11 Estimation Subject to Long-Run Restrictions

This chapter illustrates the estimation of structural VAR models subject to long-run identifying restrictions. A variety of estimation methods have been proposed to estimate these models, including the method of moments, instrumental variable (IV) estimators, and full information maximum likelihood (FIML) estimators. We first focus on structural models based on long-run identifying restrictions only, followed by structural models that impose both long-run and short-run identifying restrictions.

11.1 Model Setup

Let y_t be a K-dimensional vector of variables that may be integrated of order 1 and possibly cointegrated. Consider the structural form

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t = BY_{t-1} + w_t, \tag{11.1.1}$$

where $Y'_{t-1} \equiv (y'_{t-1}, \ldots, y'_{t-p})$ and $B \equiv [B_1, \ldots, B_p]$. Deterministic terms are neglected and the structural errors w_t are assumed to have a unit covariance matrix, $w_t \sim (0, I_K)$, unless stated otherwise. The corresponding reduced form is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = AY_{t-1} + u_t, \tag{11.1.2}$$

where $A \equiv [A_1, \ldots, A_p] = B_0^{-1} B$ and $u_t = B_0^{-1} w_t \sim (0, \Sigma_u = B_0^{-1} B_0^{-1\prime})$.

Suppose that the structural model is identified by long-run restrictions. As discussed in Chapter 10, there are two alternative representations for estimating structural VAR models subject to long-run restrictions. If the $I(1)$ variables are not cointegrated or the cointegrating relations are known, then all $I(1)$ variables may be transformed to $I(0)$ variables by expressing them in first differences or as cointegration relations. As in the previous chapter, we denote these transformed variables as z_t. For the reduced-form and structural-form VAR representations of z_t we use the same notation as for y_t in equations (11.1.1) and (11.1.2). In other words, $z_t = A_1 z_{t-1} + \cdots + A_p z_{t-p} + u_t$ is the reduced-form

 https://doi.org/10.1017/9781108164818.012

representation and $B_0 z_t = B_1 z_{t-1} + \cdots + B_p z_{t-p} + w_t$ denotes the structural form.

In that case, identification involves imposing long-run exclusion restrictions on the elements of the $K \times K$ cumulative structural impulse response matrix,

$$\Theta(1) = \sum_{i=0}^{\infty} \Theta_i = B(1)^{-1} = A(1)^{-1} B_0^{-1},$$

of the stationary process z_t, where

$$\begin{aligned} \Theta(L) = \sum_{i=0}^{\infty} \Theta_i L^i \equiv B(L)^{-1} &= (B_0 - B_1 L - \cdots - B_p L^p)^{-1} \\ &= A(L)^{-1} B_0^{-1} = (I_K - A_1 L - \cdots - A_p L^p)^{-1} B_0^{-1} \end{aligned}$$

and L is the lag operator, $A(L) \equiv I_K - A_1 L - \cdots - A_p L^p$, $B(L) \equiv B_0 - B_1 L - \cdots - B_p L^p$, $A(1)$ is $A(L)$ evaluated at $L = 1$ and $B(1)$ is $B(L)$ evaluated at $L = 1$.

An alternative and more general reduced-form representation of the VAR model (11.1.2) that also allows for unknown cointegrating relations is the VECM

$$\Delta y_t = \alpha \beta' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t, \qquad (11.1.3)$$

where α and β are $K \times r$ matrices of rank r. In the latter case, the long-run effects of the reduced-form shocks on the level of the model variables are given by the $K \times K$ matrix

$$\boldsymbol{\Xi} = \beta_\perp \left[\alpha_\perp' \left(I_K - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \right) \beta_\perp \right]^{-1} \alpha_\perp',$$

where $\beta_\perp$ and $\alpha_\perp$ are orthogonal complements of β and α, respectively. As was shown in Chapter 10, the long-run effects of the structural shocks on the level variables in the VEC representation are given by the $K \times K$ matrix

$$\boldsymbol{\Upsilon} = \boldsymbol{\Xi} B_0^{-1}.$$

Thus, long-run restrictions may also be imposed on the elements of the long-run multiplier matrix $\boldsymbol{\Upsilon}$ of the VECM. Since $\boldsymbol{\Xi}$ is determined by the reduced-form parameters of the VAR model, all constraints on $\boldsymbol{\Upsilon}$ ultimately restrict B_0^{-1} and, hence B_0.

For estimation purposes it is sometimes useful to consider a structural representation of the VECM, which is obtained by pre-multiplying the reduced form (11.1.3) by B_0,

$$B_0 \Delta y_t = \alpha^\dagger \beta' y_{t-1} + \boldsymbol{\Gamma}_1^\dagger \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}^\dagger \Delta y_{t-p+1} + w_t,$$

where $\alpha^\dagger = B_0\alpha$ and $\mathbf{\Gamma}_i^\dagger = B_0\mathbf{\Gamma}_i$, $i = 1, \ldots, p-1$. Note that in the error correction term this transformation only affects the loading coefficients, but not the cointegration relationships $\beta' y_{t-1}$. We are interested in the estimation of B_0 and/or B_0^{-1}. Obviously, any estimate of B_0 can be inverted to obtain an estimate of B_0^{-1} and vice versa.

11.2 Models Subject to Long-Run Restrictions Only

In this section, we focus on an illustrative example based on Galí (1999) who studies the relationship between technology shocks, employment, and aggregate economic fluctuations using quarterly U.S. data. The observables are the log of productivity (denoted by $prod_t$) and the log of total employee hours in nonagricultural establishments, averaged to quarterly frequency (denoted by h_t). The productivity variable is constructed by subtracting the log of hours from the log of real GDP. Let $y_t = (prod_t, h_t)'$ and suppose that both variables are $I(1)$ but not cointegrated. Galí postulates that y_t evolves according to a VAR(5) process with intercept.

Stationary Representation of the Model. The stationary VAR representation of Galí's model is a VAR(4) model with intercept for $z_t = \Delta y_t = (\Delta prod_t, \Delta h_t)'$,

$$z_t = \nu + A_1 z_{t-1} + \cdots + A_4 z_{t-4} + u_t.$$

The sample period for the transformed data is 1947q2-1998q3. The unrestricted reduced-form LS estimates are

$$\widehat{A}_1 = \begin{bmatrix} -0.1288 & -0.1283 \\ 0.2955 & 0.5809 \end{bmatrix},$$

$$\widehat{A}_2 = \begin{bmatrix} 0.0881 & -0.1258 \\ 0.1833 & -0.1060 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} -0.0240 & -0.0464 \\ 0.1190 & 0.1545 \end{bmatrix},$$

$$\widehat{A}_4 = \begin{bmatrix} 0.0251 & -0.0697 \\ -0.0052 & -0.1112 \end{bmatrix},$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.4596 & -0.0469 \\ -0.0469 & 0.5343 \end{bmatrix}.$$

There is only one identifying restriction in the model of Galí (1999). The technology shock is identified by the assumption that only technology shocks have a long-run effect on the level of productivity. Thus, if the technology shock is the first structural shock, the matrix of structural long-run multipliers is lower triangular such that

$$\Theta(1) = \left(I_2 - \sum_{i=1}^{4} A_i\right)^{-1} B_0^{-1} = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}.$$

The second structural shock is interpreted as a non-technology shock. The structure of this model mirrors that in Blanchard and Quah (1989).

Vector Error Correction Representation of the Model. Equivalently, this model can be written in the general reduced-form representation of the VAR model (11.1.2) as a VECM. In that case, we have

$$\Delta y_t = \nu + \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_4 \Delta y_{t-4} + u_t.$$

Given that both variables in y_t are $I(1)$ but not cointegrated, the cointegration rank is $r = 0$, so that $\alpha = \beta = 0$ and $\alpha_\perp = \beta_\perp = I_2$. Hence, the VECM reduces to a VAR model in first differences, $\Delta y_t = (\Delta prod_t, \Delta h_t)'$, and the VEC representation coincides with the stationary VAR representation for z_t.

The unrestricted reduced-form estimates are

$$\widehat{\mathbf{\Gamma}}_1 = \begin{bmatrix} -0.1288 & -0.1283 \\ 0.2955 & 0.5809 \end{bmatrix},$$

$$\widehat{\mathbf{\Gamma}}_2 = \begin{bmatrix} 0.0881 & -0.1258 \\ 0.1833 & -0.1060 \end{bmatrix},$$

$$\widehat{\mathbf{\Gamma}}_3 = \begin{bmatrix} -0.0240 & -0.0464 \\ 0.1190 & 0.1545 \end{bmatrix},$$

$$\widehat{\mathbf{\Gamma}}_4 = \begin{bmatrix} 0.0251 & -0.0697 \\ -0.0052 & -0.1112 \end{bmatrix},$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.4596 & -0.0469 \\ -0.0469 & 0.5343 \end{bmatrix}.$$

In Galí's model $\alpha_\perp = \beta_\perp = I_2$, so that

$$\mathbf{\Xi} = \mathbf{\Gamma}(1)^{-1} = \left(I_2 - \sum_{i=1}^{4} \mathbf{\Gamma}_i\right)^{-1}.$$

Thus, the identifying restriction that the non-technology shock has no long-run effects on the level of real GDP implies that

$$\mathbf{\Upsilon} = \mathbf{\Xi} B_0^{-1} = \left(I_2 - \sum_{i=1}^{4} \mathbf{\Gamma}_i\right)^{-1} B_0^{-1} = \begin{bmatrix} \zeta_{11} & 0 \\ \zeta_{21} & \zeta_{22} \end{bmatrix}$$

is lower triangular. In this particular case, $\Theta(1)$ coincides with $\mathbf{\Upsilon}$.

11.2.1 Method-of-Moments Estimation

Estimating the Model Using the Cholesky Decomposition

Stationary Representation. Recall from Chapter 10 that $\Theta(1)\Theta(1)' = A(1)^{-1}\Sigma_u A(1)^{-1\prime}$, where $A(1) = (I_2 - A_1 - A_2 - A_3 - A_4)$ and where we have imposed $\Sigma_w = I_2$. Given that $\Theta(1)$ has a lower-triangular structure, it can be estimated as

$$\begin{aligned}\widehat{\Theta}(1) &= \begin{bmatrix}\widehat{\theta}_{11}(1) & 0 \\ \widehat{\theta}_{21}(1) & \widehat{\theta}_{22}(1)\end{bmatrix} = \text{chol}\left(\widehat{A}(1)^{-1}\widehat{\Sigma}_u\widehat{A}(1)^{-1\prime}\right) \\ &= \begin{bmatrix}0.6157 & 0 \\ -0.2745 & 1.1125\end{bmatrix},\end{aligned}$$

where

$$\widehat{A}(1)^{-1} = \begin{bmatrix}0.6689 & -0.5141 \\ 0.8227 & 1.4435\end{bmatrix}$$

and where chol denotes a function that returns the lower-triangular Cholesky decomposition. Suitable functions that implement this Cholesky decomposition are available in commonly used software.

