9 Estimation Subject to Short-Run Restrictions

This chapter illustrates the estimation of structural VAR models subject to short-run identifying restrictions. A variety of estimation methods have been proposed to estimate these models, including the method of moments, instrumental variable (IV) estimation, full information maximum likelihood (FIML) estimation, and Bayesian estimation.

9.1 Model Setup

We consider the K-dimensional structural-form VAR(p) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t = BY_{t-1} + w_t, \tag{9.1.1}$$

where $Y'_{t-1} \equiv (y'_{t-1}, \ldots, y'_{t-p})$, $B \equiv [B_1, \ldots, B_p]$, and $w_t \sim (0, \Sigma_w)$. All deterministic terms are neglected because they are of no importance for the following discussion. It is straightforward to add these terms, as needed, in the derivations of this chapter. The structural errors have a diagonal covariance matrix Σ_w and are serially uncorrelated.

The corresponding reduced form is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = AY_{t-1} + u_t, \tag{9.1.2}$$

where $A \equiv [A_1, \ldots, A_p] = B_0^{-1} B$ and $u_t = B_0^{-1} w_t \sim (0, \Sigma_u)$ is a white noise error term with positive definite covariance matrix $\Sigma_u = B_0^{-1} \Sigma_w B_0^{-1\prime}$.

For the purpose of this chapter it is sometimes useful to assume in addition that $\Sigma_w = I_K$. Alternatively, it is sometimes assumed that the diagonal elements of B_0 are standardized to one and that the covariance matrix of w_t is a diagonal matrix such that

$$B_0 = \begin{bmatrix} 1 & b_{12,0} & \ldots & b_{1K,0} \\ b_{21,0} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{K-1,K,0} \\ b_{K1,0} & \ldots & b_{K,K-1,0} & 1 \end{bmatrix}$$

 https://doi.org/10.1017/9781108164818.010

and

$$\Sigma_w = \begin{bmatrix} \sigma^2_{w_1} & & 0 \\ & \ddots & \vdots \\ 0 & \dots & \sigma^2_{w_K} \end{bmatrix}.$$

Which assumption is preferred depends on the nature of the identifying assumptions, which in turn restricts the choice of the estimation method. As we illustrate in this chapter, some estimation methods are designed for models with identifying restrictions on B_0, whereas others are designed for estimating models with identifying restrictions on B_0^{-1}. There also are estimation methods that allow for combinations of restrictions on B_0 and B_0^{-1}.

9.2 Method-of-Moments Estimation

Many structural VAR models in the empirical literature are just identified. For such models there are a number of two-step estimation methods that may be viewed as alternative implementations of the method of moments. The method-of-moments approach to estimating B_0^{-1} is based on the second-moment matrix of the VAR innovations, Σ_u, which may be expressed in terms of the structural model parameters as

$$\Sigma_u = B_0^{-1}\Sigma_w B_0^{-1\prime}, \tag{9.2.1}$$

where Σ_w is diagonal by assumption. Because there are $K(K+1)/2$ unique elements in Σ_u, expression (9.2.1) may be viewed as defining $K(K+1)/2$ moment conditions involving the unknown parameters of B_0^{-1} and Σ_w. After replacing Σ_u by a consistent estimate $\widehat{\Sigma}_u$ and imposing suitable normalizing and identifying restrictions on the elements of Σ_w and B_0^{-1} (or alternatively on B_0), one can uniquely solve for the remaining unknown structural parameters.

In Section 9.2.1 we focus on recursively identified structural models with a lower-triangular structure for B_0^{-1}, followed by structural models identified based on nonrecursive short-run identifying restrictions in Section 9.2.2. Extensions to overidentified models, which may be estimated by the generalized method of moments (GMM), are discussed in Section 9.2.3.

9.2.1 Recursively Identified Models

The simplest approach to estimating a recursively identified model employs the Cholesky decomposition of $\widehat{\Sigma}_u$. Consider the following example. A popular argument in macroeconomics has been that oil price shocks in particular may act as domestic supply shocks for the U.S. economy. Thus, the question of how oil price shocks affect U.S. real GDP and inflation has a long tradition in macroeconomics (see, e.g., Blanchard 2002; Barsky and Kilian 2004; Kilian 2008c). We address this question by postulating a VAR(4) model with intercept for the percent changes in the real WTI price of crude oil ($\Delta rpoil_t$), the U.S. GDP deflator inflation rate (Δp_t), and U.S. real GDP growth (Δgdp_t). The data are quarterly and the estimation period is 1987q1-2013q2. The model is identified recursively with the real price of oil ordered first such that the real price of oil is predetermined with respect to the U.S. economy. We focus on the effect of an unanticipated increase in the real price of oil on the real price of oil, on the change in inflation, and on U.S. real GDP growth. The model is partially identified in that only the oil price shock can be given an economic interpretation.

Let $y_t = (\Delta rpoil_t, \Delta p_t, \Delta gdp_t)'$. The LS estimates of the reduced-form slope parameters are

$$\widehat{A}_1 = \begin{bmatrix} -0.0064 & 0.9365 & 1.7510 \\ 0.0018 & 0.5991 & 0.0282 \\ -0.0024 & -0.2187 & 0.3324 \end{bmatrix},$$

$$\widehat{A}_2 = \begin{bmatrix} -0.1822 & 12.6454 & -3.4391 \\ 0.0043 & 0.1276 & -0.0388 \\ -0.0063 & 0.3877 & 0.1459 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} -0.0073 & -3.5986 & 1.7337 \\ 0.0021 & 0.0477 & 0.0234 \\ -0.0002 & -0.1413 & -0.0439 \end{bmatrix},$$

$$\widehat{A}_4 = \begin{bmatrix} -0.1194 & -8.1807 & 0.9309 \\ 0.0007 & 0.1488 & 0.0706 \\ -0.0081 & -0.0812 & 0.0116 \end{bmatrix},$$

and the estimate of the reduced-form error covariance matrix is

$$\widehat{\Sigma}_u = \begin{bmatrix} 312.5246 & 0.7736 & 0.9193 \\ 0.7736 & 0.0515 & 0.0149 \\ 0.9193 & 0.0149 & 0.5570 \end{bmatrix}.$$

Estimating the Model Using the Cholesky Decomposition. If B_0^{-1} is lower triangular, as in this empirical example, the easiest way of estimating the structural VAR model involves two steps. First, we estimate the reduced-form VAR parameters and compute the residual variance-covariance matrix $\widehat{\Sigma}_u$. Then we estimate the structural impact multiplier matrix B_0^{-1} based on a lower-triangular Cholesky decomposition of the residual variance-covariance matrix $\widehat{\Sigma}_u$. Given the recursive structure of B_0^{-1} and the normalizing assumption that $\Sigma_w = I_K$, the system of nonlinear equations, $\Sigma_u = B_0^{-1}B_0^{-1\prime}$, with Σ_u replaced by $\widehat{\Sigma}_u$, implicitly defines the unknown elements of B_0^{-1}.

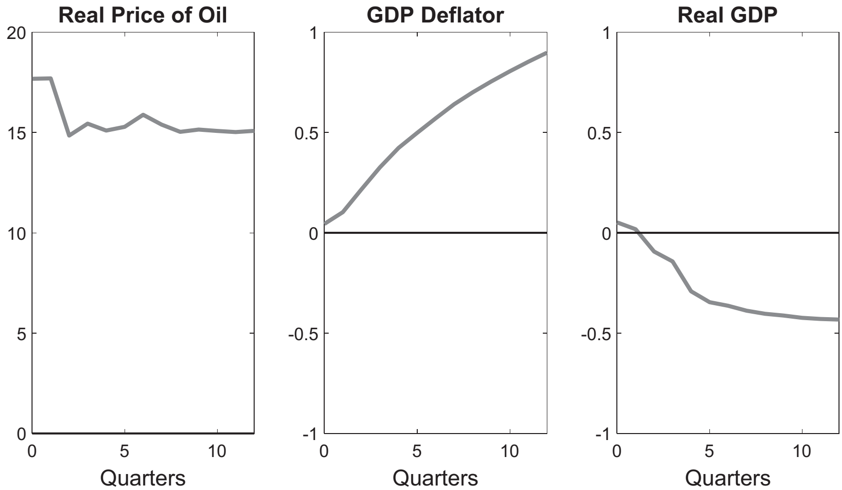

Figure 9.1. Responses of the U.S. economy to an unexpected increase in the real price of oil.

In the example above, this approach yields the estimate

$$\widehat{B}_0^{-1} = \text{chol}(\widehat{\Sigma}_u) = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix},$$

where chol denotes a function that returns the lower-triangular Cholesky decomposition such that $\widehat{B}_0^{-1}\widehat{B}_0^{-1\prime} = \widehat{\Sigma}_u$. Such functions are available in commonly used software. Note that the first column of $\widehat{B}_0^{-1}$ contains the responses of the three model variables in the impact period to an unexpected change in the real price of oil.

Given an estimate of B_0^{-1} and of the reduced-form parameters, it is straightforward to compute estimates of the structural impulse responses (see Chapter 4). Under the maintained assumption of predetermined oil prices, only the responses to the oil price innovation are economically identified. Figure 9.1 shows that an unexpected increase in the real price of crude oil is associated with an increase in the U.S. GDP deflator and a decline in U.S. real GDP.

