EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

EUI Working Paper **ECO** No. 2004 / 25

Forecasting with VARMA Models

HELMUT LÜTKEPOHL

BADIA FIESOLANA, SAN DOMENICO (FI)

Published in Italy in September 2004
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

*Revisions of this work can be found at: http://www.iue.it/Personal/Luetkepohl/.

August 30, 2004

Forecasting with VARMA Models

by

Helmut Lütkepohl

Department of Economics, European University Institute, Via della Piazzuola 43, I-50133 Firenze, ITALY, email: helmut.luetkepohl@iue.it

Contents

1 Introduction and Overview

In this chapter linear models for the conditional mean of a stochastic process are considered. These models are useful for producing linear forecasts of time series variables. Suppose that K related time series variables are considered, $y_{1t}, \ldots, y_{Kt}$, say. Defining $y_t = (y_{1t}, \ldots, y_{Kt})'$, a linear model for the conditional mean of the data generation process (DGP) of the observed series may be of the vector autoregressive (VAR) form,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{1.1}$$

where the A_i's $(i = 1, \ldots, p)$ are $(K \times K)$ coefficient matrices and u_t is a K-dimensional error term. If u_t is independent over time (i.e., u_t and u_s are independent for $t \neq s$), the conditional mean of y_t, given past observations, is

$$y_{t|t-1} \equiv E(y_t|y_{t-1}, y_{t-2}, \ldots) = A_1 y_{t-1} + \cdots + A_p y_{t-p}.$$

Thus, the model can be used directly for forecasting one period ahead and forecasts with larger horizons can be computed recursively. Therefore, variants of this model will be the basic forecasting models in this chapter.

For practical purposes the simple VAR model of order p may have some disadvantages, however. The A_i parameter matrices will be unknown and have to be replaced by estimators. For an adequate representation of the DGP of a set of time series of interest a rather large VAR order p may be required. Hence, a large number of parameters may be necessary for an adequate description of the data. Given limited sample information this will usually result in low estimation precision and also forecasts based on VAR processes with estimated coefficients may suffer from the uncertainty in the parameter estimators. Therefore it is useful to consider the larger model class of vector autoregressive moving-average (VARMA) models which may be able to represent the DGP of interest in a more parsimonious way. In this chapter the analysis of models from that class will be discussed although special case results for VAR processes will occasionally be noted explicitly.

The VARMA class has the further advantage of being closed with respect to linear transformations, that is, a linearly transformed finite order VARMA process has again a finite order VARMA representation. Therefore linear aggregation issues can be studied within this class. In this chapter special attention will be given to results related to forecasting contemporaneously and temporally aggregated processes.

VARMA models can be parameterized in different ways. In other words, different parameterizations describe the same stochastic process. Although this is no problem for forecasting

purposes because we just need to have one adequate representation of the DGP, nonunique parameters are a problem at the estimation stage. Therefore the *echelon form* of a VARMA process is presented as a unique representation. Estimation and specification of this model form will be considered.

These models have first been developed for stationary variables. In economics and also other fields of applications many variables are generated by nonstationary processes, however. Often they can be made stationary by considering differences or changes rather than the levels. A variable is called integrated of order d ($I(d)$) if it is still nonstationary after taking differences $d-1$ times but it can be made stationary or asymptotically stationary by differencing d times. In most of the following discussion the variables will be assumed to be stationary ($I(0)$) or integrated of order 1 ($I(1)$) and they may be cointegrated. In other words, there may be linear combinations of $I(1)$ variables which are $I(0)$. If cointegration is present, it is often advantageous to separate the cointegration relations from the short-run dynamics of the DGP. This can be done conveniently by allowing for an error correction or equilibrium correction (EC) term in the models and *EC echelon forms* will also be considered.

The model setup for stationary and integrated or cointegrated variables will be presented in the next section where also forecasting with VARMA models will be considered under the assumption that the DGP is known. In practice it is, of course, necessary to specify and estimate a model for the DGP on the basis of a given set of time series. Model specification, estimation and model checking are discussed in Section 3 and forecasting with estimated models is considered in Section 4. Conclusions follow in Section 5.