Given estimates $\widehat{A}(1)$ and $\widehat{\Theta}(1)$, the structural impact multiplier matrix can be computed as

$$\widehat{B}_0^{-1} = \widehat{A}(1)\widehat{\Theta}(1) = \begin{bmatrix}0.5384 & 0.4119 \\ -0.4971 & 0.5359\end{bmatrix}.$$

Note that the use of long-run identifying restrictions allows $\widehat{B}_0^{-1}$ to be nonrecursive. Given the estimate of B_0^{-1}, it is straightforward to compute the mutually uncorrelated structural residuals as

$$\widehat{w}_t = \widehat{B}_0\widehat{u}_t$$

and to compute the implied structural impulse responses from the model in differences. All structural responses in Figure 11.1 have been cumulated and correspond to the response in the log-level. A positive technology shock is associated with a persistent upward shift in productivity and a decline in hours worked. A positive non-technology shock, in contrast, implies a short-lived increase in productivity and a positive response of hours worked.

Vector Error Correction Representation. Equivalently, this model may be estimated as a VECM. In this particular example, the VECM reduces to a VAR model in differences because $r = 0$. Note that

$$\begin{aligned}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}' &= \boldsymbol{\Gamma}(1)^{-1}B_0^{-1}B_0^{\prime -1}\boldsymbol{\Gamma}(1)^{\prime -1} \\ &= \boldsymbol{\Gamma}(1)^{-1}\Sigma_u\boldsymbol{\Gamma}(1)^{\prime -1},\end{aligned}$$

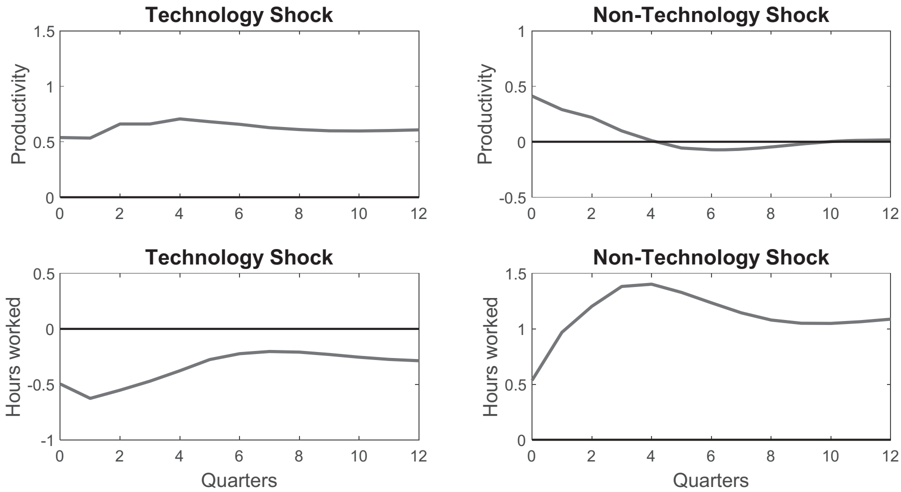

Figure 11.1. Responses to technology and non-technology shocks.

where $\boldsymbol{\Upsilon}$ is lower triangular. Replacing the parameters in the latter expression by the reduced-form estimates and imposing $\Sigma_w = I_K$, an estimate of $\boldsymbol{\Upsilon}$ can then be obtained from the lower-triangular Cholesky decomposition of

$$\widehat{\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'} \equiv \widehat{\boldsymbol{\Gamma}}(1)^{-1}\widehat{\Sigma}_u\widehat{\boldsymbol{\Gamma}}(1)'^{-1},$$

such that

$$\widehat{\boldsymbol{\Upsilon}} = \text{chol}(\widehat{\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'}).$$

The implied estimate of the structural impact multiplier matrix B_0^{-1} is given by

$$\widehat{B_0^{-1}} = \widehat{\boldsymbol{\Gamma}}(1)\text{chol}(\widehat{\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'}).$$

For our empirical example we obtain

$$\widehat{\boldsymbol{\Gamma}}(1)^{-1} = (I_2 - \widehat{\boldsymbol{\Gamma}}_1 - \widehat{\boldsymbol{\Gamma}}_2 - \widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_4)^{-1} = \begin{bmatrix} 0.6689 & -0.5141 \\ 0.8227 & 1.4435 \end{bmatrix},$$

$$\widehat{\boldsymbol{\Upsilon}} = \begin{bmatrix} 0.6157 & 0 \\ -0.2745 & 1.1125 \end{bmatrix},$$

and

$$\widehat{B_0^{-1}} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix},$$

which matches the estimate obtained using the stationary representation.

The approach of taking the Cholesky decomposition is computationally convenient, but cannot be employed unless the long-run multiplier matrix is lower triangular. A case in point is the baseline three-variable model of King, Plosser, Stock, and Watson (1991), in which the structural long-run multiplier matrix takes the form

$$\boldsymbol{\Upsilon} = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix},$$

where $*$ denotes an unrestricted element (see Section 10.3.1). Generalizations of the VECM estimation approach that can handle this situation are discussed in Section 11.3.

Estimating the Model Using a Nonlinear Equation Solver. An alternative approach is to solve the system of nonlinear equations implicitly defining the elements of B_0^{-1} using a nonlinear equation solver that finds the vector x such that $F(x) = 0$, where $F(x)$ is a system of nonlinear equations in x.[1] As before, we normalize $\Sigma_w = I_2$, while leaving the diagonal elements of B_0 unrestricted. We vectorize

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

and impose the additional identifying restriction that $\theta_{12}(1) = 0$. The objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \text{vech}(B_0^{-1} B_0^{-1\prime} - \widehat{\Sigma}_u) \\ \theta_{12}(1) \end{bmatrix} = 0, \tag{11.2.1}$$

given estimates of the reduced-form parameters $\widehat{\Sigma}_u$ and $\widehat{\boldsymbol{\Gamma}}(1)^{-1}$ and an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{\boldsymbol{\Gamma}}(1)^{-1} B_0^{-1}$. The nonlinear equation solver iterates on expression (11.2.1) until convergence, given an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{\boldsymbol{\Gamma}}(1)^{-1} B_0^{-1}$. As noted in Chapter 9, 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. The estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix}$$

obtained using this procedure exactly matches the earlier estimate based on $\widehat{\Theta}(1)$. Indeed, in this simple example, there is no advantage to using a nonlinear equation solver, given the recursive structure of $\Theta(1)$.

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, we have $\Theta(1) = \boldsymbol{\Gamma}(1)^{-1} B_0^{-1} \Sigma_w^{1/2}$, where $\Sigma_w^{1/2}$ is obtained by taking the square

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

root of the diagonal elements of Σ_w. Having directly imposed the zero restrictions on the off-diagonal elements of Σ_w and ones on the diagonal elements of B_0, the objective is to find the unknown elements of B_0 and Σ_w such that

$$\begin{bmatrix} \text{vech}(B_0^{-1}\Sigma_w B_0^{-1\prime} - \widehat{\Sigma}_u) \\ \theta_{12}(1) \end{bmatrix} = 0, \tag{11.2.2}$$

given estimates of the reduced-form parameters and an initial guess for the off-diagonal elements of B_0 and for the diagonal elements of Σ_w.

Expression (11.2.2) may be iterated until convergence using a nonlinear equation solver, yielding $\widehat{\Theta}(1) = \widehat{\mathbf{\Gamma}}(1)^{-1}\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$. The implied estimate of the structural impact multiplier matrix, $\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$, is the same as when using the Cholesky decomposition.

Estimating the Model Using the Algorithm of Rubio-Ramírez et al. (2010). Rubio-Ramírez, Waggoner, and Zha (2010) propose an algorithm for estimating exactly identified structural VAR models with possibly nonrecursive structure that tends to be computationally more efficient than using a nonlinear equation solver. This algorithm may be adapted easily to allow for the imposition of long-run restrictions. Let $\Sigma_w = I_K$. When imposing restrictions on the response of variable i at the infinite horizon, it is assumed that the i^{th} variable is expressed in first differences. The long-run structural impulse response of variable k to structural shock l corresponds to the element in row k and column l of the matrix

$$\Theta(1) = \left(I_K - \sum_{i=1}^{p} \mathbf{\Gamma}_i \right)^{-1} B_0^{-1},$$

which has K columns corresponding to the K structural shocks and K rows because we need to allow for restrictions on the long-run responses of any of the K variables. A candidate draw for $\Theta(1)$, denoted L_∞, is constructed from the reduced-form estimates $\widehat{\mathbf{\Gamma}}_i$, given an initial guess of the structural impact multiplier matrix of $B_0^{-1} = \text{chol}(\widehat{\Sigma}_u)$, denoted by L_0.

Zero restrictions on the long-run structural responses are represented by matrices Z_j for $1 \leq j \leq K$. The number of columns in Z_j is equal to the number of rows in L_∞. 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 is $z = \sum_{j=1}^{K} z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j L_\infty e_j = 0$ for $1 \leq j \leq K$, where e_j is the j^{th} column of I_K. In what follows, let Z_j represent the zero restrictions with the equations of the model ordered such that $z_j \leq K - j$, where $1 \leq j \leq K$. Observe that $Z_j L_\infty e_j = 0$ and $\bar{Z}_j L_\infty e_j = 0$ 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 Long-Run Restrictions.

1. Let $j = 1$.
2. If $j = 1$, then $Q'_j = \bar{Z}_1 L_\infty$. For $j > 1$, form the matrix
$$Q'_j = \begin{bmatrix} \bar{Z}_j L_\infty \\ q'_1 \\ \vdots \\ q'_{j-1} \end{bmatrix}.$$
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

$$Q' = \begin{bmatrix} q_1 & \cdots & q_K \end{bmatrix}$$

such that $L_0 Q'$ represents a 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.

Applying this algorithm to the example of Galí (1999), consider the candidate solution $L_0 = \text{chol}(\widehat{\Sigma}_u)$ and hence

$$L_\infty = \begin{bmatrix} 0.4890 & -0.3741 \\ 0.4579 & 1.0504 \end{bmatrix},$$

obtained as $L_\infty = (I_2 - \widehat{\boldsymbol{\Gamma}}_1 - \widehat{\boldsymbol{\Gamma}}_2 - \widehat{\boldsymbol{\Gamma}}_3 - \widehat{\boldsymbol{\Gamma}}_4)^{-1}\text{chol}(\widehat{\Sigma}_u)$. Let the first column refer to the long-run responses of the two model variables to the non-technology shock and the second column to the long-run responses to the technology shock.

