 + A_1^6 + \cdots)\Sigma$$
$$= [I - A_1^2]^{-1}\Sigma$$

where it is assumed that the stability condition holds, so A_1^n approaches zero as n approaches infinity.

If we can abstract from an initial condition, the $\{y_t\}$ and $\{z_t\}$ sequences will be jointly covariance stationary if the stability condition holds. Each sequence has a finite and time-invariant mean and a finite and time-invariant variance.

In order to get another perspective on the stability condition, use lag operators to rewrite the VAR model of (5.20) and (5.21) as

$$y_t = a_{10} + a_{11}Ly_t + a_{12}Lz_t + e_{1t}$$
$$z_t = a_{20} + a_{21}Ly_t + a_{22}Lz_t + e_{2t}$$

or

$$(1 - a_{11}L)y_t = a_{10} + a_{12}Lz_t + e_{1t}$$
$$(1 - a_{22}L)z_t = a_{20} + a_{21}Lz_t + e_{2t}$$

If we use this last equation to solve for z_t, it follows that Lz_t is

$$Lz_t = L(a_{20} + a_{21}Ly_t + e_{2t})/(1 - a_{22}L)$$

so that

$$(1 - a_{11}L)y_t = a_{10} + a_{12}L[(a_{20} + a_{21}Ly_t + e_{2t})/(1 - a_{22}L)] + e_{1t}$$

Notice that we have transformed the first-order VAR in the $\{y_t\}$ and $\{z_t\}$ sequences into a second-order stochastic difference equation in the $\{y_t\}$ sequence. Explicitly solving for y_t, we get

$$y_t = \frac{a_{10}(1 - a_{22}) + a_{12}a_{20} + (1 - a_{22}L)e_{1t} + a_{12}e_{2t-1}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2} \tag{5.28}$$

In the same manner, you should be able to demonstrate that the solution for z_t is

$$z_t = \frac{a_{20}(1 - a_{11}) + a_{21}a_{10} + (1 - a_{11}L)e_{2t} + a_{21}e_{1t-1}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2} \tag{5.29}$$

Equations (5.28) and (5.29) both have the same characteristic equation; convergence requires that the roots of the polynomial $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$ must lie outside the unit circle. (If you have forgotten the stability conditions for second-order difference equations, you might want to refresh your memory by reexamining Chapter 1.) Just as in any second-order difference equation, the roots may be real or complex and may be convergent or divergent. Notice that both y_t and z_t have the same characteristic equation; as long as a_{12} and a_{21} do not both equal zero, the solutions for the two sequences have the same characteristic roots. Hence, both will exhibit similar time paths.

2134904 - Wiley US ©

Dynamics of a VAR Model

Figure 5.6 shows the time paths of four simple systems. For each system, 100 sets of normally distributed random numbers representing the $\{e_{1t}\}$ and $\{e_{2t}\}$ sequences were drawn. The initial values of y_0 and z_0 were set equal to zero, and the $\{y_t\}$ and $\{z_t\}$ sequences were constructed as in (5.20) and (5.21). Panel (a) uses the values:

$$a_{10} = a_{20} = 0; a_{11} = a_{22} = 0.7; \text{and } a_{12} = a_{21} = 0.2$$

When we substitute these values into (5.27), it is clear that the mean of each series is zero. From the quadratic formula, the two roots of the inverse characteristic equation $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$ are 1.111 and 2.0. Since both are outside the unit circle, the system is stationary; the two characteristic roots of the solution for $\{y_t\}$ and $\{z_t\}$ are 0.9 and 0.5, respectively. Since these roots are positive, real, and less than unity, convergence will be direct. As you can see in the figure, there is a tendency for the sequences to move together. Since a_{21} is positive, a large realization in y_t induces a large realization of z_{t+1}; since a_{12} is positive, a large realization of z_t induces a large realization of y_{t+1}. The cross-correlations between the two series are positive.

Panel (b) illustrates a stationary process with $a_{10} = a_{20} = 0$, $a_{11} = a_{22} = 0.5$, and $a_{12} = a_{21} = -0.2$. Again, the mean of each series is zero, and the characteristic roots are 0.7 and 0.3. However, in contrast to the previous case, a_{21} and a_{12} are both negative so that positive realizations of y_t can be associated with negative realizations of z_{t+1}

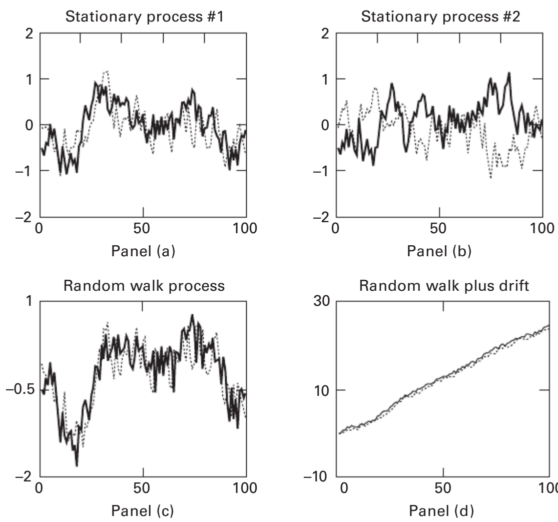

FIGURE 5.6 Four VAR Processes

2134904 - Wiley US ©

and vice versa. As can be seen from comparing the second panel, the two series appear to be negatively correlated.

In contrast, Panel (c) shows a process possessing a unit root; here, $a_{11} = a_{22} = a_{12} = a_{21} = 0.5$. You should take a moment to find the characteristic roots. Undoubtedly, there is little tendency for either of the series to revert to a constant long-run value. Here, the intercept terms a_{10} and a_{20} are equal to zero so that Panel (c) represents a multivariate generalization of the random walk model. You can see how the series seem to meander together. In Panel (d), the VAR process of Panel (c) also contains a nonzero intercept term ($a_{10} = 0.5$ and $a_{20} = 0$) that takes the role of a "drift." As you can see from Panel (d), the two series appear to move closely together. The drift term adds a deterministic time trend to the nonstationary behavior of the two series. Combined with the unit characteristic root, the $\{y_t\}$ and $\{z_t\}$ sequences are joint random walk plus drift processes. Notice that the presence of the drift dominates the long-run behavior of the series.

