EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

EUI Working Paper **ECO** No. 2005 /2

Structural Vector Autoregressive Analysis for Cointegrated Variables

HELMUT LÜTKEPOHL

BADIA FIESOLANA, SAN DOMENICO (FI)

Published in Italy in February 2005
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

February 21, 2005

Structural Vector Autoregressive Analysis for Cointegrated Variables

by

Helmut Lütkepohl
Department of Economics
European University Institute
Via della Piazzuola 43
I-50133 Firenze
ITALY

Abstract. Vector autoregressive (VAR) models are capable of capturing the dynamic structure of many time series variables. Impulse response functions are typically used to investigate the relationships between the variables included in such models. In this context the relevant impulses or innovations or shocks to be traced out in an impulse response analysis have to be specified by imposing appropriate identifying restrictions. Taking into account the cointegration structure of the variables offers interesting possibilities for imposing identifying restrictions. Therefore VAR models which explicitly take into account the cointegration structure of the variables, so-called vector error correction models, are considered. Specification, estimation and validation of reduced form vector error correction models is briefly outlined and imposing structural short- and long-run restrictions within these models is discussed.

Key Words: Cointegration, vector autoregressive process, vector error correction model

JEL classification: C32

1 Introduction

In an influential article, Sims (1980) advocated the use of vector autoregressive (VAR) models for macro econometric analysis as an alternative to the large simultaneous equations models that were in common use at the time. The latter models often did not account for the rich dynamic structure in time series data of quarterly or monthly frequency. Given that such data became more common in macro economic studies in the 1960s and 1970s, it was plausible to emphasize modelling of the dynamic interactions of the variables of interest. Sims also criticized the way the classical simultaneous equations models were identified and questioned the exogeneity assumptions for some of the variables which often reflect the preferences and prejudices of the model builders and are not necessarily fully backed by theoretical considerations. In contrast, in VAR models all observed variables are typically treated as a priori endogenous. Restrictions are imposed to a large extent by statistical tools rather than by prior beliefs based on controversial theories.

In a VAR analysis, the dynamic interactions between the variables are usually investigated by impulse responses or forecast error variance decompositions. These quantities are not unique, however. To identify those shocks or innovations and the associated impulse responses that reflect the actual ongoings in a given system of variables, usually also requires a priori assumptions which cannot be checked by statistical tools. Therefore *structural* VAR (SVAR) models were developed as a framework for incorporating identifying restrictions for the innovations to be traced out in an impulse response analysis.

In a parallel development it was discovered that the trending properties of the variables under consideration are of major importance for both econometric modelling and the associated statistical analysis. The spurious regression problem pointed out by Granger & Newbold (1974) made it clear that ignoring stochastic trends can lead to seriously misleading conclusions when modelling relations between time series variables. Consequently, the stochastic trends, unit roots or order of integration of the variables of interest became of major concern to time series econometricians and the concept of cointegration was developed by Granger (1981), Engle & Granger (1987), Johansen (1995) and many others. In this framework, the long-run relations are now often separated from the short-run dynamics. The cointegration or long-run relations are often of particular interest because they can be associated with relations derived from economic theory. It is therefore useful to construct models which explicitly separate the long-run and short-run parts of a stochastic process. Vector error correction or equilibrium correction models (VECMs) offer a convenient framework for this purpose. They also open up the possibility to separate shocks or innovations with permanent and transitory effects. This distinction may be helpful in identifying impulse responses of interest. Therefore these models will be used as the framework in the following exposition.

A variable will be called *integrated of order d* ($I(d)$) if stochastic trends or unit roots can be removed by differencing the variable d times and a stochastic trend still remains after differencing only $d-1$ times. In line with this terminology, a variable without a stochastic trend or unit root is sometimes called $I(0)$. In the following, all variables are assumed to be either $I(0)$ or $I(1)$ to simplify matters. Hence, for a time series variable y_{kt}, it is assumed that the first differences, $\Delta y_{kt} \equiv y_{kt} - y_{k,t-1}$, have no stochastic trend. A set of $I(1)$ variables is called *cointegrated* if a linear combination exists which is $I(0)$. If a system consists of both $I(0)$ and $I(1)$ variables, any linear combination which is $I(0)$ is called a cointegration relation. Admittedly, this terminology is not in the spirit of the original idea of cointegration because in this case it can happen that a linear combination of $I(0)$ variables is called a cointegration relation. In the present context, this terminology is a convenient simplification, however. Therefore it is used here.

