European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2011/30
DEPARTMENT OF ECONOMICS

VECTOR AUTOREGRESSIVE MODELS

Helmut Luetkepohl

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Vector Autoregressive Models

HELMUT LUETKEPOHL

EUI Working Paper **ECO** 2011/30

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

October 13, 2011

Vector Autoregressive Models

Helmut Lütkepohl[1]
European University Institute, Florence

Contents

[1]Prepared for the *Handbook of Research Methods and Applications on Empirical Macroeconomics*. Helpful comments by Lutz Kilian are gratefully acknowledged.

1 Introduction

Multivariate simultaneous equations models were used extensively for macroeconometric analysis when Sims (1980) advocated vector autoregressive (VAR) models as alternatives. At that time longer and more frequently observed macroeconomic time series called for models which described the dynamic structure of the variables. VAR models lend themselves for this purpose. They typically treat all variables as a priori endogenous. Thereby they account for Sims' critique that the exogeneity assumptions for some of the variables in simultaneous equations models are ad hoc and often not backed by fully developed theories. Restrictions, including exogeneity of some of the variables, may be imposed on VAR models based on statistical procedures.

VAR models are natural tools for forecasting. Their setup is such that current values of a set of variables are partly explained by past values of the variables involved. They can also be used for economic analysis, however, because they describe the joint generation mechanism of the variables involved. Structural VAR analysis attempts to investigate structural economic hypotheses with the help of VAR models. Impulse response analysis, forecast error variance decompositions, historical decompositions and the analysis of forecast scenarios are the tools which have been proposed for disentangling the relations between the variables in a VAR model.

Traditionally VAR models are designed for stationary variables without time trends. Trending behavior can be captured by including deterministic polynomial terms. In the 1980s the discovery of the importance of stochastic trends in economic variables and the development of the concept of cointegration by Granger (1981), Engle and Granger (1987), Johansen (1995) and others have shown that stochastic trends can also be captured by VAR models. If there are trends in some of the variables it may be desirable to separate the long-run relations from the short-run dynamics of the generation process of a set of variables. Vector error correction models offer a convenient framework for separating long-run and short-run components of the data generation process (DGP). In the present chapter levels VAR models are considered where cointegration relations are not modelled explicitly although they may be present. Specific issues related to trending variables will be mentioned occasionally throughout the chapter. The advantage of levels VAR models over vector error correction models is that they can also be used when the cointegration structure is unknown. Cointegration analysis and error correction models are discussed specifically in the next chapter.

1.1 Structure of the Chapter

Typically a VAR analysis proceeds by first specifying and estimating a reduced form model for the DGP and then checking its adequacy. Model deficiencies detected at the latter stage are resolved by modifying the model. If the reduced form model passes the checking stage, it may be used for forecasting, Granger-causality or structural analysis. The main steps of this modelling approach are depicted in Figure 1. The basic VAR model will be introduced in Section 2. Estimation and model specification issues are discussed in Sections 3 and 4, respectively, and model checking is considered in Section 5. Sections 6, 7 and 8 address forecasting, Granger-causality analysis and structural modelling including impulse response analysis, forecast error variance decomposition, historical decomposition of time series and analysis of forecast scenarios. Section 9 concludes and discusses extensions.

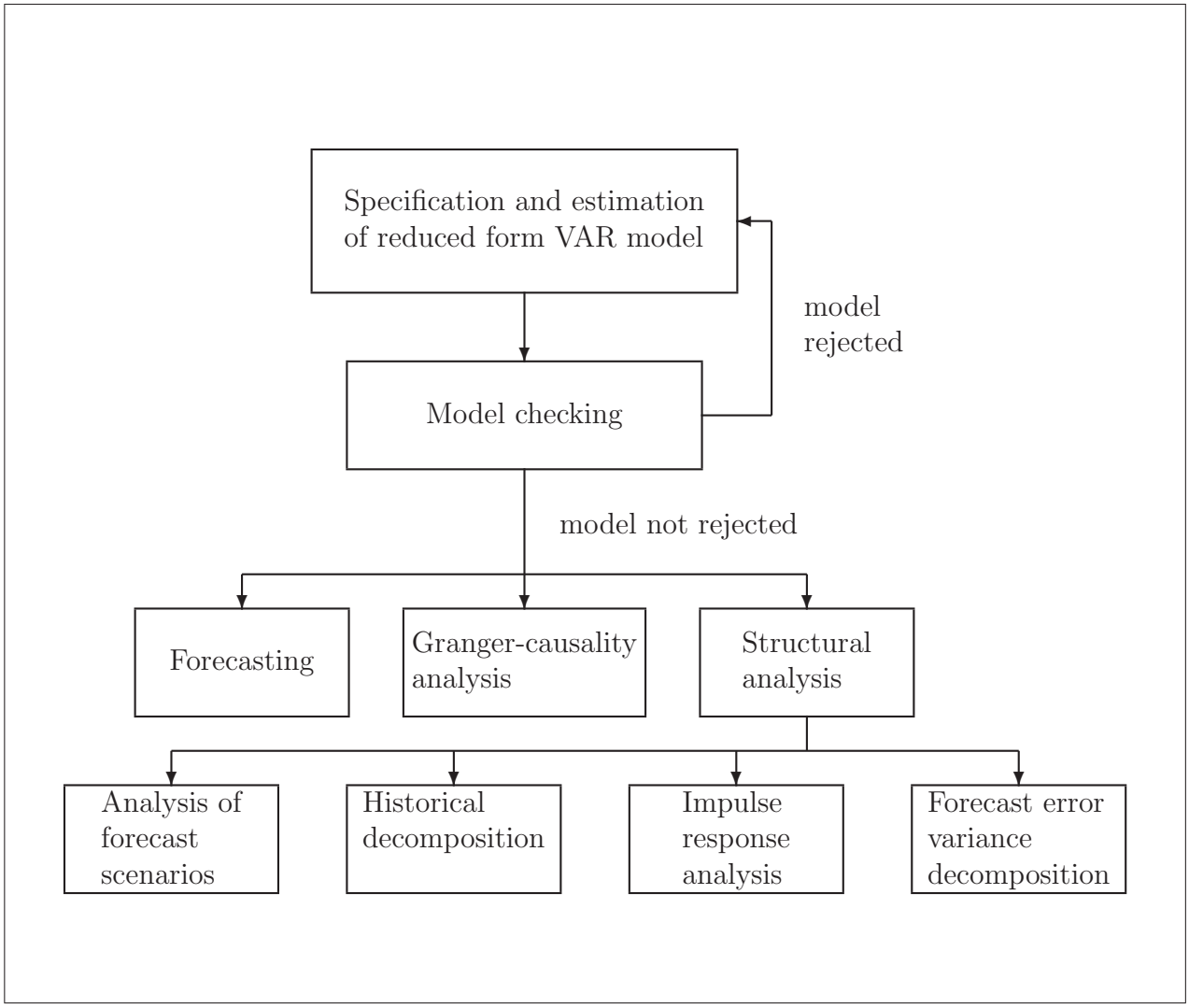

Figure 1: VAR analysis.

A number of textbooks and review articles deal with VAR models. Examples of books are Hamilton (1994), Johansen (1995), Hatanaka (1996), Lütkepohl and Krätzig (2004) and in particular Lütkepohl (2005). More formal and more detailed treatments of some of the issues discussed in the present chapter can be found in these references. The present chapter draws heavily on Lütkepohl and Krätzig (2004), Lütkepohl (2005) and earlier survey articles by Lütkepohl (2006b, 2009).

1.2 Terminology, Notation and General Assumptions

Given the importance of stochastic trends it is useful to have a special terminology in dealing with them. 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. Defining the differencing operator Δ 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. A more formal definition of an integrated variable or process can be found in Johansen (1995). In this chapter all variables are assumed to be either $I(0)$ (i.e., they do not have a stochastic trend)

or $I(1)$ (if there are stochastic trends) if not explicitly stated otherwise. A K-dimensional vector of time series variables $y_t = (y_{1t}, \ldots, y_{Kt})'$ is called $I(d)$, in short, $y_t \sim I(d)$, if at least one of its components is $I(d)$. Using this terminology, it is possible that some components of y_t may be $I(0)$ individually if $y_t \sim I(1)$. A set of $I(d)$ variables is called *cointegrated* if a linear combination exists which is of lower integration order. In that case the variables have a common trend component.

