Lütkepohl, Helmut

Working Paper

Vector autoregressions

SFB 373 Discussion Paper, No. 1999,4

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Lütkepohl, Helmut (1999) : Vector autoregressions, SFB 373 Discussion Paper, No. 1999,4, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10056014

This Version is available at:
https://hdl.handle.net/10419/61762

January 12, 1999

Vector Autoregressions

Helmut Lütkepohl[1]

Institut für Statistik und Ökonometrie
Humboldt–Universität zu Berlin
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: (+49-30) 2093-5718
Fax.: (+49-30) 2093-5712
email: luetke@wiwi.hu-berlin.de

1 Introduction

The last sixty years have witnessed a rapid development in the field of econometrics. In the 1940s and 50s the foundations were laid by the Cowles Commission researchers for analyzing econometric simultaneous equations models. Once the basic statistical theory was available many such models were constructed for empirical analysis. The parallel development of the computer technology in the 1950s and 60s has resulted in simultaneous equations models of increasing size in the hope that more detailed models would result in better approximations to the underlying data generation mechanisms. It turned out, however, that increasing the number of variables and equations of the models did not generally lead to improvements in the performance in terms of forecasting, for instance. In fact, in some forecast comparisons univariate time series models were found to be superior to large scale econometric models. One explanation of this failure of the latter models is their insufficient representation of the dynamic interactions in a system of variables.

The poor performance of standard macroeconometric models in some respects resulted in a critical assessment of econometric simultaneous equations modeling as summarized by Sims (1980) who advocated using vector autoregressive (VAR) models as alternatives. In these models often all variables are treated as a priori endogenous and allowance is made for rich dynamics. Restrictions are imposed to a large extent by statistical tools rather than by prior believes based on uncertain theoretical considerations. Although VAR models are by now standard instruments in econometric analyses it has become apparent that certain types of interpretations and economic investigations are not possible without incorporating nonstatistical a priori information. Therefore, so-called *structural* VAR models are now often

[1]I thank Jörg Breitung for helpful comments on an earlier draft of this chapter and the Deutsche Forschungsgemeinschaft, SFB 373, for financial support.

used in practice. Moreover, the invention of cointegration by Granger (1981) and Engle & Granger (1987) has resulted in specific parameterizations which support the analysis of the cointegration structure. The cointegrating relations are often interpreted as the connecting link to the relations derived from economic theory. Therefore they are of particular interest in an analysis of a set of time series variables.

In the following I will first discuss some of the related models which are now in common use. I will then consider estimation and specification issues in Sections 3 and 4, respectively. Possible uses of VAR models are presented in Section 5. Conclusions and extensions are considered in Section 6. Nowadays a number of books are available which treat modern developments in VAR modeling and dynamic econometric analysis more generally in some detail (e.g., Lütkepohl (1991), Banerjee, Dolado, Galbraith & Hendry (1993), Hamilton (1994), Hendry (1995), Johansen (1995), Hatanaka (1996)). Surveys of vector autoregressive modeling include Watson (1994) and Lütkepohl & Breitung (1997). In this chapter some more recent developments are also included.

2 VAR Models

2.1 Characteristics of Variables

The characteristics of the variables involved determine to some extent which model is a suitable representation of the data generation process (DGP). For instance, the trending properties of the variables and their seasonal fluctuations are of importance in setting up a suitable model. In the following a variable is called *integrated* of order d ($I(d)$) if stochastic trends or unit roots can be removed by differencing the variable d times. In the present chapter it is assumed that all variables are at most $I(1)$ if not otherwise stated. In other words, for any time series variable y_{kt} it is assumed that $\Delta y_{kt} \equiv y_{kt} - y_{k,t-1}$ has no stochastic trend. Note, however, that Δy_{kt} may still have deterministic components such as a polynomial trend and a seasonal component whereas seasonal unit roots are excluded. Note also that a variable without a stochastic trend or unit root is sometimes called $I(0)$. A set of $I(1)$ variables is called *cointegrated* if a linear combination exists which is $I(0)$. Occasionally it is convenient to consider systems with both $I(1)$ and $I(0)$ variables. In this case the concept of cointegration is extended by calling any linear combination which is $I(0)$ a cointegration relation although this terminology is not in the spirit of the original definition because it can result in a linear combination of $I(0)$ variables being called a cointegration relation.

As mentioned earlier, we allow for deterministic polynomial trends. For these terms we assume for convenience that they are at most linear. In other words, we exclude higher order polynomial trend terms. For practical purposes this assumption is not a severe limitation.

2.2 Alternative Models and Model Representations

Given a set of K time series variables $y_t = (y_{1t}, \ldots, y_{Kt})'$, the basic VAR model is of the form

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.1}$$

where $u_t = (u_{1t}, \ldots, u_{Kt})'$ is an unobservable zero mean independent white noise process with time invariant positive definite covariance matrix $E(u_t u_t') = \Sigma_u$ and the A_i are $(K \times K)$

coefficient matrices. This model is often briefly referred to as a VAR(p) process because the number of lags is p.

The process is *stable* if

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } |z| \leq 1. \tag{2.2}$$

Assuming that it has been initiated in the infinite past, it generates stationary time series which have time invariant means, variances and covariance structure. If the determinantal polynomial in (2.2) has a root for $z = 1$ (i.e., a unit root), then some or all of the variables are $I(1)$ and they may also be cointegrated. Thus, the present model is general enough to accommodate variables with stochastic trends. On the other hand, it is not the most suitable type of model if interest centers on the cointegration relations because they do not appear explicitly in the VAR version (2.1). They are more easily analyzed within the so-called *vector error correction model* (VECM)

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{2.3}$$

where $\Pi = -(I_K - A_1 - \cdots - A_p)$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ for $i = 1, \ldots, p-1$. This representation of the process is obtained from (2.1) by subtracting y_{t-1} from both sides and rearranging terms. Because Δy_t does not contain stochastic trends by our assumption that all variables can be at most $I(1)$, the term Πy_{t-1} is the only one which includes $I(1)$ variables. Hence, Πy_{t-1} must also be $I(0)$. Thus, it contains the cointegrating relations. The Γ_j ($j = 1, \ldots, p-1$) are often referred to as the short-term or short-run parameters while Πy_{t-1} is sometimes called long-run or long-term part. The model in (2.3) will be abbreviated as VECM(p) because p is the largest lag of the levels y_t that appear in the model. To distinguish the VECM from the VAR model the latter is sometimes called the levels version. Of course, it is also possible to determine the A_j levels parameter matrices from the coefficients of the VECM as $A_1 = \Gamma_1 + \Pi + I_K$, $A_i = \Gamma_i - \Gamma_{i-1}$ for $i = 2, \ldots, p-1$, and $A_p = -\Gamma_{p-1}$.

If the VAR(p) process has unit roots, that is, $\det(I_K - A_1 z - \cdots - A_p z^p) = 0$ for $z = 1$, the matrix Π is singular. Suppose it has rank r, that is, $\mathrm{rk}(\Pi) = r$. Then it is well-known that Π can be written as a product $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices with $\mathrm{rk}(\alpha) = \mathrm{rk}(\beta) = r$. Premultiplying an $I(0)$ vector by some matrix results again in an $I(0)$ process. Hence, premultiplying $\Pi y_{t-1} = \alpha\beta' y_{t-1}$ by $(\alpha'\alpha)^{-1}\alpha'$ shows that $\beta' y_{t-1}$ is $I(0)$ and, therefore, contains the cointegrating relations. Hence, there are $r = \mathrm{rk}(\Pi)$ linearly independent cointegrating relations among the components of y_t. The matrices α and β are not unique so that there are many possible β matrices which contain the cointegrating relations or linear transformations of them. Consequently, cointegrating relations with economic content cannot be extracted purely from the observed time series. Some nonsample information is required to identify them uniquely.

Special cases included in (2.3) are $I(0)$ processes for which $r = K$ and systems that have a stable VAR representation in first differences. In the latter case, $r = 0$ and the term Πy_{t-1} disappears in (2.3). These boundary cases do not represent cointegrated systems in the usual sense. There are also other cases where no cointegration in the original sense is present although the model (2.3) has a cointegrating rank strictly between 0 and K. Suppose, for instance, that all variables but one are $I(0)$ then the cointegrating rank is $K - 1$ although the $I(1)$ variable is not cointegrated with the other variables. Similarly, there could be $K - r$ unrelated $I(1)$ variables and r $I(0)$ components. Generally, for each $I(0)$ variable in

the system there can be a column in the matrix β with a unit in one position and zeros elsewhere. In these cases there is no cointegration in the strict sense. Still it is convenient to include these cases in the present framework because they can be accommodated easily as far as estimation and inference is concerned. Of course, the special properties of the variables may be important in the interpretation of a system and, hence, a different treatment of the special cases may be necessary in this respect.

In practice the basic models (2.1) and (2.3) are usually too restrictive to represent the main characteristics of the data. In particular, deterministic terms such as an intercept, a linear trend term or seasonal dummy variables may be required for a proper representation of the data. There are two ways to include deterministic terms. The first possibility is to represent the observed variables y_t as a sum of a deterministic term and a stochastic part,

$$y_t = \mu_t + x_t, \tag{2.4}$$

where μ_t is the deterministic part and x_t is a stochastic process which may have a VAR or VECM representation as in (2.1) or (2.3), that is, $x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t$ or $\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t$. In that case, if μ_t is a linear trend term, that is, $\mu_t = \mu_0 + \mu_1 t$, then y_t has a VAR(p) representation of the form

$$y_t = \nu_0 + \nu_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.5}$$

where $\nu_0 = -\Pi\mu_0 + (\sum_{j=1}^{p} jA_j)\mu_1$ and $\nu_1 = -\Pi\mu_1$. In other words, ν_0 and ν_1 satisfy a set of restrictions. Note, however, that if (2.5) is regarded as the basic model without restrictions for ν_i, $i = 0, 1$, the model can in principle generate quadratic trends if $I(1)$ variables are included, whereas in (2.4) with a deterministic term $\mu_t = \mu_0 + \mu_1 t$ a linear trend term is permitted only. The fact that in (2.4) a clear partitioning of the process in a deterministic and a stochastic component is available is sometimes advantageous in theoretical derivations. Also, in practice, it may be possible to subtract the deterministic term first and then focus the analysis on the stochastic part which usually contains the behavioural relations. Therefore this part is often of primary interest in econometric analyses. Of course, a VECM(p) representation equivalent to (2.5) also exists.

