March 19, 1999

Vector Autoregressive Analysis

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

Abstract

An introduction to vector autoregressive (VAR) analysis is given with special emphasis on cointegration. The models, estimating their parameters and specifying the autoregressive order, the cointegrating rank and other restrictions are discussed. Possibilities for model validation are also considered. Causality tests, impulse responses and forecast error variance decompositions are presented as tools for analyzing VAR models.

JEL classification: C32
Keywords: Cointegration, dynamic econometric models, forecasting, impulse responses

[1] I thank the Deutsche Forschungsgemeinschaft, SFB 373, for financial support.

1 Introduction

Over the last two decades vector autoregressive (VAR) processes have become popular tools for econometric analyses. The poor forecast performance of some large scale econometric simultaneous equations models has shed doubt on their usefulness for econometric analysis in general and has resulted in Sims' (1980) critique of macroeconometric modeling and his recommendation to use VAR models instead. There are some major differences between classical econometric models and VAR models. In VAR analyses it is not uncommon to treat all variables as endogenous a priori. Exogeneity is not assumed at the outset but may be the result of a detailed statistical analysis whereas in classical econometric models a large number of variables is typically assumed to be exogenous. Moreover, VAR models are usually constructed for a small number of variables only and emphasis is placed on rich dynamics with potentially many lags of the endogenous variables under study. In its basic form the model is set up in reduced form. For example, in Sims' (1980) classical article he uses a system for quarterly data consisting of the variables money, real GNP, unemployment, wages, price level and import prices and he includes four lags of each variable in each of the equations. In contrast, classical econometric models often contain dozens or even hundreds of equations with very few lags only and many a priori restrictions to identify the model (see Uebe (1995) for examples). In particular, the 'incredible' identifying a priori restrictions were criticized by Sims. In VAR analyses, impulse responses of the variables are often used for analyzing the interactions between the variables under consideration. In the past years it has become apparent that many problems of interest to econometricians cannot be analyzed in this way without any identifying restrictions. Therefore, *structural* VAR models are now often used in practice. Moreover, the invention of cointegration has resulted in specific parameterizations which support the analysis of the cointegration structure.

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 following 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 $\Delta y_{kt} = y_{kt} - y_{k,t-1}$ has no stochastic trend. Note, however, that Δy_{kt} may still have deterministic components such as a polynomial trend or a deterministic seasonal component. 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 $I(0)$ (Granger (1981), Engle & Granger (1987)). The cointegrating relations are often interpreted as the connecting links 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 models which are now in common use in VAR analyses. Estimation and specification of these models will be considered in Sections 3 and 4, respectively. Forecasting, impulse response analysis and other 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. For example, Lütkepohl (1991) gives a broad overview of many aspects of VAR models and their analysis. Hendry (1995) treats recent developments in general dynamic econometric modeling. Banerjee, Dolado, Galbraith & Hendry

(1993), Johansen (1995) and Hatanaka (1996) focus on models for integrated and cointegrated variables and Hamilton (1994) contains an introductory treatment of VAR models and related issues. Articles which survey vector autoregressive modeling include Watson (1994) and Lütkepohl & Breitung (1997). These references may be consulted for further details on some of the issues discussed in the following, for examples and further references.

2 Vector Autoregressive and Error Correction Models

The characteristics of the variables involved determine to some extent which model is a suitable representation of the data generation process (DGP). For instance, trending properties and seasonal fluctuations are of importance in setting up a suitable model. In the following we will focus on systems which contain potentially $I(0)$ and $I(1)$ variables. For convenience, the original 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.

In some of the following review, we allow for deterministic components such as 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. For simplicity we ignore seasonal dummy variables and other deterministic seasonal terms although they are often used in practice. Including them does not change the results and analysis in any essential way.

Given a set of K time series variables $y_t = (y_{1t}, \dots, y_{Kt})'$, the basic VAR model without deterministic components has the form

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

where $A = [A_1 : \cdots : A_p]$, the A_i are $(K \times K)$ coefficient matrices, $Y_{t-1}^{t-p} = (y'_{t-1}, \dots, y'_{t-p})'$ and $u_t = (u_{1t}, \dots, u_{Kt})'$ is an unobservable zero mean white noise process with time invariant positive definite covariance matrix Σ_u. That is, the u_t are serially uncorrelated or independent. The model (2.1) is briefly referred to as a VAR(p) process because the number of lags is p.

A VAR(p) process is *stable* if

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

Assuming that it has been initiated in the infinite past, it generates stationary time series which have time invariant means, variances and autocovariance 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 convenient representation if interest centers on the cointegrating relations because they do not appear explicitly in (2.1). They are more easily analyzed within a *vector error correction model* (VECM) which is obtained from (2.1) by subtracting y_{t-1} from both sides of the equality sign and rearranging terms. A VECM has the form