The $I(d)$ terminology refers only to the stochastic properties of the variables. There can also be deterministic terms. For simplicity I assume that deterministic components 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. To further simplify matters it is occasionally assumed that there is no deterministic term so that $\mu_t = 0$. Other deterministic terms which are important in practice are seasonal dummies and other dummy variables. Including them in VAR models is a straightforward extension which is not considered explicitly in this chapter.

The following matrix notation is used. 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 matrices A $(n \times m)$ and B $(n \times k)$, $[A : B]$ or (A, B) denotes the $(n \times (m+k))$ matrix which has A as its first m columns and B as the last k columns. For an $(n \times m)$ matrix A of full column rank $(n > m)$, an orthogonal complement is denoted by $A_\perp$, that is, $A_\perp' A = 0$ and $[A : A_\perp]$ is a nonsingular square matrix. 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, $\otimes$ signifies the Kronecker product and I_n is an $(n \times n)$ identity matrix.

The sets of all integers, positive integers and complex numbers are denoted by $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{C}$, respectively. The lag operator L shifts the time index backward by one period, that is, for a time series variable or vector y_t, $Ly_t = y_{t-1}$. Using this notation, the previously defined differencing operator may be written as $\Delta = 1 - L$. For a number x, $|x|$ denotes the absolute value or modulus. A sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

The following conventions are used with respect to distributions and stochastic processes. 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 is denoted as $\xrightarrow{d}$ and plim stands for 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, $E(u_t) = 0$ and positive definite covariance matrix $\Sigma_u = E(u_t u_t')$.

The following abbreviations are used: DGP, VAR, SVAR and MA for data generation process, vector autoregression, structural vector autoregression and moving average, respectively; ML, OLS, GLS, LM, LR and MSE for maximum likelihood, ordinary least squares, generalized least squares, Lagrange multiplier, likelihood ratio and mean squared error, respectively. The natural logarithm is abbreviated as log.

2 VAR Processes

2.1 The Reduced Form

Suppose the investigator is interested in a set of K related time series variables collected in $y_t = (y_{1t}, \ldots, y_{Kt})'$. Given the importance of distinguishing between stochastic and deterministic components of the DGPs of economic variables, it is convenient to separate the two components by assuming that

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

where μ_t is the deterministic part and x_t is a purely stochastic process with zero mean. The deterministic term μ_t is at most a linear trend ($\mu_t = \mu_0 + \mu_1 t$) and may also be zero ($\mu_t = 0$) or just a constant ($\mu_t = \mu_0$) for simplicity. Deterministic trend terms have implausible implications in the context of forecasting. Hence, they are not recommendable in applied VAR analysis. The issue will be further discussed in Section 6.1. The purely stochastic part, x_t, may be $I(1)$ and, hence, may include stochastic trends and cointegration relations. It has mean zero and a VAR representation. The properties of the observable process y_t are determined by those of μ_t and x_t. In particular, the order of integration and the cointegration relations are determined by x_t.

Suppose the stochastic part x_t is 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 $E(u_t u_t') = \Sigma_u$, that is, $u_t \sim (0, \Sigma_u)$. Using the lag operator and defining the matrix polynomial in the lag operator $A(L)$ as $A(L) = I_K - A_1 L - \cdots - A_p L^p$, the process (2.2) can be equivalently written as

$$A(L)x_t = u_t. \tag{2.3}$$

The VAR process (2.2)/(2.3) is *stable* if

$$\det A(z) = \det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } z \in \mathbb{C}, |z| \leq 1. \tag{2.4}$$

In other words, x_t is stable if all roots of the determinantal polynomial are outside the complex unit circle. In that case x_t is $I(0)$. Under usual assumptions a stable process x_t has time invariant means, variances and covariance structure and is, hence, stationary. If, however, $\det A(z) = 0$ for $z = 1$ (i.e., the process has a *unit root*) and all other roots of the determinantal polynomial are outside the complex unit circle, then some or all of the variables are integrated, the process is, hence, nonstationary and the variables may be cointegrated. Recall that all variables are either $I(0)$ or $I(1)$ by default.

Also, recall that x_t is the (typically unobserved) stochastic part whereas y_t is the vector of observed variables. Pre-multiplying (2.1) by $A(L)$, that is, considering $A(L)y_t = A(L)\mu_t + u_t$, shows that y_t inherits the VAR(p) representation from x_t. In other words, if $\mu_t = \mu_0 + \mu_1 t$, $A(L)y_t = \nu_0 + \nu_1 t + u_t$ 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 $\nu_1 = (I_K - \sum_{j=1}^{p} A_j)\mu_1$. Since all variables appear in levels, this form is known as the *levels form of the VAR process*. Alternatively, some or all variables may appear in first differences if the variables are $I(1)$ and not cointegrated.

If the parameters ν_i, $i = 0, 1$, are unrestricted in (2.5), the variables may have quadratic trends if $y_t \sim I(1)$. Thus, the additive model setup (2.1) imposes restrictions on the deterministic parameters in (2.5). Generally the additive setup makes it necessary to think about the deterministic terms at the beginning of the analysis and allow for the appropriate polynomial order. Sometimes trend-adjustments are performed prior to a VAR analysis. The reason is that the stochastic part of the variables is often of main interest in econometric analysis because it is viewed as describing the behavioral relations. In that case there may be no deterministic term in the levels VAR form (2.5).

Using terminology from the simultaneous equations literature, the model (2.5) is in *reduced form* because all right-hand side variables are lagged or predetermined. The instantaneous relations between the variables are summarized in the residual covariance matrix. In economic analysis it is often desirable to model the contemporaneous relations between the variables directly. This may be done by setting up a structural form which is discussed next.

2.2 Structural Forms

In *structural form models* contemporaneous variables may appear as explanatory variables in some equations. For example,

$$\mathsf{A}y_t = \nu_0^* + \nu_1^* t + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + v_t, \tag{2.6}$$

is a structural form. Here the $(K \times K)$ matrix A reflects the instantaneous relations, $\nu_i^* = \mathsf{A}\nu_i$ $(i = 0, 1)$ 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 matrix A usually has ones on its main diagonal so that the set of equations in (2.6) can be written such that each of the variables appears on the left-hand side of one of the equations and may depend on contemporaneous values of some or all of the other variables. Moreover, A is typically chosen such that Σ_v is a diagonal matrix. Structural VAR models are discussed in more detail in Chapter 24 of this volume (Kilian (2011)). Therefore they are only sketched briefly here. Other expository treatments are Amisano and Giannini (1997), Watson (1994), Breitung, Brüggemann and Lütkepohl (2004) and Lütkepohl (2005).

Multiplying (2.6) by any nonsingular matrix results in a representation of the form (2.6). This shows that the parameters of the structural form (2.6) are not identified without further restrictions. Imposing restrictions on A and Σ_v to identify the structural form is a main focus of structural VAR (SVAR) analysis (see Chapter 24, Kilian (2011)). Often zero restrictions are placed on A directly. In other words, some variables are not allowed to have an instantaneous impact on some other variables. For example, A may be lower-triangular if there is a recursive relation between the variables.

Alternatively, in SVAR analyses researchers sometimes think of specific shocks hitting the system. A suitable structural model setup for that case is obtained by pre-multiplying (2.6) by $\mathsf{B} = \mathsf{A}^{-1}$ and considering

$$y_t = \nu_0 + \nu_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \mathsf{B}v_t. \tag{2.7}$$

This setup makes it easy to specify that a certain structural shock v_{it} does not have an instantaneous effect on one of the observed variables by restricting the corresponding element of $\mathsf{B} = \mathsf{A}^{-1}$ to be zero. In other words, zero restrictions are placed on $\mathsf{B} = \mathsf{A}^{-1}$.

