4 Structural Vector Autoregressive Modelling and Impulse Response Analysis

Jörg Breitung
Universität Bonn

Ralf Brüggemann
Humboldt Universität zu Berlin

Helmut Lütkepohl
European University Institute, Florence, and Humboldt Universität zu Berlin

this version: July 14, 2003

4.1 Introduction

In the previous chapter we have seen how a model for the DGP of a set of economic time series variables can be constructed. When such a model is available, it can be used for analyzing the dynamic interactions between the variables. This kind of analysis is usually done by tracing the effect of an impulse in one of the variables through the system. In other words, an impulse response analysis is performed. Although this is technically straightforward, there are some problems related to impulse response analysis that have been the subject of considerable discussion in the literature.

As argued forcefully by Cooley & LeRoy (1985) vector autoregressions have the status of "reduced form" models and, therefore, are merely vehicles to summarize the dynamic properties of the data. Without referring to a specific economic structure such reduced form VAR models are difficult to understand. For example, it is often difficult to draw any conclusion from the large number of coefficient estimates in a VAR system. As long as such parameters are not related to "deep" structural parameters characterizing preferences, technologies, and optimization behavior, the parameters do not have an economic meaning and are subject to the so-called "Lucas critique".

Sims (1981), Sims (1986), Bernanke (1986) and Shapiro & Watson (1988) put forward a new class of econometric models that is now known as *structural vector autoregression* (SVAR) or *identified VAR*. Instead of identifying the (autoregressive) coefficients, identification focusses on the errors of the

system, which are interpreted as (linear combinations of) exogenous shocks. In the early applications of Sargent (1978) and Sims (1980), the innovations of the VAR were orthogonalized using a Choleski decomposition of the covariance matrix. Thereby a recursive structure is imposed on the instantaneous relations between the variables. Unless there is a justification from subject matter theory for a specific recursive structure, there is some degree of arbitrariness when constructing shocks in this manner. In general, choosing a different ordering of the variables produces different shocks and, thus, the effects of the shocks on the system depend on the way the variables are arranged in the time series vector y_t. To account for this difficulty, Sims (1981) recommends to try out various triangular orthogonalizations and to check whether the results are robust to the ordering of the variables.

An alternative approach to the identification of the shocks is to formulate structural equations for the errors of the system. In this case it is convenient to think of the equations as an IS curve or a money demand relation, for example, where the equations apply to the unexpected part of the variables (the *innovations*) instead of the variables themselves. If the equations are valid for the system variables, then they also apply for the unexpected part of the variables. Therefore, the identification using a set of simultaneous equations is appealing with respect to the traditional approach advocated by the Cowles Commission and it is not surprising that this kind of identification is used widely in empirical work with SVAR models.

In recent work, the identification of shocks using restrictions on their long-run effects has become popular. In many cases economic theory suggests that the effects of some shocks are zero in the long-run, that is, the shocks have transitory effects with respect to particular variables. For example, classical economic theory implies that the effect of nominal shocks on real variables like output or unemployment vanishes as time goes by. Such assumptions give rise to nonlinear restrictions on the parameters and may be used to identify the structure of the system.

An important difference of SVAR models to traditional simultaneous equation models is that the latter usually employ many more restrictions than necessary to identify the system, that is, these models are often highly over-identified. In his famous critique, Sims (1980) qualifies these overly restricted models as "incredible". SVAR proponents therefore try to avoid oversimplifying the structure and impose just enough restrictions needed to identify the parameters. Accordingly, most SVAR models are just identified.

The impulse responses obtained from an SVAR or structural VECM (SVECM) are typically highly nonlinear functions of the model parameters. This property has implications for doing inference related to the impulse responses. It is known that for the functions of interest here, standard asymptotic inference may be a poor guide in small samples. We will discuss the problem and possible solutions in Section 4.5.

The structure of this chapter is as follows. In the next section, the model framework for impulse response analysis will be reviewed briefly. The principles of impulse response analysis for a given SVAR model with identified shocks are considered in Section 4.3, the estimation of the structural parameters is considered in Section 4.4 and statistical inference for impulse responses is treated in Section 4.5. Finally, forecast error variance decomposition as a special way to summarize impulse responses is discussed in Section 4.6. Empirical examples are discussed in some detail in Section 4.7 and conclusions follow in Section 4.8.

4.2 The Models

The most general model we have considered in the previous chapter is a structural VECM form

$$\mathsf{A}\Delta y_t = \Pi^* y_{t-1} + \Gamma_1^* \Delta y_{t-1} + \cdots + \Gamma_{p-1}^* \Delta y_{t-p+1} + C^* D_t + B^* z_t + v_t, \quad (4.1)$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a $(K \times 1)$ vector of endogenous variables, z_t is a vector of exogenous or unmodelled stochastic variables, D_t contains all deterministic terms, the Π^*, Γ_j^* $(j = 1, \ldots, p-1)$, C^* and B^* are structural form parameter matrices and v_t is a $(K \times 1)$ structural form error which is a zero mean white noise process with time invariant covariance matrix Σ_v. The invertible $(K \times K)$ matrix A allows to model instantaneous relations among the variables in y_t.

As already noted, structural shocks are the central quantities in an SVAR model. These shocks are unpredictable with respect to the past of the process and are the input of a linear dynamic system generating the K-dimensional time series vector y_t. They are hence related to the residuals in (4.1). The shocks are associated with an economic meaning such as an oil price shock, exchange rate shock or a monetary shock. Because the shocks are not directly observed, assumptions are needed to identify them. There seems to be a consensus that structural shocks should be mutually uncorrelated (and thus *orthogonal*). This assumption is required to consider the dynamic impact of an isolated shock. If the shocks were correlated, we would have to take into account the relationship between the shocks. Moreover, the decomposition into orthogonal components has a long tradition in statistical analysis, and is also used in factor analysis, for example. The shocks or structural innovations, say ε_t, are assumed to be related to the model residuals by linear relations $v_t = \mathsf{B}\varepsilon_t$, where B is a $(K \times K)$ matrix.

For our present purposes, the deterministic terms are of no importance because they are not affected by impulses hitting the system. Moreover, they do not affect such impulses themselves. Therefore, for notational convenience the deterministic term is often dropped from the model. In practice this may be done by adjusting the variables or the model for deterministic terms before

an analysis of the dynamic interactions between the variables is carried out. Similarly, exogenous variables, if they are under the control of some policy maker, may not react to stochastic shocks of the system and may therefore be ignored for the present purposes. Also, in macroeconometric analysis, exogeneity of a variable is often regarded as too strong a condition and therefore all observable stochastic variables are modelled as endogenous. Instead, the error variables are viewed as the actual exogenous variables, although they are not under the control of any economic agents. Hence, instead of (4.1) we consider

$$\mathsf{A}\Delta y_t = \Pi^* y_{t-1} + \Gamma_1^* \Delta y_{t-1} + \cdots + \Gamma_{p-1}^* \Delta y_{t-p+1} + \mathsf{B}\varepsilon_t. \tag{4.2}$$

Such a model has an equivalent VAR representation for the levels variables of the form

$$\mathsf{A}y_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \mathsf{B}\varepsilon_t. \tag{4.3}$$

It is also often useful for our purposes, especially if the y_t are stationary I(0) variables. Here the A_i^* $(i = 1, \ldots, p)$ are $(K \times K)$ coefficient matrices, as usual.

The reduced forms corresponding to the structural forms (4.2) and (4.3), respectively, are obtained by premultiplying with A^{-1},

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{4.4}$$

and

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{4.5}$$

where $\Pi = \mathsf{A}^{-1}\Pi^*$, $\Gamma_j = \mathsf{A}^{-1}\Gamma_j^*$ $(j = 1, \ldots, p-1)$ and $A_j = \mathsf{A}^{-1}A_j^*$ $(j = 1, \ldots, p)$. Moreover,

$$u_t = \mathsf{A}^{-1}\mathsf{B}\varepsilon_t, \tag{4.6}$$

which relates the reduced form disturbances u_t to the underlying structural shocks ε_t.

To identify the structural form parameters, restrictions on the parameter matrices are necessary. Even if the matrix A which specifies the instantaneous relations between the variables is set to an identity matrix ($\mathsf{A} = I_K$), the assumption of orthogonal shocks ε_t is not sufficient to achieve identification. For a K-dimensional system, $K(K-1)/2$ restrictions are necessary for orthogonalizing the shocks because there are $K(K-1)/2$ potentially different instantaneous covariances. These restrictions can be obtained from a "timing scheme" for the shocks. Such an identification scheme assumes that the shocks may affect a subset of variables directly within the current time period, whereas another subset of variables is affected with a time lag only. An example of such an identification scheme is the triangular (or recursive) identification suggested by Sims (1980). In this model the shocks enter the equations successively so that the additional shock of the second equation does not affect the variable explained by the first equation in the same period. Similarly, the third shock does not affect the variables explained by the

first and second equation in the current time period. Such a scheme is also called a *Wold causal chain system* (Wold (1960)) and is often associated with a causal chain from the first to the last variable in the system. Because the impulse responses computed from these models depend on the ordering of the variables, nonrecursive identification schemes which also allow for instantaneous effects of the variables ($\mathsf{A} \neq I_K$) have been suggested in the literature (see, e.g., Sims (1986) and Bernanke (1986)). Moreover, restrictions on the long-run effects of some shocks are also sometimes used to identify SVAR models (see, e.g., Blanchard & Quah (1989), Galí (1999) and King, Plosser, Stock & Watson (1991)). In empirical applications such restrictions are suggested by economic theory or are imposed just for convenience.

In the following we discuss different types of SVAR models that have been used in applied work. The most popular kinds of restrictions can be classified as follows:

(i) The structural errors are assumed to be mutually uncorrelated such that $\Sigma_\varepsilon = \mathsf{E}(\varepsilon_t \varepsilon_t')$ is a diagonal matrix. In some applications, the shocks are normalized to have unit variances, i.e., $\Sigma_\varepsilon = I_K$. Using $\Sigma_\varepsilon = I_K$ and (4.6) we find

$$\Sigma_u = \mathsf{E}[u_t u_t'] = \mathsf{A}^{-1}\mathsf{B}\,\mathsf{E}[\varepsilon_t \varepsilon_t']\mathsf{B}'\mathsf{A}'^{-1} = \mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1}. \tag{4.7}$$

The symmetry of Σ_u and the orthonormality assumption of the structural shocks impose $K(K+1)/2$ nonlinear restrictions on the elements of A and B. Because these matrices have $2K^2$ elements, these restrictions are not sufficient to identify all structural parameters in A and B.

(ii) $\mathsf{B} = I_K$. The vector of innovations, ε_t, is modelled as an interdependent system of linear equations such that $\mathsf{A}u_t = \varepsilon_t$. Linear restrictions on A can be written in explicit form as $\mathrm{vec}(\mathsf{A}) = R_\mathsf{A}\gamma_\mathsf{A} + r_\mathsf{A}$, where γ_A contains all unrestricted elements of A, R_A is a suitable matrix with 0-1 elements and r_A is a vector of normalizing constants.

(iii) $\mathsf{A} = I_K$. In this case the model for the innovations is $u_t = \mathsf{B}\varepsilon_t$ and to exclude some (linear combinations of the) structural shocks in particular equations, restrictions of the form $\mathrm{vec}(\mathsf{B}) = R_\mathsf{B}\gamma_\mathsf{B} + r_\mathsf{B}$ are imposed, where γ_B contains the unrestricted elements of B and R_B is the corresponding selection matrix with 0-1 elements.

(iv) The so-called AB-model of Amisano & Giannini (1997) combines the restrictions for A and B from (ii) and (iii) such that the model for the innovations is $\mathsf{A}u_t = \mathsf{B}\varepsilon_t$. Accordingly, the two sets of restrictions $\mathrm{vec}(\mathsf{A}) = R_\mathsf{A}\gamma_\mathsf{A} + r_\mathsf{A}$ and $\mathrm{vec}(\mathsf{B}) = R_\mathsf{B}\gamma_\mathsf{B} + r_\mathsf{B}$ are used to identify the system.

(v) There may be prior information on the long-run effects of some shocks. They are measured by considering the responses of the system variables to the shocks. Therefore it is useful to discuss impulse responses and then also consider the long-run effects in more detail. Impulse response analysis is presented in the next section.

As an example consider a recursive identification scheme. In this case, $u_t = \mathsf{B}\varepsilon_t$ with B being lower-triangular and $\mathsf{A} = I_K$. Hence, in addition to the $K(K+1)/2$ restrictions from (4.7) ($\Sigma_u = \mathsf{B}\mathsf{B}'$) and $\mathsf{A} = I_K$ we have $K(K-1)/2$ zero restrictions for B because all the elements above the diagonal are zero. In sum we have $2K^2$ restrictions for A and B.