6. ESTIMATION AND IDENTIFICATION

One explicit aim of the Box–Jenkins approach is to provide a methodology that leads to parsimonious models. The ultimate objective of making accurate short-term forecasts is best served by purging insignificant parameter estimates from the model. Sims's (1980) criticisms of the "incredible identification restrictions" inherent in structural models argue for an alternative estimation strategy. Consider the following multivariate generalization of an autoregressive process:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + e_t \tag{5.30}$$

where x_t = an $(n \times 1)$ vector containing each of the n variables included in the VAR
A_0 = an $(n \times 1)$ vector of intercept terms
A_i = $(n \times n)$ matrices of coefficients
e_t = an $(n \times 1)$vector of error terms.

Sims's methodology entails little more than a determination of the appropriate variables to include in the VAR and a determination of the appropriate lag length. The variables to be included in the VAR are selected according to the relevant economic model. Lag length tests (to be discussed below) select the appropriate lag length. Otherwise, no explicit attempt is made to "pare down" the number of parameter estimates. The matrix A_0 contains n parameters, and each matrix A_i contains n^2 parameters; hence, $n + pn^2$ coefficients need to be estimated. Unquestionably, a VAR will be *overparameterized* in that many of these coefficient estimates will be insignificant. However, the goal is to find the important interrelationships among the variables. Improperly imposing zero restrictions may waste important information. Moreover, the regressors are likely to be highly collinear so that the t-tests on individual coefficients are not reliable guides for paring down the model.

Note that the right-hand side of (5.30) contains only predetermined variables and that the error terms are assumed to be serially uncorrelated with constant variance. Hence, *each equation in the system can be estimated using OLS*. Moreover, OLS estimates are consistent and asymptotically efficient. Even though the errors are correlated

2134904 - Wiley US ©

across equations, seemingly unrelated regressions (SUR) do not add to the efficiency of the estimation procedure since all regressions have identical right-hand side variables.

There is an issue of whether the variables in a VAR need to be stationary. Sims (1980) and Sims, Stock, and Watson (1990) recommended against differencing *even if the variables contain a unit root*. They argued that the goal of a VAR analysis is to determine the interrelationships among the variables, *not* to determine the parameter estimates. The main argument against differencing is that it "throws away" information concerning the comovements in the data (such as the possibility of cointegrating relationships). Similarly, it is argued that the data need not be detrended. In a VAR, a trending variable will be well approximated by a unit root plus drift. However, the majority view is that the form of the variables in the VAR should mimic the true data-generating process. This is particularly true if the aim is to estimate a structural model. We return to these issues in Chapter 6; for now, it is assumed that all variables are stationary. Questions 9 and 10 at the end of this chapter ask you to compare a VAR in levels to a VAR in first differences.

Forecasting

Once the VAR has been estimated, it can be used as a multiequation forecasting model. Suppose you estimate the first-order model $x_t = A_0 + A_1 x_{t-1} + e_t$ so as to obtain the values of the coefficients in A_0 and A_1. If your data run through period T, it is straightforward to obtain the one-step-ahead forecasts of your variables using the relationship $E_T x_{T+1} = A_0 + A_1 x_T$. Similarly, a two-step-ahead forecast can be obtained recursively from $E_T x_{T+2} = A_0 + A_1 E_T x_{T+1} = A_0 + A_1[A_0 + A_1 x_T]$. Nevertheless, in a higher-order model, there can be a large number of coefficient estimates. Since unrestricted VARs are overparameterized, the forecasts may be unreliable. In order to obtain a parsimonious model, many forecasters would purge the insignificant coefficients from the VAR. After reestimating the so-called **near-VAR** model using *SUR*, it could be used for forecasting purposes. Others might use a Bayesian approach by combining a set of prior beliefs with the traditional VAR methods presented in this text. West and Harrison (1989) provided an approachable introduction to the Bayesian approach. Litterman (1980) proposed a sensible set of Bayesian priors that have become the standard in Bayesian VAR models.

An interesting use of forecasting with a VAR is provided by the four-equation VAR of Eckstein and Tsiddon (2004). The aim of the study was to investigate the effects of terrorism (T) on the growth rates of Israeli real per capita GDP (ΔGDP_t), investment (ΔI_t), exports (ΔEXP_t), and nondurable consumption (ΔNDC_t). The authors use quarterly data running from 1980Q1 to 2003Q3 so that there are 95 total observations. The measure of terrorism is a weighted average of the number of Israeli fatalities, injuries, and noncasualty incidents due to both domestic and transnational attacks occurring in Israel. Consider a simplified version of their VAR model:

$$\begin{bmatrix} \Delta GDP_t \\ \Delta I_t \\ \Delta EXP_t \\ \Delta NDC_t \end{bmatrix} = \begin{bmatrix} A_{11}(L) & \dots & A_{14}(L) \\ \vdots & \ddots & \vdots \\ A_{41}(L) & \dots & A_{44}(L) \end{bmatrix} \begin{bmatrix} \Delta GDP_{t-1} \\ \Delta I_{t-1} \\ \Delta EXP_{t-1} \\ \Delta NDC_{t-1} \end{bmatrix} + \begin{bmatrix} c_1 T_{t-1} \\ c_2 T_{t-1} \\ c_3 T_{t-1} \\ c_4 T_{t-1} \end{bmatrix} + \dots + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{bmatrix}$$

2134904 - Wiley US ©

where the expressions $A_{ij}(L)$ are polynomials in the lag operator L, the c_i measure the influence of lagged terrorism on variable i, and the e_i are the regression errors. The other right-hand side variables (not shown) are the first difference of the real interest rate, three quarterly seasonal dummies, and an intercept.

The nature of the VAR is such that ΔGDP_t, ΔI_t, ΔEXP_t, and ΔNDC_t are all jointly determined. In contrast, the terrorism variable acts as an independent variable in the system. Notice that the magnitude of T_{t-1} is allowed to affect the four macroeconomic variables, but there is no feedback from these variables to the level of terrorism. The authors report that lagging the terrorism variable for a single period provided a better fit than the use of other lag lengths.