Other popular identifying restrictions are placed on the accumulated long-run effects of shocks. For example, if some variables represent rates of change of some underlying quantity, one may postulate that a shock has no long-run effect on the level of a variable by enforcing that the accumulated changes in the variable induced by the shock add to zero. For instance, in a seminal article Blanchard and Quah (1989) consider a bivariate model consisting of output growth rates (y_{1t}) and an unemployment rate (y_{2t}). They assume that demand shocks have no long-run effects on output. In other words, the accumulated effects of a demand shock on the output growth rates are assumed to be zero. Such restrictions are effectively restrictions for A or/and B.

The SVAR models (2.6) and (2.7) are sometimes referred to as A- and B-models, respectively (see Lütkepohl (2005)). They can also be combined to an AB-model of the form

$$\mathsf{A}y_t = \nu_0^* + \nu_1^* t + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \mathsf{B}v_t, \tag{2.8}$$

which makes it easy to impose restrictions on the instantaneous effects of changes in observed variables and unobserved shocks. On the other hand, it involves many more parameters in A and B and, hence, requires more identifying restrictions. In the B- and AB-models, the residuals are usually assumed to be standardized to have identity covariance matrix, that is, $\Sigma_v = I_K$. In that case the reduced form covariance matrix is $\Sigma_u = \mathsf{BB}'$ for the B-model and $\Sigma_u = \mathsf{A}^{-1}\mathsf{BB}'\mathsf{A}^{-1\prime}$ for the AB-model.

As mentioned earlier, identifying the structural relations between the variables or identifying the structural shocks is a main concern of SVAR analysis. Other types of information and restrictions for identification than those mentioned previously have also been proposed. For instance, sign restrictions, using information from higher-frequency data or heteroskedasticity may be considered (see Chapter 24, Kilian (2011) for details).

Prior to a structural analysis, a reduced form model as a valid description of the DGP is usually constructed. The stages of reduced form VAR model construction are discussed in the following. Before model specification is considered, estimation of VAR models will be discussed because estimation is typically needed at the specification stage.

3 Estimation of VAR Models

Reduced form VAR models can be estimated with standard methods. Classical least squares and maximum likelihood (ML) methods are discussed in Section 3.1 and Bayesian estimation is considered in Section 3.2. Estimation of structural models is treated in Section 3.3.

3.1 Classical Estimation of Reduced Form VARs

Consider the levels VAR(p) model (2.5) written in the more compact form

$$y_t = [\nu_0, \nu_1, A_1, \ldots, A_p]Z_{t-1} + u_t, \tag{3.1}$$

where $Z_{t-1} = (1, t, y'_{t-1}, \ldots, y_{t-p})'$. The deterministic terms may be adjusted accordingly if there is just a constant in the model or no deterministic component at all. Given a sample of size T, $y_1, \ldots, y_T$, and p presample vectors, $y_{-p+1}, \ldots, y_0$, the parameters can be estimated efficiently by ordinary least squares (OLS) for each equation separately. The estimator is

easily seen to be

$$[\hat{\nu}_0, \hat{\nu}_1, \hat{A}_1, \ldots, \hat{A}_p] = \left(\sum_{t=1}^{T} y_t Z_{t-1}' \right) \left(\sum_{t=1}^{T} Z_t Z_{t-1}' \right)^{-1}. \tag{3.2}$$

This estimator is identical to the generalized least squares (GLS) estimator, if no restrictions are imposed on the parameters. For a normally distributed (Gaussian) process y_t, where $u_t \sim \mathcal{N}(0, \Sigma_u)$, this estimator is also identical to the ML estimator, conditional on the initial presample values. Thus, the estimator has the usual desirable asymptotic properties of standard estimators. It is asymptotically normally distributed with smallest possible asymptotic covariance matrix and the usual inference procedures are available if the process is stable. In other words, in this case t-statistics can be used for testing individual coefficients and for setting up confidence intervals. Moreover, F-tests can be used for testing statistical hypotheses for sets of parameters. Of course, in the present framework these procedures are only valid asymptotically and not in small samples.

If there are integrated variables so that $y_t \sim I(1)$, the process is not stable and the variables may be cointegrated. In that case the OLS/ML estimator can still be used and it is still asymptotically normal under general conditions (see Park and Phillips (1988, 1989), Sims, Stock and Watson (1990), Lütkepohl (2005, Chapter 7)). However, in that case the covariance matrix of the asymptotic distribution is singular because some estimated parameters or linear combinations of them converge with a faster rate than the usual $\sqrt{T}$ rate when the sample size goes to infinity. This result implies that t-, χ^2- and F-tests for inference regarding the VAR parameters may be invalid asymptotically (Toda and Phillips (1993)). Although these properties require caution in doing inference for integrated processes, there are many situations where standard inference still holds (see Toda and Yamamoto (1995), Dolado and Lütkepohl (1996), Inoue and Kilian (2002a)). In particular, asymptotic inference on impulse responses as discussed in Section 8.1 remains valid if the order of the VAR process is greater than one.

If restrictions are imposed on the parameters, OLS estimation may be inefficient. In that case GLS estimation may be beneficial. Let $\boldsymbol{\alpha} = \text{vec}[\nu_1, \nu_2, A_1, \ldots, A_p]$ and suppose that there are linear restrictions for the parameters such as zero restrictions which exclude some of the lagged variables from some of the equations. Linear restrictions can often be written in the form

$$\boldsymbol{\alpha} = R\boldsymbol{\gamma}, \tag{3.3}$$

where R is a suitable, known $((K^2p + 2K) \times M)$ restriction matrix with rank M which typically consists of zeros and ones and $\boldsymbol{\gamma}$ is the $(M \times 1)$ vector of unrestricted parameters. The GLS estimator for $\boldsymbol{\gamma}$ is then

$$\hat{\boldsymbol{\gamma}} = \left[R' \left(\sum_{t=1}^{T} Z_t Z_{t-1}' \otimes \Sigma_u^{-1} \right) R \right]^{-1} R' \text{vec} \left(\Sigma_u^{-1} \sum_{t=1}^{T} y_t Z_{t-1}' \right). \tag{3.4}$$

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 and usual methods for inference are valid asymptotically.

In practice, the white noise covariance matrix is usually unknown and has to 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 under general conditions. The corresponding feasible GLS estimator of $\boldsymbol{\alpha}$, $\hat{\hat{\boldsymbol{\alpha}}} = R\hat{\hat{\boldsymbol{\gamma}}}$, is also consistent and asymptotically normal and allows for standard asymptotic inference. For Gaussian white noise u_t, ML estimation may be used alternatively. Its asymptotic properties are the same as those of the GLS estimator under standard assumptions.

For $I(1)$ processes a specific analysis of the integration and cointegration properties of the left-hand and right-hand side variables of the individual equations is necessary to determine the asymptotic properties of the estimators and the associated inference procedures.

3.2 Bayesian Estimation of Reduced Form VARs

Standard Bayesian methods for estimating linear regression models can be applied for estimating the parameters of reduced form VAR models. They are not discussed here in detail because they are considered elsewhere in this volume. In the VAR literature specific priors have been used, however, which may be worth noting at this point. Assuming a normal distribution for the residuals and, hence, for the observed y_t together with a normal-Wishart prior distribution for the VAR coefficients results in a normal-Wishard posterior distribution. Such a setup is rather common in the SVAR literature (see Uhlig (2005, Appendix B)). The so-called Minnesota prior is a specific example of a prior which has been used quite often in practice (see Doan, Litterman and Sims (1984), Litterman (1986)). It shrinks the VAR towards a random walk for each of the variables. Extensions and alternatives were proposed by Kadiyala and Karlsson (1997), Villani (2005), Sims, Waggoner and Zha (2008), Giannone, Lenza and Primiceri (2010) and others. Other, recent proposals include shrinking towards some dynamic stochastic general equilibrium model (e.g., Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004)). A more detailed exposition of Bayesian methods in VAR analysis may be found in Canova (2007, Chapters 9 - 11).