For the construction of the structural impulse responses we require an estimate of B_0^{-1} rather than B_0. If an estimate of B_0 is of interest, such an estimate may be generated by inverting $\widehat{B}_0^{-1}$. It should be noted that B_0^{-1} being lower triangular implies that B_0 is lower triangular (and vice versa). Hence the approach described above may also be used without loss of generality when estimating models that impose a recursive structure on B_0.

Estimating the Model Using a Nonlinear Equation Solver. An alternative approach is to solve the system of nonlinear equations that implicitly defines the elements of $B_0^{-1} = [b_0^{ij}]$ using a nonlinear equation solver that finds the vector x such that $F(x) = 0$, where $F(x)$ denotes a system of nonlinear equations in x.[1] After normalizing $\Sigma_w = I_3$, while leaving the diagonal elements of B_0 unrestricted, we vectorize the system of equations

$$B_0^{-1} B_0^{-1\prime} - \widehat{\Sigma}_u = 0.$$

The objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \operatorname{vech}\left(B_0^{-1} B_0^{-1\prime} - \widehat{\Sigma}_u\right) \\ b_0^{12} \\ b_0^{13} \\ b_0^{23} \end{bmatrix} = 0, \tag{9.2.2}$$

where the vech operator is used to select the set of unique elements of $B_0^{-1} B_0^{-1\prime} - \widehat{\Sigma}_u$. The nonlinear equation solver iterates on expression (9.2.2) until convergence, given an initial guess for B_0^{-1}. It should be noted that the sign of the columns in $\widehat{B}_0^{-1}$ is not unique and may flip, depending on how this numerical procedure is initialized. In practice, an additional normalization may be required to match the results from the Cholesky decomposition, which imposes by default that the diagonal elements of $\widehat{B}_0^{-1}$ are positive (see also Taylor 2004).

For example, we may obtain the solution

$$\widehat{B}_0^{-1} = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & -0.2227 & 0 \\ 0.0520 & -0.0566 & 0.7424 \end{bmatrix},$$

which exactly matches that based on the Cholesky decomposition only after flipping the signs of the second column to ensure that all diagonal elements are positive. Indeed, in this simple example, there is no advantage to using a nonlinear equation solver, given the recursive structure of B_0^{-1}. The advantage of the nonlinear equation solver is that it can also accommodate nonrecursive structures, as discussed in Section 9.2.2.

Under the alternative normalization that Σ_w is diagonal with positive elements on the diagonal, while the diagonal of B_0 is restricted to a vector of ones, the structural impact multiplier matrix is $B_0^{-1} \Sigma_w^{1/2}$, where $\Sigma_w^{1/2}$ is obtained by

[1] An example of such a nonlinear equation solver is *fsolve* in the optimization toolbox of MATLAB.

taking the square root of the diagonal elements of Σ_w. We solve

$$\begin{bmatrix} \operatorname{vech}\left(B_0^{-1}\Sigma_w B_0^{-1\prime} - \widehat{\Sigma}_u\right) \\ b_{11,0}-1 \\ b_{22,0}-1 \\ b_{33,0}-1 \\ b_0^{12} \\ b_0^{13} \\ b_0^{23} \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \end{bmatrix} = 0 \tag{9.2.3}$$

for the unknown elements of B_0 and Σ_w, where $b_{ij,0}$ and b_0^{ij} denote the ij^{th} elements of B_0 and B_0^{-1}, respectively, and $\sigma_{ij,w}$ denotes the ij^{th} element of Σ_w. Expression (9.2.3) may be solved by iterating until convergence using a nonlinear equation solver, given initial guesses for B_0 and for Σ_w. The estimate of the structural impact multiplier matrix is the same as when using the Cholesky decomposition. One advantage of this alternative normalization is that the signs of the implied structural impact multiplier matrix $\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$ automatically match those in the lower triangular Cholesky decomposition.

If the identifying restrictions are imposed on the elements of B_0 rather than B_0^{-1}, it is simpler and more computationally attractive to solve

$$\begin{bmatrix} \operatorname{vech}\left(B_0\widehat{\Sigma}_u B_0' - \Sigma_w\right) \\ b_{11,0}-1 \\ b_{22,0}-1 \\ b_{33,0}-1 \\ b_{12,0} \\ b_{13,0} \\ b_{23,0} \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \end{bmatrix} = 0. \tag{9.2.4}$$

Because B_0^{-1} being lower triangular implies that B_0 is lower triangular, the numerical solution will be identical to that from expression (9.2.3).

The computational efficiency may be increased by directly imposing the restrictions that the off-diagonal elements of Σ_w are zero and that the diagonal elements of B_0 are unity, and only solving for the remaining elements in (9.2.3) and (9.2.4). In this case we only require an initial guess for the off-diagonal elements of B_0 and for the diagonal elements of Σ_w.

Estimating the Model Using the Algorithm of Rubio-Ramirez et al. (2010). Rubio-Ramírez, Waggoner, and Zha (2010) propose an alternative algorithm for VAR models with possibly nonrecursive structure that are exactly identified by short-run exclusion restrictions on the structural impulse responses. This algorithm avoids having to numerically solve a system of nonlinear equations by brute force.

Let L_0 denote an initial guess for the structural impact multiplier matrix B_0^{-1} such that $L_0 L_0' = \widehat{\Sigma}_u$. Possible choices include $L_0 = \widehat{\Sigma}_u^{1/2}$ or $L_0 = \text{chol}(\widehat{\Sigma}_u)$. Zero restrictions on B_0^{-1} can be represented by matrices Z_j for $1 \leq j \leq K$. The dimension of Z_j equals that of L_0. The $K \times K$ matrix Z_j summarizes the exclusion restrictions on the impact effect of structural shock j on variable $i = 1, \ldots, K$. The presence of an exclusion restriction is indicated by a 1 in Z_j and its absence by a zero. If there are no restrictions for a given structural shock j, $Z_j = 0_{K\times K}$. Exclusion restrictions on the impact effect of the i^{th} variable correspond to a 1 in the i^{th} column of a given row. There is at most one restriction per row. Restrictions are imposed starting with the first row of Z_j.

If shock j implies multiple restrictions, each of these exclusion restrictions is imposed in a different row, starting with the first variable to be restricted in the first row of Z_j, then the second variable to be restricted in the second row of Z_j, etc., in ascending order. A shock j that restricts the impact effect of the second and third variable, but not of the first variable, for example, would result in

$$Z_j = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If the rank of Z_j is z_j, then z_j is the number of zero restrictions associated with the j^{th} shock. The total number of zero restrictions on B_0^{-1} is $z = \sum_{j=1}^{K} z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j L_0 e_j = 0$ for $1 \leq j \leq K$. Here e_j denotes the j^{th} column of the identity matrix I_K. In what follows, let Z_j represent the zero restrictions with the equations of the VAR model ordered such that $z_j \leq K - j$, where $1 \leq j \leq K$. Observe that $Z_j L_0 e_j = 0$ and $\bar{Z}_j L_0 e_j = 0$ are equivalent statements, provided $\bar{Z}_j$ exists, where $\bar{Z}_j$ is defined by deleting all rows of zeros in Z_j.

An Algorithm for Solving Exactly Identified Models Based on Short-Run Restrictions.

1. Let $j = 1$.
2. If $j = 1$, then $Q_j' = \bar{Z}_1 L_0$. For $j > 1$, form the matrix

$$Q_j' = \begin{bmatrix} \bar{Z}_j L_0 \\ q_1' \\ \vdots \\ q_{j-1}' \end{bmatrix}.$$

3. There exists a vector q_j of unit length such that $Q_j' q_j = 0$. To find q_j such that $Q_j' q_j = 0$, use the QR decomposition $Q_j = \widetilde{Q}R$, where $\widetilde{Q}$ is orthogonal and R is upper triangular. Choose q_j to be the last column of $\widetilde{Q}$.
4. If $j = K$, stop. Otherwise set $j = j + 1$ and go to step 2.

If the model is exactly identified, this algorithm produces a $K \times K$ orthogonal rotation matrix

$$Q' = [q_1 \dots q_K]$$

such that $L_0 Q'$ represents a unique solution for B_0^{-1}, given the estimates of the reduced-form VAR model and the identifying restrictions. As in the case of estimates based on nonlinear equation solvers, it may be necessary to normalize the sign of the columns of this solution. It should be noted that Rubio-Ramírez, Waggoner, and Zha (2010) also propose a computationally more efficient algorithm designed specifically for triangular systems. We focus on the more general algorithm above because it also accommodates models with nonrecursive B_0^{-1}.