The four equations of the model were estimated through 2003Q3 and used to obtain 1 through 12-step-ahead forecasts of ΔGDP_t, ΔI_t, ΔEXP_t, and ΔNDC_t. Unlike forecasting with a pure VAR (in which all variables are jointly determined), it was necessary for Eckstein and Tsiddon (2004) to specify the time path of the terrorism variable. Consider the VAR representation of their model $x_t = A_0 + A_1 x_{t-1} + cT_{t-1} + e_t$, where c is the 4×1 vector $[c_1, c_2, c_3, c_4]'$. The one-step-ahead forecast is $E_T x_{T+1} = A_0 + A_1 x_T + cT_T$, and two-step-ahead forecast is $E_T x_{T+2} = A_0 + A_1 E_T[x_{T+1} + cT_{T+1}]$. Hence, in order to forecast the values of x_{T+2} and beyond, it is necessary to know the magnitude of the terrorism variable over the forecast period. Toward this end, Eckstein and Tsiddon supposed that all terrorism actually ended in 2003Q4 (so that all values of $T_j = 0$ for $j >$ 2003Q4. Under this assumption, the annual growth rate of GDP was estimated to be 2.5% through 2005Q3. Instead, when they set the values of T_j at the 2000Q4–2003Q4 period average, the growth rate of GDP was estimated to be zero. Thus, a steady level of terrorism would have cost the Israeli economy all of its real output gains. In actuality, the largest influence of terrorism was found to be on investment. The impact of terrorism on investment was twice as large as the impact on real GDP.

Identification

Suppose that you want to recover the structural VAR from your estimate of the model in standard form. To illustrate the identification procedure, return to the two-variable/first-order VAR of the previous section. Due to the feedback inherent in a VAR process, the primitive equations (5.17) and (5.18) cannot be estimated directly. The reason is that z_t is correlated with the error term ε_{yt} and that y_t is correlated with the error term ε_{zt}. Standard estimation techniques require that the regressors be uncorrelated with the error term. Note that there is no such problem in estimating the VAR system in the form of (5.20) and (5.21). OLS can provide estimates of the two elements of A_0 and the four elements of A_1. Moreover, obtaining the residuals from the two regressions, it is possible to calculate estimates of the variance of e_{1t}, e_{2t}, and the covariance between e_{1t} with e_{2t}. The issue is whether it is possible to recover all of the information present in the primitive system given by (5.17) and (5.18). In other words, is the primitive system identifiable given the OLS estimates of the VAR model in the form of (5.20) and (5.21)?

The answer to this question is, "No, unless we are willing to appropriately restrict the primitive system." The reason is clear if we compare the number of parameters

2134904 - Wiley US ©

of the primitive system with the number of parameters recovered from the estimated VAR model. Estimating (5.20) and (5.21) yields six coefficient estimates (a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, and a_{22}) and the calculated values of var(e_{1t}), var(e_{2t}), and cov(e_{1t}, e_{2t}). However, the primitive system (5.17) and (5.18) contains 10 parameters. In addition to the two intercept coefficients b_{10} and b_{20}, the four autoregressive coefficients γ_{11}, γ_{12}, γ_{21}, and γ_{22}, and the two feedback coefficients b_{12} and b_{21}, there are the two standard deviations σ_y and σ_z. In all, the primitive system contains 10 parameters, whereas the VAR estimation yields only 9 parameters. Unless one is willing to restrict one of the parameters, it is not possible to identify the primitive system; equations (5.17) and (5.18) are underidentified.

One way to identify the model is to use the type of **recursive** system proposed by Sims (1980). Suppose that you are willing to impose a restriction on the primitive system such that the coefficient b_{21} is equal to zero. Of course, forcing $b_{21} = 0$ imposes an asymmetry on the system in that z_t has a contemporaneous effect on y_t but y_t affects the $\{z_t\}$ sequence with a one-period lag. Nevertheless, it should be clear that this restriction (which might be suggested by a particular economic model) results in an exactly identified system. Writing (5.22) and (5.23) with the constraint imposed yields

$$e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt}$$
$$e_{2t} = \varepsilon_{zt}$$

so that

$$\mathrm{var}(e_1) = \sigma_y^2 + b_{12}^2\sigma_z^2 \quad (5.31)$$
$$\mathrm{var}(e_2) = \sigma_z^2 \quad (5.32)$$
$$\mathrm{cov}(e_1, e_2) = -b_{12}\,\sigma_z^2 \quad (5.33)$$

Equations (5.32) through (5.33) consist of three equations in three unknowns. Since the estimated variance/covariance matrix, Σ, contains var(e_1), var(e_2), and cov(e_1, e_2), the values of b_{12}, σ_z^2, and σ_y^2 can be identified recursively as $\sigma_z^2 = \mathrm{var}(e_2)$, $b_{12} = -\mathrm{cov}(e_1, e_2)/\sigma_z^2$, and $\sigma_y^2 = \mathrm{var}(e_1) - b_{12}^2\sigma_z^2$. To put matters another way, imposing the constraint means that the primitive system of (5.17) and (5.18) is given by

$$\begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Now, premultiplication of the primitive system by B^{-1} yields

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

or

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} - b_{12}b_{20} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} - b_{12}\gamma_{21} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\ \varepsilon_{zt} \end{bmatrix}$$

Estimating the system using OLS yields the parameter estimates from:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

2134904 - Wiley US ©

where $a_{10} = b_{10} - b_{12}b_{20}$, $a_{11} = \gamma_{11} - b_{12}\gamma_{21}$, $a_{12} = \gamma_{12} - b_{12}\gamma_{22}$, $a_{20} = b_{20}$, $a_{21} = \gamma_{21}$, and $a_{22} = \gamma_{22}$.

Along with (5.31) through (5.33), we have nine parameter estimates a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22}, var(e_1), var(e_2), and cov(e_1, e_2), which can be substituted into the nine equations above in order to simultaneously solve for b_{10}, b_{12}, γ_{11}, γ_{12}, b_{20}, γ_{21}, γ_{22}, σ_y^2, and σ_z^2.

Note also that the estimates of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences can be recovered. The residuals from the second equation (i.e., the $\{e_{2t}\}$ sequence) are estimates of the $\{\varepsilon_{zt}\}$ sequence. Combining these estimates along with the solution for b_{12} allows us to calculate the estimates of the $\{\varepsilon_{yt}\}$ sequence using the relationship $e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt}$.