In practice, these representations with possibly additional deterministic terms may still not be general enough. At times one may wish to include stochastic exogenous variables on top of the deterministic part. A fairly general VECM form which includes all these terms is

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + CD_t + Bz_t + u_t, \tag{2.6}$$

where the z_t are exogenous variables, D_t contains all regressors associated with deterministic terms and C and B are parameter matrices.

Notice that different concepts of exogeneity have been considered in the literature (see Engle, Hendry & Richard (1983)). A set of variables z_t is called *weakly exogenous* for a parameter vector of interest, say θ, if estimating θ within a conditional model, conditional on z_t, does not entail a loss of information relative to estimating the vector in a full model which does not condition on z_t. Furthermore, z_t is called *strongly exogenous* if it is weakly exogenous for the parameters of the conditional model and forecasts of y_t can be made conditional on z_t without loss of forecast precision. Finally, z_t is called *super exogenous* for θ if z_t is weakly exogenous for θ and policy actions that affect the marginal process of z_t do not affect the parameters of the conditional process. Hence, weak, strong and super exogeneity are the relevant concepts for estimation, forecasting and policy analysis, respectively (Ericsson,

Hendry & Mizon (1998)). In the following the term exogeneity refers to the relevant concept for the respective context if no specific form of exogeneity is mentioned.

All the models we have presented so far are reduced form models in that they do not include instantaneous relations between the endogenous variables y_t. In practice it is often desirable to model the contemporaneous relations as well and therefore it is useful to consider a structural form

$$\Gamma_0^*\Delta y_t = \Pi^* y_{t-1} + \Gamma_1^*\Delta y_{t-1} + \cdots + \Gamma_{p-1}^*\Delta y_{t-p+1} + C^* D_t + B^* z_t + v_t, \tag{2.7}$$

where v_t is a $(K \times 1)$ zero mean white noise process with covariance matrix Σ_v and the Π^*, Γ_j^* $(j = 0, \ldots, p-1)$, C^* and B^* are structural form parameter matrices. The reduced form corresponding to the structural model (2.7) is given in (2.6) with $\Gamma_j = (\Gamma_0^*)^{-1}\Gamma_j^*$ $(j = 1, \ldots, p-1)$, $C = (\Gamma_0^*)^{-1}C^*$, $B = (\Gamma_0^*)^{-1}B^*$ and $u_t = (\Gamma_0^*)^{-1}v_t$. Of course, a number of restrictions are usually imposed on the general forms of our models. These restrictions are important at the estimation stage which will be discussed next.

3 Estimation

Since estimation of some of the special case models is particularly easy some of these cases will be considered in more detail in the following. We begin with the levels VAR representation (2.1) under the condition that no restrictions are imposed. Then estimation of the VECM (2.3) is treated and finally more general model variants are discussed.

3.1 Estimation of an Unrestricted VAR

Given a sample $y_1, \ldots, y_T$ and presample values $y_{-p+1}, \ldots, y_0$, the K equations of the VAR (2.1) may be estimated separately by least squares (LS) without loosing efficiency relative to generalized LS (GLS) approaches. In fact, in this case LS is identical to GLS as shown by Zellner (1962). Defining

$$Y = [y_1, \ldots, y_T] \quad \text{and} \quad Z = [Z_0, \ldots, Z_{T-1}], \quad \text{where} \quad Z_{t-1} = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix}, \tag{3.1}$$

the LS estimator of $A = [A_1 : \cdots : A_p]$ can be written compactly as

$$\hat{A} = [\hat{A}_1 : \cdots : \hat{A}_p] = YZ'(ZZ')^{-1}. \tag{3.2}$$

Under standard assumptions (see, e.g., Lütkepohl (1991)), the LS estimator $\hat{A}$ is consistent and asymptotically normally distributed,

$$\sqrt{T}\text{vec}(\hat{A} - A) \xrightarrow{d} N(0, \Sigma_{\hat{A}}) \quad \text{or, more intuitively,} \quad \text{vec}(\hat{A}) \overset{a}{\sim} N(\text{vec}(A), \Sigma_{\hat{A}}/T). \tag{3.3}$$

Here vec denotes the column stacking operator which stacks the columns of a matrix in a column vector, $\xrightarrow{d}$ signifies convergence in distribution and $\overset{a}{\sim}$ indicates 'asymptotically distributed as'. The covariance matrix of the asymptotic distribution is $\Sigma_{\hat{A}} = \text{plim}(ZZ'/T)^{-1} \otimes \Sigma_u$ so that an even more intuitive albeit imprecise way of writing the result in (3.3) is

$$\text{vec}(\hat{A}) \approx N(\text{vec}(A), (ZZ')^{-1} \otimes \Sigma_u).$$

Although this result also holds for $I(1)$ cointegrated systems (see Park & Phillips (1988, 1989), Sims, Stock & Watson (1990), Lütkepohl (1991, Chapter 11)) it is important to note that in this case the covariance matrix $\Sigma_{\hat{A}}$ is singular whereas it is nonsingular in the usual $I(0)$ case. In other words, if there are integrated or cointegrated variables, some estimated coefficients or linear combinations of coefficients converge with a faster rate than $T^{1/2}$. Therefore, the usual t-, χ^2- and F-tests for inference regarding the VAR parameters, may not be valid in this case as shown, e.g., by Toda & Phillips (1993). As an example consider a univariate first order autoregressive process $y_t = \alpha y_{t-1} + u_t$. If y_t is $I(1)$ and, hence, $\alpha = 1$, the LS estimator $\hat{\alpha}$ of α has a nonstandard limiting distribution. As is well-known from the unit root literature (e.g., Fuller (1976), Dickey & Fuller (1979)), the quantity $\sqrt{T}(\hat{\alpha}-\alpha)$ converges to zero in probability, that is, the limiting distribution has zero variance and is degenerate, whereas $T(\hat{\alpha}-\alpha)$ has a nondegenerate nonnormal limiting distribution. Although inference problems may arise in VAR models with $I(1)$ variables, there are also many unproblematic cases. Dolado & Lütkepohl (1996) show that if all variables are $I(1)$ or $I(0)$ and if a null hypothesis is considered which does not restrict elements of each of the A_i $(i = 1, \ldots, p)$ the usual tests have their standard asymptotic properties. For example, if the VAR order $p \geq 2$, the t-ratios have their usual asymptotic standard normal distribution (see also Toda & Yamamoto (1995) for a related result).

If the white noise process u_t is normally distributed (Gaussian) and the process y_t is $I(0)$, then the LS estimator in (3.2) is identical to the maximum likelihood (ML) estimator conditional on the initial values. It is also straightforward to include deterministic terms such as polynomial trends in the model (2.1). In this case the asymptotic properties of the VAR coefficients remain essentially the same as in the case without deterministic terms (Sims, Stock & Watson (1990)).

The covariance matrix Σ_u may be estimated in the usual way. Denoting by $\hat{u}_t$ the LS residuals, that is, $\hat{u}_t = y_t - \hat{A}Z_{t-1}$,

$$\hat{\Sigma}_u = \frac{1}{T - Kp}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t' \quad \text{and} \quad \tilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t' \tag{3.4}$$

are possible estimators. Both estimators are consistent and asymptotically normally distributed independent of $\hat{A}$, that is, $\sqrt{T}(\hat{\Sigma}_u - \Sigma_u)$ and $\sqrt{T}(\tilde{\Sigma}_u - \Sigma_u)$ have asymptotic normal distributions independent of $\hat{A}$ if sufficient moment conditions are imposed (see Lütkepohl (1991) and Lütkepohl & Saikkonen (1997a)). These properties are convenient for inference purposes.

3.2 Estimation of a VECM

If the cointegrating rank of the system under consideration is known and one wishes to impose a corresponding restriction, working with the VECM form (2.3) is convenient. In deriving estimators for the parameters of (2.3) the following additional notation is used:

$$\Delta Y = [\Delta y_1, \ldots, \Delta y_T], \quad Y_{-1} = [y_0, \ldots, y_{T-1}], \quad U = [u_1, \ldots, u_T], \quad \Gamma = [\Gamma_1 : \cdots : \Gamma_{p-1}]$$

$$\text{and} \quad X = [X_0, \ldots, X_{T-1}] \quad \text{with} \quad X_{t-1} = \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}. \tag{3.5}$$

For a sample with T observations and p presample values, the VECM (2.3) can now be written compactly as

$$\Delta Y = \Pi Y_{-1} + \Gamma X + U. \tag{3.6}$$

Given a specific matrix Π, the LS estimator of Γ is easily seen to be

$$\hat{\Gamma} = (\Delta Y - \Pi Y_{-1})X'(XX')^{-1}. \tag{3.7}$$

Substituting in (3.6) gives

$$\Delta Y M = \Pi Y_{-1} M + \hat{U},$$

where $M = I - X'(XX')^{-1}X$. For a given r, $0 < r < K$, an estimator $\hat{\Pi}$ of Π with $\mathrm{rk}(\hat{\Pi}) = r$ can be obtained by a canonical correlation analysis or, equivalently, a reduced rank regression based on the latter model. Following Johansen (1995), the estimator may be determined by defining

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

and solving the eigenvalue problem

$$\det(\lambda S_{11} - S'_{01}S_{00}^{-1}S_{01}) = 0. \tag{3.8}$$

Let the ordered eigenvalues be $\lambda_1 \geq \cdots \geq \lambda_K$ with corresponding eigenvectors $V = [v_1, \ldots, v_K]$ satisfying $\lambda_i S_{11} v_i = S'_{01}S_{00}^{-1}S_{01}v_i$ and normalized such that $V'S_{11}V = I_K$. Choose

$$\hat{\beta} = [v_1, \ldots, v_r]$$

and

$$\hat{\alpha} = \Delta Y M Y'_{-1}\hat{\beta}(\hat{\beta}'Y_{-1}MY'_{-1}\hat{\beta})^{-1}$$

that is, $\hat{\alpha}$ is obtained as LS estimator from the model

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

An estimator of Π is then $\hat{\Pi} = \hat{\alpha}\hat{\beta}'$. Using (3.7), a feasible estimator of Γ is obtained as $\hat{\Gamma} = (\Delta Y - \hat{\Pi}Y_{-1})X'(XX')^{-1}$. Under Gaussian assumptions these estimators are ML estimators conditional on the presample values (Johansen (1988, 1991)).

