European University Institute

Department of Economics

EUI Working Papers

ECO 2007/11

Econometric Analysis with Vector Autoregressive Models

Helmut Lütkepohl

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Econometric Analysis with Vector Autoregressive Models

HELMUT LÜTKEPOHL

EUI Working Paper **ECO** 2007/11

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

April 26, 2007

Econometric Analysis with Vector Autoregressive Models

Helmut Lütkepohl[1]
European University Institute, Florence

Contents

[1]Prepared for the *Handbook of Computational Econometrics*. I thank two anonymous readers for helpful comments on an earlier draft of this chapter.

Abstract. Vector autoregressive (VAR) models for stationary and integrated variables are reviewed. Model specification and parameter estimation are discussed and various uses of these models for forecasting and economic analysis are considered. For integrated and cointegrated variables it is argued that vector error correction models offer a particularly convenient parameterization both for model specification and for using the models for economic analysis.

1 Introduction

Vector autoregressive (VAR) models have a long tradition as tools for multiple time series analysis (e.g., Quenouille (1957)). Being linear models, they are relatively easy to work with both in theory and practice. Although the related computations are relatively straightforward, they are sufficiently involved to make applied work cumbersome before powerful computers were in widespread use. VAR models became popular for economic analysis when Sims (1980) advocated them as alternatives to simultaneous equations models. The latter models were used extensively since the 1950s. The availability of longer and more frequently observed time series emphasized the need for models which focussed on the dynamic structure of the variables, however. Sims also criticized the exogeneity assumptions for some of the variables in simultaneous equations models as ad hoc and often not backed by fully developed theories. In contrast, in VAR models often all observed variables are treated as a priori endogenous. Statistical procedures rather than subject matter theory are used for imposing restrictions.

VAR models are easy to use for forecasting and can also be applied for economic analysis. Impulse response analysis or forecast error variance decompositions are typically used for disentangling the relations between the variables in a VAR model. To investigate structural hypotheses based on economic theory usually requires a priori assumptions which may not be testable with statistical methods. Therefore *structural* VAR (SVAR) models were developed as a framework for incorporating such restrictions. Moreover, the discovery of the importance of stochastic trends in economic variables and the development of cointegration analysis by Granger (1981), Engle and Granger (1987), Johansen (1995) and many others has led to important new developments in analyzing the relations between economic variables. In particular, it is often desirable to separate the long-run relations from the short-run dynamics of the generation process of a set of variables. The long-run or cointegration relations are often associated with specific economic relations which are of particular interest whereas the short-run dynamics describe the adjustment to the long-run relations when disturbances have occurred. Vector error correction or equilibrium correction models (VECMs) offer a convenient framework for separating long-run and short-run components of the data generation process (DGP) and they will also be discussed in this chapter. The importance of the

trending properties of the variables makes a special terminology useful. It will be introduced next.

1.1 Integrated Variables

A stochastic trend is usually thought of as a component of a stochastic process with properties similar to those of a discrete random walk. In other words, it is viewed as a nonstationary stochastic component which does not have a tendency to revert to a fixed mean. An important characteristic of such components is that they can be removed by differencing a variable. In this chapter a time series variable y_t is called *integrated of order d* ($I(d)$) if stochastic trends can be removed by differencing the variable d times and a stochastic trend still remains after differencing only $d-1$ times. Using the differencing operator Δ which is defined such that $\Delta y_t = y_t - y_{t-1}$, the variable y_t is $I(d)$ if $\Delta^d y_t$ is stationary while $\Delta^{d-1} y_t$ still has a stochastic trend.

I follow Johansen (1995) in defining this terminology more formally for linear processes. Suppose u_t, $t \in \mathbb{Z}$, is a zero mean K-dimensional stochastic process of independently, identically distributed (iid) random variables with nonsingular covariance matrix $E(u_t u_t') = \Sigma_u$ and let Φ_i, $i = 0, 1, \ldots$, be an absolutely summable sequence of $(K \times K)$ matrices. Then the K-dimensional vector stochastic process $y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}$, $t \in \mathbb{N}$, is called a *linear process*. It is called *integrated of order 0* ($I(0)$) if $\sum_{i=0}^{\infty} \Phi_i \neq 0$ and it is called $I(d)$, $d = 1, 2, \ldots$, if $\Delta^d y_t$ is $I(0)$. To simplify matters, all variables are assumed to be either $I(0)$ or $I(1)$ in the following if not explicitly stated otherwise.

Notice that a K-dimensional vector of time series variables $y_t = (y_{1t}, \ldots, y_{Kt})'$ is $I(d)$, in short, $y_t \sim I(d)$, if at least one of its components is $I(d)$. In that case, $\Delta^{d-1} y_t$ will still have a stochastic trend while $\Delta^d y_t$ does not. This definition does not exclude the possibility that some components of y_t may be $I(0)$ individually if $y_t \sim I(1)$. Moreover, it is often convenient to define an $I(0)$ process y_t for $t \in \mathbb{Z}$ rather than $t \in \mathbb{N}$ in the same way as before and I will repeatedly make use of this possibility in the following. In fact, I will assume that $I(0)$ processes are defined for $t \in \mathbb{Z}$ in a stationary context if not otherwise stated. On the other hand, if $I(d)$ processes y_t with $d > 0$ are involved, it is often easier to define them for $t \in \mathbb{N}$ and this is therefore done here.

More general definitions of integrated processes can be given. In particular, an $I(0)$ process does not have to be a linear process and $I(d)$ processes for non-integer d can be defined (see also Johansen (1995, Chapter 3) for further discussion). For our purposes the definitions given here are general enough, however.

A set of $I(d)$ variables is called *cointegrated* if a linear combination exists which is of lower integration order. Because I mostly consider $I(1)$ and $I(0)$ variables, the leading case will be a vector y_t which is $I(1)$ and for which a vector c exists such that $c'y_t$ is $I(0)$. The linear combination $c'y_t$ is then called a cointegration relation. If y_t consists of two components only which are both individually $I(1)$ and cointegrated, then they will be driven by a common stochastic trend and hence have a particularly strong relation. In the long-run they move together although they may drift apart in the short-run. The concept extends directly to more than two variables. If y_t contains components which are individually $I(0)$, then there may be trivial linear combinations which are called cointegration relations according to our terminology. For example, if the first component y_{1t} is $I(0)$, then $c = (1, 0, \ldots, 0)'$ gives a trivial cointegration relation $c'y_t = y_{1t}$. Although our terminology is in this case not quite in the spirit of the original idea of cointegration, it will be convenient in the following.

Notice also that integratedness of the variables refers only to their stochastic properties. In addition there can be deterministic terms. More precisely, a process y_t with nonzero mean term will be called $I(d)$ if $y_t - E(y_t)$ is $I(d)$ in the sense defined earlier. In particular, $I(0)$ variables may still have deterministic trend components. With respect to deterministic terms I assume that they will usually be at most linear trends of the form $E(y_t) = \mu_t = \mu_0 + \mu_1 t$. If $\mu_1 = 0$ there is just a constant or intercept term in the process. Occasionally, $\mu_t = 0$ will in fact be assumed to simplify matters. Extensions to other deterministic terms such as seasonal dummies are straightforward and are therefore avoided for simplicity.

1.2 Structure of the Chapter

A typical VAR analysis proceeds by specifying and estimating a model and then checking its adequacy. If model defects are detected at the latter stage, model revisions are made until a satisfactory model has been found. Then the model may be used for forecasting, causality or structural analysis. Figure 1 depicts the main steps of a VAR analysis and this chapter

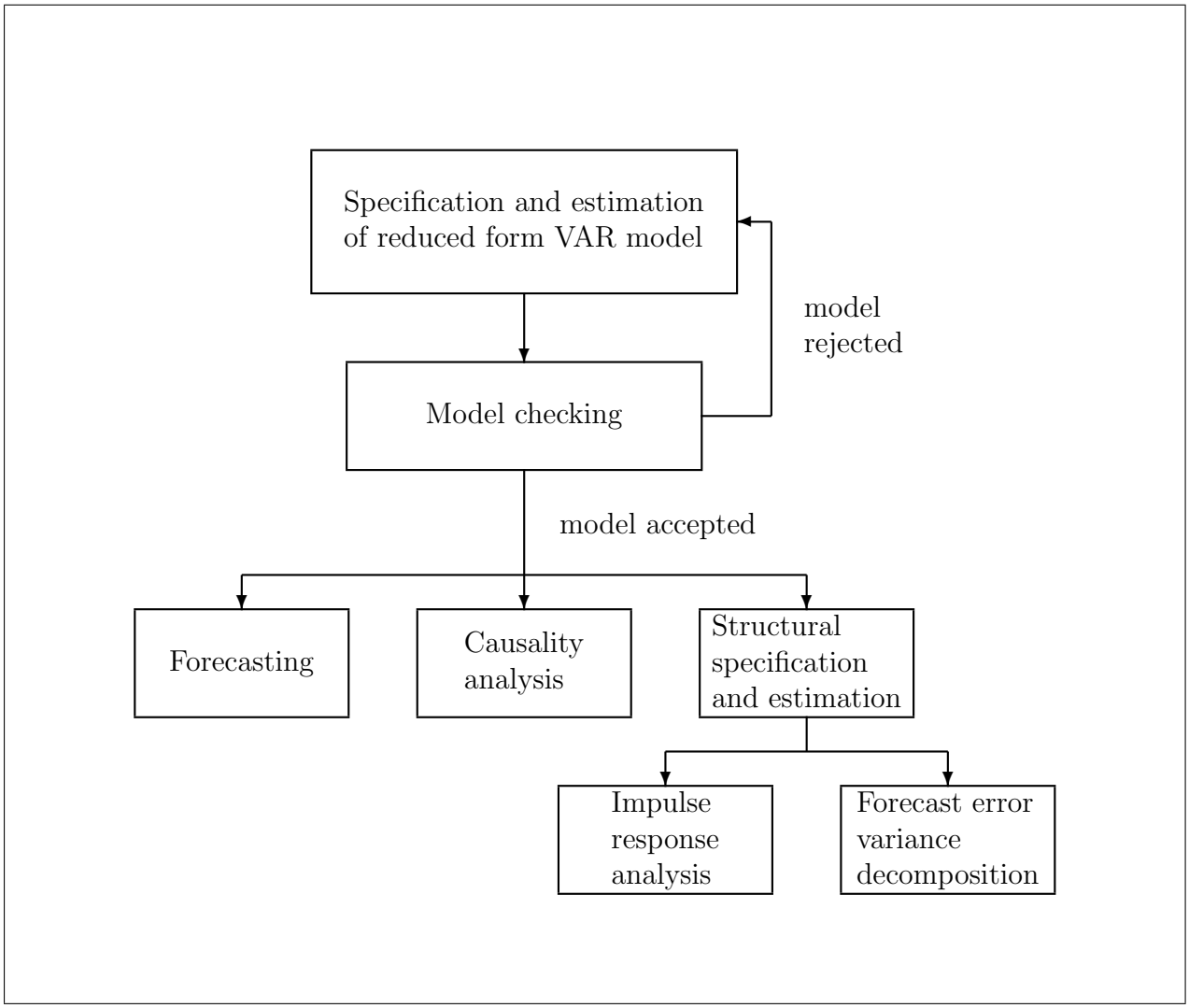

Figure 1: VAR analysis (figure adapted from Lütkepohl (2006b)).

is organized accordingly. The basic VARs and VECMs will be introduced in Section 2. Estimation and model specification issues will be treated in Sections 3 and 4, respectively, and Section 5 is devoted to model checking. Forecasting, Granger-causality analysis and structural modelling will be considered in Sections 6, 7 and 8, respectively. Conclusions and extensions are presented in Section 9.

A multiple time series analysis with VAR models can in principle be done with fairly straightforward computational algorithms. Recently computer intensive methods such as bootstrap and Bayesian simulation techniques have been developed for some stages of the analysis, however. They will be pointed out in the following even though the main focus is on the general concepts underlying a VAR analysis.

There are a number of textbooks and review articles which treat some of the issues in

more depth than the present exposition. Examples of related books are Banerjee, Dolado, Galbraith and Hendry (1993), Hamilton (1994), Hendry (1995), Johansen (1995), Hatanaka (1996), Lütkepohl and Krätzig (2004) and in particular Lütkepohl (2005). The present chapter draws heavily on the latter book and partly also on Lütkepohl (2006b).

1.3 Terminology and Notation

The following general terminology and notation will be used in this chapter. DGP, VAR, SVAR, VECM and MA abbreviate data generation process, vector autoregression, structural vector autoregression, vector error correction model and moving average, respectively. DSGE model is short for dynamic stochastic general equilibrium model. ML, LS, GLS, RR, LM, LR and MSE are used for maximum likelihood, least squares, generalized least squares, reduced rank, Lagrange multiplier, likelihood ratio and mean squared error, respectively. The natural logarithm is abbreviated as log. The sets of all integers and positive integers are denoted by $\mathbb{Z}$ and $\mathbb{N}$, respectively. The lag operator L is defined such that for a time series variable y_t, $Ly_t = y_{t-1}$, that is, it shifts the time index backward by one period. The differencing operator $\Delta = 1 - L$ is defined such that $\Delta y_t = y_t - y_{t-1}$. For a number x, $|x|$ denotes the absolute value or modulus. As usual, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

The following notation is used in relation to matrices. The transpose, inverse, trace, determinant and rank of the matrix A are denoted by A', A^{-1}, $\mathrm{tr}(A)$, $\det(A)$ and $\mathrm{rk}(A)$, respectively. For an $(n \times m)$ matrix A of full column rank $(n > m)$, an orthogonal complement is denoted by $A_\perp$. The zero matrix is the orthogonal complement of a nonsingular square matrix and an identity matrix of suitable dimension is the orthogonal complement of a zero matrix. The symbol vec denotes the column vectorization operator and vech is the corresponding operator which stacks the columns of a symmetric matrix from the main diagonal downwards. $\otimes$ signifies the Kronecker product and I_n is an $(n \times n)$ identity matrix.