Although in practice the variables will usually have nonzero means, polynomial trends or other deterministic components, it will be assumed in the following that deterministic terms are absent. The reason is that deterministic terms do not play a role in impulse response analysis which is the focus of this study. Moreover, augmenting the models with deterministic terms is usually straightforward.

In the next section the model setup for structural modelling with cointegrated VAR processes will be presented. Estimation of the models is discussed in Section 3 and issues related to model specification are considered in Section 4. Conclusions follow in Section 5. The structural VECM framework of the present article was proposed by King, Plosser, Stock & Watson (1991) and a recent more general survey of structural VAR and VECM analysis with some examples was given by Breitung, Brüggemann & Lütkepohl (2004). Further references will be given in the following. The present article draws heavily on Lütkepohl (2005, Chapter 9), where further details can be found.

The following general notation will be used. The natural logarithm is abbreviated as log. For a suitable matrix A, $\text{rk}(A)$, $\det(A)$ and $A_\perp$ denote the rank, the determinant and an orthogonal complement of A, respectively. Moreover, vec is the column stacking operator which stacks the columns of a matrix in a column vector and vech is the column stacking operator for symmetric square matrices which stacks the column from the main diagonal downwards only. The $(n \times n)$ identity matrix is signified as I_n and $\mathbf{D}_n$ denotes the $(n^2 \times \frac{1}{2}n(n+1))$ duplication matrix defined such that for a symmetric $(n \times n)$ matrix A, $\text{vec}(A) = \mathbf{D}_n \text{vech}(A)$.

2 The Model Setup

As mentioned earlier, it is assumed that all variables are at most $I(1)$ and that the data generation process can be represented as a VECM of the form

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \quad t = 1, 2, \ldots, \tag{2.1}$$

where y_t is a K-dimensional vector of observable variables and α and β are $(K \times r)$ matrices of rank r. More precisely, β is the cointegration matrix and r is the cointegrating rank of the process. The term $\alpha\beta' y_{t-1}$ is sometimes referred to as error correction term. The Γ_j's, $j = 1, \ldots, p-1$, are $(K \times K)$ short-run coefficient matrices and u_t is a white noise error vector with mean zero and nonsingular covariance matrix Σ_u, $u_t \sim (0, \Sigma_u)$. Moreover, $y_{-p+1}, \ldots, y_0$ are assumed to be fixed initial conditions.

2.1 The Identification Problem

Impulse responses are often used to study the relationships between the variables of a dynamic model such as (2.1). In other words, the marginal effect of an impulse to the system is traced out over time. The residuals u_t are the 1-step ahead forecast errors associated with the VECM (2.1). Tracing the marginal effects of a change in one component of u_t through the system may not reflect the actual responses of the variables because in practice an isolated change in a single component of u_t is not likely to occur if the component is correlated with the other components. Hence, in order to identify structural innovations which induce informative responses of the variables, uncorrelated or orthogonal impulses or shocks or innovations are usually considered.

The so-called B-model setup is typically used in this context. In that setup it is assumed that the structural innovations, say ε_t, have zero mean and identity covariance matrix, $\varepsilon_t \sim (0, I_K)$, and they are linearly related to the u_t such that

$$u_t = B\varepsilon_t.$$

Hence, $\Sigma_u = BB'$. Without further restrictions, the $(K \times K)$ matrix B is not uniquely specified by these relations. In fact, due to the symmetry of the covariance matrix, $\Sigma_u = BB'$

represents only $\frac{1}{2}K(K+1)$ independent equations. For a unique specification of the K^2 elements of B we need at least $\frac{1}{2}K(K-1)$ further restrictions. Some of them may be obtained via a more detailed examination of the cointegration structure of the model, as will be seen in the following.