Take a minute to examine the restriction. The assumption $b_{21} = 0$ means that y_t does not have a contemporaneous effect on z_t. The restriction manifests itself such that both ε_{yt} and ε_{zt} shocks affect the contemporaneous value of y_t but only ε_{zt} shocks affect the contemporaneous value of z_t. The observed values of e_{2t} are completely attributed to pure shocks to the $\{z_t\}$ sequence. Decomposing the residuals in this triangular fashion is called a **Choleski** decomposition.

In fact, the result is quite general. In an n-variable VAR, B is an $n \times n$ matrix since there are n regression residuals and n structural shocks. As shown in Section 10, exact identification requires that $(n^2 - n)/2$ restrictions be placed on the relationship between the regression residuals and the structural innovations. Since the Choleski decomposition is triangular, it forces exactly $(n^2 - n)/2$ values of the B matrix to equal zero.

7. THE IMPULSE RESPONSE FUNCTION

Just as an autoregression has a moving average representation, a vector autoregression can be written as a vector moving average (VMA). In fact, equation (5.27) is the VMA representation of (5.19) in that the variables (i.e., y_t and z_t) are expressed in terms of the current and past values of the two types of shocks (i.e., e_{1t} and e_{2t}). The VMA representation is an essential feature of Sims's (1980) methodology in that it allows you to trace out the time path of the various shocks on the variables contained in the VAR system. For illustrative purposes, continue to use the two-variable/first-order model analyzed in the previous two sections. Writing the two-variable VAR in matrix form,

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{5.34}$$

or, using (5.27), we get

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \tag{5.35}$$

Equation (5.35) expresses y_t and z_t in terms of the $\{e_{1t}\}$ and $\{e_{2t}\}$ sequences. However, it is insightful to rewrite (5.35) in terms of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences. From (5.22) and (5.23), the vector of errors can be written as

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \tag{5.36}$$

2134904 - Wiley US ©

so that (5.35) and (5.36) can be combined to form

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

Since the notation is getting unwieldy, we can simplify by defining the 2×2 matrix ϕ with elements $\phi_{jk}(i)$:

$$\phi_i = \frac{A_1^i}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$$

Hence, the moving average representation of (5.35) and (5.36) can be written in terms of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

or, more compactly,

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i} \tag{5.37}$$

The moving average representation is an especially useful tool to examine the interaction between the $\{y_t\}$ and $\{z_t\}$ sequences. The coefficients of ϕ_i can be used to generate the effects of ε_{yt} and ε_{zt} shocks on the entire time paths of the $\{y_t\}$ and $\{z_t\}$ sequences. If you understand the notation, it should be clear that the four elements $\phi_{jk}(0)$ are **impact multipliers**. For example, the coefficient $\phi_{12}(0)$ is the instantaneous impact of a one-unit change in ε_{zt} on y_t. In the same way, the elements $\phi_{11}(1)$ and $\phi_{12}(1)$ are the one-period responses of unit changes in ε_{yt-1} and ε_{zt-1} on y_t, respectively. Updating by one period indicates that $\phi_{11}(1)$ and $\phi_{12}(1)$ also represent the effects of unit changes in ε_{yt} and ε_{zt} on y_{t+1}.

The accumulated effects of unit impulses in ε_{yt} and/or ε_{zt} can be obtained by the appropriate summation of the coefficients of the impulse response functions. For example, note that, after n periods, the effect of ε_{zt} on the value of y_{t+n} is $\phi_{12}(n)$. Thus, after n periods, the cumulated sum of the effects of ε_{zt} on the $\{y_t\}$ sequence is

$$\sum_{i=0}^{n} \phi_{12}(i)$$

Letting n approach infinity yields the total cumulated effect. If the $\{y_t\}$ and $\{z_t\}$ sequences are assumed to be stationary, it must be the case that for all j and k, the values of $\phi_{jk}(i)$ converge to zero as i gets large. This follows as shocks cannot have a permanent effect on a stationary series. It also follows that

$$\sum_{i=0}^{\infty} \phi_{jk}^2(i) \text{ is finite}$$

The four sets of coefficients $\phi_{11}(i)$, $\phi_{12}(i)$, $\phi_{21}(i)$, and $\phi_{22}(i)$ are called the **impulse response functions.** Plotting the impulse response functions [i.e., plotting the coefficients of $\phi_{jk}(i)$ against i] is a practical way to visually represent the behavior of

2134904 - Wiley US ©

the $\{y_t\}$ and $\{z_t\}$ series in response to the various shocks. In principle, it might be possible to know all of the parameters of the primitive system (5.17) and (5.18). With such knowledge, it would be possible to trace out the time paths of the effects of pure ε_{yt} or ε_{zt} shocks. However, this methodology is not available to the researcher since an estimated VAR is underidentified. As explained in the previous section, knowledge of the various a_{ij} and the variance/covariance matrix Σ is not sufficient to identify the primitive system. Hence, the econometrician must impose an additional restriction on the two-variable VAR system in order to identify the impulse responses.

One possible identification restriction is to impose the recursive ordering (or Choleski decomposition) used in (5.31), such that y_t does not have a contemporaneous effect on z_t. Formally, this restriction is represented by setting $b_{21} = 0$ in the primitive system. In terms of (5.36), the error terms can be decomposed as follows:

$$e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt} \tag{5.38}$$

$$e_{2t} = \varepsilon_{zt} \tag{5.39}$$

As already noted, if we use (5.39), all of the observed errors from the $\{e_{2t}\}$ sequence are attributed to ε_{zt} shocks. Given the calculated $\{\varepsilon_{zt}\}$ sequence, knowledge of the values of the $\{e_{1t}\}$ sequence and the correlation coefficient between e_{1t} and e_{2t} allows for the calculation of the $\{\varepsilon_{yt}\}$ sequence using (5.38). Although this Choleski decomposition constrains the system such that an ε_{yt} shock has no direct effect on z_t, there is an indirect effect in that lagged values of y_t affect the contemporaneous value of z_t. The key point is that the decomposition forces a potentially important asymmetry on the system since an ε_{zt} shock has contemporaneous effects on both y_t and z_t. For this reason, (5.38) and (5.39) are said to be an **ordering** of the variables. An ε_{zt} shock directly affects e_{1t} and e_{2t}, but an ε_{yt} shock does not affect e_{2t}. Hence, z_t is said to be "causally prior" to y_t.