In this approach the parameter estimator $\hat{\beta}$ is made unique by the normalization of the eigenvectors and $\hat{\alpha}$ is adjusted accordingly. However, these are not econometric identification restrictions. Therefore only the cointegration space but not the cointegration parameters are estimated consistently in this approach. In order to estimate the matrices α and β consistently, it is necessary to impose identifying restrictions. Without such restrictions only the product $\alpha\beta' = \Pi$ 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, $\beta' = [I_r : \beta'_1]$, where β_1 is a $((K-r)\times r)$ matrix. For $r = 1$, this restriction amounts to normalizing the coefficient of the first variable. If uniqueness restrictions are imposed it can be shown that $T(\hat{\beta} - \beta)$ and $\sqrt{T}(\hat{\alpha} - \alpha)$ 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 $\sqrt{T}$.

The estimators of Γ and Π are consistent and asymptotically normal under general assumptions,

$$\sqrt{T}\mathrm{vec}([\hat{\Gamma}_1 : \cdots : \hat{\Gamma}_{p-1}] - [\Gamma_1 : \cdots : \Gamma_{p-1}]) \xrightarrow{d} N(0, \Sigma_{\hat{\Gamma}})$$

and

$$\sqrt{T}\mathrm{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} N(0, \Sigma_{\hat{\Pi}}).$$

Here the asymptotic distribution of $\hat{\Gamma}$ 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 $\Sigma_{\hat{\Pi}}$ can be shown to have rank Kr and is therefore singular if $r < K$. This result is due to two forces. On the one hand, imposing the rank constraint in estimating Π restricts the parameter space and on the other hand, Π involves the cointegrating relations which are estimated superconsistently.

It is perhaps interesting to note that an estimator of A can be computed via the estimates of Π and Γ. That estimator has the advantage of imposing the cointegrating restrictions on the levels version of the VAR process. However, its asymptotic distribution is the same as in (3.3) where no restrictions have been imposed in estimating A. Computing the covariance matrix estimator $\tilde{\Sigma}_u$ from the residuals of the VECM estimation the asymptotic distribution is the same as for the levels VAR form treated in the previous subsection and it is asymptotically independent of $\hat{\Pi}$ and $\hat{\Gamma}$. Extensions of these results for the case where the true DGP is an infinite order VAR process are considered by Saikkonen (1992) and Saikkonen & Lütkepohl (1996).

3.3 Estimation of Restricted Models and Structural Forms

Efficient estimation of a general structural form model such as (2.7) with restrictions on the parameter matrices is more complicated. Of course, identifying restrictions are necessary for consistent estimation. In practice, various overidentifying restrictions are usually available, typically in the form of zero restrictions on Γ_j^* $(j = 0, \ldots, p-1)$, C^* and B^*. In addition there may be a rank restriction for Π^* given by the number of cointegrating relations. Alternatively, Π^* may be replaced by the product $\alpha^*\beta^{*\prime}$, if identifying restrictions are available for the cointegrating relations and/or the loading matrix α^*. Restrictions for α^* are typically zero constraints, meaning that some cointegrating relations are excluded from some of the equations of the system. In some cases it is possible to estimate β^* in a first stage, for example, using a reduced form procedure which ignores some or all of the structural restrictions on the short-term parameters. Let the estimator be $\hat{\beta}^*$. Because the estimators of the cointegrating parameters converge at a better rate than the estimators of the short-term parameters they may be treated as fixed in a second stage procedure for the structural form. In other words, a systems estimation procedure may be applied to

$$\Gamma_0^*\Delta y_t = \alpha^*\hat{\beta}^{*\prime} y_{t-1} + \Gamma_1^*\Delta y_{t-1} + \cdots + \Gamma_{p-1}^*\Delta y_{t-p+1} + C^*D_t + B^*z_t + \hat{v}_t. \tag{3.9}$$

If only exclusion restrictions are imposed on the parameter matrices in this form, standard three stage LS or similar methods may be applied which result in estimators of the short-term parameters with the usual asymptotic properties.

Some care is necessary with respect to the treatment of exogenous and deterministic variables. If all exogenous variables are $I(0)$, no problems arise and parameter estimators with usual properties are obtained. If z_t contains $I(1)$ variables, however, the properties of the estimators depend on the cointegration properties of z_t. In particular, cointegration between exogenous and endogenous variables has to be taken into account appropriately (see, e.g., Boswijk (1995)). Important results on estimating models with integrated variables are also due to Phillips and his co-workers (e.g., Phillips (1987, 1991), Phillips & Durlauf (1986), Phillips & Hansen (1990), Phillips & Loretan (1991)).

If deterministic variables are to be included in the cointegration relations this requires a suitable reparameterization of the model. Such reparameterizations for intercepts and linear trend terms are presented in Section 4.3, where tests for the cointegrating rank are discussed. In that context a proper treatment of deterministic terms is of particular importance. Therefore a more detailed discussion is deferred to Section 4.3. In a subsequent analysis of the model the parameters of the deterministic terms are often of minor interest and therefore the properties of the corresponding estimators are not treated in detail here (see, however, Sims, Stock & Watson (1990)).

4 Model Specification and Model Checking

4.1 Testing for the Model Order and Exclusion Restrictions

Unrestricted VAR models usually involve a substantial number of parameters which in turn results in rather imprecise estimators. Therefore it is desirable to impose restrictions that reduce the dimensionality of the parameter space. Such restrictions may be based on economic theory or other nonsample information and on statistical procedures. Of course, for structural models nonsample information is required for imposing identifying constraints. On top of that there may be further overidentifying constraints on the basis of a priori knowledge.

Tests are common statistical procedures for detecting possible restrictions. For example, t-ratios and F-tests are available for this purpose. These tests retain their usual asymptotic properties if they are applied to the short-run parameters in a VECM whereas problems may arise in the levels VAR representation as explained in the previous section. A particular set of restrictions where such problems occur is discussed in more detail in Section 5.1. In case of doubt it may be preferable to work on the VECM form.

In practice, it is not uncommon to start from a model with some prespecified maximum lag length, say $p_{\max}$, and apply tests sequentially, eliminating one or more variables in each step until a relatively parsimonious representation with significant parameter estimates has been found. For example, in a VECM a sequence of null hypotheses $H_0 : \Gamma_{p_{\max}-1} = 0$, $H_0 : \Gamma_{p_{\max}-2} = 0$ etc. may be tested until the null hypothesis is rejected. Similarly, single coefficients in individual equations may be tested. Before such a procedure can be used, a decision on $p_{\max}$ has to be made. Occasionally this quantity is chosen by some theoretical or institutional argument. For instance, one may want to include lags of at least one year so that four lags have to be included for quarterly data and twelve lags may be used for a monthly model. An inappropriate choice of $p_{\max}$ may not be very severe in some respect because if the order is chosen too small such a problem may be discovered later when the final model is subjected to a series of specification tests (see Section 4.4). On the other hand, too large a value of $p_{\max}$ may be problematic due to its impact on the overall error probability of a sequential procedure. If a very large order $p_{\max}$ is used, a long sequence of tests may be necessary which will have an impact on the overall Type I error of the testing sequence, that is, the choice of $p_{\max}$ will have an impact on the probability of an inadequate selection of p.

Of course, it is also possible that the actual DGP does not have a finite order VAR representation. Ng & Perron (1995) consider some consequences for choosing the lag order by sequential testing procedures in univariate models in this context. Instead of sequential

tests one may alternatively choose the lag length or determine exclusion restrictions by model selection procedures which will be discussed next.

4.2 Determining the Autoregressive Order by Model Selection Criteria

Because the cointegrating rank r is usually unknown when the choice of p is made, it is useful to focus on the VAR form (2.1) at this stage. Various model selection criteria are available that can be used in this context. The general approach is to fit VAR(m) models with orders $m = 0, \ldots, p_{\max}$ and choose an estimator of the order p which minimizes some criterion.

Many of the criteria in current use have the general form

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

where $\det(\cdot)$ denotes the determinant, log is the natural logarithm, $\tilde{\Sigma}_u(m) = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ is the residual covariance matrix estimator for a model of order m, c_T is a sequence indexed by the sample size, 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 term $\log\det(\tilde{\Sigma}_u(m))$ measures the fit of a model with order m. Since there is no correction for degrees of freedom in the covariance matrix estimator the log determinant decreases (or at least does not increase) when m increases. Note that the sample size is assumed to be held constant and, hence, the number of presample values set aside for estimation is determined by the maximum order $p_{\max}$. The estimator $\hat{p}$ of p is chosen to be the order which minimizes the criterion (4.1) so that the two terms in the sum on the right hand side of (4.1) are balanced optimally for $Cr(\hat{p})$.

The following criteria are popular in applied work:

$$\mathrm{AIC}(m) = \log\det(\tilde{\Sigma}_u(m)) + \frac{2}{T}mK^2, \quad \text{where } \varphi(m) = mK^2 \text{ and } c_T = \frac{2}{T},$$

(see Akaike (1973, 1974)),

$$\mathrm{HQ}(m) = \log\det(\tilde{\Sigma}_u(m)) + \frac{2\log\log T}{T}mK^2, \quad \text{where } \varphi(m) = mK^2 \text{ and } c_T = \frac{2\log\log T}{T},$$

proposed by Hannan & Quinn (1979) and Quinn (1980) and

$$\mathrm{SC}(m) = \log\det(\tilde{\Sigma}_u(m)) + \frac{\log T}{T}mK^2, \quad \text{where } \varphi(m) = mK^2 \text{ and } c_T = \frac{\log T}{T},$$

due to Schwarz (1978) and Rissanen (1978). The AIC criterion asymptotically overestimates the order with positive probability whereas the last two criteria estimate the order consistently (plim $\hat{p} = p$ or $\hat{p} \to p$ a.s.) under quite general conditions, if the actual DGP has a finite VAR order and the maximum order $p_{\max}$ is larger than the true order. These results not only hold for $I(0)$ processes but also for $I(1)$ processes with cointegrated variables (Paulsen (1984)). Denoting the orders selected by the three criteria by $\hat{p}(\mathrm{AIC}), \hat{p}(\mathrm{HQ})$ and $\hat{p}(\mathrm{SC})$, respectively, the following relations hold even in small samples of fixed size $T \geq 16$ (see Lütkepohl (1991, Chapters 4 and 11)):

$$\hat{p}(\mathrm{SC}) \leq \hat{p}(\mathrm{HQ}) \leq \hat{p}(\mathrm{AIC}).$$

Model selection criteria may also be used for identifying single coefficients that may be replaced by zero or other exclusion restrictions. After a model has been set up, a series of checks may be employed to confirm the model's adequacy. Some such checks will be mentioned briefly in a subsequent section. Before that issue is taken up, procedures for specifying the cointegrating rank will be reviewed.