The symbol '$\sim (\mu, \Sigma)$' abbreviates 'has a distribution with mean (vector) μ and (co)variance (matrix) Σ' and $\mathcal{N}(\mu, \Sigma)$ denotes a (multivariate) normal distribution with mean (vector) μ and (co)variance (matrix) Σ. Convergence in distribution and 'is asymptotically distributed as' is denoted as $\xrightarrow{d}$ and plim abbreviates the probability limit. Independently, identically

distributed is abbreviated as iid. A stochastic process u_t with $t \in \mathbb{Z}$ or $t \in \mathbb{N}$ is called *white noise* if the u_t's are iid with mean zero, $E(u_t) = 0$, positive definite covariance matrix $\Sigma_u = E(u_t u_t')$ and finite fourth order moments. Notice that the iid assumption is stronger than elsewhere in the literature. It is actually not necessary for many of the forthcoming results to hold but is made here for convenience.

2 VAR Processes

In the following the DGP of the K time series variables $y_t = (y_{1t}, \ldots, y_{Kt})'$ is assumed to be the sum of a deterministic term and a purely stochastic part,

$$y_t = \mu_t + x_t. \tag{2.1}$$

Here μ_t is the deterministic part and x_t is a purely stochastic process with zero mean. As mentioned earlier, the deterministic term μ_t will be assumed to be zero ($\mu_t = 0$), a constant ($\mu_t = \mu_0$) or a linear trend ($\mu_t = \mu_0 + \mu_1 t$) for simplicity. The (usually unobservable) purely stochastic part, x_t, includes stochastic trends and cointegration relations. It is assumed to have mean zero and a VAR or VECM representation. The observable process y_t inherits its deterministic and stochastic properties from μ_t and x_t. In particular, the order of integration and the cointegration relations are determined by x_t. The precise relation between x_t and y_t will be seen in the next subsections.

2.1 The Levels VAR Representation

The stochastic part x_t is assumed to be generated by a VAR process of order p (VAR(p)) of the form

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t, \tag{2.2}$$

where the A_i ($i = 1, \ldots, p$) are $(K \times K)$ parameter matrices and the error process $u_t = (u_{1t}, \ldots, u_{Kt})'$ is a K-dimensional zero mean white noise process with covariance matrix $E(u_t u_t') = \Sigma_u$. In short, $u_t \sim \text{iid}(0, \Sigma_u)$. The VAR process (2.2) is *stable* if

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } |z| \leq 1, \tag{2.3}$$

that is, if all roots of the determinantal polynomial are outside the complex unit circle. In that case x_t is $I(0)$ and it is convenient to assume that the DGP is defined for all $t \in \mathbb{Z}$. The process x_t then has time invariant means, variances and covariance structure. Therefore, without prior notice, I assume $t \in \mathbb{Z}$ if x_t is stable. If, however, the determinantal polynomial in (2.3) has a root for $z = 1$ (i.e., a *unit root*) and all other roots outside the complex unit circle, then some or all of the variables are integrated and there may be cointegration. In that case, assuming that the specification in (2.2) holds for $t \in \mathbb{N}$ is convenient. The initial values $x_{-p+1}, \dots, x_0$ are then assumed to be fixed values. For simplicity it may in fact be assumed that they are zero. This will be assumed if x_t is $I(1)$ and nothing else is explicitly specified. Recall that all variables are either $I(0)$ or $I(1)$ by default. Also, recall that y_t is the vector of observed variables, whereas x_t is the (typically unobserved) stochastic part.

Using the lag operator, the process (2.2) can be written more compactly as

$$A(L)x_t = u_t, \tag{2.4}$$

where $A(L) = I_K - A_1 L - \cdots - A_p L^p$ is a matrix polynomial in the lag operator of order p. If $\mu_t = \mu_0 + \mu_1 t$, then pre-multiplying (2.1) by $A(L)$ shows that y_t has the VAR(p) representation $A(L)y_t = A(L)\mu_t + u_t$ or

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

where $\nu_0 = (I_K - \sum_{j=1}^{p} A_j)\mu_0 + (\sum_{j=1}^{p} jA_j)\mu_1$ and $\nu_1 = (I_K - \sum_{j=1}^{p} A_j)\mu_1$. This model form is often called the *levels form of the VAR process* because it is purely in the levels of the y_t variables.

If the process (2.5) is the point of departure with unrestricted parameters ν_i, $i = 0, 1$, the variables may in fact have quadratic trends if $y_t \sim I(1)$. Thus, the additive model setup (2.1) is restrictive and in particular imposes restrictions on the deterministic parameters in (2.5). The setup is useful in theoretical derivations, however. Generally it may also be a good idea to think about the necessary deterministic terms at the beginning of the analysis and allow for the appropriate polynomial order. Sometimes subtracting the deterministic term before analyzing the stochastic part is helpful and the latter part is often of main interest in econometric analysis because it usually describes the behavioral relations.

2.2 The VECM Representation

If $y_t \sim I(1)$ and the variables are potentially cointegrated, the levels form of the VAR in (2.5) may not be the most useful representation because it does not contain the cointegration relations explicitly and these relations are often of particular interest. In that case it may be advantageous to reparameterize the model (2.2) by subtracting x_{t-1} on both sides of the equality sign and rearranging terms so as to obtain

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t. \tag{2.6}$$

Here $\Pi = -(I_K - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ for $j = 1, \ldots, p-1$. This representation is known as the *vector error correction model* (VECM) form of the VAR(p). Notice that Δx_t does not contain stochastic trends because $x_t \sim I(1)$ by assumption. Thus, the term Πx_{t-1} is the only one which includes $I(1)$ variables and, consequently, Πx_{t-1} must also be $I(0)$. Hence, it must contain the cointegration relations. The term Πx_{t-1} is often referred to as the *long-run* or *long-term part* or the *error correction* or *equilibrium correction term* of the model. On the other hand, the short-run movements of the variables are determined by the Γ_j's ($j = 1, \ldots, p-1$) which are sometimes called *short-term* or *short-run parameters*.

If $y_t \sim I(1)$ and the polynomial in (2.3) has a unit root, that is, $\det(I_K - A_1 z - \cdots - A_p z^p) = 0$ for $z = 1$, the matrix Π is singular. Let $\mathrm{rk}(\Pi) = r$, $0 < r < K$. Then it is known from matrix theory that there exist $(K \times r)$ matrices α and β with $\mathrm{rk}(\alpha) = \mathrm{rk}(\beta) = r$ such that $\Pi = \alpha\beta'$. Premultiplying $\Pi x_{t-1} = \alpha\beta' x_{t-1}$ by $(\alpha'\alpha)^{-1}\alpha'$ shows that $\beta' x_{t-1}$ is $I(0)$ and, hence, there are r linearly independent cointegration relations among the components of x_t. The rank r of Π is called the *cointegrating rank* of the process. Clearly, the matrices α and β are not unique. Choosing any nonsingular $(r \times r)$ matrix P, another decomposition $\Pi = \alpha^* \beta^{*\prime}$ with $\alpha^* = \alpha P^{-1}$ and $\beta^* = \beta P'$ is obtained. Sometimes unique or, in econometric terminology, identified cointegration relations are obtained from subject matter theory.

The cointegration relations are meant to represent the long-run or equilibrium economic relations. This may be helpful knowledge in specifying restrictions. For example, if there is just one cointegration relation it suffices to normalize one coefficient to 1 and thereby identification of the cointegration parameters is achieved. Suppose, for instance, that the system of interest consists of the variables log income (gdp_t), log money stock (m_t) and an

interest rate (r_t) and there is one cointegration relation, $ec_t = \beta_1 gdp_t + \beta_2 m_t + \beta_3 r_t$. If this relation represents a money demand function, it makes sense to normalize the coefficient of m_t so that

$$m_t = \gamma_1 gdp_t + \gamma_2 r_t + ec_t^*,$$

where $\gamma_1 = -\beta_1/\beta_2$, $\gamma_2 = -\beta_3/\beta_2$ and $ec_t^* = ec_t/\beta_2$.

As another example consider a four-dimensional system of interest rates, $(R_{1t}, r_{1t}, R_{2t}, r_{2t})'$, where R_{it} and r_{it} are a long-term and a short-term interest rate, respectively, of country i, $i = 1, 2$. Suppose that all variables are $I(1)$. In this case the expectations hypothesis of the term-structure suggests that the interest rate spreads, $R_{1t} - r_{1t}$ and $R_{2t} - r_{2t}$, are $I(0)$ and uncovered interest rate parity implies that $R_{1t} - R_{2t}$ is $I(0)$. Thus, there may be three cointegration relations with a cointegration matrix

$$\beta' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}.$$

Even if the expected long-run relations do not hold precisely, there may, for instance, be cointegration relations $R_{1t} + \beta_{21} r_{1t}$, $R_{2t} + \beta_{42} r_{2t}$ and $R_{1t} + \beta_{33} R_{2t}$, so that

$$\beta' = \begin{bmatrix} 1 & \beta_{21} & 0 & 0 \\ 0 & 0 & 1 & \beta_{42} \\ 1 & 0 & \beta_{33} & 0 \end{bmatrix}$$

with unknown parameters β_{21}, β_{42}, β_{33} and normalized parameters $\beta_{11} = \beta_{32} = \beta_{13} = 1$. The zero restrictions on the remaining parameters ensure identification. Johansen (1995, Section 5.3) discusses general conditions for the cointegration matrix β to be identified.

If no suitable restrictions are available from elsewhere, a purely statistical normalization of the form

$$\beta' = [I_r : \beta'_{(K-r)}] \tag{2.7}$$

with a $((K-r) \times r)$ matrix $\beta_{(K-r)}$, is always possible if the variables are arranged appropriately. In other words, we can always choose β such that its upper part is an $(r \times r)$ identity matrix.

The cointegrating rank r is necessarily between 0 and K. If $\text{rk}(\Pi) = K$, the process x_t is $I(0)$. Moreover, for $r = 0$, the term Πx_{t-1} disappears in (2.6) and, hence, x_t has a stable VAR$(p-1)$ representation in first differences. Although these limit cases do not represent cointegrated systems in the usual sense, they are included for convenience. Other cases where no cointegration in a strict sense is present although the representation (2.6) has a cointegrating rank strictly between 0 and K arise, for instance, if all variables but one are $I(0)$ then the cointegrating rank is $K-1$ although the $I(1)$ variable is not really cointegrated with the other variables.

Pre-multiplying (2.1) by the matrix operator $\Delta I_K - \alpha\beta' L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1}\Delta L^{p-1}$ and rearranging terms shows that y_t has the VECM representation

$$\Delta y_t = \nu_0 + \nu_1 t + \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{2.8}$$

if $\mu_t = \mu_0 + \mu_1 t$. Here ν_0 and ν_1 can be obtained from μ_0 and μ_1 by the same expressions as in (2.5). Thus, $\nu_1 = -\Pi\mu_1 = -\alpha\beta'\mu_1$ so that the trend term can be absorbed into the cointegration relations,

$$\Delta y_t = \nu + \alpha[\beta', \eta] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{2.9}$$

with $\eta = -\beta'\mu_1$ being an $(r \times 1)$ vector and $\nu = \nu_0 + \nu_1$.

If μ_1 happens to be orthogonal to β, that is, $\beta'\mu_1 = 0$, then

$$\Delta y_t = \nu + \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{2.10}$$

and there is just an intercept in this model although some variables may have deterministic linear trends. This property is due to the fact that stochastic trends can generate linear deterministic trends if there is just a constant in the model. If none of the variables has a linear trend so that $\mu_1 = 0$, then $\nu = \nu_0 = -\Pi\mu_0 = -\alpha\beta'\mu_0$ and the constant term can be absorbed into the cointegration relations,

$$\Delta y_t = \alpha[\beta', \eta_0] \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{2.11}$$

where $\eta_0 = -\beta'\mu_0$. No separate intercept term is needed in the model in this case.

