Chapter 8

VECTOR AUTOREGRESSION TECHNIQUES

This chapter discusses econometric techniques for vector autoregressions (VAR). In most cases, the variables in VAR are assumed to be stationary.[1]

Let $\mathbf{y}_t$ be an n-dimensional vector stochastic process that is covariance stationary. Because $\mathbf{y}_t$ is covariance stationary, it has a Wold representation:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\epsilon}_{t-2} + \cdots = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_t, \tag{8.1}$$

where $\boldsymbol{\Psi}(L) = \mathbf{I}_n + \sum_{s=1}^{\infty} \boldsymbol{\Psi}_s L^s$ and L is the lag operator. Assuming that $\boldsymbol{\Psi}(L)$ is invertible, $\mathbf{y}_t$ has a VAR representation. Assuming that the VAR representation is of order p:

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{\delta}_\epsilon + \boldsymbol{\epsilon}_t, \tag{8.2}$$

[1] A VAR model may include nonstationary variables. Chapter 16 treats the case where some of the variables in VAR are difference stationary and cointegrated, terms that will be introduced later. When the difference stationary variables are not cointegrated, we can take the first difference to make them stationary for VAR.

where

$$\begin{aligned} \boldsymbol{\delta}_\epsilon &= \boldsymbol{\Psi}(1)^{-1}\boldsymbol{\mu} = \mathbf{A}(1)\boldsymbol{\mu}, \\ \mathbf{A}(L) &= \boldsymbol{\Psi}(L)^{-1} = \mathbf{I}_n - \sum_{i=1}^{p} \mathbf{A}_i L^i, \\ \boldsymbol{\epsilon}_t &= \mathbf{y}_t - \hat{E}(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \mathbf{y}_{t-3}, \cdots) \end{aligned} \tag{8.3}$$

and

$$E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \boldsymbol{\Sigma}_\epsilon. \tag{8.4}$$

Here $\hat{E}(\cdot|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \mathbf{y}_{t-3}, \cdots)$ is defined to be the linear projection operator onto the linear space spanned by a constant (say, 1) and $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \mathbf{y}_{t-3}, \cdots$. In virtually all applications, $\boldsymbol{\Sigma}_\epsilon$ is not diagonal. However, the Seemingly Unrelated Regression Estimator (SUR) coincides with the OLS estimator for (8.2) because the regressors are identical for all regressions when OLS is applied to each row of (8.2).

8.1 OLS Estimation

The VAR (8.2) gives a system of regression equations. It may appear that the SUR estimator should be used to estimate these equations because the error terms are contemporaneously correlated. However, the OLS and SUR estimators coincide because the regressors are the same for all equations. Hence, we can estimate each equation by OLS.

It is often convenient to use a matrix expression to write the OLS estimators for the VAR system. For this purpose, rewrite (8.2) by staking it from $t = 1, \cdots, T$ after transpose:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U} \tag{8.5}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_t' \\ \vdots \\ \mathbf{y}_T' \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 \;\; \mathbf{y}_{1-1}' \cdots \mathbf{y}_{1-p}' \\ \vdots \\ 1 \;\; \mathbf{y}_{t-1}' \cdots \mathbf{y}_{t-p}' \\ \vdots \\ 1 \;\; \mathbf{y}_{T-p}' \cdots \mathbf{y}_{T-p}' \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \boldsymbol{\delta}_\epsilon' \\ \mathbf{A}_1' \\ \vdots \\ \mathbf{A}_p' \end{bmatrix}, \text{ and } \mathbf{U} = \begin{bmatrix} \boldsymbol{\epsilon}_1' \\ \vdots \\ \boldsymbol{\epsilon}_t' \\ \vdots \\ \boldsymbol{\epsilon}_T' \end{bmatrix}.$$

In order to apply OLS techniques, express (8.5) in its vector form:

$$\mathbf{y} = (\mathbf{I}_n \otimes \mathbf{X})\mathbf{b} + \mathbf{u}, \tag{8.6}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{b} = \text{vec}(\mathbf{B})$, $\mathbf{u} = \text{vec}(\mathbf{U})$, and $E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma}_\epsilon \otimes \mathbf{I}_T$. Applying OLS techniques, we get

$$\hat{\mathbf{b}} = (\mathbf{I}_n \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \tag{8.7}$$

and

$$\text{var}(\hat{\mathbf{b}}) = \boldsymbol{\Sigma}_\epsilon \otimes (\mathbf{X}'\mathbf{X})^{-1}. \tag{8.8}$$

In many applications, we express the asymptotic variance in (8.8) using the notation $\mathbf{a} = \text{vec}(\boldsymbol{\delta}_\epsilon \; \mathbf{A}_1 \cdots \mathbf{A}_p)$. Let $\mathbf{K}_{rc}$ be the $rc \times rc$ dimensional commutation matrix that has the property of $\text{vec}(\mathbf{M}') = \mathbf{K}_{rc}\text{vec}(\mathbf{M})$ for any $r \times c$ matrix $\mathbf{M}$. Then, we can show that

$$\hat{\mathbf{a}} = \mathbf{K}_{(np+1)n}\hat{\mathbf{b}} \tag{8.9}$$

and

$$\text{var}(\hat{\mathbf{a}}) = (\mathbf{X}'\mathbf{X})^{-1} \otimes \boldsymbol{\Sigma}_\epsilon. \tag{8.10}$$

8.2 Granger Causality

Let $\mathbf{y}_t = (x_t, y_t)'$ be a two dimensional covariance stationary process. We say that y fails to *Granger-cause* x if for all $s > 0$,

$$\hat{E}(x_{t+s}|x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots) = \hat{E}(x_{t+s}|x_t, x_{t-1}, \cdots). \tag{8.11}$$

We also say that y is not *linearly informative* about future x, or x is exogenous in the time series sense with respect to y.

One can test the null hypothesis that y fails to Granger-cause x by applying the OLS to

$$x_t = \delta_{\epsilon_1} + a_{1,11}x_{t-1} + \cdots + a_{p,11}x_{t-p} + a_{1,12}y_{t-1} + \cdots + a_{p,12}y_{t-p} + \epsilon_{1t}. \tag{8.12}$$

If y fails to Granger-cause x, then $a_{i,12} = 0$ for $i = 1, \cdots, p$ in (8.12). Conversely, if $a_{i,12} = 0$ for $i = 1, \cdots, p$ in (8.12), then

$$\hat{E}(x_{t+1}|x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots) = \delta_{\epsilon_1} + a_{1,11}x_t + \cdots + a_{p,11}x_{t-p+1} \tag{8.13}$$

and

$$\begin{aligned} & \hat{E}(x_{t+2}|x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots) \\ = \ & \delta_{\epsilon_1} + a_{1,11}\hat{E}(x_{t+1}|x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots) + a_{2,11}x_t + \cdots + a_{p,11}x_{t-p+2}. \end{aligned} \tag{8.14}$$

Repeating this argument, we see that y fails to Granger-cause x. Hence we test the null hypothesis

$$H_0 : a_{i,12} = 0 \quad \text{for} \quad i = 1, \cdots, p \tag{8.15}$$

in (8.12) in order to test for Granger causality.

