

Lütkepohl, Helmut

Working Paper

Forecasting cointegrated VARMA processes

SFB 373 Discussion Paper, No. 1999,68

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,
Humboldt University Berlin

Suggested Citation: Lütkepohl, Helmut (1999) : Forecasting cointegrated VARMA processes, SFB 373 Discussion Paper, No. 1999,68, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
<https://nbn-resolving.de/urn:nbn:de:kobv:11-10046579>

This Version is available at:

<https://hdl.handle.net/10419/61731>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Forecasting Cointegrated VARMA Processes

Helmut Lütkepohl¹

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

1 Motivation and Chapter Outline

In forecasting time series variables the information in past values of the series is used to extract potential future developments. In other words, a forecast is a function of past values of a time series variable. If linear functions are considered only this leads to autoregressive (AR) models where a variable at time t , y_t say, depends on lagged values plus the forecast error, $y_t = \nu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t$. Here u_t represents the forecast error term, α_i ($i = 1, \dots, p$) are autoregressive parameters and ν is an intercept term. Similarly, if more than one variable is involved, say y_{1t}, \dots, y_{Kt} , each component is viewed as a linear function of lags of all the variables plus an error term in order to use information from all the past variables. Using vector and matrix notation, this leads to vector autoregressive (VAR) models of the form $y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$, where $y_t = (y_{1t}, \dots, y_{Kt})'$, the A_i ($i = 1, \dots, p$) are $(K \times K)$ coefficient matrices, ν is a $(K \times 1)$ intercept term and u_t is a K -dimensional error term.

In small samples the precision of forecasts based on these models will be affected by the precision of the parameter estimators. Since an adequate description of the data generation process (DGP) of a multiple time series may require a fairly large VAR order and hence a large number of parameters, the estimation precision in such models may be low. Consequently, forecasts based on VAR processes may suffer from the uncertainty in the parameter estimators. In such a situation it may be worth considering a larger model class with processes which may be able to represent the DGP of interest in a more parsimonious way. Vector autoregressive moving average (VARMA) models constitute such a class. In this chapter the analysis of models from that class will be discussed.

One of the problems in dealing with VARMA models is that generally their parameterization is not unique. For inference purposes it is necessary to focus on a unique representation of a DGP. In the following a variant of the *echelon form* of a VARMA process will be considered. It has the advantage of providing a relatively parsimonious parameterization in general and, following Poskitt (1992), we use the acronym ARMA_E for the echelon form.

¹I thank the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) for financial support.

In the next section this representation of a VARMA process will be described formally. Generally, the variables are allowed to be integrated of order one ($I(1)$) and they may be cointegrated. Therefore it will be convenient to supplement the echelon form with an error correction (EC) term. Thereby it is possible to separate the long-run cointegration relations from the short-term dynamics. This results in an EC-ARMA_E form which turns out to be a convenient framework for modeling and forecasting cointegrated variables. The model and some of its pertinent properties will be discussed in Sec. 2.

In using these models for forecasting a specific candidate from the general model class has to be chosen, its parameters have to be estimated and then point forecasts may be computed and, if desired, forecast intervals may be established. In this procedure we have to pass through the usual model specification, estimation and model checking cycles before the forecasts are determined. The necessary steps and procedures will be considered in this chapter. In Sec. 3 estimation procedures will be discussed under the assumption that a well specified model is available. The ARMA_E form of a VARMA process is characterized by a set of integer parameters called Kronecker indices. A procedure for consistently estimating these quantities and also the cointegrating rank from a given multiple time series will be discussed in Sec. 4. Moreover, some comments on model checking will be given in that section. Thereby a complete strategy for analyzing EC-ARMA_E models is provided. Once a well specified model is available it can be used for forecasting which is considered in Sec. 5. Conclusions follow in Sec. 6.

This chapter draws heavily on material from Lütkepohl & Claessen (1997) who introduced the error correction echelon form of a VARMA process, Poskitt & Lütkepohl (1995) who presented consistent estimation procedures for the Kronecker indices, and Bartel & Lütkepohl (1998) who explored the small sample properties of some such procedures. To simplify the exposition, some important ideas and concepts will be discussed in the framework of VAR models first which are treated as special VARMA models.

2 The VARMA Framework

2.1 Characteristics of Variables

The characteristics of the variables involved determine to some extent which model is a suitable representation of the DGP. For instance, the trending properties of the variables and their seasonal fluctuations are of importance in setting up a suitable model. In the following a variable is called *integrated of order d* ($I(d)$) if stochastic trends or unit roots can be removed by differencing the variable d times. A variable without a stochastic trend or unit root is sometimes called $I(0)$. In the present chapter it is assumed that all variables are at most $I(1)$ if not otherwise stated. In other words, for any time series variable y_{kt} it is assumed that $\Delta y_{kt} \equiv y_{kt} - y_{k,t-1}$ has no stochastic trend. Note, however, that in general Δy_{kt} may still have deterministic components such as a polynomial trend and seasonal components. For convenience it is assumed in the following that there are no deterministic terms. In other words, polynomial trend terms, seasonal dummy variables or nonzero mean terms are not considered. It is straightforward to include a nonzero mean term or regard the variables as being mean-adjusted and for most procedures discussed in the following other deterministic terms can be accommodated in a straightforward manner.

A set of $I(1)$ variables is called *cointegrated* if a linear combination exists which is $I(0)$

(Engle & Granger (1987)). Generally it is convenient to consider systems with both $I(1)$ and $I(0)$ variables. In this case the concept of cointegration is extended by calling any linear combination which is $I(0)$ a cointegration relation. Clearly, this terminology is not in the spirit of the original definition because it leads to a linear combination of $I(0)$ variables being called a cointegration relation. For our purposes a distinction between genuine cointegration and more general forms is not necessary, however.

2.2 VAR and Error Correction Models

As mentioned in the introduction, for a system of K time series variables $y_t = (y_{1t}, \dots, y_{Kt})'$ a VAR model is a useful forecasting tool. The basic VAR model is of the form

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad (2.1)$$

where $u_t = (u_{1t}, \dots, u_{Kt})'$ is an unobservable zero mean independent white noise process with time invariant positive definite covariance matrix $E(u_t u_t') = \Sigma_u$, the A_i ($i = 1, \dots, p$) are $(K \times K)$ coefficient matrices. This model is often briefly referred to as a VAR(p) process because the number of lags is p .

The process is said to be *stable* if

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

Here I_K denotes the $(K \times K)$ identity matrix. Assuming that a stable process has been initiated in the infinite past, it generates *stationary* time series which have time invariant means, variances and covariance structure. If the determinantal polynomial in (2.2) has unit roots (i.e., roots for $z = 1$), then some or all of the variables are $I(1)$ and they may also be cointegrated. Thus, the present model accommodates variables with stochastic trends. On the other hand, it is not the most suitable setup for analyzing cointegration relations because these relations do not appear explicitly in (2.1). They are more easily analyzed within a model obtained by rewriting (2.1) as

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

where $\Pi = -(I_K - A_1 - \dots - A_p)$ and $\Gamma_i = -(A_{i+1} + \dots + A_p)$ for $i = 1, \dots, p-1$. This reparameterization of the process is obtained from (2.1) by subtracting y_{t-1} from both sides and rearranging terms. The model form (2.3) is the so-called *vector error correction model* (VECM). Because Δy_t does not contain stochastic trends by our assumption that all variables can be at most $I(1)$, the term Πy_{t-1} is the only one which may include $I(1)$ variables. Hence, Πy_{t-1} must also be $I(0)$. Thus, it contains the cointegrating relations. These relations represent long-run or equilibrium relations and the model is written in a form which shows how the variables respond to deviations from the long-run relations. In other words, it can be seen how equilibrium errors are corrected which explains the name of the model. The Γ_j ($j = 1, \dots, p-1$) in (2.3) are often referred to as the short-term or short-run parameters.