Suppose that estimates of equations (5.20) and (5.21) yield the values $a_{10} = a_{20} = 0$, $a_{11} = a_{22} = 0.7$, and $a_{12} = a_{21} = 0.2$. You will recall that this is precisely the model used in the simulation reported in Panel (a) of Figure 5.6. Also, suppose that the elements of the Σ matrix are such that $\sigma_1^2 = \sigma_2^2$ and that $\text{cov}(e_{1t}, e_{2t})$ is such that the correlation coefficient between e_{1t} and e_{2t} (denoted by ρ_{12}) is 0.8. Hence, the decomposed errors can be represented by[4]

$$e_{1t} = \varepsilon_{yt} + 0.8\varepsilon_{zt} \tag{5.40}$$

$$e_{2t} = \varepsilon_{zt} \tag{5.41}$$

Panels (a) and (b) of Figure 5.7 trace out the effects of one-unit shocks to ε_{zt} and ε_{yt} on the time paths of the $\{y_t\}$ and $\{z_t\}$ sequences. As shown in Panel (a), a one-unit shock in ε_{zt} causes z_t to jump by one unit and y_t to jump by 0.8 units. [From (5.40), 80% of the ε_{zt} shock has a contemporaneous effect on e_{1t}.] In the next period, ε_{zt+1} returns to zero, but the autoregressive nature of the system is such that y_{t+1} and z_{t+1} do not immediately return to their long-run values. Since $z_{t+1} = 0.2y_t + 0.7z_t + \varepsilon_{zt+1}$, it follows that $z_{t+1} =$ 0.86 $[0.2(0.8) + 0.7(1) = 0.86]$. Similarly, $y_{t+1} = 0.7y_t + 0.2z_t = 0.76$. As you can see from the figure, the subsequent values of the $\{y_t\}$ and $\{z_t\}$ sequences converge to their long-run levels. This convergence is assured by the stability of the system; as found earlier, the two characteristic roots are 0.5 and 0.9.

2134904 - Wiley US ©

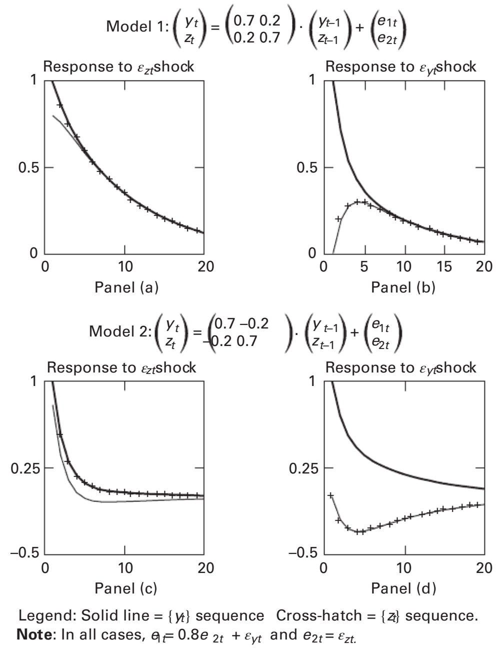

FIGURE 5.7 Two Impulse Response Functions

The effects of a one-unit shock in ε_{yt} are shown in Panel (b) of the figure. You can see the asymmetry of the decomposition immediately by comparing the two upper graphs. A one-unit shock in ε_{yt} causes the value of y_t to increase by one unit; however, *there is no contemporaneous effect on the value of* z_t so that $y_t = 1$ and $z_t = 0$. In the subsequent period, ε_{yt+1} returns to zero. The autoregressive nature of the system is such that $y_{t+1} = 0.7y_t + 0.2z_t = 0.7$ and $z_{t+1} = 0.2y_t + 0.7z_t = 0.2$. The remaining points in the figure are the impulse responses for periods $t + 2$ through $t + 20$. Since the system is stationary, the impulse responses ultimately decay.

Can you figure out the consequences of reversing the Choleski decomposition in such a way that b_{12}, rather than b_{21}, is constrained to equal zero? Since matrix A_1 is symmetric (i.e., $a_{11} = a_{22}$ and $a_{12} = a_{21}$), the impulse responses of an ε_{yt} shock would be similar to those in Panel (a) and the impulse responses of an ε_{zt} would be similar to those in Panel (b). The only difference would be that the solid line would represent the

2134904 - Wiley US ©

time path of the $\{z_t\}$ sequence and the hatched line would represent the time path of the $\{y_t\}$ sequence.

As a practical matter, how does the researcher decide which of the alternative decompositions is most appropriate? In some instances, there might be a theoretical reason to suppose that one variable has no contemporaneous effect on the other. In the terrorism/tourism example, knowledge that terrorist incidents affect tourism with a lag suggests that terrorism does not have a contemporaneous effect on tourism. Usually, there is no such *a priori* knowledge. Moreover, the very idea of imposing a structure on a VAR system seems contrary to the spirit of Sims's argument against "incredible identifying restrictions." Unfortunately, there is no simple way to circumvent the problem; identification necessitates imposing *some* structure on the system. The Choleski decomposition provides a minimal set of assumptions that can be used to identify the structural model.

It is crucial to note that *the importance of the ordering depends on the magnitude of the correlation coefficient between* e_{1t} *and* e_{2t}. Let this correlation coefficient be denoted by ρ_{12} so that $\rho_{12} = \sigma_{12}/(\sigma_1\sigma_2)$. Now suppose that the estimated model yields a value of Σ such that ρ_{12} is found to be equal to zero. In this circumstance, the ordering would be immaterial. Formally, (5.38) and (5.39) become $e_{1t} = \varepsilon_{yt}$ and $e_{2t} = \varepsilon_{zt}$. Since there is no correlation across equations, the residuals from the y_t and z_t equations are necessarily equivalent to the ε_{yt} and ε_{zt} shocks, respectively. The point is that if $Ee_{1t}e_{2t} = 0$, b_{12} and b_{21} can both be set equal to zero. At the other extreme, if ρ_{12} is found to be unity, there is a single shock that contemporarily affects both variables. When $\rho_{12} = 0$ and maintaining the assumption $b_{21} = 0$, (5.38) and (5.39) become $e_{1t} = \varepsilon_{zt}$ and $e_{2t} = \varepsilon_{zt}$, and under the alternative assumption $b_{12} = 0$, it follows that $e_{1t} = \varepsilon_{yt}$ and $e_{2t} = \varepsilon_{yt}$. Usually, the researcher will want to test the significance of ρ_{12}. As in a univariate model, you can test the null hypothesis $\rho_{12} = 0$ using a normal distribution with a mean of zero and a standard deviation of $T^{-0.5}$. As such, with 100 usable observations, if $|\rho_{12}| > 0.2$, the correlation is deemed to be significant at conventional levels. If ρ_{12} is significant, the usual procedure is to obtain the impulse response function using a particular ordering. Compare the results to the impulse response function obtained by reversing the ordering. If the implications are quite different, additional investigation into the relationships between the variables is necessary.