Historical Notes

The successful use of univariate ARMA models for forecasting has motivated researchers to extend the model class to the multivariate case. It is plausible to expect that using more information by including more interrelated variables in the model improves the forecast precision. This is actually the idea underlying Granger's influential definition of causality (Granger (1969)). It turned out, however, that generalizing univariate models to multivariate ones is far from trivial in the ARMA case. Early on Quenouille (1957) considered multivariate VARMA models. It became quickly apparent, however, that the specification and estimation of such models was much more difficult than for univariate ARMA models. The success of the Box-Jenkins modelling strategy for univariate ARMA models in the 1970s (Box & Jenkins (1976)) triggered further attempts of using the corresponding mul-

tivariate models and developing estimation and specification strategies. In particular, the possibility of using autocorrelations, partial autocorrelations and cross-correlations between the variables for model specification were explored. Because modelling strategies based on such quantities had been to some extent successful in the univariate Box-Jenkins approach, it was plausible to try multivariate extensions. Examples of such attempts are Tiao & Box (1981), Tiao & Tsay (1983, 1989), Tsay (1989a, b), Wallis (1977), Zellner & Palm (1974), Granger & Newbold (1977, Chapter 7), Jenkins & Alavi (1981). It became soon clear, however, that these strategies were at best promising for very small systems of two or perhaps three variables. Moreover, the most useful setup of multiple time series models was under discussion because VARMA representations are not unique or, to use econometric terminology, they are not identified. Important early discussions of the related problems are due to Hannan (1970, 1976, 1979, 1981), Dunsmuir & Hannan (1976) and Akaike (1974). A rather general solution to the structure theory for VARMA models was later presented by Hannan & Deistler (1988). Understanding the structural problems contributed to the development of complete specification strategies. By now textbook treatments of modelling, analyzing and forecasting VRAMA processes are available (Lütkepohl (1991), Reinsel (1993)).

The problems related to VARMA models were perhaps also relevant for a parallel development of pure VAR models as important tools for economic analysis and forecasting. Sims (1980) launched a general critique of classical econometric modelling and proposed VAR models as alternatives. A short while later the concept of cointegration was developed by Granger (1981) and Engle & Granger (1987). It is conveniently placed into the VAR framework as shown by Johansen (1995a). Therefore it is perhaps not surprising that VAR models dominate time series econometrics although the methodology and software for working with more general VARMA models is nowadays available. A recent previous overview of forecasting with VARMA processes is given by Lütkepohl (2002). The present review draws partly on that article and on a monograph by Lütkepohl (1987).

Notation, Terminology, Abbreviations

The following notation and terminology is used in this chapter. The *lag operator* also sometimes called *backshift operator* is denoted by L and it is defined as usual by $Ly_t \equiv y_{t-1}$. The *differencing operator* is denoted by Δ, that is, $\Delta y_t \equiv y_t - y_{t-1}$. For a random variable or random vector x, $x \sim (\mu, \Sigma)$ signifies that its mean (vector) is μ and its variance (covariance matrix) is Σ. The $(K \times K)$ identity matrix is denoted by I_K and the determinant and

trace of a matrix A are denoted by $\det A$ and $\mathrm{tr}A$, respectively. For quantities $A_1, \ldots, A_p$, $\mathrm{diag}[A_1, \ldots, A_p]$ denotes the diagonal or block-diagonal matrix with $A_1, \ldots, A_p$ on the diagonal. The natural logarithm of a real number is signified by log. The symbols $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{C}$ are used for the integers, the positive integers and the complex numbers, respectively.

DGP stands for data generation process. VAR, AR, MA, ARMA and VARMA are used as abbreviations for vector autoregressive, autoregressive, moving-average, autoregressive moving-average and vector autoregressive moving-average (process). Error correction is abbreviated as EC and VECM is short for vector error correction model. The echelon forms of VARMA and EC-VARMA processes are denoted by ARMA$_E$ and EC-ARMA$_E$, respectively. OLS, GLS, ML and RR abbreviate ordinary least squares, generalized least squares, maximum likelihood and reduced rank, respectively. LR and MSE are used to abbreviate likelihood ratio and mean squared error.