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t = \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t, \quad (2.3)$$

where $\Pi = -(I_K - A_1 - \cdots - A_p)$, $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ $(i = 1, \ldots, p-1)$, $\Gamma = [\Gamma_1 : \cdots : \Gamma_{p-1}]$ and $\Delta Y_{t-1}^{t-p+1} = Y_{t-1}^{t-p+1} - Y_{t-2}^{t-p}$. Because Δy_t does not contain stochastic trends by our assumption that all variables are at most $I(1)$, the term Πy_{t-1} is the only one which contains $I(1)$ 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 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 corresponding levels VAR version of the model. Given a VECM(p) it is easy to see that the A_j parameter matrices of the levels VAR form may be obtained 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, i.e., $\text{rk}(\Pi) = r$. 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 $\text{rk}(\alpha) = \text{rk}(\beta) = r$. 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 $r = \text{rk}(\Pi)$ linearly independent cointegrating relations among the components of y_t. The matrices α and β are not unique because, for any nonsingular $(r \times r)$ matrix R, defining $\alpha^* = \alpha R'$ and $\beta^* = \beta R^{-1}$ gives $\alpha^* \beta^{*'} = \Pi$. Hence, there are many possible β matrices that contain the cointegrating relations or some linear transformation 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 cases where no genuine cointegration is present even if the model (2.3) has a cointegrating rank strictly between 0 and K. Suppose, for instance, that all variables but one are stationary. 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 nonstationary variables and r stationary components. Generally, for each stationary 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 genuine cointegration. 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 term 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 $\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t$. If μ_t is a linear trend term, that is, $\mu_t = \mu_0 + \mu_1 t$, for example, y_t has a VAR(p) representation of the form

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

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 the deterministic term $\mu_t = \mu_0 + \mu_1 t$ in (2.4) enforces a linear trend term. 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 behavioral relations. Therefore this part is often of foremost interest in econometric analyses. Of course, a VECM(p) representation equivalent to (2.5) also exists (see (4.5)).

Clearly, the models considered so far are in reduced form because all right-hand side variables are predetermined or deterministic and no instantaneous relations are modeled. Sometimes it is of interest to model also the instantaneous relations. In that case it may be useful to consider a structural form model,

$$A_0 y_t = \nu_0 + \nu_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.6}$$

or a corresponding VECM,

$$\Gamma_0 \Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t. \tag{2.7}$$

Of course, restrictions have to be imposed to identify the parameters of these models.

3 Estimation

Because estimation of some of the special case models is computationally particularly easy 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 reduced form models with parameter restrictions are discussed.

3.1 Estimation of an Unrestricted VAR

Given a sample of size T, $y_1, \dots, y_T$, and p presample values, $y_{-p+1}, \dots, y_0$, it is well-known that the K equations of the VAR model (2.1) may be estimated separately by least squares (LS) without loosing efficiency relative to generalized LS (GLS). The LS estimator of $A = [A_1 : \cdots : A_p]$ is

$$\hat{A} = [\hat{A}_1 : \cdots : \hat{A}_p] = \sum_{t=1}^{T} y_t Y_{t-1}^{t-p'} \left(\sum_{t=1}^{T} Y_{t-1}^{t-p} Y_{t-1}^{t-p'} \right)^{-1}. \tag{3.1}$$

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

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

or, written in an alternative way,

$$\mathrm{vec}(\hat{A}) \overset{a}{\sim} N(\mathrm{vec}(A), \Sigma_{\hat{A}}/T). \tag{3.2}$$

Here vec denotes the column stacking operator which stacks the columns of a matrix in a column vector. Moreover, $\xrightarrow{d}$ and $\overset{a}{\sim}$ signify convergence in distribution. The covariance matrix of the asymptotic distribution is

$$\Sigma_{\hat{A}} = \mathrm{plim} \left(T^{-1} \sum_{t=1}^{T} Y_{t-1}^{t-p} Y_{t-1}^{t-p'} \right)^{-1} \otimes \Sigma_u$$

so that the result in (3.2) may be written in the somewhat imprecise but intuitive way as

$$\mathrm{vec}(\hat{A}) \approx N \left(\mathrm{vec}(A), \left(\sum_{t=1}^{T} Y_{t-1}^{t-p} Y_{t-1}^{t-p'} \right)^{-1} \otimes \Sigma_u \right). \tag{3.3}$$

Although these results also hold for cointegrated systems 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 (see Park & Phillips (1988, 1989), Sims, Stock & Watson (1990), Lütkepohl (1991, Chapter 11)). Consequently, some estimated coefficients or linear combinations of coefficients converge with a faster rate than $T^{1/2}$ if there are integrated or cointegrated variables. Therefore, in this case the usual t-, χ^2- and F-tests used for inference regarding the VAR parameters, may not be valid as shown, e.g., by Toda & Phillips (1993). This result is a generalization of the famous unit root case of a univariate first order autoregressive process, $y_t = \rho y_{t-1} + u_t$. It is well-known that the LS estimator $\hat{\rho}$ of ρ has a nonstandard limiting distribution if $\rho = 1$ and, hence, y_t is $I(1)$. From the unit root literature (e.g., Fuller (1976), Dickey & Fuller (1979)), the quantity $\sqrt{T}(\hat{\rho} - \rho)$ is known to converge to zero in probability, that is, the limiting distribution has zero variance and is degenerate, whereas $T(\hat{\rho} - \rho)$ has a nondegenerate nonnormal limiting distribution. Despite these results there are also many cases, where t-, χ^2- and F-tests can be applied in the usual manner in VAR models with $I(1)$ variables. Dolado & Lütkepohl (1996) and Toda & Yamamoto (1995) show that if a null hypothesis is considered which does not restrict elements of each of the A_i ($i = 1, \ldots, p$) the usual Wald tests have their standard asymptotic properties. For example, t-ratios have their usual asymptotic standard normal distribution if the VAR order p is greater than one.

If the process y_t is normally distributed (Gaussian) and stationary, then the LS estimator in (3.1) is identical to the maximum likelihood (ML) estimator conditional on the presample values. Moreover, deterministic terms such as polynomial trends may be included in the model. 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)).

In order to work with the asymptotic results an estimator of the covariance matrix Σ_u is needed. The usual estimators may be used for that purpose, that is,

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

are possible candidates. Here the $\hat{u}_t = y_t - \hat{A} Y_{t-1}^{t-p}$ $(t = 1, \ldots, T)$ are the LS residuals. Both estimators in (3.4) are consistent and asymptotically normally distributed under general assumptions. Furthermore, they are asymptotically independent of $\hat{A}$ (see Lütkepohl (1991) and Lütkepohl & Saikkonen (1997a)). These properties ensure that the estimators can be used in the usual way in setting up test statistics, for example.