The simple IS-LM model discussed by Pagan (1995) is another example of an AB-model. Let q_t, i_t and m_t denote output, an interest rate and real money, respectively. The errors of the corresponding reduced form VAR are denoted by $u_t = [u_t^q, u_t^i, u_t^m]'$. A structural model reflecting a traditional Keynesian view is

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

where the structural shocks are assumed to be mutually uncorrelated. The first equation represents a traditional IS curve with a negative parameter for the interest rate innovation u_t^i. The second equation results from solving a Keynesian money demand relationship with respect to interest rate innovations. In other words, the point of departure is a relation $u_t^m = \beta_1 u_t^q + \beta_2 u_t^i + \varepsilon_t^{LM}$, where β_1 is positive because more money is used to finance a larger transactions volume. Moreover, β_2 is negative because higher interest rates lead to lower money holdings and, hence, less demand for money. Accordingly, it is expected that a_{21} is negative while a_{23} is positive. Finally, the third equation postulates that the innovations of the money base are driven by exogenous money supply shocks. Obviously, this model reflects a very stylized view of the economy. A more realistic representation of the economic system would involve further equations and parameter restrictions (see, e.g., Galí (1992)). The present three equations correspond to an AB-model, which can be written as $\mathsf{A}u_t = \mathsf{B}\varepsilon_t$:

$$\begin{bmatrix} 1 & \mathsf{a}_{12} & 0 \\ \mathsf{a}_{21} & 1 & \mathsf{a}_{23} \\ 0 & 0 & 1 \end{bmatrix} u_t = \begin{bmatrix} \mathsf{b}_{11} & 0 & 0 \\ 0 & \mathsf{b}_{22} & 0 \\ 0 & 0 & \mathsf{b}_{33} \end{bmatrix} \varepsilon_t. \tag{4.8}$$

To illustrate how the restrictions implied by this model can be written using the notation defined above, we give the linear restrictions in matrix form:

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

and

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

It can be easily verified that these restrictions are just-identifying by noting that relation (4.7) provides $K(K+1)/2 = 6$ linearly independent restrictions on the parameters in A and B. Moreover, A and B are $(K \times K)$ matrices and without additional restrictions the $2K^2$ parameters could not be uniquely determined from (4.7). In fact, $2K^2 - K(K+1)/2$ additional restrictions need to be imposed for just-identification. In our example, $K = 3$ and consequently we need $2K^2 - K(K+1)/2 = 12$ restrictions on A and B. Counting the restrictions given by (4.8), we find the model to be just-identified. There are 6 restrictions for A (3 zeros and 3 ones) and an additional 6 zero restrictions for B. Given that there are enough identifying assumptions, the parameters of the SVAR model can be estimated by methods discussed in Section 4.4. Within a SVAR model the dynamic effects of structural shocks are typically investigated by an impulse response analysis, which is discussed in the following section.

4.3 Impulse Response Analysis

4.3.1 Stationary VAR Processes

If the process y_t is I(0), the effects of shocks in the variables of a given system are most easily seen in its Wold moving average (MA) representation

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots \,, \tag{4.9}$$

where $\Phi_0 = I_K$ and the

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j} A_j, \quad s = 1, 2, \ldots, \tag{4.10}$$

can be computed recursively as in (**??**) from the reduced form coefficients of the VAR in levels specified in (4.5). The coefficients of this representation may be interpreted as reflecting the responses to impulses hitting the system. The (i,j)th elements of the matrices Φ_s, regarded as a function of s, trace out the expected response of $y_{i,t+s}$ to a unit change in y_{jt} holding constant all

past values of y_t. Since the change in y_{it} given $\{y_{t-1}, y_{t-2}, \ldots\}$ is measured by the innovation u_{it}, the elements of Φ_s represent the impulse responses of the components of y_t with respect to the u_t innovations. In the presently considered I(0) case, $\Phi_s \to 0$ as $s \to \infty$. Hence, the effect of an impulse is transitory as it vanishes over time. These impulse responses are sometimes called *forecast error impulse responses* because the u_t are the 1-step ahead forecast errors.

Occasionally interest centers on the accumulated effects of the impulses. They are easily obtained by adding up the Φ_s matrices. For example, the accumulated effects over all periods, the total long-run effects, are given by

$$\Phi = \sum_{s=0}^{\infty} \Phi_s = (I_K - A_1 - \cdots - A_p)^{-1}. \tag{4.11}$$

This matrix exists if the VAR process is stable (see the stability condition in (**??**)).

A critique that has been raised against forecast error impulse responses is that the underlying shocks are not likely to occur in isolation if the components of u_t are instantaneously correlated, that is, if Σ_u is not diagonal. Therefore, orthogonal innovations are preferred in an impulse response analysis, as mentioned previously. One way to get them is to use a Choleski decomposition of the covariance matrix Σ_u. If B is a lower-triangular matrix such that $\Sigma_u = \mathsf{B}\mathsf{B}'$, the orthogonalized shocks are given by $\varepsilon_t = \mathsf{B}^{-1}u_t$. Hence, we get from (4.9),

$$y_t = \Psi_0\varepsilon_t + \Psi_1\varepsilon_{t-1} + \cdots , \tag{4.12}$$

where $\Psi_i = \Phi_i\mathsf{B}$ $(i = 0, 1, 2, \ldots)$. Here $\Psi_0 = \mathsf{B}$ is lower triangular so that an ε shock in the first variable may have an instantaneous effect on all the variables, whereas a shock in the second variable cannot have an instantaneous impact on y_{1t} but only on the other variables and so on. Thus, we have a Wold causal chain. Given that the ε shocks are instantaneously uncorrelated (orthogonal), the corresponding impulse responses are often referred to as *orthogonalized impulse responses.*

Because many matrices B exist which satisfy $\mathsf{B}\mathsf{B}' = \Sigma_u$, using a Choleski decomposition approach is to some extent arbitrary, unless there are good reasons for a particular recursive structure specified by a given B. As mentioned earlier, if B is found by a lower triangular Choleski decomposition, choosing a different ordering of the variables in the vector y_t may produce different shocks. Hence, the effects of a shock may depend on the way the variables are arranged in the vector y_t. In view of this difficulty, Sims (1981) recommends trying various triangular orthogonalizations and checking the robustness of the results with respect to the ordering of the variables if no particular ordering is suggested by subject matter theory. Using information based on the latter results in SVAR models, of course.

As discussed earlier, in an SVAR such as (4.3), the residuals are represented as $\mathsf{B}\varepsilon_t$ and ε_t is a $(K \times 1)$ vector of structural shocks with (diagonal) covariance matrix $\mathsf{E}(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon$ which is often specified to be an identity matrix. In any case, the structural shocks are instantaneously uncorrelated. In the AB-model the relation to the reduced form residuals is given by $\mathsf{A}u_t = \mathsf{B}\varepsilon_t$. Therefore, the impulse responses in a general SVAR model may be obtained from (4.12) with $\Psi_j = \Phi_j\mathsf{A}^{-1}\mathsf{B}$. If restrictions on the long-run effects are available, they may be placed on $\Psi = \Phi\mathsf{A}^{-1}\mathsf{B}$, where Φ is the matrix specified in (4.11). For example, one may want to impose the restriction that some shocks do not have any long-run effects. This is achieved by setting the respective elements of the long-run impact matrix $\Psi = \Psi_0 + \Psi_1 + \cdots$ equal to zero.

As an example we consider the model suggested by Blanchard & Quah (1989). Based on a simple economic model, Blanchard and Quah identify supply shocks to have persistent effects on output whereas demand shocks are transitory. Suppose in a VAR model for $y_t = (\Delta Q_t, U_t)'$, where Q_t denotes the log of output and U_t is the unemployment rate, we wish to identify innovations that can be interpreted as demand shocks and supply shocks, $\varepsilon_t = (\varepsilon_t^s, \varepsilon_t^d)'$. Because we have $K = 2$ variables and we can specify $\mathsf{A} = I_2$, we need $K(K-1) = 1$ restriction to identify the structural shocks from the VAR residuals. The effects of these shocks on the output growth rates ΔQ_t and the unemployment U_t are obtained from the Ψ_i matrices. Accumulating them gives the effects on Q_t and the stock of unemployment. Thus, the restriction that demand shocks have no long-run impact on output can be imposed by constraining the (1,2)-element of the matrix $\Psi = \sum_{i=0}^{\infty}\Psi_i$ to be equal to zero. Estimation of models with long-run restrictions is discussed in Section 4.4.

4.3.2 Impulse Response Analysis of Nonstationary VARs and VECMs

Although the Wold representation does not exist for nonstationary cointegrated processes it is easy to see from Section **??** that the Φ_s impulse response matrices can be computed in the same way as in (4.10) based on VARs with integrated variables or the levels version of a VECM (Lütkepohl (1991, Chapter 11), Lütkepohl & Reimers (1992)). In this case, the Φ_s may not converge to zero as $s \to \infty$ and, consequently, some shocks may have permanent effects. Of course, one may also consider orthogonalized or accumulated impulse responses. Because the Φ_s and Ψ_s may not approach zero as $s \to \infty$, the total accumulated impulse responses will generally not exist, however. Recall that for cointegrated systems the matrix $(I_K - A_1 - \cdots - A_p)$ is singular. From Johansen's version of Granger's Representation Theorem (see Johansen (1995)) it is known, however, that if y_t is generated by a reduced form VECM $\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t$, it has the following MA representation:

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L)u_t + y_0^*,$$

where $\Xi = \beta_\perp \left(\alpha_\perp'(I_K - \sum_{i=1}^{p-1} \Gamma_i)\beta_\perp\right)^{-1} \alpha_\perp'$, $\Xi^*(L) = \sum_{j=0}^{\infty} \Xi_j^* L^j$ is an infinite order polynomial in the lag operator with coefficient matrices Ξ_j^* that go to zero as $j \to \infty$. The term y_0^* contains all initial values. Notice that Ξ has rank $K-r$ if the cointegrating rank of the system is r. It represents the long-run effects of forecast error impulse responses, whereas the Ξ_i contain transitory effects.

Because the forecast error impulse responses based on Ξ and the Ξ_j^* are subject to the same criticism as for stable VAR processes, appropriate shocks have to be identified for a meaningful impulse response analysis. Replacing u_t by $\mathsf{A}^{-1}\mathsf{B}\varepsilon_t$ the orthogonalized 'short-run' impulse responses may be obtained as $\Xi_j^*\mathsf{A}^{-1}\mathsf{B}$ analogously to the stationary VAR case. Moreover, the long-run effects of ε shocks are given by

$$\Xi\mathsf{A}^{-1}\mathsf{B}. \tag{4.13}$$

Because the rank of this matrix is $K-r$, there must be r shocks with transitory effects (zero long-run impact) and $k^* = K-r$ shocks with permanent effects. Thus, we will have r columns of zeros in the matrix (4.13) corresponding to the transitory shocks. Due to the reduced rank of the matrix, these zeros represent k^*r independent restrictions only. To exactly identify the permanent shocks we need $k^*(k^*-1)/2$ additional restrictions. Similarly, $r(r-1)/2$ additional contemporaneous restrictions identify the transitory shocks (see, e.g., King et al. (1991)). Together these are a total of $k^*r + k^*(k^*-1)/2 + r(r-1)/2 = K(K-1)/2$ restrictions. Hence, assuming $\mathsf{A} = I_K$ and taking into account the $K(K+1)/2$ restrictions from $\mathsf{B}\mathsf{B}' = \Sigma_u$, we have just enough restrictions to identify B. In addition, restrictions for A may be available if there is information on instantaneous relations between the variables. In any case, long-run restrictions imply in general nonlinear restrictions on A and/or B.

An example Long-run identifying assumptions for the ε_t shocks are typically derived from economic theory. To illustrate this point we briefly describe a simple macroeconomic model of the labour market that is used by Jacobson, Vredin & Warne (1997) to investigate the effects of shocks to Scandinavian unemployment. This model consists of a production function, a labour demand relation, a labour supply, and a wage setting relation. All variables are expressed in natural logarithms. The production function relates output gdp_t to employment e_t

$$gdp_t = \rho e_t + \theta_{1,t}, \tag{4.14}$$

where ρ measures the returns to scale. The quantity $\theta_{1,t}$ is a stochastic technology trend that follows a random walk,

$$\theta_{1,t} = \theta_{1,t-1} + \varepsilon_t^{gdp},$$

and ε_t^{gdp} is the pure technology shock. Labour demand relates employment to output and real wages $(w-p)_t$:

$$e_t = \lambda gdp_t - \eta(w-p)_t + \theta_{2,t}, \tag{4.15}$$

with an error process

$$\theta_{2,t} = \phi_d \theta_{2,t-1} + \varepsilon_t^d.$$

If $|\phi_d| < 1$ the labour demand is stationary. In that case the pure labour demand innovation ε_t^d has only temporary effects on employment. Jacobson et al. (1997) assume $\phi_d = 0$ a priori, which implies that the labour demand shock has no long-run effects. Within a cointegration analysis, stationarity of labour demand can be tested and hence the a priori assumption is not needed here. In the third equation of the model the labour force l_t is related to real wages according to

$$l_t = \pi(w-p)_t + \theta_{3,t}. \tag{4.16}$$

The exogenous labour supply trend $\theta_{3,t}$ follows a random walk

$$\theta_{3,t} = \theta_{3,t-1} + \varepsilon_t^s,$$

where ε_t^s is the underlying labour supply shock. Finally, we have the wage setting relation

$$(w-p)_t = \delta(gdp_t - e_t) - \gamma(l_t - e_t) + \theta_{4,t}, \tag{4.17}$$

stating that real wages are a function of productivity $(gdp_t - e_t)$ and unemployment $(l_t - e_t)$. The wage setting trend $\theta_{4,t}$ can be stationary or nonstationary depending on ϕ_w in

$$\theta_{4,t} = \phi_w \theta_{4,t-1} + \varepsilon_t^w.$$

If $|\phi_w| < 1$, the wage setting trend is stationary. Again, results from empirical analysis can be used to determine whether wage setting is stationary.