The result that y fails to Granger-cause x if and only if (8.15) holds in (8.12) can be used to find restrictions on the VAR representation for $\mathbf{y} = (x, y)'$. Suppose that y fails to Granger-cause x, and x Granger-causes y.[2] Let the VAR representation of $\mathbf{y}$ be given by (8.2). Then the restrictions (8.15) hold if and only if $\mathbf{A}_i$ is lower triangular for each i:

$$\mathbf{A}_i = \begin{bmatrix} a_{i,11} & 0 \\ a_{i,21} & a_{i,22} \end{bmatrix}. \tag{8.16}$$

Hence y fails to Granger-cause x, and x Granger-causes y if and only if the VAR representation for $\mathbf{y} = (x, y)'$ given by (8.2) satisfies the restrictions that $\mathbf{A}_i$ is lower triangular for each i as in (8.16).

Suppose that an econometrician finds evidence for the hypothesis that y fails to Granger-cause x, but x Granger-causes y (i.e., the null hypothesis that y fails to Granger-cause x cannot be rejected, but the null hypothesis that x fails to Granger-cause y can be rejected). For example, researchers have found some evidence that real GDP fails to Granger-cause the money supply, and the money supply Granger-causes real GDP. This type of finding is consistent with some economic models which predict that a decrease in the money supply causes real GDP to fall.

It should be noted, however, that Granger-causality relationships can be very different from causal relationships when economic variables respond to future expected values of other variables as in the rational expectations models. Hence Granger-causality test results must be interpreted with caution.

For example, consider the present value model of a stock price:

$$p_t = E(\sum_{i=1}^{\infty} b^i d_{t+i} | I_t). \tag{8.17}$$

[2] Having defined the meaning of "x fails to Ganger-cause y," we define "x Granger-causes y," to mean "x does not fail to Granger-cause y."

where p_t is the stock price and d_t is the dividend. In order to illustrate the point in a simple example, assume that

$$d_t = u_t + \delta u_{t-1} + v_t, \tag{8.18}$$

where u_t and v_t are normal i.i.d. and are independent of each other. Here, the mean of the log of the dividend is normalized to be zero. Then

$$E_t(d_{t+i}) = \begin{cases} \delta u_t & \text{for} \quad i = 1 \\ 0 & \text{for} \quad i > 1 \end{cases}, \tag{8.19}$$

which implies $p_t = b\delta u_t$. Therefore, $\delta u_{t-1} = b^{-1} p_{t-1}$. Hence, the VAR representation for $\mathbf{y}_t = (p_t, d_t)'$ is

$$\begin{bmatrix} p_t \\ d_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ b^{-1} & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ d_{t-1} \end{bmatrix} + \begin{bmatrix} b\delta u_t \\ u_t + v_t \end{bmatrix}. \tag{8.20}$$

Since the VAR coefficient matrix is lower triangular, the dividend fails to Granger-cause the stock price, and the stock price Granger-causes the dividend in this example.

Since the changes in the future expected dividends cause the stock price to change in the present value model, the causal relationship is the opposite of the Granger-causality relationship. This result occurs because the stock price responds to the future expected values of the dividends in the present value model. When future dividends are expected to rise, the current stock price rises. Hence, the stock price tends to move before the dividend moves. This result does not mean that the stock price causes the dividend to move, but can mean that the stock price Granger-causes the dividend as in the example. In this sense, Granger "causality" is a misnomer.[3] It is safer to interpret Granger causality test results in terms of linear informativeness.

[3] Leamer (1985) suggests to use the word "precedence" instead of "causality". He argues that what is tested in "Granger Causality" is whether one variable regularly precedes another and that "precedence" is not sufficient for causality.

An example with this interpretation is Stock and Watson's (1989) application to search for economic variables that forecast business cycle movements.

8.3 The Impulse Response Function

Consider a moving average representation

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}_0^* \boldsymbol{\epsilon}_t^* + \boldsymbol{\Psi}_1^* \boldsymbol{\epsilon}_{t-1}^* + \boldsymbol{\Psi}_2^* \boldsymbol{\epsilon}_{t-2}^* + \cdots = \boldsymbol{\mu} + \boldsymbol{\Psi}^*(L)\boldsymbol{\epsilon}_t^*.$$

Let y_{it} be the i-th element of $\mathbf{y}_t$, ϵ_{jt}^* be the j-th element of $\boldsymbol{\epsilon}_t^*$, and $\psi_{s,ij}^*$ be the (i,j)-th element of $\boldsymbol{\Psi}_s^*$. If ϵ_{jt}^* is increased by one unit while holding all the other elements of $\boldsymbol{\epsilon}_{t+\tau}^*$ constant for all positive and negative τ, then $y_{i,t+s}$ will increase by $\psi_{s,ij}^*$ for $s > 0$. In this sense,

$$\frac{\partial y_{i,t+s}}{\partial \epsilon_{jt}^*} = \psi_{s,ij}^*, \tag{8.21}$$

or, using matrix notation,

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\epsilon}_t^{*\prime}} = \boldsymbol{\Psi}_s^*, \tag{8.22}$$

A plot of $\psi_{s,ij}^*$ for $s = 1, 2, \cdots$ is the *impulse response function* of y_i with respect to ϵ_{jt}^*.

One convenient way to estimate the impulse response function is to choose the Wold representation (8.1):

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\epsilon}_{t-2} + \cdots = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_t,$$

estimate the VAR representation by applying OLS to each row of $\mathbf{y}_t$, and simulate the estimated VAR representation to obtain an estimate of $\boldsymbol{\Psi}_s$.

There exist two difficulties in interpreting the impulse response function. The first difficulty is that $\boldsymbol{\Sigma}_\epsilon = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t')$ is not diagonal. This property means that the other elements of $\boldsymbol{\epsilon}_t$ tend to move with ϵ_{jt} when ϵ_{jt} changes. Hence, it is not very meaningful to consider the effect of changes in ϵ_{jt} on $y_{i,t+s}$ while holding the other elements of $\boldsymbol{\epsilon}_t$ constant. Computing an orthogonalized impulse response function is one method to avoid this difficulty. We assume that $\boldsymbol{\Sigma}_\epsilon$ is positive definite. Then, given the ordering of variables in $\mathbf{y}_t$, there exists a unique lower triangular matrix $\boldsymbol{\Phi}_0$ with 1's along the principal diagonal and a unique diagonal matrix $\boldsymbol{\Lambda}$ with positive entries along the principal diagonal such that

$$\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Phi}_0 \boldsymbol{\Lambda} \boldsymbol{\Phi}_0'. \tag{8.23}$$

Let

$$\mathbf{e}_t = \boldsymbol{\Phi}_0^{-1} \boldsymbol{\epsilon}_t. \tag{8.24}$$

Then $E(\mathbf{e}_t \mathbf{e}_t') = \boldsymbol{\Phi}_0^{-1} \boldsymbol{\Sigma}_\epsilon (\boldsymbol{\Phi}_0^{-1})' = \boldsymbol{\Lambda}$ which is diagonal. Since

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Phi}_0 \mathbf{e}_t, \tag{8.25}$$