The lower half of Figure 5.7, Panels (c) and (d), show the impulse response functions for a second model; the sole difference between models 1 and 2 is the change in the values of a_{12} and a_{21} to -0.2. Notice that this model was used in the simulation reported in Panel (b) of Figure 5.6. The negative off-diagonal elements of A_1 weaken the tendency for the two series to move together. Panel (c) traces out the effect of a one-unit ε_{zt} shock using ordering represented by (5.40) and (5.41). In period t, z_t rises by one unit and y_t rises by 0.8 units. In period $(t + 1)$, ε_{zt+1} returns to zero but the value of y_{t+1} is $0.7y_t - 0.2z_t = 0.36$ and the value of z_{t+1} is $-0.2y_t + 0.7z_t = 0.54$. The points represented by $t = 2$ through 20 show that the impulse responses converge to zero. Panel (d) traces the responses of a one-unit ε_{yt} shock. Since the value of z_t is unaffected by the shock, in period $(t + 1)$, $y_{t+1} = 0.7y_t - 0.2z_t = 0.7$ and

2134904 - Wiley US ©

$z_{t+1} = -0.2y_t + 0.7z_t = -0.2$. In the same way, $y_{t+2} = 0.7 * 0.7 - 0.2 * (-0.2) = 0.53$ and $z_{t+2} = -0.2 * (0.7) + 0.7 * (-0.2) = -0.28$. Since the system is stable, both sequences eventually converge to zero.

Confidence Intervals and Impulse Responses

One key issue concerning the impulse response functions is that they are constructed using the estimated coefficients. Since each coefficient is estimated imprecisely, the impulse responses also contain error. The issue is to construct confidence intervals around the impulse responses that allow for the parameter uncertainty inherent in the estimation process. To illustrate the methodology, consider the following estimate of an *AR*(1) model:

$$y_t = \underset{(4.00)}{0.60}y_{t-1} + \varepsilon_t$$

Given the t-statistic of 4.00, the AR(1) coefficient seems to be well estimated. It is easy to form the impulse response function: For any given level of y_{t-1}, a one-unit shock to ε_t will increase y_t by one unit. In subsequent periods, y_{t+1} will be 0.60 and y_{t+2} will be $(0.60)^2$. As you can easily verify, the impulse response function can be written as $\phi(i) = (0.60)^i$.

Notice that the point estimate of the AR(1) coefficient is 0.6 with a standard deviation of 0.15 ($0.15 = 0.60/4.00$). If we are willing to assume that the coefficient is normally distributed, there is a 95% chance that the actual value lies within the two standard deviation interval 0.3–0.9. As such, the decay pattern could be anywhere between $\phi(i) = (0.90)^i$ and $\phi(i) = (0.30)^i$. The problem is much more complicated in higher-order systems since the estimated coefficients will be correlated. Moreover, you may not want to assume normality. One way to obtain the desired confidence intervals from the AR(p) process $y_t = a_0 + a_1y_{t-1} + \cdots + a_py_{t-p} + \varepsilon_t$ is to perform the following Monte Carlo study:

1. Estimate the coefficients a_0 through a_p using OLS and save the residuals. Let $\hat{a}_i$ denote the estimated value of a_i and let $\{\hat{\varepsilon}_i\}$ denote the estimated residuals.

2. For a sample size of T, draw T random numbers so as to represent the $\{\varepsilon_t\}$ sequence. Most software packages will draw the numbers using randomly selected values of $\hat{\varepsilon}_i$ (with replacement). In this way, they actually generate bootstrap confidence intervals. Thus, you will have a simulated series of length T, called ε_t^s, which should have the same properties as the true error process. Use these random numbers to construct the simulated $\{y_t^s\}$ sequence as

$$y_t^s = \hat{a}_0 + \hat{a}_1y_{t-1}^s + \cdots + \hat{a}_py_{t-p}^s + \varepsilon_t^s$$

Be sure that you appropriately initialize the series so as to eliminate the effects of the initial conditions.

2134904 - Wiley US ©

3. Now act as if you did not know the coefficient values used to generate the y_t^s series. Estimate y_t^s as an AR(p) process and obtain the impulse response function. If you repeat the process several thousand times, you can generate several thousand impulse response functions. You use these impulse response functions to construct the confidence intervals. For example, you can construct the interval that excludes the lowest 2.5% and highest 2.5% of the responses to obtain a 95% confidence interval.

The benefit of this method is that you do not need to make any special assumptions concerning the distribution of the autoregressive coefficients. The actual calculation of confidence intervals is only a bit more complicated in a VAR. Consider the two-variable system:

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$
$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

The complicating issue is that the regression residuals are correlated. As such, you need to draw e_{1t} and e_{2t} so as to maintain the appropriate error structure. A simple method is to draw e_{1t} and use the value of e_{2t} that corresponds to that same period. If you use a Choleski decomposition such that $b_{21} = 0$, construct ε_{1t} and ε_{2t} using (5.38) and (5.39). Figure 5.8 reports confidence intervals from a two-variable VAR that has been estimated using the domestic and transnational terrorism data shown in Figure 5.1.