Under standard assumption for the shocks $\varepsilon^{gdp}, \varepsilon^d, \varepsilon^s$, and ε^w the solution of the model (4.14) – (4.17) in terms of the variables used in the empirical analysis is given by

$$\begin{aligned}
\begin{bmatrix} gdp_t - e_t \\ e_t \\ l_t - e_t \\ (w-p)_t \end{bmatrix} &= \psi \begin{bmatrix} (1-\lambda)(1+\gamma\pi) + \eta\gamma \\ \lambda(1+\gamma\pi) - \eta\delta \\ \eta\delta - \lambda + (1-\lambda)\pi\delta \\ \lambda\gamma + \delta(1-\lambda) \end{bmatrix} \theta_{1,t} + \psi \begin{bmatrix} (\rho-1)(1+\gamma\pi) \\ 1+\gamma\pi \\ (\rho-1)\delta\pi - 1 \\ \gamma - \delta(1-\rho) \end{bmatrix} \theta_{2,t} \\
&\quad + \psi \begin{bmatrix} (\rho-1)\eta\gamma \\ \eta\gamma \\ 1 - \rho\lambda + (\rho-1)\delta\eta \\ (\rho\lambda - 1)\gamma \end{bmatrix} \theta_{3,t} + \psi \begin{bmatrix} \eta(1-\rho) \\ -\eta \\ \eta + (1-\rho\lambda)\pi \\ 1 - \rho\lambda \end{bmatrix} \theta_{4,t}
\end{aligned} \tag{4.18}$$

with

$$\psi = \frac{1}{(1-\rho\lambda)(1+\gamma\pi)+\eta\gamma+(\rho-1)\eta\delta}.$$

From this solution it is obvious that productivity, employment, unemployment and real wages are driven by two random walks in productivity and labour supply. As explained earlier, the labour demand and the wage setting component can be stationary or nonstationary. In terms of the common trends literature, there are at least two and at most four common trends in this model. This implies at most two cointegration relations: a labour demand relation and a wage setting relation. The model together with results from a cointegration analysis implies a set of identifying assumptions for the structural VECM. Suppose, for example, that two cointegration relations ($r = 2$) have been found in a four dimensional VAR ($K = 4$) for productivity, employment, unemployment and real wages. Consequently, only $k^* = K - r = 2$ shocks may have permanent effects. We associated the technology, the labour demand, the labour supply and the wage setting shocks with the equations for productivity, employment, unemployment and real wages, respectively, such that $\varepsilon_t = (\varepsilon_t^{gdp}, \varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^w)'$.

For stationarity of labour demand and wage setting our model implies that labour demand and wage setting shocks have no long-run impact on the system variables and, hence, the second and fourth columns of the long-run impact matrix $\Xi\mathsf{A}^{-1}\mathsf{B}$ are zero. To identify the two permanent shocks we have to impose $k^*(k^*-1)/2 = 1$ additional restriction. Assuming constant returns to scale ($\rho = 1$) implies that productivity is only driven by productivity shocks ε^{gdp} in the long-run (see (4.18)). Thus, assuming $\mathsf{A} = I_K$, these sets of restrictions can be expressed as follows in terms of the long-run impact matrix:

$$\Xi\mathsf{B} = \begin{pmatrix} * & 0 & 0 & 0 \\ * & 0 & * & 0 \\ * & 0 & * & 0 \\ * & 0 & * & 0 \end{pmatrix}. \tag{4.19}$$

Here unrestricted elements are indicated by an asterisk. Note that due to the reduced rank of Ξ, we cannot simply count the zeros in (4.19) in order to determine the number of restrictions imposed on the model. As explained earlier, the two zero columns represent $k^*r = 4$ linearly independent retrictions only. Hence, the zeros in (4.19) stand for only 5 linearly independent restrictions. In addition, we need $r(r-1)/2 = 1$ contemporaneous restriction to disentangle the effects of the two transitory shocks. For instance, we may choose the restriction that labour demand shocks do not affect real wages on impact, i.e., we set $\mathsf{B}_{42} = 0$. In our example, a typical problem within the SVAR modeling class arises: The theoretical model does not suggest contemporaneous restrictions and thus, defending this type of restriction may be difficult. In practice, a sensitivity analysis with respect to different contemporaneous identifying assumptions can be useful. We will employ the presented

theoretical model to derive identifying assumptions in a structural VECM for Canadian labour market data in Section 4.7.3.

To compute the impulse responses we need not only the reduced form parameters but we also need the structural parameters. How to estimate them will be discussed in the next section before we consider inference for impulse responses in Section 4.5.

4.4 Estimation of Structural Parameters

4.4.1 SVAR Models

The estimation of the SVAR model is equivalent to the problem of estimating a simultaneous equation model with covariance restrictions. First consider a model without restrictions on the long-run effects of the shocks. It is assumed that ε_t is white noise with $\varepsilon_t \sim N(0, I_K)$ and the basic model is a VAR(p) so that the structural form is

$$\mathsf{A}y_t = \mathsf{A}[A_1, \ldots, A_p]Y_{t-1} + \mathsf{B}\varepsilon_t, \tag{4.20}$$

where $Y'_{t-1} = [y'_{t-1}, \ldots, y'_{t-p}]$. Defining $A = [A_1, \ldots, A_p]$ the corresponding reduced form is $y_t = AY_{t-1} + \mathsf{A}^{-1}\mathsf{B}\varepsilon_t$. Notice that the error term $u_t = \mathsf{A}^{-1}\mathsf{B}\varepsilon_t$. Hence, for a sample $y_1, \ldots, y_T$, using the notation $Y = [y_1, \ldots, y_T]$ and $Z = [Y_0, \ldots, Y_{T-1}]$, the log likelihood function is

$$\begin{aligned}\mathcal{L}(A, \mathsf{A}, \mathsf{B}) &= -\tfrac{KT}{2}\log 2\pi - \tfrac{T}{2}\log|\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1}| \\ &\quad -\tfrac{T}{2}\mathrm{tr}\{(Y - AZ)'[\mathsf{A}^{-1}\mathsf{B}\mathsf{B}'\mathsf{A}'^{-1}]^{-1}(Y - AZ)\} \\ &= \text{constant} + \tfrac{T}{2}\log|\mathsf{A}|^2 - \tfrac{T}{2}\log|\mathsf{B}|^2 \\ &\quad -\tfrac{T}{2}\mathrm{tr}\{\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}(Y - AZ)(Y - AZ)'\}.\end{aligned} \tag{4.21}$$

Here we have used 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 $\mathrm{tr}(VW) = \mathrm{tr}(WV)$ for matrices V and W with suitable dimensions. If there are no restrictions on A, we know from Chapter **??**, Section **??**, that for any given A and B the function $\mathcal{L}(\mathsf{A}, \mathsf{B}, A)$ is maximized with respect to A by $\hat{A} = YZ'(ZZ')^{-1}$. Substituting this expression in (4.21) gives the concentrated log likelihood

$$\mathcal{L}_C(\mathsf{A}, \mathsf{B}) = \text{constant} + \frac{T}{2}\log|\mathsf{A}|^2 - \frac{T}{2}\log|\mathsf{B}|^2 - \frac{T}{2}\mathrm{tr}(\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}\widetilde{\Sigma}_u), \tag{4.22}$$

where $\widetilde{\Sigma}_u = T^{-1}(Y - \hat{A}Z)(Y - \hat{A}Z)'$ is just the estimated covariance matrix of the VAR residuals (cf. Breitung (2001)**?**).

It is easy to extend this approach to the case where deterministic terms are present in the original model. Moreover, the same concentrated likelihood

function is obtained if there are restrictions for A, for example, if a subset model is fitted. In this case, theoretically we just have to replace $\widetilde{\Sigma}_u$ by the residual covariance matrix of the restricted ML estimator of the reduced form. For practical purposes, the exact ML estimator is usually not used, however, because its computation requires the use of nonlinear optimization methods in general. Instead, the feasible GLS estimator may be used which has the same asymptotic properties as the ML estimator. In that case, (4.22) is not strictly the concentrated log likelihood, however. For simplicity we still refer to this function as the concentrated log likelihood.

The function (4.22) is maximized with respect to A and B subject to the restrictions resulting from the structural form of the system, which for the AB-model can be written compactly as

$$\begin{bmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{bmatrix} = \begin{bmatrix} R_{\mathsf{A}} & 0 \\ 0 & R_{\mathsf{B}} \end{bmatrix} \begin{bmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{bmatrix} + \begin{bmatrix} r_{\mathsf{A}} \\ r_{\mathsf{B}} \end{bmatrix}. \tag{4.23}$$

Because no closed form solution is available, this has to be done by employing numerical optimization methods. For this purpose, Amisano & Giannini (1997) suggest to use a scoring algorithm. More precisely, estimates for γ_{A} and γ_{B} are found by iterating on

$$\begin{bmatrix} \tilde{\gamma}_{\mathsf{A}} \\ \tilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i+1} = \begin{bmatrix} \tilde{\gamma}_{\mathsf{A}} \\ \tilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i} + \ell \, \mathcal{I} \left(\begin{bmatrix} \tilde{\gamma}_{\mathsf{A}} \\ \tilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i} \right)^{-1} \mathcal{S} \left(\begin{bmatrix} \tilde{\gamma}_{\mathsf{A}} \\ \tilde{\gamma}_{\mathsf{B}} \end{bmatrix}_{i} \right),$$

where ℓ is the step length, $\mathcal{I}(\cdot)$ denotes the information matrix of the free parameters γ_{A}, γ_{B}, $\mathcal{S}(\cdot)$ is the score vector and the subscript refers to the iteration number where the signified estimator is obtained. Using

$$\begin{aligned} &\mathcal{I} \left(\begin{bmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{bmatrix} \right) \\ &\quad = T \begin{bmatrix} [\mathsf{B}^{-1}\mathsf{A}]^{-1} \otimes \mathsf{B}'^{-1} \\ -(I_K \otimes \mathsf{B}'^{-1}) \end{bmatrix} (I_{K^2} + \mathbf{K}_{KK}) \Big[([\mathsf{B}^{-1}\mathsf{A}]'^{-1} \otimes \mathsf{B}^{-1}) \vdots -(I_K \otimes \mathsf{B}^{-1}) \Big] \end{aligned}$$

with $\mathbf{K}_{KK}$ being a $(K^2 \times K^2)$ commutation matrix defined such that for any $(K \times K)$ matrix M, $\operatorname{vec}(M') = \mathbf{K}_{KK} \operatorname{vec}(M)$, the information matrix for the free parameters can be shown to be

$$\mathcal{I} \left(\begin{bmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{bmatrix} \right) = \begin{bmatrix} R'_{\mathsf{A}} & 0 \\ 0 & R'_{\mathsf{B}} \end{bmatrix} \mathcal{I} \left(\begin{bmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{bmatrix} \right) \begin{bmatrix} R_{\mathsf{A}} & 0 \\ 0 & R_{\mathsf{B}} \end{bmatrix}.$$

The score vector for the free parameters is given by

$$\mathcal{S} \left(\begin{bmatrix} \gamma_{\mathsf{A}} \\ \gamma_{\mathsf{B}} \end{bmatrix} \right) = \begin{bmatrix} R'_{\mathsf{A}} & 0 \\ 0 & R'_{\mathsf{B}} \end{bmatrix} \mathcal{S} \left(\begin{bmatrix} \operatorname{vec} \mathsf{A} \\ \operatorname{vec} \mathsf{B} \end{bmatrix} \right),$$

where we have used

$$\mathcal{S}\left(\begin{bmatrix}\text{vec } \mathsf{A} \\ \text{vec } \mathsf{B}\end{bmatrix}\right) = \begin{bmatrix}(I_K \otimes \mathsf{B}'^{-1}) \\ -(\mathsf{B}^{-1}\mathsf{A} \otimes \mathsf{B}'^{-1})\end{bmatrix} \mathcal{S}(\text{vec}[\mathsf{B}^{-1}\mathsf{A}])$$

and

$$\mathcal{S}(\text{vec}[\mathsf{B}^{-1}\mathsf{A}]) = T\text{vec}([\mathsf{B}^{-1}\mathsf{A}]'^{-1}) - T(\widetilde{\Sigma}_u \otimes I_K)\text{vec}(\mathsf{B}^{-1}\mathsf{A}).$$

Although straightforward in theory, sometimes the optimization turns out to be difficult in practice and the choice of appropriate starting values is crucial.