According to Granger's representation theorem (see Johansen (1995)), the process y_t has the representation

$$y_t = \Xi \sum_{i=1}^{t} u_i + \sum_{j=0}^{\infty} \Xi_j^* u_{t-j} + y_0^*, \quad t = 1, 2, \dots, \tag{2.2}$$

where the term y_0^* contains the initial values and the Ξ_j^*'s are absolutely summable so that the infinite sum is well-defined. Absolute summability of the Ξ_j^* implies that these matrices converge to zero for $j \to \infty$. Notice that the term $x_t \equiv \sum_{i=1}^{t} u_i = x_{t-1} + u_t$, $t = 1, 2, \dots$, is a K-dimensional random walk. The long-run effects of shocks are represented by the term $\Xi \sum_{i=1}^{t} u_i$ which captures the common stochastic trends. The matrix Ξ can be shown to be of the form

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

It has rank $K-r$. Thus, there are $K-r$ independent common trends. Substituting $B\varepsilon_i$ for u_i in the common trends term in (2.2) gives $\Xi \sum_{i=1}^{t} u_i = \Xi B \sum_{i=1}^{t} \varepsilon_i$. Clearly, the long-run effects of the structural innovations are given by ΞB because the effects of an ε_t impulse vanish in $\sum_{j=0}^{\infty} \Xi_j^* B \varepsilon_{t-j}$ in the long-run.

The structural innovations ε_t represent a regular random vector with nonsingular covariance matrix. Hence, the matrix B has to be nonsingular. Thus, $\text{rk}(\Xi B) = K - r$ and there can at most be r zero columns in the matrix ΞB. In other words, at most r of the structural innovations can have transitory effects and at least $K-r$ of them must have permanent effects. If a cointegrating rank r is diagnosed and r transitory shocks can be justified, r columns of ΞB can be restricted to zero. Because the matrix has reduced rank $K-r$, each column of zeros stands for $K-r$ independent restrictions only. Thus, the r transitory shocks represent $r(K-r)$ independent restrictions only. Still, it is useful to note that restrictions can be imposed on the basis of knowledge of the cointegrating rank of the system which can be determined by statistical means, provided as many transitory shocks can be justified as there are linearly independent cointegration relations. For a unique specification of B, further theoretical considerations are required for imposing additional restrictions, however.

For just-identification of the structural innovations in the B-model we need a total of $K(K-1)/2$ independent restrictions. Given that $r(K-r)$ restrictions can be derived from the cointegration structure of the model, this leaves us with $\frac{1}{2}K(K-1) - r(K-r)$ further restrictions for just-identifying the structural innovations. More precisely, $r(r-1)/2$ additional restrictions are required for the transitory shocks and $(K-r)((K-r)-1)/2$ restrictions are needed to identify the permanent shocks (see, e.g., King et al. (1991), Gonzalo & Ng (2001)). Together this gives a total of $\frac{1}{2}r(r-1)+\frac{1}{2}(K-r)((K-r)-1) = \frac{1}{2}K(K-1)-r(K-r)$ restrictions. The transitory shocks may be identified, for example, by placing zero restrictions on B directly and thereby specifying that certain shocks have no instantaneous impact on some of the variables. Clearly, it is not sufficient to impose arbitrary restrictions on B or ΞB. They have to be such that they identify the transitory and permanent shocks. For instance, the transitory shocks cannot be identified through restrictions on ΞB because they correspond to zero columns in that matrix. In other words, $r(r-1)/2$ of the restrictions have to be imposed on B directly. Generally, identifying restrictions are often of the form

$$C_{\Xi B}\text{vec}(\Xi B) = c_l \quad \text{and} \quad C_s\text{vec}(B) = c_s, \tag{2.3}$$

where $C_{\Xi B}$ and C_s are appropriate selection matrices to specify the long-run and contemporaneous restrictions, respectively, and c_l and c_s are vectors of suitable dimensions. In applied work, they are typically zero vectors. In other words, zero restrictions are specified in (2.3) for ΞB and B. The first set of restrictions can be written alternatively as

$$C_l\text{vec}(B) = c_l, \tag{2.4}$$

where $C_l \equiv C_{\Xi B}(I_K \otimes \Xi)$ is a matrix of long-run restrictions on B.