Given that there is only one identifying restriction on the long-run response of productivity to the non-technology shock, the identifying restrictions can be expressed as

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

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

$$\bar{Z}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Since all rows of Z_2 are zeros, there is no $\bar{Z}_2$.

For $j = 1$, we obtain

$$Q_1' = \bar{Z}_1 L_\infty = \begin{bmatrix} 0.4890 & -0.3741 \end{bmatrix}.$$

The QR decomposition for Q_1 yields

$$\widetilde{Q} = \begin{bmatrix} -0.7942 & 0.6076 \\ 0.6076 & 0.7942 \end{bmatrix},$$

so $q_1' = (0.6076, 0.7942)$, where q_1 by construction relates to the non-technology shock. For $j = 2$, we obtain

$$Q_2' = q_1' = \begin{bmatrix} 0.6076 & 0.7942 \end{bmatrix}$$

because $\bar{Z}_2$ is the empty matrix. The QR decomposition for Q_2 yields

$$\widetilde{Q} = \begin{bmatrix} -0.6076 & -0.7942 \\ -0.7942 & 0.6076 \end{bmatrix},$$

so $q_2' = (-0.7942, 0.6076)$, where q_2 by construction relates to the technology shock.

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

$$Q' = \begin{bmatrix} q_2 & q_1 \end{bmatrix} = \begin{bmatrix} -0.7942 & 0.6076 \\ 0.6076 & 0.7942 \end{bmatrix},$$

where the ordering of q_1 and q_2 has been reversed to match the ordering of the structural shocks in the original model specification of Galí (1999). The implied solution for the structural impact multiplier matrix is

$$L_0 Q' = \begin{bmatrix} -0.5384 & 0.4119 \\ 0.4971 & 0.5359 \end{bmatrix}.$$

After flipping the sign of the first column such that positive technology shocks are associated with increases in productivity, we obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix},$$

which matches the earlier solution. Of course, there is no advantage over the use of the Cholesky decomposition of the long-run variance-covariance matrix in the simple example considered here, but a similar algorithm could be used if the model is identified by both short-run and long-run restrictions or if the long-run variance covariance matrix does not have a lower-triangular representation, as illustrated in Section 11.3.

11.2.2 Full Information Maximum Likelihood Estimation

In this section we postulate that $\Sigma_w = I_K$, while the diagonal elements of B_0 remain unrestricted. If the errors are normally distributed, the reduced-form

VECM can be estimated using Johansen's ML method presented in Chapter 3. If the errors are not Gaussian, this method under suitable conditions can be interpreted as a quasi-ML method. The log-likelihood function is

$$\log l = \text{constant} - \frac{T}{2}\log(\det(\Sigma_u)) - \frac{1}{2}\sum_{t=1}^{T} u_t' \Sigma_u^{-1} u_t.$$

Substituting $\Sigma_u = B_0^{-1}B_0^{-1\prime}$ and concentrating out the intercept and slope parameters of the reduced form by replacing them by their ML estimates yields the concentrated log likelihood as a function of the structural parameters

$$\log l_c(B_0) = \text{constant} + \frac{T}{2}\log(\det(B_0)^2) - \frac{T}{2}\mathrm{tr}(B_0' B_0 \widetilde{\Sigma}_u), \qquad (11.2.3)$$

where $\widetilde{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$ is the usual ML estimator of the reduced-form residual covariance matrix. This function has to be maximized with respect to B_0 subject to identifying restrictions on specific elements of $\boldsymbol{\Upsilon} = \widehat{\boldsymbol{\Xi}}B_0^{-1}$, where

$$\widehat{\boldsymbol{\Xi}} = \widehat{\beta}_\perp \left[\widehat{\alpha}_\perp' \left(I_K - \sum_{i=1}^{p-1}\widehat{\boldsymbol{\Gamma}}_i\right)\widehat{\beta}_\perp\right]^{-1}\widehat{\alpha}_\perp',$$

if $0 < r < K$ and $\widehat{\boldsymbol{\Xi}} = \widehat{\boldsymbol{\Gamma}}(1)^{-1}$ if $r = 0$. Maximizing the log likelihood may require iterative methods.

For example, for the Galí model $r = 0$ and the long-run restriction can be written as

$$\widehat{\gamma}(1)_{12}b_{11,0} + \widehat{\gamma}(1)_{22}b_{12,0} = 0,$$

where $\widehat{\gamma}(1)_{ij}$ denotes the ij^{th} element of $\widehat{\boldsymbol{\Gamma}}(1)$, where $b_{ij,0}$ is the ij^{th} element of B_0, and where we have exploited the fact that the triangularity of $\boldsymbol{\Upsilon}$ is equivalent to $\boldsymbol{\Upsilon}^{-1} = B_0\boldsymbol{\Gamma}(1)$ being triangular. In practice, the ML estimate of B_0 in the Galí model can be constructed as

$$\widetilde{B}_0 = \mathrm{chol}(\widehat{\boldsymbol{\Gamma}}(1)\widetilde{\Sigma}_u^{-1}\widehat{\boldsymbol{\Gamma}}(1)')\widehat{\boldsymbol{\Gamma}}(1)^{-1}.$$

We obtain

$$\widetilde{B}_0 = \begin{bmatrix} 1.1115 & -0.8542 \\ 1.0308 & 1.1166 \end{bmatrix}.$$

Except for the degrees-of-freedom adjustment for $\widehat{\Sigma}_u$, this estimate equals the inverse of the estimate $\widehat{B}_0^{-1}$ discussed earlier in this section.

11.2.3 Instrumental Variable Estimation

Pagan and Pesaran (2008) point out that in some cases models with cointegration relations and long-run restrictions may also be estimated by IV estimation methods with the cointegration relations used as instruments. They consider a system with cointegrating rank r, $0 < r < K$, with exactly r structural shocks with purely transitory effects only and $K - r$ shocks with permanent effects. If the cointegration relations are known, the cointegration relations can be used as instruments for estimating some of the structural parameters. Suppose we place the permanent shocks first in the w_t vector and the transitory shocks last by partitioning w_t into $w_t = (w_t^{p\prime}, w_t^{tr\prime})'$, where w_t^p is $(K - r)$-dimensional and w_t^{tr} is r-dimensional. Then the $K \times r$ matrix of loadings, $\alpha^\dagger$, in the structural VECM has a $(K - r) \times r$ block of zeros,

$$\alpha^\dagger = \begin{bmatrix} 0_{(K-r)\times r} \\ \alpha^\dagger_{(r\times r)} \end{bmatrix}.$$

In other words, the cointegration relations do not appear in the first $K - r$ structural equations of the system. Partitioning the structural VECM,

$$B_0 \Delta y_t = \alpha^\dagger \beta' y_{t-1} + \mathbf{\Gamma}_1^\dagger \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}^\dagger \Delta y_{t-p+1} + w_t,$$

where the diagonal elements of B_0 have been normalized to unity, into the subset of equations that include an error correction term and the subset of equations that do not, yields

$$B_0^p \Delta y_t = \mathbf{\Gamma}_1^{\dagger p} \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}^{\dagger p} \Delta y_{t-p+1} + w_t^p, \tag{11.2.4}$$

$$B_0^{tr} \Delta y_t = \alpha^\dagger_{(r\times r)} \beta' y_{t-1} + \mathbf{\Gamma}_1^{\dagger tr} \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}^{\dagger tr} \Delta y_{t-p+1} + w_t^{tr}, \tag{11.2.5}$$

where w_t^p refers to shocks with permanent effects on the level of y_t and w_t^{tr} to shocks with purely transitory effects.

The objective is to estimate the unknown elements of B_0^p and B_0^{tr}. This involves instrumenting for the elements of Δy_t with unknown coefficients. We first estimate the set of equations (11.2.4). If the r cointegration relations $\beta' y_{t-1}$ are known, they can be used as instruments for the elements of Δy_t with unknown coefficients in equation (11.2.4). Instrumental variable estimation requires the instrument to be correlated with the variable to be instrumented, but uncorrelated with the error term of the regression to be estimated. The instruments $\beta' y_{t-1}$ satisfy the first condition because they are correlated with the lagged Δy_t and hence with the relevant elements of Δy_t. They satisfy the second condition because they are lagged and, hence, uncorrelated with w_t^p in equation (11.2.4). Having estimated equation (11.2.4), we next instrument for the elements of Δy_t with unknown coefficients in equation (11.2.5). For this purpose, the residuals $\widehat{w}_t^p$ from equation (11.2.4) are natural instruments. These instruments are valid because they are uncorrelated with w_t^{tr}.

In this setup, the diagonal elements of B_0 are usually normalized to one and the variances of the structural errors are left unrestricted. Hence, the

$(K-r) \times K$ matrix B_0^p has $(K-1)(K-r)$ unknown elements that must be estimated. The r cointegration relations may not provide enough instruments to achieve full identification, and other restrictions may be required for estimating the structural parameters. One may, for example, use exclusion restrictions for the impact effects of the structural shocks. Estimation under joint restrictions on long-run and short-run effects is discussed in Section 11.3. A case in which enough instruments are available for estimating B_0^p arises when the cointegrating rank $r = K - 1$ and, hence, there is just one common trend and one permanent shock.

Pagan and Pesaran (2008) use the Blanchard-Quah model as an example. This example differs from the model in Galí (1999) in that the second variable in the level representation is $I(0)$ rather than $I(1)$. Recall from Chapter 10 that the Blanchard-Quah model consists of the log of U.S. real GDP (gdp_t) and the unemployment rate (ur_t). Thus, $y_t = (gdp_t, ur_t)'$. The first variable, gdp_t, is $I(1)$, whereas the second variable, ur_t, is $I(0)$. Thus, there is one trivial cointegration relation consisting of the second variable only. We write

$$\beta' y_t = [0, 1] y_t = ur_t.$$

Blanchard and Quah (1989) consider an aggregate supply shock (w_t^s) with permanent effects on gdp_t and a transitory aggregate demand shock (w_t^d). Using the insights from Pagan and Pesaran (2008), the two structural equations can be written as

$$\Delta gdp_t = -b_{12,0} \Delta ur_t + \sum_{i=1}^{p-1} \gamma_{11,i}^{\dagger} \Delta gdp_{t-i} + \sum_{i=1}^{p-1} \gamma_{12,i}^{\dagger} \Delta ur_{t-i} + w_t^s,$$

$$\begin{aligned} \Delta ur_t = {} & -b_{21,0} \Delta gdp_t + \alpha^{\dagger} ur_{t-1} + \sum_{i=1}^{p-1} \gamma_{21,i}^{\dagger} \Delta gdp_{t-i} \\ & + \sum_{i=1}^{p-1} \gamma_{22,i}^{\dagger} \Delta ur_{t-i} + w_t^d. \end{aligned}$$

In the first equation, ur_{t-1} can be used as an instrument for Δur_t in estimating $b_{12,0}$. The structural residual $\widehat{w}_t^s$ can be used as an instrument for Δgdp_t in estimating $b_{21,0}$ in the second equation. Obviously, w_t^s is correlated with Δgdp_t and uncorrelated with the error term of the second equation by construction.

