9

Structural VARs and VECMs

In Chapters 2 and 6, we have seen that, on the one hand, impulse responses are an important tool to uncover the relations between the variables in a VAR or VECM and, on the other hand, there are some obstacles in their interpretation. In particular, impulse responses are generally not unique and it is often not clear which set of impulse responses actually reflects the ongoings in a given system. Because the different sets of impulses can be computed from the same underlying VAR or VECM, it is clear that nonsample information has to be used to decide on the proper set for a particular given model. In econometric terminology, VARs are reduced form models and structural restrictions are required to identify the relevant innovations and impulse responses. In this chapter, different possible restrictions that have been proposed in the literature will be considered. The resulting models are known as *structural VAR* (SVAR) models (see, e.g., Sims (1981, 1986), Bernanke (1986), Shapiro & Watson (1988), Blanchard & Quah (1989)) or *structural VECMs* (SVECMs) (e.g., King, Plosser, Stock & Watson (1991), Jacobson, Vredin & Warne (1997), Gonzalo & Ng (2001), Breitung, Brüggemann & Lütkepohl (2004)).

In the next section, structural restrictions will be discussed for stationary processes. Some of them will also be relevant for VARs with integrated variables. Such variables are explicitly taken into account in VECMs for which structural restrictions will be discussed in Section 9.2. It will be seen that VECMs offer additional possibilities for structural restrictions. The general modelling strategy for both SVARs and SVECMs is to specify and estimate a reduced form model first and then focus on the structural parameters and the resulting structural impulse responses. Estimation of structural VARs and VECMs will be discussed in Section 9.3 and impulse response analysis and forecast error variance decomposition based on such models are considered in Section 9.4. Some extensions of the setup used in this chapter are pointed out in Section 9.5.

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*

9.1 Structural Vector Autoregressions

Our point of departure is a K-dimensional stationary, stable VAR(p) process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{9.1.1}$$

where, as usual, y_t is a $(K \times 1)$ vector of observable time series variables, the A_j's ($j = 1, \ldots, p$) are $(K \times K)$ coefficient matrices and u_t is K-dimensional white noise with $u_t \sim (0, \Sigma_u)$. Deterministic terms have been excluded for simplicity. In other words, we just consider the stochastic part of a data generation process because it is the part of interest from the point of view of structural modelling and impulse response analysis. From Chapter 2, it is known that the process (9.1.1) has a Wold MA representation

$$y_t = u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots , \tag{9.1.2}$$

where

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j} A_j, \quad s = 1, 2, \ldots, \tag{9.1.3}$$

with $\Phi_0 = I_K$.

In Chapter 2, we have also seen that the elements of the Φ_j matrices are the forecast error impulse responses. They may not reflect the relations between the variables properly because the components of u_t may be instantaneously correlated, that is, Σ_u may not be a diagonal matrix. Thus, isolated shocks in the components of u_t may not be likely in practice. From Chapter 2, we also know that there are different ways to orthogonalize the impulses. One possibility is based on a Choleski decomposition of the white noise covariance matrix, $\Sigma_u = PP'$, where P is a lower-triangular matrix with positive elements on the main diagonal. Again such an approach is arbitrary and therefore unsatisfactory, unless there are special reasons for a recursive structure. We will now discuss different ways to use nonsample information in specifying unique innovations and, hence, unique impulse responses. The relevant models will be referred to as A-model, B-model and AB-model. The latter label was also used by Amisano & Giannini (1997). The models will be considered in turn in the following.

9.1.1 The A-Model

A conventional approach to finding a model with instantaneously uncorrelated residuals is to model the instantaneous relations between the observable variables directly. That may be done by considering a structural form model,

$$\mathsf{A} y_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t, \tag{9.1.4}$$

where $A_j^* := \mathsf{A}A_j$ $(j = 1,\dots,p)$ and $\varepsilon_t := \mathsf{A}u_t \sim (0, \Sigma_\varepsilon = \mathsf{A}\Sigma_u\mathsf{A}')$. Thus, for a proper choice of A, ε_t will have a diagonal covariance matrix. An MA representation based on the ε_t is given by

$$y_t = \Theta_0\varepsilon_t + \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \cdots, \tag{9.1.5}$$

where $\Theta_j = \Phi_j\mathsf{A}^{-1}$ $(j = 0, 1, 2, \dots)$. The elements of the Θ_j matrices represent the responses to ε_t shocks. If an identified structural form (9.1.4) can be found, the corresponding impulse responses will be unique.

It may be worth reflecting a little on the restrictions required for a unique matrix A of instantaneous effects. From the relation

$$\Sigma_\varepsilon = \mathsf{A}\Sigma_u\mathsf{A}'$$

and the assumption of a diagonal Σ_ε matrix, we get $K(K-1)/2$ independent equations, that is, all $K(K-1)/2$ off-diagonal elements of $\mathsf{A}\Sigma_u\mathsf{A}'$ are equal to zero. To solve uniquely for all K^2 elements of A, we need a set of K^2 equations, however. In other words, we need $K(K+1)/2$ additional equations. They may be set up in the form of restrictions for the elements of A. Clearly, we may want to choose the diagonal elements of A to be unity. This normalization enables us to write the k-th equation of (9.1.4) with y_{kt} as the left-hand variable. In addition to this normalization, we still need another $K(K-1)/2$ restrictions. Such restrictions have to come from nonsample sources. For example, if a Wold causal ordering is possible, where y_{1t} may have an instantaneous impact on all the other variables, y_{2t} may have an instantaneous impact on all other variables except y_{1t}, and so on (see Section 2.3.2), then

$$\mathsf{A} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \mathsf{a}_{21} & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ \mathsf{a}_{K1} & \mathsf{a}_{K2} & \dots & 1 \end{bmatrix}$$

is a lower-triangular matrix. Thus, we have just enough restrictions ($K(K-1)/2$ zeros above the main diagonal) so that the innovations and the associated impulse responses are just-identified. The zeros can also appear in a different arrangement as off-diagonal elements of A. There can also be more than $K(K-1)/2$ restrictions, of course. In SVAR modelling it is common, however, that just-identified models are considered. In other words, only as few restrictions are imposed as are necessary for obtaining unique impulse responses. If at some stage of the analysis it turns out that further restrictions are compatible with the data, it is also possible to impose them, of course.

In the presently considered model, the identifying restrictions are imposed on the matrix A such that $\varepsilon_t = \mathsf{A}u_t$ has a diagonal covariance matrix. This model will be called the A-model in the following. Given the way we have introduced the associated restrictions, it is plausible to assume that A has a unit main diagonal. In that case $K(K-1)/2$ restrictions are required for the off-diagonal elements of A to ensure just-identified shocks ε_t and, hence, just-identified impulse responses. If the restrictions are such that A is lower-triangular, the same is true for A^{-1}. Thus, the resulting Θ_j impulse responses are qualitatively the same as the orthogonalized impulse responses based on a Choleski decomposition of Σ_u which were considered in Chapter 2. The only difference is that, for the latter case, the w_t impulses have unit variances which may not be the case for the presently considered ε_t impulses.

Regarding the restrictions for A, it should be understood that they cannot be arbitrary restrictions. Writing them in the form $C_{\mathsf{A}}\mathrm{vec}(\mathsf{A}) = c_{\mathsf{A}}$, where C_{A} is a $(\frac{1}{2}K(K+1) \times K^2)$ selection matrix and c_{A} is a suitable $(\frac{1}{2}K(K+1) \times 1)$ fixed vector, the restrictions have to be such that the system of equations

$$\mathsf{A}^{-1}\Sigma_\varepsilon \mathsf{A}'^{-1} = \Sigma_u \quad \text{and} \quad C_{\mathsf{A}}\mathrm{vec}(\mathsf{A}) = c_{\mathsf{A}} \tag{9.1.6}$$

has a unique solution, at least locally. Clearly, this system is nonlinear in A. Therefore, we can only hope for local uniqueness or identification in general. The following proposition gives a necessary and sufficient condition for (9.1.6) to have a locally unique solution and, thus, for local identification of the structural parameters.

Proposition 9.1 (*Identification of the* A*-Model*)
Let Σ_ε be a $(K \times K)$ positive definite diagonal matrix and let A be a $(K \times K)$ nonsingular matrix. Then, for a given symmetric, positive definite $(K \times K)$ matrix Σ_u, an $(N \times K^2)$ matrix C_{A} and a fixed $(N \times 1)$ vector c_{A}, the system of equations in (9.1.6) has a locally unique solution for A and the diagonal elements of Σ_ε if and only if

$$\mathrm{rk}\begin{bmatrix} -2\mathbf{D}_K^{+}(\Sigma_u \otimes \mathsf{A}^{-1}) & \mathbf{D}_K^{+}(\mathsf{A}^{-1} \otimes \mathsf{A}^{-1})\mathbf{D}_K \\ C_{\mathsf{A}} & 0 \\ 0 & C_\sigma \end{bmatrix} = K^2 + \tfrac{1}{2}K(K+1).$$

Here $\mathbf{D}_K$ is a $(K^2 \times \frac{1}{2}K(K+1))$ duplication matrix, $\mathbf{D}_K^{+} := (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$, and C_σ is a $(\frac{1}{2}K(K-1) \times \frac{1}{2}K(K+1))$ selection matrix which selects the elements of $\mathrm{vech}(\Sigma_\varepsilon)$ below the main diagonal. ■

Proof: For an n-dimensional function $\varphi(x)$ of the m-dimensional vector x, the system of equations $\varphi(x) = 0$ can be solved locally uniquely for x in a neighborhood of a given vector x_0 if and only if $\mathrm{rk}(\partial\varphi/\partial x'|_{x=x_0}) = m$ (see, e.g., Rothenberg (1971, Theorem 6)). Hence, considering the function

$$\begin{bmatrix} \mathrm{vec}(\mathsf{A}) \\ \mathrm{vech}(\Sigma_\varepsilon) \end{bmatrix} \mapsto \begin{bmatrix} \mathrm{vec}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1} - \Sigma_u) \\ C_{\mathsf{A}}\mathrm{vec}(\mathsf{A}) - c_{\mathsf{A}} \\ C_\sigma \mathrm{vech}(\Sigma_\varepsilon) \end{bmatrix},$$

a locally unique solution for A and $\mathrm{vech}(\Sigma_\varepsilon)$ exists for a given Σ_u if and only if

$$\mathrm{rk}\begin{bmatrix} \dfrac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} & \dfrac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vech}(\Sigma_\varepsilon)'} \\ C_{\mathsf{A}} & 0 \\ 0 & C_\sigma \end{bmatrix} = K^2 + \tfrac{1}{2}K(K+1).$$