3.3 Estimation of Structural VARs

Properly identified structural form VAR models are also usually estimated by least squares, ML or Bayesian methods. The specific estimation algorithm depends to some extent on the type of restrictions used for identification. For example, if a just-identified A-model is used with ones on the main diagonal and diagonal residual covariance matrix Σ_v, equationwise OLS can be used for estimation. For the B-model (2.7) without restrictions on $\nu_0, \nu_1, A_1, \ldots, A_p$, the latter parameters can be concentrated out of the likelihood function by replacing them with their OLS estimators, using $\Sigma_u = \mathsf{B}\mathsf{B}'$ and estimating B by maximizing the concentrated Gaussian log-likelihood

$$l(\mathsf{B}) = \text{constant} - \frac{T}{2}\log\det(\mathsf{B})^2 - \frac{T}{2}\text{tr}(\mathsf{B}'^{-1}\mathsf{B}^{-1}\hat{\Sigma}_u), \tag{3.5}$$

where $\hat{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ is the estimator of Σ_u based on the OLS residuals (cf. Breitung et al. (2004)). If the actual distribution of y_t (and, hence, of u_t) is not normal, the resulting estimators are quasi- or pseudo-ML estimators. They still allow for standard asymptotic inference under general conditions.

In the AB-model the concentrated log-likelihood function in terms of A and B is

$$l(\mathsf{A}, \mathsf{B}) = \text{constant} + \frac{T}{2}\log\det(\mathsf{A})^2 - \frac{T}{2}\log\det(\mathsf{B})^2 - \frac{T}{2}\text{tr}(\mathsf{A}'\mathsf{B}'^{-1}\mathsf{B}^{-1}\mathsf{A}\hat{\Sigma}_u). \tag{3.6}$$

Numerical methods can be used for optimizing the functions in (3.5) and (3.6) with respect to the free parameters in B or A and B. The resulting estimators have the usual asymptotic properties of ML estimators (see, e.g., Lütkepohl (2005, Chapter 9) for details). Hence, asymptotic inference proceeds in the usual way. Alternatively, one may use Bayesian estimation methods (see, e.g., Sims et al. (2008)). The estimates will be of importance in the structural VAR analysis discussed in Section 8 and Chapter 24 (Kilian (2011)).

4 Model Specification

Model specification in the present context involves selecting the VAR order and possibly imposing restrictions on the VAR parameters. Notably zero restrictions on the parameter matrices may be desirable because the number of parameters in a VAR model increases with the square of the VAR order. Lag order specification is considered next and some comments on setting zero restrictions on the parameters are provided at the end of this section.

The VAR order is typically chosen by sequential testing procedures or model selection criteria. Sequential testing proceeds by specifying a maximum reasonable lag order, say $p_{\max}$, and then testing the following sequence of null hypotheses: $H_0 : A_{p_{\max}} = 0$, $H_0 : A_{p_{\max}-1} = 0$, etc.. The procedure stops when the null hypothesis is rejected for the first time. The order is then chosen accordingly. For stationary processes the usual Wald or LR χ^2 tests for parameter restrictions can be used in this procedure. If there are $I(1)$ variables these tests are also asymptotically valid as long as the null hypothesis $H_0 : A_1 = 0$ is not tested. Unfortunately, the small sample distributions of the tests may be quite different from their asymptotic counterparts, in particular for systems with more than a couple of variables (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)).

Alternatively, model selection criteria can be used. Some of them have the general form

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

where $\hat{\Sigma}_m = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ is the OLS residual covariance matrix estimator for a reduced form VAR 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. Popular examples 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)),

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

with $c_T = 2\log\log T/T$, 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 all these criteria $\varphi(m) = mK^2$ is the number of VAR parameters in a model with order m. The VAR order is chosen such that the respective criterion is minimized over the possible orders $m = 0, \ldots, p_{\max}$. Among these three criteria, AIC always suggests the largest order, SC chooses the smallest order and HQ is in between (Lütkepohl (2005, Chapters 4 and 8)). Of course, the criteria may all suggest the same lag order. The HQ and SC criteria are both consistent, that is, under general conditions the order estimated with these criteria converges in probability or almost surely to the true VAR order p if $p_{\max}$ is at least as large as the true lag order. AIC tends to overestimate the order asymptotically with a small probability. These results hold for both $I(0)$ and $I(1)$ processes (Paulsen (1984)).

The lag order obtained with sequential testing or model selection criteria depends to some extent on the choice of $p_{\max}$. Choosing a small $p_{\max}$, an appropriate model may not be in the set of possibilities and choosing a large $p_{\max}$ may result in a large value which is spurious. At an early stage of the analysis, using a moderate value for $p_{\max}$ appears to be a sensible strategy. An inadequate choice should be detected at the model checking stage (see Section 5).

Once the model order is determined zero restrictions may be imposed on the VAR coefficient matrices to reduce the number of parameters. Standard testing procedures can be used for that purpose. The number of possible restrictions may be very large, however, and searching over all possibilities may result in excessive computations. Therefore a number of shortcuts have been proposed in the related literature under the name of subset model selection procedures (see Lütkepohl (2005, Section 5.2.8)).

If a model is selected by some testing or model selection procedure, that model is typically treated as representing the true DGP in the subsequent statistical analysis. Recent research is devoted to the problems and possible errors associated with such an approach (e.g., Leeb and Pötscher (2005)). This literature points out that the actual distribution which does not condition on the model selected by some statistical procedure may be quite different from the conditional one. Suppose, for example, that the VAR order is selected by the AIC, say, the order chosen by this criterion is $\hat{p}$. Then a typical approach in practice is to treat a VAR($\hat{p}$) model as the true DGP and perform all subsequent analysis under this assumption. Such a conditional analysis can be misleading even if the true order coincides with $\hat{p}$ because the properties of the estimators for the VAR coefficients are affected by the post-model selection step. Conditioning on $\hat{p}$ ignores that this quatity is also a random variable based on the same data as the estimators of the VAR parameters. Since no general procedures exist for correcting the error resulting from this simplification, there is little to recommend for improving applied work in this respect.

5 Model Checking

Procedures for checking whether the VAR model represents the DGP of the variables adequately range from formal tests of the underlying assumptions to informal procedures such as inspecting plots of residuals and autocorrelations. Since a reduced form is underlying every structural form, model checking usually focusses on reduced form models. If a specific reduced form model is not an adequate representation of the DGP, any structural form based on it cannot represent the DGP well. Formal tests for residual autocorrelation, nonnormality and conditional heteroskedasticity for reduced form VARs are briefly summarized in the following. For other procedures see, e.g., Lütkepohl (2004).

5.1 Tests for Residual Autocorrelation

Portmanteau and Breusch-Godfrey-LM tests are standard tools for checking residual autocorrelation in VAR models. The null hypothesis of the protmanteau test is that all residual autocovariances are zero, that is, $H_0 : E(u_t u'_{t-i}) = 0$ $(i = 1, 2, \dots)$. The alternative is that at least one autocovariance and, hence, one autocorrelation is nonzero. The test statistic is based on the residual autocovariances, $\hat{C}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}'_{t-j}$, where the $\hat{u}_t$'s are the mean-adjusted estimated residuals. The portmanteau statistic is given by

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

or the modified version

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

may be used. The two statistics have the same asymptotic properties. For an unrestricted stationary VAR(p) process their null distributions can be approximated by a $\chi^2(K^2(h-p))$ distribution if T and h approach infinity such that $h/T \to 0$. For VAR models with 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)). Brüggemann, Lütkepohl and Saikkonen (2006) show that this approximation is unsatisfactory for integrated and cointegrated processes. For such processes the degrees of freedom also depend on the cointegrating rank. Thus, portmanteau tests are not recommended for levels VAR processes with unknown cointegrating rank.

The choice of h is crucial for the small sample properties of the test. If h is chosen too small the χ^2 approximation to the null distribution may be very poor while a large h reduces the power of the test. Using a number of different h values is not uncommon in practice.

The portmanteau test should be applied primarily to test for autocorrelation of high order. For low order autocorrelation the Breusch-Godfrey LM test is more suitable. It may be viewed as a test for zero coefficient matrices in a VAR model for the residuals,

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

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

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

may be used for checking that u_t is white noise. The precise form of the statistic can be found, e.g., in Lütkepohl (2005, Section 4.4.4). 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)). As a consequence, the LM test is applicable for levels VAR processes with unknown cointegrating rank.