More generally, if the structural parameters in the first set of equations (11.2.4) can all be estimated, the corresponding residuals $\widehat{w}_t^p$ can be used as instruments for estimating the structural parameters in B_0^{tr}, because the permanent shocks w_t^p are uncorrelated with the transitory shocks w_t^{tr}.

It is worth remembering that this IV method for estimating the structural parameters in the presence of long-run restrictions hinges on a set of important assumptions:

- The number of permanent shocks is equal to the number of common trends and the number of transitory shocks is equal to the cointegrating rank of the system. This condition is violated, for example, if the cointegration rank is greater than zero, but all structural shocks in the model have permanent effects (see Section 10.2.2).
- The cointegration relations are known (or super-consistently estimated).
- The r instruments provided by the cointegration relations suffice to estimate the structural parameters in the first $K - r$ equations.
- The permanent shocks suffice as instruments for estimating the structural parameters in the last r equations.

If these conditions are not satisfied, further identifying assumptions or instruments from other sources are needed.

11.3 Models Subject to Long-Run and Short-Run Restrictons

In this section, we consider two representative examples of structural VAR models that combine long-run and short-run identifying restrictions. The first example illustrates how these restrictions may be combined when estimating the structural model in its VAR representation. The second example illustrates how long-run and short-run restrictions may be imposed within the VECM framework.

11.3.1 Estimating the Model in VAR Representation

Consider a stylized VAR(4) model of U.S. monetary policy with only three quarterly variables. This example is based on Rubio-Ramírez, Waggoner, and Zha (2010). Let $z_t = (\Delta gnp_t, i_t, \Delta p_t)' \sim I(0)$, where gnp_t denotes the log of U.S. real GNP, p_t the corresponding GNP deflator in logs, and i_t the federal funds rate, averaged by quarter. The estimation period is restricted to 1954q4-2007q4 in order to exclude the recent period of unconventional monetary policy measures.

The Federal Reserve Board is assumed to control the interest rate by setting the policy innovation after observing the forecast errors for deflator inflation and real GNP growth. The model is fully identified and includes an aggregate demand shock and an aggregate supply shock in addition to the monetary policy shock. The monetary policy shock does not affect real GNP either within the current quarter or in the long run. The only shock to affect the log-level of real GNP in the long run is the aggregate supply shock. Defining $w_t = (w_t^{\text{policy}}, w_t^{\text{AD}}, w_t^{\text{AS}})'$, the identifying restrictions can be summarized

 https://doi.org/10.1017/9781108164818.012

as

$$B_0^{-1} = \begin{bmatrix} 0 & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$$

and

$$\Theta(1) = A(1)^{-1}B_0^{-1} = \begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

Since z_t is $I(0)$, the long-run restrictions are imposed on the cumulated impulse responses.

The unrestricted reduced-form estimates of a VAR(4) model with constant term are

$$\widehat{A}_1 = \begin{bmatrix} 0.2230 & 0.0097 & 0.3969 \\ 0.3147 & 1.0969 & 0.5979 \\ 0.0012 & 0.0636 & 0.4096 \end{bmatrix},$$

$$\widehat{A}_2 = \begin{bmatrix} 0.2143 & -0.3862 & 0.1360 \\ 0.1867 & -0.4860 & 0.5037 \\ -0.0174 & -0.0510 & 0.2350 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} -0.0053 & 0.3407 & -0.5354 \\ 0.0275 & 0.4832 & -0.3212 \\ 0.0115 & -0.0052 & 0.0815 \end{bmatrix},$$

$$\widehat{A}_4 = \begin{bmatrix} -0.0411 & 0.0013 & -0.0268 \\ -0.0226 & -0.1642 & -0.3320 \\ 0.0667 & -0.0137 & 0.2463 \end{bmatrix},$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.6031 & 0.0795 & -0.0214 \\ 0.0795 & 0.6565 & 0.0375 \\ -0.0214 & 0.0375 & 0.0684 \end{bmatrix}.$$

Method-of-Moments Estimation

Estimating the Model Using a Nonlinear Equation Solver. In situations in which there are additional restrictions beyond the long-run identifying restriction, the structural VAR can no longer be estimated based on a lower-triangular Cholesky decomposition of the long-run variance-covariance matrix. However, estimates may still be constructed with the help of a nonlinear equation solver.

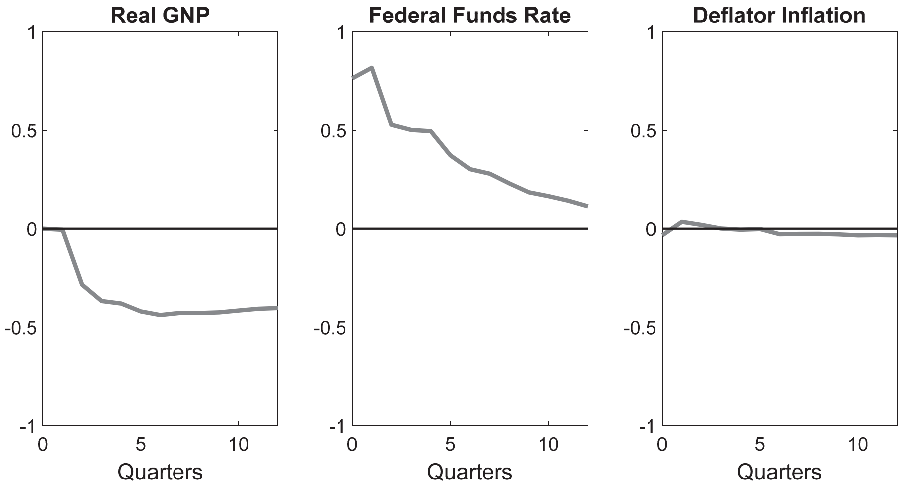

Figure 11.2. Responses to an unexpected U.S. monetary policy tightening.

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

$$\begin{bmatrix} \text{vech}(B_0^{-1}B_0^{-1\prime} - \widehat{\Sigma}_u) \\ b_0^{11} \\ \theta_{11}(1) \\ \theta_{12}(1) \end{bmatrix} = 0, \tag{11.3.1}$$

where the normalization $\Sigma_w = I_3$ has been imposed. The nonlinear equation solver iterates on expression (11.3.1) until convergence, given an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{A}(1)^{-1}B_0^{-1}$, where $\widehat{A}(1)^{-1} = (I_3 - \sum_{i=1}^{4} \widehat{A}_i)^{-1}$. The implied estimate of the structural impact multiplier matrix is

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0 & 0.5845 & 0.5113 \\ 0.7625 & 0.2445 & -0.1239 \\ -0.0332 & 0.1491 & -0.2123 \end{bmatrix},$$

where the signs of the second column have been normalized such that a positive aggregate demand shock does not lower real GNP and inflation. Figure 11.2 shows the implied structural impulse response functions. An unexpected monetary policy tightening is associated with a persistent decline in real GNP but little response in inflation.

Estimating the Model as in Rubio-Ramírez et al. (2010). The same model may also be estimated using a variant of the algorithm proposed in Rubio-Ramírez, Waggoner, and Zha (2010). As before, $\Sigma_w = I_K$. When imposing restrictions on the infinite horizon, it is assumed that the i^{th} variable is expressed in first differences. The long-run structural impulse responses correspond to the element in row i and column j of the matrix

$$\Theta(1) = \left(I_K - \sum_{i=1}^{p} A_i\right)^{-1} B_0^{-1}.$$

A candidate draw for $\Theta(1)$, denoted L_∞, is constructed from the reduced-form estimates $\widehat{A}_i$, given an initial guess of the structural impact multiplier matrix of $B_0^{-1} = \text{chol}(\widehat{\Sigma}_u)$, denoted by L_0.

It is convenient to stack the structural impulse response functions into a single matrix denoted by $\mathbf{L}$. If restrictions are imposed at horizons zero and infinity, then

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_\infty \end{bmatrix}$$

has K columns corresponding to the K structural shocks and $2K$ rows because we consider restrictions at two horizons for K variables.

Zero restrictions on the structural responses can be represented by matrices Z_j for $1 \leq j \leq K$. The number of columns in Z_j is equal to the number of rows in $\mathbf{L}$. 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 is $z = \sum_{j=1}^{K} z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j \mathbf{L} e_j = 0$ for $1 \leq j \leq K$, where e_j is the j^{th} column of 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 $Z_j \mathbf{L} e_j = 0$ and $\bar{Z}_j \mathbf{L} e_j = 0$ 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 Models That Are Exactly Identified by a Combination of Short-Run and Long-Run Restrictions.

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

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

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

$$Q' = \begin{bmatrix} q_1 & \cdots & q_K \end{bmatrix}$$

such that $L_0 Q'$ represents a 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.

Returning to the example of the monetary policy VAR model, we need to impose restrictions on the structural impulse response functions at horizons 0 and ∞. We first define a candidate solution for the structural impulse response functions of interest as

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_\infty \end{bmatrix} = \begin{bmatrix} 0.7766 & 0 & 0 \\ 0.1024 & 0.8037 & 0 \\ -0.0276 & 0.0501 & 0.2552 \\ 0.9396 & -0.3398 & -1.0189 \\ 6.2126 & 6.3718 & 15.0626 \\ -0.3180 & -0.4124 & 3.5063 \end{bmatrix}.$$

The exclusion restrictions for the monetary policy shock and the aggregate demand shock can be written as

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$Z_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad Z_3 = 0_{3\times 6},$$

where Z_1 embodies the restriction that w_t^{policy} does not affect real GDP in the short run or in the long run, and Z_2 embodies the restriction that w_t^{AD} does not affect real GDP in the long run. By deleting the rows of zeros from the Z_js, we obtain

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

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

For $j = 1$, we obtain

$$Q_1' = \bar{Z}_1 \mathbf{L} = \begin{bmatrix} 0.7766 & 0 & 0 \\ 0.9396 & -0.3398 & -1.0189 \end{bmatrix}.$$

The QR decomposition for Q_1 yields

$$\widetilde{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.3163 & -0.9487 \\ 0 & -0.9487 & 0.3163 \end{bmatrix},$$

so $q_1' = (0, -0.9487, 0.3163)$. For $j = 2$, we obtain

$$Q_2' = \begin{bmatrix} \bar{Z}_2 \mathbf{L} \\ q_1' \end{bmatrix} = \begin{bmatrix} 0.9396 & -0.3398 & -1.0189 \\ 0 & -0.9487 & 0.3163 \end{bmatrix}.$$

Applying the QR decomposition to Q_2 as before yields an orthogonal matrix, the last column of which is $q_2 = (0.7527, 0.2083, 0.6246)'$. Finally, we form

$$Q_3' = \begin{bmatrix} q_1' \\ q_2' \end{bmatrix} = \begin{bmatrix} 0 & -0.9487 & 0.3163 \\ 0.7527 & 0.2083 & 0.6246 \end{bmatrix}$$

because $\bar{Z}_3$ is the empty matrix. The last column of the orthogonal matrix obtained by applying the QR decomposition to Q_3 yields $q_3 = (-0.6584, 0.2381, 0.7140)'$.