Taking into account that the off-diagonal elements of Σ_ε are uniquely determined by $C_\sigma\mathrm{vech}(\Sigma_\varepsilon) = 0$, a locally unique solution for A and the diagonal elements of Σ_ε exists if and only if the rank condition is satisfied. Thus, the proposition follows by using the rules for matrix and vector differentiation from Appendix A.13 and noting that

$$\begin{aligned}\frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vech}(\Sigma_\varepsilon)'} &= \mathbf{D}_K^+\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vech}(\Sigma_\varepsilon)'} \\ &= \mathbf{D}_K^+(\mathsf{A}^{-1}\otimes\mathsf{A}^{-1})\frac{\partial\,\mathrm{vec}(\Sigma_\varepsilon)}{\partial\,\mathrm{vech}(\Sigma_\varepsilon)'} \\ &= \mathbf{D}_K^+(\mathsf{A}^{-1}\otimes\mathsf{A}^{-1})\mathbf{D}_K\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} &= \mathbf{D}_K^+\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\Sigma_\varepsilon\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1})'}\,\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\ &= \mathbf{D}_K^+\left[(\mathsf{A}^{-1}\Sigma_\varepsilon\otimes I_K)\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1})'}\right. \\ &\qquad \left.+\,(I_K\otimes\mathsf{A}^{-1}\Sigma_\varepsilon)\frac{\partial\,\mathrm{vec}(\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1})'}\right]\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\ &= -\mathbf{D}_K^+(I_{K^2}+\mathbf{K}_{KK})(\mathsf{A}^{-1}\Sigma_\varepsilon\otimes I_K)(\mathsf{A}'^{-1}\otimes\mathsf{A}^{-1}) \\ &= -\mathbf{D}_K^+(I_{K^2}+\mathbf{K}_{KK})(\Sigma_u\otimes\mathsf{A}^{-1}) \\ &= -2\mathbf{D}_K^+(\Sigma_u\otimes\mathsf{A}^{-1}),\end{aligned}$$

where $\mathbf{K}_{KK}$ denotes a $(K^2\times K^2)$ commutation matrix and the last equality sign holds because $\mathbf{D}_K^+\mathbf{K}_{KK} = \mathbf{D}_K^+$ (see Appendix A.12.2). ■

Although this proposition provides a condition for local identification of the A-model only, a globally unique solution is obtained if the diagonal elements of A are restricted to 1. A discussion of the nonuniqueness problem resulting from sign changes of some elements will be deferred to Section 9.1.2.

For practical purposes, it is problematic that the identification condition in Proposition 9.1 involves unknown parameters. Therefore, strictly speaking, it can only be checked when the true parameters are known. In practice, the unknown quantities may be replaced by estimates and the condition may be checked using the estimated matrix because it can be shown that the rank of the relevant matrix is either smaller than $K^2 + \frac{1}{2}K(K+1)$ everywhere in the parameter space or the rank condition is satisfied almost everywhere. In the latter case, it can fail only on a set of Lebesgue measure zero. Thus, if a randomly drawn vector from the parameter space is considered, it should satisfy the rank condition with probability one, if the model is locally identified. In any case, C_{A} must have at least $K(K+1)/2$ rows to ensure identification. In other words, having $K(K+1)/2$ restrictions is a necessary condition for identification.

Although we have stated the restrictions for the A matrix in the form $C_{\mathsf{A}}\mathrm{vec}(\mathsf{A}) = c_{\mathsf{A}}$ in the foregoing, we note that they can be written alternatively in the form

$$\mathrm{vec}(\mathsf{A}) = R_{\mathsf{A}}\gamma_{\mathsf{A}} + r_{\mathsf{A}},$$

where R_{A} and r_{A} are a suitable fixed matrix and a suitable vector, respectively, and γ_{A} is the vector of unrestricted parameters (see Chapter 5, Section 5.2.1).

9.1.2 The B-Model

Generally, in impulse response analysis the emphasis has shifted from specifying the relations between the observable variables directly to interpreting the unexpected part of their changes or the shocks. Therefore, it is not uncommon to identify the structural innovations ε_t directly from the forecast errors or reduced form residuals u_t. One way to do so is to think of the forecast errors as linear functions of the structural innovations. In that case, we have the relations $u_t = \mathsf{B}\varepsilon_t$. Hence, $\Sigma_u = \mathsf{B}\Sigma_\varepsilon\mathsf{B}'$. Normalizing the variances of the structural innovations to one, i.e., assuming $\varepsilon_t \sim (0, I_K)$, gives

$$\Sigma_u = \mathsf{B}\mathsf{B}'. \tag{9.1.7}$$

Due to the symmetry of the covariance matrix, these relations specify only $K(K+1)/2$ different equations and we need again $K(K-1)/2$ further relations to identify all K^2 elements of B. As in the previous A-model case, choosing B to be lower-triangular, for example, provides sufficiently many restrictions. Hence, choosing B by a Choleski decomposition solves the identification or uniqueness problem, as we have also seen in Chapter 2, Section 2.3.2. Now it is assumed, however, that this recursive structure is chosen only if it has some theoretical justification so that the ε_t's can be regarded as structural innovations. This property makes them potentially different from the w_t innovations in Chapter 2 which were obtained by a mechanical application of the Choleski decomposition. In principle, there could be other zero restrictions for B in the present context. The triangular form is just an example. In practice, it is perhaps the most important case (e.g., Eichenbaum & Evans (1995), Christiano, Eichenbaum & Evans (1996)).

The present model with

$$u_t = \mathsf{B}\varepsilon_t$$

and $\varepsilon_t \sim (0, I_K)$ will be called B-model in the following and it is worth remembering that at least $K(K-1)/2$ restrictions have to be imposed to identify B. If there are just zero restrictions they can be written in the form

$$C_{\mathsf{B}}\text{vec}(\mathsf{B}) = 0, \tag{9.1.8}$$

where C_{B} is an $(N \times K^2)$ selection matrix. A necessary and sufficient rank condition for local identification of the model is given in the next proposition.

Proposition 9.2 (*Local Identification of the* B*-Model*)
Let B be a nonsingular $(K \times K)$ matrix. Then, for a given symmetric, positive definite $(K \times K)$ matrix Σ_u and an $(N \times K^2)$ matrix C_{B}, the system of equations in (9.1.7)/(9.1.8) has a locally unique solution if and only if

$$\text{rk}\begin{bmatrix} 2\mathbf{D}_K^+(\mathsf{B} \otimes I_K) \\ C_{\mathsf{B}} \end{bmatrix} = K^2.$$

■

Proof: Using the same kind of reasoning as in the proof of Proposition 9.1, the result of Proposition 9.2 follows by noting that

$$\frac{\partial\, \text{vech}(\mathsf{B}\mathsf{B}')}{\partial\, \text{vec}(\mathsf{B})'} = \mathbf{D}_K^+(I_{K^2} + \mathbf{K}_{KK})(\mathsf{B} \otimes I_K) = 2\mathbf{D}_K^+(\mathsf{B} \otimes I_K).$$

■

A necessary condition for the $((\frac{1}{2}K(K+1) + N) \times K^2)$ matrix

$$\begin{bmatrix} 2\mathbf{D}_K^+(\mathsf{B} \otimes I_K) \\ C_{\mathsf{B}} \end{bmatrix}$$

to have rank K^2 is that $N = \frac{1}{2}K(K-1)$. In other words, we need $\frac{1}{2}K(K-1)$ restrictions for identification, as mentioned earlier.

It is easy to see that the solution of the system (9.1.7)/(9.1.8) will not be globally unique because for any matrix B satisfying the equations, $-\mathsf{B}$ will also be a solution. This result is due to the fact that B enters the equations (9.1.7) in "squared" form. In fact, for any solution B, the matrix $\mathsf{B}\Lambda$ will also be a solution for any diagonal matrix Λ which has only 1 and -1 elements on the main diagonal. Obviously, if B is such that (9.1.7) and (9.1.8) hold, $\Sigma_u = \mathsf{B}\Lambda\Lambda'\mathsf{B}'$ also holds because $\Lambda\Lambda' = I_K$. Moreover,

$$C_{\mathsf{B}}\text{vec}(\mathsf{B}\Lambda) = C_{\mathsf{B}}(\Lambda \otimes I_K)\text{vec}(\mathsf{B}) = 0,$$

because for each element $\mathsf{b}_{ij} = 0$ we have $-\mathsf{b}_{ij} = 0$. Thus, each column of B can be replaced by a column with opposite sign. Hence, the restrictions in (9.1.8) identify B only locally in general. Uniqueness can potentially be obtained by fixing the signs of the diagonal elements, however. The signs of the diagonal elements of B determine the signs of shocks. Thus, if we want to study the effect of a positive shock to a particular variable while the corresponding diagonal element of B is negative, we can just reverse the signs of all elements in the relevant column of B or, in other words, we can just reverse the signs of all instantaneous responses to the corresponding shock to find the desired result.

For later purposes, it is also worth noting that the restrictions can be expressed in the alternative form

$$\mathrm{vec}(\mathsf{B}) = R_{\mathsf{B}}\gamma_{\mathsf{B}}, \tag{9.1.9}$$

where γ_{B} contains all the unrestricted coefficients of B and R_{B} is a fixed matrix of zeros and ones (see Section 5.2.1).

9.1.3 The AB-Model

It is also possible to consider both types of restrictions of the previous subsections simultaneously. That is, we may consider the so-called AB-model,

$$\mathsf{A}u_t = \mathsf{B}\varepsilon_t, \qquad \varepsilon_t \sim (0, I_K). \tag{9.1.10}$$

In this case, a simultaneous equations system is formulated for the errors of the reduced form model rather than the observable variables directly. Thereby the model accounts for the shift from specifying direct relations for the observable variables to formulating relations for the innovations. Applications of this methodology can, for instance, be found in Galí (1992) and Pagan (1995) (see also Breitung et al. (2004) for further discussion and an illustration).

In this model, we get from (9.1.10), $u_t = \mathsf{A}^{-1}\mathsf{B}\varepsilon_t$ and, hence, $\varSigma_u = \mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}^{-1\prime}$. Thus, we have $K(K+1)/2$ equations

$$\mathrm{vech}(\varSigma_u) = \mathrm{vech}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}^{-1\prime}), \tag{9.1.11}$$

whereas the two matrices A and B have K^2 elements each. Thus, we need additionally $2K^2 - \frac{1}{2}K(K+1)$ restrictions to identify all $2K^2$ elements of A and B at least locally. Even if the diagonal elements of A are set to one, $2K^2 - K - \frac{1}{2}K(K+1)$ further restrictions are needed for identification. Therefore, it is perhaps not surprising that most applications consider special cases with $\mathsf{A} = I_K$ (B-model) or $\mathsf{B} = I_K$ (A-model). Still, the general model is a useful framework for SVAR analysis. The restrictions are typically normalization or zero restrictions which can be written in the form of linear equations,

$$\mathrm{vec}(\mathsf{A}) = R_{\mathsf{A}}\gamma_{\mathsf{A}} + r_{\mathsf{A}} \qquad \text{and} \qquad \mathrm{vec}(\mathsf{B}) = R_{\mathsf{B}}\gamma_{\mathsf{B}} + r_{\mathsf{B}}, \tag{9.1.12}$$

where R_{A} and R_{B} are suitable fixed matrices of zeros and ones, γ_{A} and γ_{B} are vectors of free parameters and r_{A} and r_{B} are vectors of fixed parameters which allow, for instance, to normalize the diagonal elements of A. Although r_{B} is typically zero, as in (9.1.9), we present the restrictions for B here with a general r_{B} vector because this additional term will not complicate the analysis.