So far we have just discussed a "counting rule" and, hence, a necessary condition for identification. Even though the restrictions in (2.4) are linear restrictions, the full set of equations we have for B is a nonlinear one because the relation $\Sigma_u = BB'$ is nonlinear. Hence, the matrix B will only be identified locally in general. In particular, we may reverse the signs of the columns of B to find another valid matrix. If restrictions of the form

$$C_l\text{vec}(B) = c_l \quad \text{and} \quad C_s\text{vec}(B) = c_s \tag{2.5}$$

are available for B, a necessary and sufficient condition for local identification is that

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

where $\mathbf{D}_K^+$ is the Moore-Penrose inverse of the $(K^2 \times \frac{1}{2}K(K+1))$ duplication matrix $\mathbf{D}_K$ (see Lütkepohl (2005, Proposition 9.4)). Although the unknown parameter matrix B appears in this condition, it is useful in practice because it will fail everywhere in the parameter space or be satisfied everywhere except on a set of Lebesgue measure zero. Thus, if a single admissible B matrix can be found which satisfies the restrictions in (2.5) and for which also the rank condition holds, then local identification is ensured almost everywhere in the parameter space. Thus, trying an arbitrary admissible B matrix is a possibility for checking identification.

As an example, consider a model for $K = 3$ variables. Assuming that all variables are $I(1)$ and the cointegrating rank $r = 2$, then there can be two transitory shocks and one permanent shock. If two transitory shocks are assumed, the permanent shock is identified in this situation without further assumptions because $K - r = 1$ and, hence, the number of additional restrictions is $(K-r)((K-r)-1)/2 = 0$. Moreover, only 1 $(= r(r-1)/2)$ further restriction is necessary to identify the two transitory shocks. Assuming a recursive structure for the two transitory shocks and placing the permanent shock first in the ε_t vector, the following restrictions are obtained:

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

In these matrices the asterisks denote unrestricted elements. The two zero columns in ΞB represent two independent restrictions only because ΞB has rank $K - r = 1$. A third restriction is placed on B. The way it is specified, the third shock does not have an instantaneous effect on the second variable. Hence, there are $K(K-1)/2 = 3$ independent restrictions in total and the structural innovations are locally just-identified. In this case, uniqueness can be obtained, for instance, by fixing the signs of the diagonal elements of B.

In this example, with two zero columns in ΞB, it is also easy to see that it does not suffice to impose a further restriction on this matrix to identify B locally. To disentangle the two transitory shocks, we have to impose a restriction on B. In fact, it is necessary to restrict an element in the last two columns of B.

In the standard B-model with three variables which does not take into account the cointegration structure, at least $\frac{1}{2}K(K-1) = 3$ restrictions are needed for identification. In contrast, in the present VECM case, assuming that $r = 2$ and that there are two transitory shocks, only one restriction is required because two columns of ΞB are zero. Thus, the long-run restrictions from the cointegration structure of the variables may help in the identification of shocks of interest. As another example consider a bivariate system with one cointegrating relation. No further restriction is required to identify the permanent and transitory shocks in this case, if, say, the first shock is allowed to have permanent effects and the second one can have transitory effects only. Further examples may be found in Breitung et al. (2004) and more discussion of partitioning the shocks in permanent and transitory ones is given in Gonzalo & Ng (2001) and Fisher & Huh (1999) among others.

2.2 Computation of Impulse Responses and Forecast Error Variance Decompositions

If the matrix B is uniquely specified, impulse responses can be computed easily from the structural form parameters. Rewriting the VECM (2.1) in levels VAR form as

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t,$$

where $A_1 = \alpha\beta' + I_K + \Gamma_1$, $A_i = \Gamma_i - \Gamma_{i-1}$, $i = 2, \ldots, p-1$, and $A_p = -\Gamma_{p-1}$, and computing matrices

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \ldots,$$

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$, the structural impulse response coefficients can be shown to be the elements of the matrices

$$\Theta_j = \Phi_j B, \quad j = 0, 1, 2, \ldots \tag{2.6}$$

(see Lütkepohl (2005) for details).

Forecast error variance decompositions are alternative tools for analyzing the dynamic interactions between the variables of a VAR model. Denoting by $\omega_{kj}(h)$ the percentage contribution of variable j to the h-step forecast error variance of variable k, it can be shown that

$$\omega_{kj}(h) = (\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2) \Big/ \sum_{j=1}^{K} (\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2),$$

where $\theta_{kj,l}$ is the kj-th element of Θ_l. Because these forecast error variance components depend on the structural impulse responses, they also require identified innovations, that is, a uniquely specified matrix B, for a meaningful interpretation.