$\mathbf{y}_t$ has an MA representation in terms of $\mathbf{e}_t$:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_0 \mathbf{e}_t + \boldsymbol{\Psi}_1 \boldsymbol{\Phi}_0 \mathbf{e}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\Phi}_0 \mathbf{e}_{t-2} + \cdots = \boldsymbol{\mu} + \boldsymbol{\Phi}(L) \mathbf{e}_t, \tag{8.26}$$

where $\boldsymbol{\Phi}(L) = \sum_{s=0}^{\infty} \boldsymbol{\Phi}_s L^s$ and $\boldsymbol{\Phi}_s = \boldsymbol{\Psi}_s \boldsymbol{\Phi}_0$. Let e_{jt} be the j-th element of $\mathbf{e}_t$ and $\phi_{s,ij}$ be the (i,j)-th element of $\boldsymbol{\Phi}_s$. Then (8.26) implies that

$$\frac{\partial y_{i,t+s}}{\partial e_{jt}} = \phi_{s,ij}. \tag{8.27}$$

A plot of (8.27) as a function of $s \geq 0$ is an *orthogonalized impulse response function.*

The sample counterparts of $\boldsymbol{\Psi}_s$ and $\boldsymbol{\Phi}_0$ can be used to estimate the orthogonalized impulse response function. For example, the Cholesky factorization, which

GAUSS can be used to compute, of the estimate of $\mathbf{\Sigma}_\epsilon$ can be used to estimate $\mathbf{\Phi}_0$. If $\mathbf{P}$ is the Cholesky factorization of $\mathbf{\Sigma}_\epsilon$, then $\mathbf{P} = \mathbf{\Phi}_0\mathbf{\Lambda}^{\frac{1}{2}}$, and the principal diagonal of $\mathbf{P}$ is the principal diagonal of $\mathbf{\Lambda}^{\frac{1}{2}}$. Hence, $\mathbf{\Phi}_0 = \mathbf{P}\mathbf{\Lambda}^{-\frac{1}{2}}$. This formula can be used to construct a sample counterpart of $\mathbf{\Phi}_0$.

The second difficulty in interpreting the impulse response function is that it is not possible to interpret $\boldsymbol{\epsilon}_t$ or $\mathbf{e}_t$ as shocks to the economy without imposing any economic structure to the VAR representation. For example, if the first element in $\mathbf{y}_t$ is the money supply, it is tempting to interpret the first element of $\boldsymbol{\epsilon}_t$ as the money supply shock which represents random changes in the money supply. With this interpretation, one can learn about how endogenous variables respond to the money supply shock by examining the impulse response functions. However, without any economic model, $\boldsymbol{\epsilon}_t$ is simply the forecast error when the linear forecasting rule is used with the past values of $\mathbf{y}_t$ as the information set. In some linear rational expectations models, $\boldsymbol{\epsilon}_t$ is simply the difference between the economic agents' forecast and the linear forecast based on the past values of $\mathbf{y}_t$. When the economic agents use a nonlinear forecasting rule with a larger information set, their forecast can be very different from $\hat{E}(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \cdots)$. In these models, it is not clear what we learn from the impulse response functions. Section 8.5 will discuss structural models that provide economically meaningful shocks with various restrictions. Under the recursive assumptions introduced in Section 8.5, the orthogonalized impulse response function discussed above can be used to compute impulse response functions of the structural shocks. In the majority of the VAR applications, the recursive assumptions are used. Under other assumptions, alternative methods are used to compute impulse response functions for the structural shocks as explained below.

We provide three traditional methods of computing confidence intervals of impulse responses: asymptotic normal approximation (see, e.g., Lütkepohl, 1990), bootstrap (see, e.g., Runkle, 1987; Kilian, 1998), and Monte Carlo integration (see, e.g., Doan, 1992; Sims and Zha, 1999). All three methods are asymptotically valid in stationary models but not the same in small samples. Kilian (1998) shows from his Monte Carlo simulation that bootstrap-after-bootstrap method performs better than others in small samples, while Sims and Zha (1999) argue that the Bayesian intervals have a firmer theoretical foundation and show how to obtain correct intervals for over-identified models.

8.4 Forecast error decomposition

Denoting the h-step forecast error by

$$\begin{aligned} \mathbf{y}_{t+h} - \hat{E}_t \mathbf{y}_{t+h} &= \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s (\boldsymbol{\epsilon}_{t+h-s} - \hat{E}_t \boldsymbol{\epsilon}_{t+h-s}) \\ &= \sum_{s=0}^{h-1} \boldsymbol{\Psi}_s \boldsymbol{\epsilon}_{t+h-s}, \end{aligned} \tag{8.28}$$

the forecast error variance is computed from the diagonal components of

$$E(\mathbf{y}_{t+h} - \hat{E}_t \mathbf{y}_{t+h})^2 = \sum_{s=0}^{h-1} \boldsymbol{\Psi}_s \boldsymbol{\Sigma}_\epsilon \boldsymbol{\Psi}'_s. \tag{8.29}$$

In particular, the forecast error variance of the i-th variable, $y_{i,t+h}$, is defined by

$$\sum_{s=0}^{h-1} \boldsymbol{\Psi}_{s,i\cdot} \boldsymbol{\Sigma}_\epsilon \boldsymbol{\Psi}'_{s,i\cdot} \tag{8.30}$$

where $\boldsymbol{\Psi}_{s,i\cdot}$ denotes the i-th row of $\boldsymbol{\Psi}_s$.

The same two difficulties concerning the interpretation of the impulse response function exist for the forecast variance decomposition. As with the impulse response function, the recursive assumptions have been employed in many VAR applications so that the orthogonalized shocks $\mathbf{e}_t$ in (8.24) are structural shocks.

The contribution of orthogonalized shocks to forecast error variance of the h-step forecast is defined by the diagonal components of

$$\sum_{s=0}^{h-1} \mathbf{\Phi}_s \mathbf{\Lambda} \mathbf{\Phi}_s'. \tag{8.31}$$

In particular, the contribution of the j-th orthogonalized shock, e_j, to the forecast error variance of the i-th variable, $y_{i,t+h}$, is[4]

$$\sum_{s=0}^{h-1} (\phi_{s,ij})^2 d_{jj}, \tag{8.32}$$

where d_{jj} is the variance of the j-th orthogonalized shock. The sample counterparts of $\mathbf{\Phi}$ and d_{jj} can be used to estimate this contribution.

Finally, dividing (8.32) by (8.30) yields the fraction of the h-step forecast error variance of the i-th variable attributed to the j-th orthogonalized shock.

8.5 Structural VAR Models

This section discusses structural economic models in which the orthogonalized impulse response functions are meaningful. A class of structural models can be written in the following form of a structural dynamic model:

$$\mathbf{B}_0 \mathbf{y}_t = \boldsymbol{\delta} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{B}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{B}_p \mathbf{y}_{t-p} + \mathbf{e}_t \tag{8.33}$$

where $\mathbf{B}_i$ is a $n \times n$ matrix, and $\boldsymbol{\delta}$ is a $n \times 1$ vector. Here $\mathbf{B}_0$ is a nonsingular matrix of real numbers with 1's along its principal diagonal, and $\mathbf{e}_t$ is a stationary n-dimensional

[4] By virtue of the assumption that orthogonalized shocks are mutually uncorrelated, we can separate the contribution of each orthogonalized shock.

vector of random variables with $E(\mathbf{e}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \cdots) = \mathbf{0}$. This structural model is related to its reduced form with $\mathbf{e}_t = \mathbf{B}_0\boldsymbol{\epsilon}_t$, $\boldsymbol{\delta} = \mathbf{B}_0\boldsymbol{\delta}_\epsilon$, $\mathbf{B}_i = \mathbf{B}_0\mathbf{A}_i$ for $i = 1, \cdots, p$. In many applications, it is assumed that the shocks are mutually uncorrelated so that the covariance matrix of $\mathbf{e}_t$ is diagonal.