To distinguish the VECM from the VAR model the latter is sometimes called the levels version. Of course, it is also possible to determine the A_j levels parameter matrices from the coefficients of the VECM as $A_1 = \Gamma_1 + \Pi + I_K$, $A_i = \Gamma_i - \Gamma_{i-1}$ for $i = 2, \dots, p-1$, and $A_p = -\Gamma_{p-1}$. Thus, the two forms (2.1) and (2.3) are equivalent representations of a

stochastic process. Which one of the two representations is used for a specific analysis is mainly a question of convenience with respect to model specification, estimation or analysis.

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

Special cases included in (2.3) are $I(0)$ processes for which $r = K$ and systems that have a stable, stationary VAR representation in first differences. In the latter case, $r = 0$ and the term Πy_{t-1} disappears in (2.3). These boundary cases do not represent cointegrated systems in the usual sense. There are also other cases which are not in line with the original idea of the concept of cointegration even if the cointegrating rank is strictly between zero and K . An obvious example is obtained if all variables but one are $I(0)$. In that case the cointegrating rank is $K - 1$ although the only $I(1)$ variable is not cointegrated with the other variables. For our purposes, it is no problem to include those cases here.

So far I have not been precise about the range of the index t . If the process is $I(0)$ than assuming that the range consists of all integers is convenient. In that case, in each period t the process is already in operation for an infinite time and, hence, the moments and distributions have stabilized and are therefore time invariant. On the other hand, if there are $I(1)$ variables, it is usually more convenient from a theoretical point of view to assume that the process has been initialized in some period, say in $t = 0$ or $t = 1$ and has a finite past only because otherwise the variables may have infinite variances. Since $I(1)$ variables are generally allowed for in the following, it will be assumed that the process under consideration has been initiated at some finite time if not otherwise stated. All variables indexed by integers smaller than the initial period may be assumed to be zero if not otherwise specified. For instance, if the process y_t is initialized at $t = 1$, the variables y_0, y_{-1} etc. which may appear in a VAR model are assumed to be zero. Again this is just a convenient simplification for expository purposes. More general assumptions are possible for the theoretical results of the following sections to hold.

2.3 VARMA Models

As mentioned in the introduction, the VAR class of processes occasionally has the disadvantage that a quite large order p is necessary for a proper representation of the DGP. In fact, in some cases theoretical considerations lead to infinite order VAR processes (see also Sec. 2.3.4). Therefore we consider the following more general process class which includes infinite order VAR processes as well.

2.3.1 General Form

It is assumed that the DGP of the K -dimensional multiple time series y_1, \dots, y_T is from the VARMA class,

$$A_0 y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + M_0 u_t + M_1 u_{t-1} + \dots + M_p u_{t-p}, \quad t = 1, 2, \dots, \quad (2.4)$$

or

$$A(L)y_t = M(L)u_t, \quad t = 1, 2, \dots, \quad (2.5)$$

where u_t is again a white noise process with zero mean and nonsingular, time invariant covariance matrix $E(u_t u_t') = \Sigma_u$. As mentioned previously, the initial values are assumed to be zero for convenience, that is, $u_t = y_t = 0$ for $t \leq 0$. Moreover, $A(L) = A_0 - A_1 L - \dots - A_p L^p$ and $M(L) = M_0 + M_1 L + \dots + M_p L^p$ are matrix polynomials in the lag or backshift operator L , which is defined as usual by $Ly_t = y_{t-1}$. The zero order matrices A_0 and M_0 are assumed to be nonsingular. Later on it will be argued that $A_0 = M_0$ may be assumed without loss of generality. Some of the A_i and M_j coefficient matrices may be zero so that the AR or MA order may actually be less than p . The matrix polynomials are assumed to satisfy

$$\det A(z) \neq 0, |z| \leq 1, z \neq 1, \quad \text{and} \quad \det M(z) \neq 0, |z| \leq 1. \quad (2.6)$$

The second part of this condition is the usual *invertibility condition* for the MA operator which is imposed to ensure the existence of a pure (possibly infinite order) VAR representation of the type discussed below. As in the pure VAR case, the possibility that the operator $A(z)$ can have zeros for $z = 1$ implies that the components of y_t may be integrated and perhaps cointegrated variables.

Clearly, without further restrictions on the operators $A(L)$ and $M(L)$ the representation (2.4)/(2.5) is not unique. This phenomenon is easily seen by writing (2.5) in AR form

$$y_t = \sum_{i=1}^{t-1} \Xi_i y_{t-i} + u_t, \quad (2.7)$$

where $A_0 = M_0$ is assumed and $\Xi(z) = I_K - \sum_{i=1}^{\infty} \Xi_i z^i = M(z)^{-1} A(z)$. Every pair of operators $A(z), M(z)$ which leads to the same transfer function $\Xi(z)$ defines an equivalent VARMA representation for y_t . For instance, premultiplying by some nonsingular matrix results in an equivalent representation. In fact, premultiplying by some operator $D(L) = D_0 + D_1 L + \dots + D_q L^q$ satisfying $\det D_0 \neq 0$ and $\det D(z) \neq 0$ for $|z| \leq 1$, results in an equivalent VARMA representation of the process. In other words, cancellation of factors of $A(L)$ and $M(L)$ is possible. Therefore it is usually assumed that there are no unnecessary redundancies. This condition is imposed by assuming that the operator $[A(z) : M(z)]$ is left-coprime, meaning that only unimodular operators $D(z)$ can be factored from $[A(z) : M(z)]$. In other words, $[A(z) : M(z)]$ is left-coprime if the existence of $(K \times K)$ matrix polynomials $D(z), \bar{A}(z), \bar{M}(z)$ such that $[A(z) : M(z)] = D(z)[\bar{A}(z) : \bar{M}(z)]$ implies that $D(z)$ is unimodular.

Recall that a matrix polynomial $D(z)$ is unimodular if $\det D(z)$ is a constant which does not depend on z . For example,

$$D(z) = D_0 \quad \text{or} \quad D(z) = \begin{bmatrix} 1 & \delta z \\ 0 & 1 \end{bmatrix}$$

are unimodular matrix polynomials. (See Lütkepohl (1996) for definitions and properties of matrix polynomials.) Obviously, requiring that $[A(z) : M(z)]$ is left-coprime therefore does not solve the nonuniqueness problem completely because the possibility of premultiplying by some nonsingular matrix, for example, is still not excluded. Even if this possibility were excluded by assuming $A_0 = M_0 = I_K$, the nonuniqueness problem is not solved because there are unimodular operators $D(z) = I_K + D_1 z + \dots + D_q z^q$ with zero order matrix I_K (see the foregoing example). Premultiplying with such an operator still leaves $[A(z) : M(z)]$ left-coprime. Clearly more restrictions are needed for uniqueness. One possible set of restrictions is given by the echelon form which is discussed next.

2.3.2 The Echelon Form

In order to obtain a unique representation we denote the kl -th elements of $A(z)$ and $M(z)$ by $\alpha_{kl}(z)$ and $m_{kl}(z)$, respectively, and impose the following constraints. First of all, let $[A(z) : M(z)]$ be left-coprime. Moreover,

$$m_{kk}(L) = 1 + \sum_{i=1}^{p_k} m_{kk,i} L^i, \quad \text{for } k = 1, \dots, K, \quad (2.8a)$$

$$m_{kl}(L) = \sum_{i=p_k-p_{kl}+1}^{p_k} m_{kl,i} L^i, \quad \text{for } k \neq l, \quad (2.8b)$$

and

$$\alpha_{kl}(L) = \alpha_{kl,0} - \sum_{i=1}^{p_k} \alpha_{kl,i} L^i, \quad \text{with } \alpha_{kl,0} = m_{kl,0} \quad \text{for } k, l = 1, \dots, K. \quad (2.8c)$$

Here

$$p_{kl} = \begin{cases} \min(p_k + 1, p_l) & \text{for } k > l, \\ \min(p_k, p_l) & \text{for } k < l, \end{cases} \quad k, l = 1, \dots, K.$$

The process is said to be in echelon form or, briefly, ARMA_E form if $A(L)$ and $M(L)$ satisfy these restrictions. The row degrees p_k in this representation are the *Kronecker indices* (see Hannan and Deistler (1988) and Lütkepohl (1991)). In (2.4)/(2.5), $p = \max(p_1, \dots, p_K)$, that is, p is the maximum row degree or Kronecker index. $\text{ARMA}_E(p_1, \dots, p_K)$ denotes an echelon form with Kronecker indices p_1, \dots, p_K . The sum of the Kronecker indices $p_1 + \dots + p_K$ is said to be the *McMillan degree*.