In practice, the parameters of the VECM are unknown and have to be estimated from the given time series data. This issue will be considered next. Computing the impulse responses and forecast error variance components from estimated rather than known parameters gives estimates of these quantities. Some implications of working with estimated impulse responses will also be considered in the next section.

3 Estimation

If the lag order $p-1$ and the cointegrating rank r as well as structural identifying restrictions are given, estimation of a VECM can proceed by first estimating the reduced form parameters and then estimating B as described in the following.

3.1 Estimating the Reduced Form

The parameters of the reduced form VECM (2.1) can be estimated by the Johansen (1995) Gaussian maximum likelihood (ML) procedure. In presenting the estimators, the following notation will be used, where a sample of size T and p presample values are assumed to be available: $\Delta Y = [\Delta y_1, \ldots, \Delta y_T]$, $Y_{-1} = [y_0, \ldots, y_{T-1}]$, $U = [u_1, \ldots, u_T]$, $\Gamma = [\Gamma_1 : \cdots : \Gamma_{p-1}]$ and $X = [X_0, \ldots, X_{T-1}]$ with

$$X_{t-1} = \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}.$$

Using this notation, the VECM (2.1) can be written compactly as

$$\Delta Y = \alpha\beta' Y_{-1} + \Gamma X + U. \tag{3.1}$$

Given a specific matrix $\alpha\beta'$, the equationwise least squares estimator of Γ is easily seen to be

$$\widehat{\Gamma} = (\Delta Y - \alpha\beta' Y_{-1})X'(XX')^{-1}. \tag{3.2}$$

Substituting this matrix for Γ in (3.1) and rearranging terms gives

$$\Delta Y M = \alpha\beta' Y_{-1} M + \widehat{U}, \tag{3.3}$$

where $M = I - X'(XX')^{-1}X$. Estimators for α and β can be obtained by canonical correlation analysis (see Anderson (1984)) or, equivalently, a reduced rank regression based on the model (3.3). Following Johansen (1995), the estimator may be determined by defining

$$S_{00} = T^{-1}\Delta Y M \Delta Y', \quad S_{01} = T^{-1}\Delta Y M Y_{-1}', \quad S_{11} = T^{-1} Y_{-1} M Y_{-1}',$$

and solving the generalized eigenvalue problem

$$\det(\lambda S_{11} - S_{01}' S_{00}^{-1} S_{01}) = 0. \tag{3.4}$$

Denote the ordered eigenvalues by $\lambda_1 \geq \cdots \geq \lambda_K$ and the associated matrix of eigenvectors by $V = [b_1, \ldots, b_K]$. The generalized eigenvectors satisfy $\lambda_i S_{11} b_i = S_{01}' S_{00}^{-1} S_{01} b_i$ and they are assumed to be normalized such that $V'S_{11}V = I_K$. An estimator of β is then obtained by choosing

$$\widehat{\beta} = [b_1, \ldots, b_r]$$

and α is estimated as

$$\widehat{\alpha} = \Delta Y M Y_{-1}' \widehat{\beta} (\widehat{\beta}' Y_{-1} M Y_{-1}' \widehat{\beta})^{-1}, \tag{3.5}$$

that is, $\widehat{\alpha}$ may be regarded as the least squares estimator from the model

$$\Delta Y M = \alpha \widehat{\beta}' Y_{-1} M + \tilde{U}.$$

Using (3.2), a feasible estimator of Γ is $\widehat{\Gamma} = (\Delta Y - \widehat{\alpha}\widehat{\beta}' Y_{-1}) X'(XX')^{-1}$. Under Gaussian assumptions, these estimators are ML estimators conditional on the presample values (Johansen (1988, 1991, 1995)). They are consistent and jointly asymptotically normal under more general assumptions than Gaussianity,

$$\sqrt{T}\text{vec}([\widehat{\Gamma}_1 : \cdots : \widehat{\Gamma}_{p-1}] - [\Gamma_1 : \cdots : \Gamma_{p-1}]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\Gamma}}) \tag{3.6}$$

and

$$\sqrt{T}\text{vec}(\widehat{\alpha}\widehat{\beta}' - \alpha\beta') \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\alpha}\widehat{\beta}'}). \tag{3.7}$$

The asymptotic distribution of $\widehat{\Gamma}$ is nonsingular so that standard inference may be used for the short-term parameters Γ_j. On the other hand, the $(K^2 \times K^2)$ covariance matrix $\Sigma_{\widehat{\alpha}\widehat{\beta}'}$ can be shown to have reduced rank Kr. Hence, $\mathcal{N}(0, \Sigma_{\widehat{\alpha}\widehat{\beta}'})$ is a singular normal distribution if $r < K$. Moreover, the distribution in (3.7) provides an asymptotic distribution for the product matrix $\alpha\beta'$ and not for α and β separately.