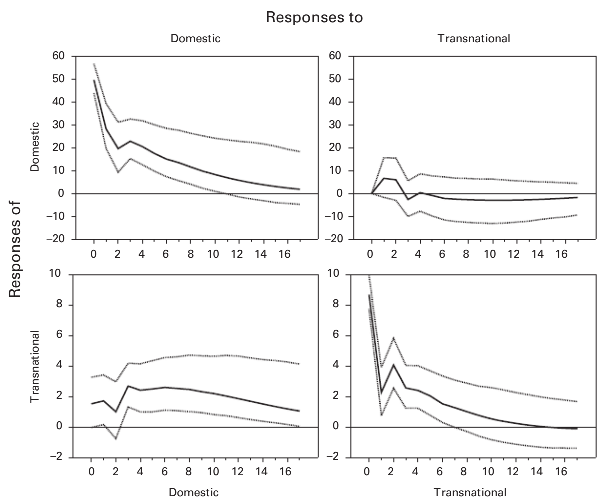

FIGURE 5.8 Impulse Responses of Terrorism

2134904 - Wiley US ©

You can see that the responses of domestic terrorism to transnational shocks are never significant.

Variance Decomposition

Another useful aid in uncovering interrelationships among the variables in the system is a forecast error variance decomposition. Suppose that we knew the coefficients of A_0 and A_1 and wanted to forecast the various values of x_{t+i} conditional on the observed value of x_t. Updating (5.19) one period (i.e., $x_{t+1} = A_0 + A_1x_t + e_{t+1}$) and taking the conditional expectation of x_{t+1}, we obtain

$$E_t x_{t+1} = A_0 + A_1x_t$$

Note that the one-step-ahead forecast error is $x_{t+1} - E_tx_{t+1} = e_{t+1}$. Similarly, updating two periods, we get

$$\begin{aligned} x_{t+2} &= A_0 + A_1x_{t+1} + e_{t+2} \\ &= A_0 + A_1(A_0 + A_1x_t + e_{t+1}) + e_{t+2} \end{aligned}$$

If we take conditional expectations, the two-step-ahead forecast of x_{t+2} is

$$E_tx_{t+2} = (I + A_1)A_0 + A_1^2x_t$$

The two-step-ahead forecast error (i.e., the difference between the realization of x_{t+2} and the forecast) is $e_{t+2} + A_1e_{t+1}$. More generally, it is easily verified that the n-step-ahead forecast is

$$E_tx_{t+n} = (I + A_1 + A_1^2 + \cdots + A_1^{n-1})A_0 + A_1^nx_t$$

and that the associated forecast error is

$$e_{t+n} + A_1e_{t+n-1} + A_1^2e_{t+n-2} + \cdots + A_1^{n-1}e_{t+1} \tag{5.42}$$

We can also consider these forecast errors in terms of (5.37) (i.e., the VMA form of the structural model). Of course, the VMA and the VAR models contain exactly the same information but it is convenient (and a good exercise) to describe the properties of the forecast errors in terms of the $\{\varepsilon_t\}$ sequence. If we use (5.37) to conditionally forecast x_{t+1}, one step ahead the forecast error is $\phi_0\varepsilon_{t+1}$. In general,

$$x_{t+n} = \mu + \sum_{i=0}^{\infty}\phi_i\varepsilon_{t+n-i}$$

so that the n-period forecast error $x_{t+n} - E_tx_{t+n}$ is

$$x_{t+n} - E_tx_{t+n} = \sum_{i=0}^{n-1}\phi_i\varepsilon_{t+n-i}$$

Focusing solely on the $\{y_t\}$ sequence, we see that the n-step-ahead forecast error is

$$\begin{aligned} y_{t+n} - E_ty_{t+n} = {} & \phi_{11}(0)\varepsilon_{yt+n} + \phi_{11}(1)\varepsilon_{yt+n-1} + \cdots + \phi_{11}(n-1)\varepsilon_{yt+1} \\ & + \phi_{12}(0)\varepsilon_{zt+n} + \phi_{12}(1)\varepsilon_{zt+n-1} + \cdots + \phi_{12}(n-1)\varepsilon_{zt+1} \end{aligned}$$

2134904 - Wiley US ©

Denote the n-step-ahead forecast error variance of y_{t+n} as $\sigma_y(n)^2$:

$$\sigma_y(n)^2 = \sigma_y^2[\phi_{11}(0)^2 + \phi_{11}(1)^2 + \cdots + \phi_{11}(n-1)^2] + \sigma_z^2[\phi_{12}(0)^2 + \phi_{12}(1)^2 + \cdots + \phi_{12}(n-1)^2]$$

Because all values of $\phi_{jk}(i)^2$ are necessarily nonnegative, the variance of the forecast error increases as the forecast horizon n increases. Note that it is possible to decompose the n-step-ahead forecast error variance into the proportions due to each shock. The proportions of $\sigma_y(n)^2$ due to shocks in the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences are

$$\frac{\sigma_y^2[\phi_{11}(0)^2 + \phi_{11}(1)^2 + \cdots + \phi_{11}(n-1)^2]}{\sigma_y(n)^2}$$

and

$$\frac{\sigma_z^2[\phi_{12}(0)^2 + \phi_{12}(1)^2 + \cdots + \phi_{12}(n-1)^2]}{\sigma_y(n)^2}$$

respectively.

The **forecast error variance decomposition** tells us the proportion of the movements in a sequence due to its "own" shocks versus shocks to the other variable. If ε_{zt} shocks explain none of the forecast error variance of $\{y_t\}$ at all forecast horizons, we can say that the $\{y_t\}$ sequence is exogeneous. In this circumstance, $\{y_t\}$ evolves independently of the ε_{zt} shocks and the $\{z_t\}$ sequence. At the other extreme, ε_{zt} shocks could explain all of the forecast error variance in the $\{y_t\}$ sequence at all forecast horizons, so that $\{y_t\}$ would be entirely endogeneous. In applied research, it is typical for a variable to explain almost all of its forecast error variance at short horizons and smaller proportions at longer horizons. We would expect this pattern if ε_{zt} shocks had little contemporaneous effect on y_t but acted to affect the $\{y_t\}$ sequence with a lag.

Note that the variance decomposition contains the same problem inherent in impulse response function analysis. In order to identify the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences, it is necessary to restrict the B matrix. The Choleski decomposition used in (5.38) and (5.39) necessitates that all of the one-period forecast error variance of z_t is due to ε_{zt}. If we use the alternative ordering, all of the one-period forecast error variance of y_t would be due to ε_{yt}. The effects of these alternative assumptions are reduced at longer forecasting horizons. In practice, it is useful to examine the variance decompositions at various forecast horizons. As n increases, the variance decompositions should converge. Moreover, if the correlation coefficient ρ_{12} is significantly different from zero, it is customary to obtain the variance decompositions under various orderings.