3.2 Estimation of a VECM

If the cointegrating rank of y_t is known and one wishes to impose the corresponding restrictions, it is convenient to work with the VECM form (2.3). Following Johansen (1988, 1995), we denote the residuals from a regression of Δy_t and y_{t-1} on ΔY_{t-1}^{t-p+1} by R_{0t} and R_{1t}, respectively, and define

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}', \quad i, j = 0, 1.$$

The parameter estimators under the restriction $\text{rk}(\Pi) = r$ are then obtained by solving the eigenvalue problem

$$\det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) = 0. \tag{3.5}$$

Let the ordered eigenvalues be $\lambda_1 \geq \cdots \geq \lambda_K$ with corresponding eigenvectors $V = [v_1, \ldots, v_K]$ satisfying $\lambda_i S_{11} v_i = S_{10} S_{00}^{-1} S_{01} v_i$ $(i = 1, \ldots, K)$ and normalized such that $V' S_{11} V = I_K$. Then β and α may be estimated as

$$\hat{\beta} = [v_1, \ldots, v_r] \quad \text{and} \quad \hat{\alpha} = S_{01} \hat{\beta} (\hat{\beta}' S_{11} \hat{\beta})^{-1},$$

respectively, that is, $\hat{\alpha}$ may be viewed as LS estimator from the model

$$R_{0t} = \alpha \hat{\beta}' R_{1t} + \tilde{u}_t.$$

An estimator of Π is $\hat{\Pi} = \hat{\alpha}\hat{\beta}'$ and, using $\Delta y_t - \hat{\Pi} y_{t-1} = \Gamma \Delta Y_{t-1}^{t-p+1} + \tilde{u}_t$, $\Gamma = [\Gamma_1 : \cdots : \Gamma_{p-1}]$ may be estimated as

$$\hat{\Gamma} = [\hat{\Gamma}_1 : \cdots : \hat{\Gamma}_{p-1}] = \left(\sum_{t=1}^{T} (\Delta y_t - \hat{\Pi} y_{t-1}) \Delta Y_{t-1}^{t-p+1'} \right) \left(\sum_{t=1}^{T} \Delta Y_{t-1}^{t-p+1} \Delta Y_{t-1}^{t-p+1'} \right)^{-1}.$$

Under Gaussian assumptions these estimators are ML estimators conditional on the presample values (Johansen (1995)).

In this approach the parameter estimator $\hat{\beta}$ is made unique by normalizing the eigenvectors from the eigenvalue problem (3.5) and $\hat{\alpha}$ is adjusted accordingly. However, these are not econometric identification restrictions. Without such restrictions only the product $\alpha\beta' = \Pi$ can be estimated consistently. For consistent estimation of the matrices α and β, identifying restrictions have to be imposed. For example, in a specific model it may be reasonable to assume that the first part of β is an identity matrix, so that $\beta' = [I_r : \beta_1']$, where β_1 is a $((K-r) \times r)$ matrix. For $r = 1$, this restriction amounts to normalizing the coefficient of the first variable. This identifying restriction has attracted some attention in the cointegration literature. 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)). In other words, the estimator of β converges with the fast rate T. It is therefore sometimes called *superconsistent* whereas 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 and converge at the usual $\sqrt{T}$ rate, $\sqrt{T}\text{vec}(\hat{\Gamma} - \Gamma) \xrightarrow{d} N(0, \Sigma_{\hat{\Gamma}})$ and $\sqrt{T}\text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} N(0, \Sigma_{\hat{\Pi}})$. The asymptotic distribution of $\hat{\Gamma}$ is nonsingular and, hence, standard inference may be used for Γ. In contrast, the $(K^2 \times K^2)$ covariance matrix $\Sigma_{\hat{\Pi}}$ has rank Kr. It is singular if $r < K$. This result is obtained because Π involves the cointegrating relations which are estimated superconsistently.

Interestingly, if an estimator of the levels parameters A is computed via the estimates of Π and Γ and thereby satisfies the cointegration restriction, that estimator has the same asymptotic distribution as in (3.2) where no restrictions have been imposed in estimating A. Moreover, computing the covariance matrix estimators in (3.4) from the residuals of the VECM estimation results in the same asymptotic properties as for the levels VAR form. Important results on estimating models with integrated variables are due to Phillips and his co-workers (e.g., Phillips & Durlauf (1986), Phillips (1987, 1991), Phillips & Hansen (1990), Phillips & Loretan (1991)). Extensions of the foregoing results to 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

In practice it is often desirable to place restrictions on the parameters to reduce the dimensionality of the parameter space. For instance, it is quite common that different lags of the differenced variables appear in the individual equations. In other words, there may be zero restrictions on the short-run parameters Γ. Moreover, some of the cointegrating relations may be confined to specific equations by imposing zero constraints on the loading matrix α. Efficient estimation of a model with parameter restrictions is more complicated than in the restricted case because LS is no longer identical to GLS in general. A possible estimation procedure is to estimate β in a first stage, for example, using the reduced form which ignores the restrictions on the short-run 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-run parameters the former may be treated as fixed in a second stage of the estimation procedure. In other words, a systems estimation procedure may be applied to

$$\Delta y_t = \alpha\hat{\beta}' y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + \tilde{u}_t. \tag{3.6}$$

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

4 Statistical Tools for Specifying VAR Models

4.1 Testing for Model Reduction

Because unrestricted VAR models usually involve a substantial number of parameters which in turn results in rather imprecise estimators, it is desirable to impose restrictions in order to improve the estimation precision. Statistical tests are commonly used for detecting possible restrictions. As mentioned previously, t-ratios and F-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. A particular set of restrictions where such problems occur is discussed in more detail in Section 5.2. In case of doubt it may be preferable to work on the VECM form.