As an example consider a three-dimensional process with Kronecker indices $(p_1, p_2, p_3) = (1, 2, 1)$. In this case

$$[p_{kl}] = \begin{bmatrix} \bullet & 1 & 1 \\ 1 & \bullet & 1 \\ 1 & 2 & \bullet \end{bmatrix}.$$

Hence, an $\text{ARMA}_E(1,2,1)$ has the following form:

$$\begin{aligned} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} y_t &= \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} & \alpha_{13,1} \\ \alpha_{21,1} & \alpha_{22,1} & \alpha_{23,1} \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ \alpha_{21,2} & \alpha_{22,2} & \alpha_{23,2} \\ 0 & 0 & 0 \end{bmatrix} y_{t-2} \\ &+ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & m_{12,1} & m_{13,1} \\ 0 & m_{22,1} & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ m_{21,2} & m_{22,2} & m_{23,2} \\ 0 & 0 & 0 \end{bmatrix} u_{t-2}. \end{aligned} \quad (2.9)$$

Note that in the formulation of the echelon form in (2.8) the autoregressive operator is unrestricted except for the constraints imposed by the maximum row degrees or Kronecker indices and the zero order matrix ($A_0 = M_0$) whereas additional zero restrictions are placed on the moving average coefficient matrices attached to low lags of the u_t . For example, in (2.9), there are two zero restrictions on M_1 . This representation of the echelon form was proposed by Lütkepohl & Claessen (1997) for processes with integrated and cointegrated variables. It differs from the ARMA_E form usually found in the literature on stationary processes where the restrictions on low order lags are imposed on the AR coefficient matrices (e.g., Hannan & Deistler (1988), Lütkepohl (1991)). The advantage of (2.8) in the present context is that it can be combined easily with the error correction form as will be seen shortly.

It may be worth noting that an ARMA_E form may have more zero coefficients than those specified in (2.8). Further zero restrictions may lead to models where the AR and MA orders are not identical. Such constraints are not excluded in an echelon form. However, the echelon form does not need them to ensure uniqueness of the representation of the operator $\Xi(L)$. Note also that the echelon form does not exclude processes. In other words, every VARMA process can be written in echelon form.

2.3.3 The Error Correction Echelon Form

An EC form may be obtained from (2.4) by subtracting $A_0 y_{t-1}$ on both sides and re-arranging terms as for the VECM representation of a VAR in Section 2.2:

$$\begin{aligned} A_0 \Delta y_t &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} \\ &\quad + M_0 u_t + M_1 u_{t-1} + \cdots + M_p u_{t-p} \end{aligned} \quad (2.10)$$

where $\Pi = -(A_0 - A_1 - \cdots - A_p)$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ ($i = 1, \dots, p-1$) as before. Again Πy_{t-1} is the error correction term and $r = \text{rk}(\Pi)$ is the cointegrating rank of the system which indicates the number of linearly independent cointegration relations.

If the operators $A(L)$ and $M(L)$ satisfy the echelon form restrictions, it is easily seen that the Γ_i satisfy similar identifying constraints as the A_i . More precisely, Γ_i obeys the same zero restrictions as A_{i+1} for $i = 1, \dots, p-1$ because a zero restriction on an element $\alpha_{kl,i}$ of A_i implies that the corresponding elements $\alpha_{kl,j}$ of A_j are also zero for $j > i$. For the same reason the zero restrictions on Π are the same as those on $A_0 - A_1$. This means in particular that there are no echelon form zero restrictions on Π if all Kronecker indices $p_k \geq 1$ ($k = 1, \dots, K$) because in that case the echelon form does not impose zero restrictions on A_1 . On the other hand, if there are zero Kronecker indices this has implications for the integration and cointegration structure of the variables. A specific analysis of the relations between the variables is called for in that case. Denoting by ϱ the number of Kronecker indices which are zero, it can be shown that

$$\text{rk}(\Pi) \geq \varrho. \quad (2.11)$$

This result has to be taken into account in the procedure for specifying the cointegrating rank of a VARMA system discussed in Section 6.

If the model (2.10) satisfies the echelon form restrictions it is denoted as EC-ARMA_E in

the following. As an example consider again the system (2.9). Its EC-ARMA_E form is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} \Delta y_t = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ \gamma_{21,2} & \gamma_{22,2} & \gamma_{23,2} \\ 0 & 0 & 0 \end{bmatrix} \Delta y_{t-1} \\ + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & m_{12,1} & m_{13,1} \\ 0 & m_{22,1} & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ m_{21,2} & m_{22,2} & m_{23,2} \\ 0 & 0 & 0 \end{bmatrix} u_{t-2}.$$

As a further example consider the three-dimensional ARMA_E(0,0,1) system

$$y_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1} + u_t + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1}. \quad (2.12)$$

Because two of the Kronecker indices are zero, the cointegrating rank of this system must be at least 2. The EC-ARMA_E form is easily seen to be

$$\Delta y_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} y_{t-1} + u_t + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1}.$$

Obviously, the rank of

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}$$

is at least 2.

Specifying an EC-ARMA_E model requires that the cointegrating rank r is determined, the Kronecker indices p_1, \dots, p_K are specified and possibly further overidentifying zero restrictions are placed on the coefficient matrices Γ_i and M_j . Before we discuss strategies for this task, some useful properties of linearly transformed VARMA processes are considered in the next section.

2.3.4 Linear Transformations

In many cases systems of linearly transformed variables are of interest. For example the gross national product of a country is the sum of the gross products of the different regions or the money stock variable M1 consists of currency plus sight deposits. Moreover, temporal aggregation is often just a linear transformation. For instance, the quarterly value of a variable may be the sum or the average of the monthly values or it may just be the value of the last month of a quarter. Furthermore, seasonal adjustment procedures are sometimes linear transformations such as finite moving averages of the unadjusted variables. In other words, quite often a system of variables z_t is of interest which is obtained as a linear transformation of the original system y_t , say $z_t = F y_t$, where F is a suitable transformation matrix. Notice that this notation also covers temporal aggregation if the process y_t is defined in an appropriate way (see, e.g., Lütkepohl (1987; 1991, Chapter 6)). Also any marginal process of y_t can be written in the form $F y_t$ by a suitable choice of F .

An important result regarding linear transformations of VARMA processes is that they are again VARMA processes. More precisely, if y_t has a VARMA representation, the same is true for z_t . Hence, the VARMA class of processes is closed with respect to linear transformations (see Lütkepohl (1984a, 1987)). In general, the autoregressive and moving average orders or Kronecker indices associated with z_t will be different from those of y_t , however. Furthermore, if y_t is a finite order VAR(p) process, z_t will in general not have a finite order VAR representation but will be of a mixed VARMA type. In particular, if the individual component series of a VAR process y_t are considered, they will usually not have a finite order AR representation. Therefore, for internal consistency in the modeling process it is sometimes useful to consider the more general VARMA class of processes rather than pure VAR models. In practice, pure VAR processes are often used because they are usually more easily dealt with in terms of inference than mixed VARMA processes as we will see in the following sections. More discussion of the structure of linearly transformed VARMA processes may be found in Lütkepohl (1987).