Multiplying the two sets of equations in (9.1.12) by orthogonal complements of R_{A} and R_{B}, $R_{\mathsf{A}\perp}$ and $R_{\mathsf{B}\perp}$, respectively, it is easy to see that they can be written alternatively in the form

$$C_{\mathsf{A}}\mathrm{vec}(\mathsf{A}) = c_{\mathsf{A}} \qquad \text{and} \qquad C_{\mathsf{B}}\mathrm{vec}(\mathsf{B}) = c_{\mathsf{B}}, \tag{9.1.13}$$

where $C_{\mathsf{A}} = R_{\mathsf{A}\perp}$, $C_{\mathsf{B}} = R_{\mathsf{B}\perp}$, $c_{\mathsf{A}} = R_{\mathsf{A}\perp}r_{\mathsf{A}}$ and $c_{\mathsf{B}} = R_{\mathsf{B}\perp}r_{\mathsf{B}}$ (see Appendix A.8.2 for the definition of an orthogonal complement of a matrix). The matrices C_{A} and C_{B} may be thought of as appropriate selection matrices. Again, in general, the restrictions will ensure only local uniqueness of A and B due to the nonlinear nature of the full set of equations from which to solve for the two matrices. The following proposition states a rank condition for local identification.

Proposition 9.3 (*Local Identification of the* AB-*Model*)
Let A and B be nonsingular $(K \times K)$ matrices. Then, for a given symmetric, positive definite $(K \times K)$ matrix Σ_u, the system of equations in (9.1.11)/(9.1.13) has a locally unique solution if and only if

$$\mathrm{rk}\begin{bmatrix} -2\mathbf{D}_K^+(\Sigma_u \otimes \mathsf{A}^{-1}) & 2\mathbf{D}_K^+(\mathsf{A}^{-1}\mathsf{B} \otimes \mathsf{A}^{-1}) \\ C_{\mathsf{A}} & 0 \\ 0 & C_{\mathsf{B}} \end{bmatrix} = 2K^2. \tag{9.1.14}$$

■

Proof: Again, we can use the same reasoning as in the proof of Proposition 9.1. The result of Proposition 9.3 is then obtained by noting that

$$\begin{aligned}
\frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} &= \frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})'}\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\
&= \mathbf{D}_K^+\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})'}\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\
&= \mathbf{D}_K^+\left[(\mathsf{A}^{-1}\mathsf{B} \otimes I_K)\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})'}\right. \\
&\qquad \left. +(I_K \otimes \mathsf{A}^{-1}\mathsf{B})\frac{\partial\,\mathrm{vec}(\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})'}\right]\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\
&= \mathbf{D}_K^+\left[(\mathsf{A}^{-1}\mathsf{B} \otimes I_K) + (I_K \otimes \mathsf{A}^{-1}\mathsf{B})\mathbf{K}_{KK}\right] \\
&\qquad \times(\mathsf{B}' \otimes I_K)\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1})}{\partial\,\mathrm{vec}(\mathsf{A})'} \\
&= -\mathbf{D}_K^+(I_{K^2} + \mathbf{K}_{KK})(\Sigma_u \otimes \mathsf{A}^{-1}) \\
&= -2\mathbf{D}_K^+(\Sigma_u \otimes \mathsf{A}^{-1})
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{B})'} &= \frac{\partial\,\mathrm{vech}(\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1})}{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})'}\frac{\partial\,\mathrm{vec}(\mathsf{A}^{-1}\mathsf{B})}{\partial\,\mathrm{vec}(\mathsf{B})'} \\
&= \mathbf{D}_K^+(I_{K^2} + \mathbf{K}_{KK})(\mathsf{A}^{-1}\mathsf{B} \otimes \mathsf{A}^{-1}) \\
&= 2\mathbf{D}_K^+(\mathsf{A}^{-1}\mathsf{B} \otimes \mathsf{A}^{-1}),
\end{aligned}$$

because $\mathbf{D}_K^+\mathbf{K}_{KK} = \mathbf{D}_K^+$ (see Appendix A.12.2). ■

To illustrate the AB-model, we follow Breitung et al. (2004) and use a small macro system from Pagan (1995) for output q_t, an interest rate i_t, and real money m_t. The residuals of the reduced form VAR model will be denoted by $u_t = (u_t^q, u_t^i, u_t^m)'$. Pagan (1995) uses Keynesian arguments to specify the following relations between the reduced form residuals and the structural innovations:

$$\begin{aligned} u_t^q &= -\mathsf{a}_{12} u_t^i + \mathsf{b}_{11}\varepsilon_t^{IS} && \text{(IS curve)},\\ u_t^i &= -\mathsf{a}_{21} u_t^q - \mathsf{a}_{23} u_t^m + \mathsf{b}_{22}\varepsilon_t^{LM} && \text{(inverse LM curve)},\\ u_t^m &= \mathsf{b}_{33}\varepsilon_t^m && \text{(money supply rule)}. \end{aligned}$$

Here $\varepsilon_t = (\varepsilon_t^{IS}, \varepsilon_t^{LM}, \varepsilon_t^m)'$ is the vector of structural innovations with $\varepsilon_t \sim (0, I_K)$ (see Breitung et al. (2004) for further discussion of this example system).

For our purposes the three equations can be written in AB-model form as

$$\begin{bmatrix} 1 & \mathsf{a}_{12} & 0 \\ \mathsf{a}_{21} & 1 & \mathsf{a}_{23} \\ 0 & 0 & 1 \end{bmatrix} u_t = \begin{bmatrix} \mathsf{b}_{11} & 0 & 0 \\ 0 & \mathsf{b}_{22} & 0 \\ 0 & 0 & \mathsf{b}_{33} \end{bmatrix} \varepsilon_t.$$

Thus, we have the following set of restrictions:

$$\mathrm{vec}(\mathsf{A}) = \begin{bmatrix} 1 \\ \mathsf{a}_{21} \\ 0 \\ \mathsf{a}_{12} \\ 1 \\ 0 \\ 0 \\ \mathsf{a}_{23} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathsf{a}_{21} \\ \mathsf{a}_{12} \\ \mathsf{a}_{23} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$\mathrm{vec}(\mathsf{B}) = \begin{bmatrix} \mathsf{b}_{11} \\ 0 \\ 0 \\ 0 \\ \mathsf{b}_{22} \\ 0 \\ 0 \\ 0 \\ \mathsf{b}_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathsf{b}_{11} \\ \mathsf{b}_{22} \\ \mathsf{b}_{33} \end{bmatrix}.$$

Because $K = 3$, we need $2K^2 - \frac{1}{2}K(K+1) = 12$ restrictions on A and B for identification in this example model. There are 3 zeros and 3 ones in A. Thus, we have 6 restrictions on this matrix. In addition, there are 6 zero restrictions for B.

Writing the restrictions in the form (9.1.13), we get

$$\begin{bmatrix} 1&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&1&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} \mathsf{a}_{11}\\ \mathsf{a}_{21}\\ \mathsf{a}_{31}\\ \mathsf{a}_{12}\\ \mathsf{a}_{22}\\ \mathsf{a}_{32}\\ \mathsf{a}_{13}\\ \mathsf{a}_{23}\\ \mathsf{a}_{33} \end{bmatrix} = \begin{bmatrix} 1\\0\\1\\0\\0\\1 \end{bmatrix}$$

and

$$\begin{bmatrix} 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&1&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0 \end{bmatrix} \begin{bmatrix} \mathsf{b}_{11}\\ \mathsf{b}_{21}\\ \mathsf{b}_{31}\\ \mathsf{b}_{12}\\ \mathsf{b}_{22}\\ \mathsf{b}_{32}\\ \mathsf{b}_{13}\\ \mathsf{b}_{23}\\ \mathsf{b}_{33} \end{bmatrix} = \begin{bmatrix} 0\\0\\0\\0\\0\\0 \end{bmatrix}.$$

Thus, the necessary condition for local identification is satisfied. The necessary and sufficient condition from Proposition 9.3 can be checked by selecting randomly drawn matrices A and B from the restricted parameter space and determining the rank of the corresponding matrix in (9.1.14).

9.1.4 Long-Run Restrictions à la Blanchard-Quah

Clearly, it is not always easy to find suitable and generally acceptable restrictions for the matrices A and B. Imposing the restrictions directly on these matrices is in fact not necessary to identify the structural innovations and impulse responses. Another type of restrictions was discussed by Blanchard & Quah (1989). They considered the accumulated effects of shocks to the system. In terms of the structural impulse responses in (9.1.5) they focussed on the *total impact matrix*,

$$\Xi_\infty = \sum_{i=0}^{\infty} \Theta_i = (I_K - A_1 - \cdots - A_p)^{-1}\mathsf{A}^{-1}\mathsf{B}, \tag{9.1.15}$$

and they identified the structural innovations by placing zero restrictions on this matrix. In other words, they assumed that some shocks do not have any total long-run effects. In particular, they considered a bivariate system consisting of output growth q_t and an unemployment rate ur_t (i.e., $y_t =$

$(q_t, ur_t)')$ and they assumed that the structural innovations represent supply and demand shocks. Moreover, they assumed that the demand shocks have only transitory effects on q_t and that the accumulated long-run effect of such shocks on q_t is zero. Placing the supply shocks first and the demand shocks last in the vectors of structural innovations $\varepsilon_t = (\varepsilon_t^s, \varepsilon_t^d)'$, the (1,2)-element of Ξ_∞ is restricted to be zero. In other words, we restrict the upper right-hand corner element of

$$\Xi_\infty = (I_K - A_1 - \cdots - A_p)^{-1}\mathsf{A}^{-1}\mathsf{B}$$

to zero. Given the VAR parameters, this set of equations clearly specifies a restriction for $\mathsf{A}^{-1}\mathsf{B}$. Thereby we have enough restrictions for identification of a bivariate system if we set $\mathsf{A} = I_K$, because, for $K = 2$, we have $K(K-1)/2 = 1$. Notice that $\mathsf{A} = I_K$ may be chosen because the idea is to identify the structural shocks from the reduced form residuals only and no restrictions are placed on the instantaneous effects of the observable variables directly. Thus, we have a B-model with restriction

$$(0,0,1,0)\mathrm{vec}[(I_K - A_1 - \cdots - A_p)^{-1}\mathsf{B}]$$

$$= (0,0,1,0)[I_2 \otimes (I_K - A_1 - \cdots - A_p)^{-1}]\mathrm{vec}(\mathsf{B}) = 0.$$

In summary, the AB-model offers a useful general framework for placing identifying restrictions for the structural innovations and impulse responses on a VAR process. The restrictions can be simple normalization and exclusion (zero) restrictions and may also be more general nonlinear restrictions. Clearly, before we can actually use this framework in practice, it will be necessary to estimate the reduced form and structural parameters. Estimation of the former parameters has been discussed in some detail in previous chapters. Thus, it remains to consider estimation of the A, B matrices. We will do so in Section 9.3. Before turning to inference procedures, we will consider structural restrictions for VECMs in the following section.