2 VARMA Processes

2.1 Stationary Processes

Suppose the DGP of the K-dimensional multiple time series, $y_1, \ldots, y_T$, is stationary, that is, its first and second moments are time invariant. It is a (finite order) VARMA process if it can be represented in the general form

$$A_0 y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + M_0 u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{2.1}$$

where $A_0, A_1, \ldots, A_p$ are $(K \times K)$ autoregressive parameter matrices while $M_0, M_1, \ldots, M_q$ are moving average parameter matrices also of dimension $(K \times K)$. Defining the VAR and MA operators, respectively, as $A(L) = A_0 - A_1 L - \cdots - A_p L^p$ and $M(L) = M_0 + M_1 L + \cdots + M_q L^q$, the model can be written in more compact notation as

$$A(L) y_t = M(L) u_t, \qquad t = 0, \pm 1, \pm 2, \ldots. \tag{2.2}$$

Here u_t is a white-noise process with zero mean, nonsingular, time-invariant covariance matrix $E(u_t u_t') = \Sigma_u$ and zero covariances, $E(u_t u_{t-h}') = 0$ for $h = \pm 1, \pm 2, \ldots$. The zero-order matrices A_0 and M_0 are assumed to be nonsingular. They will often be identical, $A_0 = M_0$, and in many cases they will be equal to the identity matrix, $A_0 = M_0 = I_K$. To indicate the orders of the VAR and MA operators, the process (2.1) is sometimes called a VARMA(p, q) process. Notice, however, that so far we have not made further assumptions

regarding the parameter matrices so that some or all of the elements of the A_i's and M_j's may be zero. In other words, there may be a VARMA representation with VAR or MA orders less than p and q, respectively. Obviously, the VAR model (1.1) is a VARMA$(p, 0)$ special case with $A_0 = I_K$ and $M(L) = I_K$. It may also be worth pointing out that there are no deterministic terms such as nonzero mean terms in our basic VARMA model (2.1). These terms are ignored here for convenience although they are important in practice. The necessary modifications for deterministic terms will be discussed in Section 2.5.

The matrix polynomials in (2.2) are assumed to satisfy

$$\det A(z) \neq 0, \ |z| \leq 1, \quad \text{and} \quad \det M(z) \neq 0, \ |z| \leq 1 \quad \text{for} \quad z \in \mathbb{C}. \tag{2.3}$$

The first of these conditions ensures that the VAR operator is *stable* and the process is stationary. Then it has a pure MA representation

$$y_t = \sum_{j=0}^{\infty} \Phi_i u_{t-i} \tag{2.4}$$

with MA operator $\Phi(L) = \Phi_0 + \sum_{i=1}^{\infty} \Phi_i L^i = A(L)^{-1} M(L)$. Notice that $\Phi_0 = I_K$ if $A_0 = M_0$ and in particular if both zero order matrices are identity matrices. In that case (2.4) is just the *Wold MA representation* of the process and, as we will see later, the u_t are just the one-step ahead forecast errors. Some of the forthcoming results are valid for more general stationary processes with Wold representation (2.4) which may not come from a finite order VARMA representation. In that case, it is assumed that the Φ_i's are absolutely summable so that the infinite sum in (2.4) is well-defined.

The second part of condition (2.3) is the usual *invertibility condition* for the MA operator which implies the existence of a pure VAR representation of the process,

$$y_t = \sum_{i=1}^{\infty} \Xi_i y_{t-i} + u_t, \tag{2.5}$$

where $A_0 = M_0$ is assumed and $\Xi(L) = I_K - \sum_{i=1}^{\infty} \Xi_i L^i = M(L)^{-1} A(L)$. Occasionally invertibility of the MA operator will not be a necessary condition. In that case, it is assumed without loss of generality that $\det M(z) \neq 0$, for $|z| < 1$. In other words, the roots of the MA operator are outside or on the unit circle. There are still no roots inside the unit circle, however. This assumption can be made without loss of generality because it can be shown that for an MA process with roots inside the complex unit circle an equivalent one exists which has all its roots outside and on the unit circle.