4.3 Specifying the Cointegrating Rank

In practice, the cointegrating rank r is also usually unknown. It is commonly determined by a sequential testing procedure based on likelihood ratio (LR) type tests. Because for a given cointegrating rank Gaussian ML estimates for the unrestricted reduced form VECM are easy to compute, as shown in Section 3.2, LR test statistics are also easily available. The following sequence of hypotheses may be considered:

$$H_0(r_0) : \mathrm{rk}(\Pi) = r_0 \quad \text{versus} \quad H_1(r_0) : \mathrm{rk}(\Pi) > r_0, \qquad r_0 = 0, \ldots, K-1. \tag{4.2}$$

The testing sequence terminates if the null hypothesis cannot be rejected for the first time. If the first null hypothesis, $H_0(0)$, cannot be rejected, a VAR process in first differences is considered. At the other end, if all the null hypotheses can be rejected, the process is assumed to be $I(0)$ in levels.

Although, under Gaussian assumptions, LR tests can be used here it turns out that the limiting distribution of the LR statistic under $H_0(r_0)$ is nonstandard. It depends on the difference $K - r_0$ and on the deterministic terms included in the DGP. In particular, the deterministic trend terms in the DGP have an impact on the null distribution of the LR tests. Therefore LR type tests have been derived under different assumptions regarding the deterministic trend parameters. Fortunately, the limiting null distributions do not depend on the short-term dynamics and, hence, critical values for LR type tests have been tabulated for different values of $K - r_0$ under alternative assumptions for deterministic trend terms.

In this context it turns out that the model (2.4), where the deterministic and stochastic parts are separated, is a convenient point of departure. Therefore we consider the model

$$y_t = \mu_0 + \mu_1 t + x_t \tag{4.3}$$

with

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t. \tag{4.4}$$

It is easy to see that the process y_t has a VECM representation

$$\begin{aligned} \Delta y_t &= \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t \\ &= \nu + [\Pi : \nu_1] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t \\ &= \nu + \Pi^+ y_{t-1}^+ + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t, \end{aligned} \tag{4.5}$$

where ν_0 and ν_1 are as defined below (2.5), $\nu = \nu_0 + \nu_1$, $\Pi^+ = [\Pi : \nu_1]$ and $y_{t-1}^+ = [y_{t-1}' : t-1]'$. Depending on the assumptions for μ_0 and μ_1 different tests can be obtained in this framework. An overview is given in Table 1. In the following the different cases will be discussed briefly.

Table 1. Models and LR Type Tests.

Assumption for deterministic term	Model	Reference
$\mu_0 = \mu_1 = 0$	$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Johansen (1988, 1995)
μ_0 arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Johansen (1991, 1995)
$\mu_1 = 0$	$\Delta y_t = [\Pi : \nu_0] \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Johansen & Juselius (1990)
	$\Delta y_t = \Pi(y_{t-1} - \tilde{\mu}_0) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Saikkonen & Luukkonen (1997)
μ_0 arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Johansen (1995)
$\mu_1 \neq 0, \beta'\mu_1 = 0$	$\Delta y_t - \hat{\mu}_1 = \Pi(y_{t-1} - \hat{\mu}_0) + \sum_{j=1}^{p-1} \Gamma_j (\Delta y_{t-j} - \hat{\mu}_1) + u_t$	Saikkonen & Lütkepohl (1998)
μ_0, μ_1 arbitrary	$\Delta y_t = \nu + [\Pi : \nu_1] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Johansen (1992, 1994, 1995)
	$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t$	Perron & Campbell (1993)
	$\Delta y_t - \hat{\mu}_1 = \Pi(y_{t-1} - \hat{\mu}_0 - \hat{\mu}_1(t-1))$	Saikkonen & Lütkepohl (1997)
	$+ \sum_{j=1}^{p-1} \Gamma_j (\Delta y_{t-j} - \hat{\mu}_1) + u_t$	Lütkepohl & Saikkonen (1999)

Case 1: $\mu_0 = \mu_1 = 0$

Although this case, where $y_t = x_t$, is not important for applied work because a zero mean term can rarely be assumed, it is still useful to consider it first. It is particularly easy to derive LR tests for the rank of Π under this assumption. The LR statistic can be obtained by estimating

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{4.6}$$

subject to $\mathrm{rk}(\Pi) = r_0$ and $\mathrm{rk}(\Pi) = K$ as discussed in Section 3.2. It can be shown that for a sample $y_1, \ldots, y_T$ and presample values $y_{-p+1}, \ldots, y_0$ the LR test statistics can in fact be obtained from the ordered generalized eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ from (3.8). Johansen's (1988, 1995) LR *trace statistic* for testing (4.2) is given by

$$LR(r_0) = -T \sum_{j=r_0+1}^{K} \log(1 - \lambda_j). \tag{4.7}$$

As mentioned earlier, the limiting distribution under the null hypothesis is nonstandard and depends on the dimension of the process and the cointegrating rank r_0 or, more precisely, on the difference $K - r_0$. Critical values are tabulated, for instance, in Johansen (1988, Table 1; 1995, Table 15.1). It may be worth noting that the asymptotic distribution remains valid if the u_t are not Gaussian, but have other suitable distributions. For the other cases listed in Table 1 and briefly discussed in the following, the test statistics can be computed analogously by suitable modifications of the quantities in (3.8).

Case 2: μ_0 **arbitrary,** $\mu_1 = 0$

In this case a nonzero mean term is accommodated whereas a deterministic linear trend term is excluded by assumption. Three variants of LR type tests have been considered in the literature for this situation plus a number of asymptotically equivalent modifications. As can be seen from Table 1, the three statistics can be obtained easily from VECMs. The first test is obtained by dropping the $\nu_1 t$ term in (4.5) and estimating the intercept term in the VECM in unrestricted form and hence, the estimated model may generate

linear trends. The second test enforces the restriction that there is no linear deterministic trend in computing the test statistic by absorbing the intercept $\nu = -\Pi\mu_0$ in (4.5) into the cointegration relations. Finally, in the third test the mean term μ_0 is estimated in a first step and is subtracted from y_t. Then the estimation procedure with rank restriction for Π is applied to (4.4) with x_t replaced by $\tilde{x}_t = y_t - \hat{\mu}_0$. A suitable estimator $\hat{\mu}_0$ is proposed by Saikkonen & Luukkonen (1997) who also show that the asymptotic distribution of the resulting test statistic under the null hypothesis is the same as that of the LR test for the case $\mu_0 = \mu_1 = 0$. It is shown in Saikkonen & Lütkepohl (1999) that the latter test can have considerably more local power than the other two LR tests. Thus, based on local power it is the first choice if $\mu_1 = 0$.

Case 3: μ_0 arbitrary, $\mu_1 \neq 0$, $\beta'\mu_1 = 0$

In this case at least one of the variables has a deterministic linear trend because $\mu_1 \neq 0$, whereas the cointegration relations do not have a linear trend due to the constraint $\beta'\mu_1 = 0$. The resulting tests are perhaps the most frequently used ones for determining the cointegrating rank in applied work. It may be worth emphasizing, however, that for the $(K \times r)$ matrix β to satisfy $\beta'\mu_1 = 0$, $\mu_1 \neq 0$ implies that $r < K$. Hence, if a trend is known to be present then it should also be allowed for under the alternative and consequently even under the alternative the rank must be smaller than K. In other words, in the present setting only tests of null hypotheses $\mathrm{rk}(\Pi) = r_0 < K - 1$ make sense. This result is a consequence of the fact that a linear trend is assumed in at least one of the variables ($\mu_1 \neq 0$) whereas a stable model with an intercept cannot generate a linear trend.

Both test statistics which have been proposed for the presently considered case can be obtained from VECMs with rank restriction for Π, as can be seen in Table 1. The first test uses the same intercept model as the first test for the case $\mu_1 = 0$. In the present case the asymptotic properties are different, however (see Johansen (1995)). The second test for the presently considered situation was proposed by Saikkonen & Lütkepohl (1998). In this case the mean and trend parameters are estimated in a first step by a feasible generalized LS procedure, the trend is subtracted from y_t to yield $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t$ and then the test statistic is computed via the VECM (4.4) with x_t replaced by $\hat{x}_t$, using $\Delta\hat{x}_t = \Delta y_t - \hat{\mu}_1$ and noting that $\Pi\mu_1 = 0$. Again it turns out that trend adjusting first and then performing the test may result in considerable gains in local power (see Saikkonen & Lütkepohl (1998)). The null distributions are tabulated in the references given in Table 1.

Case 4: μ_0 and μ_1 arbitrary

In this case both the variables and the cointegrating relations may have a deterministic linear trend. Three different LR type tests and some asymptotically equivalent relatives have been proposed for this situation. The corresponding models are also listed in Table 1. Again, all test statistics can be obtained conveniently via the VECMs using the techniques of Section 3.2. The first model is set up in such a way so as to impose the linearity of the trend term. The second model includes the trend term in unrestricted form. As mentioned earlier, in principle such a model can generate quadratic trends. Of course, such trends are excluded here by assumption, that is, the ν_i, $i = 0, 1$, must obey appropriate restrictions. These restrictions are not imposed in the estimation which leads to the Perron-Campbell test statistic. Finally, the last test in Table 1 is again based on prior trend adjustment and

estimation of the resulting VECM for the trend adjusted variables. The trend parameters are again estimated in a first step by a generalized LS procedure. Critical values for all these tests may be found in the references given in Table 1. In a simulation comparison Lütkepohl & Saikkonen (1999) found that none of the three tests is uniformly superior to its competitors.