2.3 Structural Forms

The previously considered model forms hide all contemporaneous relations between the observable variables in the white noise covariance matrix Σ_u. All right-hand side variables are lagged or predetermined. Such model forms are called *reduced forms*. Sometimes it is desirable to model also the contemporaneous relations between the variables. This can be done with *structural form models* where contemporaneous variables may appear as explanatory variables in some equations. For example, a structural form associated with the VAR representation (2.5) may be of the form

$$\mathsf{A}y_t = \nu_0^* + \nu_1^* t + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + v_t, \tag{2.12}$$

where $\nu_i^* = \mathsf{A}\nu_i$ $(i = 0, 1)$ and $A_j^* = \mathsf{A}A_j$ $(j = 1, \ldots, p)$. The structural form error term $v_t = \mathsf{A}u_t$ is iid white noise with covariance matrix $\Sigma_v = \mathsf{A}\Sigma_u\mathsf{A}'$. The $(K \times K)$ matrix A is nonsingular and describes the instantaneous relations between the variables. Clearly, one could multiply (2.5) by any nonsingular matrix to obtain a representation of the form (2.12). Thus, the parameters of the structural form (2.12) are not identified without further restrictions. Structural form models will play an important role in analyzing the relations between the variables with impulse responses. I will return to them in that context. Before a structural analysis is conducted, a reduced form model as a valid description of the DGP is usually constructed. I will discuss the relevant stages of a VAR analysis in the following. Before issues related to model specification are considered, it is useful to discuss estimation of fully specified models because estimation of various models is usually necessary at the specification stage.

3 Estimation of VAR Models

Estimating unrestricted reduced form VAR models is computationally straightforward. I will discuss estimators for the levels VAR form and VECMs first in Sections 3.1 and 3.2, respectively. Estimation of models with linear constraints is considered in Section 3.3 and some comments on Bayesian estimation are provided in Section 3.4.

3.1 Estimation of Unrestricted VARs

Consider the levels VAR(p) model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{3.1}$$

Deterministic terms are deleted for simplicity. Including them is straightforward and I will comment on them later. Given a sample of size T, $y_1, \ldots, y_T$, and p presample values, $y_{-p+1}, \ldots, y_0$, I define

$$Y = [y_1, \ldots, y_T], \quad \mathbf{Y} = [Y_0, \ldots, Y_{T-1}], \quad \text{where} \quad Y_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}, \tag{3.2}$$

$$A = [A_1, \ldots, A_p] \quad \text{and} \quad U = [u_1, \ldots, u_T].$$

Using this notation, the model can be written compactly as

$$Y = A\mathbf{Y} + U. \tag{3.3}$$

Estimating the K equations separately by LS results in the estimator

$$\hat{A} = [\hat{A}_1, \ldots, \hat{A}_p] = Y\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}. \tag{3.4}$$

It was shown by Zellner (1962) that this estimator is identical to GLS estimation, if no restrictions are imposed on the parameter matrix A. If the process is normally distributed (Gaussian) or, equivalently, $u_t \sim \mathcal{N}(0, \Sigma_u)$, this estimator is also identical to the ML estimator (conditional on the initial values). Consequently there is no loss in asymptotic estimation efficiency.

If the process is stable ($I(0)$), the LS estimator $\hat{A}$ has an asymptotic normal distribution under general conditions (see, e.g., Lütkepohl (2005, Chapter 3)),

$$\sqrt{T}\text{vec}(\hat{A} - A) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{A}}). \tag{3.5}$$

The covariance matrix of the asymptotic distribution is $\Sigma_{\hat{A}} = \text{plim}(\mathbf{Y}\mathbf{Y}'/T)^{-1} \otimes \Sigma_u$, which can be estimated consistently by $\hat{\Sigma}_{\hat{A}} = (\mathbf{Y}\mathbf{Y}'/T)^{-1} \otimes \hat{\Sigma}_u$, where, for example,

$$\hat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' \tag{3.6}$$

can be used as an estimator for Σ_u. Here $\hat{u}_t = y_t - \hat{A}Y_{t-1}$ are the LS residuals. Thus,

$$\text{vec}(\hat{A}) \approx \mathcal{N}(\text{vec}(A), (\mathbf{Y}\mathbf{Y}')^{-1} \otimes \hat{\Sigma}_u) \tag{3.7}$$

is an intuitive way of writing the result in (3.5). Clearly, the covariance matrix is a stochastic matrix which makes the statement imprecise. It has the advantage, however, that by pretending that the result is precise and using it in the usual way to set up t-, χ^2- and F-statistics results in asymptotically correct inference.

If $y_t \sim I(1)$ and, hence, the process is not stable and the variables may be cointegrated, (3.5) still holds (see Park and Phillips (1988, 1989), Sims, Stock and Watson (1990), Lütkepohl (2005, Chapter 7)). However, in that case the covariance matrix $\Sigma_{\hat{A}}$ is singular because some estimated parameters or linear combinations of parameters converge with a faster rate than $\sqrt{T}$, as will be seen in the next subsection. In this situation the usual t-, χ^2- and F-tests for inference regarding the VAR parameters may not be valid asymptotically anymore (Toda and Phillips (1993)). Despite this general result, there are many cases where asymptotic inference remains valid. Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) have shown that, if a null hypothesis is considered which does not restrict elements of each of the A_i's ($i = 1, \ldots, p$), then the usual tests have their standard asymptotic properties. For example, for a bivariate VAR(2) process with coefficient matrices

$$A_1 = \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} \qquad \text{and} \qquad A_2 = \begin{bmatrix} \alpha_{11,2} & \alpha_{12,2} \\ \alpha_{21,2} & \alpha_{22,2} \end{bmatrix},$$

standard Wald tests of the following hypotheses have the usual asymptotic χ^2-distributions: $H_0 : \alpha_{11,1} = 0$, $H_0 : \alpha_{11,2} = \alpha_{21,2} = 0$ or even $H_0 : \alpha_{11,1} = \alpha_{21,1} = \alpha_{12,1} = \alpha_{22,1} = 0$. On the other hand, the Wald statistic for testing $H_0 : \alpha_{12,1} = \alpha_{12,2} = \alpha_{21,2} = 0$ may not have a standard asymptotic $\chi^2(3)$ distribution, even if H_0 is true, because the null hypothesis places restrictions on both coefficient matrices. Generally, for a VAR(p) with $p \geq 2$, the t-ratios have their usual asymptotic standard normal distributions because they are test statistics for hypotheses regarding individual coefficients. I will discuss a specific, problematic case in Section 7.2, when causality tests are considered.

Deterministic terms can be included easily in the model (3.3) by augmenting the regressor vectors Y_{t-1} accordingly. For example, a one or a one and a t may be included as additional regressors. The previously stated formulas apply with these modifications. Although the

convergence rates of some of the parameter estimators may not be $\sqrt{T}$, the result in (3.7) can still be used for constructing asymptotically valid test statistics in the $I(0)$ case. In the $I(1)$ case, the situation is complicated by the fact that some deterministic terms may be absorbed into the cointegration relations (see Section 2.2). However, the asymptotic properties of the estimators of the VAR parameters remain essentially the same as in the case without deterministic terms (Sims et al. (1990)). Still, estimating the parameters in the VECM framework may be advantageous in some respects. Estimation of VECMs is therefore considered in the following section.

3.2 Estimation of VECMs

Consider a VECM without deterministic terms,

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t. \tag{3.8}$$

Using the definitions

$$\Delta Y = [\Delta y_1, \ldots, \Delta y_T], \qquad Y_{-1} = [y_0, \ldots, y_{T-1}], \qquad \Gamma = [\Gamma_1, \ldots, \Gamma_{p-1}],$$

$$\Delta \mathbf{X} = [\Delta X_0, \ldots, \Delta X_{T-1}] \quad \text{with} \quad \Delta X_{t-1} = \begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix}$$

and

$$U = [u_1, \ldots, u_T],$$

the model can be written in compact matrix form as

$$\Delta Y = \alpha\beta' Y_{-1} + \Gamma \Delta \mathbf{X} + U. \tag{3.9}$$

In the following, ML and feasible GLS estimation of the parameters of this model will be considered. A complication arises from the error correction term $\alpha\beta' Y_{-1}$ which involves a product of parameters and, hence, the usual equationwise LS estimation cannot be used directly. However, if the product $\alpha\beta'$ were known, LS estimation of the other parameters were possible. In that case, the LS estimator for Γ is

$$\hat{\Gamma}(\alpha\beta') = (\Delta Y - \alpha\beta' Y_{-1})\Delta\mathbf{X}'(\Delta\mathbf{X}\Delta\mathbf{X}')^{-1}. \tag{3.10}$$

Replacing Γ in (3.9) by this estimator gives

$$\Delta Y = \alpha\beta' Y_{-1} + (\Delta Y - \alpha\beta' Y_{-1})\Delta\mathbf{X}'(\Delta\mathbf{X}\Delta\mathbf{X}')^{-1}\Delta\mathbf{X} + U^*$$

which involves just the parameters $\alpha\beta'$. Defining

$$\mathbf{M} = I_T - \Delta\mathbf{X}'(\Delta\mathbf{X}\Delta\mathbf{X}')^{-1}\Delta\mathbf{X}, \quad R_0 = \Delta Y\mathbf{M} \quad \text{and} \quad R_1 = Y_{-1}\mathbf{M},$$

gives

$$R_0 = \alpha\beta' R_1 + U^*. \tag{3.11}$$

In the following, different methods for estimating α and β from (3.11) will be discussed. The estimator for Γ is then obtained by substituting the estimators for α and β in (3.10).

3.2.1 ML Estimation for Gaussian Processes

The first estimation method is ML under Gaussian assumptions as proposed by Johansen (1988), also known as reduced rank (RR) regression (see also Anderson (1951)). The objective is to choose the estimators for α and β such that

$$\log l = -\frac{KT}{2}\log 2\pi - \frac{T}{2}\log|\Sigma_u| - \frac{1}{2}\text{tr}\left[(R_0 - \alpha\beta' R_1)'\Sigma_u^{-1}(R_0 - \alpha\beta' R_1)\right] \tag{3.12}$$

is maximized. The solution is

$$\tilde{\beta} = [v_1, \ldots, v_r]' S_{11}^{-1/2} \quad \text{and} \quad \tilde{\alpha} = S_{01}\tilde{\beta}(\tilde{\beta}' S_{11}\tilde{\beta})^{-1}, \tag{3.13}$$

where $S_{ij} = R_i R_j'/T$, $i = 0, 1$, and $v_1, \ldots, v_K$ are the orthonormal eigenvectors of the matrix $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$ corresponding to the eigenvalues in nonincreasing order.

Normalizing β such that $\beta' = [I_r : \beta_{(K-r)}']$ as in (2.7) and the ML estimator accordingly, it can be shown that, under general conditions, $T\text{vec}(\tilde{\beta}_{(K-r)}' - \beta_{(K-r)}')$ converges in distribution to a mixture normal distribution (Johansen (1995) or Lütkepohl (2005, Chapter 7)). In other words, the estimator of the cointegration parameters converges at the faster rate T rather than just $\sqrt{T}$ as the other parameters. It is therefore sometimes called superconsistent. The fact that the asymptotic distribution of the cointegration parameter estimators is mixed normal implies that inference can be performed as with asymptotically normal estimators. For example, t-ratios have the usual asymptotic interpretation.

Notice also that the estimator for α is the LS estimator obtained from a multivariate regression model

$$R_0 = \alpha\tilde{\beta}' R_1 + U^*$$

with regressor matrix $\tilde{\beta}' R_1$. Its asymptotic properties are standard as are those of the resulting estimator for Γ,

$$\tilde{\Gamma}(\tilde{\alpha}\tilde{\beta}') = (\Delta Y - \tilde{\alpha}\tilde{\beta}' Y_{-1})\Delta\mathbf{X}'(\Delta\mathbf{X}\Delta\mathbf{X}')^{-1}.$$

In particular, the estimators converge at the usual rate $\sqrt{T}$ to an asymptotic normal distribution under general conditions. Unfortunately, in small samples the ML estimator produces occasional outlying estimates far away from the true parameter values, as shown, e.g., by Brüggemann and Lütkepohl (2005). Therefore it may be worth considering the more robust GLS estimator which is discussed next.

3.2.2 Feasible GLS Estimation

Using the normalization $\beta' = [I_r : \beta'_{(K-r)}]$ given in (2.7), the equation (3.11) can be rewritten as

$$R_0 - \alpha R_1^{(1)} = \alpha\beta'_{(K-r)} R_1^{(2)} + U^*, \tag{3.14}$$

where $R_1^{(1)}$ and $R_1^{(2)}$ denote the first r and last $K-r$ rows of R_1, respectively. For a known α, the GLS estimator of $\beta'_{(K-r)}$ based on this specification can be shown to be

$$\hat{\beta}'_{(K-r)} = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}(R_0 - \alpha R_1^{(1)})R_1^{(2)\prime}\left(R_1^{(2)}R_1^{(2)\prime}\right)^{-1} \tag{3.15}$$

(see Lütkepohl (2005, Chapter 7)). A feasible GLS estimator can be obtained by estimating the matrix Π from $R_0 = \Pi R_1 + U^*$ with unrestricted equationwise LS and using $\Pi = [\alpha : \alpha\beta'_{(K-r)}]$. Thus, the first r columns of the estimator for Π can be used as an estimator for α, say $\hat{\alpha}$. Substituting this estimator and the LS estimator of the white noise covariance matrix in (3.15) gives the feasible GLS estimator

$$\hat{\hat{\beta}}'_{(K-r)} = (\hat{\alpha}'\hat{\Sigma}_u^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\hat{\Sigma}_u^{-1}(R_0 - \hat{\alpha}R_1^{(1)})R_1^{(2)\prime}\left(R_1^{(2)}R_1^{(2)\prime}\right)^{-1}. \tag{3.16}$$

This estimator was proposed by Ahn and Reinsel (1990) and Saikkonen (1992) (see also Reinsel (1993, Chapter 6)). It has the same asymptotic properties as the ML estimator and

also the asymptotic properties of the associated estimators of α and Γ are the same as in the previous section.