Identification of the SVAR models is often checked numerically using the starting values for γ_{A} and γ_{B}. As noted by Christiano, Eichenbaum & Evans (1999), in some cases the signs of the elements in A and B are not identified even if the usual order criterion for identification is met. They suggest to normalize the sign such that the contemporaneous impact matrix $\mathsf{A}^{-1}\mathsf{B}$ has positive diagonal elements.

Iteration of the algorithm stops when some prespecified criterion, such as the relative change in the log-likelihood and the relative change in the parameters, is met. The resulting ML estimator is asymptotically efficient and normally distributed, where the asymptotic covariance matrix is estimated by the inverse of the information matrix. Moreover, a ML estimator for Σ_u is given by

$$\widetilde{\Sigma}_u^* = \tilde{\mathsf{A}}^{-1}\tilde{\mathsf{B}}\tilde{\mathsf{B}}'\tilde{\mathsf{A}}'^{-1}, \tag{4.24}$$

where $\tilde{\mathsf{A}}$ and $\tilde{\mathsf{B}}$ are estimators of A and B, respectively. Note that $\widetilde{\Sigma}_u^*$ only corresponds to the reduced form estimate $\widetilde{\Sigma}_u$ if the SVAR is exactly identified. In the presence of over-identifying restrictions, a LR test statistic for these restrictions can be constructed in the usual way as

$$LR = T(\log|\widetilde{\Sigma}_u^*| - \log|\widetilde{\Sigma}_u|). \tag{4.25}$$

This statistic has an asymptotic χ^2-distribution with degrees of freedom equal to the number of over-identifying restrictions. In the AB-model this is the number of constraints imposed on A and B minus $2K^2 - K(K+1)/2$.

At times a priori information on the effects of structural shocks is available that specifies the structural long-run impact matrix Ψ to be lower-triangular. Examples in the literature include Blanchard & Quah (1989) and Galí (1999). As discussed in Section 4.3, in a stationary VAR model these restrictions can be imposed by restricting the elements above the main diagonal of Ψ to zero. Estimation is particularly easy in this case. For the computation of the structurally identified impulse responses it is sufficient to estimate the contemporaneous impact matrix $\mathsf{C} = \mathsf{A}^{-1}\mathsf{B}$. We note that the long-run impact matrix of structural shocks Ψ can be related to Φ by the contemporaneous impact matrix C,

$$\Psi = \Phi\mathsf{C},$$

such that

$$\Psi\Psi' = \Phi\Sigma_u\Phi' = (I_K - A_1 - \cdots - A_p)^{-1}\Sigma_u(I_K - A_1' - \cdots - A_p')^{-1}.$$

Since Ψ is assumed to be lower triangular, it can be obtained from a Choleski decomposition of the matrix

$$(I_K - A_1 - \cdots - A_p)^{-1} \Sigma_u (I_K - A_1' - \cdots - A_p')^{-1}.$$

Replacing the unknown quantities with estimates, C can be estimated by

$$\hat{\mathsf{C}} = \hat{\Phi}^{-1}\hat{\Psi} = \hat{\Phi}^{-1}\text{chol}[\hat{\Phi}\hat{\Sigma}_u\hat{\Phi}'].$$

It must be emphasized that this estimation procedure works only in stationary VAR models, because $(I_K - A_1 - \cdots - A_p)^{-1}$ does not exist otherwise. Moreover, it can only be used to estimate models that are exactly identified. Although this case can be seen as a fairly special model, it is sometimes useful to have a closed form solution for estimating the structural parameters. Section 4.7 illustrates the use of different estimation techniques.

Generally, if not otherwise mentioned, the procedures described in this subsection are also applicable if some or all of the variables have unit roots and a levels VAR model is fitted without taking possible cointegration restrictions into account.

4.4.2 Structural VECMs

For VECMs the concentrated likelihood function (4.22) can also be used for estimating the structural parameters A and B. If no restrictions are imposed on the short-run parameters, the $\widetilde{\Sigma}_u$ matrix represents the residual covariance matrix obtained from a reduced rank regression as described in Chapter **??**, Section **??**. If the short-run parameters are restricted or restrictions are placed on the cointegration vectors, some other estimator may be used instead of the ML estimator and Σ_u may be estimated from the corresponding residuals.

Generally, if long-run identifying restrictions have to be considered, maximization of (4.22) is a numerically difficult task because these restrictions are typically highly nonlinear for A and/or B. In some cases, however, it is possible to express these long-run restrictions as linear restrictions and maximization can be done using the scoring algorithm defined above. In particular, consider a cointegrated VECM where $\mathsf{A} = I_K$. Then restrictions that some shocks have no long-run impact on the system variables can be written in implicit form as

$$R_\Xi \text{vec}(\Xi\mathsf{B}) = 0,$$

where R_Ξ is an appropriate restriction matrix. Following Vlaar (1998) in using the rules of the vec operator these restrictions can be reformulated as

$$R_\Xi (I_K \otimes \Xi)\text{vec}(\mathsf{B}) = R_{\mathsf{B},l}\text{vec}(\mathsf{B}) = 0.$$

Replacing Ξ by an estimator obtained from the reduced form, $R_{\mathsf{B},l} = R_\Xi(I_K \otimes \Xi)$ is a stochastic restriction matrix. These implicit restrictions can

be translated into the explicit form used in the maximization procedure of the SVECM and, moreover, can be combined with contemporaneous restrictions on the elements of B in the form $\text{vec}(\mathsf{B}) = R_{\mathsf{B}}\gamma_{\mathsf{B}}$. It is also worth noting that this method also works if there are more restrictions on the structural parameters than necessary for exact identification. In other words, it may be applied for over-identified models and the validity of these restrictions may be tested using the LR test statistic given in (4.25).

4.5 Statistical Inference for Impulse Responses

4.5.1 Asymptotic Estimation Theory

If an estimator $\hat{\theta}$, say, of the SVAR or SVECM coefficients summarized in the vector θ is available, estimators of the impulse responses may be obtained as functions of $\hat{\theta}$. Formally, we write for some arbitrary impulse response coefficient $\phi = \phi(\theta)$ so that

$$\hat{\phi} = \phi(\hat{\theta}). \tag{4.26}$$

Assuming that $\hat{\theta}$ has an asymptotic normal distribution,

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_{\hat{\theta}}), \tag{4.27}$$

the ϕ are also asymptotically normally distributed,

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2_{\hat{\phi}}), \tag{4.28}$$

where

$$\sigma^2_{\hat{\phi}} = \frac{\partial \phi}{\partial \theta'} \Sigma_{\hat{\theta}} \frac{\partial \phi}{\partial \theta} \tag{4.29}$$

and $\partial\phi/\partial\theta$ denotes the vector of first order partial derivatives of ϕ with respect to the elements of θ. The limiting result in (4.28) holds if $\sigma^2_{\hat{\phi}}$ is nonzero which in turn is guaranteed if $\Sigma_{\hat{\theta}}$ is nonsingular and $\partial\phi/\partial\theta \neq 0$. Note that the covariance matrix $\Sigma_{\hat{\theta}}$ may be singular if there are constraints on the coefficients or, as mentioned in Chapter **??**, if there are I(1) variables. The partial derivatives will also usually be zero in some points of the parameter space because the ϕ generally consist of sums of products of the VAR coefficients and, hence, the partial derivatives will also be sums of products of such coefficients which may be zero. Nonzero partial derivatives are guaranteed if all elements of θ are nonzero. In other words, fitting subset VAR models where all those coefficients are restricted to zero which are actually zero, helps to make the asymptotics work. Of course, in practice it is usually unknown which coefficients are zero. Therefore fitting subset models as described in Chapter **??**, Section **??**, may be a good idea.

4.5.2 Bootstrapping Impulse Responses

In applied work, statistical inference regarding impulse responses is often based on bootstrap methods. In particular, they are often used to construct confidence intervals (CIs) for impulse responses because thereby occasionally more reliable small sample inference is possible than by using asymptotic theory (e.g., Kilian (1998)). Moreover, the analytical expressions of the asymptotic variances of the impulse response coefficients are rather complicated (e.g., Lütkepohl (1991, Chapter 3)). Using the bootstrap for setting up CIs, the precise expressions of the variances are not needed and, hence, deriving the analytical expressions explicitly can be avoided.

Alternative bootstrap approaches have been proposed for setting up CIs for impulse responses. They use residual based bootstraps which proceed as follows: First the model of interest is estimated. Denoting the estimation residuals by $\hat{u}_t$, centered residuals, $\hat{u}_1 - \bar{\hat{u}}, \ldots, \hat{u}_T - \bar{\hat{u}}$, are computed and bootstrap residuals, $u_1^*, \ldots, u_T^*$, are generated by randomly drawing with replacement from the centered residuals. The u_t^* are used to compute bootstrap time series recursively starting from given presample values $y_{-p+1}, \ldots, y_0$ for a model with p lags. The model is reestimated and the quantities of interest are determined on the basis of the parameter estimates obtained in this way. Repeating these steps a large number of times gives the empirical bootstrap distribution of the quantities of interest. From that distribution quantiles and, hence, CIs may be obtained for the impulse responses.

In the following the symbols ϕ, $\hat{\phi}$ and $\hat{\phi}^*$ denote some general impulse response coefficient, its estimator implied by the estimators of the model coefficients and the corresponding bootstrap estimator, respectively. The following bootstrap CIs have, for instance, been considered in the literature (see, e.g., Benkwitz, Lütkepohl & Wolters (2001)):

– *Standard percentile interval*
  The most common method in setting up CIs for impulse responses in practice is to use the interval

$$CI_S = \left[s^*_{\gamma/2}, s^*_{(1-\gamma/2)}\right],$$

  where $s^*_{\gamma/2}$ and $s^*_{(1-\gamma/2)}$ are the $\gamma/2$- and $(1-\gamma/2)$-quantiles, respectively, of the empirical distribution of the $\hat{\phi}^*$. The interval CI_S is the percentile confidence interval described, e.g., by Efron & Tibshirani (1993).
– *Hall's percentile interval*
  Hall (1992) uses the usual bootstrap analogy which states that the distribution of $(\hat{\phi}-\phi)$ is approximately equal to that of $(\hat{\phi}^*-\hat{\phi})$ in large samples. Using this result, the interval

$$CI_H = \left[\hat{\phi} - t^*_{(1-\gamma/2)}, \hat{\phi} - t^*_{\gamma/2}\right]$$

  can be derived. Here $t^*_{\gamma/2}$ and $t^*_{(1-\gamma/2)}$ are the $\gamma/2$- and $(1-\gamma/2)$-quantiles, respectively, of the empirical distribution of $(\hat{\phi}^* - \hat{\phi})$.

– *Hall's studentized interval*
In some situations, using a studentized statistic $(\hat{\phi}-\phi)/(\widehat{\mathrm{var}}(\hat{\phi}))^{1/2}$ for constructing confidence intervals may be advantageous. In that case, bootstrap quantiles $t^{**}_{\gamma/2}$ and $t^{**}_{1-\gamma/2}$ from the distribution of $(\hat{\phi}^*-\hat{\phi})/(\widehat{\mathrm{var}}(\hat{\phi}^*))^{1/2}$ are used to construct an interval

$$CI_{SH} = \left[\hat{\phi} - t^{**}_{(1-\gamma/2)}(\widehat{\mathrm{var}}(\hat{\phi}))^{1/2},\ \hat{\phi} - t^{**}_{\gamma/2}(\widehat{\mathrm{var}}(\hat{\phi}))^{1/2}\right].$$

In this approach the variances are estimated by a bootstrap within each bootstrap replication.

A number of refinements and modifications of these intervals exist. Unfortunately, the bootstrap does not necessarily overcome the problems due to a singularity in the asymptotic distribution which results from a zero variance in (4.28). In other words, in these cases bootstrap CIs may not have the desired coverage even asymptotically. For a critical discussion see Benkwitz, Lütkepohl & Neumann (2000).

There are at least three possible strategies to overcome the problems resulting from the different rates of convergence in the parameter space. First, one may consider bootstrap procedures that adjust to the kind of singularity in the asymptotic distribution. A couple of different proposals of this sort are discussed in Benkwitz et al. (2000). For instance, subsampling may be used to estimate the convergence rate of the parameter estimators in addition to the model parameters. These and other methods were shown to have drawbacks, however, in empirical applications. Either they are not very practical for processes of realistic dimension and autoregressive order or they do not perform well in samples of typical size.

A second possibility to tackle the singularity problem is to single out and eliminate the points where problems occur before an impulse response analysis is carried out. In the present case this proposal amounts to determining all zero coefficients in a first stage of the analysis and enforcing the resulting zero restrictions in the next stage, where the resulting subset model is estimated and used for computing impulse responses. This solution is, for instance, considered by Benkwitz et al. (2001). Problematic in this approach may be the uncertainty with respect to the actual zero restrictions.