Example 8.1 Consider a model of money demand. Let m_t be the real money balance, m_t^d be the desired real money balance, and i_t be the nominal interest rate:

$$m_t^d = \beta_0 + \beta_1 i_t. \tag{8.34}$$

Suppose that the actual money holdings are slowly adjusted toward the desired level so that

$$m_t - m_t^d = \alpha(m_{t-1} - m_{t-1}^d) + e_t^d, \tag{8.35}$$

where $0 < \alpha < 1$, and e_t^d is a money demand shock. Substituting (8.34) into (8.35) yields

$$m_t = \beta_0(1-\alpha) + \alpha m_{t-1} + \beta_1 i_t - \alpha\beta_1 i_{t-1} + e_t^d. \tag{8.36}$$

Imagine that the central bank determines the money supply at date t so that i_t is at a desired level given by the right hand side of the following equation:

$$i_t = \gamma_0 + \gamma_1 m_{t-1} + \gamma_2 i_{t-1} + e_t^s, \tag{8.37}$$

where e_t^s is a money supply shock. Then, when we choose $(i_t, m_t)'$ as $\mathbf{y}_t$, this money demand model is of the form (8.33):

$$\begin{bmatrix} 1 & 0 \\ -\beta_1 & 1 \end{bmatrix} \begin{bmatrix} i_t \\ m_t \end{bmatrix} = \begin{bmatrix} \gamma_0 \\ \beta_0(1-\alpha) \end{bmatrix} + \begin{bmatrix} \gamma_2 & \gamma_1 \\ -\alpha\beta_1 & \alpha \end{bmatrix} \begin{bmatrix} i_{t-1} \\ m_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^s \\ e_t^d \end{bmatrix}. \tag{8.38}$$ ∎

In general, $\mathbf{B}_0$ in (8.33) is not diagonal because some variables in $\mathbf{y}_t$ are affected by other variables in $\mathbf{y}_t$ as well as lagged values of the variables in $\mathbf{y}_t$. In Example 8.1, m_t is affected by i_t as well as lagged values of i_{t-1} and m_{t-1}.

In many structural models, it is reasonable to assume that the covariance matrix of $\mathbf{e}_t$ is diagonal. In Example 8.1, e_t^d is the money demand shock and e_t^s is the money supply shock, and these shocks should be uncorrelated. In this case, the impulse response functions with respect to the elements of $\mathbf{e}_t$ can be interpreted without any problem and are of interest. We will assume that $\mathbf{\Lambda} = E(\mathbf{e}_t\mathbf{e}_t')$ is diagonal for the rest of this chapter.

When the reduced form VAR (8.2) is estimated, various restrictions can be imposed on $\mathbf{B}_0$ to compute the impulse response functions of $\mathbf{e}_t$. For example, suppose that $\mathbf{B}_0$ is known. Let $\mathbf{\Phi}_0 = \mathbf{B}_0^{-1}, \mathbf{\Psi}_s = \partial\mathbf{y}_{t+s}/\partial\boldsymbol{\epsilon}_t'$ be the impulse response function with respect to $\boldsymbol{\epsilon}_t$, and $\mathbf{\Phi}_{0,\cdot j}$ be the j-th column of $\mathbf{\Phi}_0$. By the same argument used for the orthogonalized impulse response function, $\mathbf{\Psi}_s\mathbf{\Phi}_{0,\cdot j}$ gives the impulse response function with respect to e_{jt}.

In most models, $\mathbf{B}_0$ is unknown. A restriction on $\mathbf{B}_0$ often used in applications is that it is a lower triangular matrix. Example 8.1 satisfies this restriction. In the example, i_t is determined by i_{t-1} and m_{t-1} and is not affected by m_t. Note that $\mathbf{B}_0$ would not be lower triangular if $\mathbf{y}_t$ were defined to be $(m_t, i_t)'$ rather than $(i_t, m_t)'$. Thus the order of the variables in $\mathbf{y}_t$ is important. In general, $\mathbf{B}_0$ is lower triangular when the model has a *recursive structure*: y_{1t} is determined when the past values of $\mathbf{y}_t$ are given, y_{2t} is determined by y_{1t} and the past values of $\mathbf{y}_t$, y_{3t} is determined by y_{1t}, y_{2t}, and the past values of $\mathbf{y}_t$.

When $\mathbf{B}_0$ is lower triangular, $\mathbf{B}_0^{-1}$ is a lower triangular matrix and has 1's along the principal diagonal. It is known that when a positive matrix $\mathbf{\Sigma}_\epsilon$ is given, there exists a unique lower triangular matrix $\mathbf{\Phi}_0$ which has ones along the principal diagonal such that $\mathbf{\Sigma}_\epsilon = \mathbf{\Phi}_0\mathbf{\Lambda}\mathbf{\Phi}_0'$. Hence $\mathbf{B}_0$ can be computed by the Cholesky factorization using $\mathbf{B}_0 = \mathbf{\Phi}_0^{-1}$. Thus, the standard method of computing the orthogonalized impulse response function yields the impulse response function with respect to $\mathbf{e}_t$ when $\mathbf{B}_0$ is lower triangular. On the other hand, when $\mathbf{B}_0$ is not lower triangular the Choleski decomposition cannot be used, and ML or GMM estimation is often used as discussed in Section 8.6.3.

8.6 Identification

In order to identify $\mathbf{B}_0$, we need at least n^2 restrictions. In most cases, we assume that structural shocks are mutually uncorrelated. This orthogonality condition implies the variance-covariance matrix of structural disturbances is diagonal and gives $\frac{n(n-1)}{2}$ restrictions. Second, we impose a normalization condition that the diagonal components of $\mathbf{B}_0$ are 1's, which yields n restrictions.[5] Structural VAR varies depending on how the additional $\frac{n(n-1)}{2}$ conditions are imposed for identification.

8.6.1 Short-Run Restrictions for Structural VAR

The simplest model originating with Sims (1980) assumes that $\mathbf{B}_0$ is lower triangular. This structure is called recursive assumptions. This gives $\frac{n(n-1)}{2}$ necessary conditions so that the model is just identified as shown below. Letting $\mathbf{\Phi}_0 = \mathbf{B}_0^{-1}$, it follows

[5]Instead, we can consider an alternative normalization condition that the variance-covariance matrix of structural disturbances is an identity matrix. This change does not affect the main results.

from $\mathbf{e}_t = \mathbf{B}_0\boldsymbol{\epsilon}_t$ that

$$\boldsymbol{\Phi}_0\boldsymbol{\Lambda}\boldsymbol{\Phi}_0' = \boldsymbol{\Sigma}_\epsilon, \tag{8.39}$$

where $\boldsymbol{\Phi}_0$ is also a lower triangular matrix. Let $\mathbf{P}$ be a lower triangular matrix of the Cholesky decomposition of $\boldsymbol{\Sigma}_\epsilon$ so that $\mathbf{P}\mathbf{P}' = \boldsymbol{\Sigma}_\epsilon$. From $\boldsymbol{\Phi}_0\boldsymbol{\Lambda}^{\frac{1}{2}} = \mathbf{P}$, it follows that

$$\boldsymbol{\Phi}_0 = \mathbf{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}, \tag{8.40}$$

where $\boldsymbol{\Lambda} = [diag(\mathbf{P})]^2$.