Including deterministic terms in the ML or GLS estimation procedures is straightforward. All that needs to be done is to include the required terms in the list of regressors in the short-term dynamics or the cointegration term as appropriate. The asymptotic properties of the resulting estimators are the usual ones so that standard inference can be used under usual assumptions.

3.3 Estimation with Linear Restrictions

If restrictions are imposed on the parameters of the levels VAR or VECM representations, the previously discussed estimation methods may be asymptotically inefficient. However, efficient GLS methods may be used. They are easy to implement as long as no (overidentifying) restrictions are imposed on the cointegration matrix. Zero restrictions are the most common constraints for the parameters of these models. I will therefore focus on such restrictions in the following. Suppose first that a levels VAR form (3.1) is of interest and there are zero restrictions for $A = [A_1, \ldots, A_p]$. Defining $\boldsymbol{\alpha} = \text{vec}(A)$, they can be written in the form

$$\boldsymbol{\alpha} = R\boldsymbol{\gamma}, \tag{3.17}$$

where R is a known $(K^2p \times M)$ matrix of zeros and ones with rank M and $\boldsymbol{\gamma}$ is the $(M \times 1)$ vector of all unrestricted parameters. Using the rule $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ for conformable matrices A, B, C and vectorizing the compact model form (3.3) as

$$\text{vec}(Y) = (\mathbf{Y}' \otimes I_K)\text{vec}(A) + \text{vec}(U) = (\mathbf{Y}' \otimes I_K)R\boldsymbol{\gamma} + \text{vec}(U) \tag{3.18}$$

shows that the GLS estimator for $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = [R'(\mathbf{Y}\mathbf{Y}' \otimes \Sigma_u^{-1})R]^{-1}R'(\mathbf{Y} \otimes \Sigma_u^{-1})\text{vec}(Y). \tag{3.19}$$

Here it has been used that the covariance matrix of $\text{vec}(U)$ is $I_T \otimes \Sigma_u$. The estimator $\hat{\boldsymbol{\gamma}}$ has standard asymptotic properties if $y_t \sim I(0)$, that is, the GLS estimator is consistent and asymptotically normally distributed,

$$\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0, (R'\Sigma_{\hat{A}}^{-1}R)^{-1}). \tag{3.20}$$

If the white noise covariance matrix is unknown, as is usually the case in practice, it may be replaced by an estimator based on an unrestricted estimation of the model. The resulting feasible GLS estimator, say $\hat{\hat{\boldsymbol{\gamma}}}$, has the same asymptotic properties as the GLS estimator. The corresponding feasible GLS estimator of $\boldsymbol{\alpha}$, $\hat{\hat{\boldsymbol{\alpha}}} = R\hat{\hat{\boldsymbol{\gamma}}}$, is also consistent and asymptotically normal,

$$\sqrt{T}(\hat{\hat{\boldsymbol{\alpha}}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, R(R'\Sigma_{\hat{A}}^{-1}R)^{-1}R'). \tag{3.21}$$

The feasible GLS estimator may be iterated by reestimating the white noise covariance matrix from the first round feasible GLS residuals and using that estimator in the next round. The procedure may be continued until convergence. The resulting estimators for $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ will have the same asymptotic distributions as without iteration. For Gaussian white noise u_t, ML estimation may also be used. Its asymptotic properties are also the same as those of the GLS estimator. Deterministic terms can be included by modifying the regressor matrix in the foregoing formulas and are, hence, straightforward to deal with.

If y_t is $I(1)$, it is useful to impose possible restrictions on the VECM form. In case there are no restrictions for the cointegration matrix, the cointegration parameters may be estimated in a first round as in Section 3.2, ignoring the restrictions. Let the estimator be $\hat{\beta}$. In a second stage, the remaining parameters may then be estimated from

$$\Delta y_t = \alpha\hat{\beta}' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t. \tag{3.22}$$

Conditional on $\hat{\beta}$ this is a linear system and a feasible GLS procedure may be used just like for the levels VAR form. The properties of the estimators are the same as if the cointegration matrix β were known.

If there are also restrictions for β, one may use nonlinear optimization algorithms to obtain ML estimators of all parameters simultaneously, provided the β parameters are identified. It is also possible to use a two-step procedure which estimates the restricted β matrix first and in a second step conditions on that estimator. Restricted estimation of β is treated, for instance, by Johansen (1995), Boswijk and Doornik (2002) and Lütkepohl (2005, Chapter 7).

3.4 Bayesian Estimation of VARs

Because the levels VAR form is linear in the parameters, standard Bayesian methods for estimating linear regression models can be applied for estimating its parameters. Since Bayesian methods are treated elsewhere in this volume, I will not discuss these methods in detail here. It is worth noting, however, that specific priors have been popular in VAR analysis. For example, in the earlier literature the so-called Minnesota prior was quite popular (see Doan, Litterman and Sims (1984), Litterman (1986)). It shrinks the VAR towards a random walk for each of the variables. Recently proposals have been made to shrink towards some dynamic stochastic general equilibrium (DSGE) model (e.g., Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004)). A more detailed presentation of Bayesian methods in VAR analysis is also given by Canova (2007, Chapters 9 - 11).

For VECMs the situation is complicated by the matrix product in the error correction term. It makes straightforward application of linear Bayesian methods problematic. These problems and possible solutions are well documented in the literature. Important publications on the subject are Kleibergen and van Dijk (1994), Bauwens and Lubrano (1996), Geweke (1996), Kleibergen and Paap (2002), Strachan and Inder (2004). Koop, Strachan, van Dijk and Villani (2005) provide a survey with many more references.

4 Model Specification

Model specification in the present context involves selection of the VAR order and, in VECMs, also choosing the cointegrating rank. Because the number of parameters in these models increases with the square of the number of variables it is also often desirable to impose zero restrictions on the parameter matrices and thereby eliminate some lagged variables from some of the equations of the system. Various algorithms exist that can assist in specifying these so-called subset restrictions. They are treated in detail in Chapter ?? of this Handbook and are therefore not considered here. Lag order selection and testing for the cointegrating rank of a VAR process are discussed next.

4.1 Choosing the Lag Order

The most common procedures for VAR order selection are sequential testing procedures and application of model selection criteria. These approaches will be discussed in turn.

4.1.1 Sequential Testing

Given a maximum reasonable order, say $p_{\max}$, for a VAR model, the following sequence of null hypotheses can be tested to determine the lag order: $H_0 : A_{p_{\max}} = 0$, $H_0 : A_{p_{\max}-1} = 0$, etc.. The testing procedure stops and the lag order is chosen accordingly when the null hypothesis is rejected for the first time. Because the parameter estimators have standard asymptotic properties, the usual Wald or LR χ^2 tests for parameter restrictions can be used for this purpose if the process is stationary. In fact, the discussion in Section 3.1 implies that even if some of the variables are $I(1)$ these tests will have standard asymptotic properties as long as the null hypothesis $H_0 : A_1 = 0$ is not tested. There is evidence, however, that the tests have small sample distributions which are quite different from their asymptotic χ^2 counterparts, especially if systems with more than a couple of variables are studied (e.g., Lütkepohl (2005, Section 4.3.4)). Therefore it may be useful to consider small sample adjustments, possibly based on bootstrap methods (e.g., Li and Maddala (1996), Berkowitz and Kilian (2000)).

The lag order obtained with such a procedure depends to some extent on the choice of $p_{\max}$. If $p_{\max}$ is chosen quite small, an appropriate model may not be in the set of possibilities and, hence, it cannot be found. If, on the other hand, $p_{\max}$ is chosen excessively large, there is a high chance to end up with too large an order because in each test there is a chance of rejecting a true null hypothesis and, hence, for committing a Type I error. Generally, at an early stage of the analysis, using a moderate value for $p_{\max}$ appears to be a sensible strategy because any problems caused by an inadequate choice should be detected at the model checking stage (see Section 5).

The procedure can in fact also be used if the DGP under consideration does not have a finite order VAR representation although in that case a 'true' finite VAR order does not exist. Ng and Perron (1995) have considered this case and discuss some consequences.

4.1.2 Model Selection Criteria

The standard model selection criteria which are used in this context choose the VAR order which minimizes them over a set of possible orders $m = 0, \ldots, p_{\max}$. The general form of a set of such criteria is

$$C(m) = \log\det(\hat{\Sigma}_m) + c_T\varphi(m), \tag{4.1}$$

where $\hat{\Sigma}_m = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ is the residual covariance matrix estimator for a model of order m, $\varphi(m)$ is a function of the order m which penalizes large VAR orders and c_T is a sequence which may depend on the sample size and identifies the specific criterion. The term $\log\det(\tilde{\Sigma}_m)$ is a nonincreasing function of the order m while $\varphi(m)$ increases with m. The lag order is chosen which optimally balances these two forces.

Examples of criteria of this type are Akaike's information criterion (Akaike (1973, 1974)),

$$\mathrm{AIC}(m) = \log\det(\hat{\Sigma}_m) + \frac{2}{T}mK^2,$$

where $c_T = 2/T$, the Hannan-Quinn criterion (Hannan and Quinn (1979), Quinn (1980)),

$$\mathrm{HQ}(m) = \log\det(\hat{\Sigma}_m) + \frac{2\log\log T}{T}mK^2,$$

for which $c_T = 2\log\log T/T$, and the Schwarz (or Rissanen) criterion (Schwarz (1978), Rissanen (1978)),

$$\mathrm{SC}(m) = \log\det(\hat{\Sigma}_m) + \frac{\log T}{T}mK^2,$$

with $c_T = \log T/T$. In each case $\varphi(m) = mK^2$ is the number of VAR parameters in a model with order m. Denoting by $\hat{p}(\mathrm{AIC})$, $\hat{p}(\mathrm{HQ})$ and $\hat{p}(\mathrm{SC})$ the orders selected by AIC, HQ and SC, respectively, the following relations hold for samples of fixed size $T \geq 16$:

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

Thus, AIC always suggests the largest order, SC chooses the smallest order and HQ is in between (Lütkepohl (2005, Chapters 4 and 8)). Of course, this does not preclude the possibility that all three criteria agree in their choice of VAR order. The HQ and SC criteria are both consistent, that is, the order estimated with these criteria converges in probability or almost surely to the true VAR order p under quite general conditions, if $p_{\max}$ exceeds the true order. On the other hand, the AIC criterion tends to overestimate the order asymptotically. These results hold for both $I(0)$ and $I(1)$ processes (Paulsen (1984)).

4.2 Choosing the Cointegrating Rank of a VECM

A great number of proposals have been made for determining the cointegrating rank of a VAR process. Many of them are reviewed and compared in Hubrich, Lütkepohl and Saikkonen (2001). Generally, there is a good case for using the Johansen (1995) likelihood ratio (LR) approach based on Gaussian assumptions and its modifications because all other approaches were found to have shortcomings in some situations. Even if the actual DGP is not Gaussian, the resulting pseudo LR tests may have better properties than many competitors. These tests are also attractive from a computational point of view because, for a given cointegrating rank r, ML estimates and, hence, the likelihood maximum are easy to compute (see Section 3.2). Of course, if there are specific reasons, for example, due to special data properties, it may be worth using alternative tests.

Denoting the matrix $\alpha\beta'$ in the error correction term by Π, the following sequence of hypotheses may be considered for selecting the cointegrating rank:

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

The cointegrating rank specified in the first null hypothesis which cannot be rejected is then chosen as estimate for the true cointegrating rank r. If $H_0(0)$, the first null hypothesis in this sequence, cannot be rejected, a VAR process in first differences is considered. If all the null hypotheses can be rejected including $H_0(K-1)$, the process is treated as $I(0)$ and a levels VAR model is specified.

The LR statistics corresponding to the null hypotheses in (4.2) have nonstandard asymptotic distributions. They depend on the difference $K - r_0$ and on the deterministic terms included in the DGP but not on the short-term dynamics. Critical values for various possible sets of deterministic components such as constants and linear trends have been computed by simulation methods and are available in the literature (e.g., Johansen (1995, Chapter 15)).

The power of the tests can be improved by specifying the deterministic terms as tightly as possible. For example, if there is no deterministic linear trend term, it is desirable to perform the cointegration rank tests without such terms. On the other hand, leaving them out if they are part of the DGP can lead to major distortions in the tests. Johansen (1995) also provided the asymptotic theory for testing hypotheses regarding the deterministic terms which can be helpful in this respect.

Test versions have also been developed for the case where a structural break occurs in the deterministic term either in the form of a level shift or a break in the trend slop or both. In this case the critical values of the LR tests depend also on the timing of the break. This feature is inconvenient if the break point is not known a priori and also has to be estimated. In that case a test variant proposed by Saikkonen and Lütkepohl (2000a, b) may be preferable. They suggest to estimate the deterministic term first by a GLS procedure and adjust the series before an LR type test is applied to the adjusted process. The advantage is that the asymptotic null distribution of the test statistic does not depend on the break point if just a level shift is considered. This fact makes it possible to develop procedures which work when the break date is not known (e.g., Lütkepohl, Saikkonen and Trenkler (2004), Saikkonen, Lütkepohl and Trenkler (2006)).