Applying this algorithm to our empirical example, consider the initial guess

$$L_0 = \widehat{\Sigma}_u^{1/2} = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ 0.0432 & 0.2225 & 0.0132 \\ 0.0499 & 0.0132 & 0.7446 \end{bmatrix},$$

where $\widehat{\Sigma}_u^{1/2}$ denotes the matrix square root of $\widehat{\Sigma}_u$. We define the restrictions to be imposed on L_0, starting with the structural shock that implies the most exclusion restrictions, which is the third structural shock in the original specification, followed by the structural shock that implies the second-largest number of exclusion restrictions, which is the second structural shock. Given the order of the variables in $y_t = (\Delta rpoil_t, \Delta p_t, \Delta gdp_t)'$, the first column of Z_j, $j = 1, 2, 3$, corresponds to the variable $\Delta rpoil_t$, the second column to Δp_t, and the third column to Δgdp_t. Hence,

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Z_3 = 0_{3\times 3}\,.$$

By deleting the rows of zeros from the Z_js, we obtain

$$\bar{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$
$$\bar{Z}_2 = (1, 0, 0).$$

Since all rows of Z_3 are zeros, there is no $\bar{Z}_3$.

For $j = 1$, we obtain

$$Q_1' = \bar{Z}_1 L_0 = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ 0.0432 & 0.2225 & 0.0132 \end{bmatrix}.$$

The QR decomposition for Q_1 yields

$$\widetilde{Q} = \begin{bmatrix} -1.0000 & 0.0026 & -0.0027 \\ -0.0024 & -0.9983 & -0.0585 \\ -0.0028 & -0.0585 & 0.9983 \end{bmatrix},$$

so $q_1' = \begin{bmatrix} -0.0027 & -0.0585 & 0.9983 \end{bmatrix}$, where q_1 is the last column of $\widetilde{Q}$. For $j = 2$, we obtain

$$Q_2' = \begin{bmatrix} \bar{Z}_2 L_0 \\ q_1' \end{bmatrix} = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ -0.0027 & -0.0585 & 0.9983 \end{bmatrix}.$$

The QR decomposition of Q_2 yields

$$\widetilde{Q} = \begin{bmatrix} -1.0000 & -0.0027 & -0.0026 \\ -0.0024 & -0.0585 & 0.9983 \\ -0.0028 & 0.9983 & 0.0585 \end{bmatrix},$$

and $q_2' = \begin{bmatrix} -0.0026 & 0.9983 & 0.0585 \end{bmatrix}$ is obtained from the last column of $\widetilde{Q}$. For $j = 3$, we obtain

$$Q_3' = \begin{bmatrix} q_1' \\ q_2' \end{bmatrix} = \begin{bmatrix} -0.0027 & -0.0585 & 0.9983 \\ -0.0026 & 0.9983 & 0.0585 \end{bmatrix},$$

because $\bar{Z}_3 L_0$ is the empty matrix. Finally, the QR decomposition of Q_3 yields

$$\widetilde{Q} = \begin{bmatrix} -0.0027 & 0.0026 & 1.0000 \\ -0.0585 & -0.9983 & 0.0024 \\ 0.9983 & -0.0585 & 0.0028 \end{bmatrix},$$

and $q_3' = \begin{bmatrix} 1.0000 & 0.0024 & 0.0028 \end{bmatrix}$ is obtained from the last column of $\widetilde{Q}$.

Upon completing this subroutine, we obtain the following solution for the restricted rotation matrix:

$$Q' = \begin{bmatrix} q_3 & q_2 & q_1 \end{bmatrix} = \begin{bmatrix} 1.0000 & -0.0026 & -0.0027 \\ 0.0024 & 0.9983 & -0.0585 \\ 0.0028 & 0.0585 & 0.9983 \end{bmatrix},$$

where the ordering of q_1, q_2, and q_3 has been reversed to match the ordering of the structural shocks in the original recursive model specification. The implied solution for the structural impact multiplier matrix is

$$L_0 Q' = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix}.$$

In this example there is no need to flip the signs of any of the columns to ensure that the diagonal elements of $\widehat{B}_0^{-1}$ are positive. We obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix},$$

which matches the lower-triangular Cholesky decomposition. Of course, there is no advantage over the use of the Cholesky decomposition in the recursive setting considered here, but this example is nevertheless instructive and helps prepare for more general applications of this algorithm to nonrecursive structural models, which cannot be estimated by applying a Cholesky decomposition.

9.2.2 Nonrecursively Identified Models

For expository purposes, we focus on a quarterly model of U.S. monetary policy due to Keating (1992) that has already been reviewed in Chapter 8. Let $y_t = (\Delta p_t, \Delta gnp_t, i_t, \Delta m_t)'$, where p_t refers to the log of the GNP deflator, gnp_t to the log of real GNP, i_t to the federal funds rate, averaged by quarter, and m_t to the log of M1. The VAR model includes four lags and an intercept. The estimation period is restricted to 1959q2-2007q4 in order to exclude the recent period of unconventional monetary policy measures. The unrestricted reduced-form estimates are

$$\widehat{A}_1 = \begin{bmatrix} 0.4885 & -0.0240 & 0.0693 & 0.0290 \\ 0.1695 & 0.1826 & 0.0124 & 0.0013 \\ 0.5739 & 0.3156 & 1.1325 & 0.0656 \\ 0.0003 & -0.0671 & -0.2889 & 0.1904 \end{bmatrix},$$

$$\widehat{A}_2 = \begin{bmatrix} 0.1255 & -0.0187 & -0.0543 & -0.0134 \\ 0.0885 & 0.2456 & -0.3570 & 0.0859 \\ 0.5174 & 0.2097 & -0.5362 & -0.0638 \\ 0.2851 & -0.0118 & 0.3221 & 0.3370 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} 0.1491 & 0.0187 & -0.0092 & 0.0173 \\ -0.3633 & -0.0011 & 0.3065 & -0.0240 \\ -0.3880 & 0.0201 & 0.5237 & 0.0620 \\ -0.3367 & 0.0909 & -0.1357 & 0.0702 \end{bmatrix},$$

$$\widehat{A}_4 = \begin{bmatrix} 0.1962 & 0.0607 & -0.0160 & 0.0041 \\ 0.1320 & 0.0341 & -0.0126 & -0.0135 \\ -0.2784 & -0.0327 & -0.1876 & -0.0102 \\ 0.1903 & 0.0846 & 0.1359 & -0.0013 \end{bmatrix},$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.0611 & -0.0153 & 0.0424 & 0.0038 \\ -0.0153 & 0.5230 & 0.0797 & 0.0306 \\ 0.0424 & 0.0797 & 0.7169 & -0.2451 \\ 0.0038 & 0.0306 & -0.2451 & 1.1093 \end{bmatrix}.$$

The structural shock vector $w_t = (w_t^{AS}, w_t^{IS}, w_t^{MS}, w_t^{MD})'$ includes an aggregate supply shock, an IS shock, a money supply shock, and a money demand shock. The structural model can be written as

$$\begin{pmatrix} u_t^p \\ u_t^{gnp} \\ u_t^i \\ u_t^m \end{pmatrix} = \begin{pmatrix} w_t^{AS} \\ -b_{21,0}u_t^p - b_{23,0}u_t^i - b_{24,0}u_t^m + w_t^{IS} \\ -b_{34,0}u_t^m + w_t^{MS} \\ -b_{41,0}(u_t^{gnp} + u_t^p) - b_{43,0}u_t^i + w_t^{MD} \end{pmatrix}.$$

As discussed earlier, the first equation represents a horizontal AS curve. The second equation can be interpreted as an IS curve, allowing real output to respond to all other model variables. The third equation represents a simple money supply function, according to which the central bank adjusts the rate of interest in relation to the money stock, and the fourth equation is a money demand function in which short-run money holdings rise in proportion to nominal income, yielding the final restriction required for exact identification.

The identifying restrictions may be summarized as

$$B_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ b_{21,0} & 1 & b_{23,0} & b_{24,0} \\ 0 & 0 & 1 & b_{34,0} \\ b_{41,0} & b_{41,0} & b_{43,0} & 1 \end{bmatrix}.$$

Note that the model imposes the normalizing assumption that the diagonal elements of B_0 equal unity and leaves the diagonal elements of Σ_w unrestricted, except for their positive sign. All identifying restrictions are imposed on B_0 rather than B_0^{-1}. Clearly, B_0 is not recursive, preventing the use of a Cholesky decomposition.

Of the method-of-moments estimation methods we discussed so far only the use of a nonlinear equation solver is practical in this nonrecursive example.