Typically, researchers decide the order of variables to use from the type of restrictions, but do not use a tightly specified economic model to derive these restrictions in applications. Instead, impulse responses estimated from recursive assumptions are compared with implications of economic models. Some researchers make a more explicit connection between estimated impulse responses and an economic model. Rotemberg and Woodford (1999) minimize a distance measure between impulse responses estimated from recursive assumptions and impulse responses implied by a monetary model by choosing parameters of the model. Their monetary model incorporates an optimum monetary policy rule that is similar to the rule proposed by Taylor (1993).

Blanchard and Watson (1986) consider the case where $\mathbf{B}_0$ is not lower triangular. As their four-variable model includes eight unknown parameters in $\mathbf{B}_0$, they use a priori theoretical and empirical information about the private sector behavior and policy reaction functions on two of the parameters, and impose four zero restrictions to achieve identification on the remaining six ($=\frac{n(n-1)}{2}$) unknown parameters. Given these restrictions, their model is just identified. From $\mathbf{e}_t = \mathbf{B}_0\boldsymbol{\epsilon}_t$ it follows that

$$\boldsymbol{\Lambda} = \mathbf{B}_0\boldsymbol{\Sigma}_\epsilon\mathbf{B}_0', \tag{8.41}$$

which yields unique solutions for $\mathbf{B}_0$ and $\mathbf{\Lambda}$. Gordon and Leeper (1994) use full information maximum likelihood estimation to study liquidity effects in their over-identified model. To identify their model, they impose conventional exclusion restrictions and plausible informational assumptions from a traditional view of monetary policy and private sector behavior, such as which variables enter demand and supply for the reserve market.

Bernanke (1986) considers a model that allows more than one structural shock in an equation. The structural form is

$$\mathbf{B}(L)\mathbf{y}_t = \mathbf{F}\mathbf{e}_t. \tag{8.42}$$

Assume that $\mathbf{B}_0$ is not lower triangular but that there are $\frac{n(n-1)}{2}$ unknown parameters in $\mathbf{B}_0$ and $\mathbf{F}$. From $\mathbf{F}\mathbf{e}_t = \mathbf{B}_0\boldsymbol{\epsilon}_t$ it follows that

$$\mathbf{\Lambda} = \mathbf{F}^{-1}\mathbf{B}_0\mathbf{\Sigma}_\epsilon\mathbf{B}_0'\mathbf{F}^{-1\prime}, \tag{8.43}$$

which yields the unique solutions for $\mathbf{B}_0$, $\mathbf{F}$ and $\mathbf{\Lambda}$.

8.6.2 Identification of block recursive systems

Christiano, Eichenbaum, and Evans (1999) provide a theoretical background and illustrate identification of block recursive systems. Partitioning $\mathbf{y}_t$ into three blocks is convenient to illustrate the block recursive structure:

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1t} \\ s_t \\ \mathbf{y}_{2t} \end{bmatrix}, \tag{8.44}$$

where $\mathbf{y}_t$ is a vector of $n(= n_1 + 1 + n_2)$ variables of interest, s_t is a monetary policy variable, $\mathbf{y}_{1t}$ includes n_1 variables which are in the information set when the Fed implements a monetary policy, and $\mathbf{y}_{2t}$ contains n_2 variables which are excluded from the information set. Alternatively, $\mathbf{y}_{1t}$ does not respond to a monetary policy shock contemporaneously, while $\mathbf{y}_{2t}$ does. The block recursive assumption imposes zero restrictions on the following partitioned $\mathbf{B}_0$:

$$\mathbf{B}_0 = \begin{bmatrix} \underset{(n_1 \times n_1)}{b_{11}} & \underset{(n_1 \times 1)}{0} & \underset{(n_1 \times n_2)}{0} \\ \underset{(1 \times n_1)}{b_{21}} & \underset{(1 \times 1)}{b_{22}} & \underset{(1 \times n_2)}{0} \\ \underset{(n_2 \times n_1)}{b_{31}} & \underset{(n_2 \times 1)}{b_{32}} & \underset{(n_2 \times n_2)}{b_{33}} \end{bmatrix} \tag{8.45}$$

Two zero restrictions, $b_{12} = b_{13} = 0$, are required for the monetary policy shock to be orthogonal to other structural shocks, while the restriction $b_{23} = 0$ implies the assumption that the Fed does not have information about variables in y_{2t} when it makes a monetary policy decision.

The following property may help explain the block recursive system:

$$\begin{bmatrix} \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}_{11}^{-1} & \mathbf{0} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} & \mathbf{B}_{22}^{-1} \end{bmatrix} \tag{8.46}$$

The block recursive structure gives sufficient conditions to identify a monetary policy shock, and the ordering within y_{1t} and y_{2t} does not affect the results if one is interested in the effects of a monetary policy shock. Instead, the ordering across two groups might affect the results substantially.

8.6.3 Two-step ML estimation

When $\mathbf{B}_0$ is not lower triangular, maximum likelihood estimation or GMM estimation can be used once the structural model is identified as discussed in the following section. As VAR models involve a large number of parameters, two-step estimation is often used. The reduced form VAR model is estimated in the first step, and ML or GMM estimation is used in the second step focusing on the relation of $\mathbf{B}_0\hat{\mathbf{\Sigma}}_\epsilon\mathbf{B}_0' = \mathbf{\Lambda}$ to estimate $\mathbf{B}_0$ and $\mathbf{\Lambda}$ from the first step estimate of $\mathbf{\Sigma}_\epsilon$. The two-step ML estimation is discussed by Giannini (1992) in detail, while two-step GMM estimation is used by Bernanke and Mihov (1998).

Suppose that the model is identified with short-run economic restrictions:

$$\text{vec}(\mathbf{B}_0) = \mathbf{S}_b\mathbf{b}_s + \mathbf{s}_b, \tag{8.47}$$

where $\mathbf{b}_s$ be a n_s ($\leq \frac{n(n+1)}{2}$) dimensional vector of free parameters in $\mathbf{B}_0$, and the restrictions are expressed by an $n^2 \times n_s$ matrix of $\mathbf{S}$ and $n^2 \times 1$ vector of $\mathbf{s}_b$.