Although the short-term dynamics do not matter for the asymptotic theory, they have a substantial impact in small and moderate samples. Therefore the choice of the lag order p is quite important. Choosing p rather large to be on the save side as far as missing out on important short-term dynamics is concerned, may lead to a drastic loss in power of the cointegrating rank tests. On the other hand, choosing the lag order too small may lead to dramatic size distortions even for well-behaved DGPs. In a small sample simulation study, Lütkepohl and Saikkonen (1999) found that using the AIC criterion for order selection may be a good compromise.

There are many other interesting suggestions for modifying and improving the Johansen approach to cointegration testing. For example, to improve the performance of the Johansen cointegration tests in small samples, Johansen (2002) presents a Barlett correction. Also, there are a number of proposals based on different ideas. As mentioned previously, much of the earlier literature is reviewed in Hubrich et al. (2001). Generally, at present it appears that the Johansen approach should be the default and only if there are particular reasons other proposals are worth contemplating.

Clearly, the Johansen approach has its drawbacks. In particular, in large dimensional systems and if long lag orders are necessary to capture the short-term dynamics, it may not find all the cointegration relations (see Gonzalo and Pitarakis (1999)). In other words, its power may not suffice to reject some small cointegration rank if the true rank is a bit larger. Therefore it may be useful to apply cointegration tests to all possible subsystems as well and

check whether the results are consistent with those for the full model.

5 Model Checking

A wide range of procedures is available for checking the adequacy of VARs and VECMs. They should be applied before a model is used for a specific purpose to ensure that it represents the DGP adequately. A number of procedures considers the estimated residuals and checks whether they are in line with the white noise assumption. Another set of tests checks the stability of the model over time. In the following I will first consider residual based tests for autocorrelation, nonnormality and conditional heteroskedasticity. Then I will discuss tests for structural stability. In addition to these more formal procedures there are also many informal procedures based, e.g., on plots of residuals and autocorrelations. For some of these procedures see Lütkepohl (2004).

5.1 Tests for Residual Autocorrelation

5.1.1 Portmanteau Test

The portmanteau test for residual autocorrelation checks the null hypothesis that all residual autocovariances are zero, that is, $H_0 : E(u_t u'_{t-i}) = 0$ $(i = 1, 2, \dots)$. It is tested against the alternative that at least one autocovariance and, hence, one autocorrelation is nonzero. The test statistic is based on the residual autocovariances and has the form

$$Q_h = T \sum_{j=1}^{h} \mathrm{tr}(\hat{C}'_j \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}), \tag{5.1}$$

where $\hat{C}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}'_{t-j}$ and the $\hat{u}_t$'s are the estimated residuals. For an unrestricted stationary VAR(p) process the null distribution of Q_h can be approximated by a $\chi^2(K^2(h-p))$ distribution if T and h approach infinity such that $h/T \to 0$. If there are parameter restrictions, the degrees of freedom of the approximate χ^2 distribution are obtained as the difference between the number of (non-instantaneous) autocovariances included in the statistic ($K^2 h$) and the number of estimated VAR parameters (e.g., Ahn (1988), Hosking (1980, 1981a, 1981b), Li and McLeod (1981) or Lütkepohl (2005, Section 4.4)). This approximation is unsatisfactory for integrated and cointegrated processes, as pointed out by Brüggemann,

Lütkepohl and Saikkonen (2006). For such processes the degrees of freedom also depend on the cointegrating rank. Thus, portmanteau tests are not recommendable for levels VAR processes with unknown cointegrating rank. For VECMs with appropriately specified cointegrating rank r and no restrictions on α and $\Gamma_1, \ldots, \Gamma_{p-1}$, the proper approximate distribution is $\chi^2(K^2h - K^2(p-1) - Kr)$.

All these approximations of the test distributions may be poor in small samples. To improve the match between actual and approximating distribution, Hosking (1980) proposed to use the modified statistic

$$Q_h^* = T^2 \sum_{j=1}^{h} \frac{1}{T-j} \mathrm{tr}(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1})$$

instead of the original version (5.1).

The choice of h is also important for the small sample properties of the test. This quantity should be considerably larger than p to get a good approximation to the null distribution. Unfortunately, choosing h too large may reduce the power of the test. Often a number of different values of h are considered in practice. The portmanteau test should be applied primarily to test for autocorrelation of high order. The LM test considered in the next section is more suitable to check for low order residual autocorrelation.

5.1.2 LM Test

The LM test, also known as Breusch-Godfrey test for residual autocorrelation of order h may be viewed as a test for zero coefficient matrices in the model

$$u_t = B_1 u_{t-1} + \cdots + B_h u_{t-h} + e_t.$$

The quantity e_t denotes a white noise error term. Thus, a test of

$$H_0 : B_1 = \cdots = B_h = 0 \quad \text{versus} \quad H_1 : B_i \neq 0 \text{ for at least one } i \in \{1, \ldots, h\}$$

is called for. The relevant LM statistic can be computed easily by considering the auxiliary model

$$\hat{u}_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_1 \hat{u}_{t-1} + \cdots + B_h \hat{u}_{t-h} + e_t^*, \tag{5.2}$$

for the levels VAR form or

$$\hat{u}_t = \alpha \hat{\beta}' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + B_1 \hat{u}_{t-1} + \cdots + B_h \hat{u}_{t-h} + e_t^*, \tag{5.3}$$

for the VECM form. Here the $\hat{u}_t$'s are again the estimated residuals from the original model, $\hat{\beta}$ is the RR estimator of the cointegration matrix and e_t^* is an auxiliary error term. The $\hat{u}_t$'s with $t \leq 0$ should be replaced by zero. If the model contains deterministic terms they should also be added to the auxiliary model.

Denoting the residuals from the estimated auxiliary model by $\hat{e}_t^*$ $(t = 1, \ldots, T)$ and defining $\hat{\Sigma}_e = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t^* \hat{e}_t^{*\prime}$, the LM statistic may be computed as

$$Q_{LM} = T\left(K - \mathrm{tr}(\hat{\Sigma}_u^{-1}\hat{\Sigma}_e)\right).$$

It has an asymptotic $\chi^2(hK^2)$-distribution under the null hypothesis for both $I(0)$ and $I(1)$ systems (Brüggemann et al. (2006)). Thus, the LM test can also be applied to levels VAR processes with unknown cointegrating rank.

Edgerton and Shukur (1999) have performed a large Monte Carlo study for stationary processes and found that this statistic also may have a small sample distribution which differs considerably from its asymptotic χ^2-distribution. They propose an F version with better small sample properties.

5.2 Tests for Nonnormality

Although normality is not a necessary condition for the validity of many of the statistical procedures related to VAR models, deviations from the normality assumption may indicate that model improvements are possible. Therefore nonnormality tests are common in applied work. Multivariate versions can be applied to the full residual vector of the VAR model and univariate versions can be used for the errors of the individual equations.

Multivariate tests for nonnormality may be constructed to check whether the third and fourth moments of the residuals are conformable with those of a normal distribution. This approach extends ideas of Lomnicki (1961) and Jarque and Bera (1987) for univariate models. In the multivariate case, the residual vector of a VAR or VECM is first transformed to make the individual components independent. Then the moments are compared with those of normal distributions. For given residuals $\hat{u}_t$ $(t = 1, \ldots, T)$ of an estimated VAR process or VECM, the residual covariance matrix $\hat{\Sigma}_u$ is determined and a matrix P such that $PP' = \hat{\Sigma}_u$ is computed. The tests for nonnormality can then be based on the skewness and kurtosis of the standardized residuals $\hat{u}_t^s = P^{-1}\hat{u}_t$.

The nonnormality tests depend to some extent on the transformation matrix P which is used to standardize the residuals. Doornik and Hansen (1994) proposed to use the square root matrix of $\hat{\Sigma}_u$ whereas Lütkepohl (2005, Chapter 4) considered a Choleski decomposition of the residual covariance matrix. Although the literature on testing for nonnormality is extensive and many other tests are available, the ones mentioned here are probably the most popular tests in the context of VAR analysis. Nonnormality is also a likely problem if the residuals are conditionally heteroskedastic. Special tests for this feature are discussed next.

5.3 ARCH Tests

A test for multivariate autoregressive conditional heteroskedasticity (ARCH) can be based on similar ideas as the LM test for residual autocorrelation. A multivariate ARCH model of order q for the residual vector u_t has the form

$$\text{vech}(\Sigma_{t|t-1}) = \beta_0 + B_1\text{vech}(u_{t-1}u_{t-1}') + \cdots + B_q\text{vech}(u_{t-q}u_{t-q}'),$$

where vech is the column stacking operator for symmetric matrices which stacks the columns from the main diagonal downwards and $\Sigma_{t|t-1}$ is the conditional covariance matrix of u_t given $u_{t-1}, u_{t-2}, \ldots$. Moreover, β_0 is a $\frac{1}{2}K(K+1)$-dimensional parameter vector and the B_j's are $(\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1))$ coefficient matrices for $(j = 1, \ldots, q)$. For this model one may want to test the pair of hypotheses

$$H_0 : B_1 = \cdots = B_q = 0 \quad \text{versus} \quad H_1 : B_i \neq 0 \text{ for at least one } i \in \{1, \ldots, q\}.$$

If H_0 is true, there is no ARCH in the residuals. The relevant LM statistic can be obtained by using the auxiliary model

$$\text{vech}(\hat{u}_t\hat{u}_t') = \beta_0 + B_1\text{vech}(\hat{u}_{t-1}\hat{u}_{t-1}') + \cdots + B_q\text{vech}(\hat{u}_{t-q}\hat{u}_{t-q}') + \text{error}_t \tag{5.4}$$

and computing

$$LM_{ARCH}(q) = \frac{1}{2}TK(K+1)\left(1 - \frac{2}{K(K+1)}\text{tr}(\hat{\Omega}\hat{\Omega}_0^{-1})\right),$$

where $\hat{\Omega}$ is the residual covariance matrix of the $\frac{1}{2}K(K+1)$-dimensional regression model (5.4) with $q > 0$ and $\hat{\Omega}_0$ is the corresponding matrix for the case $q = 0$. The statistic is similar to the one described by Doornik and Hendry (1997, Sec. 10.9.2.4) and may be used with critical values from a $\chi^2(qK^2(K+1)^2/4)$-distribution. Alternatively, an F version may be considered which may be advantageous in small samples.

5.4 Stability Analysis

A wide range of procedures for checking the stability or time invariance of a given model exists (e.g., Doornik and Hendry (1997), Lütkepohl (2004, 2005, Chapter 17)). They may be used to detect potential structural breaks during the sample period. For example, possible breaks in monetary models for the U.S. have been discussed extensively in the literature (e.g., Bernanke and Mihov (1998), Christiano, Eichenbaum and Evans (1999), Cogley and Sargent (2001), Lubik and Schorfheide (2004), Primiceri (2005), Sims and Zha (2006), Lanne and Lütkepohl (2006)). Therefore it is important to have statistical instruments for investigating possible changes in the structure of VARs and VECMs. Here I will just discuss Chow tests which are standard tools for stability analysis for time series models.

Different types of Chow tests exist. They check the null hypothesis of time invariant parameters throughout the sample period against the possibility of a change in the parameter values in some period T_B, say. In one type of test the model under consideration is estimated from the full sample of T observations and from the first T_1 and the last T_2 observations, where $T_1 < T_B$ and $T_2 \leq T - T_B$. The test is based on the LR principle under Gaussian assumptions. In other words, the likelihood maximum from the constant parameter model is compared to the one with different parameter values before and after period T_B, leaving out the observations between T_1 and $T - T_2 + 1$. Denoting the conditional log-density of the t-th observation vector by l_t, i.e., $l_t = \log f(y_t|y_{t-1}, \ldots, y_1)$, a version of the Chow test statistic can be written as

$$\lambda_{Chow} = 2\left[\sup\left(\sum_{t=1}^{T_1} l_t\right) + \sup\left(\sum_{t=T-T_2+1}^{T} l_t\right) - \sup\left(\sum_{t=1}^{T_1} l_t + \sum_{t=T-T_2+1}^{T} l_t\right)\right]. \tag{5.5}$$

If the model is time invariant, the statistic has an asymptotic χ^2-distribution. The degrees of freedom are given by the number of restrictions imposed by assuming a constant coefficient model for the full sample period, that is, it is the difference between the sum of the number of free coefficients estimated in the first and last subperiods and the number of free coefficients in the full sample model.

For a K-dimensional VECM, $\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t$, with cointegrating rank r, counting all parameters in the model apart from those in Σ_u gives $2Kr + (p-1)K^2$. The degrees of freedom for the test based on λ_{Chow} are, however, $2Kr - r^2 + (p-1)K^2$, where r^2 is subtracted because normalizing $\beta' = [I_r : \beta'_{(K-r)}]$ shows that r^2

of its parameters are fixed throughout the sample (see Hansen (2003)).

From the point of view of asymptotic theory there is no need to leave out any observations between the two subsamples. So $T_1 = T_B - 1$ and $T_2 = T - T_B$ is a possible choice. In practice, if the parameter change has not occurred instantaneously at the beginning of period T_B, but is spread out over a few periods or its exact timing is unknown, leaving out some observations may improve the small sample power of the test.