5.2 Other Popular Tests for Model Adequacy

Nonnormality tests are often used for model checking, although normality is not a necessary condition for the validity of many of the statistical procedures related to VAR models. However, nonnonormality of the residuals may indicate other model deficiencies such as nonlinearities or structural change. Multivariate normality tests are often applied to the residual vector of the VAR model and univariate versions are used to check normality of the errors of the individual equations. The standard tests check whether the third and fourth moments of the residuals are in line with a normal distribution, as proposed by Lomnicki (1961) and Jarque and Bera (1987) for univariate models. For details see Lütkepohl (2005, Section 4.5) and for small sample corrections see Kilian and Demiroglu (2000).

Conditional heteroskedasticity is often a concern for models based on data with monthly or higher frequency. Therefore suitable univariate and multivariate tests are available to check for such features in the residuals of VAR models. Again much of the analysis can be done even if there is conditional heteroskedasticity. Notice that the VAR model represents the conditional mean of the variables which is often of primary interest. Still, it may be useful to check for conditional heteroskedasticity to better understand the properties of the underlying data and to improve inference. Also, heteroskedastic residuals can indicate structural changes. If conditional heteroskedasticity is found in the residuals, modelling them by multivariate GARCH models or using heteroskedasticity robust inference procedures may be useful to avoid distortions in the estimators of the conditional mean parameters. For a proposal to robustify inference against conditional heteroskedasticity see Goncalves and Kilian (2004).

There are a number of tests for structural stability which check whether there are changes in the VAR parameters or the residual covariances throughout the sample period. Prominent examples are so-called Chow tests. They consider 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. One possible test version compares the likelihood maximum of the constant parameter model to the one with different parameter values before and after period T_B. If the model is time invariant, the resulting LR statistic has an asymptotic χ^2-distribution under standard assumptions. See Lütkepohl (2005, Section 4.6) for details and other tests for structural stability of VARs.

Stability tests are sometimes performed for a range of potential break points T_B. Using the maximum of the test statistics, that is, rejecting stability if one of the test statistics exceeds some critical value, the test is no longer asymptotically χ^2 but has a different asymptotic distribution (see Andrews (1993), Andrews and Ploberger (1994) and Hansen (1997)).

If a reduced form VAR model has passed the adequacy tests, it can be used for forecasting and structural analysis which are treated next.

6 Forecasting

Since reduced form VAR models represent the conditional mean of a stochastic process, they lend themselves for forecasting. For simplicity forecasting with known VAR processes will be discussed first and then extensions for estimated processes will be considered.

6.1 Forecasting Known VAR Processes

If y_t is generated by a VAR(p) process (2.5), the conditional expectation of y_{T+h} given y_t, $t \leq T$, is

$$y_{T+h|T} = E(y_{T+h}|y_T, y_{T-1}, \dots) = \nu_0 + \nu_1(T+h) + A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T}, \quad (6.1)$$

where $y_{T+j|T} = y_{T+j}$ for $j \leq 0$. If the white noise process u_t is iid, $y_{T+h|T}$ is the optimal, minimum mean squared error (MSE) h-step ahead forecast in period T. The forecasts can easily be computed recursively for $h = 1, 2, \dots$. The forecast error associated with an h-step forecast is

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}, \quad (6.2)$$

where the Φ_i matrices may be obtained recursively as

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

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$ (e.g., Lütkepohl (2005, Chapter 2)). In other words, the Φ_i are the coefficient matrices of the infinite order polynomial in the lag operator $A(L)^{-1} = \sum_{j=0}^{\infty} \Phi_j L^j$. Obviously, the reduced form VAR residual u_t is the forecast error for a 1-step forecast in period $t-1$. The forecasts are unbiased, that is, the errors have mean zero and the forecast error covariance or MSE matrix is

$$\Sigma_y(h) = E[(y_{T+h} - y_{T+h|T})(y_{T+h} - y_{T+h|T})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j', \quad (6.4)$$

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

In fact, the conditional expectation in (6.1) is obtained whenever the conditional expectation of u_{T+h} is zero or in other words, if u_t is a martingale difference sequence. Even if the u_t's are just uncorrelated and do not have conditional mean zero, the forecasts obtained recursively from (6.1) are still best *linear* forecasts but may do not be minimum MSE forecasts in a larger class which includes nonlinear forecasts.

These results are valid even if the VAR process has $I(1)$ components. However, if y_t is $I(0)$ (stationary) the forecast MSEs are bounded as the horizon h goes to infinity. In contrast, for $I(1)$ processes the forecast MSE matrices are unbounded and, hence, forecast uncertainty increases without bounds for increasing forecast horizon.

Notice the major difference between considering deterministic and stochastic trends in a VAR model. The deterministic time trend in (6.1) does not add to the inaccuracy of the forecasts in this framework, where no estimation uncertainty is present, while stochastic trends have a substantial impact on the forecast uncertainty. Many researchers find it implausible that trending behavior is not reflected in the uncertainty of long-term forecasts. Therefore deterministic trend components should be used with caution. In particular, higher order polynomial trends or even linear trends should be avoided unless there are very good reasons for them. Using them just to improve the fit of a VAR model can be counterproductive from a forecasting point of view.

For Gaussian VAR processes y_t with $u_t \sim$ iid $\mathcal{N}(0, \Sigma_u)$, the forecast errors are also multivariate normal, $y_{T+h} - y_{T+h|T} \sim \mathcal{N}(0, \Sigma_y(h))$, and forecast intervals can be set up in

the usual way. For non-Gaussian processes y_t with unknown distribution other methods for setting up forecast intervals are called for, for instance, bootstrap methods may be considered (see, e.g., Findley (1986), Masarotto (1990), Grigoletto (1998), Kabaila (1993), Kim (1999) and Pascual, Romo and Ruiz (2004)).

6.2 Forecasting Estimated VAR Processes

If the DGP is unknown and, hence, the VAR model only approximates the true DGP, the previously discussed forecasts will not be available. Let $\hat{y}_{T+h|T}$ denote a forecast based on a VAR model which is specified and estimated based on the available data. Then the forecast error is

$$y_{T+h} - \hat{y}_{T+h|T} = (y_{T+h} - y_{T+h|T}) + (y_{T+h|T} - \hat{y}_{T+h|T}). \tag{6.5}$$

If the true DGP is a VAR process, the first term on the right-hand side is $\sum_{j=0}^{h-1} \Phi_j u_{T+h-j}$. It includes residuals u_t with $t > T$ only, whereas the second term involves just $y_T, y_{T-1}, \ldots$, if only variables up to time T have been used for model specification and estimation. Consequently, the two terms are independent or at least uncorrelated so that the MSE matrix has the form

$$\Sigma_{\hat{y}}(h) = E[(y_{T+h} - \hat{y}_{T+h|T})(y_{T+h} - \hat{y}_{T+h|T})'] = \Sigma_y(h) + \text{MSE}(y_{T+h|T} - \hat{y}_{T+h|T}). \tag{6.6}$$

If the VAR model specified for y_t properly represents the DGP, the last term on the right-hand side approaches zero as the sample size gets large because the difference $y_{T+h|T} - \hat{y}_{T+h|T}$ vanishes asymptotically in probability under standard assumptions. Thus, if the theoretical model fully captures the DGP, specification and estimation uncertainty is not important asymptotically. On the other hand, in finite samples the precision of the forecasts depends on the precision of the estimators. Suitable correction factors for MSEs and forecast intervals for stationary processes are given by Baillie (1979), Reinsel (1980), Samaranayake and Hasza (1988) and Lütkepohl (2005, Chapter 3). A discussion of extensions with a number of further references may be found in Lütkepohl (2009).

7 Granger-Causality Analysis

Because VAR models describe the joint generation process of a number of variables, they can be used for investigating relations between the variables. A specific type of relation was pointed out by Granger (1969) and is known as *Granger-causality*. Granger called a variable y_{2t} causal for a variable y_{1t} if the information in past and present values of y_{2t} is helpful for improving the forecasts of y_{1t}. This concept is especially easy to implement in a VAR framework. Suppose that y_{1t} and y_{2t} are generated by a bivariate VAR(p) process,

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

Then y_{2t} is not Granger-causal for y_{1t} if and only if $\alpha_{12,i} = 0$, $i = 1, 2, \ldots, p$. In other words, y_{2t} is not Granger-causal for y_{1t} if the former variable does not appear in the y_{1t} equation of the model. This result holds for both stationary and integrated processes.