Estimating the Model Using a Nonlinear Equation Solver. The objective is to find the unknown elements of B_0 and Σ_w such that

$$\begin{bmatrix} \text{vech}\left(B_0 \widehat{\Sigma}_u B_0' - \Sigma_w\right) \\ b_{12,0} \\ b_{13,0} \\ b_{14,0} \\ b_{31,0} \\ b_{32,0} \\ b_{41,0} - b_{42,0} \\ b_{11,0} - 1 \\ b_{22,0} - 1 \\ b_{33,0} - 1 \\ b_{44,0} - 1 \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \\ \sigma_{41,w} \\ \sigma_{42,w} \\ \sigma_{43,w} \end{bmatrix} = 0, \tag{9.2.5}$$

where the elements on the diagonal of Σ_w are restricted to be positive. The nonlinear equation solver iterates on expression (9.2.5) until convergence, given initial guesses for B_0 and Σ_w. As noted earlier, the restrictions on the off-diagonal elements of Σ_w and on the diagonal elements of B_0 should be imposed directly to increase computational efficiency. The solution for B_0 is

$$\widehat{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.2669 & 1 & 0.7288 & 0.1784 \\ 0 & 0 & 1 & -11.2057 \\ -3.2443 & -3.2443 & 3.4133 & 1 \end{bmatrix},$$

$$\widehat{\Sigma}_w = \begin{bmatrix} 0.0611 & 0 & 0 & 0 \\ 0 & 0.9981 & 0 & 0 \\ 0 & 0 & 145.4997 & 0 \\ 0 & 0 & 0 & 10.6879 \end{bmatrix},$$

and the implied structural impact multiplier matrix is

$$\widehat{B}_0^{-1} \widehat{\Sigma}_w^{1/2} = \begin{bmatrix} 0.2471 & 0 & 0 & 0 \\ -0.0618 & 0.5912 & -0.0218 & -0.4114 \\ 0.1716 & 0.5476 & 0.2871 & 0.5524 \\ 0.0153 & 0.0489 & -1.0508 & 0.0493 \end{bmatrix}.$$

Figure 9.2 shows the implied responses to an unexpected upward shift of the aggregate supply curve. Given that the aggregate supply equation is normalized on the price level, the aggregate supply shock raises the price deflator and lowers real GNP. The effect on the federal funds rate is positive and the effect on M1 is ultimately positive. Of course, this example is just an illustration.

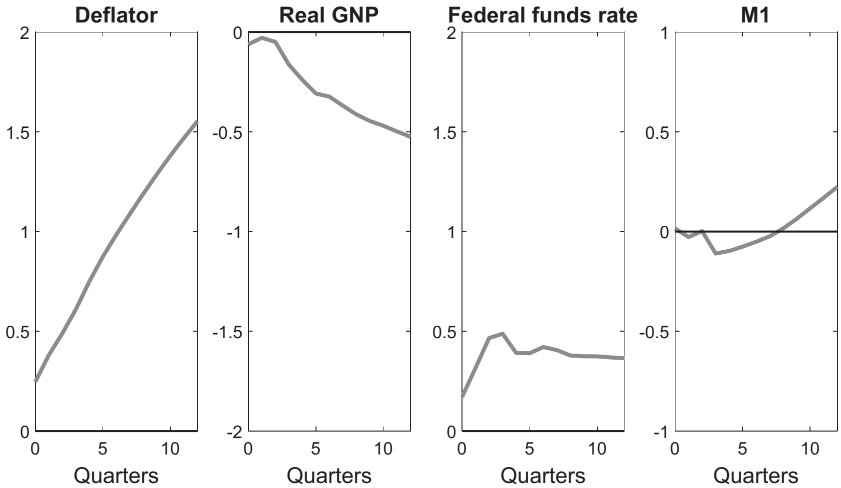

Figure 9.2. Responses of the U.S. economy to an aggregate supply shock in the Keating (1992) model.

It may not be appropriate for the U.S. economy during the sample period in question.

9.2.3 GMM Estimation of Overidentified Models

So far we have focused on models that are just identified. There are situations, however, in which there are more identifying restrictions on B_0^{-1} (or B_0) than are necessary to achieve identification. A common approach to estimating such models is to rely on the generalized method-of-moments (GMM) estimator originally proposed by Hansen (1982). Because there are more moment conditions than unknown parameters, the method-of-moments estimator cannot satisfy all moment conditions at the same time. Instead, the objective is to search for the structural parameter values that minimize a weighted average of the moment conditions where the asymptotically optimal set of weights corresponds to the inverse of the variance-covariance matrix of the sample moment conditions. This GMM estimator may be implemented in two steps, much like the method-of-moments estimator discussed earlier. The first step is to obtain consistent estimates of the reduced-form VAR model. The second step is to equate the sample second moment of the reduced-form VAR innovations with the same second moment written as function of the structural parameters in B_0^{-1}. Building on Hansen (1982), Bernanke and Mihov (1995) show that this GMM estimator is consistent and asymptotically normal. It is also efficient even without accounting for the uncertainty about the estimates of the variance-covariance matrix of the VAR residuals underlying the weighting matrix (see also Watson 1994).

Restrictions on B_0^{-1}**.** The relevant moment condition for estimating B_0^{-1} is

$$\mathbb{E}\left(\mathrm{vech}(u_t u_t') - \mathrm{vech}(B_0^{-1} B_0^{-1\prime})\right) = 0.$$

where the normalizing assumption $\Sigma_w = I_K$ is imposed. Using the empirical counterpart of the expectation, we have

$$\mathrm{vech}\left(\frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t' - B_0^{-1} B_0^{-1\prime}\right) = \mathrm{vech}(\widetilde{\Sigma}_u) - \mathrm{vech}(B_0^{-1} B_0^{-1\prime}).$$

This means that we are searching for B_0^{-1} such that this expression is as close to zero as possible. We therefore choose the unknown elements of B_0^{-1} to minimize

$$J = T(\mathrm{vech}(\widetilde{\Sigma}_u) - \mathrm{vech}(B_0^{-1} B_0^{-1\prime}))' \widehat{W} (\mathrm{vech}(\widetilde{\Sigma}_u) - \mathrm{vech}(B_0^{-1} B_0^{-1\prime})),$$

where all identifying and overidentifying restrictions have been imposed on B_0^{-1} and $\widehat{W}$ is the inverse of the estimate of the variance-covariance matrix of the sample moment restrictions. More specifically,

$$\widehat{W} = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{vech}(\widehat{u}_t \widehat{u}_t') - \overline{\mathrm{vech}(\widehat{u}\widehat{u}')}\right)\left(\mathrm{vech}(\widehat{u}_t \widehat{u}_t') - \overline{\mathrm{vech}(\widehat{u}\widehat{u}')}\right)'\right]^{-1}$$

where

$$\overline{\mathrm{vech}(\widehat{u}\widehat{u}')} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{vech}(\widehat{u}_t \widehat{u}_t').$$

Implementing the GMM estimator for overidentified models thus necessarily requires iteration. An empirical example for this GMM estimator can be found in Chapter 12.

Hansen (1982) shows that under the null hypothesis that the overidentifying restrictions are correct, under general conditions,

$$\hat{J} \xrightarrow{d} \chi^2_n,$$

where n is the number of overidentifying restrictions and $\hat{J}$ is the J-statistic evaluated at $\widehat{B}_0^{-1}$. This result allows a formal test of overidentifying restrictions conditional on the premise that the remaining restrictions are correct.

Restrictions on B_0**.** If identifying restrictions are imposed on B_0 rather than B_0^{-1}, it is possible to employ an alternative GMM estimation algorithm. This alternative estimator requires the user to take a stand on which restrictions are exactly identifying the structural model and which are overidentifying. We start with the exactly identifying restrictions and write the method-of-moments estimator in terms of the structural VAR representation.

For expository purposes, let us for the moment assume that we are dealing with a VAR(0) model such that $B_0 y_t = w_t$, and that the model can be exactly identified using only exclusion restrictions.[2] Then the k^{th} equation of the structural-form VAR model can be written as

$$y_{kt} = \mathbf{y}'_{kt} b_k + w_{kt},$$

where $\mathbf{y}_{kt}$ is the column vector of the elements of y_t that appear on the right-hand side of the k^{th} equation, where y_{kt} is the left-hand side variable in the k^{th} equation, and where b_k is the vector of associated structural parameters. For example, if B_0 is a lower-triangular matrix, there are no parameters to estimate in the first equation. For the remaining equations, $k = 2, \ldots, K$, we have $\mathbf{y}'_{kt} = (y_{1t}, \ldots, y_{k-1,t})$ and $b_k = (-b_{k1,0}, \ldots, -b_{k,k-1,0})'$.

For $t = 1, \ldots, T$, we obtain in matrix notation

$$\mathbf{y}_k = Y_k b_k + \mathbf{w}_k, \tag{9.2.6}$$

where $\mathbf{y}_k \equiv (y_{k1}, \ldots, y_{kT})'$, $Y_k \equiv [\mathbf{y}_{k1}, \ldots, \mathbf{y}_{kT}]'$ and $\mathbf{w}_k \equiv (w_{k1}, \ldots, w_{kT})'$. Assuming that the k^{th} equation is identified such that all right-hand side regressors are uncorrelated with w_{kt}, we can use the moment conditions

$$\mathbb{E}(\mathbf{y}_{kt} w_{kt}) = 0$$

to estimate b_k. In other words, $\widehat{b}_k$ minimizes the objective function

$$J(b_k) = (\mathbf{y}_k - Y_k b_k)'(\mathbf{y}_k - Y_k b_k).$$

To incorporate the overidentifying restrictions into the estimation requires a modification of this objective function. If the overidentifying restrictions are exclusion restrictions, we can express the linear restrictions on the parameters in the k^{th} equation as

$$b_k = R_k \gamma_k, \tag{9.2.7}$$

where R_k is a fixed given matrix that specifies the restrictions and γ_k is the vector of unrestricted parameters. For example, if the overidentifying restriction involves a zero restriction on the first parameter in the third equation of B_0 in a lower-triangular model, we have

$$R_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \gamma_3 = -b_{32,0}.$$

[2] The estimation method must be modified if the exactly identifying restrictions involve nonexclusion restrictions. This complicates the exposition. For illustrative purposes we focus on the more common case of exclusion restrictions.