9.2 Structural Vector Error Correction Models

If all or some of the variables of interest are integrated, the previously discussed AB-model can still be used together with the levels VAR form of the data generation process. In most of the analysis of Section 9.1, the stationarity of the process was not used. Only in the treatment of the Blanchard-Quah restrictions, stability of the VAR operator is required because otherwise the matrix of total accumulated long-run effects does not exist. This result follows from the fact that the matrix $(I_K - A_1 - \cdots - A_p)$ is singular for cointegrated processes, as we have seen in Chapter 6. In other cases, we may use the AB-model even for integrated variables. In fact, we can even specify and fit a reduced form VECM, convert that model to the levels VAR form and then use it as a basis for an AB-analysis, as discussed in the previous section. There are, however, advantages in utilizing the cointegration properties of the variables. They provide restrictions which can be taken into account beneficially in identifying the structural shocks. Therefore, it is useful to treat SVECMs separately.

As in the previous chapters, we assume that all variables are at most $I(1)$ and that the data generation process can be represented as a VECM with cointegration rank r of the form

$$\Delta y_t = \alpha\beta' y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t, \tag{9.2.1}$$

where all symbols have their usual meanings. In other words, y_t is a K-dimensional vector of observable variables, α is a $(K \times r)$ matrix of loading coefficients, β is the $(K \times r)$ cointegration matrix, $\mathbf{\Gamma}_j$ is a $(K \times K)$ short-run coefficient matrix for $j = 1, \dots, p-1$, and u_t is a white noise error vector with $u_t \sim (0, \Sigma_u)$.

In Chapter 6, Proposition 6.1, we have seen that the process has the Beveridge-Nelson MA representation

$$y_t = \Xi \sum_{i=1}^{t} u_i + \sum_{j=0}^{\infty} \Xi_j^* u_{t-j} + y_0^*, \tag{9.2.2}$$

where the Ξ_j^* are absolutely summable so that the infinite sum is well-defined and the term y_0^* contains the initial values. Absolute summability of the Ξ_j^* implies that these matrices converge to zero for $j \to \infty$. Thus, the long-run effects of shocks are captured by the common trends term $\Xi \sum_{i=1}^{t} u_i$. The matrix

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

has rank $K - r$. Thus, there are $K - r$ common trends and if the structural innovations embodied in the u_i can be recovered, at most r of them can have transitory effects only because the matrix Ξ or a nonsingular transformation of this matrix cannot have more than r columns of zeros. Thus, by knowing the cointegrating rank of the system, we know already the maximum number of transitory shocks.

In this context, the focus of interest is usually on the residuals and, hence, in order to identify the structural innovations, the B-model setup is typically used. In other words, we are looking for a matrix B such that

$$u_t = \mathsf{B}\varepsilon_t \quad \text{with} \quad \varepsilon_t \sim (0, I_K).$$

Substituting this relation in the common trends term gives $\Xi\mathsf{B}\sum_{i=1}^{t}\varepsilon_i$. Hence, the long-run effects of the structural innovations are given by

$\Xi\mathsf{B}$.

Because the structural innovations represent a regular random vector with nonsingular covariance matrix, the matrix B has to be nonsingular. Recall that $\Sigma_u = \mathsf{B}\mathsf{B}'$. Thus, $\mathrm{rk}(\Xi\mathsf{B}) = K - r$ and there can be at most r zero columns in this matrix. In other words, r of the structural innovations can have transitory effects and $K - r$ of them must have permanent effects. If there are r transitory shocks, we can restrict r columns of $\Xi\mathsf{B}$ to zero. Because the matrix has reduced rank $K - r$, each column of zeros stands for $K - r$ independent restrictions only. Thus, the r transitory shocks represent $r(K-r)$ independent restrictions only. Still, it is useful to note that restrictions can be imposed on the basis of our knowledge of the cointegrating rank of the system which can be determined by statistical means. Further theoretical considerations are required for imposing additional restrictions, however.

For local just-identification of the structural innovations in the B-model, we need a total of $K(K-1)/2$ restrictions. Assuming that there are r shocks with transitory effects only, we have already $r(K-r)$ restrictions from the cointegration structure of the model, this leaves us with $\frac{1}{2}K(K-1)-r(K-r)$ further restrictions for just-identifying the structural innovations. In fact, $r(r-1)/2$ additional contemporaneous restrictions are needed to disentangle the transitory shocks and $(K-r)((K-r)-1)/2$ restrictions identify the permanent shocks (see, e.g., King et al. (1991), Gonzalo & Ng (2001)). Then we have a total of $\frac{1}{2}r(r-1)+\frac{1}{2}(K-r)((K-r)-1) = \frac{1}{2}K(K-1)-r(K-r)$ restrictions, as required. Thus, it is not sufficient to impose arbitrary restrictions on B or $\Xi\mathsf{B}$, but we have to choose them to identify the transitory and permanent shocks at least locally. In fact, the transitory shocks can only be identified through restrictions directly on B because they correspond to zero columns in $\Xi\mathsf{B}$. Thus, $r(r-1)/2$ of the restrictions have to be imposed on B directly. Generally, the restrictions have the form

$$C_{\Xi\mathsf{B}}\mathrm{vec}(\Xi\mathsf{B}) = c_l \text{ or } C_l\mathrm{vec}(\mathsf{B}) = c_l \quad \text{and} \quad C_s\mathrm{vec}(\mathsf{B}) = c_s, \tag{9.2.3}$$

where $C_l := C_{\Xi\mathsf{B}}(I_K \otimes \Xi)$ is a matrix of long-run restrictions, that is, $C_{\Xi\mathsf{B}}$ is a suitable selection matrix such that $C_{\Xi\mathsf{B}}\mathrm{vec}(\Xi\mathsf{B}) = c_l$, and C_s specifies short-run or instantaneous constraints by restricting elements of B directly. Here c_l and c_s are vectors of suitable dimensions. In applied work, they are typically zero vectors. In other words, zero restrictions are specified in (9.2.3) for $\Xi\mathsf{B}$ and B.

As discussed for the stationary case in Section 9.1.2, the matrix B will only be locally identified. In particular, in general we may reverse the signs of the columns of B to find another valid matrix. Formal necessary and sufficient conditions for local identification are given in the following proposition.

Proposition 9.4 (*Local Identification of a SVECM*)
Suppose the reduced form model (9.2.1) with Beveridge-Nelson MA representation (9.2.2) is given. Let B be a nonsingular $(K \times K)$ matrix. Then, the set of equations

$$\Sigma_u = \mathsf{B}\mathsf{B}', \quad C_l \mathrm{vec}(\mathsf{B}) = c_l \quad \text{and} \quad C_s \mathrm{vec}(\mathsf{B}) = c_s,$$

with C_l, c_l, C_s, and c_s as in (9.2.3), has a locally unique solution for B if and only if

$$\mathrm{rk} \begin{bmatrix} 2\mathbf{D}_K^+(\mathsf{B} \otimes I_K) \\ C_l \\ C_s \end{bmatrix} = K^2.$$

■

Proof: The model underlying Proposition 9.4 is a B-model. Therefore the proposition can be shown using the same arguments as for Proposition 9.2. Details are omitted. ■

As an example, we consider a small model discussed by King et al. (1991). They specified a model for the logarithms of private output (q_t), consumption (c_t), and investment (i_t). Assuming that all three variables are $I(1)$ with cointegrating rank $r = 2$ and that there are two transitory shocks and one permanent shock, the permanent shock is identified without further assumptions because $K - r = 1$ and, hence, $(K - r)((K - r) - 1)/2 = 0$. Moreover, only 1 $(= r(r-1)/2)$ further restriction is necessary to identify the two transitory shocks. Placing the permanent shock first in the ε_t vector and allowing the first transitory shock to have instantaneous effects on all variables, we may use the following restrictions:

$$\Xi\mathsf{B} = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathsf{B} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & * \end{bmatrix}. \tag{9.2.4}$$

Here asterisks denote unrestricted elements. The two zero columns in $\Xi\mathsf{B}$ represent two independent restrictions only because $\Xi\mathsf{B}$ has rank 1. A third restriction is placed on B in such a way that the third shock does not have an instantaneous effect on the second variable. Hence, there are $K(K-1)/2 = 3$ independent restrictions in total and the structural innovations are locally just-identified. Uniqueness can be obtained by fixing the signs of the diagonal elements of B.

In our three-dimensional example with two zero columns in $\Xi\mathsf{B}$, it does not suffice to impose a further restriction on this matrix to ensure local uniqueness of B. For that we need to disentangle the two transitory shocks which cannot be identified by restrictions on the long-run matrix $\Xi\mathsf{B}$. Thus, we have to impose a restriction directly on B. In fact, it is necessary to restrict an element in the last two columns of B (see also Problem 9.1 for further details).

In the standard B-model with three variables, we need to specify at least 3 restrictions for identification. In contrast, in the present VECM case, assuming that $r = 2$ and there are two transitory shocks, only one restriction is needed because two columns of $\Xi\mathsf{B}$ are zero. Thus, taking into account the long-run restrictions from the cointegration properties of the variables may result in substantial simplifications. In fact, for a bivariate system with one cointegrating relation, no further restriction is required to identify the permanent and transitory shocks. It is enough to specify that the first shock is allowed to have permanent effects while the second one can only have transitory effects or vice versa. A more detailed higher-dimensional example may be found in Breitung et al. (2004). Further discussion of partitioning the shocks in permanent and transitory ones is also given in Gonzalo & Ng (2001) and Fisher & Huh (1999).

9.3 Estimation of Structural Parameters

We will first consider estimation of the AB-SVAR model and then discuss SVECMs. The A- and B-models are straightforward special cases which are not treated separately in detail. For both SVARs and SVECMs, ML methods are typically used and they will therefore be presented here.