Various generalizations of these tests are possible. For example, one could test for more than one break or one could check constancy of a subset of parameters keeping the remaining ones fixed. Moreover, there may be deterministic terms in the cointegration relations or the number of cointegration relations may change in different subperiods. These generalizations are also treated by Hansen (2003). A Chow forecast test version for multivariate time series models was proposed by Doornik and Hendry (1997). It tests the null hypothesis that the forecasts from a model fitted to the first T_B observations are in line with the actually observed data. Doornik and Hendry (1997) also proposed F versions of the tests to alleviate small sample distortions. Candelon and Lütkepohl (2001) pointed out that especially for multivariate time series models the asymptotic χ^2-distribution may be a poor guide for small sample inference. Even adjustments based on F approximations can lead to distorted test sizes. Therefore they proposed to use bootstrap versions of the Chow tests in order to improve their small sample properties.

Chow tests are sometimes performed repeatedly for a range of potential break points T_B. If the test decision is based on the maximum of the test statistics, the test is effectively based on the test statistic $\sup_{T_B \in \mathsf{T}} \lambda_{Chow}$, where $\mathsf{T} \subset \{1, \ldots, T\}$ is the set of periods for which the test statistic is determined. The asymptotic distribution of the sup test statistic is not χ^2 but of a different type (see Andrews (1993), Andrews and Ploberger (1994) and Hansen (1997)).

Various versions of these tests may be useful for checking parameter constancy. If the short-term dynamics are expected to be stable and a test of parameter change in the long-run part only is desired, one may first concentrate out the short-term parameters based on the full sample. Then one may focus on recursive estimation of α and β. Hansen and Johansen (1999) derive tests which may be used to test stability of the cointegration space separately.

Once an adequate model for the reduced form has been found, this can be used as the

basis for forecasting and structural analysis. These issues will be considered next.

6 Forecasting

Although VARs and VECMs are both well suited for forecasting, I focus on the levels VAR form in the following for convenience. The discussion is equally valid for VECMs which have been converted to VAR form. The somewhat unrealistic situation of a known DGP will be considered first to separate model inherent uncertainty from issues related to uncertainty due to model specification and parameter estimation.

6.1 Known Processes

Because the future values of deterministic terms are known with certainty by their very nature, I will first focus on the stochastic part. Let x_t be generated by a VAR(p) process as in (2.2), where u_t is an iid white noise process. Under these assumptions the conditional expectation given x_t, $t \leq T$,

$$x_{T+h|T} = E(x_{T+h}|x_T, x_{T-1}, \dots) = A_1 x_{T+h-1|T} + \cdots + A_p x_{T+h-p|T}, \tag{6.1}$$

is the optimal (minimum MSE) h-steps ahead forecast in period T. Here $x_{T+j|T} = x_{T+j}$ for $j \leq 0$. The forecasts can easily be computed recursively for $h = 1, 2, \dots$. The associated forecast error is

$$x_{T+h} - x_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}, \tag{6.2}$$

where the Φ_i weighting matrices may be obtained recursively as

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \dots, \tag{6.3}$$

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$ (e.g., Lütkepohl (2005, Chapter 2)). For $h = 1$, u_t is seen to be the forecast error in period $t - 1$. The forecast errors have mean zero and, hence, the forecasts are unbiased. The forecast error covariance or MSE matrix is

$$\Sigma_x(h) = E[(x_{T+h} - x_{T+h|T})(x_{T+h} - x_{T+h|T})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'. \tag{6.4}$$

If the u_t's are just uncorrelated and not independent, the forecasts obtained recursively as in (6.1) are still best *linear* forecasts but do not necessarily minimize the MSE in a larger class which includes nonlinear forecasts as well.

These results are valid for both $I(0)$ and $I(1)$ processes. Yet there are important differences in the forecasts for these alternative types of variables. For $I(0)$ processes the forecast MSEs are bounded as the horizon h goes to infinity whereas for $I(1)$ processes the forecast uncertainty and, hence, the MSE increases without bounds for increasing forecast horizon.

Deterministic components can be added easily if $y_t = \mu_t + x_t$ with nonzero μ_t. In this case the h-step forecast of y_t at origin T is $y_{T+h|T} = \mu_{T+h} + x_{T+h|T}$. Obviously, the forecast errors are identical to those of the x_t process. In other words,

$$y_{T+h} - y_{T+h|T} \sim (0, \Sigma_x(h)).$$

If the process y_t is Gaussian, that is, $u_t \sim$ iid $\mathcal{N}(0, \Sigma_u)$, the forecast errors are also multivariate normal and forecast intervals can be set up in the usual way. If y_t is non-Gaussian or if the distribution is unknown, other methods for setting up forecast intervals are called for. Findley (1986), Masarotto (1990), Grigoletto (1998), Kabaila (1993), Kim (1999) and Pascual, Romo and Ruiz (2004) considered bootstrap methods for forecasting nonnormal processes. For nonnormally distributed variables, forecast intervals may not represent the best way to report the forecast uncertainty, however. A survey of related issues is given by Tay and Wallis (2002).

6.2 Estimated Processes

If unknown parameters are replaced by estimators in the previous formulas, this has implications for the forecast precision. Signifying forecasts based on estimated parameters with a hat, the forecast error becomes

$$\begin{aligned} y_{T+h} - \hat{y}_{T+h|T} &= (y_{T+h} - y_{T+h|T}) + (y_{T+h|T} - \hat{y}_{T+h|T}) \\ &= \sum_{j=0}^{h-1} \Phi_j u_{T+h-j} + (y_{T+h|T} - \hat{y}_{T+h|T}). \end{aligned} \tag{6.5}$$

The first term on the right-hand side involves residuals u_t with $t > T$ only and the second term involves just $y_T, y_{T-1}, \ldots$, if only variables up to time T have been used for estimation.

Consequently, the two terms are independent and

$$\Sigma_{\hat{y}}(h) = E[(y_{T+h} - \hat{y}_{T+h|T})(y_{T+h} - \hat{y}_{T+h|T})'] = \Sigma_x(h) + o(1) \tag{6.6}$$

is obtained, where $o(1)$ denotes a term which approaches zero as the sample size tends to infinity. The latter result follows because the difference $y_{T+h|T} - \hat{y}_{T+h|T}$ is small in probability under standard assumptions. Thus, asymptotically the forecast uncertainty implied by estimation uncertainty may be ignored. In finite samples the precision of the forecasts will depend on the precision of the estimators, however. Hence, precise forecasts require precise estimators. For the stationary case, possible correction factors for MSEs and forecast intervals are given, e.g., in Yamamoto (1980), Baillie (1981) and Lütkepohl (2005, Chapter 3).

A number of extensions of these results are worth mentioning. For example, constructing separate models for different forecast horizons may be considered if estimation uncertainty is an issue (e.g., Bhansali (2002)). Moreover, Lewis and Reinsel (1985) and Lütkepohl (1985) considered the case of a stationary infinite order VAR DGP which is approximated by a finite order VAR, thereby extending earlier univariate results by Bhansali (1978). Moreover, Reinsel and Lewis (1987), Basu and Sen Roy (1987), Engle and Yoo (1987), Sampson (1991), Reinsel and Ahn (1992) and Clements and Hendry (1998, 2001) presented results for processes with unit roots. Stock (1996) and Kemp (1999) assumed that the forecast horizon h and the sample size T both go to infinity simultaneously. Clements and Hendry (1998, 2001) also considered various other sources of possible forecast errors.

7 Causality Analysis

7.1 Intuition and Theory

VAR models also open up the possibility for analyzing the relation between the variables involved. Analyzing the causal relations is of particular interest. Granger (1969) presented a definition of causality in the time series context which has become quite popular in applied work. He called a variable y_{1t} causal for a variable y_{2t} if the information in y_{1t} is helpful for improving the forecasts of y_{2t}. Because this is a special concept of causality, it is often

referred to as *Granger-causality*. In a bivariate VAR(p) setting,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + u_t, \tag{7.1}$$

it turns out that y_{1t} is not Granger-causal for a variable y_{2t} if and only if

$$\alpha_{21,i} = 0, \qquad i = 1, 2, \ldots, p, \tag{7.2}$$

that is, if it does not appear in the y_{2t} equation of the model. Although this result holds for stationary and integrated processes alike, for $I(1)$ variables it may still be desirable to investigate Granger-causal relations within a VECM. This case is discussed by Mosconi and Giannini (1992). An interesting implication of cointegration between two $I(1)$ variables is that there must be Granger-causality in at least one direction (from y_{1t} to y_{2t} or from y_{2t} to y_{1t} or both). For other characterizations of Granger-causality see also Gourieroux and Monfort (1997).

Because economic systems of interest usually consist of more than two variables, it is desirable to extend the concept of Granger-causality to higher dimensional processes. Such extensions have been discussed by Lütkepohl (1993) and Dufour and Renault (1998). In one approach the vector of all variables is partitioned into two subvectors so that $y_t = (y_{1t}', y_{2t}')'$ and Granger-causality from y_{1t} to y_{2t} is considered in this context. It turns out that in the VAR(p) framework Granger-noncausality is still characterized by zero restrictions on the parameter matrices similar to the bivariate case (Lütkepohl (2005, Section 2.3.1)).

This type of generalization is not satisfactory if a causal relation between two variables within a higher dimensional system is of interest. In that case more complex restrictions may have to be considered (Dufour and Renault (1998), Lütkepohl (2005, Section 2.3.1)). Causality has been discussed extensively in the econometrics literature and is an area of ongoing research. It is obviously also of interest in other than VAR models. In fact, causality is of interest not only in the context of time series models but also in describing economic relations more generally (e.g., Granger (1982), Angrist, Imbens and Rubin (1996), Heckman (2000), Pearl (2000), Hoover (2001)).

7.2 Testing for Granger-Causality

Because Granger-noncausality is characterized by zero restrictions on the levels VAR representation of the DGP, standard Wald χ^2- or F-tests can be applied for causality analysis. Unfortunately, these tests may have nonstandard asymptotic properties if $y_t \sim I(1)$ (Toda and Phillips (1993)). There is a simple modification that fixes the problem in this case, however. Recall from Section 3.2 that standard inference is possible whenever the elements in at least one of the complete coefficient matrices A_i are not restricted at all under the null hypothesis. Thus, adding an extra redundant lag in estimating the parameters of the process ensures standard asymptotics for the Wald tests if elements from all matrices $A_1, \ldots, A_p$ are involved in the restrictions as, for instance, in the noncausality restrictions in (7.2) (see Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996)). Clearly, a VAR($p+1$) is an appropriate model if the same is true for a VAR(p). The test may then be performed on the $A_1, \ldots, A_p$ only. As a consequence of results due to Park and Phillips (1989) and Sims et al. (1990) the procedure remains valid if deterministic terms are included in the VAR model. Although this device leads to a known limiting distribution of the test statistic, it is not a fully efficient procedure because of the redundant lag.

Extensions to testing for Granger-causality in infinite order VAR processes were considered by Lütkepohl and Poskitt (1996b) and Saikkonen and Lütkepohl (1996). Moreover, tests for other types of causality restrictions were considered by Lütkepohl and Burda (1997) and Dufour, Pelletier and Renault (2006).

8 Structural VARs and Impulse Response Analysis

Traditional econometric simultaneous equations models are sometimes used to predict the responses of the endogenous variables to changes in exogenous variables or to derive optimal policy responses to changes in the economic conditions. In VAR models there are typically no exogenous variables. In these models the effects of shocks are usually studied and used to link the VAR models to economic models. For example, they may be used to investigate the effects of monetary policy shocks, that is, of unexpected changes in the variables which were not anticipated by the economic agents. The relevant tool is known as impulse response analysis and will be considered in the following.

For a meaningful analysis it is important to isolate the actual shocks of interest. This requires the imposition of some structure on the reduced forms which we have discussed predominantly so far. The relevant structural VARs and VECMs will be discussed in Sections 8.1 and 8.2, respectively. Estimation of structural parameters and impulse responses will be discussed in Section 8.3. Based on structural innovations, forecast error variance decompositions are often computed and used to study the structure of economic systems. They are briefly presented in Section 8.4.

8.1 Levels VARs

Impulse response analysis is a standard tool for investigating the relations between the variables in a VAR model. If the VAR(p) process y_t is $I(0)$, it has a Wold moving average (MA) representation of the form

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \tag{8.1}$$

where deterministic terms are ignored to simplify the exposition. In this representation, $\Phi_0 = I_K$ and the Φ_j's ($j = 1, 2, \dots$) are $(K \times K)$ coefficient matrices which can be computed using the formulas in (6.3). The marginal response of $y_{n,t+j}$ to a unit change in y_{mt}, holding constant all past values of y_t, are given by the (n, m)th elements of the matrices Φ_j, viewed as a function of j. More precisely, the elements of Φ_j represent responses to u_t innovations, that is, to forecast errors. Therefore these quantities are sometimes called *forecast error impulse responses* (Lütkepohl (2005, Section 2.3.2)). Since $\Phi_j \to 0$ as $j \to \infty$ for stationary processes, the effect of an impulse vanishes over time. In other words, it is *transitory*.