As another example, if the overidentifying restriction forces to zero the first parameter in the second equation of the lower-triangular model, there are no parameters left to be estimated in that equation.

Overidentifying restrictions on B_0 other than exclusion restrictions may be imposed as well. For expository purposes, suppose that there are further linear restrictions on the parameters in the k^{th} equation. Such restrictions can also be expressed in the form of equation (9.2.7). For example, if we wish to impose in a four-dimensional model that $b_{41,0} = b_{42,0}$,

$$R_4 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \gamma_4 = \begin{pmatrix} -b_{41,0} \\ -b_{43,0} \end{pmatrix}.$$

Having defined the matrix R_k, let $Y_k^\dagger = Y_k R_k$ and estimate γ_k in the regression equation

$$\mathbf{y}_k = Y_k^\dagger \gamma_k + \mathbf{w}_k.$$

Because we have more moment conditions than parameters in the k^{th} equation, we estimate γ_k by minimizing the GMM objective function

$$\begin{aligned} J(\gamma_k) &= \left(\frac{1}{T}\mathbf{w}_k' Y_k\right)\left(\frac{1}{T}Y_k' Y_k\right)^{-1}\left(\frac{1}{T}Y_k'\mathbf{w}_k\right) \\ &= \left(\frac{1}{T}(\mathbf{y}_k - Y_k^\dagger\gamma_k)' Y_k\right)\left(\frac{1}{T}Y_k' Y_k\right)^{-1}\left(\frac{1}{T}Y_k'(\mathbf{y}_k - Y_k^\dagger\gamma_k)\right). \end{aligned}$$

This GMM estimator has the closed-form solution

$$\widehat{\gamma}_k^{GMM} = (\widehat{Y}_k^{\dagger\prime} Y_k^\dagger)^{-1}\widehat{Y}_k^{\dagger\prime}\mathbf{y}_k, \tag{9.2.8}$$

where $\widehat{Y}_k^\dagger \equiv Y_k(Y_k'Y_k)^{-1}Y_k'Y_k^\dagger$. As is well known, if the errors are iid, this GMM estimator is efficient.

So far we have only discussed the estimation of the k^{th} equation of the structural VAR model. Since the components of w_t are instantaneously uncorrelated, single-equation GMM is identical to estimating the full system

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \begin{bmatrix} Y_1^\dagger & & 0 \\ & \ddots & \\ 0 & & Y_K^\dagger \end{bmatrix}\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_K \end{pmatrix} + \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{pmatrix}$$

by GMM. Note that there may be equations without any parameters to estimate. Such equations are dropped from the system.

In practice, we typically do not deal with VAR(0) processes, and this procedure must be adapted to allow for lagged variables. In that case the k^{th} equation becomes

$$y_{kt} = \mathbf{y}_{kt}' R\gamma_k + Y_{t-1}'\mathbf{b}_k + w_{kt}, \tag{9.2.9}$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$, as before, and $\mathbf{b}'_k$ is the k^{th} row of $B = [B_1, \dots, B_p]$. Defining $Z \equiv [Y_0, \dots, Y_{T-1}]$, the same analysis as for the VAR(0) case can be applied to

$$\mathbf{y}_k = [Y_k^\dagger, Z'] \begin{pmatrix} \gamma_k \\ \mathbf{b}_k \end{pmatrix} + \mathbf{w}_k. \tag{9.2.10}$$

The GMM estimator of the parameters then becomes

$$\begin{pmatrix} \widehat{\gamma}_k^{GMM} \\ \widehat{\mathbf{b}}_k^{GMM} \end{pmatrix} = \left(\widehat{Z}_k [Y_k^\dagger, Z'] \right)^{-1} \widehat{Z}_k \mathbf{y}_k,$$

where

$$\widehat{Z}_k{}' = [Y_k, Z'] \left(\begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k, Z'] \right)^{-1} \begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k^\dagger, Z'].$$

This setup facilitates not only the computation of the estimates, but also the derivation of the asymptotic distribution. Under conditions that are satisfied for stationary VAR processes with martingale difference innovations, these estimators are known to be consistent and asymptotically normally distributed. The joint distribution of the GMM estimator of all parameters of the system is also asymptotically normal and can be obtained by considering the system of all K equations jointly. In practice, the GMM estimator may also be bootstrapped, as discussed in Chapter 12.

Equivalently the estimator for γ_k may be computed by concentrating out the lags and replacing the observed variables by the corresponding residuals. In other words, we estimate the reduced-form model

$$y_t = [A_1, \dots, A_p] Y_{t-1} + u_t$$

by LS and use the residuals $\widehat{u}_t$ instead of the y_t in the expressions for the VAR(0) model. It is important to note, however, that standard estimates of the covariance matrix of this estimator will be invalid because replacing the observations in the VAR(0) model by residuals obtained from a VAR(p) model introduces a generated-regressor problem, which invalidates conventional estimates of the standard errors of the structural parameters (see Pagan 1984).

Having obtained estimates of γ_k, we construct the corresponding estimates of the elements of B_0 from $b_k = R_k \gamma_k$. As discussed earlier, the construction of the structural impact multiplier matrix requires not only an estimate of B_0 but also an estimate of the diagonal matrix Σ_w. The diagonal elements of this matrix are estimated as

$$\hat{\sigma}_{w_k}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{w}_{kt}^2.$$

Restrictions on B_0^{-1} ***and*** B_0**.** If there are identifying restrictions on both B_0^{-1} and B_0 in an overidentified structural VAR model, one minimizes the same objective function as in the case when there are only restrictions on B_0^{-1},

$$J = T(\mathrm{vech}(\widehat{\Sigma}_u) - \mathrm{vech}(B_0^{-1}B_0^{-1\prime}))'\widehat{W}(\mathrm{vech}(\widehat{\Sigma}_u) - \mathrm{vech}(B_0^{-1}B_0^{-1\prime})).$$

The difference is that the objective function must be minimized subject to the additional identifying restrictions on B_0 rather than by unconstrained minimization of J. This involves nonlinear restrictions on B_0^{-1}. Unlike in the case of restrictions on B_0 only, there is no closed-form solution in this case. Under the conditions invoked in the discussion of the GMM estimator without restrictions on B_0, this procedure generates a consistent and asymptotically normal estimator of B_0^{-1} and hence B_0.

9.3 Instrumental Variable Estimation

If there are identifying restrictions on B_0^{-1} that cannot be expressed as exclusion restrictions on B_0, the GMM estimator must be constructed by iterative techniques. In contrast, if the identifying restrictions can be imposed on B_0, GMM estimates may also be constructed using traditional IV estimators based on linear regressions.

IV Analysis of the Structural VAR(0) Model. For expository purposes we again assume for the moment that we are dealing with a VAR(0) model. Then, as in the previous section, the k^{th} equation of the structural-form VAR model can be written as

$$y_{kt} = \mathbf{y}_{kt}'b_k + w_{kt},$$

where $\mathbf{y}_{kt}$ is the column vector of the elements of y_t that appear on the right-hand side of the k^{th} equation, where y_{kt} is the left-hand side variable in the k^{th} equation, and where b_k is the vector of associated structural parameters. Equivalently, in matrix notation

$$\mathbf{y}_k = Y_k b_k + \mathbf{w}_k, \tag{9.3.1}$$

where all notation is defined as in equation (9.2.6). Assuming that the k^{th} equation is identified such that all right-hand side variables are uncorrelated with w_{kt}, we can use Y_k as the instrument, which is equivalent to estimating equation (9.3.1) by LS:

$$\hat{b}_k^{IV} = \hat{b}_k^{LS} = (Y_k'Y_k)^{-1}Y_k'\mathbf{y}_k. \tag{9.3.2}$$

Note that this is also the method-of-moments estimator given moment conditions $\mathbb{E}(\mathbf{y}_{kt}'w_{kt}) = 0$. Thus, for just-identified models, we have

$$\hat{b}_k^{LS} = \hat{b}_k^{IV} = \hat{b}_k^{GMM}.$$

This IV estimation method also works if there are overidentifying restrictions. Suppose that there are further linear restrictions on the parameters in the k^{th} equation such that

$$b_k = R_k \gamma_k, \tag{9.3.3}$$

where R_k is a matrix that relates the unrestricted parameters γ_k to the restricted parameter vector b_k.

Having defined R_k, we recover estimates of γ_k from the regression

$$\mathbf{y}_k = Y_k^{\dagger} \gamma_k + \mathbf{w}_k,$$

where $Y_k^{\dagger} = Y_k R_k$. Because we have more instruments in Y_k than parameters in the k^{th} equation, we can estimate this regression model by two-stage LS (2SLS) using the instruments $\widehat{Y}_k^{\dagger} \equiv Y_k (Y_k' Y_k)^{-1} Y_k' Y_k^{\dagger}$:

$$\widehat{\gamma}_k^{IV} = \widehat{\gamma}_k^{2SLS} = (\widehat{Y}_k^{\dagger\prime} Y_k^{\dagger})^{-1} \widehat{Y}_k^{\dagger\prime} \mathbf{y}_k. \tag{9.3.4}$$

As is well known, if the errors are iid, this is also the GMM estimator based on the moment conditions $\mathbb{E}(\mathbf{y}_{kt}' w_{kt}) = 0$ and the optimal weighting matrix, because

$$\widehat{\gamma}_k^{GMM} = \widehat{\gamma}_k^{2SLS}$$

is precisely the solution that minimizes the GMM objective function.