Further issues

A comprehensive survey of the properties of LR type tests for the cointegrating rank as well as a substantial number of other tests that have been proposed in the literature is given by Hubrich, Lütkepohl & Saikkonen (1998). We refer the interested reader to that article for further details. Here we will only add a few specific remarks.

Instead of the pair of hypotheses in (4.2) one may alternatively test $H_0(r_0) : \text{rk}(\Pi) = r_0$ versus $H_1^*(r_0) : \text{rk}(\Pi) = r_0 + 1$. LR tests for this pair of hypotheses were also pioneered by Johansen (1988, 1991) and are known as *maximum eigenvalue tests*. They are based on a statistic

$$LR_{\max}(r_0) = -T \log(1 - \lambda_{r_0+1})$$

and can be applied for all the different cases listed in Table 1. They also have nonstandard limiting distributions. Critical values can be found in the literature cited in the foregoing.

For univariate processes ($K = 1$) the testing sequences reduce to a test of $H_0 : r = 0$ against $H_1 : r = 1$. In other words, the null hypothesis of the process being $I(1)$ is tested against the $I(0)$ alternative. All the tests can be generalized to this situation except those for the case $\beta'\mu_1 = 0$ because the latter tests require that the alternative is at most $r = K-1$ which can obviously not hold for $K = 1$. LR tests corresponding to the other cases were proposed by Dickey & Fuller (1979) and Fuller (1976). They are very popular in applied work and are known as augmented Dickey-Fuller (ADF) tests.

The limiting distributions of the LR statistics are not only valid for normally distributed (Gaussian) processes but also under more general distributional assumptions even if the LR statistics are computed under Gaussian assumptions. In that situation these tests are, of course, just pseudo LR tests. Saikkonen & Luukkonen (1997) show that some of the tests (based on finite order VAR processes) remain asymptotically valid even if the true DGP has an infinite VAR order. This result is of interest because in practice tests for unit roots and cointegration are usually applied to the univariate series or subsystems first to determine the order of integration for the individual variables or the cointegrating properties of a subset of variables. However, if the full system of variables is driven by a finite order VAR process, then the generating process of the individual variables may be of infinite order autoregressive type (see Lütkepohl (1991, Sec. 6.6)). Hence, for the sake of consistency it is reassuring to know that the tests remain valid for this case. Lütkepohl & Saikkonen (1997b) analyze this situation in more detail. In particular, they consider the impact of lag length selection in this context.

There is a notable difference between the asymptotic properties of the tests and their actual performance for samples of the size typically available in economics. A large scale simulation study comparing the small sample properties of many of the tests is also performed by Hubrich, Lütkepohl & Saikkonen (1998).

Instead of the sequential testing procedures model selection criteria may be used for determining the cointegrating rank. This possibility is considered, for instance, by Lütkepohl & Poskitt (1998).

4.4 Model Checking

Once a model has been specified and estimated its adequacy is usually checked with a range of tests and other statistical procedures. Many of these model checking tools are based on the residuals of the final model. Some of them are applied to the residuals of individual equations and others are based on the full residual vectors. Examples of specification checking tools are visual inspection of the plots of the residuals and their autocorrelations. In addition, autocorrelations of squared residuals may be considered to check for possible autoregressive conditional heteroscedasticity (ARCH). Although it may be quite insightful to inspect the autocorrelations visually, formal statistical tests for remaining residual autocorrelation should also be applied. Such tests are often based on LM (Lagrange Multiplier) or Portmanteau statistics. Moreover, normality tests of the Lomnicki-Jarque-Bera type may be applied to the residuals (see, e.g., Lütkepohl (1991), Doornik & Hendry (1997)).

There are also procedures for checking the stability and possible nonlinearity of a model (Granger & Teräsvirta (1993)). These procedures are used for detecting potential structural shifts during the sample period. They range from performing prediction tests as discussed in Lütkepohl (1991) to computing recursive residuals (Doornik & Hendry (1997)) or applying CUSUM type tests (Krämer, Ploberger & Alt (1988)) or recursive tests for cointegration (Hansen & Johansen (1993)). In addition out-of-sample forecasts are sometimes used for model checking when new data become available. If rival models for the same economic relations are available, encompassing tests may be applied to compare them (Hendry (1995)). Although the significance of individual coefficients or groups of parameters is often investigated at the model specification stage, these tools are also available for model checking. For a more detailed discussion of model checking see Doornik & Hendry (1997).

If model defects such as residual autocorrelation or ARCH effects are detected at the checking stage this is usually regarded as an indication of the model being a poor representation of the DGP and efforts are made to find a better representation by adding other variables or lags to the model, by including nonlinear terms or changing the functional form, by modifying the sampling period or getting other data.

5 Uses of Vector Autoregressive Models

When an adequate model for the DGP of a system of variables has been found it may be used for forecasting and economic analysis. Different tools have been proposed for the latter purpose. For instance, there has been an extensive discussion of how to analyze causal relations between the variables of a system of interest. In this section forecasting VAR processes will be discussed first and then the concept of Granger-causality will be introduced which is based on forecast performance. It has received considerable attention in the theoretical and empirical literature. In Subsection 5.3 impulse responses are considered. They may also be regarded as instruments for analyzing causal relations between variables. Finally, forecast error variance decompositions and policy analysis are discussed in Subsections 5.4 and 5.5, respectively.

5.1 Forecasting VAR Processes

5.1.1 Known Processes

Neglecting deterministic terms and exogenous variables the levels VAR form (2.1) is particularly convenient to use in forecasting the variables y_t. Suppose the u_t are generated by an independent rather than just uncorrelated white noise process. Then the optimal (minimum MSE) 1-step forecast in period T is the conditional expectation,

$$y_{T+1|T} = E(y_{T+1}|y_T, y_{T-1}, \ldots) = A_1 y_T + \cdots + A_p y_{T+1-p}. \tag{5.1}$$

Forecasts for larger horizons $h \geq 1$ may be obtained recursively as

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T}, \tag{5.2}$$

where $y_{T+j|T} = y_{T+j}$ for $j \leq 0$. The corresponding forecast errors are

$$\begin{array}{rcl} y_{T+1} - y_{T+1|T} & = & u_{T+1}, \\ y_{T+2} - y_{T+2|T} & = & u_{T+2} + A_1 u_{T+1}, \\ & \vdots & \\ y_{T+h} - y_{T+h|T} & = & u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}, \end{array} \tag{5.3}$$

where it is easy to see by successive substitution that

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j} A_j, \quad s = 1, 2, \ldots, \tag{5.4}$$

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$ (see Lütkepohl (1991, Sec. 11.3)). Hence, u_t is the 1-step forecast error in period $t-1$ and the forecasts are unbiased, that is, the forecast errors have expectation 0. As mentioned earlier, these are the minimum MSE forecasts. The MSE matrix of an h-step forecast is

$$\Sigma_y(h) = E\{(y_{T+h} - y_{T+h|T})(y_{T+h} - y_{T+h|T})'\} = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'. \tag{5.5}$$

For any other h-step forecast with MSE matrix $\Sigma_y^*(h)$, say, the difference $\Sigma_y^*(h) - \Sigma_y(h)$ is a positive semidefinite matrix. This result relies on the assumption that u_t is independent white noise, i.e., u_t and u_s are independent for $s \neq t$. If u_t is uncorrelated white noise and not necessarily independent over time, the forecasts obtained recursively as

$$y_T(h) = A_1 y_T(h-1) + \cdots + A_p y_T(h-p), \qquad h = 1, 2, \ldots, \tag{5.6}$$

with $y_T(j) = y_{T+j}$ for $j \leq 0$, are just best linear forecasts (see Lütkepohl (1991, Sec. 2.2.2) for an illustrative example).

It may be worth pointing out that the forecast MSEs for integrated processes are generally unbounded as the horizon h goes to infinity. Thus the forecast uncertainty increases without bounds for forecasts of the distant future. This contrasts with the case of $I(0)$ variables for which the forecast MSEs are bounded by the unconditional covariance Σ_y of y_t. This means, in particular, that forecasts of cointegration relations have bounded MSEs even for horizons approaching infinity.

The corresponding forecast intervals reflect this property as well. Assuming that the process y_t is Gaussian, that is, $u_t \sim$ iid $N(0, \Sigma_u)$, the forecast errors are also multivariate normal. This result may be used to set up forecast intervals of the form

$$[y_{k,T+h|T} - c_{1-\gamma/2}\sigma_k(h),\ y_{k,T+h|T} + c_{1-\gamma/2}\sigma_k(h)] \tag{5.7}$$

where $c_{1-\gamma/2}$ is the $(1-\frac{\gamma}{2})100$ percentage point of the standard normal distribution, $y_{k,T+h|T}$ denotes the kth component of $y_{T+h|T}$ and $\sigma_k(h)$ denotes the square root of the kth diagonal element of $\Sigma_y(h)$, that is, $\sigma_k(h)$ is the standard deviation of the h-step forecast error for the kth component of y_t. Obviously, if $\sigma_k(h)$ is unbounded for $h \to \infty$, the same is true for the length of the interval in (5.7).

To illustrate these issues consider the following bivariate VAR(1) example process with cointegrating rank 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{5.8}$$

The corresponding VECM representation is

$$\Delta y_t = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} y_{t-1} + u_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1, -1] y_{t-1} + u_t,$$

that is,

$$\alpha = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \beta' = [1, -1].$$

For this process it is easily seen that $\Phi_0 = I_2$ and

$$\Phi_j = A_1^j = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \qquad j = 1, 2, \ldots,$$

which implies

$$\Sigma_y(h) = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j' = \Sigma_u + (h-1) \begin{bmatrix} \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 \end{bmatrix}, \qquad h = 1, 2, \ldots,$$

where σ_2^2 is the variance of u_{2t}. Moreover, the conditional expectations are $y_{k,T+h|T} = y_{2,T}$ $(k = 1, 2)$. Hence, the forecast intervals are

$$\left[y_{2,T} - c_{1-\gamma/2}\sqrt{\sigma_k^2 + (h-1)\sigma_2^2},\ y_{2,T} + c_{1-\gamma/2}\sqrt{\sigma_k^2 + (h-1)\sigma_2^2}\right], \qquad k = 1, 2.$$

Obviously the length of this interval approaches infinity for $h \to \infty$.