In practice, one often starts from a model with some prespecified maximum lag length $p_{\max}$ and applies tests sequentially, eliminating one or more variables or lags of variables in each step until a relatively parsimonious representation with significant parameter estimates has been found. For instance, in a VECM one may first test the null hypothesis $H_0 : \Gamma_{p_{\max}-1} = 0$. If H_0 cannot be rejected, the lag length is reduced by one and $H_0 : \Gamma_{p_{\max}-2} = 0$ may be tested. This procedure is repeated 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 the maximum lag order to start with has to be made. Occasionally this quantity is chosen using some theoretical or institutional argument. For example, one may want to include lags of at least one year so that four lags are included initially for quarterly data and twelve lags for a monthly model. In some respect an inappropriate choice of $p_{\max}$ may not have severe consequences because starting with too small a $p_{\max}$ this may be discovered later when the final model is subjected to a series of specification tests (see Section 4.4). On the other hand, overspecifying $p_{\max}$ 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 before all insignificant lags are eliminated. The number of tests has an impact on the overall Type I error of the testing sequence. Hence, the choice of $p_{\max}$ will have an impact on the probability of choosing an overspecified model with redundant variables.

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. Alternatively, model selection procedures may be used for choosing the lag length or for determining exclusion restrictions. They will be discussed next.

4.2 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 levels VAR form (2.1) at this stage. A number of model selection criteria are available that can be used for choosing p. They proceed by fitting 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

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

where $\det(\cdot)$ denotes the determinant, log is the natural logarithm,

$$\widetilde{\Sigma}_u(m) = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$$

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 penalty function which penalizes large numbers of parameters in a model. 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 and the same sample size T is used for all orders the log determinant decreases (or at least does not increase) when m increases. The estimator $\hat{p}$ of p is chosen so as to balance the two terms in the sum on the right hand side of (4.1).

Examples of popular criteria in empirical work are Akaike's (1973, 1974) AIC which is obtained by defining $\varphi(m) = mK^2$ and $c_T = 2/T$, the HQ criterion of Hannan & Quinn (1979) and Quinn (1980) which uses $\varphi(m) = mK^2$ and $c_T = 2\log\log T/T$, and the SC with $\varphi(m) = mK^2$ and $c_T = \log T/T$, which was proposed by Schwarz (1978) and Rissanen (1978). The AIC asymptotically overestimates the order with positive probability whereas the last two criteria estimate the order consistently under quite general conditions, if the actual DGP has a finite VAR order and the maximum order $p_{\max}$ is at least as large as the true order. These results not only hold for stationary processes but also for nonstationary integrated and cointegrated processes (Paulsen (1984)). Denoting the orders selected by the three criteria by $\hat{p}(\text{AIC}), \hat{p}(\text{HQ})$ and $\hat{p}(\text{SC})$, respectively, it can be shown that $\hat{p}(\text{SC}) \leq \hat{p}(\text{HQ}) \leq \hat{p}(\text{AIC})$ for $T \geq 16$ (see Lütkepohl (1991, Chapters 4 and 11)).

Appropriately modified versions of the criteria may also be used for imposing other exclusion restrictions. In addition to specifying the model order and zero restrictions for the short-run parameters, the cointegrating rank also has to be determined. Possible tests are discussed next.

4.3 Tests for the Cointegrating Rank

The cointegrating rank r of a system of variables y_t is usually investigated by a sequential testing procedure based on likelihood ratio (LR) type tests. Because for a given cointegrating rank r, Gaussian ML estimates for the VECM are easy to compute, as shown in Section 3.2, LR test statistics are also easily available. The following hypotheses are typically tested sequentially,

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

Table 1. Model forms underlying LR type tests.

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

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. In contrast, if all the null hypotheses can be rejected including $H_0(K-1)$, 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 null distribution of the LR statistics are nonstandard. They depend on the difference $K - r_0$ and on deterministic trend terms included in the DGP. Therefore LR type tests have been derived under different assumptions regarding the deterministic term. The limiting null distributions do not depend on the short-run dynamics so that critical values for LR type tests can be tabulated for different values of $K - r_0$ under alternative assumptions for the deterministic terms.

For the present purposes the model (2.4), with clear separation of deterministic and stochastic terms turns out to be convenient. Therefore we consider the model

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

with

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

Using this stochastic part, it is easy to see that the process y_t has a VECM representation

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

where ν_0 and ν_1 are as defined below (2.5), $\nu = \nu_0 + \nu_1$, $\Pi^+ = [\nu_1 : \Pi]$ and $y_{t-1}^+ = [t-1 : y_{t-1}']'$. 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. A brief discussion of the different cases follows.