Then the following are used in the second step for ML estimation:

(a) Likelihood function:

$$L(\mathbf{B}_0) = T \log |\mathbf{B}_0| - \frac{T}{2}\text{trace}(\mathbf{B}_0'\mathbf{B}_0\hat{\mathbf{\Sigma}}) \tag{8.48}$$

(b) Gradient:

$$\mathbf{g}(\mathbf{B}_0) = T[\text{vec}(\mathbf{B}_0'^{-1}) - (\hat{\mathbf{\Sigma}} \otimes \mathbf{I}_{n_2})\text{vec}(\mathbf{B}_0)] \tag{8.49}$$

(c) Information matrix:

$$\mathbf{I}_T(\mathbf{B}_0) = 2T(\mathbf{B}_0^{-1} \otimes \mathbf{I}_{n_2})\mathbf{N}_{n_2}(\mathbf{B}_0'^{-1} \otimes \mathbf{I}_{n_2}) \tag{8.50}$$

(d) Score algorithm:

$$\mathbf{b}_{s,i+1} = \mathbf{b}_{s,i} + [\mathbf{I}_T(\mathbf{b}_{s,i})]^{-1}\mathbf{g}(\mathbf{b}_{s,i}), \tag{8.51}$$

where $\mathbf{g}(\mathbf{b}_s) = \mathbf{S}_b'\mathbf{g}(\mathbf{B}_0)$, $\mathbf{I}_T(\mathbf{b}_s) = \mathbf{S}_b'\mathbf{I}_T(\mathbf{B}_0)\mathbf{S}_b$, and i denotes the iteration step.

In addition, if $\mathbf{B}_0$ is over-identified, the over-identifying restrictions can be tested using

$$LRT = 2(L(\hat{\mathbf{\Sigma}}) - L(\hat{\mathbf{B}}_{0,ML})), \tag{8.52}$$

where $L(\hat{\mathbf{\Sigma}}) = -\frac{T}{2}\log|\hat{\mathbf{\Sigma}}| - \frac{nT}{2}$, and LRT is asymptotically $\chi^2_{(q)}$-distributed, where q is the number of over-identification.

Appendix

This appendix provides three traditional methods of computing confidence intervals of impulse responses, which are widely used as standard tools for economic analysis in the applied VAR literature (see, e.g., Baillie, 1987; Runkle, 1987).

8.A Asymptotic Interval Method

Let $\boldsymbol{\theta} = (\mathbf{a}', \boldsymbol{\sigma}')'$, where $\mathbf{a} = \text{vec}(\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_p)$ and $\boldsymbol{\sigma} = \text{vech}(\boldsymbol{\Sigma})$.[6] It is well known that $\boldsymbol{\theta}$ is asymptotically normally distributed

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\theta),$$

where

$$\boldsymbol{\Sigma}_\theta = \begin{bmatrix} \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\sigma \end{bmatrix} = \begin{bmatrix} [E(\mathbf{x}_t\mathbf{x}_t')]^{-1} \otimes \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{D}_n^+(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{D}_n^{+\prime} \end{bmatrix},$$

$\mathbf{x}_t = \left[\mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \cdots, \mathbf{y}_{t-p}'\right]'$, and $\mathbf{D}_n^+$ is the Moore-Penrose inverse of $\mathbf{D}_n$. Refer to Hamilton (1994) for its derivation and extended discussion.

In addition to impulse responses derived in the text, it is often of interest to trace the accumulated responses

$$\boldsymbol{\Psi}_{ci} = \sum_{j=0}^{i} \boldsymbol{\Psi}_j, \quad \boldsymbol{\Phi}_{ci} = \boldsymbol{\Psi}_{ci}\boldsymbol{\Phi}_0$$

and the total accumulated responses

$$\boldsymbol{\Psi}(1) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j = \mathbf{A}(1)^{-1}, \quad \boldsymbol{\Phi}(1) = \boldsymbol{\Psi}(1)\boldsymbol{\Phi}_0.$$

[6] Note that we define $\mathbf{a}$ slightly differently from Section 8.1 which includes the constant term $\boldsymbol{\delta}_\epsilon$.

Let $\boldsymbol{\ell}_p$ be the p-dimensional vector with ones and denote

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix} \text{ and } \mathbf{J}_{np} = \begin{bmatrix} \mathbf{I}_n \vdots \mathbf{0}_{n\times n(p-1)} \end{bmatrix}.$$

Consider a VAR model with short-run restrictions of the form $\text{vec}(\mathbf{B}_0) = \mathbf{S}_b\mathbf{b}_s + \mathbf{s}_b$ and define $\mathbf{G}_{\phi\sigma} = \mathbf{G}_{\phi b}\mathbf{G}^{+}_{\phi\phi b}$ where $\mathbf{G}_{\phi b} = \left[-\mathbf{B}_0'^{-1} \otimes \mathbf{B}_0^{-1} \vdots \mathbf{I}_n \otimes \mathbf{B}_0^{-1}\right]\mathbf{S}_b$ and $\mathbf{G}_{\phi\phi b} = 2\mathbf{D}_n^{+}\left[-\mathbf{B}_0^{-1}\mathbf{B}_0'^{-1} \otimes \mathbf{B}_0^{-1} \vdots \mathbf{B}_0^{-1} \otimes \mathbf{B}_0^{-1}\right]\mathbf{S}_b$. With this notation, we obtain the asymptotic distributions of the impulse responses in the next proposition. See Lütkepohl (1990) for just-identified recursive VARs and Jang (2004) for more generalized VARs including non-recursive and over-identified models.

Proposition 8.A.1 *Suppose* $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$ *and* $\text{vec}(\mathbf{B}_0) = \mathbf{S}_b\mathbf{b}_s + \mathbf{s}_b$. *Then*

(a) $$\sqrt{T}\text{vec}(\hat{\boldsymbol{\Psi}}_i - \boldsymbol{\Psi}_i) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Psi ai}\boldsymbol{\Sigma}_a\mathbf{G}'_{\Psi ai}), \quad i = 1, 2, \cdots,$$

where

$$\mathbf{G}_{\Psi ai} = \frac{\partial \text{vec}(\boldsymbol{\Psi}_i)}{\partial \mathbf{a}'} = \sum_{j=0}^{i-1} \mathbf{J}_{np}(\mathbf{A}')^{i-1-j} \otimes \boldsymbol{\Psi}_j;$$

(b) $$\sqrt{T}\text{vec}(\hat{\boldsymbol{\Psi}}_{ci} - \boldsymbol{\Psi}_{ci}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Psi cai}\boldsymbol{\Sigma}_a\mathbf{G}'_{\Psi cai}), \quad i = 1, 2, \cdots,$$

where

$$\mathbf{G}_{\Psi cai} = \frac{\partial \text{vec}(\boldsymbol{\Psi}_{ci})}{\partial \mathbf{a}'} = \sum_{j=0}^{i} \mathbf{G}_{\Psi aj};$$

(c) $$\sqrt{T}\text{vec}(\hat{\boldsymbol{\Psi}}(1) - \boldsymbol{\Psi}(1)) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Psi 1a}\boldsymbol{\Sigma}_a\mathbf{G}'_{\Psi 1a})$$

where

$$\mathbf{G}_{\Psi 1a} = \frac{\partial \text{vec}(\boldsymbol{\Psi}(1))}{\partial \mathbf{a}'} = \boldsymbol{\ell}_p{}' \otimes \boldsymbol{\Psi}(1)' \otimes \boldsymbol{\Psi}(1);$$