On the other hand, multiplying (5.8) by

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

gives

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} y_t = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} y_{t-1} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} u_t$$

which implies that the cointegration relation $z_t = y_{1t} - y_{2t} = u_{1t} - u_{2t}$ is zero mean white noise. Thus, the forecast intervals for z_t for any forecast horizon h are of constant length,

$$[z_{T+h|T} - c_{1-\gamma/2}\sigma_z(h),\ z_{T+h|T} + c_{1-\gamma/2}\sigma_z(h)] = [-c_{1-\gamma/2}\sigma_z,\ c_{1-\gamma/2}\sigma_z],$$

where $\sigma_z^2 = \mathrm{Var}(u_{1t}) + \mathrm{Var}(u_{2t}) - 2\mathrm{Cov}(u_{1t}, u_{2t})$ is the variance of z_t and $z_{T+h|T} = 0$ for $h \geq 1$ has been used.

5.1.2 Estimated Processes

In practice the parameters of a VAR process are usually estimated. We will now consider the implications for the forecast precision. For that purpose, we denote the optimal h-step forecast by $y_{T+h|T}$ as in (5.2) and furnish its counterpart based on estimated coefficients by a hat,

$$\hat{y}_{T+h|T} = \hat{A}_1 \hat{y}_{T+h-1|T} + \cdots + \hat{A}_p \hat{y}_{T+h-p|T}, \tag{5.9}$$

where, of course, $\hat{y}_{T+j|T} = y_{T+j}$ for $j \leq 0$ and the $\hat{A}_i$ $(i = 1, \ldots, p)$ are estimated parameters. The corresponding forecast error is

$$\begin{aligned} y_{T+h} - \hat{y}_{T+h|T} &= [y_{T+h} - y_{T+h|T}] + [y_{T+h|T} - \hat{y}_{T+h|T}] \\ &= \sum_{j=1}^{h-1} \Phi_j u_{T+h-j} + [y_{T+h|T} - \hat{y}_{T+h|T}]. \end{aligned} \tag{5.10}$$

At the forecast origin T the first term on the right-hand side involves future residuals only whereas the second term involves present and past variables only, provided only past variables have been used for estimation. Consequently, if u_t is independent white noise, the two terms are independent. Moreover, under standard assumptions, the difference $y_{T+h|T} - \hat{y}_{T+h|T}$ is small in probability as the sample size used for estimation gets large. Hence, the forecast error covariance matrix in this case is

$$\begin{aligned} \Sigma_{\hat{y}}(h) &= E\{[y_{T+h} - \hat{y}_{T+h|T}][y_{T+h} - \hat{y}_{T+h|T}]'\} \\ &= \Sigma_y(h) + o(1), \end{aligned} \tag{5.11}$$

where $o(1)$ denotes a term which approaches zero as the sample size tends to infinity. Thus, for large samples the estimation uncertainty may be ignored in evaluating the forecast precision and setting up forecast intervals. In small samples, including a correction term is preferable, however. In this case the precision of the forecasts will depend on the precision of the estimators. Hence, if precise forecasts are desired, it is a good strategy to look for precise parameter estimators. For the stationary case, see Lütkepohl (1991, Chapter 3) for further details and, for the nonstationary case, see Reimers (1991), Engle & Yoo (1987) and Basu & Sen Roy (1987) for further discussion.

5.2 Granger-Causality Analysis

5.2.1 The Concept

The causality concept introduced by Granger (1969) is perhaps the most widely discussed form of causality in the econometrics literature. Granger defines a variable y_{1t} to be causal for another time series variable y_{2t} if the former helps predicting the latter. Formally, denoting by $y_{2,t+h|\Omega_t}$ the optimal h-step predictor of y_{2t} at origin t based on the set of all the relevant information in the universe Ω_t, y_{1t} may be defined to be Granger-noncausal for y_{2t} if and only if

$$y_{2,t+h|\Omega_t} = y_{2,t+h|\Omega_t \setminus \{y_{1,s}|s \leq t\}}, \qquad h = 1, 2, \ldots. \tag{5.12}$$

Here $\Omega_t \setminus \mathcal{A}$ denotes the set containing all elements of Ω_t which are not in the set $\mathcal{A}$. In other words, y_{1t} is not causal for y_{2t} if removing the past of y_{1t} from the information set does not change the optimal forecast for y_{2t} at any forecast horizon. In turn, y_{1t} is Granger-causal for y_{2t} if the equality in (5.12) is violated for at least one h and, thus, a better forecast of y_{2t}

is obtained for some forecast horizon by including the past of y_{1t} in the information set. If $\Omega_t = \{(y_{1,s}, y_{2,s})' | s \leq t\}$ and $(y_{1t}, y_{2t})'$ is generated by a bivariate VAR(p) process,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + u_t, \tag{5.13}$$

then (5.12) is easily seen to be equivalent to

$$\alpha_{21,i} = 0, \qquad i = 1, 2, \ldots, p. \tag{5.14}$$

It is perhaps worth mentioning that Granger-causality can also be investigated in the framework of the VECM. Writing that model for the presently considered bivariate case as

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \alpha\beta' \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \sum_{i=1}^{p-1} \begin{bmatrix} \gamma_{11,i} & \gamma_{12,i} \\ \gamma_{21,i} & \gamma_{22,i} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-i} \\ \Delta y_{2,t-i} \end{bmatrix} + u_t,$$

it is easy to see that (5.14) is equivalent to $\gamma_{21,i} = 0$ $(i = 1, \ldots, p-1)$ and the element in the lower left hand corner of $\alpha\beta'$ is also zero. Of course, in a bivariate situation the cointegrating rank r can only be 0, 1 or 2, where $r = 1$ is the only case which may involve genuine cointegration. In that case, α and β are (2×1) vectors and

$$\alpha\beta' = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [\beta_1, \beta_2] = \begin{bmatrix} \alpha_1\beta_1 & \alpha_1\beta_2 \\ \alpha_2\beta_1 & \alpha_2\beta_2 \end{bmatrix}.$$

Thus, in this case, $\alpha_2\beta_1 = 0$ needs to be checked in addition to $\gamma_{21,i} = 0$ $(i = 1, \ldots, p-1)$ (see also Mosconi & Giannini (1992)).

Economic systems usually consist of more than two relevant variables. Hence, it is desirable to extend the concept of Granger-causality to higher dimensional systems. Different possible extensions have been considered (see, e.g., Lütkepohl (1993), Dufour & Renault (1998)). One possible generalization assumes that the vector of all variables, y_t, is partitioned into two subvectors so that $y_t = (y_{1t}', y_{2t}')'$. Then the definition in (5.12) may be used for the two subvectors y_{1t}, y_{2t} rather than two individual variables. If $\Omega_t = \{y_s | s \leq t\}$ and y_t is a VAR process of the form (5.13), where the $\alpha_{kh,i}$ are now matrices of appropriate dimensions, the restrictions for noncausality are the same as in the bivariate case so that y_{1t} is Granger-noncausal for y_{2t} if $\alpha_{21,i} = 0$ for $i = 1, \ldots, p$ (Lütkepohl (1991, Sec. 2.3.1)).

This approach is not satisfactory if interest centers on a causal relation between two variables within a higher dimensional system because a set of variables being causal for another set of variables does not necessarily imply that each member of the former set is causal for each member of the latter set. Therefore it is of interest to consider causality of y_{1t} to y_{2t} if there are further variables in the system. In this context, different causality concepts have been proposed which are most easily explained in terms of the three-dimensional VAR process

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} & \alpha_{13,i} \\ \alpha_{21,i} & \alpha_{22,i} & \alpha_{23,i} \\ \alpha_{31,i} & \alpha_{32,i} & \alpha_{33,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ y_{3,t-i} \end{bmatrix} + u_t. \tag{5.15}$$

Within this system causality of y_{1t} for y_{2t} is sometimes checked by testing

$$H_0 : \alpha_{21,i} = 0, \ i = 1, \ldots, p. \tag{5.16}$$

These restrictions are not equivalent to (5.12), however. They are equivalent to $y_{2,t+1|\Omega_t} = y_{2,t+1|\Omega_t \setminus \{y_{1,s}|s \leq t\}}$. The information in past y_{1t} may still help improving the forecasts of y_{2t} more than one period ahead if (5.16) holds (Lütkepohl (1993)). Intuitively this happens because there may be indirect causal links, e.g., y_{1t} may have an impact on y_{3t} which in turn may affect y_{2t}. Thus, the definition of noncausality corresponding to the restrictions in (5.16) is not in line with an intuitive notion of the term. For higher dimensional processes the definition based on (5.12) results in more complicated nonlinear restrictions for the VAR coefficients. Details are given in Dufour & Renault (1998).

5.2.2 Testing for Granger-Causality

Wald tests are standard tools for testing restrictions on the coefficients of VAR processes because the test statistics are easy to compute in this context. Unfortunately, they may have nonstandard asymptotic properties if the VAR contains $I(1)$ variables. In particular, Wald tests for Granger-causality are known to result in nonstandard limiting distributions depending on the cointegration properties of the system and possibly on nuisance parameters (see Toda & Phillips (1993)).

Dolado & Lütkepohl (1996) and Toda & Yamamoto (1995) point out a simple way to overcome the problems with these tests in the present context. As mentioned in Section 3.1, the nonstandard asymptotic properties of the standard tests on the coefficients of cointegrated VAR processes are due to the singularity of the asymptotic distribution of the LS estimators. Hence, the idea is to get rid of the singularity by fitting a VAR process whose order exceeds the true order. It can be shown that this device leads to a nonsingular asymptotic distribution of the relevant coefficients, overcoming the problems associated with standard tests and their complicated nonstandard limiting properties.

More generally, as mentioned in Section 3.1, Dolado & Lütkepohl (1996) show that whenever the elements in at least one of the complete coefficient matrices A_i are not restricted at all under the null hypothesis, the Wald statistic has its usual limiting χ^2-distribution. Thus, if elements from all A_i, $i = 1, \ldots, p$, are involved in the restrictions as, for instance, in the noncausality restrictions in (5.14) or (5.16), simply adding an extra (redundant) lag in estimating the parameters of the process, ensures standard asymptotics for the Wald test. Of course, if the true DGP is a VAR(p) process, then a VAR($p+1$) with $A_{p+1} = 0$ is also an appropriate model. The test is then performed on the A_i, $i = 1, \ldots, p$, only.