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

Although the case, where $y_t = x_t$ and hence there is no deterministic term at all is of limited importance for applied work because a zero mean term can rarely be assumed, it is still useful to consider it first because it is particularly easy to derive LR tests for the rank of Π under this assumption. The LR statistic can be obtained by estimating the VECM

$$\Delta y_t = \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t \tag{4.6}$$

under $H_0(r_0)$ with $\text{rk}(\Pi) = r_0$ and under $H_1(r_0)$ with $\text{rk}(\Pi) = K$ as discussed in Section 3.2 to get the likelihood maximum of the restricted and unrestricted models, respectively. It turns out that for a sample $y_1, \dots, y_T$ and presample values $y_{-p+1}, \dots, y_0$ the LR test statistic reduces to

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

where $\lambda_{r_0+1}, \dots, \lambda_K$ are the eigenvalues obtained from solving (3.5) (Johansen (1988, 1995)). As mentioned earlier, the limiting distribution under the null hypothesis is nonstandard and depends on the difference $K - r_0$. Critical values may be found in Johansen (1995, Table 15.1). Although it is convenient here to assume a Gaussian process y_t, the asymptotic distribution of the test statistic $LR(r_0)$ may be derived under more general assumptions for the process distribution. For the other cases listed in Table 1 the test statistics can be computed analogously by suitable modifications of the quantities in (3.5). These cases will be discussed briefly in the following.

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

In this case, where the mean term is allowed to be nonzero whereas a deterministic linear trend term is excluded by assumption, there are three variants of LR type tests that have been considered in the literature plus a number of asymptotically equivalent modifications. As can be seen from Table 1, the three statistics may be computed easily by using the reduced rank (RR) regression technique described in Section 3.2. 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 because a VAR model with integrated variables can in principle generate a linear trend if there is an intercept term. The second test enforces the restriction that there is no linear deterministic trend in computing the test statistic by absorbing the intercept into the cointegrating relations. Finally, in the third test the mean term μ_0 is estimated in a first stage and is subtracted from y_t. Then a RR regression is applied to (4.4) with x_t replaced by $\tilde{x}_t = y_t - \hat{\mu}_0$ to determine the test statistic. A suitable estimator $\hat{\mu}_0$ is proposed by Saikkonen & Luukkonen (1997). These authors 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 demonstrated in Saikkonen & Lütkepohl (1999) that the latter test can have considerably more local power than the other two LR tests that have been proposed for the present case with unrestricted mean term. Thus, based on local power the Saikkonen-Luukkonen variant of the LR test is the first choice if $\mu_1 = 0$.

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

In this case it is assumed that at least one of the variables has a deterministic linear trend so that $\mu_1 \neq 0$, whereas the constraint $\beta'\mu_1 = 0$ ensures that the cointegrating relations do not have a linear trend. It may be worth emphasizing, however, that for the $(K \times r)$ matrix β to satisfy $\beta'\mu_1 = 0$, the assumption $\mu_1 \neq 0$ 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 under the present assumptions. As a consequence, only tests of null hypotheses $\text{rk}(\Pi) = r_0 < K - 1$ make sense in this case. Intuitively, this result is plausible because a linear trend is assumed in at least one of the variables ($\mu_1 \neq 0$) whereas a stable model ($\text{rk}(\Pi) = K$) with an intercept cannot generate a linear trend.

From Table 1 it can be seen that both test statistics which have been proposed for the presently considered case can be obtained from a RR regression. The first test uses the same intercept model as the first test for the previous case where $\mu_1 = 0$ was assumed. In the present situation 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 GLS procedure, the trend is subtracted from y_t to yield $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t$. The test statistic is then computed via a RR regression applied to (4.4) with x_t replaced by $\hat{x}_t$ and using $\Pi\mu_1 = 0$. Note that $\Delta\hat{x}_t = \Delta y_t - \hat{\mu}_1$. The null distributions are tabulated in the references given in Table 1. Again it turns out that trend adjusting first and then performing the test may result in considerable gains in local power (Saikkonen & Lütkepohl (1998)).

Case 4: Arbitrary mean and trend parameters

If μ_0 and μ_1 are unconstrained parameter vectors, both the variables and the cointegrating relations may have a deterministic linear trend. In Table 1 three different LR type tests are listed that have been proposed for this situation. Again, all test statistics can be obtained conveniently via RR regression techniques. In the setup of the first model the linearity of the trend term is enforced. The second model includes the trend term in unrestricted form. As mentioned earlier, in principle such a model can generate quadratic trends. Because such trends are excluded here by our assumptions, the ν_i, $i = 0, 1$, must obey appropriate restrictions. These restrictions are not imposed in the RR regression underlying the Perron-Campbell test statistic. The last test in Table 1 is again based on prior trend adjustment and application of RR regression techniques to the trend adjusted data. The trend parameters may again be estimated by a GLS procedure. Critical values for all these tests may be found in the references given in Table 1. In a simulation comparison of the local power properties, Lütkepohl & Saikkonen (1999) found that none of the three tests is uniformly best.

Remarks on related issues

A comprehensive survey of the properties of LR and other tests for the cointegrating rank is given by Hubrich, Lütkepohl & Saikkonen (1998). We refer the interested reader to that article for further details. Small sample properties are also considered in that article. In the following a few specific remarks on some related issues will be added.

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 are known as maximum eigenvalue tests. They were also pioneered by Johansen (1988, 1991). The test statistics are of the form

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

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

For univariate processes ($K = 1$) testing $H_0 : r = 0$ against $H_1 : r = 1$ means testing that the process is $I(1)$ ($r = 0$) against the alternative of stationarity ($r = 1$). All the tests can be generalized to this situation except those for the case $\beta'\mu_1 = 0$ because the latter tests are meaningful only for alternatives $r \leq K - 1$ which is obviously not a possible alternative for $K = 1$. LR tests corresponding to the other cases were proposed by Dickey & Fuller (1979) and Fuller (1976). They are known as augmented Dickey-Fuller (ADF) tests and are closely related to the tests considered here.