3 Estimation

Although model specification precedes estimation I discuss the latter first because estimates are needed in the specification procedures considered later.

3.1 Maximum Likelihood Estimation

If the distribution of u_t is Gaussian the log-likelihood function of the model (2.4) for a given multiple time series y_1, \dots, y_T is

$$l(\theta) = \sum_{t=1}^T l_t(\theta), \quad (3.1)$$

where θ represents the vector of all parameters to be estimated and

$$l_t(\theta) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma_u - \frac{1}{2} u_t' \Sigma_u^{-1} u_t$$

with

$$u_t = M_0^{-1}(A_0 y_t - A_1 y_{t-1} - \dots - A_p y_{t-p} - M_1 u_{t-1} - \dots - M_q u_{t-p}).$$

In general, maximization of $l(\theta)$ is a nonlinear optimization problem which is complicated by the fact that inequality constraints have to be observed to ensure invertibility of the MA operator. Iterative optimization algorithms have to be used for the maximization. Suitable start-up values may be obtained from one of the algorithms considered in the following.

If the DGP is stable and invertible and the parameters are identified, the ML estimators have standard limiting properties, that is,

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

where \xrightarrow{d} signifies convergence in distribution and $\Sigma_{\hat{\theta}}$ is the inverse asymptotic information matrix. On the other hand, if the variables are cointegrated, the estimators of the long-run and short-run parameters have different convergence rates. In special case models simplifications are possible. Some of them will be considered in the following.

3.2 VAR Models

Because estimation of some special case models is computationally particularly easy, these cases will be considered in more detail in the following. We begin with the levels VAR representation (2.1) under the condition that no parameter restrictions are imposed. Then estimation of the unrestricted VECM (2.3) is treated and VECMs with parameter restrictions are discussed.

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

$$y_t = (Y_{t-1}^{t-p'} \otimes I_K) a + u_t, \quad (3.2)$$

where $Y_{t-1}^{t-p'} = (y'_{t-1}, \dots, y'_{t-p})$. Under standard assumptions, the resulting estimator \hat{a} is consistent and asymptotically normally distributed (see, e.g., Lütkepohl (1991)),

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

or, written in an alternative way,

$$\hat{a} \stackrel{d}{\sim} N(a, \Sigma_{\hat{a}}/T). \quad (3.3)$$

Here vec denotes the column stacking operator which stacks the columns of a matrix in a column vector. The covariance matrix of the asymptotic distribution is

$$\Sigma_{\hat{a}} = \text{plim} \left(T^{-1} \sum_{t=1}^T Y_{t-1}^{t-p} Y_{t-1}^{t-p'} \right)^{-1} \otimes \Sigma_u. \quad (3.4)$$

Although these results also hold for cointegrated systems it is important to note that in this case the covariance matrix $\Sigma_{\hat{a}}$ is singular whereas it is nonsingular in the usual $I(0)$ case (see Park & Phillips (1988, 1989), Sims, Stock & Watson (1990), Lütkepohl (1991, Chapter 11)). In fact, some estimated coefficients or linear combinations of coefficients converge with a faster rate than $T^{1/2}$ if there are integrated or cointegrated variables. Therefore, in this case the usual t -, χ^2 - and F -tests used for inference regarding the VAR parameters, may not be valid as shown, e.g., by Toda & Phillips (1993).

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

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

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

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

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

$$\hat{\beta} = [v_1, \dots, v_r] \quad \text{and} \quad \hat{\alpha} = S_{01} \hat{\beta} (\hat{\beta}' S_{11} \hat{\beta})^{-1}, \quad (3.6)$$

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

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

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

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

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

In this approach the parameter estimator $\hat{\beta}$ is made unique by normalizing the eigenvectors from the eigenvalue problem (3.5) and $\hat{\alpha}$ is adjusted accordingly. However, these are not econometric identification restrictions. Without such restrictions only the product $\alpha \beta' = \Pi$ can be estimated consistently. For consistent estimation of the matrices α and β , identifying restrictions have to be imposed. For example, in a specific model it may be reasonable to assume that the first part of β is an identity matrix, so that $\beta' = [I_r : \beta_1']$, where β_1 is a $((K - r) \times r)$ matrix. For $r = 1$, this restriction amounts to normalizing the coefficient of the first variable. This identifying restriction has attracted some attention in the cointegration literature. If uniqueness restrictions are imposed it can be shown that $T(\hat{\beta} - \beta)$ and $\sqrt{T}(\hat{\alpha} - \alpha)$ converge in distribution (Johansen (1995)). In other words, the estimator of β converges with the fast rate T . It is therefore sometimes called *superconsistent* whereas the estimator of α converges with the usual rate \sqrt{T} .

The estimators of Γ and Π are consistent and asymptotically normal under general assumptions and converge at the usual \sqrt{T} rate, $\sqrt{T} \text{vec}(\hat{\Gamma} - \Gamma) \xrightarrow{d} N(0, \Sigma_{\hat{\Gamma}})$ and $\sqrt{T} \text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} N(0, \Sigma_{\hat{\Pi}})$. The asymptotic distribution of $\hat{\Gamma}$ is nonsingular and, hence, standard inference may be used for Γ . In contrast, the $(K^2 \times K^2)$ covariance matrix $\Sigma_{\hat{\Pi}}$ has rank Kr . It is singular if $r < K$. This result is obtained because Π involves the cointegrating relations which are estimated superconsistently.

Interestingly, if an estimator of the levels parameters A is computed via the estimates of Π and Γ and thereby satisfies the cointegration restriction, that estimator has the same asymptotic distribution as in (3.3) where no restrictions have been imposed in estimating A . Important results on estimating models with integrated variables are due to Phillips and his co-workers (e.g., Phillips & Durlauf (1986), Phillips (1987, 1991)).

In practice it is often desirable to place restrictions on the parameters to reduce the dimensionality of the parameter space. For instance, it is quite common that different lags of the differenced variables appear in the individual equations. In other words, there may be zero restrictions on the short-run parameters Γ . Moreover, some of the cointegrating relations may be confined to specific equations by imposing zero constraints on the loading matrix α . Efficient estimation of a model with parameter restrictions is more complicated than in the unrestricted case because LS is no longer identical to GLS and ML in general. A possible estimation procedure estimates β in a first stage, ignoring the restrictions on the short-run parameters. Let the estimator be $\hat{\beta}$. Because the estimators of the cointegrating parameters converge at a better rate than the estimators of the short-run parameters the former may be treated as fixed in a second stage of the estimation procedure. In other words, a systems estimation procedure may be applied to

$$\Delta y_t = (y_{t-1}' \hat{\beta} \otimes I_K) \text{vec}(\alpha) + (\Delta Y_{t-1}^{t-p+1} \otimes I_K) \text{vec}(\Gamma) + \tilde{u}_t. \quad (3.7)$$

If only exclusion restrictions are imposed on the parameter vectors $\text{vec}(\alpha)$ and $\text{vec}(\Gamma)$ in this form, standard GLS or similar methods may be applied. They result in estimators of the short-run parameters with the usual asymptotic properties. Feasible GLS estimation of more general VARMA models is discussed next.

3.3 Feasible GLS Estimation of VARMA Models

For given Kronecker indices, an ARMA_E model can be estimated even if the cointegrating rank is unknown. The Kronecker indices are assumed to be known in the following. Strategies for specifying them will be described in Section 4.