For this procedure to work it is neither necessary to know the cointegration properties of the system nor the order of integration of the variables. Thus, if there is uncertainty with respect to the integration properties of the variables an extra lag may simply be added and the test may be performed on the lag augmented model to be on the safe side. Unfortunately, the procedure is not fully efficient due to the redundant parameters.

It may be worth noting that the procedure remains valid if an intercept or other deterministic terms are included in the VAR model, as a consequence of results due to Park & Phillips (1989) and Sims, Stock & Watson (1990). A generalization of these ideas to Wald tests for nonlinear restrictions representing, for instance, other causality definitions, is discussed by Lütkepohl & Burda (1997). Testing for Granger-causality in infinite order VAR processes is considered by Lütkepohl & Poskitt (1996a) and Saikkonen & Lütkepohl (1996).

5.3 Impulse Response Analysis

5.3.1 Concepts and Ideas

Tracing out the effects of shocks in the variables of a given system may also be regarded as a type of causality analysis. If the process y_t is $I(0)$, it has a Wold moving average (MA) representation

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots \ , \tag{5.17}$$

where $\Phi_0 = I_K$ and the Φ_s can be computed recursively as in (5.4). 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.

Although the Wold representation does not exist for nonstationary cointegrated processes it is easy to see from Section 5.1.1 that the Φ_s impulse response matrices can be computed in the same way as in (5.4) (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. Assuming that all variables are $I(1)$, it is also reasonable to consider the Wold representation of the stationary process Δy_t,

$$\Delta y_t = \Xi_0 u_t + \Xi_1 u_{t-1} + \Xi_2 u_{t-2} + \cdots, \tag{5.18}$$

where $\Xi_0 = I_K$ and $\Xi_j = \Phi_j - \Phi_{j-1}$ $(j = 1, 2, \ldots)$. Again, the coefficients of this representation may be interpreted as impulse responses. Because $\Phi_s = \sum_{j=0}^{s} \Xi_j$, $s = 1, 2, \ldots$, the Φ_s may be regarded as accumulated impulse responses of the representation in first differences.

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 not instantaneously uncorrelated, that is, if Σ_u is not diagonal. Therefore, in many applications the innovations of the VAR are orthogonalized using a Cholesky decomposition of the covariance matrix Σ_u. Denoting by P a lower triangular matrix such that $\Sigma_u = PP'$, the orthogonalized shocks are given by $\varepsilon_t = P^{-1} u_t$. Hence, in the stationary case we get from (5.17),

$$y_t = \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \cdots \ , \tag{5.19}$$

where $\Psi_i = \Phi_i P$ $(i = 0, 1, 2, \ldots)$. Here $\Psi_0 = P$ 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.

Since many matrices P exist which satisfy $PP' = \Sigma_u$, using this approach is to some extent arbitrary. Even if P 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 to try various triangular orthogonalizations and check the robustness of the results with respect to the ordering of the variables. He

also recommends using a priori hypotheses about the structure if possible. The resulting models are known as *structural VARs*. They are of the general form (2.7). In addition, the residuals may be represented as $v_t = R\varepsilon_t$ and ε_t is a $(K \times 1)$ vector of structural shocks with covariance matrix $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$. Usually it is assumed that Σ_ε is a diagonal matrix so that the structural shocks are instantaneously uncorrelated. The relation to the reduced form residuals is given by $\Gamma_0^* u_t = R\varepsilon_t$.

In recent years, different types of identifying restrictions were considered (see, e.g., Watson (1994) and Lütkepohl & Breitung (1997) for discussions). The aforementioned triangular system is a special case of such a class of structural models with $P = \Gamma_0^{*-1} R$. Obviously, identifying restrictions are required to obtain a unique structural representation. In the early literature, linear restrictions on Γ_0^* or R were used to identify the system (e.g., Pagan (1995)). Later Blanchard & Quah (1989), King, Plosser, Stock & Watson (1991), Gali (1992) and others introduced nonlinear restrictions. To motivate the nonlinear constraints it is useful to consider the moving average representation (5.18) and write it in terms of the structural residuals:

$$\Delta y_t = \Theta_0 \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2} + \cdots, \tag{5.20}$$

where $\Theta_s = \Xi_s \Gamma_0^{*-1} R$ $(s = 0, 1, \ldots)$. The long run impact of the structural shocks on y_t is given by $\lim_{n\to\infty} \partial y_{t+n} / \partial \varepsilon_t' = \lim_{n\to\infty} \Phi_n \Gamma_0^{*-1} R = \sum_{s=0}^{\infty} \Theta_s \equiv \bar{\Theta}$. If the shock ε_{jt} has a transitory effect on y_{it}, then the (i,j)th element of $\bar{\Theta}$ is zero. Hence, the restriction that ε_{jt} does not affect y_{it} in the long run may be written as the nonlinear constraint

$$e_i' \bar{\Theta} e_j = e_i' (I_K + \Xi_1 + \Xi_2 + \cdots) \Gamma_0^{*-1} R e_j = 0.$$

Here e_i (e_j) is the ith $(j$th) column of the identity matrix. It can be shown that for a cointegrated system with cointegrating rank r, the matrix $\bar{\Theta}$ has rank $n - r$ so that there exist $n - r$ shocks with permanent effects (e.g., Engle & Granger (1987)).

Imposing this kind of nonlinear restrictions in the estimation procedure requires that nonlinear procedures are used. For instance, generalized methods of moments (GMM) estimation may be applied (see Watson (1994)). In the following subsection we will discuss the estimation of impulse responses when estimators of the model parameters are available.

5.3.2 Asymptotic Theory and Bootstrapping

If an estimator $\hat{\alpha}$, say, of the VAR coefficients summarized in the vector α is available, estimators of the impulse responses may be obtained as

$$\hat{\phi}_{ij,h} = \phi_{ij,h}(\hat{\alpha}). \tag{5.21}$$

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

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \Sigma_{\hat{\alpha}}), \tag{5.22}$$

the $\phi_{ij,h}$ are also asymptotically normally distributed,

$$\sqrt{T}(\hat{\phi}_{ij,h} - \phi_{ij,h}) \xrightarrow{d} N(0, \sigma_{ij,h}^2), \tag{5.23}$$

where

$$\sigma_{ij,h}^2 = \frac{\partial \phi_{ij,h}}{\partial \alpha'} \Sigma_{\hat{\alpha}} \frac{\partial \phi_{ij,h}}{\partial \alpha} \tag{5.24}$$

and $\partial\phi_{ij,h}/\partial\alpha$ denotes the vector of first order partial derivatives of $\phi_{ij,h}$ with respect to the elements of α. The limiting result in (5.23) holds if $\sigma^2_{ij,h}$ is nonzero which in turn is guaranteed if $\Sigma_{\hat{\alpha}}$ is nonsingular and $\partial\phi_{ij,h}/\partial\alpha \neq 0$. Note that the covariance matrix $\Sigma_{\hat{\alpha}}$ may be singular if there are constraints on the coefficients or, as mentioned earlier, if there are $I(1)$ variables. The partial derivatives will also usually be zero in parts of the parameter space because the $\phi_{ij,h}$ 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 α 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 some pretesting is applied in finding a parsimoniously parameterized model.

In practice, bootstrap methods are often used to construct confidence intervals (CIs) for impulse responses because these methods occasionally lead to more reliable small sample inference than asymptotic theory. Moreover, the analytical expressions of the asymptotic variances of the impulse response coefficients are rather complicated. 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. Unfortunately, the bootstrap does not necessarily overcome the problems due to a singularity in the asymptotic distribution which results from a zero variance in (5.23). In other words, in these cases bootstrap CIs may not have the desired coverage. For a critical discussion see Benkwitz, Lütkepohl & Neumann (1997).

5.4 Forecast Error Variance Decomposition

In practice forecast error variance decompositions are also popular tools for interpreting VAR models. Expressing the h-step forecast error from (5.3) in terms of the orthogonalized impulse responses $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})' = P^{-1}u_t$ from (5.19), where P is a lower triangular matrix such that $PP' = \Sigma_u$, gives

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

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

5.5 Policy Analysis

If there are superexogenous variables in the system (2.7), the model may also be used directly for policy analysis. In other words, if a policy maker affects the values or properties of z_t the effect on the endogenous variables may be investigated within the conditional model (2.7). If the policy maker sets the values of z_t the effect of such an action can be analyzed by considering the resulting dynamic effects on the endogenous variables similar to an impulse response analysis. In general, if z_t represents stochastic variables, it is more natural to think of policy actions as changes in the distribution of z_t. For instance, a policy maker may shift the mean of z_t. Again, such changes can be analyzed in the context of our extended VAR models. For details see, for example, Hendry & Mizon (1998).

6 Conclusions and Extensions

Since the publication of Sims' (1980) critique of classical econometric modeling VAR processes have become standard tools for macroeconometric analyses. A brief introduction to these models, their estimation, specification and analysis has been provided. Special attention has been given to cointegrated systems. Forecasting, causality, impulse response and policy analysis are discussed as possible uses of VAR models. In some of the discussion exogenous variables and deterministic terms are explicitly allowed for and, hence, the model class is generalized slightly relative to standard pure VAR processes.

There are now different software packages that support VAR analyses. For example, PcFiml (see Doornik & Hendry (1997)) and EVIEWS may be used. Furthermore, packages programmed in GAUSS exist which simplify a VAR analysis (see, e.g., Haase et al. (1992)).

In practice, further model generalizations are often useful. For instance, to obtain a more parsimonious parameterization allowing for MA terms as well and, hence, considering the class of vector autoregressive moving average processes may be desirable (see Hannan & Deistler (1988), Lütkepohl & Poskitt (1996b)). Extensions of these models to cointegrated systems are discussed by Lütkepohl & Claessen (1997), Bartel & Lütkepohl (1998) and Poskitt & Lütkepohl (1995). Especially for financial time series modeling the conditional second moments is sometimes of primary interest. Multivariate ARCH type models that can be used for this purpose are, for instance, discussed by Engle & Kroner (1995). Generally, nonlinearities of unknown functional form may be treated nonparametrically, semiparametrically or seminonparametrically. A large body of literature is currently developing on these issues. For some references see Härdle, Lütkepohl & Chen (1997).