IV Analysis of the Structural VAR(p) Model. The discussion so far focused on the VAR(0) model. There are two ways of generalizing this method to the VAR(p) model, $p > 0$. The first approach is to adapt equation (9.3.1) to include p lags of the model variables and to augment the set of instruments accordingly. In that case, the k^{th} equation becomes

$$y_{kt} = \mathbf{y}_{kt}' R_k \gamma_k + Y_{t-1}' \mathbf{b}_k + w_{kt}, \tag{9.3.5}$$

where $Y_{t-1}' \equiv (y_{t-1}', \dots, y_{t-p}')$ and $\mathbf{b}_k'$ is the k^{th} row of $B = [B_1, \dots, B_p]$, as before. Defining $Z \equiv [Y_0, \dots, Y_{T-1}]$, the same analysis as for the VAR(0) case can be applied to

$$\mathbf{y}_k = [Y_k^{\dagger}, Z'] \begin{pmatrix} \gamma_k \\ \mathbf{b}_k \end{pmatrix} + \mathbf{w}_k. \tag{9.3.6}$$

The IV estimator of the parameters then becomes

$$\begin{pmatrix} \widehat{\gamma}_k^{IV} \\ \widehat{\mathbf{b}}_k^{IV} \end{pmatrix} = \left(\widehat{Z}_k [Y_k^{\dagger}, Z'] \right)^{-1} \widehat{Z}_k \mathbf{y}_k,$$

where

$$\widehat{Z}_k{}' = [Y_k, Z'] \left(\begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k, Z'] \right)^{-1} \begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k^{\dagger}, Z']. \tag{9.3.7}$$

An alternative approach to constructing the IV estimator of the VAR(p), $p > 0$, model is to concentrate out the lagged regressors and to replace the observed variables in the VAR(0) model by the residuals obtained from the VAR(p) model. Although this residual-based IV approach is simple, such regressions do not produce valid estimates of the covariance matrix of the IV estimator. Using residuals as instruments introduces a generated regressor problem (see Pagan 1984). Corrections of the covariance estimator are discussed in King and Watson (1997). An alternative is the use of the IV estimator as the starting value in constructing a single-iteration GMM estimator, which allows standard software to be used to consistently estimate the covariance of the IV/GMM estimator. The easiest solution, however, would be to include the lagged observables in IV estimation, as discussed earlier.

As in the discussion of single-equation GMM, having obtained estimates of γ_k, we construct the corresponding estimates of the elements of B_0 from $b_k = R_k\gamma_k$, and estimates of the diagonal elements of the matrix Σ_w are obtained as

$$\hat{\sigma}^2_{w_k} = \frac{1}{T}\sum_{t=1}^{T}\widehat{w}^2_{kt}.$$

To conclude, the advantage of the GMM approach is that it can handle restrictions on B_0^{-1}, whereas the IV approach cannot. The advantage of the IV approach is that it does not require iteration.

As in the GMM framework, the asymptotic properties of the IV estimator follow from standard theoretical arguments for IV estimators (see, e.g., Judge, Griffiths, Hill, Lütkepohl, and Lee 1985). If the instruments are valid and strong, the IV estimator under our assumptions is consistent and asymptotically normally distributed. One advantage of the IV estimator is that the strength of the instrument in the first stage may be evaluated empirically (see, e.g., Pagan and Robertson 1998).

The strength of the instrument increases with the correlation between the instruments and the regressors. For example, in model (9.3.6) the contemporaneous regressors are contained in the matrix $Y_k^{\dagger}$ and the lagged regressors are in Z'. Both regressors are instrumented by $\widehat{Z}_k$, as defined in equation (9.3.7). For the structural VAR model to be identified, the matrix of correlations between the regressors and the instruments must be of full rank. The strength of the identification is measured by the conditioning number of this matrix. Evidence of this matrix being ill-conditioned would be indicative of the instruments being weak and would cast doubt on the reliability of the empirical estimates. Similar problems of imprecise estimates may, of course, also arise when estimating structural VAR models by the method of moments or by ML. Exploiting the equivalence between IV estimators and many non-IV estimators can be helpful in diagnosing such problems.

An Empirical Illustration. For the example of the three-dimensional recursive model used in Section 9.2.1, the IV estimator is identical to the GMM estimator. We obtain the following estimates:

$$
\begin{aligned}
y_{1t} &= \widehat{\nu}_1^{IV} + \sum_{i=1}^{4} (\widehat{b}_{11,i}^{IV} y_{1,t-i} + \widehat{b}_{12,i}^{IV} y_{2,t-i} + \widehat{b}_{13,i}^{IV} y_{3,t-i}) + \widehat{w}_{1t} \\
&= -0.6807 - 0.0064 y_{1,t-1} + 0.9365 y_{2,t-1} + 1.7510 y_{3,t-1} \\
&\quad - 0.1822 y_{1,t-2} + 12.6454 y_{2,t-2} - 3.4391 y_{3,t-2} \\
&\quad - 0.0073 y_{1,t-3} - 3.5986 y_{2,t-3} + 1.7337 y_{3,t-3} \\
&\quad - 0.1194 y_{1,t-4} - 8.1807 y_{2,t-4} + 0.9309 y_{3,t-4} + \widehat{w}_{1t},
\end{aligned}
$$

$$
\widehat{\sigma}_{w_1}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{w}_{1t}^2 = 286.8106,
$$

$$
\begin{aligned}
y_{2t} &= \widehat{\nu}_2^{IV} - \widehat{b}_{21,0}^{IV} y_{1t} \\
&\quad + \sum_{i=1}^{4} (\widehat{b}_{21,i}^{IV} y_{1,t-i} + \widehat{b}_{22,i}^{IV} y_{2,t-i} + \widehat{b}_{23,i}^{IV} y_{3,t-i}) + \widehat{w}_{2t} \\
&= -0.0135 + 0.0025 y_{1t} \\
&\quad + 0.0018 y_{1,t-1} + 0.5968 y_{2,t-1} + 0.0239 y_{3,t-1} \\
&\quad + 0.0047 y_{1,t-2} + 0.0963 y_{2,t-2} - 0.0303 y_{3,t-2} \\
&\quad + 0.0022 y_{1,t-3} + 0.0566 y_{2,t-3} + 0.0191 y_{3,t-3} \\
&\quad + 0.0010 y_{1,t-4} + 0.1690 y_{2,t-4} + 0.0683 y_{3,t-4} + \widehat{w}_{2t},
\end{aligned}
$$

$$
\widehat{\sigma}_{w_2}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{w}_{2t}^2 = 0.0455,
$$

$$
\begin{aligned}
y_{3t} &= \widehat{\nu}_3^{IV} - \widehat{b}_{31,0}^{IV} y_{1t} - \widehat{b}_{32,0}^{IV} y_{2t} \\
&\quad + \sum_{i=1}^{4} (\widehat{b}_{31,i}^{IV} y_{1,t-i} + \widehat{b}_{32,i}^{IV} y_{2,t-i} + \widehat{b}_{33,i}^{IV} y_{3,t-i}) + \widehat{w}_{3t} \\
&= 0.4418 + 0.0023 y_{1t} + 0.2540 y_{2t} \\
&\quad - 0.0028 y_{1,t-1} - 0.3731 y_{2,t-1} + 0.3211 y_{3,t-1} \\
&\quad - 0.0070 y_{1,t-2} + 0.3261 y_{2,t-2} + 0.1637 y_{3,t-2} \\
&\quad - 0.0007 y_{1,t-3} - 0.1451 y_{2,t-3} - 0.0538 y_{3,t-3} \\
&\quad - 0.0080 y_{1,t-4} - 0.1001 y_{2,t-4} - 0.0085 y_{3,t-4} + \widehat{w}_{3t},
\end{aligned}
$$

$$
\widehat{\sigma}_{w_3}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{w}_{3t}^2 = 0.5058.
$$

Since the model is recursive and hence just identified, the estimates are identical to the LS estimates for the three equations. For example, the estimates in the first equation are the same as those in the first rows of the estimated $\widehat{A}_j$ reduced-form coefficient matrices in Section 9.2.1. The estimates of the variances of the structural errors are computed without degrees-of-freedom adjustment.

Collecting the estimates associated with unlagged regressors in the second and third equations, we obtain an estimate for B_0,

$$\widehat{B}_0^{IV} = \begin{bmatrix} 1 & 0 & 0 \\ -0.0025 & 1 & 0 \\ -0.0023 & -0.2540 & 1 \end{bmatrix},$$

with diagonal elements standardized to 1. Defining the IV estimator

$$\widehat{\Sigma}_w = \begin{bmatrix} 286.8106 & 0 & 0 \\ 0 & 0.0455 & 0 \\ 0 & 0 & 0.5058 \end{bmatrix},$$

the quantity $(\widehat{B}_0^{IV})^{-1}\widehat{\Sigma}_w^{1/2}$ produces the same structural impact multiplier matrix as the Cholesky decomposition apart from the degrees-of-freedom adjustment we used in earlier sections.