Let $a = S_a \text{vec}(A_1 : \dots : A_p)$, $\lambda = S_\lambda \text{vec}(A_0 - I_K)$ and $m = S_m \text{vec}(M_1 : \dots : M_p)$ denote the freely varying elements in the coefficient matrices. The selection matrices S_a , S_λ and S_m are defined such that they have a one in the column corresponding to the unrestricted coefficient being chosen. Then $\theta = (\nu' : a' : \lambda' : m')'$ contains the freely varying coefficients of $[A(z) : M(z)]$ not restricted to be zero or one and we can rewrite the system (2.4) in the following compact regression model form:

$$y_t = R_t' \theta + u_t, \quad (3.8)$$

where

$$R_t = \begin{bmatrix} S_a[Y_{t-1}^{t-p} \otimes I_K] \\ S_\lambda[(y_t - u_t) \otimes I_K] \\ S_m[U_{t-1}^{t-p} \otimes I_K] \end{bmatrix},$$

and $U_{t-1}^{t-p} = (u'_{t-1}, \dots, u'_{t-p})'$. Clearly, since R_t involves unknown residuals, estimating the parameters θ directly by regression from (3.8) is not feasible. Therefore, in a first step an unrestricted long VAR model of order h_T , say, is fitted by LS as in (3.2). Denoting the residuals by \hat{u}_t , we let \hat{R}_t be defined as R_t except that the residuals \hat{u}_t are substituted for the unknown u_t 's. Moreover, let $\hat{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$ be the corresponding estimator of the white noise covariance matrix Σ_u . Then a feasible GLS estimator of θ is given by

$$\hat{\theta} = \left(\sum_{t=1}^T \hat{R}_t \hat{\Sigma}_u^{-1} \hat{R}_t' \right)^{-1} \sum_{t=1}^T \hat{R}_t \hat{\Sigma}_u^{-1} y_t. \quad (3.9)$$

Here the choice of h_T should be such that significant residual autocorrelation is largely eliminated in the empirical model. Poskitt & Lütkepohl (1995) show that the estimator $\hat{\theta}$ is strongly consistent under suitable conditions for y_t and h_T .

If the cointegrating rank is known or has been determined by some procedure and one wishes to take it into account in the estimation procedure, it is preferable to estimate the EC-ARMA_E form (2.10). This may be done by first estimating θ in the ARMA_E form as in (3.9). The estimator $\hat{\theta}$ implies, of course, estimators \hat{A}_i of the A_i . It is natural to consider estimating the error correction specification in (2.10) via its relation to the A_i . Let $\hat{\Pi} = -(\hat{A}_0 - \hat{A}_1 - \dots - \hat{A}_p)$ and $\hat{\Gamma}_i = -(\hat{A}_{i+1} + \dots + \hat{A}_p)$ ($i = 1, \dots, p-1$). These estimators are also strongly consistent, of course. If identifying restrictions for α and β are available, we may then obtain estimators for these parameters via their relation to $\Pi = \alpha\beta'$. For instance, if β' has the form $\beta' = [I_r : \beta'_1]$, we get estimators

$$\hat{\alpha} = \hat{\Pi}_1 \quad \text{and} \quad \hat{\beta}'_1 = (\hat{\alpha}'[\hat{\Psi}\hat{\Sigma}_u\hat{\Psi}']^{-1}\hat{\alpha})^{-1}(\hat{\alpha}'[\hat{\Psi}\hat{\Sigma}_u\hat{\Psi}']^{-1}\hat{\Pi}_2), \quad (3.10)$$

where $[\hat{\Pi}_1 : \hat{\Pi}_2]$ denotes a partition of $\hat{\Pi}$ into submatrices of dimension $(K \times r)$ and $(K \times (K - r))$, respectively, and $\hat{\Psi} = \hat{A}_0 + \hat{M}_1 + \dots + \hat{M}_p$. Again these estimators are strongly consistent. Alternatively, if no specific identifying restrictions for β are available, we may estimate this matrix as in the Johansen procedure in (3.6) on the basis of a long VAR(h_T). In the next step this estimator may be fixed and substituted in (2.10). The remaining parameters of the EC-ARMA_E form may then be estimated by a feasible GLS procedure similar to the one described in the foregoing.

These estimators may be used as starting values for a Gaussian maximum likelihood or pseudo maximum likelihood procedure. Using a scoring algorithm, for example, will result in asymptotically fully efficient estimators after one iteration. In the scoring iterations the estimator $\hat{\beta}$ may be fixed or iterated together with the other parameters because this estimator has a higher rate of convergence than the estimators of the autoregressive and moving average parameters if identifying restrictions are available for β . In this case, the other parameter estimators have an asymptotic normal distribution which is the same as if β were known. This, of course, is analogous to the pure VAR case considered in the previous chapter (see also Phillips (1991), Yap & Reinsel (1995) or Lütkepohl & Claessen (1997)).

As mentioned earlier, before a model can be estimated, the Kronecker indices and possibly the cointegrating rank have to be specified. How to do that is discussed in the next section.

4 Model Specification

4.1 Specification of the Kronecker Indices

For stationary processes a number of proposals have been made for specifying the Kronecker indices of an ARMA_E model. For example, Hannan & Kavalieris (1984), Poskitt (1992), and Nsiri & Roy (1992) are important contributions where practical specification and analysis tools for stationary processes are introduced. In Lütkepohl & Poskitt (1996) several specification strategies are surveyed and extensions to integrated and cointegrated processes are considered by Lütkepohl & Claessen (1997), Claessen (1995) and Poskitt & Lütkepohl (1995). The strategies for specifying the Kronecker indices of cointegrated ARMA_E processes presented in this section are proposed in the latter paper where they are also shown to result in consistent estimators of the Kronecker indices under suitable conditions. In a simulation study, Bartel & Lütkepohl (1998) found that they work reasonably well in small samples at least for the processes explored in the Monte Carlo study.

The specification procedures may be partitioned in two stages. The first stage is the same in the procedures considered here. It consists of fitting a long VAR of order h_T , say, by least squares in order to obtain estimates of the unobservable innovations u_t as in the estimation procedure of the previous section. In a second stage the residuals are then substituted for the unknown lagged u_t in the individual equations of an ARMA_E form which may be estimated by linear LS procedures. Based on the equations estimated in this way, a choice of the Kronecker indices is made using model selection criteria. Poskitt & Lütkepohl (1995), Guo, Huang & Hannan (1990) and Huang & Guo (1990) show that the estimation residuals \hat{u}_t are “good” estimates of the true residuals if the VAR order h_T approaches infinity at a suitable rate as T goes to infinity.

The methods differ in the way they choose the Kronecker indices in the next step. An obvious idea is to search over all models associated with Kronecker indices which are smaller

than some prespecified upper bound p_{\max} , $\{(p_1, \dots, p_K) | 0 \leq p_k \leq p_{\max}, k = 1, \dots, K\}$, and choose the set of Kronecker indices which optimizes some model selection criterion. Unfortunately, this procedure is extremely computer intensive for systems of moderate or large dimension. Therefore, procedures have been proposed which are computationally more efficient. One possibility is to use linear regressions to estimate the individual equations separately for different lag lengths. A choice of the optimal lag length is then based on some prespecified model selection criterion which includes the residual variance as a measure of goodness of fit. For example, a criterion of the general form

$$\Lambda_{k,T}(n) = \log \hat{\sigma}_{k,T}^2(n) + C_T n/T, \quad n = 0, 1, \dots, P_T,$$

may be used, where C_T is a suitable function of the sample size T and $T\hat{\sigma}_{k,T}^2(n)$ is the residual sum of squares from a regression of y_{kt} on $(y_{jt} - \hat{u}_{jt})$ ($j = 1, \dots, K, j \neq k$) and y_{t-s} and \hat{u}_{t-s} ($s = 1, \dots, n$). Here the maximum lag length P_T may depend on the sample size.