9.3.1 Estimating SVAR Models

Suppose we wish to estimate the following SVAR model

$$\mathsf{A}y_t = \mathsf{A}AY_{t-1} + \mathsf{B}\varepsilon_t, \tag{9.3.1}$$

where $Y'_{t-1} := [y'_{t-1}, \dots, y'_{t-p}]$, $A := [A_1, \dots, A_p]$, and ε_t is assumed to be Gaussian white noise with covariance matrix I_K, $\varepsilon_t \sim \mathcal{N}(0, I_K)$. The normality assumption is just made for convenience to derive the estimators. The asymptotic properties of the estimators will be the same under more general distributional assumptions, as usual. The reduced form residuals corresponding to (9.3.1) have the form $u_t = \mathsf{A}^{-1}\mathsf{B}\varepsilon_t$.

From Chapter 3, Section 3.4, the log-likelihood function for a sample $y_1, \dots, y_T$ is seen to be

$$\begin{aligned}
\ln l(A, \mathsf{A}, \mathsf{B}) &= -\tfrac{KT}{2}\ln 2\pi - \tfrac{T}{2}\ln|\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1}| \\
&\quad -\tfrac{1}{2}\mathrm{tr}\{(Y - AX)'[\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1}]^{-1}(Y - AX)\} \\
&= \text{constant} + \tfrac{T}{2}\ln|\mathsf{A}|^2 - \tfrac{T}{2}\ln|\mathsf{B}|^2 \\
&\quad -\tfrac{1}{2}\mathrm{tr}\{\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}(Y - AX)(Y - AX)'\},
\end{aligned} \tag{9.3.2}$$

where, as usual, $Y := [y_1, \dots, y_T]$, $X := [Y_0, \dots, Y_{T-1}]$, and the matrix rules $|\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'(\mathsf{A}^{-1})'| = |\mathsf{A}^{-1}|^2|\mathsf{B}|^2 = |\mathsf{A}|^{-2}|\mathsf{B}|^2$ and $\mathrm{tr}(VW) = \mathrm{tr}(WV)$ have been used (see Appendix A).

Suppose there are no restrictions on the reduced form parameters A. Then, it follows from Section 3.4 that for any given A and B, the log-likelihood function $\ln l(A, \mathsf{A}, \mathsf{B})$ is maximized with respect to A by $\widehat{A} = YX'(XX')^{-1}$. Thus, replacing A with $\widehat{A}$ in (9.3.2) gives the concentrated log-likelihood

$$\ln l_c(\mathsf{A}, \mathsf{B}) = \text{constant} + \frac{T}{2} \ln |\mathsf{A}|^2 - \frac{T}{2} \ln |\mathsf{B}|^2 - \frac{T}{2} \text{tr}(\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}\widetilde{\Sigma}_u), \quad (9.3.3)$$

where $\widetilde{\Sigma}_u = T^{-1}(Y - \widehat{A}X)(Y - \widehat{A}X)'$. Maximization of this function with respect to A and B, subject to the structural restrictions (9.1.12) or (9.1.13), has to be done by numerical methods because a closed form solution is usually not available. If the restrictions are of the form (9.1.12), restricted maximization of the concentrated log-likelihood amounts to maximization with respect to γ_A and γ_B. If these parameters are locally identified, the ML estimators have standard asymptotic properties which are summarized in the following proposition.

Proposition 9.5 (*Properties of the SVAR ML Estimators*)
Suppose y_t is a stationary Gaussian VAR(p) as in (9.1.1) and structural restrictions of the form (9.1.12) are available such that γ_A and γ_B are locally identified. Then the ML estimators $\widetilde{\gamma}_\mathsf{A}$ and $\widetilde{\gamma}_\mathsf{B}$ are consistent and asymptotically normally distributed,

$$\sqrt{T}\left(\begin{bmatrix} \widetilde{\gamma}_\mathsf{A} \\ \widetilde{\gamma}_\mathsf{B} \end{bmatrix} - \begin{bmatrix} \gamma_\mathsf{A} \\ \gamma_\mathsf{B} \end{bmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_a\begin{pmatrix} \gamma_\mathsf{A} \\ \gamma_\mathsf{B} \end{pmatrix}^{-1}\right),$$

where $\mathcal{I}_a(\cdot)$ is the asymptotic information matrix. It has the form

$$\mathcal{I}_a\begin{pmatrix} \gamma_\mathsf{A} \\ \gamma_\mathsf{B} \end{pmatrix} = \begin{bmatrix} R_\mathsf{A}' & 0 \\ 0 & R_\mathsf{B}' \end{bmatrix} \mathcal{I}_a\begin{pmatrix} \text{vec}\,\mathsf{A} \\ \text{vec}\,\mathsf{B} \end{pmatrix} \begin{bmatrix} R_\mathsf{A} & 0 \\ 0 & R_\mathsf{B} \end{bmatrix}$$

and

$$\begin{aligned} &\mathcal{I}_a\begin{pmatrix} \text{vec}\,\mathsf{A} \\ \text{vec}\,\mathsf{B} \end{pmatrix} \\ &= \begin{bmatrix} \mathsf{A}^{-1}\mathsf{B} \otimes \mathsf{B}'^{-1} \\ -(I_K \otimes \mathsf{B}'^{-1}) \end{bmatrix} (I_{K^2} + \mathbf{K}_{KK}) \\ &\qquad \times \left[(\mathsf{B}'\mathsf{A}'^{-1} \otimes \mathsf{B}^{-1}) : -(I_K \otimes \mathsf{B}^{-1})\right] \end{aligned} \quad (9.3.4)$$

■

Proof: The proposition follows from the general ML theory (see Appendix C.6). For the derivation of the asymptotic information matrix see Problem 9.4. ■

If γ_A and γ_B are identified, the same is true for A and B. Estimating these matrices such that $\text{vec}(\widetilde{\mathsf{A}}) = R_\mathsf{A}\widetilde{\gamma}_\mathsf{A} + r_\mathsf{A}$ and $\text{vec}(\widetilde{\mathsf{B}}) = R_\mathsf{B}\widetilde{\gamma}_\mathsf{B} + r_\mathsf{B}$, respectively, we get the following immediate implication of Proposition 9.5.

Corollary 9.5.1
Under the conditions of Proposition 9.5,

$$\sqrt{T}\left(\begin{bmatrix} \operatorname{vec}\widetilde{\mathsf{A}} \\ \operatorname{vec}\widetilde{\mathsf{B}} \end{bmatrix} - \begin{bmatrix} \operatorname{vec}\mathsf{A} \\ \operatorname{vec}\mathsf{B} \end{bmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathsf{AB}}),$$

where

$$\Sigma_{\mathsf{AB}} = \begin{bmatrix} R_{\mathsf{A}} & 0 \\ 0 & R_{\mathsf{B}} \end{bmatrix} \mathcal{I}_a \begin{pmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{pmatrix}^{-1} \begin{bmatrix} R'_{\mathsf{A}} & 0 \\ 0 & R'_{\mathsf{B}} \end{bmatrix}.$$

■

If only just-identifying restrictions are imposed on the structural parameters, we have for the ML estimator of Σ_u,

$$\widetilde{\Sigma}_u = T^{-1}(Y - \widehat{A}X)(Y - \widehat{A}X)' = \widetilde{\mathsf{A}}^{-1}\widetilde{\mathsf{B}}\widetilde{\mathsf{B}}'\widetilde{\mathsf{A}}'^{-1}.$$

If, however, over-identifying restrictions have been imposed on A and/or B, the corresponding estimator for Σ_u,

$$\widetilde{\Sigma}_u^r := \widetilde{\mathsf{A}}^{-1}\widetilde{\mathsf{B}}\widetilde{\mathsf{B}}'\widetilde{\mathsf{A}}'^{-1}, \tag{9.3.5}$$

will differ from $\widetilde{\Sigma}_u$. In fact, the LR statistic,

$$\lambda_{LR} = T(\ln|\widetilde{\Sigma}_u^r| - \ln|\widetilde{\Sigma}_u|), \tag{9.3.6}$$

can be used to check the over-identifying restrictions. Under the null hypothesis that the restrictions are valid, it has an asymptotic χ^2-distribution with degrees of freedom equal to the number of over-identifying restrictions. In other words, the number of degrees of freedom is equal to the number of independent constraints imposed on A and B minus $2K^2 - \frac{1}{2}K(K+1)$.

Computation of ML Estimates

Because the structural parameters A and B are nonlinearly related to the reduced form parameters, no closed form of the ML estimates exists in general and an iterative optimization algorithm may be used for actually computing the ML estimates. Amisano & Giannini (1997) proposed to use a scoring algorithm for this purpose. The i-th iteration of this algorithm is of the form

$$\begin{bmatrix} \widetilde{\gamma}_{\mathsf{A}} \\ \widetilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i+1} = \begin{bmatrix} \widetilde{\gamma}_{\mathsf{A}} \\ \widetilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i} + \ell\, \mathcal{I}\left(\begin{bmatrix} \widetilde{\gamma}_{\mathsf{A}} \\ \widetilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i}\right)^{-1} \mathbf{s}\left(\begin{bmatrix} \widetilde{\gamma}_{\mathsf{A}} \\ \widetilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i}\right), \tag{9.3.7}$$

where $\mathcal{I}(\cdot)$ denotes the information matrix of the free parameters γ_{A}, γ_{B}, that is, in this case $\mathcal{I}(\cdot) = T\mathcal{I}_a(\cdot)$, $\mathbf{s}(\cdot)$ is the score vector and ℓ is the step length (see also Chapter 12, Section 12.3.2, for further discussion of optimization algorithms of this type).

The score vector can be obtained using the rules for matrix and vector differentiation (Appendix A.13). Applying the chain rule for vector differentiation, it is seen to be

$$\mathbf{s}\begin{pmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{pmatrix} = \frac{\partial \ln l}{\partial(\gamma_{\mathsf{A}}', \gamma_{\mathsf{B}}')'} = \begin{bmatrix} R_{\mathsf{A}}' & 0 \\ 0 & R_{\mathsf{B}}' \end{bmatrix} \mathbf{s}\begin{pmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{pmatrix}, \tag{9.3.8}$$

and

$$\mathbf{s}\begin{pmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{pmatrix} = \frac{\partial \ln l}{\partial \begin{pmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{pmatrix}} = \begin{bmatrix} (I_K \otimes \mathsf{B}'^{-1}) \\ -(\mathsf{B}^{-1}\mathsf{A} \otimes \mathsf{B}'^{-1}) \end{bmatrix} \mathbf{s}(\operatorname{vec}[\mathsf{B}^{-1}\mathsf{A}])$$

with

$$\mathbf{s}(\operatorname{vec}[\mathsf{B}^{-1}\mathsf{A}]) = T\operatorname{vec}([\mathsf{B}^{-1}\mathsf{A}]'^{-1}) - T(\widetilde{\Sigma}_u \otimes I_K)\operatorname{vec}(\mathsf{B}^{-1}\mathsf{A})$$

(see Problem 9.3 for further details). In practice, the iterations of the scoring algorithm terminate if prespecified convergence criteria, such as the relative change in the log-likelihood and the parameters, are satisfied. For this algorithm to work, the inverse of the information matrix has to exist which is guaranteed by the identification of the parameters, at least in a neighborhood of the true parameter values. Giannini (1992) used this property to derive alternative conditions for identification of the models presented in Section 9.1. More precisely, he derived identification conditions from the fact that, for instance, the AB-model is locally identified if and only if the matrix

$$\begin{bmatrix} \mathcal{I}_a\begin{pmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{pmatrix} \\ \begin{bmatrix} C_{\mathsf{A}} & 0 \\ 0 & C_{\mathsf{B}} \end{bmatrix} \end{bmatrix} \tag{9.3.9}$$

has full column rank when $\mathcal{I}_a(\cdot)$ is evaluated at the true parameter values (see Rothenberg (1971)).