A third way out of the singularity problem is to consider a different type of modelling approach based on the assumption of a potentially infinite VAR order. In this approach it is assumed that the model order is increased when more sample information becomes available. In other words, the model order is assumed to approach infinity with the sample size at a suitable rate. An asymptotic theory is developed by Lütkepohl (1988, 1996), Lütkepohl & Poskitt (1991, 1996), Lütkepohl & Saikkonen (1997) and Saikkonen & Lütkepohl (1996, 2000) based on work by Lewis & Reinsel (1985) and Saikkonen (1992). It turns out that this asymptotic theory avoids the kind of singularity in the asymptotic distribution which causes the failure of the bootstrap procedures. On the other hand, the greater generality of the model results in

an inefficiency relative to the model with finite fixed order. In practice, using subset models may be the preferred solution.

4.5.3 An Illustration

To illustrate the use of an impulse response analysis for studying the relations between the variables of a model, we consider again the German interest rate-inflation system of Chapter **??**. We use the subset VECM (**??**) to account for the problems related to constructing bootstrap confidence intervals for the impulse responses. Because no instantaneous residual correlation was found in testing for instantaneous causality, it may be reasonable for this system to consider the forecast error impulse responses. We plot them in Figure 4.1 together with bootstrap confidence intervals obtained with all three methods mentioned in the foregoing. The standard and Hall confidence intervals are based on 2000 bootstrap replications and the studentized Hall intervals are based in addition on 50 replications for estimating the variances $\widehat{\text{var}}(\hat{\phi}^*)$ in each of the outer replication rounds. As a consequence, the latter CIs are computationally rather demanding. The bootstrap literature suggests that the number of bootstrap replications has to be quite large in order to obtain reliable results. In the present case it is therefore a question whether the chosen number of 2000 replications is adequate. One way to check this would be to simulate CIs with increasing numbers of replications and determine whether an increase of the number of replications leads to changes in the CIs. For the present example, similar results are also obtained with smaller numbers of replications, say 1000. Therefore, using 2000 replications may be sufficient.

In this particular case all three methods for computing CIs produce qualitatively similar results. Both variables react permanently to a one-time impulse in their own residuals. Also they both have an impact on each other. Clearly, this result is not surprising from the perspective of economic theory which suggests that the nominal interest rate is partly determined by the inflation rate. Moreover, interest rate changes should impact on economic activity and thereby eventually also have an effect on the development of the price level. Although these relations are reflected in the impulse responses of our final model, they were not seen as clearly in other kinds of analyses. For instance, a Granger-causality analysis in Chapter **??** produced only much weaker evidence for a feedback relation. One of the problems in this kind of analysis may be that the sample information is too scarce to produce clear results in relatively unrestricted models.

To illustrate this point we also provide the impulse responses of an unrestricted VAR(4) model for the two variables of interest in the lower part of Figure 4.1. Taking into account the size of the CIs the reactions of the variables are much less clear in this experiment than in the restricted subset VECM. In particular, imposing no cointegration restrictions, the long-run reactions of the variables to impulses to the system are not very clear and

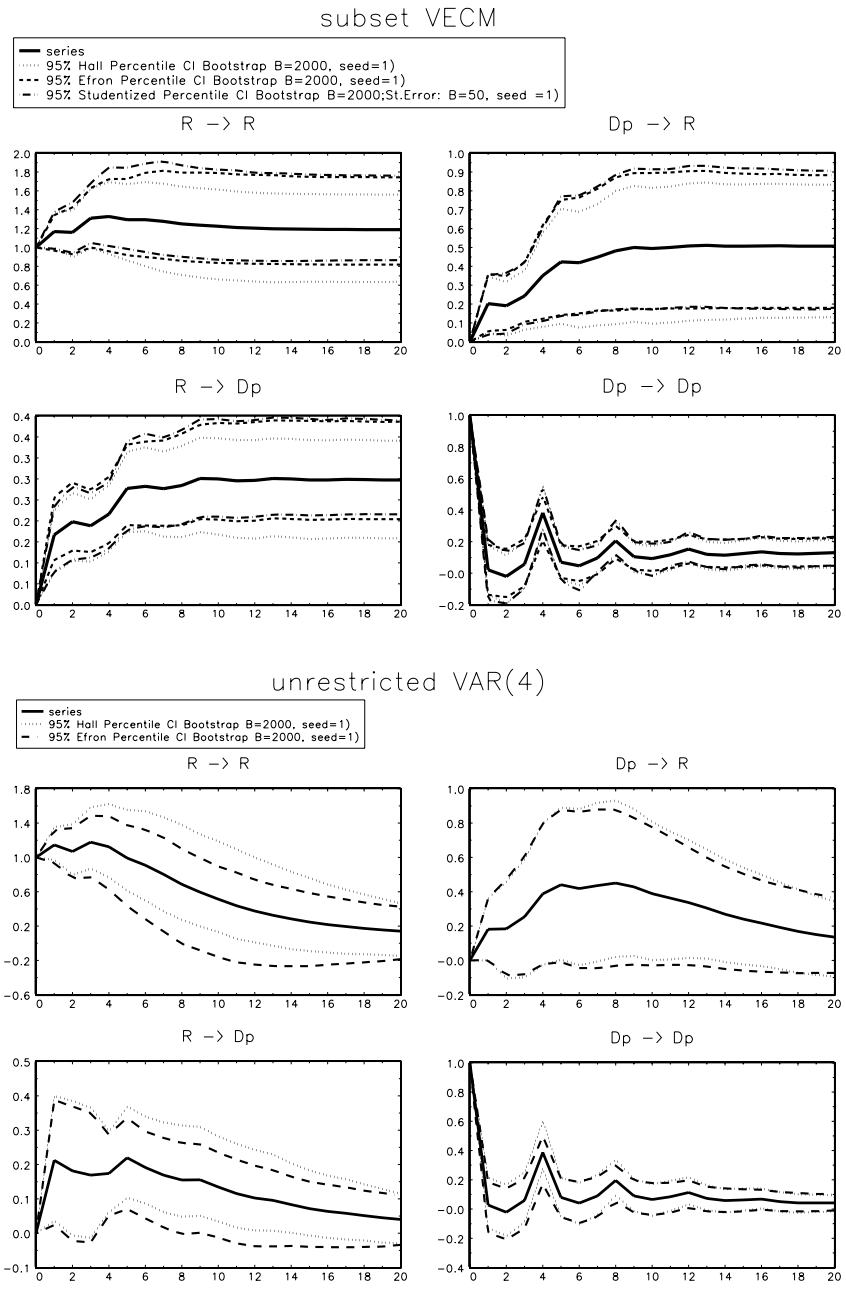


Fig. 4.1. Forecast error impulse responses of German interest rate-inflation system based on subset VECM (**??**) (upper panel) and an unrestricted VAR(4) model (lower panel).

may not even be diagnosed to be permanent. Of course, it has to be kept in mind that our modelling effort including the specification of the cointegration relation is based on a statistical analysis which cannot produce definite results with certainty. Therefore, some uncertainty remains with respect to the adequacy of the restrictions we have imposed on the final model (**??**). Consequently, also the conclusions drawn from the model have some degree of uncertainty. Still the analysis shows that the data are compatible with the conclusions drawn from the subset VECM.

4.6 Forecast Error Variance Decomposition

Forecast error variance decompositions are also popular tools for interpreting VAR models and VECMs. Recall that the h-step forecast error from a VAR model is

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}$$

(see (**??**) in Chapter **??**). Expressing this error in terms of structural innovations $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})' = \mathsf{B}^{-1}\mathsf{A}u_t$ gives

$$y_{T+h} - y_{T+h|T} = \Psi_0 \varepsilon_{T+h} + \Psi_1 \varepsilon_{T+h-1} + \cdots + \Psi_{h-1} \varepsilon_{T+1},$$

where $\Psi_j = \Phi_j \mathsf{A}^{-1}\mathsf{B}$. Denoting the ijth element of Ψ_n by $\psi_{ij,n}$, the kth element of the forecast error vector becomes

$$y_{k,T+h} - y_{k,T+h|T} = \sum_{n=0}^{h-1} (\psi_{k1,n}\varepsilon_{1,T+h-n} + \cdots + \psi_{kK,n}\varepsilon_{K,T+h-n}).$$

Using that the ε_{kt} are contemporaneously and serially uncorrelated and have unit variances by construction, it follows that the corresponding forecast error variance is

$$\sigma_k^2(h) = \sum_{n=0}^{h-1} (\psi_{k1,n}^2 + \cdots + \psi_{kK,n}^2) = \sum_{j=1}^{K} (\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2).$$

The term $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)$ is interpreted as the contribution of variable j to the h-step forecast error variance of variable k. This interpretation makes sense if the ε_{it} can be viewed as shocks in variable i. Dividing the above terms by $\sigma_k^2(h)$ gives the percentage contribution of variable j to the h-step forecast error variance of variable k,

$$\omega_{kj}(h) = (\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)/\sigma_k^2(h).$$

These quantities, computed from estimated parameters, are often reported for various forecast horizons. Clearly, their interpretation as forecast error

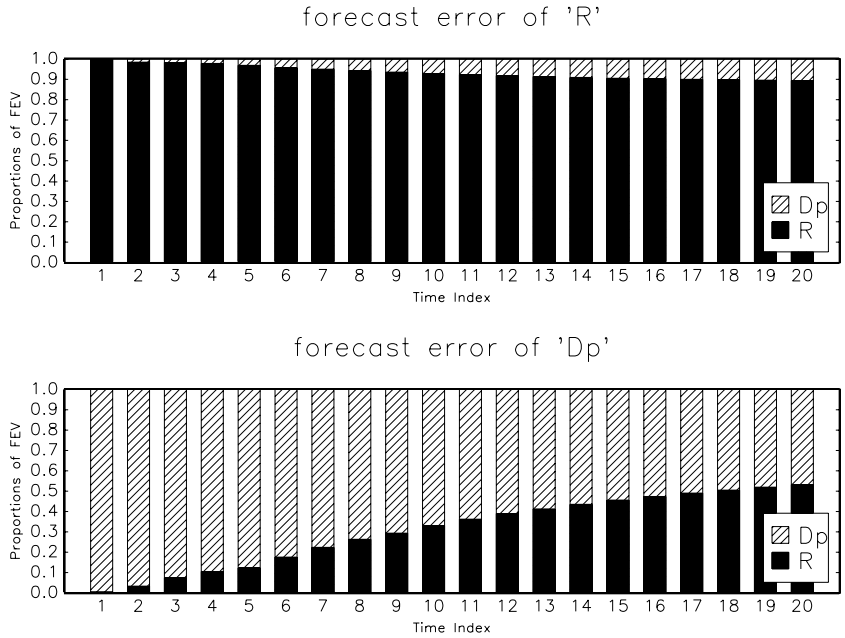


Fig. 4.2. Forecast error variance decomposition of German interest rate-inflation system based on subset VECM (**??**).

variance components may be criticized on the same grounds as structural impulse responses because they are based on the latter quantities.

In Figure 4.2 a forecast error variance decomposition of the German interest rate-inflation system based on the subset VECM (**??**) from Chapter **??** is shown. It is based on recursively orthogonalizing the innovations via a Choleski decomposition of the covariance matrix. In Figure 4.2 it appears that the interest rate dominates the system to some extent. Its forecast errors are largely attributable to own innovations whereas the forecast error of the inflation rate are partly determined by interest rate changes, at least for longer term forecasts. This interpretation has to be viewed with caution, however, because the forecast error variance components are computed from estimated quantities and are therefore uncertain. Also, in the present case, one may wonder whether the picture changes if the order of the variables is interchanged. Recall that the forecast error variance decomposition is based on the orthogonalized impulse responses for which the order of the variables matters. Although the instantaneous residual correlation is small in our subset VECM, it will have some impact on the outcome of a forecast error variance decomposition. For the present example the impact is small,

however. Therefore we don't show the result but encourage the reader to perform his/her own analysis.

4.7 Examples

In this section we illustrate the use of SVAR models by applying the methods discussed in previous sections. We start with the simple IS-LM model considered earlier to illustrate the AB-model. Then we discuss the Blanchard-Quah model as an example for using long-run restrictions for the effects of shocks in the SVAR framework. The main emphasis in these two examples is on illustrating specific SVAR issues. In the final example we consider a Canadian labour market model and go through the different steps of an analysis in more detail.

4.7.1 A Simple AB-Model

In the first example we follow Breitung (2000)**?** and estimate the stylized IS-LM model considered in Section 4.2 using U.S. macroeconomic time series. The empirical model is a simple trivariate VAR which includes the log of real GDP (q_t), a three month interbank interest rate (i_t) and the log of the real monetary base (m_t). Therefore the vector of time series variables is $y_t = (q_t, i_t, m_t)'$. We use quarterly, seasonally adjusted data for the period of $1970Q1$ until $1997Q4$ from the Federal Reserve Economic Data (FRED) database maintained at the Federal Reserve Bank of St. Louis. Investigating the trending properties reveals evidence for a unit root in each of the three time series. As mentioned earlier, the estimation procedure from subsection 4.4.1 may be applied to a VAR model fitted to the levels even if variables have unit roots, hence, ignoring possible cointegration restrictions. This is frequently done in SVAR modelling to avoid imposing too many restrictions and we follow that road here. The specification of the reduced form model used in our analysis is the model used in Breitung (2000)**?**. More precisely, the starting point of the analysis is a reduced form VAR(4) including a trend and a constant as deterministic terms. Using the first four observations as presample values, we have $T = 108$ observations for estimation. Alternative reduced form model specifications may be used for the subsequent structural analysis. For instance, lag length criteria point to a smaller lag length ($p = 2$ or $p = 3$), while results from residual autocorrelation tests indicate a lag length of $p = 6$. We encourage the reader to explore these possibilities as well. As mentioned before, the main purpose of our analysis is an illustration of the AB-model. Therefore we do not consider alternative specifications here.