As mentioned earlier, the limiting distributions of the test statistics are not only valid for normally distributed (Gaussian) processes but also under more general distributional assumptions even if the LR statistics computed under Gaussian assumptions are used. In that situation these tests are, of course, 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. 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)). Consequently, for the sake of consistency it is reassuring to know that the tests remain valid for this case. This situation is analyzed in more detail by Lütkepohl & Saikkonen (1997b). In particular, these authors consider the impact of lag length selection in this context.

Instead of the sequential testing procedures presented in the foregoing, Lütkepohl & Poskitt (1998) among others consider the possibility of determining the cointegrating rank by model selection criteria.

4.4 Model Validation

Once a model has been set up, its adequacy is usually checked with a range of tests and other statistical procedures. Many of these tools for model validation are based on estimation residuals. Some procedures are applied to the residuals of individual equations whereas others are based on the full residual vectors. For example, plots of the residual series may be visually inspected and their autocorrelations may be checked. Moreover, autocorrelations of squared residuals may be analyzed for possible autoregressive conditional heteroscedasticity (ARCH). In addition to visual inspection, formal statistical tests for remaining residual autocorrelation or ARCH are also applied routinely. For instance, LM (Lagrange Multiplier) or Portmanteau statistics may be used for that purpose. Furthermore, Lomnicki-Jarque-Bera tests for nonnormality may be applied to the residuals (see, e.g., Lütkepohl (1991), Doornik & Hendry (1997)).

Procedures for checking the stability and possible nonlinearity of a model are also available. They are used, e.g., for detecting potential structural shifts during the sample period and range from prediction tests to assessing recursive residuals or CUSUM type tests as well as recursive tests for cointegration (see, e.g., Granger & Teräsvirta (1993), Lütkepohl (1991), Doornik & Hendry (1997), Krämer, Ploberger & Alt (1988), Hansen & Johansen (1993)). If rival models for the same economic relations are available, encompassing tests may be applied to compare them (Hendry (1995)). For a more detailed discussion of model checking see also Doornik & Hendry (1997).

If model defects such as residual autocorrelation or ARCH effects are detected at the validation stage, model improvements are usually considered. For instance, adding further variables or lags of variables to the model or some of its equations may be considered. Moreover, including nonlinear terms or changing the functional form may result in improvements. It is also possible to modify the sampling period or to get other data.

5 Uses of Vector Autoregressive Models

Once an adequate model for the DGP of a system of variables is available it may be used for forecasting and economic analysis. For the latter purpose causality investigations, impulse response analysis and forecast error variance decompositions have been used. In the following, forecasting VAR processes will be discussed first. Then the concept of Granger-causality will be introduced and impulse response analysis and forecast error variance decompositions are considered.

5.1 Forecasting VAR Processes

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

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

Forecasts for larger horizons may be obtained recursively for $h = 1, 2, \dots$, 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

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}, \tag{5.3}$$

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 corresponding MSE matrices are

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

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.

The forecast MSEs for integrated processes are generally unbounded as the horizon h increases. Consequently, the forecast uncertainty increases without bounds for forecasts of the distant future. In contrast, for an $I(0)$ variable the forecast MSE is bounded by the unconditional variance of the variable. This result implies that forecasts of cointegration relations have bounded MSEs even for horizons approaching infinity.

The corresponding forecast intervals reflect these properties. If y_t is Gaussian and, thus, $u_t \sim$ iid $N(0, \Sigma_u)$, the forecast errors are also multivariate normally distributed. Using this result gives forecast intervals of the form

$$[y_{k,T+h|T} - c_{1-\gamma/2}\sigma_k(h),\ y_{k,T+h|T} + c_{1-\gamma/2}\sigma_k(h)], \tag{5.6}$$

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 standard deviation of the h-step forecast error for the kth component of y_t, that is, $\sigma_k(h)$ is the square root of the kth diagonal element of $\Sigma_y(h)$. If $\sigma_k(h)$ is unbounded for $h \to \infty$, the same is obviously true for the interval length in (5.6).

In practice, processes with estimated parameters are usually used for forecasting. To investigate the implications for the forecast precision, we denote the h-step forecast based on estimated parameters by $\hat{y}_{T+h|T}$, that is,

$$\hat{y}_{T+h|T} = \hat{A}_1 \hat{y}_{T+h-1|T} + \cdots + \hat{A}_p \hat{y}_{T+h-p|T}, \quad h = 1, 2, \ldots, \tag{5.7}$$

where, of course, $\hat{y}_{T+j|T} = y_{T+j}$ for $j \le 0$. The corresponding forecast error is

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

If T marks the end of the sample period used for estimation and is at the same time the forecast origin, then the first term on the right-hand side of the foregoing expression consists of future residuals only whereas the second term involves present and past variables only. Hence, assuming that 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 T gets large. Consequently, the forecast error covariance matrix is

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

Here $o(1)$ 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. On the other hand, in small samples the forecast precision will depend to some extent on the quality of the parameter estimators. Hence, if precise forecasts are desired, it is a good strategy to look for precise parameter estimators.

5.2 Granger-Causality

5.2.1 The Concept

Granger (1969) has introduced a concept of causality which has received considerable attention in the econometrics literature. He defines a time series variable y_{1t} to be causal for another variable y_{2t} if the information in the former helps improving the predictions of the latter. 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} is noncausal for y_{2t} if and only if

$$y_{2,t+h|\Omega_t} = y_{2,t+h|\Omega_t \setminus \{y_{1,s}|s \leq t\}}, \qquad h = 1, 2, \ldots. \tag{5.8}$$

Here $\Omega_t \setminus \mathcal{A}$ denotes the set containing all elements of Ω_t which are not in the set $\mathcal{A}$. Hence, y_{1t} is Granger-causal for y_{2t} if the equality in (5.8) is violated for at least one h. In order to deduce a useful concept from these general ideas, the information set Ω_t and the DGP of the variables involved have to be specified more precisely. If $\Omega_t = \{(y_{1,s}, y_{2,s})'|s \leq t\}$ and $(y_{1t}, y_{2t})'$ is generated by the bivariate VAR(p) process,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + u_t, \tag{5.9}$$

then (5.8) is equivalent to

$$\alpha_{21,i} = 0, \qquad i = 1, 2, \ldots, p, \tag{5.10}$$

(e.g., Lütkepohl (1991, Sec. 2.3.1)).