Forecast error impulse responses may not reflect the actual reactions of a given system properly because, if the components of u_t are contemporaneously correlated and, hence, Σ_u is not diagonal, the shocks are not likely to occur in isolation in practice. Therefore orthogonalized shocks are often considered in impulse response analysis. Any nonsingular matrix P with the property that $PP' = \Sigma_u$ can be used to define orthogonalized shocks as $\varepsilon_t = P^{-1}u_t$. Clearly, these shocks have the property that $\varepsilon_t \sim (0, I_K)$ and they are contemporaneously uncorrelated. The responses to such shocks are given by the coefficients

of the MA representation

$$y_t = \sum_{j=1}^{\infty} \Phi_j P P^{-1} u_{t-j} = \sum_{j=1}^{\infty} \Psi_j \varepsilon_{t-j}. \tag{8.2}$$

The problem here is that the matrix P is not unique and, hence, many different impulse responses $\Psi_j = \Phi_j P$ exist. To identify those impulses which are interesting from an economic point of view is the objective of structural VAR (SVAR) analysis. It uses subject matter theory to impose restrictions on P which result in unique impulse responses.

Historically, a popular choice of P has been a lower triangular matrix obtained by a Choleski decomposition of Σ_u. The recent SVAR literature has considered many other possibilities as well. A number of them can be placed within the so-called AB-model of Giannini (1992) and Amisano and Giannini (1997) which may be represented as

$$\mathsf{A} y_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \mathsf{B}\varepsilon_t, \tag{8.3}$$

where $\varepsilon_t \sim (0, I_K)$, as before. In this setup instantaneous relations between the components of y_t may be modelled via A as in (2.12) and relations between the residuals may be taken into account via B (see also Breitung, Brüggemann and Lütkepohl (2004) and Lütkepohl (2005, Chapter 9)). The error term in (8.3) is related to the reduced form error term by

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

and, hence, $P = \mathsf{A}^{-1}\mathsf{B}$ in (8.2).

Different types of identifying restrictions for the matrices A and B have been considered. Examples can be found, for instance, in Sims (1986), Bernanke (1986), Blanchard and Quah (1989), Galí (1999) and Pagan (1995).

8.2 Structural VECMs

Forecast error impulse responses can also be computed for $I(1)$ processes (Lütkepohl and Reimers (1992), Lütkepohl (2005, Chapter 6)). The shocks may have *permanent* effects, however. Although the AB-model may also be used for $I(1)$ systems to specify identifying restrictions for impulse responses for levels VAR processes, it is often of interest to distinguish between shocks with permanent and transitory effects. This can be done more easily in the VECM framework. To simplify matters I focus on a B-model setup. That is, I assume that

$u_t = \mathsf{B}\varepsilon_t$. Hence, $\Sigma_u = \mathsf{B}\mathsf{B}'$. Due to the symmetry of the covariance matrix, $\Sigma_u = \mathsf{B}\mathsf{B}'$ represents only $\frac{1}{2}K(K+1)$ independent equations. For a unique specification of the K^2 elements of B, at least $\frac{1}{2}K(K-1)$ further restrictions are needed and, as will be seen now, the cointegration structure may help in setting them up.

Granger's representation theorem (see Johansen (1995, Theorem 4.2)) states that y_t has the representation

$$y_t = \Xi \sum_{i=1}^{t} u_i + \sum_{j=0}^{\infty} \Xi_j^* u_{t-j} + y_0^*, \quad t = 1, 2, \ldots, \tag{8.4}$$

where the term y_0^* contains the initial values. The term $\sum_{j=0}^{\infty} \Xi_j^* u_{t-j}$ is an $I(0)$ component, while $z_t = \sum_{i=1}^{t} u_i = z_{t-1} + u_t$, $t = 1, 2, \ldots$, is a K-dimensional random walk and is thus $I(1)$. Hence, $\Xi \sum_{i=1}^{t} u_i$ represents a stochastic trend component. The matrix Ξ can be shown to be of the form

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

For a VECM with cointegrating rank r it has rank $K - r$. Thus, the representation in (8.4) decomposes y_t in $K - r$ common trends and a stationary cyclical part.

Substituting $\mathsf{B}\varepsilon_i$ for u_i in (8.4) shows that the long-run effects of the structural innovations are given by $\Xi\mathsf{B}$. As this matrix also has rank $K - r$, it can have at most r zero columns. In other words, at most r of the structural innovations can have transitory effects only and at least $K - r$ of them must have permanent effects at least on some of the variables. Thereby cointegration analysis can help in suggesting how many transitory shocks there can be at most. In this framework, linear restrictions for $\Xi\mathsf{B}$ and B are typically specified to identify the structural shocks (see, e.g., King, Plosser, Stock and Watson (1991), Gonzalo and Ng (2001)). The transitory shocks may be identified, for example, by placing zero restrictions on B directly and thereby specifying that certain shocks have no instantaneous impact on some of the variables. For further examples see Breitung et al. (2004) and more discussion of partitioning the shocks in permanent and transitory ones is given in Gonzalo and Ng (2001), Fisher and Huh (1999) and others.

Identifying the impulse responses within a given model is but one problem related to a proper interpretation. Other problems are due to omitted variables, filtering and adjusting series prior to using them for a VAR analysis. Moreover, aggregated or transformed data

can lead to major changes in the dynamic structure of the variables and hence, the impulse responses. These issues should be taken into account in an impulse response analysis.

8.3 Estimating Impulse Responses

Forecast error impulse responses can be estimated straightforwardly by substituting estimated reduced form parameters in the formulas for computing them. Estimation of the structural impulse responses is also straightforward if estimates of the structural parameters A and B are available. To estimate these parameters, ML or quasi ML methods under normality assumptions are often used. If no restrictions are imposed on the VAR model or VECM (apart from the rank restriction for the error correction term) the parameters can be concentrated out of the likelihood function by replacing them by their LS or RR estimators. The concentrated log-likelihood function in terms of A and B then becomes

$$l(\mathsf{A},\mathsf{B}) = \text{constant} + \frac{T}{2}\log\det(\mathsf{A})^2 - \frac{T}{2}\log\det(\mathsf{B})^2 - \frac{T}{2}\mathrm{tr}(\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}\hat{\Sigma}_u), \tag{8.5}$$

where $\hat{\Sigma}_u = T^{-1}(Y - \hat{A}\mathbf{Y})(Y - \hat{A}\mathbf{Y})'$ is the usual estimator of Σ_u (cf. Breitung et al. (2004)). Numerical methods can be used to optimize this function with respect to the free parameters in A and B. Under standard assumptions the resulting estimators have the usual asymptotic properties of ML estimators (see, e.g., Lütkepohl (2005, Chapter 9) for details).

Denoting the vector of all structural form parameters by $\boldsymbol{\alpha}$ and its estimator by $\hat{\boldsymbol{\alpha}}$, a vector of impulse response coefficients ψ is a (nonlinear) function of $\boldsymbol{\alpha}$, $\psi = \psi(\boldsymbol{\alpha})$, which can be estimated as $\hat{\psi} = \psi(\hat{\boldsymbol{\alpha}})$. If $\hat{\boldsymbol{\alpha}}$ is asymptotically normal, the same will hold for $\hat{\psi} = \psi(\hat{\boldsymbol{\alpha}})$ by appealing to the delta method. More precisely, if

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

then

$$\sqrt{T}(\hat{\psi} - \psi) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\psi}}), \tag{8.6}$$

where

$$\Sigma_{\hat{\psi}} = \frac{\partial\psi}{\partial\boldsymbol{\alpha}'}\Sigma_{\hat{\boldsymbol{\alpha}}}\frac{\partial\psi'}{\partial\boldsymbol{\alpha}},$$

provided the matrix of partial derivatives $\partial\psi/\partial\boldsymbol{\alpha}'$ is such that none of the variances is zero and, in particular, $\partial\psi/\partial\boldsymbol{\alpha}' \neq 0$. However, inference problems may arise if $\partial\psi/\partial\boldsymbol{\alpha}'$ does

not have full row rank. In that case, $\Sigma_{\hat{\psi}}$ and, hence, the asymptotic distribution of $\hat{\psi}$ will be singular. This may in fact occur at specific points in the parameter space in the present situation because the function $\psi(\boldsymbol{\alpha})$ consists of sums of products of elements of $\boldsymbol{\alpha}$. Moreover, $\Sigma_{\hat{\boldsymbol{\alpha}}}$ will in general be singular if y_t is $I(1)$ which in turn may imply singularity of $\Sigma_{\hat{\psi}}$ even if $\partial\psi/\partial\boldsymbol{\alpha}'$ has full row rank. As a further complication, both problems may be present simultaneously. For further discussion see also Benkwitz, Lütkepohl and Neumann (2000). Thus, standard asymptotic confidence intervals around impulse responses based on the asymptotic normal distribution in (8.6) may be misleading.

Generally, even in parts of the parameter space where standard asymptotic theory works, it may not provide good approximations to the small sample distributions of impulse responses. Therefore, bootstrap methods are often used in applied work to construct confidence intervals for these quantities (e.g., Kilian (1998b), Benkwitz, Lütkepohl and Wolters (2001)). They are computer intensive but have the advantage that complicated analytical expressions of the asymptotic variances are not needed and they may improve inference procedures. On the other hand, it is important to note that they are also justified by asymptotic theory. In general the bootstrap does not overcome the problems due to a singularity in the asymptotic distribution. Thus, in these cases bootstrap confidence intervals may also be unreliable and may have a coverage which does not correspond to the nominal level (see Benkwitz et al. (2000) for further details). A solution to these problems may be possible by using subset VAR techniques to single out all zero coefficients and estimate only the remaining nonzero parameters.

Several authors have compared the relative merits of confidence intervals for impulse responses obtained by asymptotic theory on the one hand and bootstrap methods on the other hand using Monte Carlo simulation methods (e.g., Fachin and Bravetti (1996), Kilian (1998a, b), Kilian and Chang (2000)). The results are rather mixed and depend to some extent on the Monte Carlo design. Generally, in some cases simple asymptotic intervals outperformed standard bootstrap intervals in terms of coverage, i.e., their actual coverage probabilities were closer to the nominal ones chosen by the user. Bootstrap or simulation methods may improve substantially, however, once adjustments are made as, e.g., proposed by Kilian (1998b). Also a Bayesian Monte Carlo integration method proposed by Sims and Zha (1999) performed well in one of the simulation comparisons.

Generally, inference related to impulse responses has been discussed extensively in the recent literature. For example, Sims and Zha (1999) questioned the practice of reporting confidence intervals around individual impulse response coefficients and proposed likelihood-characterizing error bands as alternatives. Moreover, Koop (1992) considered confidence bands for impulse responses constructed with Bayesian methods. Generally, Bayesian methods are popular in structural VAR analysis. For example, Del Negro and Schorfheide (2004) used Baysesian methods in a structural analysis which combines DSGE models with VARs and, in particular, derived the structure from a simple DSGE model. Christiano, Eichenbaum and Evans (2005) estimated the parameters of a DSGE model by minimizing the distance of impulse responses from a DSGE model and a VAR. Other ideas for identifying impulse responses have been advanced by Uhlig (2005) who proposed to use inequality constraints and also used Bayesian methods, Lanne and Lütkepohl (2005, 2006) suggested to utilize information from the residual distribution and Lee, Pesaran and Pierse (1992) and Pesaran and Shin (1996) considered persistence profiles which measure the persistence of certain shocks without imposing structural identification restrictions.

8.4 Forecast Error Variance Decompositions

Using the structural innovations to express the h-step forecast error from (6.2) gives

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

From this expression the forecast error variance of the kth component can be shown to be

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

where $\psi_{nm,j}$ denotes the (n,m)th element of Ψ_j. The term $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)$ may be interpreted as the contribution of the jth innovation to the h-step forecast error variance of variable k. The relative contributions obtained as the ratios $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)/\sigma_k^2(h)$ are often reported for various variables and forecast horizons. Estimation of these quantities can easily be done by replacing unknown parameters by their estimators. Standard asymptotic and bootstrap evaluations of their sampling uncertainty is problematic, however, because some of these quantities may assume the boundary values of zero and one. Clearly,

interpretation of these quantities is informative only if they are based on meaningful and economically relevant structural innovations.

9 Conclusions and Extensions

In this chapter the specification and estimation of finite order VAR models have been reviewed. Furthermore the possible uses of these models for forecasting, causality and impulse response analysis have been discussed. Special attention has been paid to integrated and cointegrated variables for which vector error correction models have been considered throughout. Using these models in practice requires extensive computations because asymptotic inference is often not very precise in small samples and improved methods based on bootstraps, say, are often used. There are some software packages which contain programs for many of the procedures discussed in this chapter. Examples of specialized software are *PcGive* (Doornik and Hendry (1997)), *EViews* (EViews (2000)) and *JMulTi* (Krätzig (2004)).

There are different possible directions for generalizing the methods and results discussed in this chapter. For example, one may consider infinite order VAR processes and study the consequences of fitting finite order approximations to time series generated by such DGPs. For an exposition of this topic with further references see Lütkepohl (2005, Chapter 15). For a more parsimonious parameterization the class of vector autoregressive moving average processes may be worth considering. In these models the residuals of the VAR part are allowed to have a MA structure rather than being white noise. For the stationary case these models have been considered, for instance, by Hannan and Deistler (1988), Lütkepohl (2005, Part IV) and Lütkepohl and Poskitt (1996a). Extensions to cointegrated systems have been discussed by Lütkepohl and Claessen (1997), Bartel and Lütkepohl (1998), Poskitt (2003) and Lütkepohl (2005, Chapter 14). A recent survey with many more references was given by Lütkepohl (2006a). Furthermore, nonlinear components may be included in VAR models (e.g., Balke and Fomby (1997), Granger (2001), Escribano and Mira (2002), Saikkonen (2005)) or specific modelling of strong seasonal fluctuations may be desirable (e.g., Ghysels and Osborn (2001)). Also VAR models may be amended by ARCH residuals (e.g., Lütkepohl (2005, Chapter 16)) and higher order integration may be accounted for (e.g., Boswijk (2000), Johansen (1997, 2006)).