Using the reduced form VAR(4) we impose the just-identifying restrictions discussed in Section 4.2 and estimate the structural parameters using the ML estimator considered in Section 4.4. The resulting structural parameter estimates of the matrices A and B are given by

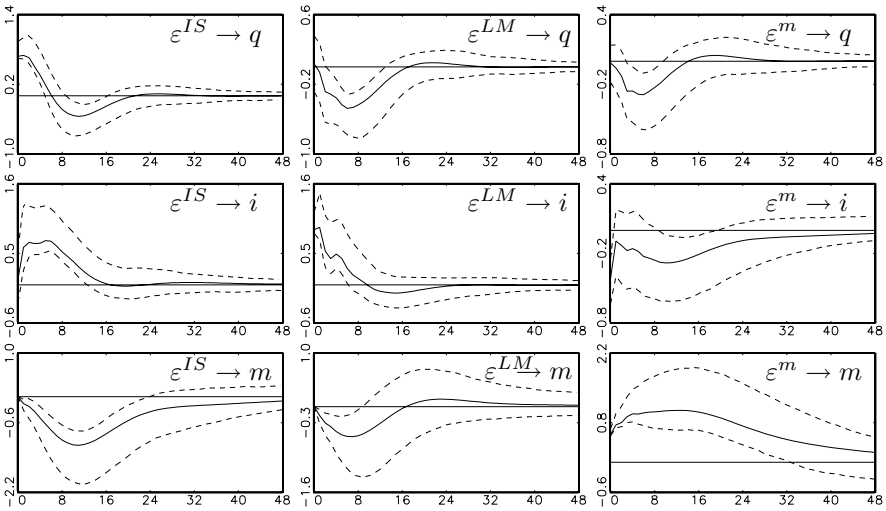


Fig. 4.3. Responses of output (q), interest rate (i) and real money (m) in IS-LM model with 95% confidence intervals based on 2000 replications of the Hall bootstrap.

$$\tilde{\mathsf{A}} = \begin{bmatrix} 1 & \underset{(-0.25)}{-0.04} & 0 \\ \underset{(-0.51)}{-0.14} & 1 & \underset{(4.92)}{0.73} \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \tilde{\mathsf{B}} = \begin{bmatrix} \underset{(13.9)}{0.0068} & 0 & 0 \\ 0 & \underset{(14.5)}{0.0087} & 0 \\ 0 & 0 & \underset{(14.7)}{0.0056} \end{bmatrix},$$

where we list asymptotic t-ratios in parentheses below the coefficient estimates. It turns out that the estimated coefficient a_{12} is negative and, thus, has the wrong sign ($\hat{\mathsf{a}}_{12} = -0.04$) but is statistically insignificant. The parameters of the inverted money demand relation have the expected sign, however, the coefficient for the output innovation (-0.14) is not significant at conventional levels.

Once the structural model has been identified and estimated the effects of the structural shocks ε_t can be investigated within an impulse response analysis. The results of the impulse response analysis are often more informative than the structural parameter estimates itself. For this purpose, the estimated contemporaneous impact matrix can be obtained from the ML estimates of the structural parameters:

$$\tilde{\mathsf{A}}^{-1}\tilde{\mathsf{B}} = \begin{bmatrix} 0.69 & 0.03 & -0.02 \\ 0.10 & 0.88 & -0.42 \\ 0.00 & 0.00 & 0.56 \end{bmatrix} \times 10^{-2}.$$

Figure 4.3 gives the responses (multiplied by 100) of the three system variables to the identified structural shocks together with 95% Hall bootstrap confidence intervals based on 2000 bootstrap replications. According to our estimated SVAR model an IS or spending shock (ε^{IS}) increases output immediately and for about 8 quarters, increases interest rates with a maximal response after two years and gradually drives down real money holdings. These effects are predicted by the IS-LM model. Moreover, an LM shock (ε^{LM}) leads to an increase in the interest rate and a decrease in real money holdings as well as a decrease of output for about three years. A positive money supply shock (ε^{m}) drives output down after roughly a year, which is at odds with economic theory. Although this effect is only borderline significant, it might suggest that money supply shocks are not appropriately captured by our stylized model. However, this money supply shock leads to an immediate drop in the interest rate and to a gradual increase of real money balances. This reaction is known as the 'liquidity effect' and has also been observed in more complex SVAR models. This examples demonstrate that even a very simple model produces results that are largely in line with predictions of a basic IS-LM model.

4.7.2 The Blanchard-Quah Model

Our second example provides estimates for the Blanchard-Quah model introduced in Section 4.3. Their model is a simple bivariate VAR model for the growth rates of output and the unemployment rate U_t. The growth rate is measured as the first differences of log output, denoted as ΔQ_t. Consequently, $y_t = (\Delta Q_t, U_t)'$. We use quarterly U.S. time series data for the period between $1948Q2$ and $1987Q4$ as in the original study by Blanchard & Quah (1989). The data are available from the data archive of the Journal of Applied Econometrics (see data for Weber (1995)). In the data set, the growth rate of output is adjusted for a constant mean with a structural break at $1974Q1$, whereas the unemployment rate is adjusted for a linear time trend. ADF unit root tests for both variables clearly reject the hypothesis of a unit root indicating that both series are stationary.

Recalling the discussion in Section 4.3, Blanchard & Quah (1989) identify aggregate demand shocks by assuming that they have merely transitory effects on the level of output. In contrast, aggregate supply shocks may have permanent effects on Q_t. In other words, the vector of structural shocks $\varepsilon_t = (\varepsilon_t^s, \varepsilon_t^d)'$ is identified by restricting the long-run effect of the demand shock ε_t^d on output (i.e. the (1,2)-element of the matrix Ψ) to be zero. This restriction implies a lower triangular structure for Ψ, such that $\hat{\Psi}$ can be obtained by a Choleski decomposition of the long-run covariance matrix (see Section 4.4). Results from lag length information criteria suggest a VAR order of $p = 2$ and, apart from some nonnormality in the residuals of the unemployment equation, a VAR(2) is a well specified model. While the reader

is invited to perform the following analysis based on a VAR(2) specification, we follow Blanchard & Quah (1989) and use a VAR(8) with a constant term to estimate the reduced form model in order to replicate the original results as closely as possible. Checking the t-values of the VAR(8) reveals many insignificant coefficient estimates, such that a subset model may be another alternative reduced form model. Based on the unrestricted VAR(8) we estimate the structural decomposition by computing the contemporaneous impact matrix B, as well as the identified total long-run impact matrix Ψ:

$$\hat{\mathsf{B}} = \begin{bmatrix} \underset{(0.270)}{0.075} & \underset{(\text{-}7.02)}{-0.930} \\ \underset{(3.18)}{0.220} & \underset{(2.98)}{0.208} \end{bmatrix} \quad \text{and} \quad \widehat{\Psi} = \begin{bmatrix} \underset{(4.29)}{0.519} & 0 \\ \underset{(0.005)}{0.008} & \underset{(3.01)}{4.044} \end{bmatrix}, \tag{4.30}$$

where we give t-statistics based on 2000 bootstrap draws in parentheses. The t-statistics have been obtained using the standard bootstrap from Section 4.5.2. These estimates suggest that supply shocks have a significant, positive long-run effect on output, whereas the long-run effect of a (negative) demand shock is restricted to zero.

Figure 4.4 shows the implied impulse response functions for the Blanchard-Quah example together with approximate 95% confidence intervals obtained from Hall's bootstrap method using 2000 replications. Because we are interested in the effects of the structural shocks on the level of output, the first row in Figure 4.4 shows the accumulated impulse responses for ΔQ_t which are the responses of Q_t. As already suggested by the estimate of $\hat{\Psi}$, positive supply shocks increase output in the long-run. The interval estimate indicates, however, that output does not react significantly to supply shocks for about one year after the shock has occurred. Interestingly, a positive supply shock increases unemployment for about one year before it returns to its initial level. The response of output to a negative demand shock given in the upper right panel of Figure 4.4 represents the identifying long-run restriction imposed. After the initial decrease, output returns to its initial level after about 2 years. Similarly, unemployment goes up reaching its peak one year after the shock has occurred and then returns to its equilibrium level. Overall, the results seem compatible with the usual interpretation of demand and supply effects.

Just to illustrate the impact of imposing subset restrictions on the present model, we have also determined estimates of B and Ψ based on a subset model obtained with the sequential selection procedure in conjunction with the AIC described in Section **??**. The resulting estimates are

$$\hat{\mathsf{B}} = \begin{bmatrix} \underset{(0.55)}{0.129} & \underset{(\text{-}10.0)}{-0.896} \\ \underset{(3.23)}{0.199} & \underset{(4.31)}{0.214} \end{bmatrix} \quad \text{and} \quad \widehat{\Psi} = \begin{bmatrix} \underset{(6.63)}{0.507} & 0 \\ \underset{(1.09)}{0.990} & \underset{(4.69)}{2.842} \end{bmatrix}.$$

The corresponding impulse responses are shown in Figure 4.5. Obviously, they do not look very different from those in Figure 4.4. In the latter figure the

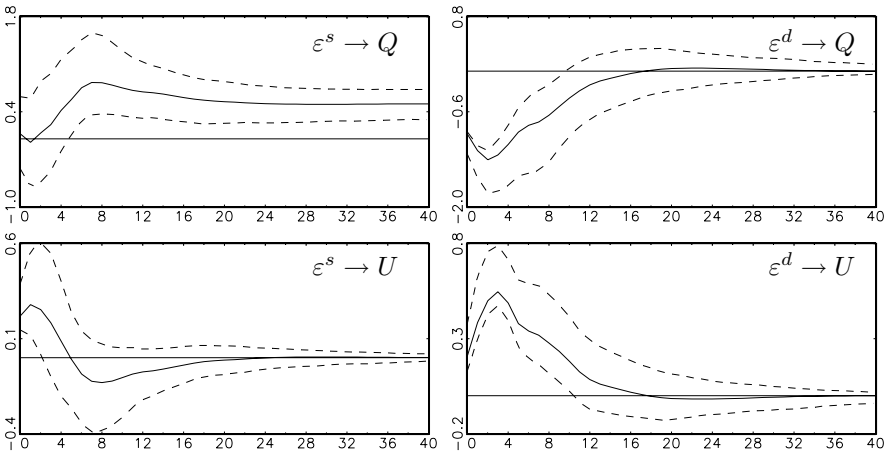


Fig. 4.4. Responses of output and unemployment to supply and (negative) demand shock based on an unrestricted VAR(8) with 95% Hall bootstrap confidence intervals using 2000 bootstrap replications.

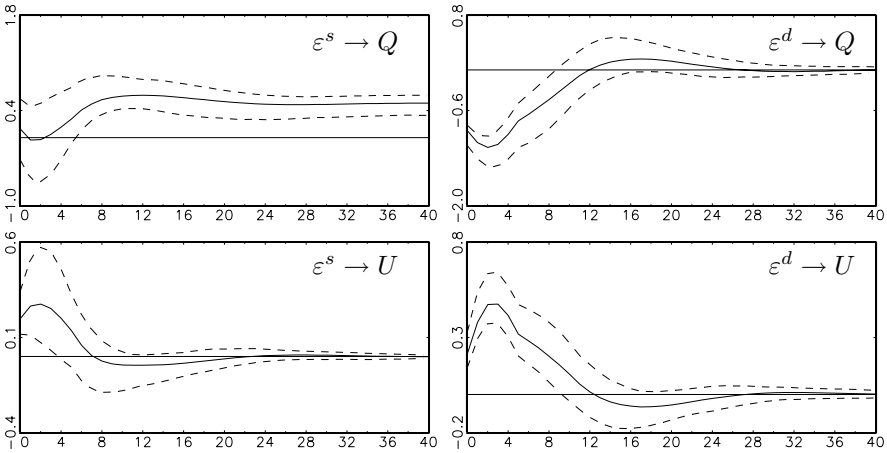


Fig. 4.5. Responses of output and unemployment to supply and (negative) demand shock based on a subset VAR model with 95% Hall bootstrap confidence intervals using 2000 bootstrap replications.

confidence intervals are a bit wider in general, however. Thus, imposing subset restrictions may result in more precise estimates of the impulse responses, provided, of course, that the subset restrictions are valid.