Although we have discussed models without deterministic terms and restrictions on the reduced form parameters, the ML estimation procedure for the structural parameters can be extended easily to more general situations which cover these complications. Again, estimation of the structural parameters can be based on the concentrated likelihood function. If there are restrictions for the reduced form parameters A, for example, if a subset model is considered, one may even use the EGLS estimator instead of the ML estimator for these parameters in estimating the structural parameters. Clearly, in that case, the white noise covariance estimator $\widetilde{\Sigma}_u$ will not be the exact ML estimator and the exact concentrated log-likelihood is obtained only if ML estimators are substituted for the reduced form parameters A. Asymptotically, the corresponding estimators $\widetilde{\mathsf{A}}, \widetilde{\mathsf{B}}$ based on the EGLS estimators will have the same properties as the exact ML estimators, however. Even in small samples, exact ML estimation may not result in substantial gains (see, e.g., Brüggemann (2004)).

Estimation with Long-Run Restrictions à la Blanchard-Quah

If the total impact matrix Ξ_∞ is restricted to be triangular as in Blanchard & Quah (1989) and Galí (1999), estimation becomes particularly easy. Specifying $\mathsf{A} = I_K$, using the relation $\Xi_\infty = (I_K - A_1 - \cdots - A_p)^{-1}\mathsf{B}$ and noting that

$$\Xi_\infty \Xi_\infty' = (I_K - A_1 - \cdots - A_p)^{-1}\Sigma_u(I_K - A_1' - \cdots - A_p')^{-1},$$

the matrix B can be estimated by premultiplying a Choleski decomposition of the matrix

$$(I_K - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1}\widetilde{\Sigma}_u(I_K - \widehat{A}_1' - \cdots - \widehat{A}_p')^{-1}$$

by $(I_K - \widehat{A}_1 - \cdots - \widehat{A}_p)$.

This latter procedure works only if the VAR operator is stable and the process is stationary because for integrated processes the inverse of $(I_K - A_1 - \cdots - A_p)$ does not exist, as explained earlier. On the other hand, cointegrated variables do not create problems for the other estimation methods for SVAR models.

9.3.2 Estimating Structural VECMs

Suppose the structural restrictions for a VECM are given in the form of linear restrictions on $\Xi\mathsf{B}$ and B, as in (9.2.3). For computing the parameter estimates, we may replace Ξ by its reduced form ML estimator,

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

where the $\widetilde{\mathbf{\Gamma}}_i$'s are the ML estimators of the $\mathbf{\Gamma}_i$'s from Proposition 7.3 and $\widetilde{\alpha}_\perp$ and $\widetilde{\beta}_\perp$ are any orthogonal complements of the ML estimators $\widetilde{\alpha}$ and $\widetilde{\beta}$, respectively. The restricted ML estimator of B can be obtained by setting $\mathsf{A} = I_K$ and optimizing the concentrated log-likelihood function (9.3.3) with respect to B, subject to the restrictions (9.2.3), with C_l replaced by

$$\widetilde{C}_l = C_{\Xi\mathsf{B}}(I_K \otimes \widetilde{\Xi})$$

(see Vlaar (2004)). Although this procedure results in a set of stochastic restrictions, from a numerical point of view we have a standard constrained optimization problem which can be solved by a Lagrange approach (see Appendix A.14) because $\widetilde{\Xi}$ is fixed in computing the estimate of B. Due to the fact that for a just-identified structural model the log-likelihood maximum is the same as for the reduced form, a comparison of the log-likelihood values can serve as a check for a proper convergence of the optimization algorithm used for structural estimation.

The properties of the ML estimator of B follow in principle from Corollary 9.5.1. In other words, $\widetilde{\mathsf{B}}$ is consistent and asymptotically normal under standard conditions,

$$\sqrt{T}\,\mathrm{vec}(\widetilde{\mathsf{B}} - \mathsf{B}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathsf{B}}).$$

The asymptotic distribution is singular because of the restrictions that have been imposed on B. Thus, although t-ratios can be used for assessing the significance of individual parameters, F-tests based on the Wald principle will in general not be valid and have to be interpreted cautiously. Expressions for the covariance matrices of the asymptotic distributions in terms of the model parameters can be obtained in the usual way by working out the corresponding information matrices (see Vlaar (2004)). For practical purposes, it is common to use bootstrap methods for inference in this context.

In principle, the same approach can be used if there are over-identifying restrictions for B. In that case, $\widetilde{\mathsf{B}}\widetilde{\mathsf{B}}'$ will not be equal to the reduced form white noise covariance estimator $\widetilde{\Sigma}_u$, however. Still the estimator of B will be consistent and asymptotically normal under general conditions and also the LR statistic given in (9.3.6) can be used to check the validity of the over-identifying restrictions. It will have the usual asymptotic χ^2-distribution with degrees of freedom equal to the number of over-identifying restrictions.

9.4 Impulse Response Analysis and Forecast Error Variance Decomposition

Impulse response analysis can now be based on structural innovations. In other words, the impulse response coefficients are obtained from the matrices

$$\Theta_j = \Phi_j \mathsf{A}^{-1}\mathsf{B}, \quad j = 0, 1, 2, \ldots .$$

Using the same reasoning as in Chapter 3, Section 3.7, the corresponding estimated quantities are asymptotically normal as nonlinear functions of asymptotically normal parameter estimators,

$$\sqrt{T}\,\mathrm{vec}(\widehat{\Theta}_j - \Theta_j) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\Theta}_j}).$$

In practice, bootstrap methods are routinely employed for inference in this context. However, the same inference problems as in Chapter 3, Section 3.7, prevail for structural impulse responses. More precisely, the asymptotic distribution may be singular in which case confidence intervals based on asymptotic theory or bootstrap methods may not have the desired confidence level even asymptotically.

We use a set of quarterly U.S. data for the period 1947.1–1988.4 from King et al. (1991) for the three variables log private output (q_t), consumption

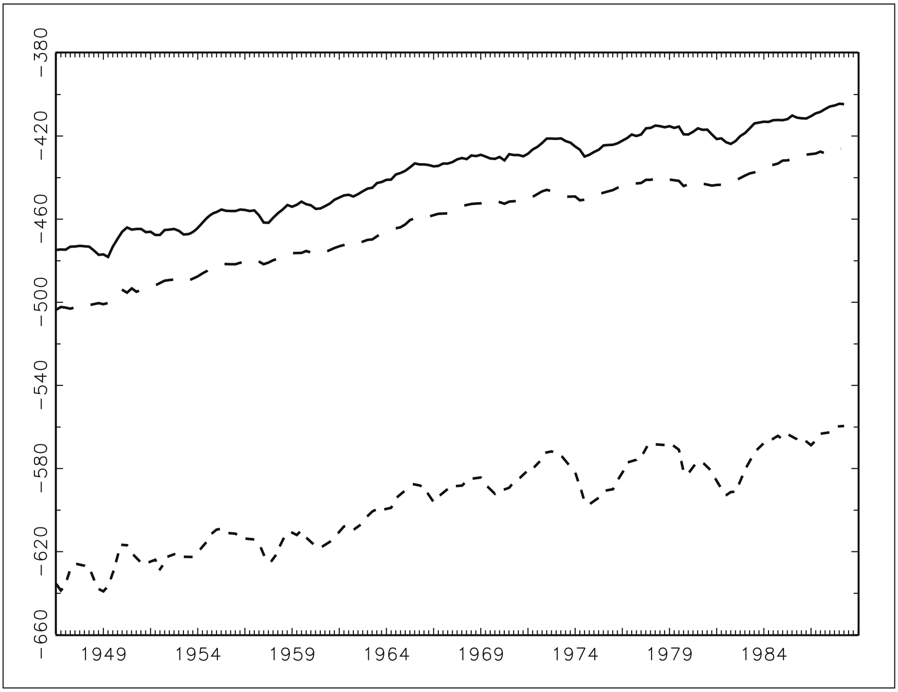

Fig. 9.1. Quarterly U.S. log private output (—), consumption (– –), and investment (- - -).

(c_t), and investment (i_t) (all multiplied by 100) to illustrate structural impulse responses.[1] The three series are plotted in Figure 9.1. They all have a trending behavior and there is some evidence that they are well modelled as $I(1)$ series. Applying LR tests for the cointegrating rank with a trend orthogonal to the cointegration relations to a model with one lagged difference of the variables, provides evidence for two cointegration relations, that is, $r = 2$ (see Section 8.2.4 for the description of the tests). Therefore we proceed from the following estimated reduced form VECM (t-statistics in parentheses):

$$\begin{bmatrix} \Delta q_t \\ \Delta c_t \\ \Delta i_t \end{bmatrix} = \begin{bmatrix} \underset{(-0.2)}{-0.88} \\ \underset{(-1.1)}{-2.83} \\ \underset{(-4.1)}{-30.07} \end{bmatrix}$$

[1] The data are available at the website http://www.wws.princeton.edu/ mwatson/.

$$+\begin{bmatrix} \underset{(-3.6)}{-0.23} & \underset{(4.6)}{0.20} \\ \underset{(-1.5)}{-0.06} & \underset{(2.4)}{0.07} \\ \underset{(-0.9)}{-0.11} & \underset{(2.9)}{0.26} \end{bmatrix} \begin{bmatrix} 1 & 0 & \underset{(-27.7)}{-1.02} \\ 0 & 1 & \underset{(-24.2)}{-1.10} \end{bmatrix} \begin{bmatrix} q_{t-1} \\ c_{t-1} \\ i_{t-1} \end{bmatrix}$$

$$+\begin{bmatrix} \underset{(1.2)}{0.12} & \underset{(0.7)}{0.09} & \underset{(3.4)}{0.16} \\ \underset{(3.2)}{0.21} & \underset{(-2.3)}{-0.21} & \underset{(0.8)}{0.02} \\ \underset{(3.6)}{0.70} & \underset{(-0.6)}{-0.17} & \underset{(3.6)}{0.33} \end{bmatrix} \begin{bmatrix} \Delta q_{t-1} \\ \Delta c_{t-1} \\ \Delta i_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}. \quad (9.4.1)$$