Notice that for any nonsingular $(r \times r)$ matrix C, we may define $\alpha^* = \alpha C'$ and $\beta^* = \beta C^{-1}$ and get $\alpha\beta' = \alpha^*\beta^{*\prime}$. In order to estimate the matrices α and β consistently, it is necessary to impose identifying (uniqueness) restrictions. Without such restrictions, only the product $\alpha\beta'$ can be estimated consistently. An example of identifying restrictions which has received some attention in the literature, assumes that the first part of β is an identity matrix, $\beta' = [I_r : \beta_{(K-r)}']$, where $\beta_{(K-r)}$ is a $((K-r) \times r)$ matrix. For instance, for $r = 1$, this restriction amounts to normalizing the coefficient of the first variable to be one. By a suitable rearrangement of the variables it can always be ensured that the normalization $\beta' = [I_r : \beta_{(K-r)}']$ is possible.

Using this normalization, the parameters $\beta_{(K-r)}$ are identified so that inference becomes possible. Generally, if uniqueness restrictions are imposed, it can be shown that $T(\widehat{\beta} - \beta)$ and $\sqrt{T}(\widehat{\alpha} - \alpha)$ converge in distribution (Johansen (1995)). Hence, the estimator of β converges with the fast rate T and is therefore sometimes called *superconsistent*. In contrast, the estimator of α converges with the usual rate $\sqrt{T}$. It has an asymptotic normal distribution under general assumptions and, hence, it behaves like usual estimators in a model with stationary variables. In fact, its asymptotic distribution is the same that would be obtained when $\widehat{\beta}$ were replaced by the true cointegration matrix β and α were estimated by least squares from (3.3).

Although inference for α and β separately requires identifying restrictions, such constraints for α and β are not necessary for the impulse response analysis. In particular, the same matrices Ξ and Θ_j, $j = 0, 1, 2, \dots$, are obtained for any pair of $(K \times r)$ matrices α and β that gives rise to the same product matrix $\alpha\beta'$.

3.2 Estimating the Structural Parameters

Replacing the reduced form parameters by their ML estimators gives the concentrated log-likelihood function

$$\log l_c(B) = \text{constant} - \frac{T}{2}\log|B|^2 - \frac{T}{2}\text{tr}(B'^{-1}B^{-1}\widetilde{\Sigma}_u), \tag{3.8}$$

where $\widetilde{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$ and the $\widehat{u}_t$'s are the estimated reduced form residuals. Maximization of this function with respect to B subject to the structural restrictions has to be done by numerical methods because a closed form solution is usually not available.

Suppose the structural restrictions for a VECM are given in the form of linear restrictions as in (2.5). For computing the parameter estimates, we may replace Ξ by its reduced form ML estimator,

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

and the restricted ML estimator of B can be obtained by optimizing the concentrated log-likelihood function (3.8) with respect to B, subject to the restrictions in (2.5), with C_l replaced by

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

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

Under usual assumptions, the ML estimator of B, $\widehat{B}$ say, is consistent and asymptotically normal,

$$\sqrt{T}\text{vec}(\widehat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{B}}).$$

Thus, the t-ratios of elements with regular asymptotic distributions can be used for assessing the significance of individual parameters. The asymptotic distribution of $\widehat{B}$ is singular, however, because of the restrictions that have been imposed on B. Therefore F-tests will in general not be valid and have to be interpreted cautiously. Expressions for the covariance matrices of the asymptotic distributions in terms of the model parameters can be obtained by working out the corresponding information matrices (see Vlaar (2004)). For practical purposes, bootstrap methods are in common use for inference in this context.