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

$$Q' = [q_1 \;\; q_2 \;\; q_3] = \begin{bmatrix} 0 & 0.7527 & -0.6584 \\ -0.9487 & 0.2083 & 0.2381 \\ 0.3163 & 0.6246 & 0.7140 \end{bmatrix}.$$

The implied solution for the structural impact multiplier matrix is

$$L_0 Q' = \begin{bmatrix} 0 & 0.5845 & -0.5113 \\ -0.7625 & 0.2445 & 0.1239 \\ 0.0332 & 0.1491 & 0.2123 \end{bmatrix}.$$

After flipping the signs of the first column such that a positive interest rate innovation raises the interest rate and flipping the sign of the last column such that a positive aggregate supply shock is not associated with lower real GNP on impact, we obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0 & 0.5845 & 0.5113 \\ 0.7625 & 0.2445 & -0.1239 \\ -0.0332 & 0.1491 & -0.2123 \end{bmatrix},$$

which matches the estimate obtained using the nonlinear equation solver.

Other Estimation Methods. It is also possible to estimate our example model by FIML. In contrast, FIML estimation is straightforward. Because the VECM representation of the structural VAR model accommodates more general models, the reader is referred to our discussion of the FIML estimator in the following section.

11.3.2 Estimating the Model in VECM Representation

Recall from Section 10.3.1 that King, Plosser, Stock, and Watson (1991) considered a baseline quarterly VAR(2) model in levels for the log of U.S. real GNP (gnp_t), the log of real consumption (c_t), and the log of real investment (inv_t). The estimation period is 1947q1-1988q4. In this model all variables are affected by the same productivity shock in the long run. This means that the VAR model for $y_t = (gnp_t, c_t, inv_t)'$ may equivalently be written as a reduced-form VECM as in the previous section with known cointegration rank $r = 2$ and unknown cointegrating matrix β:

$$\Delta y_t = \nu + \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + u_t.$$

Let $\Sigma_w = I_3$. King et al. are interested in using this model to identify the responses to the common productivity shock. Assuming that there are two transitory shocks that are placed last in the vector of structural shocks, we obtain

$$\mathbf{\Upsilon} = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix},$$

where $*$ denotes an unrestricted element. No further restrictions are necessary to identify the permanent shock.

If we are only interested in quantifying the effects of the common productivity shock, we may, without loss of generality, impose an arbitrary exclusion restriction on the last two columns of B_0^{-1}. Without loss of generality, we choose to restrict b_0^{23} to zero for expository purposes such that

$$B_0^{-1} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & * \end{bmatrix},$$

as in Lütkepohl (2005, chapter 9).

The reduced-form ML estimates of the VECM obtained by the Johansen method are

$$\widetilde{\nu} = \begin{bmatrix} -0.009 \\ -0.028 \\ -0.301 \end{bmatrix},$$

$$\widetilde{\alpha} = \begin{bmatrix} -0.225 & 0.204 \\ -0.062 & 0.072 \\ -0.112 & 0.255 \end{bmatrix},$$

$$\widetilde{\beta}' = \begin{bmatrix} 1 & 0 & -1.020 \\ 0 & 1 & -1.099 \end{bmatrix},$$

$$\widetilde{\boldsymbol{\Gamma}}_1 = \begin{bmatrix} 0.123 & 0.090 & 0.159 \\ 0.208 & -0.207 & 0.025 \\ 0.703 & -0.169 & 0.331 \end{bmatrix},$$

and

$$\widetilde{\Sigma}_u \times 1000 = \begin{bmatrix} 0.1259 & 0.0395 & 0.1580 \\ 0.0395 & 0.0547 & 0.0684 \\ 0.1580 & 0.0684 & 0.4897 \end{bmatrix},$$

where the cointegrating vectors in $\widetilde{\beta}'$ have been normalized for expository purposes.

Method-of-Moments Estimation. As in the earlier example, we proceed in two steps. Having estimated the reduced form of the VECM, the objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \operatorname{vech}(B_0^{-1}B_0^{-1\prime} - \widetilde{\Sigma}_u) \\ b_0^{23} \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{22} \\ \zeta_{23} \\ \zeta_{32} \\ \zeta_{33} \end{bmatrix} = 0, \tag{11.3.2}$$

where ζ_{ij} is the ij^{th} element of $\boldsymbol{\Upsilon}$. The nonlinear equation solver iterates on expression (11.3.2), given an initial guess for B_0^{-1} such that $\boldsymbol{\Upsilon} = \widetilde{\boldsymbol{\Xi}} B_0^{-1}$, where

$$\widetilde{\boldsymbol{\Xi}} = \widetilde{\beta}_\perp \left[\widetilde{\alpha}'_\perp \left(I_K - \sum_{i=1}^{p-1} \widetilde{\boldsymbol{\Gamma}}_i \right) \widetilde{\beta}_\perp \right]^{-1} \widetilde{\alpha}'_\perp,$$

which in this example reduces to

$$\widetilde{\boldsymbol{\Xi}} = \widetilde{\beta}_\perp \left[\widetilde{\alpha}'_\perp (I_K - \widetilde{\boldsymbol{\Gamma}}_1) \widetilde{\beta}_\perp \right]^{-1} \widetilde{\alpha}'_\perp.$$

Here $\widetilde{\beta}_\perp$ and $\widetilde{\alpha}_\perp$ are orthogonal complements of $\widetilde{\beta}$ and $\widetilde{\alpha}$, respectively. For a $K \times r$ matrix β of rank r the orthogonal complement $\beta_\perp$ is computed by evaluating the singular value decomposition

$$\beta = Q \Lambda P,$$

where Q and P are orthogonal matrices of dimension $K \times K$ and $r \times r$, respectively, and Λ is a $K \times r$ diagonal matrix. We may choose $\beta_\perp$ as the last $K - r$

columns of Q.[2] Thus,

$$\widetilde{\beta}_\perp = \begin{pmatrix} 0.566 \\ 0.610 \\ 0.555 \end{pmatrix}$$

and

$$\widetilde{\alpha}_\perp = \begin{pmatrix} 0.219 \\ -0.971 \\ 0.098 \end{pmatrix}.$$

Hence,

$$\widetilde{\boldsymbol{\Xi}} = \begin{bmatrix} -0.247 & 1.095 & -0.110 \\ -0.266 & 1.180 & -0.119 \\ -0.242 & 1.074 & -0.108 \end{bmatrix}.$$

The estimate of the structural impact multiplier matrix obtained using the nonlinear equation solver is

$$\widetilde{B}_0^{-1} = \begin{bmatrix} 0.0008 & 0.0103 & -0.0045 \\ -0.0060 & 0.0043 & 0.0000 \\ 0.0026 & 0.0196 & 0.0100 \end{bmatrix}$$

and the long-run effects of the structural shocks on the level variables in this VECM are given by

$$\widetilde{\boldsymbol{\Upsilon}} = \widetilde{\boldsymbol{\Xi}}\widetilde{B}_0^{-1} = \begin{bmatrix} -0.0071 & 0 & 0 \\ -0.0076 & 0 & 0 \\ -0.0069 & 0 & 0 \end{bmatrix}.$$

Figure 11.3 shows the implied responses of logged GNP, consumption, and investment to a common productivity shock.

Alternatively, if β' is known to be

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix},$$

after imposing these restrictions on the VECM, the reduced-form model may be estimated by unrestricted ML, which allows us to replace $\widetilde{\Sigma}_u$ in expression (11.3.2) by the corresponding estimate from the restricted reduced-form VECM and to proceed as before.

An even easier approach, when β is known, is to express the VECM as a VAR model for

$$z_t = \begin{bmatrix} \Delta gnp \\ c_t - gnp_t \\ inv_t - gnp_t \end{bmatrix} \sim I(0)$$

and to impose a lower-triangular structure on $\Theta(1)$.

[2] In MATLAB the singular value decomposition can be computed by the function *svd*.

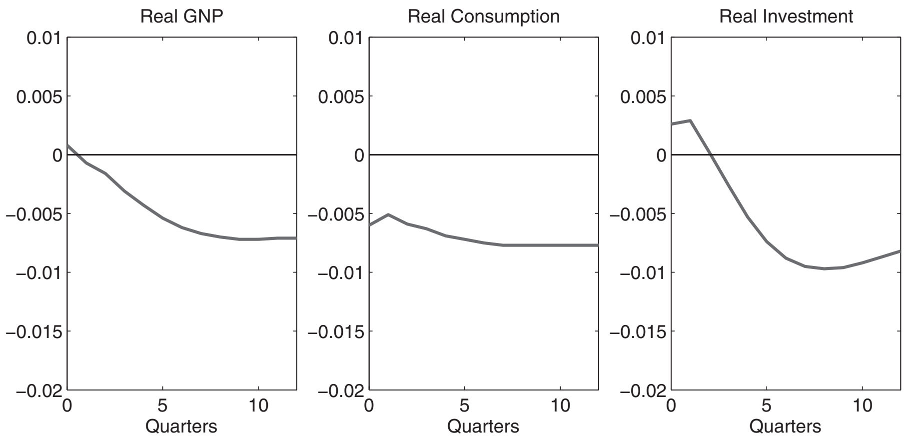

Figure 11.3. Responses to a productivity shock in the baseline model of King et al. (1991).

Full Information Maximum Likelihood Estimation. The Gaussian log-likelihood function can be set up as in Section 11.2.2. The concentrated log-likelihood is

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

where $\widetilde{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$. Maximizing this function with respect to B_0 subject to the structural restrictions on B_0^{-1} and $\boldsymbol{\Upsilon}$ is a numerical optimization problem. In the just-identified case, the solution for B_0 is obtained with a nonlinear equation solver. More generally, in the overidentified case, the likelihood has to be maximized numerically subject to all identifying restrictions.

In the empirical example based on King, Plosser, Stock, and Watson (1991) FIML estimation yields

$$\widetilde{B}_0 = \begin{bmatrix} 34.905 & -155.206 & 15.707 \\ 48.705 & 15.991 & 21.917 \\ -104.537 & 9.010 & 52.958 \end{bmatrix}$$

which is the inverse of the $\widetilde{B}_0^{-1}$ that we obtained using the method of moments.