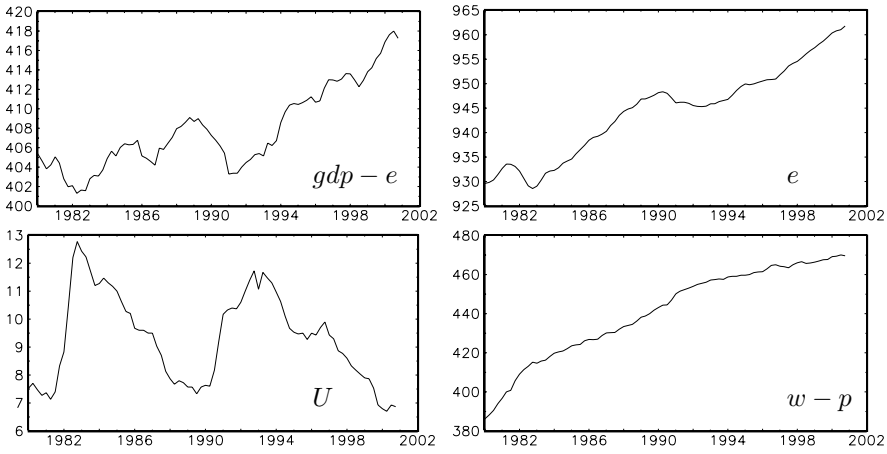


Fig. 4.6. Canadian labour market time series, $1980Q1 - 2000Q4$.

4.7.3 A SVECM for Canadian Labour Market Data

In the following, we use the theoretical framework considered in Section 4.3.2 to identify macroeconomic shocks to the labour market within a VECM for Canadian data. In particular, the empirical VAR model is specified for labour productivity, employment, unemployment rate and real wages. In our analysis we use quarterly, seasonally adjusted data for Canada constructed from data obtained from the OECD database for the period of $1980Q1$ to $2000Q4$. Productivity ($gdp - e$) is constructed by subtracting the log of employment e from the log real GDP (gdp), U denotes the unemployment rate and $w - p$ is the log of a real wage index. The time series vector used in our analysis is thus given by

$$y_t = [(gdp_t - e_t), e_t, U_t, (w_t - p_t)]'$$

and we show the time series plots in Figure 4.6.

The time series for productivity, employment and real wages show a clear upward trending behaviour. Although this upward trend cannot be observed in the unemployment series, none of the four times series look stationary. Therefore, we conduct ADF unit root tests in a first step before proceeding with the reduced form model specification. The unit root test results are summarized in Table 4.1 Given the shape of the time series in Figure 4.6 we include a constant and a deterministic time trend in the test regressions for $gdp - e$, e and $w - p$. Accordingly, we only use a constant when testing for a unit root in unemployment. As the number of lagged differences included in the Dickey-Fuller test we use values suggested by the AIC criterion when employing a maximum lag order of $p_{\max} = 8$. Given the test results in Table 4.1, unit roots cannot be rejected in the levels of all four variables. To test for unit roots in the first differences of our variables, we include a constant in the

Table 4.1. Augmented Dickey-Fuller Unit Root Tests for Canadian Data

	deterministic		test	critical values[a]		
variable	terms	lags	value	1%	5%	10%
$gdp - e$	constant,trend	2	-1.98	-3.96	-3.41	-3.13
$\Delta(gdp - e)$	constant	1	-5.16^{***}	-3.43	-2.86	-2.57
e	constant,trend	2	-1.99	-3.96	-3.41	-3.13
Δe	constant	1	-4.51^{***}	-3.43	-2.86	-2.57
U	constant	1	-2.36	-3.43	-2.86	-2.57
ΔU	-	0	-4.75^{***}	-2.56	-1.94	-1.62
$(w - p)$	constant,trend	4	-1.57	-3.96	-3.41	-3.13
$\Delta(w - p)$	constant	3	-2.62^{*}	-3.43	-2.86	-2.57
$\Delta(w - p)$	constant	0	-5.60^{***}	-3.43	-2.86	-2.57

[a] Critical values from Davidson & MacKinnon (1993, Table 20.1).

test regression of $\Delta(gdp - e)$, Δe and $\Delta(w - p)$ and no deterministic terms in the equation for ΔU. Moreover, compared to the level test regressions the number of lagged differences is decreased by one. These specifications are obtained by taking first differences of the levels model. The unit root hypothesis is rejected at the 1% level for the first differences of productivity, employment and unemployment, whereas for $\Delta(w - p)$ it can only be rejected on the 10% level. The conclusions of the ADF test for productivity, employment and the unemployment rate are quite clear and not sensitive to the choice of the lag length. The test statistic for the real wages, however, is only slightly larger than the 10% critical values but varying the number of lagged differences (e.g., to 0, 1 or 2) in the test regressions leads to the rejection of a unit root at the 1% level. As an example, we included the test result for the specification with no lagged differences as suggested by the SC criterion. Overall, the results in Table 4.1 suggest that $gdp - e$, e, U and $w - p$ can be treated as I(1) variables.

Given the integration and trending properties of the time series, cointegration between the four variables is possible. Therefore, the next step in our analysis is the specification of an initial unrestricted VAR model which forms the basis for systems cointegration tests as discussed in Chapter **??**. For this purpose we employ information criteria to select the lag length of a VAR specification including a constant and a deterministic trend. Choosing a maximum lag order of $p_{\max} = 8$, AIC and FPE suggest $p = 3$, HQ proposes $p = 2$ while SC chooses $p = 1$. For all the suggested lag lengths we conduct a series of diagnostic tests. In particular, we test against autocorrelation, nonnormality and ARCH effects in the VAR residuals (see Chapter **??** for details on the diagnostic tests). We list the results for $p = 1, 2$ and 3 in Table 4.2. For $p = 3$ and $p = 2$, none of the diagnostic tests indicates signs of misspecification. The VAR(1) suggested by the SC criterion, however, shows some signs of residual autocorrelation and residual ARCH effects. Moreover, uni-

Table 4.2. Diagnostic Tests for VAR(3) Specification for Canadian Data

	Q_{16}	Q^*_{16}	FLM_5	LJB_4	LJB^L_4	$VARCH_{LM}(5)$
$p=3$	174.0 [0.96]	202.9 [0.59]	0.99 [0.51]	8.63[0.37]	9.67 [0.29]	512.0 [0.35]
$p=2$	209.7 [0.74]	241.8 [0.20]	1.20 [0.16]	3.23[0.92]	2.28 [0.97]	528.1 [0.19]
$p=1$	233.5 [0.61]	263.1 [0.15]	1.74 [0.00]	9.71[0.24]	9.92 [0.27]	570.1 [0.02]

Note: p-values in brackets.

Table 4.3. Johansen Cointegration Tests for Canadian System

	test statistics		critical values[a]		
H_0	$p=3$	$p=2$	90%	95%	99%
$r=0$	84.92	86.12	58.96	62.61	70.22
$r=1$	36.42	37.33	39.08	42.20	48.59
$r=2$	18.71	15.64	22.95	25.47	30.65
$r=3$	3.85	4.10	10.56	12.39	16.39

Note: Deterministic terms: constant and linear trend (restricted to long-run part).
[a] Critical values from Johansen (1995, Table 15.4).

variate Lomnicki-Jarque-Bera tests (results not shown) indicate some signs of nonnormality for the employment and unemployment equation if $p = 1$ is used. Therefore, the lag length $p = 1$ seems too restrictive as an initial choice. Using the tools of stability analysis for $p = 2$ and $p = 3$ reveals some signs of parameter non-constancy. In particular, some significant Chow tests, the visual inspection of recursive coefficients and some large recursive residuals point to instability of some VAR model parameters. Although, in principle, it would be desirable to model these effects, we continue the analysis using a time invariant VAR model.

To test for cointegration we use both, the VAR(2) and VAR(3) specification. As deterministic terms we include a constant and a linear trend. Using $p = 3$, the Johansen cointegration test (see Section **??**) rejects the null hypothesis of no cointegration at the 1% level. In contrast, the null of $r = 1$ cannot be rejected on conventional significance levels. Note, however, the value of the test statistic is not too far away from the critical values. Nevertheless, formally, the Johansen test indicates a cointegrating rank of $r = 1$ and for illustrative purposes we proceed by assuming a cointegrating rank of $r = 1$. Against the background of the theoretical model discussed in Section 4.3.2 which suggests up to two cointegration relations, the reader is invited to perform a sensitivity analysis using $r = 2$ as an alternative specification.

We estimate a VECM based on the VAR(3) specification suggested by AIC and FPE under the rank restriction $r = 1$, that is, we include two lags of the differences of the variables. The diagnostic tests for this VECM do not indicate signs of misspecification. Moreover, we show the recursive eigenvalues and the corresponding τ-tests in Figure 4.7, which were computed

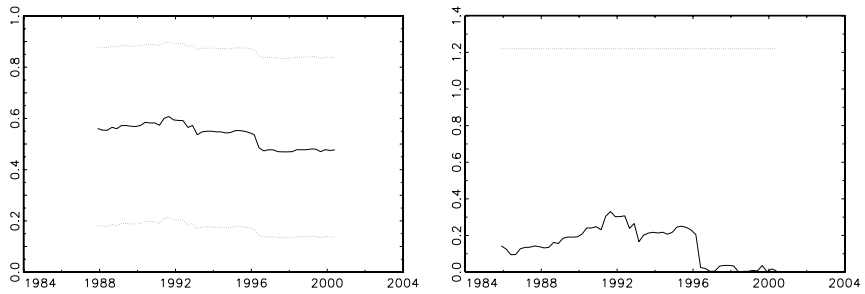


Fig. 4.7. Recursive eigenvalue (left) and τ-test with 10% critical value (right) of VECM with two lagged differences and $r = 1$.

Table 4.4. Cointegration Vector and Loading Parameters for VECM with two Lagged Differences and Cointerating Rank $r = 1$

	$gdp - e$	e	U	$w - p$	t
$\hat{\beta}'$	0.545	−0.013	1.727	1	0.709
	(1.69)	(-0.09)	(4.11)		(24.3)
$\hat{\alpha}'$	−0.012	−0.016	−0.009	−0.085	
	(-0.92)	(-2.15)	(-1.45)	(-5.71)	

Note: t-statistics in parentheses.

by choosing the recursive estimation of all parameters thereby allowing for possibly changing short-run parameters. Neither the recursive eigenvalues nor the stability tests point to instabilities of the eigenvalue. Hence, we regard the VECM(3) model as an adequate description of the data. Table 4.4 shows the Johansen ML estimate of the cointegration relation β, where we have normalized the coefficient of real wages to one. This cointegration vector may be interpreted as a stationary wage setting relation, in which real wages are related to productivity and unemployment as in equation (4.17). Rewriting the vector in the form of a wage setting relation gives

$$w_t - p_t = -0.545(gdp - e)_t + 0.013e_t - 1.727U_t + 0.709t + ec_t. \qquad (4.31)$$

Productivity enters the wage relation with the wrong sign, but this may be just a consequence of including a linear trend in (4.31). This trend may be interpreted as a proxy for productivity that picks up the effects of productivity increases. The coefficient for employment is relatively small and may as well be zero. In line with theoretical arguments, real wages are negatively related to unemployment, hence, outsiders influence the wage setting in Canada. The estimates of the adjustment coefficients α support the interpretation of the cointegration vector as a wage setting schedule. According to the significant loading estimates from Table 4.4, excess real wages drive real wage growth down and also decrease employment growth. Having specified

the reduced form model, we now turn to the structural analysis. The reader is encouraged to specify a satisfactory subset model for the present data set. We continue our analysis based on a full VECM with two lagged differences and cointegrating rank 1. For this purpose, we discuss the identification of shocks to the Canadian labour market in the following.