Of course, Granger-causality can also be investigated in the framework of the VECM. Writing that model for the presently considered bivariate case as

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

(5.10) is replaced by the restrictions

$$\gamma_{21,i} = 0, \qquad i = 1, \ldots, p-1,$$

and, in addition, the element in the lower left hand corner of $\alpha\beta'$ is also zero. In a bivariate situation the cointegrating rank r can only be 0, 1 or 2, the case $r = 1$ being the only one which may involve genuine cointegration. In that case, α and β are (2×1) vectors so that

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

Hence, 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) for further discussion).

Because economic systems usually consist of more than two relevant variables, it is desirable to extend the concept of Granger-causality to higher dimensional processes and larger information sets Ω_t. Different possible extensions have been considered in the literature (see, e.g., Lütkepohl (1993), Dufour & Renault (1998)). One possible generalization is based on partitioning y_t into two subvectors so that $y_t = (y_{1t}', y_{2t}')'$. Then the definition in (5.8) 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(p) process as in (5.9), where the $\alpha_{kn,i}$ are 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 $i = 1, \ldots, p$ (e.g., Lütkepohl (1991, Sec. 2.3.1)).

If interest centers on a causal relation between two variables within a higher dimensional system this approach may not be satisfactory 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. To illustrate the related problems consider the three-dimensional VAR process

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} & \alpha_{13,i} \\ \alpha_{21,i} & \alpha_{22,i} & \alpha_{23,i} \\ \alpha_{31,i} & \alpha_{32,i} & \alpha_{33,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ y_{3,t-i} \end{bmatrix} + u_t. \qquad (5.11)$$

Within this system the restrictions $\alpha_{21,i} = 0$ $(i = 1, \ldots, p)$ are not equivalent to the equality of the forecasts in (5.8). The restrictions imply that y_{1t} does not help improving the 1-step ahead forecasts of y_{2t}, that is, $y_{2,t+1|\Omega_t} = y_{2,t+1|\Omega_t \setminus \{y_{1,s}|s \leq t\}}$. The information in past y_{1t} may still help improving the forecasts of y_{2t} more than one period ahead (Lütkepohl (1993)). Intuitively this result is obtained 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}. For higher dimensional processes the definition based on (5.8) with Ω_t containing all variables of the system results in more complicated nonlinear restrictions for the VAR coefficients. Details are given in Dufour & Renault (1998).

5.2.2 Testing for Granger-Causality

As mentioned earlier, the usual tests for restrictions on the coefficients of VAR processes may have nonstandard asymptotic properties if the process contains integrated or cointegrated variables. In particular, Toda & Phillips (1993) show that Wald tests for Granger-causality result in test statistics with nonstandard limiting distributions depending on the cointegration properties of the system and possibly on nuisance parameters. 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 Wald tests for the coefficients of cointegrated VAR processes are due to the singularity of the asymptotic distribution of the LS estimators. 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 $(i = 1, \dots, p)$ are involved in the restrictions as in (5.10), simply adding an extra (redundant) lag in estimating the parameters of the process, ensures standard asymptotics for the Wald test. Clearly, if the true DGP is a VAR(p) process, then a VAR($p+1$) with $A_{p+1} = 0$ is also an appropriate model and the test may be performed on the A_i $(i = 1, \dots, p)$ 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 assuming that they are at most $I(1)$. Hence, if there is uncertainty with respect to the cointegration 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, due to the redundant parameters the procedure is not fully efficient.

Notice 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). Furthermore, Lütkepohl & Poskitt (1996a) and Saikkonen & Lütkepohl (1996) consider testing for Granger-causality in infinite order VAR processes.

5.3 Impulse Responses

Tracing out the effects of an impulse in one of the variables is another way of analyzing causal links between the variables of a system. If the process y_t is $I(0)$, it has a Wold moving average (MA) representation

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots , \tag{5.12}$$

where $\Phi_0 = I_K$ and the Φ_s are obtained as in (5.4). The (i,j)th elements of the matrices Φ_s, $s = 1, 2, \dots$, 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_{jt} given $\{y_{t-1}, y_{t-2}, \dots\}$ is given by the error term u_{jt}, the elements of Φ_s represent the impulse responses of the components of y_t with respect to the u_t innovations. These impulse responses are sometimes called *forecast error impulse responses* because the u_t are the 1-step ahead forecast errors. If y_t is $I(0)$, $\Phi_s \to 0$ as $s \to \infty$. Hence, the marginal effect of an impulse is transitory and vanishes over time.

Although a Wold representation for the levels does not exist for $I(1)$ processes it can be shown that the impulse response matrices can be computed in the same way as in (5.12) (Lütkepohl (1991, Chapter 11), Lütkepohl & Reimers (1992)). However, in this case the Φ_j will not converge to zero as $j \to \infty$ and, consequently, some shocks have permanent effects. If y_t is $I(1)$ then Δy_t is $I(0)$ and has a Wold representation, say

$$\Delta y_t = \Xi_0 u_t + \Xi_1 u_{t-1} + \Xi_2 u_{t-2} + \cdots , \tag{5.13}$$

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. Notice that $\Phi_s = \sum_{j=0}^{s} \Xi_j$, $s = 1, 2, \ldots$, and, hence, the Φ_s may be regarded as accumulated impulse responses of the representation in first differences.