Before we can proceed with structural estimation, we have to specify identifying restrictions. Using the zero restrictions from (9.2.4), the following estimates are obtained:

$$\widetilde{\mathsf{B}} = \begin{bmatrix} \underset{(0.4)}{0.08} & \underset{(3.9)}{1.03} & \underset{(-0.8)}{-0.45} \\ \underset{(-0.7)}{-0.60} & \underset{(4.1)}{0.43} & 0 \\ \underset{(0.6)}{0.26} & \underset{(5.1)}{1.96} & \underset{(1.9)}{1.00} \end{bmatrix} \quad (9.4.2)$$

and

$$\widetilde{\Xi}\widetilde{\mathsf{B}} = \begin{bmatrix} \underset{(-0.8)}{-0.71} & 0 & 0 \\ \underset{(-0.8)}{-0.76} & 0 & 0 \\ \underset{(-0.8)}{-0.69} & 0 & 0 \end{bmatrix}.$$

Here bootstrapped t-statistics based on 2000 bootstrap replications are given in parentheses. In other words, the standard deviations of the estimates are obtained with a bootstrap (see Appendix D.3) and then the estimated coefficients are divided by their respective bootstrap standard deviations to get the t-ratios. Clearly, some of the t-ratios are quite small. Thus, it may be possible to impose over-identifying restrictions. In fact, because all t-ratios of the nonzero long-run effects are small, it may be tempting to argue that no significant permanent effect is found. Recall, however, that, based on the unit root and cointegration analysis, there cannot be more shocks with transitory effects. We have used the just-identified model for an impulse response analysis to shed more light on this issue.

There are three structural innovations, one of which must have permanent effects if the cointegration rank is 2. In Figure 9.2, the responses of all three variables to the shock with potentially permanent effects are depicted. The 95% confidence intervals are based on 2000 replications. Considering the confidence intervals determined with Hall's percentile method (see Appendix D.3), it turns out that none of the confidence intervals associated with longer term responses contains zero. Hence, a significant long-run effect may actually

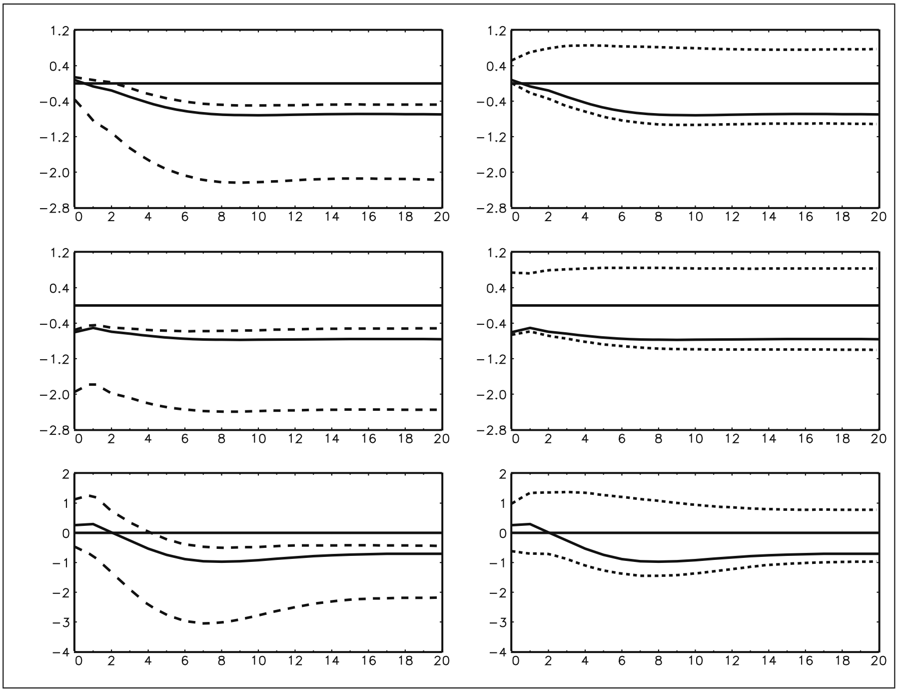

Fig. 9.2. Responses of output, consumption, and investment (top to bottom) to a permanent shock with Hall percentile (left) and standard percentile (right) 95% bootstrap confidence intervals based on 2000 bootstrap replications.

be present for each of the three variables. If, however, the standard percentile bootstrap confidence intervals are used for the impulse responses, the situation is quite different. These confidence intervals are also shown in Figure 9.2 and they all include zero for longer term horizons. Thus, the results are not very robust with respect to the methods used. Clearly, the confidence intervals are quite asymmetric around the point estimates. In such a situation the Hall percentile confidence intervals may be more reliable due to their built-in bias correction.

The estimated responses to the permanent shock are all negative in the long-run. To see the effects of an impulse which leads to positive long-run effects, we can just reverse the signs of the responses. This follows from the unidentified signs of the columns of B discussed in Sections 9.1.2 and 9.2. Generally, the effects of positive and negative shocks of the same size are identical in absolute value because our model is a linear one which does not permit asymmetric reactions to positive and negative shocks.

In Figure 9.3, the responses of the variables to the two transitory shocks are shown. All impulse responses approach zero quickly after some periods and the effects of the shocks after 20 periods are practically negligible. The

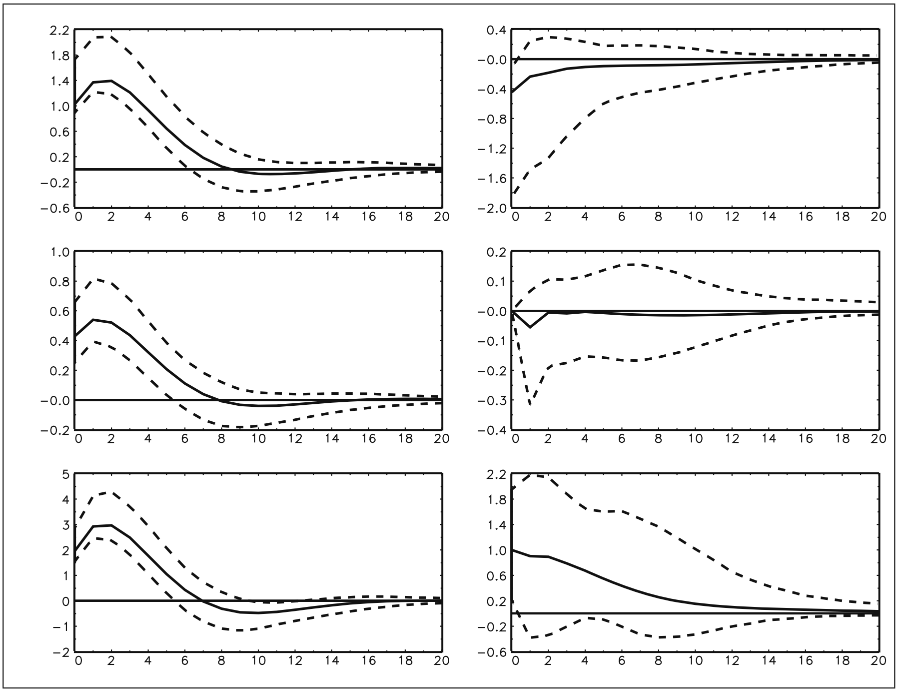

Fig. 9.3. Responses of output, consumption, and investment (top to bottom) to transitory shocks with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications (identification restriction (9.4.2)).

identifying restriction on the B matrix is clearly seen in the right-hand panel in the middle row of Figure 9.3. Here the instantaneous effect of the second transitory shock on c_t is zero. If a zero restriction is imposed instead on the upper right-hand corner element of B, the estimated matrix becomes

$$\widetilde{\mathsf{B}} = \begin{bmatrix} \underset{(0.4)}{0.08} & \underset{(5.7)}{1.12} & 0 \\ \underset{(-0.7)}{-0.60} & \underset{(2.9)}{0.39} & \underset{(1.4)}{0.17} \\ \underset{(0.6)}{0.26} & \underset{(4.5)}{1.39} & \underset{(11.2)}{1.70} \end{bmatrix} \tag{9.4.3}$$

and the corresponding structural impulse responses are depicted in Figure 9.4. Obviously, the identification restriction determines to some extent the shape of the impulse responses. At least the responses to the second transitory shock are quite different from those based on the identification restriction (9.4.2). Now, of course, q_t reacts only with a delay to the second transitory shock. The first column of $\widetilde{\mathsf{B}}$ in (9.4.3) is unchanged relative to (9.4.2) and, more generally, the responses to the permanent shock (not shown) are unaffected because that shock is identified without additional restrictions.

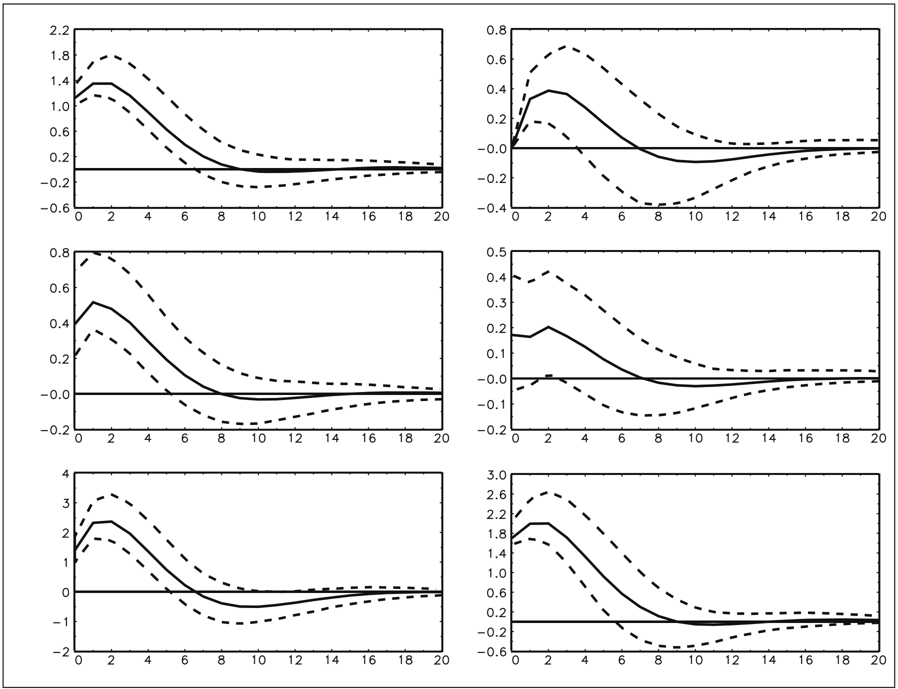

Fig. 9.4. Responses of output, consumption, and investment (top to bottom) to transitory shocks with 95% Hall percentile bootstrap confidence intervals based on 2000 bootstrap replications (identification restriction (9.4.3)).