Instrumental Variable Estimation. As in the earlier monetary policy example, the model of King, Plosser, Stock, and Watson (1991) imposes identifying restrictions on B_0^{-1}, making it impossible to estimate this model by IV methods. There are other models, however, that combine long-run restrictions with identifying restrictions on B_0 (see, e.g., Pagan and Robertson 1998). Although it is possible in principle to generalize the IV approach of Pagan and Pesaran (2008) in Section 11.2.3 to accommodate additional exclusion restrictions on B_0, such applications appear to be rare.

11.4 Practical Limitations of Long-Run Restrictions

In Chapter 10 we already reviewed a number of conceptual challenges in imposing long-run identifying restrictions. In addition, there are challenges that arise when implementing these methods in practice. To build some intuition, consider a bivariate model with $\Theta(1)$ restricted to be lower triangular. Recall that

$$\Theta(1) = \mathbf{\Gamma}(1)^{-1} B_0^{-1} = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix},$$

where $\mathbf{\Gamma}(1) = (I_2 - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i)$, which may be estimated as

$$\widehat{\Theta}(1) = \begin{bmatrix} \widehat{\theta}_{11}(1) & 0 \\ \widehat{\theta}_{21}(1) & \widehat{\theta}_{22}(1) \end{bmatrix} = \text{chol}\left(\widehat{\mathbf{\Gamma}}(1)^{-1} \widehat{\Sigma}_u \widehat{\mathbf{\Gamma}}(1)^{-1\prime}\right)$$

such that

$$\widehat{B}_0^{-1} = \widehat{\mathbf{\Gamma}}(1)\widehat{\Theta}(1).$$

Note that in constructing $\widehat{\Theta}(1)$ we have to invert $\widehat{\mathbf{\Gamma}}(1)$. There are two potential problems that affect the reliability of the estimator of the long-run impulse responses. The first concern is that $\widehat{\mathbf{\Gamma}}(1)$ and hence $\widehat{\mathbf{\Gamma}}(1)^{-1}$ may be estimated imprecisely. For example, the matrix $\mathbf{\Gamma}(1)$ may be estimated imprecisely if the finite-order VAR model is a poor approximation to an infinite-order DGP.

The second concern is that the estimate of $\mathbf{\Gamma}(1)^{-1}$ (and hence the estimate of $\Theta(1)$) may be imprecise, even when $\mathbf{\Gamma}(1)$ is estimated accurately. This problem arises when $\widehat{\mathbf{\Gamma}}(1)$ is close to singular. If the matrix $\mathbf{\Gamma}(1)$ is not well-conditioned, the estimator of its inverse may have high variance, even if the estimator of $\mathbf{\Gamma}(1)$ has low variance. This problem is analogous to estimating the inverse $1/\gamma$ of a scalar parameter γ. Suppose γ is close to zero. Even if a very precise estimator $\hat{\gamma}$ is used, which often provides estimates very close to zero, the inverse $1/\hat{\gamma}$ has very large values and hence may have very large variance. Any small deviations from the true value of γ will be magnified by the estimator $1/\hat{\gamma}$. Thus, $1/\gamma$ may be estimated very imprecisely.[3]

This near-singularity problem is known to arise, for example, if one of the variables in the stationary representation of the VAR model is persistent. In fact, in the limiting case of an exact unit root, we know already that $\mathbf{\Gamma}(1)$ is

[3] It is useful to keep in mind, however, that the inversion of a matrix may also improve the reliability of an estimator. For example, reconsider the earlier scalar example. If the true γ is large, then large deviations of $\hat{\gamma}$ from γ can result in small deviations of $1/\hat{\gamma}$ from $1/\gamma$.

singular (see Chapter 3). How empirically relevant this near-singularity problem is, of course, depends on the choice of the model variables.

These potential problems with the use of long-run restrictions have been studied in the literature from different angles. This section summarizes the main arguments.

11.4.1 Estimators of the Long-Run Multiplier Matrix May Be Unreliable

As stressed by Faust and Leeper (1997), estimating the long-run multiplier of a VAR model is akin to estimating the spectral density of the data at frequency zero based on the estimated coefficients of the finite-order VAR approximation. When this finite-order approximation is poor, the model approximation error may contaminate the implied estimate of the model's long-run behavior. One approach to this problem is to increase the autoregressive sieve lag order, possibly beyond typical choices employed in empirical work. Another response has been to develop alternative nonparametric estimators of the long-run multiplier matrix (see, e.g., Christiano, Eichenbaum, and Vigfusson 2006a, 2006b). No comprehensive comparison of these approaches in finite samples exists at this point.

Some simulation evidence based on DSGE model DGPs is provided in Mertens (2012) who concludes that nonparametric estimators of the long-run variance-covariance matrix need not improve on conventional estimators. Like much of this literature, however, Mertens' work does not make any adjustments for small-sample bias in the LS estimates, nor do his lag-order choices recognize that the approximating VAR models are autoregressive sieves.

11.4.2 Lack of Power

Obtaining an accurate estimate of the impulse response at the infinite horizon amounts to pinning down the dominant autoregressive root of the process. We know that it is not possible to estimate accurately the long-run behavior of an economic time series from a short time span of data. For that reason one would expect estimates from such structural VAR models to be imprecise. This problem may persist even in large samples, as stressed by Faust and Leeper (1997). Indeed, it is well established that properly constructed confidence intervals for VAR models based on long-run restrictions tend to be wide compared with models based on short-run identifying restrictions (see, e.g., Erceg, Guerrieri, and Gust 2005; Christiano, Eichenbaum, and Vigfusson 2006b).[4]

[4] One way of mitigating this problem is to replace the long-run identifying restrictions by finite-horizon restrictions. For example, we may identify a technology shock as the structural shock that has the most explanatory power for the level of real GDP at a large but finite horizon of 10 years (see, e.g., Francis, Owyang, Roush, and DiCecio 2014).

The mirror image of wide confidence intervals is a lack of power against alternative models. Faust and Leeper argue that, under long-run identifying assumptions, confidence intervals for structural impulse responses will asymptotically rule out false models only at a rate equal to the rejection rate of the true model (Faust and Leeper 1997, p. 347). In other words, models based on long-run identifying restrictions lack power. This statement allows for general models under the alternative.

This point is investigated further in Gust and Vigfusson (2009). They investigate the ability of impulse-response based inference in VAR models based on long-run restrictions to discriminate between alternative structural models. They make the case that, when the set of possible DGPs is restricted to a plausible class of DSGE models that satisfy the long-run identifying restrictions, the rejection probabilities for false models are greater than the rejection rates based on data coming from the true model, indicating nontrivial power. Gust and Vigfusson also observe that tests of the sign and shape of impulse response functions may have higher power than tests based on impulse response confidence intervals.

11.4.3 Near-Observational Equivalence of Shocks with Permanent Effects and Shocks with Persistent Effects

Erceg, Guerrieri, and Gust (2005) show that models based on long-run identifying restrictions have difficulty disentangling shocks with permanent effects from shocks with persistent but not permanent effects. This result is not surprising given the observational equivalence of VAR models with unit roots and with roots close to unity in small samples.

11.4.4 Weak Instrument Problems

Another concern is that the $I(0)$ variables used to achieve identification often themselves are quite persistent. The unemployment rate used in the Blanchard and Quah (1989) model is a good example. Blanchard himself makes it clear that he is conflicted about whether this variable should be considered $I(0)$ or $I(1)$ (see Blanchard 1989). The problem is that conventional asymptotics tend to become unreliable when the data are highly persistent. Gospodinov (2010) proves that the impulse responses of interest are not consistently estimable under the long-run identification scheme when the process for this variable is parameterized as local to unity, and that standard confidence intervals are invalid in this case. Gospodinov represents the Blanchard and Quah (1989) model as

$$(1 - \mathbf{C}L)y_t = \Psi(L)u_t,$$

where $y_t = (gdp_t, ur_t)$, $\Psi(L)$ is an MA operator, and

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 + \lambda/T \end{bmatrix},$$

with $\lambda \leq 0$ a fixed constant. For $\lambda = 0$, this model would reduce to a specification with both variables in differences. For $\lambda < 0$, the ur_t process is restricted to be in the neighborhood of a difference stationary process. The larger T, the closer the ur_t process becomes to a unit root process. The purpose of this alternative representation is to serve as an asymptotic device that allows us to mimic a situation in which we are unable to tell whether the unemployment rate is $I(0)$ or $I(1)$ in finite samples. For a more detailed exposition of local-to-unity asymptotics in VAR models, the reader is referred to Chapter 12.

Within this framework, Gospodinov studies the statistical properties of the impulse response estimator. He expresses the impulse response estimator as a function of parameters to be estimated by instrumental variables. He demonstrates that these parameters cannot be consistently estimated because roots local to unity in the second equation of the Blachard-Quah model necessarily cause a weak-instrument problem. It follows immediately that conventional structural impulse response estimators are inconsistent and conventional impulse response confidence intervals are invalid. In response to this problem, Gospodinov (2010) proposes an alternative estimator of the structural impulse responses in the Blanchard and Quah (1989) model that does not suffer from the weak instrument problem. This impulse response estimator has much lower bias than conventional estimators in finite samples and lower variance.[5] Related work also includes Chevillon, Mavroeidis, and Zhan (2015).

11.5 Can Structural VAR Models Recover Responses in DSGE Models?

11.5.1 The Origin of This Controversy

The reason why this seemingly arcane question has risen to great prominence in recent years is that Galí (1999) in particular used evidence from structural VAR models based on long-run restrictions to make the case that the class of real business cycle (RBC) models is inconsistent with U.S. macroeconomic data. The RBC model, as proposed by Kydland and Prescott (1982), is an example of a DSGE model. Galí (1999) observed that evaluating the fit of this model based on unconditional second moments alone, as proposed by Kydland and Prescott (1982), can be misleading in that a DSGE model may do well

[5] The problem of weak instruments in long-run identified VAR models was first discussed in Pagan and Robertson (1998) and Cooley and Dwyer (1998). As noted by Watson (2006), conventional tests for weak instruments are not designed for highly persistent data and may not be able to detect weak instrument problems of the type discussed by Gospodinov (2010).

according to this criterion and yet provide a highly distorted impression of the economy's response to a technology shock.

The situation Galí had in mind was the following. The basic RBC model, as originally proposed, implies a high positive correlation between hours worked and productivity, whereas the data show a correlation near zero. This observation led researchers subsequently to augment the RBC model with non-technology shocks, which allow the model to replicate the near-zero correlation in the data. While the same unconditional correlation near zero can alternatively be explained by a suitably specified New Keynesian model, these two theoretical models differ in their implications for the responses of hours worked to a technology shock. Whereas the New Keynesian model implies that hours worked decline in response to a positive technology shock, the augmented RBC model implies a positive response of hours. Galí (1999) provided empirical evidence that hours decline persistently and significantly in response to a positive technology shock, which contradicts the augmented RBC model.