It has been criticized that forecast error impulse responses may not reflect the actual reactions of the variables because the underlying shocks are not likely to occur in isolation if Σ_u is not diagonal and, thus, the components of u_t are instantaneously correlated. Therefore, in many applications the VAR innovations have been orthogonalized using a Cholesky decomposition of the covariance matrix Σ_u. Denoting by P a lower triangular matrix with positive diagonal elements such that $\Sigma_u = PP'$, the orthogonalized shocks are given by $\varepsilon_t = P^{-1}u_t$. Defining $\Psi_i = \Phi_i P$ $(i = 0, 1, 2, \ldots)$ we get from (5.12),

$$y_t = \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \cdots , \tag{5.14}$$

if y_t is $I(0)$. Since $\Psi_0 = P$ is lower triangular, 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 of the system and so on. In other words, if orthogonalized impulse responses are considered the ordering of the variables is of importance for the results. Notice that many matrices P exist which satisfy $PP' = \Sigma_u$. Therefore, this approach is to some extent arbitrary. Even if P is chosen by a lower triangular Choleski decomposition, a different ordering of the variables in the vector y_t may produce different responses so that 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 check various different triangular orthogonalizations and determine 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*. More generally these models are of the form (2.6) or (2.7) where the residuals may be represented as $u_t = R\varepsilon_t$ and ε_t is a $(K \times 1)$ vector of structural shocks with covariance matrix $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$. The latter covariance matrix is commonly assumed to be diagonal so that the structural shocks are instantaneously uncorrelated.

Different types of identifying restrictions have been considered in the past (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 $\Gamma_0 = I_K$ and $P = R$. 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) and others introduced nonlinear restrictions, for instance, by imposing that certain shocks are transitory and others have permanent effects. It can be shown that for a cointegrated system with cointegrating rank r, there exist r shocks with transitory and $n - r$ shocks with permanent effects (e.g., Engle & Granger (1987)).

Nonlinear procedures have to be used for imposing this kind of nonlinear restrictions in the estimation. For instance, generalized method of moments (GMM) estimation may be used (see Watson (1994)). Generally, if an estimator $\hat{\theta}$, say, of the VAR or VECM coefficients summarized in the vector θ is available, estimators of the impulse responses may be obtained as $\hat{\phi}_{ij,h} = \phi_{ij,h}(\hat{\theta})$. If $\hat{\theta}$ has an asymptotic normal distribution, then the same is true for the $\hat{\phi}_{ij,h}$. Assuming that the limiting distribution of the estimated $\phi_{ij,h}$ is regular with nonzero variances this asymptotic result may be used for setting up confidence intervals for the impulse responses. In practice, bootstrap methods are often used for this purpose because these methods occasionally lead to more reliable small sample inference than asymptotic theory. Furthermore, using the bootstrap for setting up confidence intervals, the precise expressions for the variances are not needed and, hence, deriving complicated analytical expressions explicitly can be avoided. Unfortunately, the asymptotic distributions of the estimated $\phi_{ij,h}$ may be singular. The bootstrap does not necessarily overcome this problem. In other words, in these cases bootstrap confidence intervals may not have the desired coverage. Benkwitz, Lütkepohl & Neumann (1997) discuss this problem in detail.

5.4 Forecast Error Variance Decomposition

Forecast error variance decompositions are also popular tools for interpreting VAR models. Expressing the h-step forecast error given in (5.3) in terms of the orthogonalized impulse responses $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Kt})' = P^{-1}u_t$ from (5.14) gives

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

The kth element of this vector is

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

where $\psi_{ij,n}$ is the (i,j)th element of Ψ_n. Using that the ε_{kt} are contemporaneously and serially uncorrelated and have unit variance 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 is meaningful if the ε_{it} can be viewed as shocks in variable i. Dividing the above terms by $\sigma_k^2(h)$ gives the percentage contribution of variable j to the h-step forecast error variance of variable k,

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

These quantities, computed from estimated parameters, are often reported for various forecast horizons. Because the forecast error variance components are based on the orthogonal impulse responses, they are subject to the same critique as the orthogonalized impulse responses. In other words, they may depend on the ordering of the variables.

6 Conclusions and Extensions

VAR processes have become standard tools for macroeconometric analyses since the publication of Sims' (1980) critique of classical econometric modeling. In the foregoing a brief introduction to these models is given. Moreover, their estimation, specification and analysis is discussed with special emphasis on cointegrated systems. Causality tests, impulse responses and forecast error variance decompositions are presented as possible tools for VAR analyses. A number of software packages support VAR analyses. Examples are PcFiml (see Doornik & Hendry (1997)) and EVIEWS. There are also packages programmed in GAUSS which simplify a VAR analysis (see, e.g., Haase et al. (1992)).

In practice, generalizations of pure VAR models are often used. 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 helpful (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). In a number of studies some of the variables are exogenous with respect to the parameters of interest and are therefore not modeled explicitly. Therefore VAR models are often extended to include exogenous variables (e.g., Hendry (1995)). Especially for financial time series the conditional second moments are sometimes of foremost 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 has developed around these issues. References may be found in 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.

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.

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. & K.F. Kroner (1995), Multivariate Simultaneous Generalized GARCH, *Econometric Theory*, 11, 122-150.

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

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.

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 Nonstationary 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.

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.

Uebe, G. (1995), *World of Economic Models*, Avebury, England.

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, pp. 2843-2915.