Although in structural VAR and VECM analysis just-identified models are often used to minimize the risk of misspecification, the same approach can be used if there are over-identifying restrictions for B. In that case, $\widehat{B}\widehat{B}'$ will not be equal to the reduced form white noise covariance estimator $\widetilde{\Sigma}_u$, however. Still the estimator of B will be consistent and asymptotically normal under general conditions. The LR statistic,

$$\lambda_{LR} = T(\log|\widehat{B}\widehat{B}'| - \log|\widetilde{\Sigma}_u|), \tag{3.9}$$

can be used to check the over-identifying restrictions. It has an asymptotic χ^2-distribution with degrees of freedom equal to the number of over-identifying restrictions if the null hypothesis holds.

3.3 Estimation of Impulse Responses

The impulse responses are estimated by replacing all unknown quantities in (2.6) by estimators. Suppose the structural form coefficients are collected in a vector $\boldsymbol{\alpha}$ and denote its estimator by $\widehat{\boldsymbol{\alpha}}$. For inference purposes it is important to note that any specific impulse response coefficient θ is a (nonlinear) function of $\boldsymbol{\alpha}$ and it is estimated as

$$\widehat{\theta} = \theta(\widehat{\boldsymbol{\alpha}}). \tag{3.10}$$

If $\widehat{\boldsymbol{\alpha}}$ is asymptotically normal, that is,

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\alpha}}}), \tag{3.11}$$

then $\widehat{\theta}$ is also asymptotically normally distributed,

$$\sqrt{T}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\theta}), \tag{3.12}$$

and the variance of the asymptotic distribution is

$$\sigma_\theta^2 = \frac{\partial \theta}{\partial \boldsymbol{\alpha}'} \Sigma_{\widehat{\boldsymbol{\alpha}}} \frac{\partial \theta}{\partial \boldsymbol{\alpha}}. \tag{3.13}$$

Here $\partial\theta/\partial\boldsymbol{\alpha}$ denotes the vector of first order partial derivatives of θ with respect to the elements of $\boldsymbol{\alpha}$. The result (3.13) holds if σ_θ^2 is nonzero, which follows if $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ is nonsingular and $\partial\theta/\partial\boldsymbol{\alpha} \neq 0$. In general the covariance matrix $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ will not be nonsingular for cointegrated systems, however, because, for example, $\Sigma_{\widehat{\alpha}\widehat{\beta}'}$ is singular due to the superconsistency of $\widehat{\beta}$. Moreover, the impulse responses generally consist of sums of products of the VAR coefficients and, hence, the partial derivatives will also be sums of products of such coefficients. Therefore the partial derivatives will also usually be zero in parts of the parameter space. Thus, $\sigma_\theta^2 = 0$ may hold and, hence, $\widehat{\theta}$ may actually converge at a faster rate than $\sqrt{T}$ in parts of the parameter space (cf. Benkwitz, Lütkepohl & Neumann (2000)).

It was found, however, that even under ideal conditions where the asymptotic theory holds, it may not provide a good guide for small sample inference. Therefore bootstrap methods are often used to construct confidence intervals for impulse responses (e.g., Kilian (1998), Benkwitz, Lütkepohl & Wolters (2001)). In the present context, these methods have the additional advantage that they avoid deriving exlicit forms of the rather complicated analytical expressions for the asymptotic variances of the impulse response coefficients. Unfortunately, bootstrap methods generally do not overcome the problems due to zero variances in the asymptotic distributions of the impulse responses and they may provide confidence intervals which do not have the desired coverage level even asymptotically (see Benkwitz et al. (2000) for further discussion).

Although we have discussed the estimation problems in terms of impulse responses, similar problems arise for forecast error variance components. In fact, these quantities are proportions and they are therefore always between zero and one. In other words, they are bounded from below and above. Moreover, the boundary values are possible values as well. This feature makes inference even more delicate.

So far it was assumed that a model and identifying structural restrictions are given. In practice this is usually not the case. While the structural restrictions normally come from theoretical considerations or institutional knowledge, there is a range of statistical tools for specifying the reduced form of a VECM. These tools will be summarized briefly in the next section.