Because each equation is treated separately, restrictions from the echelon structure are not explicitly taken into account in this procedure. Instead for each equation it is implicitly assumed that the current index under consideration is the smallest and thus no restrictions are imported from other equations. Still, the k -th equation will be misspecified whenever the lag order is less than the true Kronecker index because in that case lagged values required for a correct specification are omitted. On the other hand, if the lag order is greater than the true Kronecker index, the k -th equation will be correctly specified but may include redundant parameters and variables. Therefore, for an appropriate choice of C_T , the criterion function $\Lambda_{k,T}(n)$ will possess a global minimum asymptotically when n is equal to the true Kronecker index. For practical purposes, possible choices of C_T are $C_T = h_T \log T$ or $C_T = h_T^2$. More details on the procedure may be found in the aforementioned articles.

Poskitt & Lütkepohl (1995) also propose a modification of this procedure which permits to take into account coefficient restrictions derived from those equations in the system which have smaller Kronecker indices. In that modification, after specifying the Kronecker indices with the previous procedure, the smallest Kronecker index is fixed and the procedure is repeated for the remaining equations. In this second application of the procedure the restrictions implied by the smallest Kronecker index found in the first round are taken into account when the second smallest index is determined. We proceed in this way by fixing the smallest Kronecker index found in each successive round until all the Kronecker indices have been specified. The variables are ordered in such a way that the Kronecker indices of the final system are ordered from largest to smallest. That is, the variable whose equation is associated with the smallest Kronecker index is placed last in the list of variables. The one with the second smallest Kronecker index is assigned the next to the last place and so on.

Poskitt & Lütkepohl (1995) show that for a suitable choice of C_T the procedure results in consistent estimators of the Kronecker indices. In this version of the specification procedure the coefficient restrictions derived from the echelon form are directly incorporated into the identification stage. The idea is that this may result in a superior performance of the selection procedure. Bartel & Lütkepohl (1998) found in their Monte Carlo study that this is indeed the case. On the other hand, the computational burden is increased substantially relative to a procedure which treats all equations independently. Therefore the procedure may be problematic for high dimensional systems.

It should be understood that the Kronecker indices found in such a procedure for a given time series of finite length can only be expected to be a reasonable starting point for a more refined analysis of the system under consideration. Based on the specified Kronecker

indices a more efficient procedure for estimating the parameters may be applied as discussed in the previous section and the model may be modified subsequently. Before we consider possibilities for model checking, we will discuss procedures for specifying the cointegrating rank.

4.2 Specification of the Cointegrating Rank

For pure VAR processes, Johansen (1995) provides LR tests for specifying the cointegrating rank. Under Gaussian assumptions, ML estimation of unrestricted VECMs with a specific cointegrating rank r was found to be relatively easy in Sec. 3.2. Therefore the LR statistic for a pair of hypotheses $H_0 : r = r_0$ versus $H_1 : r > r_0$ is easily determined by evaluating the maxima of the likelihood functions for $r = r_0$ and for $r = K$. In fact, it can be shown that the LR statistic has the form

$$LR(r_0) = -T \sum_{j=r_0+1}^K \log(1 - \lambda_j), \quad (4.1)$$

where the λ_j are the eigenvalues from (3.5). The asymptotic distribution of the LR statistic is nonstandard but has been tabulated in Johansen (1995, Chapter 15), for example. It depends on the deterministic terms in the DGP such as intercept and trend terms. Hence, in this case it makes a difference which deterministic terms are included.

Saikkonen & Luukkonen (1997) show that Johansen's LR tests can be justified even if a finite order VAR process is fitted although the true underlying process has an infinite order VAR structure, provided the VAR order goes to infinity with the sample size. Consequently, these tests may be applied at the beginning of the specification procedure even if the true DGP is a mixed VARMA process. Lütkepohl & Saikkonen (1999) discuss the choice of the VAR order in this case. Alternatively, Yap & Reinsel (1995) have extended the likelihood ratio principle to VARMA processes and develop tests for the cointegrating rank under the assumption that the lag orders of $A(z)$ and $M(z)$ are known. Thus, these tests may be applied once the Kronecker indices have been identified. Whatever approach is adopted, for our purposes a modification is useful. It is known from (2.11) that the cointegrating rank $r \geq \varrho$, the number of zero Kronecker indices. Hence, only null hypotheses should be considered where r is greater than or equal to ϱ so that the sequence of null hypotheses $H_0 : r = \varrho, H_0 : r = \varrho + 1, \dots, H_0 : r = K - 1$, is tested. The estimator of r is chosen as the smallest value for which H_0 cannot be rejected.

4.3 Model Checking

Once a model has been specified and estimated by some efficient procedure, some checks for model adequacy are in order and possible further model reductions or modifications may be called for. For instance, insignificant parameter estimates may be restricted to zero. Here it is convenient that the t -ratios of the short-run and loading parameters have their usual asymptotic standard normal distributions under the null hypothesis due to the limiting normal distribution of the ML estimators. In this way overidentifying restrictions may be imposed.

A number of model checking tools are based on the residuals of the final model. Some of them are applied to the residuals of individual equations and others are based on the full residual vectors. Examples of specification checking tools are visual inspection of the plots

of the residuals and their autocorrelations. In addition, autocorrelations of squared residuals may be considered to check for possible autoregressive conditional heteroscedasticity (ARCH). Although it may be quite insightful to inspect the autocorrelations visually, formal statistical tests for remaining residual autocorrelation should also be applied. Such tests are often based on LM (Lagrange Multiplier) or Portmanteau statistics. Moreover, normality tests of the Lomnicki-Jarque-Bera type may be applied to the residuals (see, e.g., Lütkepohl (1991), Doornik & Hendry (1997)).

There are also procedures for checking of a model for potential structural shifts during the sample period. For example, prediction tests as discussed in Lütkepohl (1991) may be performed or recursive residuals (Doornik & Hendry (1997)) may be inspected. In addition out-of-sample forecasts are sometimes used for model checking when new data become available. For a more detailed discussion of model checking see Hendry (1995) or Doornik & Hendry (1997). If model defects are detected at the checking stage efforts have to be made to find a better representation of the DGP by adding other variables or lags to the model, by modifying the sampling period, considering nonlinear terms etc..

5 Forecasting

In discussing forecasts based on VARMA processes it is again instructive to begin with pure VAR models. An extension to mixed VARMA processes will then be relatively simple. I will first consider forecasts based on known processes and then discuss the consequences of using models with estimated rather than known parameters. Finally, in Subsection 5.4 forecasting linearly transformed processes will be considered.

5.1 VAR Processes

The VAR form (2.1) is particularly easy to use in forecasting the variables y_t . The optimal, minimum mean squared error (MSE) 1-step forecast in period T is the conditional expectation

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

where the latter equality holds if u_t is independent white noise, that is, u_t and u_s are independent random vectors for $s \neq t$. In that case, $E(u_{T+1}|y_T, y_{T-1}, \dots) = 0$. Forecasts for larger horizons $h > 1$ may be obtained recursively as

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \dots + A_p y_{T+h-p|T}, \quad (5.2)$$

where $y_{T+j|T} = y_{T+j}$ for $j \leq 0$. The corresponding forecast errors are $y_{T+1} - y_{T+1|T} = u_{T+1}$, $y_{T+2} - y_{T+2|T} = u_{T+2} + A_1 u_{T+1}$ and, more generally,

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \dots + \Phi_{h-1} u_{T+1}, \quad (5.3)$$

where it is easy to see by successive substitution that the weight matrices in the latter equation may be obtained as

$$\Phi_s = \sum_{j=1}^s \Phi_{s-j} A_j, \quad s = 1, 2, \dots, \quad (5.4)$$

with $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$ (see Lütkepohl (1991, Sec. 11.3)). Obviously, the forecasts are unbiased, that is, the forecast errors have expectation 0.