His evidence was based on structural VAR models in the tradition of Blanchard and Quah (1989), with one important difference. Given that RBC models do not generate data on the unemployment rate, Galí followed Cogley and Nason (1995) in replacing the unemployment rate in Blanchard and Quah's model with per capita hours worked, and expressed this variable in differences. Galí interpreted the impulse response estimates and related evidence from the same VAR model as implying that technology shocks do not play an important role in driving the business cycle. In view of the important role of technology shocks in explaining the business cycle in the recent literature, this evidence has been referred to as a potential paradigm shifter (see Francis and Ramey 2005).

Not surprisingly, proponents of the augmented RBC model have risen to the challenge of refuting Galí's conclusions. Some studies, including Francis and Ramey (2005), showed that the fall in hours could also be explained by RBC models that allow for even richer preference and technology structures. Others, including Christiano, Eichenbaum, and Vigfusson (2004), challenged the conclusion of Galí (1999) on the grounds that hours worked should be modeled as a stationary process in levels rather than in differences, in which case a positive shock to technology drives up per capita hours worked and real output, reversing Galí's conclusion. Finally, Chari, Kehoe, and McGrattan (2008) challenged the findings of Galí (1999) on methodological grounds. They made the case that the response estimator in Galí (1999) is prone to severe bias. Chari et al. suggested that VAR approximations to the DSGE model may generate significant declines in hours worked, even when applied to data from a DSGE model in which the response of hours is known to be positive. Their conclusion was that structural VAR models based on long-run restrictions simply should not be used to assess the empirical content of business cycle theory.

The latter conclusion has been challenged in turn by Christiano, Eichenbaum, and Vigfusson (2006b). First, Christiano et al. take issue with the realism of the RBC model specification employed by Chari et al. Second, they take issue with the specification of hours in log differences in the VAR model. They show that spurious rejections do not arise when hours are specified in levels. Third, they make the case that we ought to focus on the coverage accuracy of confidence intervals for the structural impulse responses generated from DSGE models rather than the accuracy of the impulse response point estimates. They provide evidence that users of VAR models are unlikely to spuriously reject the underlying DSGE model after allowing for sampling uncertainty in the impulse response estimates. They also provide examples in which users of structural VAR models may successfully discriminate between alternative DSGE models. Their study in turn drew a response from Kehoe (2006) who took issue with the econometric methodology for evaluating structural impulse response estimators employed in Christiano, Eichenbaum, and Vigfusson (2006b).

11.5.2 The Position of Chari et al. (2008)

Because this ongoing debate about the VAR methodology has received much attention, it is useful to outline each position in turn, starting with Chari, Kehoe, and McGrattan (2008). Notably, Kehoe (2006), building on his earlier work with Chari and McGrattan, makes the case that there is only one reasonable approach to evaluating DSGE models.

Suppose that we are interested in the question of how well a calibrated DSGE model fits the U.S. data, as measured by the impulse responses to structural shocks. Kehoe suggests that we first fit the structural VAR model of our choice to the actual U.S. data of sample length T. Then we generate repeated draws of length T from the calibrated DSGE model. For each of these draws, we fit the structural VAR model of interest, making sure to have transformed the data prior to the analysis exactly like the actual data. Finally, we compare the average of the implied structural impulse response estimates obtained conditional on the simulated data to the impulse response estimate of the structural VAR model fit to actual data. The distance between the average of the model-based structural impulse responses and the structural impulse responses based on the U.S. data becomes a measure of model fit.

Kehoe (2006) and Chari, Kehoe, and McGrattan (2008) refer to this approach as the Sims-Cogley-Nason approach with reference to Sims (1989) and, more importantly, Cogley and Nason (1995). The approach they describe actually is exactly how Kydland and Prescott would have evaluated a DSGE model, except that the latter would have focused on the cross-autocorrelation statistics rather than on structural impulse response statistics. We refer to this approach as the CKM approach.

Kehoe lists three reasons for preferring this approach to the conventional approach of comparing VAR estimates of the structural impulse responses to the population responses in the DSGE model. He observes, first, that his preferred approach does not suffer from small-sample bias. This claim is correct in that by conditioning on the actual data when evaluating the structural VAR model, one effectively takes a Bayesian approach. Conditional on the data, there is no small-sample bias.

Second, Kehoe suggests that this approach avoids lag-truncation bias by which he means that both the model fit to the U.S. data and the model fit to the simulated data are truncated in the same way. This does not mean that the truncation bias has been avoided so much as that this bias is the same whether working with simulated data or U.S. data.

Third, Kehoe emphasizes that this approach does not require the identifying assumptions used in the structural VAR to be correct, because we can compare the model-based statistics to the statistics based on U.S. data, even if those statistics have no economic meaning. This statement is technically correct, but it effectively amounts to reinterpreting the structural impulse responses as descriptive statistics only rather than as structural objects. Indeed, Kehoe hastens to add that in this case we might as well dispense with the identification of structural VAR models and revert back to using cross-autocorrelation statistics. Of course, this argument just amounts to ignoring the concern in Galí (1999) (as well as Cogley and Nason 1995) that evaluating the fit of a DSGE model based on unconditional second moments alone can be misleading.

According to Chari et al., the fit of a given DSGE model is measured by the extent to which the structural impulse responses obtained by fitting a VAR model to the U.S. data deviate from the average of the corresponding structural impulse response estimates obtained by fitting the same VAR model to each of many random draws of data generated from the calibrated DSGE model. Chari, Kehoe, and McGrattan (2008) are not interested in exploring the substantive implications of this approach, however. Rather, the objective of their analysis is to demonstrate that the fit of a DSGE model cannot be evaluated by comparing VAR model estimates of structural impulse responses based on U.S. data with the population impulse responses in DSGE models.

For this purpose, they design a thought experiment in which the data are generated from a DSGE model with known structural parameters. They then draw 1,000 data sequences of length T from this DSGE model and fit a structural VAR(p) model of the type employed by Galí (1999) to each sequence of simulated data. They compute the mean of the structural impulse responses implied by these structural VAR model estimates. Finally, they compare this average of the structural impulse responses to the corresponding population impulse response in the DSGE model. If the latter differs substantially from the average structural response estimates based on the simulated data, as measured by the sign or the magnitude of the response, the structural VAR model is considered inappropriate for assessing the validity of DSGE models. Chari et al. demonstrate that the VAR estimates of the structural impulse response may look quite different on average than the corresponding population impulse responses in their DGP.

Based on this exercise, Chari, Kehoe, and McGrattan (2008) conclude that, for common choices of p and T, a researcher comparing structural VAR estimates of the impulse responses obtained from U.S. data to the population responses in the DSGE model is likely to incorrectly conclude that the data are not generated from the DSGE model. Chari et al. attribute the apparent failure of the structural VAR approach in this thought experiment mainly to lag-order truncation bias.

11.5.3 The Position of Christiano et al. (2006)

How does the CKM approach differ from Christiano et al.'s approach? Christiano, Eichenbaum, and Vigfusson (2006b) take the position, consistent with the frequentist approach to econometrics, that the data are random, but that the DSGE model parameters are non-random, because, after all, we know their exact values when we specify the DSGE model. This fact allows one to compute the exact population value of the structural impulse responses of interest based on the VAR(∞) representation of the DSGE model.

The objective is to judge whether the population value of the statistic of interest in the DSGE model is consistent with the data. One addresses this question by fitting a structural VAR model to the U.S. data and approximating the sampling distribution of the statistic of interest by bootstrap or by asymptotic methods. After forming a confidence interval for the statistic of interest, one rejects the DSGE model at the chosen significance level γ, if the population value of the statistic of interest in the DSGE model lies outside of the $(1-\gamma)\,100\%$ confidence interval.

In the context of Christiano et al.'s study, the focus is on the response of hours worked to a technology innovation. In practice, they consider two summary statistics. The first summary statistic is the bias of the structural impulse response estimator. The bias is computed as the difference between the population impulse responses in the DSGE model and the average of the structural impulse response estimates obtained by fitting an approximating VAR model to each of 1,000 simulated data sets of sample length T generated by the DSGE model. The second summary statistic refers to the probability that pointwise bootstrap confidence intervals for the structural impulse responses, constructed from the finite-order VAR approximation, include the population value of the structural impulse response in repeated sampling (see Inoue and Kilian 2002b). The closer this probability is to the nominal coverage probability $1-\gamma$, the more accurate the interval. Christiano et al. evaluate the effective coverage probability by keeping track of how many of the bootstrap interval estimates – each obtained after fitting a VAR model to one of the 1,000 data sets of length T coming from the DSGE model – include the population value of the response.

Christiano, Eichenbaum, and Vigfusson (2006b) make the case that even if the VAR point estimates of the structural impulse responses are inaccurate in small samples, after accounting for sampling uncertainty, researchers would rarely reject a DSGE model incorrectly. Although the confidence bands may be wide, they are not so wide as to be consistent with any possible DSGE model. Christiano et al. show by example that, at least in some cases, the confidence bands are tight enough to allow the researcher to discriminate between competing DSGE models.

The approach employed by Christiano et al. is standard in frequentist econometrics. Thus, Kehoe's assertion that Christiano et al.'s approach lacks statistical foundations is without basis. If anything, the statistical foundations of his own preferred approach seem debatable, especially when compared to a proper Bayesian analysis. The "confidence intervals" reported in figure 4 and table 1 of Chari, Kehoe, and McGrattan (2008) are neither regions of highest posterior density nor confidence intervals in the frequentist sense, calling into question the authors' conclusion that the structural VAR model in question spuriously rejects the DSGE model that generated the data.

11.5.4 Understanding the Simulation Evidence

Kehoe (2006) raises several concerns about the reliability of Christiano et al.'s approach. Many of these concerns relate to the bias of the impulse response estimates obtained after imposing long-run restrictions. There are four distinct potential sources of bias. One concern is that all structural VAR models, but especially those based on long-run identifying restrictions, are subject to truncation bias in the lag order. This argument reflects a view exemplified by Cooley and Dwyer (1998, p. 76) that, as a theoretical matter, finite-order VAR models cannot accommodate the VARMA representations of DSGE models. This view is mistaken (see Chapter 2). Indeed, Chari, Kehoe, and McGrattan (2008) recognize that, under suitable conditions, a VARMA model can be represented as a VAR(∞) model. It is well known that under these conditions the finite-order VAR(p) model can be viewed as a semiparametric approximation, which becomes arbitrarily accurate asymptotically if the lag order increases with T at a suitable rate (see, e.g., Lewis and Reinsel 1985; Lütkepohl and Poskitt 1991; Inoue and Kilian 2002a). This means that the truncation bias refers to finite-sample approximation error rather than omitted variable bias. This distinction is not just a matter of semantics. It not only invalidates standard procedures for selecting the lag order and affects how one conducts inference about structural impulse responses, but it has been shown that the lag order required for reasonable sieve approximations tends to be larger than for conventional models with finite lag order (see, e.g., Inoue and Kilian 2002b). Thus, evidence that a VAR(4) approximation may yield impulse response estimates subject to large approximation error, for example, is hardly unexpected.