Because Granger-noncausality is characterized by zero restrictions on the levels VAR representation of the DGP, testing for it becomes straightforward. Standard Wald χ^2- or F-tests can be applied. If y_t contains integrated and possibly cointegrated variables these tests may not have standard asymptotic properties, however (Toda and Phillips (1993)). For the presently considered case, there is a simple way to fix the problem. In this case the problem of getting a nonstandard asymptotic distribution for Wald tests for zero restrictions can be resolved by adding an extra redundant lag to the VAR in estimating the parameters of the process and testing the relevant null hypothesis on the matrices $A_1, \ldots, A_p$ only (see Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996)). Since a VAR($p+1$) is an appropriate model with $A_{p+1} = 0$ if the true VAR order is p, the procedure is sound. It will not be fully efficient, however, due to the redundant VAR lag.

If there are more than two variables the conditions for non-causality or causality become more complicated even if the DGP is a VAR process (see, e.g., Lütkepohl (1993) and Dufour and Renault (1998)). In practice, Granger-causality is therefore often investigated for bivariate processes. It should be clear, however, that Granger-causality depends on the information set considered. In other words, even if a variable is Granger-causal in a bivariate model, it may not be Granger-causal in a larger model involving more variables. For instance, there may be a variable driving both variables of a bivariate process. When that variable is added to the model, a bivariate causal structure may disappear. In turn it is also possible that a variable is non-causal for another one in a bivariate model and becomes causal if the information set is extended to include other variables as well. There are also a number of other limitations of the concept of Granger-causality which have stimulated an extensive discussion of the concept and have promted alternative definitions. For further discussion and references see Lütkepohl (2005, Section 2.3.1) and for extensions to testing for Granger-causality in infinite order VAR processes see Lütkepohl and Poskitt (1996) and Saikkonen and Lütkepohl (1996).

8 Structural Analysis

Traditionally the interaction between economic variables is studied by considering the effects of changes in one variable on the other variables of interest. In VAR models changes in the variables are induced by nonzero residuals, that is, by shocks which may have a structural interpretation if identifying structural restrictions have been placed accordingly. Hence, to study the relations between the variables the effects of nonzero residuals or shocks are traced through the system. This kind of analysis is known as impulse response analysis. It will be discussed in Section 8.1. Related tools are forecast error variance decompositions and historical decompositions of time series of interest in terms of the contributions attributable to the different structural shocks. Moreover, forecasts conditional on a specific path of a variable or set of variables may be considered. These tools are discussed in Sections 8.2, 8.3 and 8.4, respectively.

8.1 Impulse Response Analysis

In the reduced form VAR model (2.5) impulses, innovations or shocks enter through the residual vector $u_t = (u_{1t}, \ldots, u_{Kt})'$. A nonzero component of u_t corresponds to an equivalent change in the associated left-hand side variable which in turn will induce further changes

in the other variables of the system in the next periods. The marginal effect of a single nonzero element in u_t can be studied conveniently by inverting the VAR representation and considering the corresponding moving average (MA) representation. Ignoring deterministic terms because they are not important for impulse response analysis gives

$$y_t = A(L)^{-1}u_t = \Phi(L)u_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \tag{8.1}$$

where $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j = A(L)^{-1}$. The $(K \times K)$ coefficient matrices Φ_j are precisely those given in (6.3). The marginal response of $y_{n,t+j}$ to a unit impulse u_{mt} is given by the (n,m)th elements of the matrices Φ_j, viewed as a function of j. Hence, the elements of Φ_j represent responses to u_t innovations. Because the u_t are just the 1-step forecast errors, these impulse responses are sometimes called *forecast error impulse responses* (Lütkepohl (2005, Section 2.3.2)) and the corresponding MA representation is called Wold MA representation.

The existence of the representation (8.1) is ensured if the VAR process is stable and, hence, y_t consists of stationary ($I(0)$) variables. In that case $\Phi_j \to 0$ as $j \to \infty$ and the effect of an impulse is *transitory*. If y_t has $I(1)$ components, the Wold MA representation (8.1) does not exist. However, for any finite j, Φ_j can be computed as in the stationary case, using the formula (6.3). Thus, impulse responses can also be computed for $I(1)$ processes. For such processes the marginal effects of a single shock may lead to *permanent* changes in some or all of the variables.

Because the residual covariance matrix Σ_u is generally not diagonal, the components of u_t may be contemporaneously correlated. Consequently, the u_{jt} shocks are not likely to occur in isolation in practice. Therefore tracing such shocks may not reflect what actually happens in the system if a shock hits. In other words, forecast error shocks may not be the right ones to consider if one is interested in understanding the interactions within the system under consideration. Therefore researchers typically try to determine structural shocks and trace their effects. A main task in structural VAR analysis is in fact the specification of the shocks of interest.

If an identified structural form such as (2.8) is available, the corresponding residuals are the structural shocks. For a stationary process their corresponding impulse responses can again be obtained by inverting the VAR representation,

$$y_t = (\mathsf{A} - A_1^* L - \cdots - A_p^* L^p)^{-1}\mathsf{B}v_t = \sum_{j=1}^{\infty} \Phi_j \mathsf{A}^{-1}\mathsf{B}v_{t-j} = \sum_{j=1}^{\infty} \Psi_j v_{t-j}, \tag{8.2}$$

where the $\Psi_j = \Phi_j \mathsf{A}^{-1}\mathsf{B}$ contain the structural impulse responses. The latter formulas can also be used for computing structural impulse responses for $I(1)$ processes even if the representation (8.2) does not exist.

Estimation of impulse responses is straightforward by substituting estimated reduced form or structural form parameters in the formulas for computing them. Suppose the structural form VAR parameters are collected in the vector $\boldsymbol{\alpha}$ and denote its estimator by $\hat{\boldsymbol{\alpha}}$. Moreover, let ψ be the vector of impulse response coefficients of interest. This vector is a (nonlinear) function of $\boldsymbol{\alpha}$, $\psi = \psi(\boldsymbol{\alpha})$, which can be estimated as $\hat{\psi} = \psi(\hat{\boldsymbol{\alpha}})$. Using the delta method, it is easy to see that $\hat{\psi} = \psi(\hat{\boldsymbol{\alpha}})$ is asymptotically normal if $\hat{\boldsymbol{\alpha}}$ has this property. More precisely,

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

implies

$$\sqrt{T}(\hat{\psi} - \psi) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\psi}}), \tag{8.3}$$

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, $\partial\psi/\partial\boldsymbol{\alpha}' \neq 0$. If $\partial\psi/\partial\boldsymbol{\alpha}'$ does not have full row rank, the asymptotic covariance matrix $\Sigma_{\hat{\psi}}$ is singular. This problem will arise 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 $\boldsymbol{\alpha}$. Also, $\Sigma_{\hat{\boldsymbol{\alpha}}}$ is generally 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. In the present case, both problems may occur jointly. A singular asymptotic covariance matrix may give rise to misleading inference for impulse responses. For further discussion see Benkwitz, Lütkepohl and Neumann (2000).

Even in those parts of the parameter space where standard asymptotic theory works, it is known that the actual small sample distributions of impulse responses may be quite different from their asymptotic counterparts. In particular, the accuracy of the confidence intervals tends to be low for large-dimensional VARs at longer horizons if the data are highly persistent, that is, if the process has roots close to the unit circle (see Kilian and Chang (2000)). Therefore attempts have been made to use local-to-unity asymptotics for improving inference in this situation. Earlier attempts in this context are Stock (1991), Wright (2000), Gospodinov (2004) and more recent articles using that approach are Pesavento and Rossi (2006) and Mikusheva (2011).

In practice, bootstrap methods are often used in applied work to construct impulse response confidence intervals (e.g., Kilian (1998), Benkwitz, Lütkepohl and Wolters (2001)). Although they have the advantage that complicated analytical expressions of the asymptotic variances are not needed, it is not clear that they lead to substantially improved inference. In particular, they are also justified by asymptotic theory. In general the bootstrap does not overcome the problems due to a singularity in the asymptotic distribution. Consequently bootstrap confidence intervals may have a coverage which does not correspond to the nominal level and may, hence, be unreliable (see Benkwitz et al. (2000)). Using subset VAR techniques to impose as many zero restrictions on the parameters as possible and estimating only the remaining nonzero parameters offers a possible solution to this problem.