As mentioned earlier, these are the minimum MSE forecasts. The MSE matrix of an h -step forecast is

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

For any other h -step forecast with MSE matrix $\Sigma_y^*(h)$, say, the difference $\Sigma_y^*(h) - \Sigma_y(h)$ is a positive semidefinite matrix. This result relies on the assumption that u_t is independent white noise. If u_t is just uncorrelated white noise so that u_t and u_s are uncorrelated for $s \neq t$ and not necessarily independent, the forecasts computed by the recursions in (5.2) are just best linear forecasts (see Lütkepohl (1991, Sec. 2.2.2) for an illustrative example).

It may be worth pointing out that the forecast MSEs for integrated processes are generally unbounded as the horizon h goes to infinity. Thus the forecast uncertainty increases without bounds for forecasts of the distant future. This contrasts with the case of stationary $I(0)$ variables for which the forecast MSEs are bounded by the unconditional covariance of y_t . Because cointegration relations are $I(0)$, this means, in particular, that forecasts of cointegration relations have bounded MSEs even for horizons approaching infinity. For a more detailed discussion of forecasting cointegrated processes see Clements & Hendry (1995).

Of course, this result is also reflected in the corresponding forecast intervals. Assuming that the process y_t is Gaussian and, hence, $u_t \sim \text{iid } N(0, \Sigma_u)$, the forecast errors are also multivariate normal. This result may be used to set up forecast intervals of the form

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

where $c_{1-\gamma/2}$ is the $(1 - \frac{\gamma}{2})100$ percentage point of the standard normal distribution, $y_{k,T+h|T}$ denotes the k -th component of $y_{T+h|T}$ and $\sigma_k(h)$ denotes the square root of the k -th diagonal element of $\Sigma_y(h)$, that is, $\sigma_k(h)$ is the standard deviation of the h -step forecast error for the k -th component of y_t . I will now turn to forecasts based on VARMA processes still assuming that the parameters are known.

5.2 VARMA Processes

Forecasts of the variables of the VARMA process (2.4)/(2.5) are obtained easily from the pure VAR form (2.7). Assuming again independent white noise, an optimal 1-step forecast at forecast origin T is

$$y_{T+1|T} = \sum_{i=1}^T \Xi_i y_{T+1-i}. \quad (5.7)$$

More generally, optimal h -step forecasts may be computed recursively for $h = 1, 2, \dots$, as

$$y_{T+h|T} = \sum_{i=1}^{T+h-1} \Xi_i y_{T+h-i|T}. \quad (5.8)$$

Because for given y_1, \dots, y_T the u_1, \dots, u_T may be computed from (2.7), the 1-step forecast may be obtained alternatively as

$$y_{T+1|T} = A_0^{-1}(A_1 y_T + \dots + A_p y_{T+1-p}) + A_0^{-1}(M_1 u_T + \dots + M_p u_{T+1-p}),$$

where $M_0 = A_0$ is assumed as in the echelon form. Again more generally, the h -step forecasts may be determined as

$$y_{T+h|T} = A_0^{-1}(A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T}) + A_0^{-1} \sum_{i=h}^p M_i u_{T+h-i}, \quad (5.9)$$

where, as usual, the sum vanishes if $h > p$.

Both ways to compute h -step forecasts from VARMA models are relatively computer intensive especially for long time series. Moreover, they rely on our initial value assumption which states that $y_t = 0$ for $t \leq 0$. If such an assumption is not made the true error terms cannot be computed and also (5.7) and (5.8) are only approximations. In that case, precise formulas based on y_1, \dots, y_T may be obtained via the so-called *Multivariate Innovations Algorithm* of Brockwell & Davis (1987, §11.4).

Under our assumptions, the properties of the forecast errors are easily derived by expressing the process (2.5) as

$$y_t = u_t + \sum_{i=1}^{t-1} \Phi_i u_{t-i}, \quad (5.10)$$

where

$$\Phi(z) = I_K + \sum_{i=1}^{\infty} \Phi_i z^i = A(z)^{-1} M(z). \quad (5.11)$$

Note that in general, if there are unit roots, for $A(z)^{-1}$ to exist, z must be strictly within the complex unit circle. Nevertheless, we can, of course, get the coefficient matrices Φ_i from (5.11). In terms of the representation (5.10) the optimal h -step forecast may be expressed as

$$y_{T+h|T} = \sum_{i=h}^{T+h-1} \Phi_i u_{T+h-i}. \quad (5.12)$$

Hence, the forecast errors are seen to be

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1} \quad (5.13)$$

just as in the pure VAR case in (5.3). Thus, the MSE or forecast error covariance matrix is also obtained in the same way as in that section, that is, $\Sigma_y(h)$ has precisely the form given in (5.5). Moreover, forecast intervals etc. may be derived from these results in the familiar way under Gaussian assumptions.

5.3 Forecasting Estimated Processes

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

$$\hat{y}_{T+h|T} = \hat{A}_0^{-1}(\hat{A}_1 \hat{y}_{T+h-1|T} + \cdots + \hat{A}_p \hat{y}_{T+h-p|T}) + \hat{A}_0^{-1} \sum_{i=h}^p \hat{M}_i \hat{u}_{T+h-i}, \quad h = 1, 2, \dots, \quad (5.14)$$

where, of course, $\hat{y}_{T+j|T} = y_{T+j}$ for $j \leq 0$ and the \hat{u}_{T+h-i} are estimation residuals. The corresponding forecast error is

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

If T marks the end of the sample period used for estimation and is at the same time the forecast origin, then the first term on the right-hand side of the foregoing expression consists of future residuals only whereas the second term involves present and past variables only. Hence, assuming that u_t is independent white noise, the two terms are independent. Moreover, under standard assumptions, the difference $y_{T+h|T} - \hat{y}_{T+h|T}$ is small in probability as T gets large. Consequently, the forecast error covariance matrix is

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

Here $o(1)$ denotes a term which approaches zero as the sample size tends to infinity. Thus, for large samples the estimation uncertainty may be ignored in evaluating the forecast precision and setting up forecast intervals. On the other hand, in small samples the forecast precision will depend on the quality of the parameter estimators. Hence, if precise forecasts are desired, good parameter estimators are needed. A more precise form of the difference between $\Sigma_{\hat{y}}(h)$ and $\Sigma_y(h)$ may be found, e.g., in Lütkepohl (1991) and a more complete discussion of forecasting estimated cointegrated processes with further references is given in Clements & Hendry (1995).

5.4 Forecasting Linearly Transformed and Aggregated Processes

As mentioned in Sec. 2.3.4, linearly transformed VARMA processes are again VARMA processes. Hence, if y_t has a VARMA representation, the same is true for $z_t = Fy_t$, where F is some fixed $(M \times K)$ transformation matrix. The VAR and MA orders can change considerably, however. Nevertheless, it can be shown that if forecasts of z_t are desired, it is generally preferable to forecast the untransformed process y_t first and then transform the forecasts rather than forecast z_t directly. More precisely, it can be shown that $Fy_{T+h|T}$ is generally a better forecast of z_{T+h} than $\hat{z}_{T+h|T}$ in the sense that $\Sigma_z(h) - F\Sigma_y(h)F'$ is positive semidefinite (see Lütkepohl (1984a, 1987)). Here $F\Sigma_y(h)F'$ is the MSE matrix associated with $Fy_{T+h|T}$, of course. This result implies that if forecasts of an aggregate are of interest, then it is generally useful to consider disaggregated series, forecast them in a multivariate model and aggregate the forecasts. Because temporal aggregation can also be viewed as a linear transformation of a process observed at a higher frequency this result also means that if forecasts of quarterly variables, say, are of interest forecasting monthly values and aggregating the monthly forecasts to obtain a quarterly forecast is preferable to forecasting the quarterly process directly (see also Lütkepohl (1984b, 1986a, 1986b) for examples).