(d) $\sqrt{T}\text{vec}(\hat{\mathbf{\Phi}}_i - \mathbf{\Phi}_i) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Phi ai}\mathbf{\Sigma}_a\mathbf{G}'_{\Phi ai} + \mathbf{G}_{\Phi\sigma i}\mathbf{\Sigma}_\sigma\mathbf{G}'_{\Phi\sigma i}), \quad i = 0, 1, 2, \cdots,$

where

$$\begin{aligned} \mathbf{G}_{\Phi ai} &= \frac{\partial \text{vec}(\mathbf{\Phi}_i)}{\partial \mathbf{a}'} = \begin{cases} \mathbf{0}, & i = 0 \\ (\mathbf{\Phi}'_0 \otimes \mathbf{I}_n)\mathbf{G}_{\Psi ai}, & i = 1, 2, \cdots \end{cases} \quad \textit{and} \\ \mathbf{G}_{\Phi\sigma i} &= \frac{\partial \text{vec}(\mathbf{\Phi}_i)}{\partial \boldsymbol{\sigma}'} = (\mathbf{I}_{n_2} \otimes \mathbf{\Psi}_i)\mathbf{G}_{\phi\sigma}; \end{aligned}$$

(e) $\sqrt{T}\text{vec}(\hat{\mathbf{\Phi}}_{ci} - \mathbf{\Phi}_{ci}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Phi cai}\mathbf{\Sigma}_a\mathbf{G}'_{\Phi cai} + \mathbf{G}_{\Phi c\sigma i}\mathbf{\Sigma}_\sigma\mathbf{G}'_{\Phi c\sigma i}), \quad i = 0, 1, 2, \cdots,$

where

$$\begin{aligned} \mathbf{G}_{\Phi cai} &= \frac{\partial \text{vec}(\mathbf{\Phi}_{ci})}{\partial \mathbf{a}'} = \sum_{j=0}^{i} \mathbf{G}_{\Phi aj} \quad \textit{and} \\ \mathbf{G}_{\Phi c\sigma i} &= \frac{\partial \text{vec}(\mathbf{\Phi}_{ci})}{\partial \boldsymbol{\sigma}'} = \sum_{j=0}^{i} \mathbf{G}_{\Phi\sigma j}; \end{aligned}$$

(f) $\qquad \sqrt{T}\text{vec}(\hat{\mathbf{\Phi}}(1) - \mathbf{\Phi}(1)) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\Phi 1a}\mathbf{\Sigma}_a\mathbf{G}'_{\Phi 1a} + \mathbf{G}_{\Phi 1\sigma}\mathbf{\Sigma}_\sigma\mathbf{G}'_{\Phi 1\sigma}),$

where

$$\begin{aligned} \mathbf{G}_{\Phi 1a} &= \frac{\partial \text{vec}(\mathbf{\Phi}_i)}{\partial \mathbf{a}'} = (\mathbf{\Phi}'_0 \otimes \mathbf{I}_n)\mathbf{G}_{\Psi 1a} \quad \textit{and} \\ \mathbf{G}_{\Phi 1\sigma} &= \frac{\partial \text{vec}(\mathbf{\Phi}_i)}{\partial \boldsymbol{\sigma}'} = (\mathbf{I}_{n_2} \otimes \mathbf{\Psi}(1))\mathbf{G}_{\phi\sigma}. \end{aligned}$$

Proof (a)–(c) See Lütkepohl (1990) Proposition 1.
(d)–(f) See Jang (2004) Theorem 3.2.

8.B Bias-Corrected Bootstrap Method

Kilian (1998) suggests the following algorithm for the bias-corrected bootstrap (bootstrap after bootstrap) method:

1. Estimate the VAR(p) in equation (8.2) and generate 1000 bootstrap replications $\hat{\mathbf{a}}^*$ from

$$\hat{\mathbf{A}}(L)\mathbf{y}_t^* = \hat{\boldsymbol{\delta}}_\epsilon + \boldsymbol{\epsilon}_t^*,$$

using standard nonparametric bootstrap techniques.

2. Approximate the bias term $\boldsymbol{\lambda} = E(\hat{\mathbf{a}} - \mathbf{a})$ by $\boldsymbol{\lambda}^* = E^*(\hat{\mathbf{a}}^* - \hat{\mathbf{a}})$, which suggests $\hat{\boldsymbol{\lambda}} = \bar{\mathbf{a}}^* - \hat{\mathbf{a}}$ for the bias estimate where $\bar{\mathbf{a}}^*$ is the mean of the bootstrap sample of $\hat{\mathbf{a}}^*$.

3. Adjust $\hat{\mathbf{a}}$ for stationarity correction to avoid pushing stationary impulse responses into the nonstationary region.

   (i) Compute $m(\hat{\mathbf{a}})$, the modulus of the largest root of the companion matrix associated with $\hat{\mathbf{a}}$.

   (ii) If $m(\hat{\mathbf{a}}) \geq 1$, set $\tilde{\mathbf{a}} = \hat{\mathbf{a}}$ without any adjustments.

   (iii) Otherwise, construct the bias-corrected coefficient estimate $\tilde{\mathbf{a}} = \hat{\mathbf{a}} - \hat{\boldsymbol{\lambda}}$. If $m(\tilde{\mathbf{a}}) \geq 1$, let $\hat{\boldsymbol{\lambda}}_1 = \hat{\boldsymbol{\lambda}}$ and $\nu_1 = 1$. Define $\hat{\boldsymbol{\lambda}}_{j+1} = \nu_j \hat{\boldsymbol{\lambda}}_j$ and $\nu_{j+1} = \nu_j - 0.01$. Set $\tilde{\mathbf{a}} = \tilde{\mathbf{a}}_j$ after iterating on $\tilde{\mathbf{a}}_j = \hat{\mathbf{a}} - \hat{\boldsymbol{\lambda}}_j$ for $j = 1, 2, \cdots$ until $m(\hat{\mathbf{a}}) < 1$.

4. Substitute $\tilde{\mathbf{a}}$ for $\hat{\mathbf{a}}$ and generate 2000 new bootstrap replications $\hat{\mathbf{a}}^*$ from

$$\tilde{\mathbf{A}}(L)\mathbf{y}_t^* = \tilde{\boldsymbol{\delta}}_\epsilon + \boldsymbol{\epsilon}_t^*,$$

   using standard nonparametric bootstrap techniques.

5. Compute $\tilde{\mathbf{a}}^*$ from $\hat{\mathbf{a}}^*$ and $\hat{\boldsymbol{\lambda}}^*$ with the adjustment of $\hat{\mathbf{a}}^*$ for stationarity correction as described in Step 3.

6. Compute the α and $1 - \alpha$ percentile intervals of impulse responses generated with $\tilde{\mathbf{a}}^*$ and $\hat{\boldsymbol{\sigma}}^*$.