9.4 Full Information Maximum Likelihood Estimation

The IV estimator may also be viewed as an alternative to constructing the Gaussian ML estimator of just-identified structural VAR models. The advantage of the full information ML (FIML) estimator is that (like some GMM approaches) it allows for identifying restrictions on both B_0 and B_0^{-1}. Given the structural VAR model

$$B_0 y_t = B_0 A Y_{t-1} + w_t, \tag{9.4.1}$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$ and $A \equiv [A_1, \dots, A_p]$ denotes the reduced-form slope VAR parameters, as before. Assuming that w_t is Gaussian white noise with diagonal covariance matrix Σ_w and $w_t \sim \mathcal{N}(0, \Sigma_w)$, the associated reduced-form residuals are $u_t = B_0^{-1} w_t \sim \mathcal{N}(0, \Sigma_u = B_0^{-1}\Sigma_w B_0^{-1\prime})$.

The corresponding log-likelihood function for a sample $Y \equiv [y_1, \dots, y_T]$ is

$$\begin{aligned} \log l(A, B_0, \Sigma_w) &= -\frac{KT}{2}\log(2\pi) - \frac{T}{2}\log(\det(B_0^{-1}\Sigma_w B_0'^{-1})) \\ &\quad - \frac{1}{2}\mathrm{tr}\{(Y - AZ)'[B_0^{-1}\Sigma_w B_0'^{-1}]^{-1}(Y - AZ)\} \\ &= \text{constant} + \frac{T}{2}\log(\det(B_0)^2) - \frac{T}{2}\log(\det(\Sigma_w)) \\ &\quad - \frac{1}{2}\mathrm{tr}\{B_0'\Sigma_w^{-1}B_0(Y - AZ)(Y - AZ)'\}, \end{aligned} \tag{9.4.2}$$

where $Z \equiv [Y_0, \dots, Y_{T-1}]$, as before (see Lütkepohl 2005, chapter 9).

If there are no restrictions on the reduced-form parameters, then, for given B_0 and Σ_w, the log-likelihood function $\log l(A, B_0, \Sigma_w)$ is maximized with respect to A by

$$\widehat{A} = YZ'(ZZ')^{-1} = \left(\sum_{t=1}^{T} y_t Y'_{t-1}\right)\left(\sum_{t=1}^{T} Y_{t-1}Y'_{t-1}\right)^{-1}$$

(see Section 2.3). Replacing A with this estimator results in the concentrated log-likelihood function

$$\begin{aligned}\log l_c(B_0, \Sigma_w) &= \text{constant} + \frac{T}{2}\log(\det(B_0)^2)\\ &\quad - \frac{T}{2}\log(\det(\Sigma_w)) - \frac{T}{2}\operatorname{tr}(B_0'\Sigma_w^{-1}B_0\widetilde{\Sigma}_u), \qquad (9.4.3)\end{aligned}$$

where $\widetilde{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$ and $\widehat{u}_t = y_t - \widehat{A}Y_{t-1}$ is the LS residual. In general, the concentrated log-likelihood function can be maximized by numerical methods with respect to B_0 and Σ_w, subject to the identifying restrictions. For just-identified models it can be shown that the maximum is obtained when

$$B_0'\Sigma_w^{-1}B_0 = \widetilde{\Sigma}_u^{-1}$$

(Lütkepohl 2005, chapter 9). If the variances of the structural errors are normalized to one so that $\Sigma_w = I_K$, then the ML estimator $\widetilde{B}_0$ of B_0 satisfies

$$\widetilde{B}_0'\widetilde{B}_0 = \widetilde{\Sigma}_u^{-1} \quad \text{or} \quad \widetilde{B}_0^{-1}\widetilde{B}_0'^{-1} = \widetilde{\Sigma}_u.$$

For the VAR example from Section 9.2.1, where B_0^{-1} is assumed to be lower triangular, we obtain

$$\widetilde{B}_0^{-1} = \begin{bmatrix} 16.9355 & 0 & 0\\ 0.0419 & 0.2134 & 0\\ 0.0498 & 0.0542 & 0.7112 \end{bmatrix}$$

which is identical to the Cholesky decomposition of $\widehat{\Sigma}_u$, as in Section 9.2.1, after accounting for the differences in the degrees-of-freedom adjustment. The corresponding ML estimate of B_0 is

$$\widetilde{B}_0 = \begin{bmatrix} 0.0590 & 0 & 0\\ -0.0116 & 4.6864 & 0\\ -0.0033 & -0.3572 & 1.4061 \end{bmatrix}.$$

If instead the diagonal elements of B_0 are normalized to one and the variances of the structural shocks are unrestricted, we obtain the ML estimates

$$\widetilde{B}_0 = \begin{bmatrix} 1 & 0 & 0\\ -0.0025 & 1 & 0\\ -0.0023 & -0.2540 & 1 \end{bmatrix}$$

and

$$\widetilde{\Sigma}_w = \begin{bmatrix} 286.8106 & 0 & 0 \\ 0 & 0.0455 & 0 \\ 0 & 0 & 0.5058 \end{bmatrix}.$$

such that $\widetilde{B}_0^{-1}\widetilde{\Sigma}_w^{1/2}$ produces the structural impact multiplier matrix.

Of course, ML estimates can also be computed in the same manner if there are nonrecursive identifying restrictions as in the Keating example model from Section 9.2.2. The estimates for this example are identical to the solutions in Section 9.2.2 after adjusting for the degrees of freedom. More precisely, if the diagonal elements of B_0 are normalized to unity, we have

$$\widetilde{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.2669 & 1 & 0.7288 & 0.1784 \\ 0 & 0 & 1 & -11.2056 \\ -3.2443 & -3.2443 & 3.4133 & 1 \end{bmatrix}$$

and

$$\widetilde{\Sigma}_w = \begin{bmatrix} 0.0556 & 0 & 0 & 0 \\ 0 & 0.9092 & 0 & 0 \\ 0 & 0 & 132.5486 & 0 \\ 0 & 0 & 0 & 9.7366 \end{bmatrix}.$$

In practice, if the model is just-identified, this ML estimator often is implemented by first estimating the reduced form and then applying numerical solution methods to recover $\widetilde{B}_0^{-1}$ from $\widetilde{\Sigma}_u$. An alternative is to rely on iterative methods to maximize the concentrated likelihood subject to constraints on B_0 and/or B_0^{-1}. A detailed review of these numerical methods is beyond the scope of this book (see, however, Judge, Griffiths, Hill, Lütkepohl, and Lee 1985, appendix B). The same iterative algorithms may also be used when there are overidentifying restrictions.

If y_t is a stationary Gaussian VAR(p) process, the ML estimation framework facilitates the derivation of the asymptotic properties of the estimator. Standard ML theory implies that the unrestricted parameters are consistent and asymptotically normal. Suppose that the unrestricted parameters are collected in the vector η which contains the reduced-form parameters A, the unrestricted elements of B_0, and the diagonal elements of Σ_w if the latter are not standardized to one. Denoting the ML estimator of η by $\tilde{\eta}$, under general assumptions we obtain

$$\sqrt{T}(\tilde{\eta} - \eta) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_a^{-1}(\eta)\right),$$

where $\mathcal{I}_a(\eta)$ denotes the asymptotic information matrix. Note that the diagonal elements of $\widetilde{\Sigma}_w$ also have an asymptotic normal distribution.

If there are overidentifying restrictions on B_0, the restricted ML estimator of Σ_u may not equal the unrestricted ML estimator of the reduced-form covariance matrix anymore. In that case the reduced-form residual covariance matrix Σ_u is estimated as

$$\widetilde{\Sigma}_u^r \equiv \widetilde{B}_0^{-1} \widetilde{\Sigma}_w \widetilde{B}_0^{-1\prime}, \tag{9.4.4}$$

where $\widetilde{B}_0^{-1}$ and $\widetilde{\Sigma}_w$ are the restricted ML estimators, and the LR statistic,

$$LR = T \left(\log(\det(\widetilde{\Sigma}_u^r)) - \log(\det(\widetilde{\Sigma}_u)) \right), \tag{9.4.5}$$

can be used to test the overidentifying restrictions. Under general conditions, this LR statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of overidentifying restrictions, provided the restrictions are correct.

An alternative approach to estimating structural VAR models by ML that allows for a combination of restrictions on B_0 and B_0^{-1} (as is the case, for example, in the Blanchard and Perotti (2002) model of fiscal policy discussed in Section 8.5.1), sometimes facilitates the analysis. Recall that in this class of models

$$B_0 u_t = C w_t$$

with $\Sigma_w = I_K$ such that $\Sigma_u = B_0^{-1} C C' B_0^{-1\prime}$. The implied concentrated log likelihood of this model is

$$\log l_c(B_0, C) = \text{constant} + \frac{T}{2} \log(\det(B_0)^2) - \frac{T}{2} \log(\det(C)^2) - \frac{T}{2} \text{tr}(B_0' C^{-1\prime} C^{-1} B_0 \widetilde{\Sigma}_u).$$

The advantage of this setup is that one can impose the identifying restrictions directly on B_0 and C allowing the unconstrained maximization of the log-likelihood with respect to the unrestricted elements of B_0 and C. In addition, overidentifying restrictions may also be imposed directly. The resulting Gaussian ML estimator is consistent and asymptotically normal (see Lütkepohl 2005, section 9.3.1). Tests for overidentifying restrictions may be conducted as discussed earlier with $\widetilde{\Sigma}_u^r \equiv \widetilde{B}_0^{-1} \widetilde{C} \widetilde{C}' \widetilde{B}_0^{-1\prime}$.