It may be worth noting at this stage already that every pair of operators $A(L)$, $M(L)$ which leads to the same transfer functions $\Phi(L)$ and $\Xi(L)$ defines an equivalent VARMA representation for y_t. This nonuniqueness problem of the VARMA representation will become important when parameter estimation is discussed in Section 3.

As specified in (2.1), we are assuming that the process is defined for all $t \in \mathbb{Z}$. For stable, stationary processes this assumption is convenient because it avoids considering issues related to initial conditions. Alternatively one could define y_t to be generated by a VARMA process such as (2.1) for $t \in \mathbb{N}$, and specify the initial values $y_0, \ldots, y_{-p+1}, u_0, \ldots, u_{-p+1}$ separately. Under our assumptions they can be defined such that y_t is stationary. Alternatively, one may define fixed initial values or perhaps even $y_0 = \cdots = y_{-p+1} = u_0 = \cdots = u_{-p+1} = 0$. In general, such an assumption implies that the process is not stationary but just *asymptotically stationary*, that is, the first and second order moments converge to the corresponding quantities of the stationary process obtained by specifying the initial conditions accordingly or defining y_t for $t \in \mathbb{Z}$. The issue of defining initial values properly becomes more important for the nonstationary processes discussed in Section 2.2.

Both the MA and the VAR representations of the process will be convenient to work with in particular situations. Another useful representation of a stationary VARMA process is the state space representation which will not be used in this review, however. State space representations of VARMA processes are considered, for example, by Aoki (1987), Hannan & Deistler (1988), and Wei (1990). In more general terms state space models are discussed in Chapter ??? of this Handbook.

2.2 Cointegrated $I(1)$ Processes

If the DGP is not stationary but contains some $I(1)$ variables, the levels VARMA form (2.1) is not the most convenient one for inference purposes. In that case, $\det A(z) = 0$ for $z = 1$. Therefore we write the model in EC form by subtracting $A_0 y_{t-1}$ on both sides and re-arranging terms as follows:

$$\begin{aligned} A_0 \Delta y_t &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} \\ &\quad + M_0 u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \quad t \in \mathbb{N}, \end{aligned} \tag{2.6}$$

where $\Pi = -(A_0 - A_1 - \cdots - A_p) = -A(1)$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ $(i = 1, \ldots, p-1)$ (Lütkepohl & Claessen (1997)). Here Πy_{t-1} is the EC term and $r = \mathrm{rk}(\Pi)$ is the cointegrating rank of the system which specifies the number of linearly independent cointegration relations. The process is assumed to be started at time $t = 1$ from some initial values

$y_0, \ldots, y_{-p+1}, u_0, \ldots, u_{-p+1}$ to avoid infinite moments. Thus, the initial values are now of some importance. Assuming that they are zero is convenient because in that case the process is easily seen to have a pure EC-VAR or VECM representation of the form

$$\Delta y_t = \Pi^* y_{t-1} + \sum_{j=1}^{t-1} \Theta_j \Delta y_{t-j} + A_0^{-1} M_0 u_t, \quad t \in \mathbb{N}, \tag{2.7}$$

where Π^* and Θ_j $(j = 1, 2, \ldots)$ are such that

$$I_K \Delta - \Pi^* L - \sum_{j=1}^{\infty} \Theta_j \Delta L^j = A_0^{-1} M_0 M(L)^{-1} (A_0 \Delta - \Pi L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1} \Delta L^{p-1}).$$

A similar representation can also be obtained if nonzero initial values are permitted (see Saikkonen & Lütkepohl (1996)). Furthermore, Bauer & Wagner (2003) present a state space representation which is especially suitable for cointegrated processes.