Bayesian methods provide another possible solution (e.g. Sims and Zha (1999)). If an a posteriori distribution is available for $\hat{\boldsymbol{\alpha}}$, it can be used to simulate the distribution of $\hat{\psi} = \psi(\hat{\boldsymbol{\alpha}})$ using standard Bayesian simulation techniques. That distribution can then be used for setting up confidence intervals or for inference on ψ. As Bayesian inference does not rely on asymptotic arguments, the singularity problem is not relevant. This does not mean that Bayesian estimation is necessarily more reliable. It requires extensive computations and is based on distributional assumptions which may be questionable.

8.2 Forecast Error Variance Decompositions

As mentioned earlier, forecast error variance decompositions are another tool for investigating the impacts of shocks in VAR models. In terms of the structural residuals the h-step forecast

error (6.2) can be represented as

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

Using $\Sigma_v = I_K$, the forecast error variance of the kth component of y_{T+h} 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 quantity $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)$ represents the contribution of the jth shock to the h-step forecast error variance of variable k. In practice, the relative contributions $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)/\sigma_k^2(h)$ are often reported and interpreted for various variables and forecast horizons. A meaningful interpretation of these quantities requires that the shocks considered in the decomposition are economically meaningful.

The quantities of interest here can again be estimated easily by replacing unknown parameters by their estimators. Inference is complicated by the fact, however, that the relative variance shares may be zero or one and, hence, may assume boundary values. In such cases both classical asymptotic as well as bootstrap methods have problems.

8.3 Historical Decomposition of Time Series

Another way of looking at the contributions of the structural shocks to the observed series is opened up by decomposing the series as proposed by Burbidge and Harrison (1985). Neglecting deterministic terms and considering the structural MA representation (8.2), the jth variable can be represented as

$$y_{jt} = \sum_{i=0}^{\infty} (\psi_{j1,i} v_{1,t-i} + \cdots + \psi_{jK,i} v_{K,t-i}),$$

where $\psi_{jk,i}$ is the (j,k)th element of the structural MA matrix Ψ_i, as before. Thus,

$$y_{jt}^{(k)} = \sum_{i=0}^{\infty} \psi_{jk,i} v_{k,t-i}$$

is the contribution of the kth structural shock to the jth variable y_{jt}. Ideally one would like to plot the $y_{jt}^{(k)}$ for $k = 1, \ldots, K$, throughout the sample period, that is, for $t = 1, \ldots, T$, and interpret the relative contributions of the different structural shocks to the jth variable.

In practice, such a *historical decomposition* is, of course, not feasible because the structural shocks are not available. However, we can estimate the shocks associated with the sample period and use an estimated historical decomposition by noting that by successive substitution, the VAR process (2.5) can be written as

$$\begin{aligned} y_t &= \sum_{i=0}^{t-1} \Phi_i u_{t-i} + A_1^{(t)} y_0 + \cdots + A_p^{(t)} y_{-p+1} \\ &= \sum_{i=0}^{t-1} \Psi_i v_{t-i} + A_1^{(t)} y_0 + \cdots + A_p^{(t)} y_{-p+1}, \end{aligned} \tag{8.4}$$

where the Φ_i and Ψ_i are the MA coefficient matrices defined earlier and the $A_i^{(t)}$ are such that $[A_1^{(t)}, \ldots, A_p^{(t)}]$ consists of the first K rows of the $(pK \times pK)$ matrix $\mathbf{A}^t$, where

$$\mathbf{A} = \begin{bmatrix} A_1 & \ldots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & I_K & 0 \end{bmatrix}$$

(see Lütkepohl (2005, Sec. 2.1)). Hence, the $A_i^{(t)}$ go to zero for stationary VARs when t becomes large so that the contribution of the initial state becomes negligible for stationary processes as $t \to \infty$. On the other hand, for $I(1)$ processes the contribution of the initial values $y_0, \ldots, y_{-p+1}$ will remain important. In any case, y_{jt} may be decomposed as

$$y_{jt}^{(k)} = \sum_{i=0}^{t-1} \psi_{jk,i} v_{k,t-i} + \alpha_{j1}^{(t)} y_0 + \cdots + \alpha_{jp}^{(t)} y_{-p+1},$$

where $\psi_{jk,i}$ is the (j,k)th element of Ψ_i and $\alpha_{ji}^{(t)}$ is the jth row of $A_i^{(t)}$. The series $y_{jt}^{(k)}$ represents the contribution of the kth structural shock to the jth component series of y_t, given $y_0, \ldots, y_{-p+1}$. In practice all unknown parameters have to be replaced by estimators. The corresponding series $\hat{y}_{jt}^{(k)}$, $k = 1, \ldots, K$, represent a *historical decomposition* of y_{jt}. They are typically plotted to assess the contributions of the structural shocks to the jth series. Obviously, one may start the decomposition at any point in the sample and not necessarily at $t = 0$. In fact, for t close to the starting point of the decomposition the initial values may have a substantial impact even for stationary processes. So one may only want to consider the decomposition for periods some distance away from the starting point.

8.4 Analysis of Forecast Scenarios

SVAR models have also been used for analyzing different forecast scenarios or conditional forecasts given restrictions for the future values of some of the variables. For example, in monetary policy analysis one may be interested in knowing the future development of the system under consideration for a given path of the interest rate or if the interest rate remains within a given range. Clearly, in a model where all variables are endogenous, fixing the future values of one or more variables may be problematic and one has to evaluate carefully how far the model can be stretched without being invalidated. In other words, SVAR models cannot be expected to reflect the changes induced in the future paths of the variables for arbitrary forecast scenarios (for applications see, e.g., Waggoner and Zha (1999), Baumeister and Kilian (2011)).

For describing the approach, a SVAR representation similar to (8.4) is particularly suitable,

$$y_{T+h} = \sum_{i=0}^{h-1} \Psi_i v_{T+h-i} + A_1^{(h)} y_T + \cdots + A_p^{(h)} y_{T-p+1}, \quad h = 1, 2, \ldots, \tag{8.5}$$

where deterministic terms are again ignored for simplicity and all symbols are defined as in (8.4). The standard reduced form forecast $y_{T|T+h}$ discussed in Section 6.1 is obtained

from this expression by replacing all structural residuals in the first term on the right-hand side by zero. A forecast scenario different from this baseline forecast may be obtained by assigning other values to the structural shocks. For instance, a scenario where the jth variable has future values $y^*_{j,T+h}$ for $h = 1, \ldots, H$, amounts to choosing structural shocks v^*_{T+h}, $h = 1, \ldots, H$, such that

$$\sum_{i=0}^{h-1} \sum_{k=1}^{K} \psi_{jk,i} v^*_{k,T+h-i} = y^*_{j,T+h} - \alpha^{(h)}_{j1} y_T - \cdots - \alpha^{(h)}_{jp} y_{T-p+1} \quad h = 1, \ldots, H, \tag{8.6}$$

for periods $T+1, \ldots, T+H$ or, more generally in matrix notation,

$$R_H \mathbf{v}^*_{T,T+H} = r_H, \tag{8.7}$$

where $\mathbf{v}_{T,T+H} = (v'_{T+1}, \ldots, v'_{T+H})'$ is a $(KH \times 1)$ vector of stacked future structural shocks, R_H is a suitable $(Q \times KH)$ dimensional restriction matrix representing the left-hand side relations in (8.6) with $Q \leq KH$ and r_H is a $(Q \times 1)$ vector containing the right-hand side of (8.6). The forecasts conditional on the v^*_{T+h} shocks are then computed as

$$y^{cond}_{T|T+h} = \sum_{i=0}^{h-1} \Psi_i v^*_{T+h-i} + A^{(h)}_1 y_T + \cdots + A^{(h)}_p y_{T-p+1}, \quad h = 1, \ldots, H. \tag{8.8}$$

For concreteness consider, for instance, the case of a B-model with lower-triangular initial effects matrix $\mathsf{B} = \Psi_0$. In that case, if the path of the first variable is prespecified as $y^*_{1,T+h}$, $h = 1, \ldots, H$, this amounts to choosing the first residual as $v^*_{1,T+h} = y^*_{1,T+h} - \alpha^{(1)}_{11} y^{cond}_{T+h-1|T} - \cdots - \alpha^{(1)}_{1p} y^{cond}_{T+h-p|T}$, $h = 1, \ldots, H$. In general, the values for the v^*_{T+h}'s will not be uniquely determined by the restrictions. In that case Doan et al. (1984) suggest using

$$\mathbf{v}^*_{T,T+H} = R'_H (R_H R'_H)^{-1} r_H$$

which is the least squares solution obtained by minimizing $\sum_{h=1}^{H} v'_{T+h} v_{T+h} = \mathbf{v}'_{T,T+H} \mathbf{v}_{T,T+H}$ subject to the restrictions (8.7).