References

Ahn, S. K. (1988). Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization, *Biometrika* **75**: 590–593.

Ahn, S. K. and Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* **85**: 813–823.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov and F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **AC-19**: 716–723.

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

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics* **22**: 327–351.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–856.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**: 1383–1414.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables, *Journal of the American Statistical Association* **91**: 444–472.

Baillie, R. T. (1981). Prediction from the dynamic simultaneous equation model with vector autoregressive errors, *Econometrica* **49**: 1331–1337.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration, *International Economic Review* **38**: 627–645.

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-integration, Error-Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

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

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

Bauwens, L. and Lubrano, M. (1996). Identification restrictions and posterior densities in cointegrated Gaussian VAR systems, *Advances in Econometrics, Volume 11, Part B*, JAI Press, Greenwich, CT, pp. 3–28.

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

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

Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series, *Econometric Reviews* **19**: 1–48.

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

Bernanke, B. S. and Mihov, I. (1998). Measuring monetary policy, *Quarterly Journal of Economics* **113**: 869–902.

Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain, *Annals of Statistics* **6**: 224–231.

Bhansali, R. J. (2002). Multi-step forecasting, *in* M. P. Clements and D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 206–221.

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

Boswijk, H. P. (2000). Mixed normality and ancillarity in I(2) systems, *Econometric Theory* **16**: 878–904.

Boswijk, H. P. and Doornik, J. A. (2002). Identifying, estimating and testing restricted cointegrated systems: An overview, *Paper presented at the Henri Theil Memorial Conference*, Universiteit van Amsterdam.

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

Brüggemann, R. and Lütkepohl, H. (2005). Practical problems with reduced-rank ML estimators for cointegration parameters and a simple alternative, *Oxford Bulletin of Economics and Statistics* **67**: 673–690.

Brüggemann, R., Lütkepohl, H. and Saikkonen, P. (2006). Residual autocorrelation testing for vector error correction models, *Journal of Econometrics* **134**: 579–604.

Candelon, B. and Lütkepohl, H. (2001). On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models, *Economics Letters* **73**: 155–160.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.

Christiano, L. J., Eichenbaum, M. and Evans, C. (1999). Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Vol. 1A, Elsevier, Amsterdam, pp. 65–148.

Christiano, L. J., Eichenbaum, M. and Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**: 1–45.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

Clements, M. P. and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge, Ma.

Cogley, T. and Sargent, T. J. (2001). Evolving post-world war II U.S. inflation dynamics, *NBER Macroeconomics Annual* **16**: 331–373.

Del Negro, M. and Schorfheide, F. (2004). Priors from general equilibrium models for VARs, *International Economic Review* **45**: 643–673.

Doan, T., Litterman, R. B. and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews* **3**: 1–144.

Dolado, J. J. and Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems, *Econometric Reviews* **15**: 369–386.

Doornik, J. A. and Hansen, H. (1994). A practical test of multivariate normality, unpublished paper, Nuffield College.

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

Dufour, J.-M., Pelletier, D. and Renault, E. (2006). Short run and long run causality in time series: inference, *Journal of Econometrics* **132**: 337–362.

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

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective, *Econometric Reviews* **18**: 343–386.

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

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

Escribano, A. and Mira, S. (2002). Nonlinear error correction models, *Journal of Time Series Analysis* **23**: 509–522.

EViews (2000). *EViews 4.0 User's Guide*, Quantitative Micro Software, Irvine, CA.

Fachin, S. and Bravetti, L. (1996). Asymptotic normal and bootstrap inference in structural VAR analysis, *Journal of Forecasting* **15**: 329–341.

Findley, D. F. (1986). On bootstrap estimates of forecast mean square errors for autoregressive processes, *in* D. M. Allen (ed.), *Computer Science and Statistics: The Interface*, North-Holland, Amsterdam, pp. 11–17.

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

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

Geweke, J. (1996). Bayesian reduced rank regression in econometrics, *Journal of Econometrics* **75**: 121–146.

Ghysels, E. and Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press, Cambridge.

Giannini, C. (1992). *Topics in Structural VAR Econometrics*, Springer, Heidelberg.

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

Gonzalo, J. and Pitarakis, J.-Y. (1999). Dimensionality effect in cointegration analysis, *in* R. Engle and H. White (eds), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 212–229.

Gourieroux, C. and Monfort, A. (1997). *Time Series and Dynamic Models*, Cambridge University Press, Cambridge.

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. (1982). Generating mechanisms, models, and causality, *in* W. Hildenbrand (ed.), *Advances in Econometrics*, Cambridge University Press, Cambridge, pp. 237–253.

Granger, C. W. J. (2001). Overview of nonlinear macroeconometric empirical models, *Macroeconomic Dynamics* **5**: 466–481.

Grigoletto, M. (1998). Bootstrap prediction intervals for autoregressions: Some alternatives, *International Journal of Forecasting* **14**: 447–456.

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

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

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society* **B41**: 190–195.

Hansen, B. E. (1997). Approximate asymptotic p values for structural-change tests, *Journal of Business & Economic Statistics* **15**: 60–67.

Hansen, H. and Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR-models, *Econometrics Journal* **2**: 306–333.

Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* **114**: 261–295.

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

Heckman, J. J. (2000). Causal parameters and policy analysis in econometrics: A twentieth century retrospective, *Quarterly Journal of Economics* **105**: 45–97.

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

Hoover, K. D. (2001). *Causality in Macroeconomics*, Cambridge University Press, Cambridge, MA.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic, *Journal of the American Statistical Association* **75**: 602–608.

Hosking, J. R. M. (1981a). Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society* **B43**: 261–262.

Hosking, J. R. M. (1981b). Lagrange-multiplier tests of multivariate time series models, *Journal of the Royal Statistical Society* **B43**: 219–230.

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

Ingram, B. F. and Whiteman, C. H. (1994). Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors, *Journal of Monetary Economics* **34**: 497–510.

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals, *International Statistical Review* **55**: 163–172.

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

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

Johansen, S. (1997). Likelihood analysis of the I(2) model, *Scandinavian Journal of Statistics* **24**: 433–462.

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica* **70**: 1929–1961.

Johansen, S. (2006). Statistical analysis of hypotheses on the cointegrating relations in the I(2) model, *Journal of Econometrics* **132**: 81–115.

Kabaila, P. (1993). On bootstrap predictive inference for autoregressive processes, *Journal of Time Series Analysis* **14**: 473–484.

Kemp, G. C. R. (1999). The behavior of forecast errors from a nearly integrated AR(1) model as both sample size and forecast horizon become large, *Econometric Theory* **15**: 238–256.

Kilian, L. (1998a). Confidence intervals for impulse responses under departures from normality, *Econometric Reviews* **17**: 1–29.

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

Kilian, L. and Chang, P.-L. (2000). How accurate are confidence intervals for impulse responses in large VAR models?, *Economics Letters* **69**: 299–307.

Kim, J. H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression, *International Journal of Forecasting* **15**: 393–403.

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

Kleibergen, F. and Paap, R. (2002). Priors, posteriors and bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics* **111**: 223–249.

Kleibergen, F. and van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* **10**: 514–551.

Koop, G. (1992). Aggregate shocks and macroeconomic fluctuations: A Bayesian approach, *Journal of Applied Econometrics* **7**: 395–411.

Koop, G., Strachan, R., van Dijk, H. and Villani, M. (2005). Bayesian approaches to cointegration, *in* K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, Palgrave Macmillan, Houndmills.

Krätzig, M. (2004). The software JMulTi, *in* H. Lütkepohl and M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 289–299.

Lanne, M. and Lütkepohl, H. (2005). Structural vector autoregressions with nonnormal residuals, *Working paper*, European University Institute, Florence.

Lanne, M. and Lütkepohl, H. (2006). Identifying monetary policy shocks via changes in volatility, *Working paper*, European University Institute, Florence.

Lee, K. C., Pesaran, M. H. and Pierse, R. G. (1992). Persistence of shocks and its sources in a multisectoral model of U(k) output growth, *Economic Journal* **102**: 342–356.

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

Li, H. and Maddala, G. S. (1996). Bootstrapping time series models, *Econometric Reviews* **15**: 115–158.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society* **B43**: 231–239.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience, *Journal of Business & Economic Statistics* **4**: 25–38.

Lomnicki, Z. A. (1961). Tests for departure from normality in the case of linear stochastic processes, *Metrika* **4**: 37–62.

Lubik, T. A. and Schorfheide, F. (2004). Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review* **94**: 190–217.

Lütkepohl, H. (1985). The joint asymptotic distribution of multistep prediction errors of estimated vector autoregressions, *Economics Letters* **17**: 103–106.

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

Lütkepohl, H. (2004). Vector autoregressive and vector error correction models, *in* H. Lütkepohl and M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 86–158.

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

Lütkepohl, H. (2006a). Forecasting with VARMA models, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting, Volume 1*, Elsevier, Amsterdam, pp. 287–325.

Lütkepohl, H. (2006b). Vector autoregressive models, *in* T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics, Volume 1, Econometric Theory*, Palgrave Macmillan, Houndmills, Basingstoke, pp. 477–510.

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

Lütkepohl, H. and Claessen, H. (1997). Analysis of cointegrated VARMA processes, *Journal of Econometrics* **80**: 223–239.

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

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

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

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

Lütkepohl, H. and Saikkonen, P. (1999). Order selection in testing for the cointegrating rank of a VAR process, *in* R. F. Engle and H. White (eds), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 168–199.

Lütkepohl, H., Saikkonen, P. and Trenkler, C. (2004). Testing for the cointegrating rank of a VAR process with level shift at unknown time, *Econometrica* **72**: 647–662.

Masarotto, G. (1990). Bootstrap prediction intervals for autoregressions, *International Journal of Forecasting* **6**: 229–239.

Mosconi, R. and Giannini, C. (1992). Non-causality in cointegrated systems: Representation, estimation and testing, *Oxford Bulletin of Economics and Statistics* **54**: 399–417.

Ng, S. and Perron, P. (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 and S. Sayer (eds), *Surveys in Econometrics*, Basil Blackwell, Oxford.

Park, J. Y. and Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1, *Econometric Theory* **4**: 468–497.

Park, J. Y. and Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory* **5**: 95–131.

Pascual, L., Romo, J. and Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes, *Journal of Time Series Analysis* **25**: 449–465.

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

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, MA.

Pesaran, M. H. and Shin, Y. (1996). Cointegration and speed of convergence to equilibrium, *Journal of Econometrics* **71**: 117–143.

Poskitt, D. S. (2003). On the specification of cointegrated autoregressive moving-average forecasting systems, *International Journal of Forecasting* **19**: 503–519.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**: 821–852.

Quenouille, M. H. (1957). *The Analysis of Multiple Time-Series*, Griffin, London.

Quinn, B. G. (1980). Order determination for a multivariate autoregression, *Journal of the Royal Statistical Society* **B42**: 182–185.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.

Reinsel, G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis* **13**: 353–375.

Reinsel, G. C. and Lewis, A. L. (1987). Prediction mean square error for non-stationary multivariate time series using estimated parameters, *Economics Letters* **24**: 57–61.

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. (2005). Stability results for nonlinear error correction models, *Journal of Econometrics* **127**: 69–81.

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

Saikkonen, P. and Lütkepohl, H. (2000a). Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics* **18**: 451–464.

Saikkonen, P. and Lütkepohl, H. (2000b). Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process, *Journal of Time Series Analysis* **21**: 435–456.

Saikkonen, P., Lütkepohl, H. and Trenkler, C. (2006). Break date estimation for VAR processes with level shift with an application to cointegration testing, *Econometric Theory* **22**: 15–68.

Sampson, M. (1991). The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models, *Journal of Applied Econometrics* **6**: 67–76.

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. (1986). Are forecasting models usable for policy analysis?, *Quarterly Review, Federal Reserve Bank of Minneapolis* **10**: 2–16.

Sims, C. A., Stock, J. H. and Watson, M. W. (1990). Inference in linear time series models with some unit roots, *Econometrica* **58**: 113–144.

Sims, C. A. and Zha, T. (1999). Error bands for impulse responses, *Econometrica* **67**: 1113–1155.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 54–81.

Stock, J. H. (1996). VAR, error correction and pretest forecasts at long horizons, *Oxford Bulletin of Economics and Statistics* **58**: 685–701.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model, *Journal of Econometrics* **123**: 307–325.

Tay, A. S. and Wallis, K. F. (2002). Density forecasting: A survey, *in* M. P. Clements and D. F. Hendry (eds), *Companion to Economic Forecasting*, Blackwell, Oxford, pp. 45–68.

Toda, H. Y. and Phillips, P. C. B. (1993). Vector autoregressions and causality, *Econometrica* **61**: 1367–1393.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics* **66**: 225–250.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.

Yamamoto, T. (1980). On the treatment of autocorrelated errors in the multiperiod prediction of dynamic simultaneous equation models, *International Economic Review* **21**: 735–748.

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.