To illustrate these issues consider the following example from Lütkepohl (1991). Suppose y_t is a bivariate VAR(1) process,

$$\begin{bmatrix} 1 - 0.5L & 0.66L \\ 0.5L & 1 + 0.3L \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \quad \text{with} \quad \Sigma_u = I_2. \quad (5.15)$$

Furthermore, suppose that we are interested in the aggregate $z_t = y_{1t} + y_{2t}$. It can be shown that z_t has the representation

$$(1 - 0.2L - 0.48L^2)z_t = (1 - 0.504L)v_t \quad \text{with} \quad \sigma_v^2 = 2.70.$$

Obviously, the 1-step forecast error variance of $z_{T+1|T}$ is $\Sigma_z(1) = \sigma_v^2 = 2.70$ whereas $\Sigma_y(1) = \Sigma_u = I_2$ and, hence, for $F = [1, 1]$, $F\Sigma_y(1)F' = 2$.

Unfortunately, in general aggregated forecasts are not necessarily better than forecasting the aggregate directly if the true process is unknown and the parameters have to be estimated, for example. More details and examples illustrating this case may be found in Lütkepohl (1984b, 1986a) and other references. Notice also that forecasting the disaggregate series individually using univariate time series methods and then aggregating them may result in a forecast which is inferior to a direct forecast of the aggregate.

6 Conclusions and Extensions

In this chapter the use of VARMA processes for forecasting systems of variables has been considered. In this procedure the usual steps of model specification, estimation and diagnostic checking have to be done first. Once a well specified and efficiently estimated model has been established it can be used for forecasting in a straightforward manner. Procedures are presented that may be used in the different steps of the analysis. It is assumed that all variables are at most $I(1)$ so that stochastic trends can be removed by differencing once. Possible cointegration between the variables is accommodated by including an error correction term which makes it particularly easy to analyze the cointegration relations. Moreover, it is argued that the echelon form offers a framework for a unique parameterization of the VARMA structure which may be combined conveniently with the error correction form.

It should be noted, however, that other possibilities exist for unique parameterizations of VARMA models. For instance, the scalar component models of Tiao & Tsay (1989) and Tsay (1989) have received some attention in the literature. Moreover, early attempts of econometric VARMA modeling have focussed on so-called final equations forms (Zellner & Palm (1974), Wallis (1977)). These approaches may be useful occasionally but have not found wide-spread use in applications.

Throughout I have focussed on models without deterministic terms. Adding mean terms, seasonal dummy variables or polynomial trend terms is a straightforward extension for most of the models and procedures presented in the foregoing. Some adjustments are necessary in some of the procedures, however. In particular, deterministic terms lead to changes in the asymptotic distributions of the cointegration tests. More generally, there may be additional unmodeled or exogenous variables. In that case the VARMAX class of models considered by Hannan & Deistler (1988), for example, may offer an appropriate framework.

Another limitation of the previous analysis is the exclusion of variables with higher order integration. In practice, some variables such as price indices may be modeled more appropriately as $I(2)$ variables. If variables of higher order integration are present, the analysis and modeling of the cointegration relations becomes more difficult. Except for that problem, forecasts can be set up as discussed in Sec. 5 on the basis of the levels version if that is available. A proper analysis of the cointegration properties may be very important for longer term forecasts, however.

References

- Bartel, H. & H. Lütkepohl (1998), Estimating the Kronecker indices of cointegrated echelon form VARMA models, *Econometrics Journal*, 1 C76-C99.
- Brockwell, P.J. & R.A. Davis (1987), *Time Series: Theory and Methods*, New York:

Springer-Verlag.

- Claessen, H. (1995), *Spezifikation und Schätzung von VARMA-Prozessen unter besonderer Berücksichtigung der Echelon-Form*. Bergisch Gladbach: Verlag Josef Eul.
- Clements, M.P. & D.F. Hendry (1995), Forecasting in cointegrated systems, *Journal of Applied Econometrics*, 10, 127-146.
- Doornik, J.A. & D.F. Hendry (1997), *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, London: International Thomson Business Press.
- Engle, R.F. & C.W.J. Granger (1987), Co-integration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.
- Guo, L., D.W. Huang & E.J. Hannan (1990), On $ARX(\infty)$ approximation, *Journal of Multivariate Analysis*, 32, 17-47.
- Hannan, E.J. & M. Deistler (1988), *The Statistical Theory of Linear Systems*. New York: Wiley.
- Hannan, E.J. & L. Kavalieris (1984), Multivariate time series models, *Advances in Applied Probability*, 16, 492 - 561.
- Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.
- Huang, D. & L. Guo (1990), Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method, *Annals of Statistics*, 18, 1729 - 1756.
- Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Lütkepohl, H. (1984a), Linear transformations of vector ARMA processes, *Journal of Econometrics*, 26, 283-293.
- Lütkepohl, H. (1984b), Forecasting contemporaneously aggregated vector ARMA processes, *Journal of Business & Economic Statistics*, 2, 201-214.
- Lütkepohl, H. (1986a), Forecasting vector ARMA processes with systematically missing observations, *Journal of Business & Economic Statistics*, 4, 375-390.
- Lütkepohl, H. (1986b), Forecasting temporally aggregated vector ARMA processes, *Journal of Forecasting*, 5, 85-95.
- Lütkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, Berlin: Springer-Verlag.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.
- Lütkepohl, H. (1996), *Handbook of Matrices*, Chichester: Wiley.
- Lütkepohl, H. & H. Claessen (1997), Analysis of cointegrated VARMA processes, *Journal of Econometrics*, 80, 223-239.

- Lütkepohl, H. & D.S. Poskitt (1996), Specification of echelon form VARMA models, *Journal of Business & Economic Statistics*, 14, 69 - 79.
- Lütkepohl, H. & P. Saikkonen (1999), Order selection in testing for the cointegrating rank of a VAR process, forthcoming in Granger festschrift.
- Nsiri, S. & R. Roy (1992), On the identification of ARMA echelon-form models, *Canadian Journal of Statistics*, 20, 369 - 386.
- Park, J.Y. & P.C.B. Phillips (1988), Statistical inference in regressions with integrated processes: Part 1, *Econometric Theory*, 4, 468-497.
- Park, J.Y. & P.C.B. Phillips (1989), Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory*, 5, 95-131.
- Phillips, P.C.B. (1987), Time series regression with a unit root, *Econometrica*, 55, 277-301.
- Phillips, P.C.B. (1991), Optimal inference in cointegrated systems, *Econometrica*, 59, 283-306.
- Phillips, P.C.B. & S.N. Durlauf (1986), Multiple time series regression with integrated processes, *Review of Economic Studies*, 53, 473-495.
- Poskitt, D.S. (1992), Identification of echelon canonical forms for vector linear processes using least squares, *Annals of Statistics*, 20, 195 - 215.
- Poskitt, D.S. & H. Lütkepohl (1995), Consistent specification of cointegrated autoregressive moving-average systems, Discussion Paper 54 1995, SFB 373, Humboldt-Universität zu Berlin.
- Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, 81, 93-126.
- Sims, C.A., J.H. Stock & M.W. Watson (1990), Inference in linear time series models with some unit roots, *Econometrica*, 58, 113-144.
- Tiao, G.C. & R.S. Tsay (1989), Model specification in multivariate time series (with discussion), *Journal of the Royal Statistical Society*, B51, 157-213.
- Toda, H.Y. & P.C.B. Phillips (1993), Vector autoregressions and causality, *Econometrica*, 61, 1367-1393.
- Tsay, R.S. (1989), Parsimonious parameterization of vector autoregressive moving average models, *Journal of Business & Economic Statistics*, 7, 327-341.
- Wallis, K.F. (1977), Multiple time series analysis and the final form of econometric models, *Econometrica*, 45, 1481-1497.
- Yap, S.F. & G.C. Reinsel (1995), Estimation and testing for unit roots in a partially non-stationary vector autoregressive moving average model, *Journal of the American Statistical Association*, 90, 253-267.
- Zellner, A. & F. Palm (1974), Time series analysis and simultaneous equation econometric models, *Journal of Econometrics*, 2, 17-54.