9.5 Bayesian Estimation

By definition, Bayesian estimators are obtained by evaluating the posterior distribution of the statistic of interest. They minimize the posterior expected value of a user-defined loss function. Under quadratic loss, for example, the Bayesian estimator of a scalar parameter would be the mean of the posterior distribution of this parameter, and, under absolute loss, it would be the posterior median.

As discussed in Chapter 5, if the posterior is Gaussian, minimizing a symmetric loss function such as the quadratic or absolute loss function produces a Bayesian estimator that asymptotically coincides with the usual LS/ML estimator. More generally, under alternative loss functions or when the posterior distribution is not Gaussian, the Bayesian estimator may differ from conventional frequentist estimators even asymptotically.

The construction of the posterior distribution of the structural model parameters starts with the specification of a prior. The prior distributions discussed in Chapter 5 were for reduced-form VAR model parameters. For structural VAR analysis there are two common approaches. One is to rely on a conventional reduced-form prior such as the Gaussian-inverse Wishart prior and to generate posterior draws for the structural model parameters by suitably transforming each reduced-form posterior draw, as outlined in Chapter 5 (see also, e.g., Canova 1991; Gordon and Leeper 1994). In other words, if $\Sigma_u^{(i)}$ is the i^{th} draw from the posterior distribution of Σ_u, then the corresponding posterior draw for B_0 (or, equivalently, the corresponding draws for B_0^{-1} and Σ_w) may be constructed from $\Sigma_u^{(i)}$ by any of the method-of-moments methods already discussed in this chapter. For example, if a recursive identification scheme is used, a lower-triangular Cholesky decomposition of $\Sigma_u^{(i)}$ yields the i^{th} posterior draw for B_0^{-1}.

This widely used approach, however, lacks a theoretical foundation, as stressed by Sims and Zha (1998). One concern is that standard reduced-form priors make a distinction between own lags and other lags in each VAR equation that does not exist in the structural VAR representation. Moreover, the approach employed by Canova (1991) and Gordon and Leeper (1994), among many others, is practically feasible only when the structural VAR model is just identified. When B_0^{-1} (or B_0) is overidentified, it is necessary to impose the prior directly on the structural VAR representation, because the overidentifying restrictions impose restrictions on the reduced-form model. An alternative approach that allows one to impose the priors directly on the structural VAR model parameters is developed in Sims and Zha (1998, 1999). It may be applied regardless of whether the structural model is just identified or overidentified.

Recall that the Gaussian VAR(p) model may be expressed in structural form as

$$B_0 y_t = BY_{t-1} + w_t, \qquad w_t \sim \mathcal{N}(0, I_K),$$

where $Y_{t-1}' \equiv (y_{t-1}', \ldots, y_{t-p}')$ and $B \equiv [B_1, \ldots, B_p]$. Stacking the observations for $t = 1, \ldots, T$, yields

$$B_0 Y - BZ = \mathbf{B}X = W,$$

where $Y \equiv [y_1, \ldots, y_T]$ is $K \times T$, $W \equiv [w_1, \ldots, w_T]$ is $K \times T$, $Z \equiv [Y_0, \ldots, Y_{T-1}]$ is $Kp \times T$, and

$$\mathbf{B} \equiv [B_0, B], \quad \text{and} \quad X \equiv \begin{bmatrix} Y \\ -Z \end{bmatrix}$$

are of dimension $K \times K(p+1)$ and $K(p+1) \times T$, respectively. Let $\mathbf{b} = \text{vec}(\mathbf{B})$, $b_0 = \text{vec}(B_0)$, and $b = \text{vec}(B)$. Then the likelihood function is

$$\begin{aligned} l(\mathbf{B}|Y) &\propto \det(B_0)^T \exp\left(-\frac{1}{2}\text{tr}(\mathbf{B}X)(\mathbf{B}X)'\right) \\ &= \det(B_0)^T \exp\left(-\frac{1}{2}\mathbf{b}'(XX' \otimes I_K)\mathbf{b}\right). \end{aligned}$$

The prior distribution for b_0 may have singularities caused by identifying restrictions. Let $g(b_0)$ be the prior for the parameters in B_0 and let $g(b|b_0)$ denote the density of $\mathcal{N}(\bar{b}(b_0), \bar{V}_b(b_0))$, where $\bar{b}(b_0)$ and $\bar{V}_b(b_0)$ denote the mean and the covariance matrix of the prior distribution of b conditional on b_0. Then $g(\mathbf{b}) = g(b_0)g(b|b_0)$ and the posterior density of $\mathbf{b}$ is

$$g(\mathbf{b}|Y) \propto g(\mathbf{b})l(\mathbf{B}|Y). \tag{9.5.1}$$

This posterior density is nonstandard in general. It can be shown that for fixed b_0, the conditional posterior distribution of b given b_0 is Gaussian. Sims and Zha (1998) also derive the marginal posterior density of b_0. An important question is how to specify the marginal prior of b_0 in this model. Sims and Zha (1998) recommend specifying a joint normal prior on the nonzero elements of B_0 in the case of a lower triangular normalization of B_0.[3]

As discussed in Canova (2007, chapter 10.3, pp. 391–392), it is instructive to compare the structural priors discussed in Sims and Zha (1998) with more traditional Minnesota priors on the reduced-form parameters. There are three key differences. First, since $A = [A_1, \ldots, A_p] = B_0^{-1}B$, prior beliefs about A may be correlated across equations if the prior beliefs about B_0 are. Second, there is no distinction between own lags and lags of other variables in a given equation, because in simultaneous equation models there is no unique normalization of the left-hand side variables. Third, the scale factors for the prior variances of the lag coefficients differ from those in the Minnesota prior.

It should be noted that the approach of Sims and Zha (1998) is specifically designed for structural VAR models with linear restrictions on B_0 under the assumption that the diagonal elements of B_0 are unity and that Σ_w is diagonal. Put differently, this approach cannot be used for models imposing identifying or overidentifying restrictions on B_0^{-1}. Canova and Pérez Forero

[3] MATLAB code for the implementation for this method is provided at www.econ.yale.edu/~sims.

(2015) recently have generalized the approach of Sims and Zha by allowing for certain nonlinear identifying restrictions on B_0. This allows them, for example, to accommodate restrictions on long-run responses of the type discussed in Chapters 10 and 11. It is not clear, however, whether their approach also handles direct restrictions on the elements of B_0^{-1}.

A generalization of the approach of Sims and Zha (1998) has recently been proposed by Baumeister and Hamilton (2015c). They allow for prior uncertainty about exclusion restrictions on B_0. Rather than imposing that certain elements of B_0 are exactly zero, they postulate independent Student-*t* distributions for these elements that are centered on zero, allowing the user to account for identification uncertainty in conducting Bayesian analysis. Of course, this approach presumes that the users have extraneous information allowing them to parameterize the identification uncertainty in the form of a known density. Such information may not be readily available in practice.

9.6 Summary

This chapter reviewed estimation methods for structural VAR models with short-run restrictions on the matrix B_0 or B_0^{-1}. We stressed that there is a range of alternative estimation methods, some of which are designed for models with restrictions on B_0, whereas others are designed for restrictions on B_0^{-1}. The most general estimators for models with identifying or overidentifying restrictions on both B_0 or B_0^{-1} are the Gaussian ML estimator and the GMM estimator. If the identifying or overidentifying restrictions are imposed on B_0 only, the GMM estimator may be computed in closed form. If there are restrictions on B_0^{-1} or on both B_0 and B_0^{-1}, then the GMM estimator must be computed in two steps. We first estimate Σ_u from the reduced-form representation of the VAR model, before expressing Σ_u in terms of the structural parameters of the model and solving for B_0 or B_0^{-1} numerically. If there are restrictions on B_0 only, the IV estimator for just- and overidentified models is another estimator that yields a closed-form solution. For just-identified models the method of moments can be used alternatively. Finally, for the special case of recursively identified models it is possible to estimate B_0^{-1} based on the lower-triangular Cholesky decomposition of the estimate of Σ_u.

We observed that it is also possible to estimate structural VAR models by Bayesian methods. The conventional approach of applying standard estimation methods for B_0 and/or B_0^{-1} to the posterior draws of Σ_u does not allow for overidentification. In contrast, Bayesian methods that impose priors directly on the structural form allow for both just-identifying and overidentifying restrictions, but only on B_0. At present, they do not appear to allow for restrictions on B_0^{-1} except in the recursive case, because the triangularity of B_0^{-1} implies that B_0 is triangular.