4 Model Specification and Validation

The general approach to structural VECM analysis is to specify a reduced form first and then impose structural restrictions that can be used in an impulse response analysis. To specify the reduced form VECM, the lag order and the cointegrating rank have to be chosen. Most procedures for specifying the latter quantity require that the lag order is already known whereas order selection can be done without prior knowledge of the cointegrating rank. Therefore lag order selection is typically based on a VAR process in levels without imposing a cointegration rank restriction. Standard model selection criteria of the form

$$\mathrm{Cr}(m) = \log\det(\widetilde{\Sigma}_u(m)) + c_T\varphi(m), \tag{4.1}$$

can be used for that purpose. Here $\widetilde{\Sigma}_u(m) = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$ is the residual covariance matrix estimator for a model with lag order m and $\varphi(m)$ is a function which penalizes large VAR orders. For instance, $\varphi(m)$ may represent the number of parameters which have to be estimated in a VAR(m) model. The quantity c_T is a sequence that depends on the sample size T. For example, for Akaike's AIC, $c_T = 2/T$ and for the popular Hannan-Quinn criterion, $c_T = 2\log\log T/T$. The term $\log\det(\widetilde{\Sigma}_u(m))$ measures the fit of a model with order m. It decreases (or at least does not increase) when m increases because there is no correction for degrees of freedom in the covariance matrix estimator. The criterion chosen by the analyst is evaluated for $m = 0, \dots, p_{\max}$, where $p_{\max}$ is a prespecified upper bound and the order p is estimated so as to minimize the criterion. Rewriting the levels VAR(p) model in VECM form, there are $p-1$ lagged differences that may be used in the next stage of the analysis, where the cointegrating rank is chosen.

Once the lag order is specified the cointegrating rank can be chosen by defining the matrix $\Pi = \alpha\beta'$ and testing a sequence of null hypotheses $H_0(0) : \mathrm{rk}(\Pi) = 0$, $H_0(1) : \mathrm{rk}(\Pi) = 1, \dots,$ $H_0(K-1) : \mathrm{rk}(\Pi) = K-1$ against the rank being greater than the one specified in the null hypothesis. The rank for which the null hypothesis cannot be rejected for the first time is then used in the next stages of the analysis. A range of test statistics is available for use in this testing sequence (see, e.g., Hubrich, Lütkepohl & Saikkonen (2001) for a recent survey). The most popular tests in applied work are Johansen's (1995) likelihood ratio tests. They are easy to compute because the Gaussian likelihood function is easy to maximize for any given cointegrating rank, as shown in Section 3.1.

When a reduced form model has been specified, a range of tools can be used for model checking. For example, tests for residual autocorrelation and structural stability may be used (see Lütkepohl (2005) for details). Finally, once a satisfactory reduced form is available, the structural restrictions may be imposed and the model can then be used for impulse response analysis.

5 Conclusions

In this article a brief overview of some important issues related to structural modelling based on VARs with cointegrated variables was given. Generally, using a standard VAR analysis, the impulse responses are the relevant tools for interpreting the relationships between the variables. Unfortunately, they are not unique and subject matter knowledge is required to specify those impulses and their associated responses which reflect the actual ongoings in the system of interest. It was discussed how the cointegration properties of the variables can help in specifying identifying restrictions properly. In particular, the cointegrating rank specifies the maximum number of transitory shocks in a system with cointegrated variables. This rank in turn can be determined by statistical procedures. As a final note it may be worth mentioning that the software JMulTi (Lütkepohl & Krätzig (2004)) provides easy access to the necessary computations for a structural VECM analysis.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

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.

Breitung, J., Brüggemann, R. & Lütkepohl, H. (2004). Structural vector autoregressive modeling and impulse responses, *in* H. Lütkepohl & M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 159–196.

Engle, R. F. & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing, *Econometrica* **55**: 251–276.

Fisher, L. A. & Huh, H. (1999). Weak exogeneity and long-run and contemporaneous identifying restrictions in VEC models, *Economics Letters* **63**: 159–165.

Gonzalo, J. & Ng, S. (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks, *Journal of Economic Dynamics & Control* **25**: 1527–1546.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**: 121–130.

Granger, C. W. J. & Newbold, P. (1974). Spurious regressions in eonometrics, *Journal of Econometrics* **2**: 111–120.

Hubrich, K., Lütkepohl, H. & Saikkonen, P. (2001). A review of systems cointegration tests, *Econometric Reviews* **20**: 247–318.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**: 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**: 1551–1581.

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.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.

Lütkepohl, H. & Krätzig, M. (eds) (2004). *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.

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

Vlaar, P. J. G. (2004). On the asymptotic distribution of impulse response functions with long-run restrictions, *Econometric Theory* **20**: 891–903.