2.3 Linear Transformations of VARMA Processes

As mentioned in the introduction, a major advantage of the class of VARMA processes is that it is closed with respect to linear transformations. In other words, linear transformations of VARMA processes have again a finite order VARMA representation. These transformations are very common and are useful to study problems of aggregation, marginal processes or averages of variables generated by VARMA processes etc.. In particular, the following result from Lütkepohl (1984) is useful in this context. Let

$$y_t = u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}$$

be a K-dimensional invertible MA(q) process and let F be an $(M \times K)$ matrix of rank M. Then the M-dimensional process $z_t = F y_t$ has an invertible MA($\breve{q}$) representation with $\breve{q} \leq q$. An interesting consequence of this result is that if y_t is a stable and invertible VARMA(p, q) process as in (2.1), then the linearly transformed process $z_t = F y_t$ has a stable and invertible VARMA$(\breve{p}, \breve{q})$ representation with $\breve{p} \leq (K - M + 1)p$ and $\breve{q} \leq (K - M)p + q$ (Lütkepohl (1987, Chapter 4)).

These results are directly relevant for contemporaneous aggregation of VARMA processes and they can also be used to study temporal aggregation problems. To see this suppose we wish to aggregate the variables y_t generated by (2.1) over m subsequent periods. For instance, $m = 3$ if we wish to aggregate monthly data to quarterly figures. To express the temporal

aggregation as a linear transformation we define

$$\mathbf{y}_\vartheta = \begin{bmatrix} y_{m(\vartheta-1)+1} \\ y_{m(\vartheta-1)+2} \\ \vdots \\ y_{m\vartheta} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_\vartheta = \begin{bmatrix} u_{m(\vartheta-1)+1} \\ u_{m(\vartheta-1)+2} \\ \vdots \\ u_{m\vartheta} \end{bmatrix} \tag{2.8}$$

and specify the process

$$\mathcal{A}_0\mathbf{y}_\vartheta = \mathcal{A}_1\mathbf{y}_{\vartheta-1} + \cdots + \mathcal{A}_P\mathbf{y}_{\vartheta-P} + \mathcal{M}_0\mathbf{u}_\vartheta + \mathcal{M}_1\mathbf{u}_{\vartheta-1} + \cdots + \mathcal{M}_Q\mathbf{u}_{\vartheta-Q}, \tag{2.9}$$

where

$$\mathcal{A}_0 = \begin{bmatrix} A_0 & 0 & 0 & \dots & 0 \\ -A_1 & A_0 & 0 & \dots & 0 \\ -A_2 & -A_1 & A_0 & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \\ -A_{m-1} & -A_{m-2} & -A_{m-3} & \dots & A_0 \end{bmatrix},$$

$$\mathcal{A}_i = \begin{bmatrix} A_{im} & A_{im-1} & \dots & A_{im-m+1} \\ A_{im+1} & A_{im} & \dots & A_{im-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{im+m-1} & A_{im+m-2} & \dots & A_{im} \end{bmatrix}, \quad i = 1, \dots, P,$$

with $A_j = 0$ for $j > p$ and $\mathcal{M}_0, \dots, \mathcal{M}_Q$ defined in an analogous manner. The order $P = \min\{n \in \mathbb{N} | nm \geq p\}$ and $Q = \min\{n \in \mathbb{N} | nm \geq q\}$. Notice that the time subscript of $\mathbf{y}_\vartheta$ is different from that of y_t. The new time index ϑ refers to another observation frequency than t. For example, if t refers to months and $m = 3$, ϑ refers to quarters.

Using the process (2.9), temporal aggregation over m periods can be represented as a linear transformation. In fact, different types of temporal aggregation can be handled. For instance, the aggregate may be the sum of subsequent values or it may be their average. Furthermore, temporal and contemporaneous aggregation can be dealt with simultaneously. In all of these cases the aggregate has a finite order VARMA representation if the original variables are generated by a finite order VARMA process and its structure can be analyzed using linear transformations. For another approach to study temporal aggregates see Marcellino (1999).

2.4 Forecasting

2.4.1 General Results

When forecasting a set of variables is the objective, it is useful to think about a loss function or an evaluation criterion for the forecast performance. Given such a criterion, optimal forecasts may be constructed. VARMA processes are particularly useful for producing forecasts that minimize the forecast MSE. Therefore this criterion will be used here and the reader is referred to Chapter ??? for a discussion of other forecast evaluation criteria.