The question is whether this approximation error can be mitigated by the choice of a much larger lag order. Chari et al.'s premise is that for common choices of T, VAR models necessarily require short lag orders. This is not the case. There is considerable room for increasing the lag order in practice. Chari, Kehoe, and McGrattan (2008) provide an example where increasing the lag order from 4 to about 40 greatly improves the approximation to the population responses. While this example may be extreme, it remains to be seen what the benefit is more generally from increasing the lag order beyond the customary 4 or 8 quarterly lags.

Many researchers are reluctant to increase the VAR lag order because of the perception that adding more lags necessarily greatly inflates the variance of the impulse response estimator and renders the model uninformative. This perception is not correct. Not only are there counterexamples in empirical work showing that statistically significant impulse responses may be obtained even when using many autoregressive lags (see, e.g., Kilian 2009), but it has been shown that impulse responses from VAR models that are underparameterized tend to be highly misleading (see Kilian 2001).

There are several ways of approaching this problem. For example, Christiano, Eichenbaum, and Vigfusson (2006b) experiment with alternative nonparametric estimators of the long-run variance matrix in VAR models with long-run restrictions. An obvious question is whether this problem could be solved simply by increasing the approximating lag order substantially. The answer will be case specific. Intuitively, the lag order required for an accurate approximation depends on the magnitude of the MA roots in the VARMA representation of the DSGE model. One would not expect VAR models of given lag order to work equally well for all DSGE models for that reason. It therefore would be useful to know in particular what lag order it takes to achieve a good approximation for the DSGE models of interest to policymakers.

The second source of bias is small-sample bias of the type discussed in Pope (1990) and Kilian (1998c) (see Section 2.3.3). Chari, Kehoe, and McGrattan (2008) suggest that this small-sample bias is small. This finding is not unexpected if both real output and per capita hours worked are expressed in log-differences. Small-sample bias is expected to be higher when per capita hours are expressed in levels, as in Christiano et al.'s analysis. Indeed, Christiano, Eichenbaum, and Vigfusson (2006b, p. 26) present evidence of small-sample bias. In this case, it would make sense to consider using well-known methods of correcting for this bias, at least in constructing confidence bands. In contrast, Christiano et al. rely on Runkle's (1987) bootstrap method that ignores small-sample bias, and the number of bootstrap replications used in their study is unreasonably small, raising questions about the reliability of the results (see Chapter 12).

The third source of bias is asymptotic in nature and arises in models with per capita hours worked expressed in levels. This is the type of model used in Christiano, Eichenbaum, and Vigfusson (2006b). Hours worked in levels are quite persistent. As shown by Gospodinov (2010), such estimates are biased even asymptotically when hours follow a local-to-unity process, which is expected to be a good approximation when we are unsure of whether hours are $I(0)$ or $I(1)$ in the data. Gospodinov proposes an alternative impulse response estimator that is consistent and that is more precise than conventional estimators. The latter method does not appear to have been used in the debate at hand. Although this method was designed for finite-order VAR processes, extensions to VAR(∞) models should be feasible. It is also worth pointing out that there is some tension between the RBC model used by Christiano et al., which implies that per capita hours are stationary, and the actual data which may or may not be stationary.

Fourth, we already highlighted several reasons why VAR inference based on long-run restrictions may be unreliable in practice. One concern was the difficulty of interpreting the model estimates, when the model includes only one non-technology shock as in Blanchard and Quah (1989). This result is consistent with Christiano et al.'s finding that VAR models with four rather than two variables are better at recovering the DSGE model responses. Extending the model may not be enough, however, if there are non-technology shock processes in the model that themselves are highly persistent (see also Erceg, Guerrieri, and Gust 2005).

Fifth, none of the studies in question explains how its authors dealt with the problem that the impulse response functions implied by long-run restrictions are not unique without normalizing the sign of the response function of interest (see Section 10.5.4). This question is of obvious importance, given that this debate was triggered by questions about the sign of the response of hours worked to a technology shock. It also matters because the choice of normalization affects the coverage accuracy of impulse response confidence intervals and our ability to discriminate between alternative DSGE models.

To summarize, there are many potential sources of bias in the VAR impulse response estimates, some of which are specific to VAR models based on long-run identifying assumptions. The magnitude of the bias in the response estimates can also be sensitive to the specification of the underlying DSGE model, making it difficult to generalize from specific examples. For example, Christiano, Eichenbaum, and Vigfusson (2006b) demonstrate by simulation that recursively identified VAR models are quite accurate at recovering the population responses in the DSGE model if these short-run identifying restrictions are also satisfied in the underlying DGP. Such restrictions exist in a subset of DSGE models, which does not include the standard RBC model. As Kehoe (2006) correctly points out, this result does not mean that recursively identified VAR models can be used more broadly to discriminate between alternative theoretical models. In particular, one cannot reject theoretical models merely because their implications are not consistent with the estimates of recursively identified structural VAR models, unless these models also satisfy the identifying assumptions used in the VAR models. However, the results in Christiano, Eichenbaum, and Vigfusson (2006b) and Kascha and Mertens (2009) provide further evidence that truncation bias may not be the main problem, nor for that matter small-sample bias, but the identification method.

Chari, Kehoe, and McGrattan (2008) and Kehoe (2006) also question the ability of structural VAR models to provide accurate inference about the validity of DSGE models, but that assessment is not based on the methods of inference actually being used by practitioners, but rather on their own diagnostics. The evidence in Chari, Kehoe, and McGrattan (2008) is based on a VAR model in which hours worked are log-differenced. Christiano, Eichenbaum, and Vigfusson (2006b) make the case that − from a theoretical point of view − a VAR specification with hours worked in levels is the only specification that is consistent with the underlying DSGE model. Even Chari, Kehoe, and McGrattan (2008) concede that in the latter case, no spurious rejection of the DSGE model arises, even using their nonstandard approach to inference. That conclusion is consistent with simulation evidence on the coverage accuracy of structural impulse response confidence intervals in Christiano, Eichenbaum, and Vigfusson (2006b), which shows that the use of conventional pointwise confidence intervals tends to protect researchers from spurious inference, as long as enough of the variability in hours worked is driven by the structural shocks of interest. This evidence, in short, suggests that much of this debate actually is about the question of how to specify the log of hours worked in VAR models based on long-run restrictions.

Equally importantly, it can be shown by example that these confidence intervals at least in some cases help discriminate among competing DSGE models, undermining the claim that structural VAR models based on long-run restrictions are altogether useless for developing business cycle theory. Nevertheless, much more work remains to be done to answer the questions raised by this debate. Perhaps the most useful insight provided by this literature so far is that whether or not estimates from a structural VAR model will be informative depends on the structure of the economy, on the specification of the structural VAR model, on there being enough variability in the data to allow us to identify the model parameters of interest, and potentially even on unmodeled features of the U.S. data.

11.5.5 Summary

It is important to keep in mind that this discussion relates to the ability of VAR models with long-run restrictions to recover impulse responses from a make-belief world described by a stylized DSGE model. This focus is helpful in that it allows us to rule out a number of potential misspecification issues that we would confront with actual data. The results have to be viewed with caution, however, precisely because actual data need not conform to the assumptions of the DSGE model. A case in point is the analysis in Gospodinov, Maynard, and Pesavento (2011) of how features beyond the scope of standard DSGE business cycle models (such as the low-frequency correlation between hours and productivity) may affect the reliability of alternative specifications of VAR models based on long-run restrictions.

Keeping in mind these caveats, what have we learned from this debate about the reliability of structural VAR models based on long-run restrictions? First, the answer depends on the DSGE model of interest. On the one hand, even after making sure that the VAR model has as many structural shocks as the DSGE model and that the identifying restrictions are consistent with the DGP, there are situations in which confidence intervals based on structural VAR models imposing long-run restrictions may have serious coverage deficiencies. On the other hand, there are also many DSGE model specifications for which structural VAR models based on long-run restrictions are accurate with reasonably tight intervals, allowing the rejection of economically interesting alternatives (see also Erceg, Guerrieri, and Gust 2005).

Second, when working with long-run identifying assumptions, the VAR lag order is important for determining the accuracy of the approximation. The simulations in Chari, Kehoe, and McGrattan (2008, figure 3), which represent the least favorable DSGE model example discussed in this literature, show that with four autoregressive lags, the approximation to the true impulse response is poor, but with 40 lags the bias, as defined in that paper, appears reasonably small. Of course, these numbers are specific to their example. It remains to be seen to what extent the performance of structural models based on long-run restrictions may be improved by more careful attention to the lag order and which features of the DSGE model may facilitate the use of low-order VAR approximations (see, e.g., Pagan and Robinson 2016).

Third, there are additional concerns about the use of structural VAR models based on long-run restrictions in particular. Kascha and Mertens (2009), for example, attribute the biases of the impulse response estimators reported in the literature to the use of long-run identifying restrictions rather than to any truncation bias in the approximating VAR model. If there were no alternatives to the use of long-run restrictions in practice, one could argue that the precise reason for these biases would be of little consequence. Even granting that sometimes it can be difficult to justify conventional short-run exclusion restrictions in DSGE models, however, this is not the only alternativeidentification scheme. For example, Canova and Paustian (2011) report considerable success in recovering the qualitative characteristics of the underlying structural impulse responses when fitting structural VAR models with sign restrictions to data generated from DSGE models. This alternative class of models is discussed in Chapter 13. The current debate has largely ignored this and other alternative identification strategies.

Finally, as stressed by Kehoe (2006), one must not draw conclusions based on estimates of structural VAR models derived under a given set of identifying assumptions about theoretical models that do not satisfy these assumptions. This is only half of the message, however. The other half is that it is difficult to generalize from one or two DSGE models to the ability of structural VAR models to recover population responses in general. It may seem that if the structural VAR approach does not work in a simple case, it cannot be trusted more generally. This conclusion is unwarranted, because we know that the extent to which truncation bias in particular matters will differ from one DSGE model to the next, depending on the magnitude of the MA roots. Hence, if one wants to make a case against the use of structural VAR models, it is not enough to show that for some stylized and grossly unrealistic DSGE model the structural VAR approach can be inaccurate. Rather, one needs to show that structural VAR models are unable to capture the underlying population responses of DSGE models that deserve our attention because they fit the data reasonably well and because they are economically plausible.