Identification of Labour Market Shocks Remember from Section 4.3.2 that the vector of structural shocks is given by $\varepsilon_t = (\varepsilon_t^{gdp}, \varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^w)'$. To derive the set of identifying assumptions, we make use of the theoretical labour market model discussed in Section 4.3.2. If we let $\mathsf{A} = I_K$, we need $\frac{1}{2}K(K-1) = 6$ linearly independent restrictions to uniquely identify the parameters in B. We have $k^* = K - r = 3$ shocks with permanent effects in our VECM, while $r = 1$ merely have transitory effects. The cointegration analysis suggests that the wage setting relation is stationary. Accordingly, wage shocks ε^w have no long-run impact on the variables included in y_t, which corresponds to four zero restrictions in the last column of the identified long-run impact matrix $\Xi\mathsf{B}$. Due to the reduced rank of $\Xi\mathsf{B}$, this only imposes $k^* r = 3$ linearly independent restrictions. To identify the $k^* = 3$ permanent shocks, $k^*(k^*-1)/2 = 3$ additional restrictions are necessary. Assuming constant returns to scale, productivity is only driven by technology shocks ε^{gdp} in the long-run. These restrictions are imposed by setting the elements $(\Xi\mathsf{B})_{1j}$, $j = 2, 3, 4$, equal to zero. Because $(\Xi\mathsf{B})_{14} = 0$ is already imposed by the first set of restrictions, only two additional linearly independent restrictions are provided by assuming constant returns to scale. Consequently, one additional restriction is needed for just-identification of the SVECM model. As the theoretical model does not suggest an additional long-run restriction, we choose one contemporaneous restriction, i.e., a zero restriction for B. As in Section 4.3.2, we assume that labour demand shocks do not affect real wages on impact, i.e. $\mathsf{B}_{42} = 0$. In Section 4.3.2 this restriction was needed to identify the two transitory shocks. In the present empirical analysis, only one shock is transitory and consequently no contemporaneous restrictions are needed to identify it. Here, the restriction is used to identify the permanent shocks. Taken together, these identifying assumption are exactly identifying and correspond to the following structure on the contemporaneous impact matrix B and the identified long-run impact matrix $\Xi\mathsf{B}$:

$$\mathsf{B} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & 0 & * & * \end{pmatrix} \quad \text{and} \quad \Xi\mathsf{B} = \begin{pmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \end{pmatrix}. \tag{4.32}$$

Using the estimation procedure described in Section 4.4.2 we obtain the following estimates for the contemporaneous and long-run impact matrix:

$$\tilde{\mathsf{B}} = \begin{pmatrix} \underset{(5.98)}{0.62} & \underset{(0.64)}{0.08} & \underset{(-0.66)}{-0.16} & \underset{(0.92)}{0.07} \\ \underset{(-1.69)}{-0.13} & \underset{(4.11)}{0.28} & \underset{(-0.87)}{-0.17} & \underset{(2.21)}{0.10} \\ \underset{(0.44)}{0.03} & \underset{(-5.45)}{-0.29} & \underset{(0.09)}{0.01} & \underset{(1.59)}{0.05} \\ \underset{(0.74)}{0.12} & 0 & \underset{(0.74)}{0.52} & \underset{(5.91)}{0.52} \end{pmatrix} \tag{4.33}$$

and

$$\widetilde{\Xi\mathsf{B}} = \begin{pmatrix} \underset{(5.16)}{0.85} & 0 & 0 & 0 \\ \underset{(0.85)}{0.22} & \underset{(3.26)}{0.62} & \underset{(-0.85)}{-0.53} & 0 \\ \underset{(-1.37)}{-0.17} & \underset{(-3.76)}{-0.36} & \underset{(0.92)}{0.15} & 0 \\ \underset{(-0.83)}{-0.16} & \underset{(3.75)}{0.64} & \underset{(-0.91)}{-0.27} & 0 \end{pmatrix}. \tag{4.34}$$

In parentheses we provide bootstrapped t-values obtained using 2000 bootstrap replications. The estimated long-run effects of labour market shocks on unemployment are given in by the third row in (4.34). Note that according to our estimate only labour demand shocks significantly decrease the unemployment rate in Canada in the long-run. Conditional on the set of just-identifying restrictions, it is also possible to test further restrictions on the effects of structural shocks. For instance, checking whether labour supply shocks have no long-run effects on unemployment corresponds to testing the hypothesis $\mathrm{H}_0 : (\Xi\mathsf{B})_{33} = 0$. In our example, the corresponding LR test statistic has a $\chi^2(1)$-distribution and is given by $LR = 6.07$ with a p-value $= 0.014$. Hence, the the null hypothesis is rejected on the 5% significance level. Using the estimates $\tilde{\mathsf{B}}$ in (4.33) we may also compute the responses to the structural shocks ε_t which provide a more informative picture on the dynamic effects of macroeconomic shocks to the Canadian labour market.

Figure 4.8 shows the responses of unemployment to a technology, a labour demand, a labour supply and wage setting shocks together with 95% Hall bootstrap confidence intervals based on 2000 bootstrap replications. The point estimates suggest that a technology shock (ε^{gdp}) drives unemployment down. However, this effect is not significant neither in the short-run nor in the long-run. A labour demand shock (ε^{d}) leads to a significant drop in unemployment reaching the maximum effect after about one year. A positive labour supply (ε^{s}) shock has a significant positive impact on unemployment for about two years. In the long-run there is no significant effect of the labour supply shock, which is in line with the bootstrap t-values of the estimated long-run impact matrix. Finally, unemployment increases after a shock to wage setting (ε^{w}). This effect becomes significant only about 6 quarters after the shock. In the long-run the response to wage shocks is zero as imposed by our identifying assumption. Note, however, that compared to responses to other shocks in the system, the reaction of wages is relatively small. Overall, the impulse responses are in line with what one would have expected from

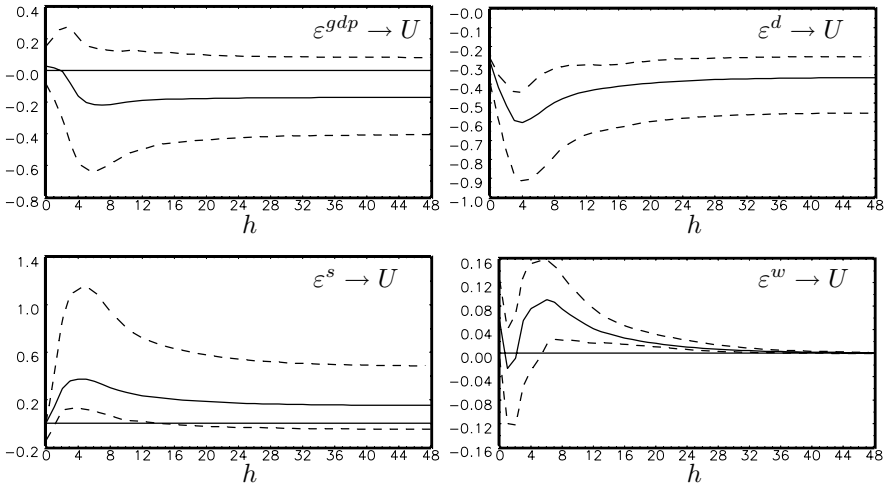


Fig. 4.8. Responses of unemployment to economic shocks with 95% Hall bootstrap confidence intervals based on 2000 bootstrap replications.

economic theory. Moreover, the adjustment to a new labour market equilibrium takes about 3-5 years.

Table 4.5. Forecast Error Variance Decomposition of Canadian Unemployment

h	ε^{gdp}	ε^{d}	ε^{s}	ε^{w}
1	0.009	0.887	0.126	0.035
4	0.007	0.926	0.026	0.006
8	0.054	0.832	0.028	0.012
12	0.075	0.813	0.024	0.012
24	0.099	0.815	0.016	0.008
48	0.120	0.824	0.010	0.005

To assess the relative importance of the identified labour market shocks we list for different horizons h the forecast error variance decomposition of unemployment in Table 4.5. Clearly, according to our estimates, labour demand shocks are the dominant source for Canadian unemployment. Even in the long-run more than 80% of the variation in the unemployment variable can be attributed to the labour demand shocks. While technology shocks are not important at short horizons, they become more and more important as h increases. After $h = 48$ periods, technology shocks explain a fraction of 12% of the variance in unemployment. In contrast, labour supply and wage setting shocks are not an important source for the determination of unemployment

in the long-run. Note, that the result for the wage setting shocks is implied by the identifying restrictions of the shocks.

As in every structural VAR analysis, the results may depend to some extend on the specification of the reduced form model and the choice of identifying assumptions. Here, the main purpose was to illustrate the use of structural VECM models. Therefore, we used only one specification. The reader is invited to check the robustness of the results with respect to different reduced form model specifications and alternative identifying assumptions.

4.8 Conclusions

In this chapter tools for analyzing dynamic time series models have been discussed. If the dynamic interactions between the variables of a given system are of interest, impulse response analysis can give useful insights. A major problem is, however, to identify those innovations that actually reflect the ongoings in the system. It was discussed how information from subject matter theory can be used to identify unique impulses that can be traced through the system and may convey useful information on the relations of the variables. Imposing such restrictions means to consider structural VARs or VECMs. Forecast error variance decompositions are alternative tools for a dynamic analysis. They can also be based on structural VARs or VECMs.

Although imposing structural restrictions resolves the nonuniqueness problem of the innovations, it may be worth remembering that Sims (1980) advocated VAR models as alternatives to econometric simultaneous equations models because he regarded the identifying restrictions used for them as "incredible". Thus structural VAR modelling may be criticized on the same grounds. Even if the restrictions imposed are firmly based on some economic theory, they may not truly reflect what goes on in the actual underlying system. Thus, all we can hope for in an impulse response analysis is getting information on the question whether some theory or theoretical concept is compatible with the data or rather whether the data can be interpreted in such a way so as to reflect a specific theory. It is much harder to reject a specific theory on the basis of a VAR or VECM analysis because if the implications of the theory are not reflected in the impulse responses, say, this can have various reasons. Choosing the identifying restrictions properly is only one source of possible error. Other sources may be a wrong choice of variables, inappropriate data or poor modelling and estimation. Therefore, conclusions from analyzing a dynamic model are conditional on a number of assumptions and choices of the analyst. These reservations should be kept in mind and they should be a challenge for improving ones analysis.

References

Amisano, G. & Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Benkwitz, A., Lütkepohl, H. & Neumann, M. (2000). Problems related to bootstrapping impulse responses of autoregressive processes, *Econometric Reviews* **19**: 69–103.

Benkwitz, A., Lütkepohl, H. & Wolters, J. (2001). Comparison of bootstrap confidence intervals for impulse responses of German monetary systems, *Macroeconomic Dynamics* **5**: 81–100.

Bernanke, B. (1986). Alternative explanations of the money–income correlation, *Carnegie-Rochester Conference Series on Public Policy*, North-Holland, Amsterdam.

Blanchard, O. & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* **79**: 655–673.

Breitung, J. (2000). *Structural Inference in Cointegrated Vector Autoregressive Models*, Habilitationsschrift, Humboldt-Universität zu Berlin.

Breitung, J. (2001). A convenient representation for structural vector autoregressions, *Empirical Economics* **26**: 447–459.

Christiano, L. J., Eichenbaum, M. & Evans, C. (1999). Monetary shocks: What have we learned and to what end?, *in* J. Taylor & M. Woodford (eds), *The Handbook of Macroeconomics*, Amsterdam: Elsevier Science Publication.

Cooley, T. F. & LeRoy, S. F. (1985). Atheoretical macroeconometrics: A critique, *Journal of Monetary Economics* **16**: 283–308.

Davidson, R. & MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, London.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Galí, J. (1992). How well does the IS-LM model fit postwar U.S. data?, *Quarterly Journal of Economics* **107**(2): 709–738.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?, *American Economic Review* **89**(1): 249–271.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Jacobson, T., Vredin, A. & Warne, A. (1997). Common trends and hysteresis in Scandinavian unemployment, *European Economic Review* **41**: 1781–1816.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* **80**: 218–230.

King, R. G., Plosser, C. I., Stock, J. H. & Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.

Lewis, R. & Reinsel, G. C. (1985). Prediction of multivarate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**: 393–411.

Lütkepohl, H. (1988). Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process, *Econometric Theory* **4**: 77–85.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.

Lütkepohl, H. (1996). Testing for nonzero impulse responses in vector autoregressive processes, *Journal of Statistical Planning and Inference* **50**: 1–20.

Lütkepohl, H. & Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models, *Econometric Theory* **7**: 487–496.

Lütkepohl, H. & Poskitt, D. S. (1996). Testing for causation using infinite order vector autoregressive processes, *Econometric Theory* **12**: 61–87.

Lütkepohl, H. & Reimers, H.-E. (1992). Impulse response analysis of cointegrated systems, *Journal of Economic Dynamics and Control* **16**: 53–78.

Lütkepohl, H. & Saikkonen, P. (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes, *Journal of Econometrics* **81**: 127–157.

Pagan, A. (1995). Three econometric methodologies: An update, *in* L. Oxley, D. A. R. George, C. J. Roberts & S. Sayer (eds), *Surveys in Econometrics*, Basil Blackwell, Oxford.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.

Saikkonen, P. & Lütkepohl, H. (1996). Infinite order cointegrated vector autoregressive processes: Estimation and inference, *Econometric Theory* **12**: 814–844.

Saikkonen, P. & Lütkepohl, H. (2000). Asymptotic inference on nonlinear functions of the coefficients of infinite order cointegrated VAR processes, *in* W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim & A. Würtz (eds), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge University Press, Cambridge, pp. 165–201.

Sargent, T. J. (1978). Estimation of dynamic labor demand schedules under rational expectations, *Journal of Political Economy* **86**: 1009–1044.

Shapiro, M. & Watson, M. W. (1988). Sources of business cycle fluctuations, *NBER Macroeconomcs Annual* **3**: 111–156.

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

Sims, C. A. (1981). An autoregressive index model for the U.S. 1948-1975, *in* J. Kmenta & J. B. Ramsey (eds), *Large-Scale Macro-Econometric Models*, North-Holland, Amsterdam, pp. 283–327.

Sims, C. A. (1986). Are forecasting models usable for policy analysis?, *Quarterly Review, Federal Reserve Bank of Minneapolis* .

Vlaar, P. (1998). On the asymptotic distribution of impulse response functions with long run restrictions, *DNB Staff reports 22*, De Nederlandsche Bank.

Weber, C. E. (1995). Cyclical output, cyclical unemployment, and Okun's coefficient: A new approach, *Journal of Applied Econometrics* **10**: 433–445.

Wold, H. (1960). A generalization of causal chain models, *Econometrica* **28**: 443–463.