Nevertheless, impulse analysis and variance decompositions (together called **innovation accounting**) can be useful tools to examine the relationships among economic variables. If the correlations among the various innovations are small, the identification problem is not likely to be especially important. The alternative orderings should yield similar impulse responses and variance decompositions. Of course, the contemporaneous movements of many economic variables are highly correlated. Sections 10–13 consider two attractive methods that can be used to identify the structural innovations. Before examining these techniques, we consider hypothesis testing in a VAR

2134904 - Wiley US ©

framework and reexamine the interrelationships between domestic and transnational terrorism.

8. TESTING HYPOTHESES

In principle, there is nothing to prevent you from incorporating a large number of variables in the VAR. It is possible to construct an n-equation VAR with each equation containing p lags of all n variables in the system. You will want to include those variables that have important economic effects on each other. As a practical matter, degrees of freedom are quickly eroded as more variables are included. For example, using monthly data with 12 lags, the inclusion of one additional variable uses an additional 12 degrees of freedom in each equation. A careful examination of the relevant theoretical model will help you to select the set of variables to include in your VAR model.

An n-equation VAR can be represented by

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ x_{nt} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ \cdot \\ A_{n0} \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) & \cdot & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \cdot & A_{2n}(L) \\ \cdot \quad \cdot & & \cdot & \cdot \\ A_{n1}(L) & A_{n2}(L) & \cdot & A_{nn}(L) \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \cdot \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ \cdot \\ e_{nt} \end{bmatrix} \tag{5.43}$$

where A_{i0} = the parameters representing intercept terms
$A_{ij}(L)$ = the polynomials in the lag operator L.

The individual coefficients of $A_{ij}(L)$ are denoted by $a_{ij}(1)$, $a_{ij}(2)$, ... Since all equations have the same lag length, the polynomials $A_{ij}(L)$ are all of the same degree. The terms e_{it} are white-noise disturbances that may be correlated with each other. Again, designate the variance/covariance matrix by Σ, where the dimension of Σ is $(n \times n)$.

In addition to the determination of the set of variables to include in the VAR, it is important to determine the appropriate lag length. One possible procedure is to allow for different lag lengths for each variable in each equation. However, in order to preserve the symmetry of the system (and to be able to use OLS efficiently), it is common to use the same lag length for all equations. As indicated in Section 6, as long as there are identical regressors in each equation, OLS estimates are consistent and asymptotically efficient. If some of the VAR equations have regressors not included in the others, seemingly unrelated regressions (SUR) provide efficient estimates of the VAR coefficients. Hence, when there is a good reason to let lag lengths differ across equations, estimate the so-called **near-VAR** using SUR.

In a VAR, long-lag lengths quickly consume degrees of freedom. If lag length is p, each of the n equations contains np coefficients plus the intercept term. Appropriate lag length selection can be critical. If p is too small, the model is misspecified; if p is too large, degrees of freedom are wasted. To check lag length, begin with the longest plausible length or the longest feasible length given degrees-of-freedom considerations. Estimate the VAR and form the variance/covariance matrix of the residuals. Using quarterly data, you might start with a lag length of 12 quarters based on the *a priori* notion that 3 years is sufficiently long to capture the system's dynamics.

2134904 - Wiley US ©

Call the variance/covariance matrix of the residuals from the 12-lag model Σ_{12}. Now suppose you want to determine whether eight lags are appropriate. After all, restricting the model from 12 to 8 lags would reduce the number of estimated parameters by $4n$ in each equation.

Since the goal is to determine whether lag 8 is appropriate for all equations, an equation by equation F-test on lags 9 through 12 is not appropriate. Instead, the proper test for this **cross-equation** restriction is a likelihood ratio test. Reestimate the VAR *over the same sample period* using eight lags and obtain the variance/covariance matrix of the residuals Σ_8. Note that Σ_8 pertains to a system of n equations with $4n$ restrictions in each equation, for a total of $4n^2$ restrictions. The likelihood ratio statistic is

$$(T)(\ln|\Sigma_8| - \ln|\Sigma_{12}|)$$

However, given the sample sizes usually found in economic analysis, Sims (1980) recommended using

$$(T - c)\left(\ln|\Sigma_8| - \ln|\Sigma_{12}|\right)$$

where T is number of usable observations, c the number of parameters estimated in each equation of the unrestricted system, and $\ln|\Sigma_n|$ = the natural logarithm of the determinant of Σ_n.

In the example at hand, $c = 1 + 12n$ since each equation of the unrestricted model has 12 lags for each variable plus an intercept.

This statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of restrictions *in the system*. In the example under consideration, there are $4n$ restrictions in each equation, for a total of $4n^2$ restrictions in the system. Clearly, if the restriction of a reduced number of lags is not binding, we would expect $\ln|\Sigma_8|$ to be equal to $\ln|\Sigma_{12}|$. Large values of this sample statistic indicate that having only eight lags is a binding restriction; hence, we can reject the null hypothesis that lag length = 8. If the calculated value of the statistic is less than χ^2 at a prespecified significance level, we will not be able to reject the null of only eight lags. At that point, we could seek to determine whether four lags were appropriate by constructing

$$(T - c)(\ln|\Sigma_4| - \ln|\Sigma_8|)$$

Considerable care should be taken in paring down lag length in this manner. Often, this procedure will not reject the null hypotheses of 8 versus 12 lags and 4 versus 8 lags, although it will reject a null of 4 versus 12 lags. The problem with paring down the model is that you may lose a small amount of explanatory power at each stage. Overall, the total loss in explanatory power can be significant. In such circumstances, it is better to use the longer lag lengths.

This type of likelihood ratio test is applicable to any type of cross-equation restriction. Let Σ_u and Σ_r be the variance/covariance matrices of the unrestricted and restricted systems, respectively. If the equations of the unrestricted model contain different regressors, let c denote the maximum number of regressors contained in the longest equation. Sims's recommendation is to compare the test statistic

$$(T - c)(\ln|\Sigma_r| - \ln|\Sigma_u|) \tag{5.44}$$

2134904 - Wiley US ©