Obviously, the conditional forecast $y^{cond}_{T|T+h}$ differs from the unconditional forecast $y_{T|T+h}$ of Section 6.1 by the first term on the right-hand side of equation (8.8). Of course, other forecast scenarios may be of interest. For example, one may not want to condition on a particular path of a variable but on its values being in a given range. Such a scenario can be investigated by using the above formulas for computing forecast ranges accordingly.

For practical purposes the unknown parameters have to be replaced by estimates, as usual. Waggoner and Zha (1999) present a Bayesian method for taking into account the related estimation uncertainty in that case.

A critical question in the context of evaluating forecast scenarios in the context of SVAR models is whether such models are suitable for that purpose or whether they are stretched too much by restricting the future developments of some variables. The problem here is that all variables are regarded as endogenous and, hence, their values should be generated endogenously by the model and not be forced upon them exogenously. Of course, there could be cases where one or more of the variables are really exogenous and not affected by feedback relations within the system under consideration. In such cases a conditional analysis as described in the foregoing is plausible. In other cases the users should be cautious in interpreting the results and carefully evaluate whether the model can be expected to be informative about the questions of interest.

9 Conclusions and Extensions

This chapter reviews VAR analysis. Specification, estimation, forecasting, causality and structural analysis are discussed. Finite order VAR models are popular for economic analysis because they are easy to use. There are several software products which can be used in performing a VAR analysis (see, e.g., *PcGive* (Doornik and Hendry (1997)), *EViews* (EViews (2000)) and *JMulTi* (Krätzig (2004)).

In many situations of practical interest the VAR models discussed in this chapter are too limited, however. For example, the assumption that the VAR order is finite is rather restrictive because theoretically omitted variables or linear transformations of the variables may lead to infinite order VAR processes. Hence, it may be appealing to extend the model class to infinite order VAR processes. Such an extension may be achieved by considering VAR models with MA error terms or by studying the implications of fitting approximate finite order VAR models to series which are generated by infinite order processes. This issue is dealt with in Lewis and Reinsel (1985), Lütkepohl and Saikkonen (1997), Inoue and Kilian (2002b) and Lütkepohl (2005, Part IV). An authoritative discussion of the theory of VARMA models is also available in Hannan and Deistler (1988) and a recent survey of the related literature is given by Lütkepohl (2006a).

As already mentioned in the introduction, if some variables are integrated and perhaps cointegrated, vector error correction models are suitable tools for modelling the cointegration relations in detail. Such models are presented in Chapter 9. A possible advantage of the levels VAR models considered in the present chapter is that they are robust to cointegration of unknown form. They can be used even if the number of cointegration relations is unknown not to speak of the precise cointegration relations. Of course, statistical tools are available for analyzing the number and type of cointegration relations. However, such pretesting procedures have their limitations as well, in particular if some roots are only near unity, as pointed out, for instance, by Elliott (1998).

Other possible extensions are the inclusion of nonlinear components (e.g., Granger (2001), Teräsvirta, Tjøstheim and Granger (2010)) or allowing the VAR coefficients to be time-varying (e.g., Primiceri (2005)). Moreover, seasonal macroeconomic data are now available for many countries. Hence, accounting specifically for seasonal fluctuations may be necessary (e.g., Ghysels and Osborn (2001)). Furthermore, heteroskedasticity or conditional heteroskedasticity in the residuals of VAR models may be of importance in practice, in particular, when higher frequency data are considered (see Lütkepohl (2005, Chapter 16) for a textbook treatment and Bauwens, Laurent and Rombouts (2006) or Silvennoinen and Teräsvirta (2009) for recent surveys). For some economic variables restricting the order of integration to zero or one may be unrealistic. Extensions to higher order integration have been considered by Boswijk (2000), Johansen (1997, 2006) and others.

References

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

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.

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.

Baillie, R. T. (1979). Asymptotic prediction mean squared error for vector autoregressive models, *Biometrika* **66**: 675–678.

Baumeister, C. and Kilian, L. (2011). Real-time forecasts of the real price of oil, *mimeo*, University of Michigan.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey, *Journal of Applied Econometrics* **21**: 79–109.

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.

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.

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

Burbidge, J. and Harrison, A. (1985). A historical decomposition of the great depression to determine the role of money, *Journal of Monetary Economics* **16**: 45–54.

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

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

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

Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots, *Econometrica* **66**: 149–158.

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

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

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.

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

Giannone, D., Lenza, M. and Primiceri, G. (2010). Prior selection for vector autoregressions, *mimeo*, Department of Economics, Free University of Brussels.

Goncalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**: 89–120.

Gospodinov, N. (2004). Asymptotic confidence intervals for impulse responses of near-integrated processes, *Econometrics Journal* **7**: 505–527.

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. (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.

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

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.

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.

Inoue, A. and Kilian, L. (2002a). Bootstrapping autoregressive processes with possible unit roots, *Econometrica* **70**: 377–391.

Inoue, A. and Kilian, L. (2002b). Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models, *International Economic Review* **43**: 309–332.

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. (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. (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.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models, *Journal of Applied Econometrics* **12**: 99–132.

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

Kilian, L. (2011). Structural vector autoregressions, *Handbook of Research Methods and Applications on Empirical Macroeconomics*, Edward Elger.

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

Kilian, L. and Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence, *Journal of Business & Economic Statistics* **18**: 40–50.

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

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

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction, *Econometric Theory* **21**: 21–59.

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.

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. (2009). Econometric analysis with vector autoregressive models, *in* D. A. Belsley and E. J. Kontoghiorghes (eds), *Handbook of Computational Econometrics*, Wiley, Chichester, pp. 281–319.

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

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

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

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

Mikusheva, A. (2011). One-dimensional inference in autoregressive models with the potential presence of a unit root, *mimeo*, Department of Economics, M.I.T.

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.

Pesavento, E. and Rossi, B. (2006). Small-sample confidence intervals for multivariate impulse response functions at long horizons, *Journal of Applied Econometrics* **21**: 1135–1155.

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

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

Reinsel, G. (1980). Asymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters, *Journal of the Royal Statistical Society* **B42**: 328–333.

Rissanen, J. (1978). Modeling by shortest data description, *Automatica* **14**: 465–471.

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

Samaranayake, V. A. and Hasza, D. P. (1988). Properties of predictors for multivariate autoregressive models with estimated parameters, *Journal of Time Series Analysis* **9**: 361–383.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**: 461–464.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreis and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 201–228.

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

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., Waggoner, D. F. and Zha, T. (2008). Methods for inference in large multiple-equation Markov-switching models, *Journal of Econometrics* **146**: 255–274.

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

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in U.S. economic time series, *Journal of Monetary Economics* **28**: 435–460.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

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.

Villani, M. (2005). Bayesian reference analysis of cointegration, *Econometric Theory* **21**: 326–357.

Waggoner, D. F. and Zha, T. (1999). Conditional forecasts in dynamic multivariate models, *Review of Economics and Statistics* **81**: 639–651.

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

Wright, J. H. (2000). Confidence intervals for univariate impulse responses with a near unit root, *Journal of Business & Economic Statistics* **18**: 368–373.


European
University
Institute
European University Press