References

Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle, In: B.N. Petrov & F. Csáki (eds.), *2nd International Symposium on Information Theory*, Budapest: Académiai Kiadó, 267–281.

Akaike, H. (1974), A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Banerjee, A., J.J. Dolado, J.W. Galbraith & D.F. Hendry (1993), *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, New York: Oxford University Press.

Bartel, H. & H. Lütkepohl (1998), Estimating the Kronecker indices of cointegrated echelon form VARMA models, *Econometrics Journal*, 1, C76-C99.

Basu, A.K. & S. Sen Roy (1987), On asymptotic prediction problems for multivariate autoregressive models in the unstable nonexplosive case, *Calcutta Statistical Association Bulletin*, 36, 29–37.

Benkwitz, A., H. Lütkepohl & M.H. Neumann (1997), Problems related to bootstrapping impulse responses of autoregressive processes, Discussion Paper 85, SFB 373, Humboldt Universität Berlin.

Blanchard, O. & D. Quah (1989), The Dynamic Effects of Aggregate Demand and Supply Disturbances, *American Economic Review*, 79, 655-673.

Boswijk, H.P. (1995), Efficient Inference on Cointegration Parameters in Structural Error Correction Models, *Journal of Econometrics*, 69, 133-158.

Dickey, D.A. & W.A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with Unit Root, *Journal of the American Statistical Association*, 74, 427-431.

Dolado, J.J. & H. Lütkepohl (1996), Making Wald Tests Work for Cointegrated Systems, *Econometric Reviews*, 15, 369-386.

Doornik, J.A. & D.F. Hendry (1997), *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, London: International Thomson Business Press.

Dufour, J.-M. & E. Renault (1998), Short run and long run causality in time series: Theory, *Econometrica*, 66, 1099 - 1125.

Engle, R.F. & C.W.J. Granger (1987), Co-Integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-276.

Engle, R.F., D.F. Hendry & J.F. Richard (1983), Exogeneity, *Econometrica*, 51, 277-304.

Engle, R.F. & K.F. Kroner (1995), Multivariate simultaneous generalized GARCH, *Econometric Theory*, 11, 122-150.

Engle, R.F. & B.S. Yoo (1987), Forecasting and testing in cointegrated systems, *Journal of Econometrics*, 35, 143-159.

Ericsson, N.R., D.F. Hendry & G.E. Mizon (1998), Exogeneity, Cointegration, and Economic Policy Analysis, *Journal of Business & Economic Statistics*, 16, 370-387.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York: John Wiley.

Gali, J. (1992), How Well does the IS-LM Model Fit Postwar U.S. Data, *Quarterly Journal of Economics*, 107, 709–738.

Granger, C.W.J. (1969), Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 37, 424-438.

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. & T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Haase, K., H. Lütkepohl, H. Claessen, M. Moryson & W. Schneider (1992), *MulTi: A Menu-Driven GAUSS Program for Multiple Time Series Analysis*, Universität Kiel, Kiel, Germany.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton: Princeton University Press.

Hannan, E.J. & M. Deistler (1988), *The Statistical Theory of Linear Systems*. New York: Wiley.

Hannan, E.J. & B.G. Quinn (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society*, B41, 190-195.

Hansen, H. & S. Johansen (1993), Recursive estimation in cointegrated VAR-models, Preprint, Institute of Mathematical Statistics, University of Copenhagen.

Härdle, W., H. Lütkepohl & R. Chen (1997), A review of nonparametric time series analysis, *International Statistical Review*, 65, 49–72.

Hatanaka, M. (1996), *Time-Series-Based Econometrics: Unit Roots and Co-Integration*, Oxford: Oxford University Press.

Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.

Hendry, D.F. & G.E. Mizon (1998), Exogeneity, Causality, and Co-breaking in Economic Policy Analysis of a Small Econometric Model of Money in the UK, *Empirical Economics*, 23, 267-294.

Hubrich, K., H. Lütkepohl & P. Saikkonen (1998), A review of systems cointegration tests, Discussion Paper, SFB 373, Humboldt Universität Berlin.

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, 1551-1580.

Johansen, S. (1992), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics*, 54, 383-397.

Johansen, S. (1994), The role of the constant and linear terms in cointegration analysis of non-stationary time series, *Econometric Reviews*, 13, 205-231.

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

Johansen, S. & K. Juselius (1990), Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991), Stochastic Trends and Economic Fluctuations, *American Economic Review*, 81, 819–840.

Krämer, W., W. Ploberger & R. Alt (1988), Testing for structural change in dynamic models, *Econometrica*, 56, 1355 - 1369.

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

Lütkepohl, H. (1993), Testing for causation between two variables in higher dimensional VAR models, in H. Schneeweiss & K.F. Zimmermann (eds.) *Studies in Applied Econometrics*, Heidelberg:Physica-Verlag, pp. 75 - 91.

Lütkepohl, H. & J. Breitung (1997), Impulse response analysis of vector autoregressive processes, in: Heij, C., Schumacher, H., Hanzon, B. & C. Praagman, *System Dynamics in Economic and Financial Models*, Chichester: John Wiley.

Lütkepohl, H. & M.M. Burda (1997), Modified Wald tests under nonregular conditions, *Journal of Econometrics*, 78, 315 - 332.

Lütkepohl, H. & H. Claessen (1997), Analysis of cointegrated VARMA processes, *Journal of Econometrics,* 80, 223-239.

Lütkepohl, H. & D.S. Poskitt (1996a), Testing for Causation Using Infinite Order Vector Autoregressive Processes, *Econometric Theory*, 12, 61-87.

Lütkepohl, H. & D.S. Poskitt (1996b), Specification of echelon form VARMA models, *Journal of Business & Economic Statistics*, 14, 69 - 79.

Lütkepohl, H. & D.S. Poskitt (1998), Consistent estimation of the number of cointegration relations in a vector autoregressive model, in R. Galata & H. Küchenhoff (eds.), *Econometrics in Theory and Practice. Festschrift for Hans Schneeweiß*, Heidelberg: Physica-Verlag, pp. 87-100.

Lütkepohl, H. & H.-E. Reimers (1992), Impulse Response Analysis of Cointegrated Systems, *Journal of Economic Dynamics and Control*, 16, 53-78.

Lütkepohl, H. & P. Saikkonen (1997a), Impulse Response Analysis in Infinite Order Cointegrated Vector Autoregressive Processes, *Journal of Econometrics*, 81, 127 - 157.

Lütkepohl, H. & P. Saikkonen (1997b), Order selection in testing for the cointegrating rank of a VAR process, Discussion Paper 93, SFB 373, Humboldt Universität Berlin.

Lütkepohl, H. & P. Saikkonen (1999), Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics*, forthcoming.

Mosconi, R. & C. Giannini (1992), Non-causality in Cointegrated Systems: Representation, Estimation and Testing, *Oxford Bulletin of Economics and Statistics*, 54, 399-417.

Ng, S. & P. Perron (1995), Unit Root Tests in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag *Journal of the American Statistical Association*, 90, 268-281.

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

Park, J.Y. & P.C.B. Phillips (1988), Statistical Inference in Regressions with Integrated Processes: Part 1, *Econometric Theory*, 4, 468-497.

Park, J.Y. & P.C.B. Phillips (1989), Statistical Inference in Regressions with Integrated Processes: Part 2, *Econometric Theory*, 5, 95-131.

Paulsen, J. (1984), Order determination of multivariate autoregressive time series with unit roots, *Journal of Time Series Analysis*, 5, 115-127.

Perron, P. & J.Y. Campbell (1993), A Note on Johansen's Cointegration Procedure when Trends are Present, *Empirical Economics*, 18, 777-789.

Phillips, P.C.B. (1987), Time series regression with a unit root, *Econometrica*, 55, 277-301.

Phillips, P.C.B. (1991), Optimal inference in cointegrated systems, *Econometrica*, 59, 283-306.

Phillips, P.C.B. & S.N. Durlauf (1986), Multiple time series regression with integrated processes, *Review of Economic Studies*, 53, 473-495.

Phillips, P.C.B. & B.E. Hansen (1990), Statistical inference in instrumental variables regression with $I(1)$ processes, *Review of Economic Studies*, 57, 99-125.

Phillips, P.C.B. & M. Loretan (1991), Estimating long-run economic equilibria, *Review of Economic Studies*, 58, 407-436.

Poskitt, D.S. & H. Lütkepohl (1995), Consistent specification of cointegrated autoregressive moving-average systems, Discussion Paper 54 1995, SFB 373, Humboldt–Universität zu Berlin.

Quinn, B.G. (1980), Order Determination for a Multivariate Autoregression," *Journal of the Royal Statistical Society*, B42, 182-185.

Reimers, H.-E. (1991), *Analyse kointegrierter Variablen mittels vektorautoregressiver Modelle*, Heidelberg: Physica-Verlag.

Rissanen, J. (1978), Modeling by Shortest Data Description, *Automatica*, 14, 465–471.

Saikkonen, P. (1992), Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation, *Econometric Theory*, 8, 1-27.

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

Saikkonen, P. & H. Lütkepohl (1997), Trend adjustment prior to testing for the cointegrating rank of a VAR process, Discussion Paper 84, SFB 373, Humboldt-Universität zu Berlin.

Saikkonen, P. & H. Lütkepohl (1998), Testing for the cointegrating rank of a VAR process with an intercept, Discussion Paper 51, SFB 373, Humboldt-Universität zu Berlin.

Saikkonen, P. & H. Lütkepohl (1999), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, forthcoming.

Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, 81, 93–126.

Schwarz, G. (1978), Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461-464.

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 (Hrsg.), *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland, pp. 283-327.

Sims, C.A., J.H. Stock & M.W. Watson (1990), Inference in Linear Time Series Models with Some Unit Roots, *Econometrica*, 58, 113-144.

Toda, H.Y. & P.C.B. Phillips (1993), Vector Autoregressions and Causality, *Econometrica*, 61, 1367-1393.

Toda, H.Y. & T. Yamamoto (1995), Statistical Inference in Vector Autoregressions with Possibly Integrated Processes, *Journal of Econometrics*, 66, 225-250.

Watson, M.W. (1994), Vector Autoregressions and Cointegration, in: Engle, R.F. and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, New York: Elsevier.

Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, *Journal of the American Statistical Association*, 57, 348-368.