Forecast error variance decompositions can also be based on the structural innovations. The computations are based on the Θ_j as in Section 2.3.3. The interpretation may be different, however. It may not be possible to associate the structural innovations uniquely with the variables of the system. Therefore, the forecast errors are not decomposed into contributions of the different variables but into the contributions of the structural innovations. For instance, for the example system with identifying restriction on B as in (9.4.2), a forecast error variance decomposition is shown in Figure 9.5. Now we can see that the permanent shocks (the first components of the ε_t's) have a growing importance with increasing forecast horizon, where the estimation uncertainty is ignored, however. In turn, the importance of the transitory shocks (shocks number 2 and 3) declines for all three variables. Actually, the third shock (the second transitory shock) does not contribute much to the forecast errors of any of the three variables.

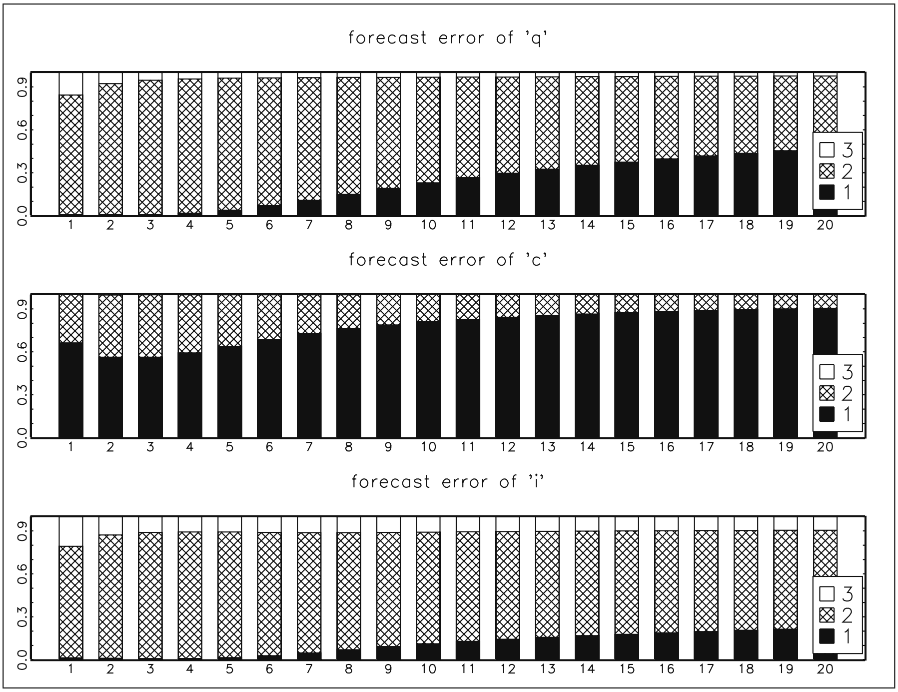

Fig. 9.5. Forecast error variance decomposition of the output, consumption, and investment system based on identification scheme (9.4.2) with relative contributions of the permanent shock (1) and the two transitory shocks (2 and 3).

9.5 Further Issues

Structural VARs and VECMs have not only found widespread use in applied work but there are also numerous further methodological contributions. For example, confidence bands for impulse responses are sometimes constructed with Bayesian methods (e.g., Koop (1992)). In fact, the practice of reporting confidence intervals around individual impulse response coefficients was questioned by Sims & Zha (1999). They proposed likelihood-characterizing error bands as alternatives.

Also other forms of identifying restrictions were considered by some authors. For example, Uhlig (1994) proposed to use inequality constraints for the impulse responses for identifying them. In contrast, Lee, Pesaran & Pierse (1992) and Pesaran & Shin (1996) considered persistence profiles which measure the persistence of certain shocks without imposing structural identification restrictions.

It may be worth remembering, however, that structural impulse responses are not immune to some of the problems discussed in Chapter 2 in the context of impulse response analysis. In particular, omitted variables, filtering and adjusting series prior to using them for a VAR analysis and using aggregated or transformed data can lead to major changes in the dynamic behavior of the model. For instance, if an important variable is omitted from a system of interest, adding it can change in principle all the impulse responses. Similarly, using seasonally adjusted and, hence, filtered data can change the dynamic structure of the variables and, thus, may lead to impulse responses which are quite different from those for unadjusted variables. These problems are not solved by imposing identifying restrictions and are worth keeping in mind also in a structural VAR analysis.

9.6 Exercises

9.6.1 Algebraic Problems

Problem 9.1
Show that for a three-dimensional VECM with cointegration rank $r = 2$, the set of restrictions

$$\Xi\mathsf{B} = \begin{bmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix}$$

is not sufficient for identification. Moreover, show that the restrictions

$$\Xi\mathsf{B} = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathsf{B} = \begin{bmatrix} 0 & * & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

do not identify B locally.
(Hint: Choose

$$\mathsf{B} = \begin{bmatrix} \mathsf{b}_{11} & 0 \\ 0 & \mathsf{B}_2 \end{bmatrix},$$

where B_2 is a (2×2) matrix. Show that B_2 is not unique.)

Problem 9.2
Suppose a four-dimensional process y_t can be written in VECM form (9.2.1) with cointegrating rank 2. Impose just-identifying restrictions on B and $\Xi\mathsf{B}$.

Problem 9.3
Define $C = \mathsf{B}^{-1}\mathsf{A}$ and write the concentrated log-likelihood (9.3.3) as

$$\ln l_c(C) = \text{constant} + T \ln |C| - \frac{T}{2}\text{tr}(C'C\widetilde{\Sigma}_u).$$

Use the rules for matrix differentiation from Appendix A.13 to show that

$$\frac{\partial \ln l_c}{\partial C} = TC'^{-1} - TC\widetilde{\Sigma}_u.$$

Next show that

$$\frac{\partial \operatorname{vec}(\mathsf{B}^{-1}\mathsf{A})}{\partial \operatorname{vec}(\mathsf{A})'} = I_K \otimes \mathsf{B}^{-1}$$

and

$$\frac{\partial \operatorname{vec}(\mathsf{B}^{-1}\mathsf{A})}{\partial \operatorname{vec}(\mathsf{B})'} = -(\mathsf{A}'\mathsf{B}'^{-1} \otimes \mathsf{B}^{-1}).$$

Use these results to derive an explicit expression for the score vector

$$\mathbf{s}\begin{pmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{pmatrix} = \frac{\partial \ln l}{\partial(\gamma'_{\mathsf{A}}, \gamma'_{\mathsf{B}})'}.$$

Problem 9.4
Define $\mathsf{a} := [\operatorname{vec}(\mathsf{A})', \operatorname{vec}(\mathsf{B})']'$ and $\gamma := (\gamma'_{\mathsf{A}}, \gamma'_{\mathsf{B}})'$ and show that, for the setup in Proposition 9.5,

$$-E\left(\frac{\partial^2 \ln l}{\partial\gamma\partial\gamma'}\right) = -\begin{bmatrix} R'_{\mathsf{A}} & 0 \\ 0 & R'_{\mathsf{B}} \end{bmatrix} E\left(\frac{\partial^2 \ln l}{\partial\mathsf{a}\partial\mathsf{a}'}\right) \begin{bmatrix} R_{\mathsf{A}} & 0 \\ 0 & R_{\mathsf{B}} \end{bmatrix}.$$

Moreover, show that (9.3.4) holds by proving that

$$E\left(\frac{\partial^2 \ln l}{\partial\mathsf{a}\partial\mathsf{a}'}\right) = \frac{\partial \operatorname{vec}(\Sigma_u)'}{\partial\mathsf{a}} E\left(\frac{\partial^2 \ln l}{\partial \operatorname{vec}(\Sigma_u)\partial \operatorname{vec}(\Sigma_u)'}\right) \frac{\partial \operatorname{vec}(\Sigma_u)}{\partial\mathsf{a}'}$$

and, for C such that $CC' = \Sigma_u$,

$$\frac{\partial \operatorname{vec}(\Sigma_u)}{\partial\mathsf{a}'} = \frac{\partial \operatorname{vec}(CC')}{\partial \operatorname{vec}(C)'} \frac{\partial \operatorname{vec}(C)}{\partial\mathsf{a}'} = (I_{K^2} + \mathbf{K}_{KK})(C \otimes I_K)\frac{\partial \operatorname{vec}(C)}{\partial\mathsf{a}'}$$

(see also Chapter 3 for related derivations).

9.6.2 Numerical Problems

Problem 9.5
Specify, estimate, and analyze a model for U.S. quarterly log output (q_t) and the unemployment rate (ur_t) for the period 1948.2–1987.4 as given in the *Journal of Applied Econometrics* data archive at
http://www.econ.queensu.ca/jae/
(see the data for Weber (1995)). Blanchard & Quah (1989) considered this system in their study.

(a) Analyze the integration and cointegration properties of the data.
(b) Fit a suitable VAR model to the bivariate series.
(c) Check the adequacy of the model.
(d) Impose an identifying restriction on the long-run total impact matrix and perform a structural impulse response analysis.

(e) Compare standard and Hall percentile confidence intervals for the impulse responses and interpret possible differences.
(f) Perform a forecast error variance decomposition and comment on the results.

(Hint: See Breitung et al. (2004) for a similar analysis.)

Problem 9.6
Analyze the Canadian labor market data from Breitung et al. (2004) (see
www.jmulti.de → datasets
for the data). The variables are:
p_t – ln productivity,
e_t – ln employment,
ur_t – unemployment rate,
w_t – ln real wage index.
Thus, $y_t = (p_t, e_t, ur_t, w_t)'$ is four-dimensional. The data are quarterly for the period 1980.1–2000.4. They are constructed as described in Breitung et al. (2004) based on data from the OECD database. Note that Breitung et al. (2004) use a slightly different notation for the variables.

(a) Analyze the integration and cointegration properties of the data.
(b) Fit a VECM with cointegration rank $r = 1$ for y_t.
(c) Check the adequacy of your model.
(d) Impose identifying restrictions of the form

$$\mathsf{B} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & 0 & * & * \end{bmatrix} \quad \text{and} \quad \Xi\mathsf{B} = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \end{bmatrix}$$

and perform a structural impulse response analysis.
(e) Compare standard and Hall percentile confidence intervals for the impulse responses and interpret possible differences.
(f) Impose another zero restriction on B and repeat the structural impulse response analysis.
(g) Perform forecast error variance decompositions based on the structural innovations for different identification schemes and comment on the results.

(Hint: See Breitung et al. (2004) for a detailed analysis of the system.)