8.C Monte Carlo Integration

Consider the VAR system in the form of (8.5). Assuming that u_t is i.i.d. and normally distributed, Zellner (1971) finds that $\mathbf{\Sigma}_\epsilon$ follows the Normal-inverse Wishart posterior distribution, with the prior, $f(\mathbf{b}, \mathbf{\Sigma}_\epsilon) \sim |\mathbf{\Sigma}_\epsilon|^{-\frac{n+1}{2}}$:

$$\text{(8.C.1)} \qquad \mathbf{\Sigma}_\epsilon^{-1} \sim Wishart((T\hat{\mathbf{\Sigma}}_\epsilon)^{-1}, T) \quad \text{with given } \hat{\mathbf{\Sigma}}_\epsilon$$

and

$$\text{(8.C.2)} \qquad \mathbf{b} \sim N(\hat{\mathbf{b}}, \mathbf{\Sigma}_\epsilon \otimes (\mathbf{X}'\mathbf{X})^{-1}).$$

Doan (1992) and Sims and Zha (1999) suggest the following parametric Monte Carlo integration method for computing impulse responses:

1. Estimate (16.17) and let $\hat{\mathbf{b}}$ and $\hat{\mathbf{\Sigma}}$ be these estimates.
2. Let $\mathbf{A}$ be a lower triangular matrix of Choleski decomposition of $(\mathbf{X}'\mathbf{X})^{-1}$.
3. Let $\mathbf{S}^{-1}$ be a lower triangular matrix of Choleski decomposition of $\hat{\mathbf{\Sigma}}_\epsilon^{-1}$.
4. Generate $n \times T$ random numbers, $\mathbf{w}_b$, from the normal distribution, $N(0, \frac{1}{T})$.
5. Generate $(n(p-1)+r+1) \times n$ random numbers, $\mathbf{u}_b$, from the standard normal distribution, $N(0,1)$.
6. Let $\mathbf{r}_b = \mathbf{w}_b'\mathbf{S}^{-1}$, and get $\mathbf{\Sigma}_b^{-1} = \mathbf{r}_b'\mathbf{r}_b$.
7. Let $\mathbf{S}_b$ be a lower triangular matrix of Choleski decomposition of $\mathbf{\Sigma}_b$.
8. Let $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}_b$, in which $\mathbf{e}_b = \mathbf{A}\mathbf{u}_b\mathbf{S}_b'$. Then, $\mathbf{b} \sim N(\hat{\mathbf{b}}, \mathbf{\Sigma}_b \otimes (\mathbf{X}'\mathbf{X})^{-1})$.
9. Draw impulse responses, $\mathbf{ir}_b$, as described in Section 16.3.3.

10. Repeat 4 ∼ 9, B times, and calculate 95% upper and lower bands of impulse responses using

$$Upper = \frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b + 2\left(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b^2 - \left(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b\right)^2\right)^{\frac{1}{2}} \tag{8.C.3}$$

and

$$Lower = \frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b - 2\left(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b^2 - \left(\frac{1}{B}\sum_{b=1}^{B}\mathbf{ir}_b\right)^2\right)^{\frac{1}{2}}. \tag{8.C.4}$$

Exercises

8.1 Let y_t and m_t be detrended log GDP and log money supply, respectively. Assume that $\mathbf{z}_t = (y_t, m_t)'$ is a covariance stationary process with a p-th order VAR representation.

(**a**) Define the concept, "y fails to Granger-cause m".

(**b**) How do you test the hypothesis that log GDP fails to Granger-cause log money supply?

(**c**) Imagine that you find empirical evidence that y fails to Granger-cause m, and m Granger-causes y. Discuss why this evidence can be consistent with a model in which money is neutral in the short run (money is neutral when changes in the level of money supply cannot affect any real economic variable such as real GDP).

(**d**) Define the orthogonalized impulse response function. Let

$$\mathbf{B}_0\mathbf{z}_t = \boldsymbol{\delta} + \mathbf{B}_1\mathbf{z}_{t-1} + \mathbf{B}_2\mathbf{z}_{t-2} + \cdots + \mathbf{B}_p\mathbf{z}_{t-p} + \mathbf{e}_t \tag{8.E.1}$$

be a structural model for $\mathbf{z}_t$, where $\mathbf{B}_i$ is a $n \times n$ matrix, and $\boldsymbol{\delta}$ is a $n \times 1$ vector. Here $\mathbf{B}_0$ is a nonsingular matrix of real numbers with 1's along its principal diagonal, and $\mathbf{e}_t$ is a stationary n-dimensional vector of normally distributed i.i.d. random variables. Discuss conditions for $\mathbf{B}_0$ under which the orthogonalized impulse response function represents the effects of each element of $\mathbf{e}_t$ on $\mathbf{z}_{t+s}$.

8.2 True or False. Briefly explain your answers.

(**a**) OLS estimation is equivalent to SUR estimation for a reduced-form VAR model because the regressors are identical.

(**b**) OLS estimation is equivalent to SUR estimation for a structural-form VAR model because structural disturbances are uncorrelated.

(**c**) In a recursive VAR model, $e_1 = \epsilon_1$.

(**d**) In a recursive VAR model, impulse responses to e_1 are the same as those of ϵ_1.

(**e**) In a recursive VAR model, $e_n = \epsilon_n$.

(**f**) In a recursive VAR model, impulse responses to e_n are the same as those of ϵ_n.

References

Baillie, R. T. (1987): "Inference in Dynamic Models Containing 'Surprise' Variables," *Journal of Econometrics*, 35, 101–117.

Bernanke, B. S. (1986): "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–100.

Bernanke, B. S., and I. Mihov (1998): "Measuring Monetary Policy," *Quarterly Journal of Economics*, 113(3), 869–902.

Blanchard, O. J., and M. W. Watson (1986): "Are Business Cycles All Alike?," in *The American Business Cycle: Continuity and Change*, ed. by R. J. Gordon, vol. 25 of *National Bureau of Economic Research Studies in Business Cycles*, pp. 123–182. University of Chicago Press, Chicago.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1, chap. 2, pp. 65–148. Elsevier Science.

Doan, T. A. (1992): *RATS User's Manual, Version 4*. Estima, Evanston, IL.

Giannini, C. (1992): *Topics in Structural VAR Econometrics*. Apringer Verlag, New York.

Gordon, D. B., and E. M. Leeper (1994): "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy*, 102(6), 1228–1247.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Jang, K. (2004): "Generalized Two-Step Maximum Likelihood Estimation of Structural Vector Autoregressive Models Partially Identified with Short-Run restrictions," Manuscript.

Kilian, L. (1998): "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80(2), 218–230.

Leamer, E. (1985): "Vector Autoregressions for Causal Inference," in *Carnegie-Rochester Conference Series On Public Policy: Understanding Monetary Regimes*, ed. by K. Brunner, and A. H. Meltzer, vol. 22, pp. 255–304. Elsevier.

Lütkepohl, H. (1990): "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics*, 72(1), 116–125.

Rotemberg, J. J., and M. Woodford (1999): "Interest Rate Rules in an Estimated Sticky Price Model," in *Monetary Policy Rules*, ed. by J. B. Taylor, vol. 31 of *NBER-Business Cycles Series*, chap. 2, pp. 57–126. The University of Chicago Press, Chicago, With Comment by Martin Feldstein.

Runkle, D. E. (1987): "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, 5, 437–442.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Sims, C. A., and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1156.

Stock, J. H., and M. W. Watson (1989): "New Indexes of Leading and Coincident Economic Indicators," in *NBER Macroeconomics Annual*, pp. 351–394.

Taylor, J. B. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.