Forecasts of the variables of the VARMA process (2.1) are obtained easily from the pure VAR form (2.5). Assuming an independent white noise process u_t, an optimal, minimum MSE h-step forecast at time τ is the conditional expectation given the y_t, $t \leq \tau$,

$$y_{\tau+h|\tau} \equiv E(y_{\tau+h}|y_\tau, y_{\tau-1}, \dots).$$

It may be determined recursively for $h = 1, 2, \dots$, as

$$y_{\tau+h|\tau} = \sum_{i=1}^{\infty} \Xi_i y_{\tau+h-i|\tau}, \tag{2.10}$$

where $y_{\tau+j|\tau} = y_{\tau+j}$ for $j \leq 0$. If the u_t do not form an independent but only uncorrelated white noise sequence, the forecast obtained in this way is still the best linear forecast although it may not be the best in a larger class of possibly nonlinear functions of past observations.

For given initial values, the u_t can also be determined under the present assumption of a known process. Hence, the h-step forecasts may be determined alternatively as

$$y_{\tau+h|\tau} = A_0^{-1}(A_1 y_{\tau+h-1|\tau} + \dots + A_p y_{\tau+h-p|\tau}) + A_0^{-1} \sum_{i=h}^{q} M_i u_{\tau+h-i}, \tag{2.11}$$

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

Both ways of computing h-step forecasts from VARMA models rely on the availability of initial values. In the pure VAR formula (2.10) all infinitely many past y_t are in principle necessary if the VAR representation is indeed of infinite order. In contrast, in order to use (2.11), the u_t's need to be known which are unobserved and can only be obtained if all past y_t or initial conditions are available. If only $y_1, \dots, y_\tau$ are given, the infinite sum in (2.10) may be truncated accordingly. For large τ, the approximation error will be negligible because the Ξ_i's go to zero quickly as $i \to \infty$. Alternatively, precise forecasting formulas based on $y_1, \dots, y_\tau$ 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 for stable, stationary processes are easily derived by expressing the process (2.1) in Wold MA form,

$$y_t = u_t + \sum_{i=1}^{\infty} \Phi_i u_{t-i}, \tag{2.12}$$

where $A_0 = M_0$ is assumed (see (2.4)). In terms of this representation the optimal h-step forecast may be expressed as

$$y_{\tau+h|\tau} = \sum_{i=h}^{\infty} \Phi_i u_{\tau+h-i}. \tag{2.13}$$

Hence, the forecast errors are seen to be

$$y_{\tau+h} - y_{\tau+h|\tau} = u_{\tau+h} + \Phi_1 u_{\tau+h-1} + \cdots + \Phi_{h-1} u_{\tau+1}. \tag{2.14}$$

Thus, the forecast is unbiased (i.e., the forecast errors have mean zero) and the MSE or forecast error covariance matrix is

$$\Sigma_y(h) \equiv E[(y_{\tau+h} - y_{\tau+h|\tau})(y_{\tau+h} - y_{\tau+h|\tau})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'.$$

If u_t is normally distributed (Gaussian), the forecast errors are also normally distributed,

$$y_{\tau+h} - y_{\tau+h|\tau} \sim N(0, \Sigma_y(h)). \tag{2.15}$$

Hence, forecast intervals etc. may be derived from these results in the familiar way under Gaussian assumptions.

It is also interesting to note that the forecast error variance is bounded by the covariance matrix of y_t,

$$\Sigma_y(h) \to_{h\to\infty} \Sigma_y \equiv E(y_t y_t') = \sum_{j=0}^{\infty} \Phi_j \Sigma_u \Phi_j'. \tag{2.16}$$

Hence, forecast intervals will also have bounded length as the forecast horizon increases.

The situation is different if there are integrated variables. The formula (2.11) can again be used for computing the forecasts. Their properties will be different from those for stationary processes, however. Although the Wold MA representation does not exist for integrated processes, the Φ_j coefficient matrices can be computed in the same way as for stationary processes from the power series $A(z)^{-1}M(z)$ which still exists for $z \in \mathbb{C}$ with $|z| < 1$. Hence, the forecast errors can still be represented as in (2.14) (see Lütkepohl (1991, Chapter 11)). Thus, formally the forecast errors look quite similar to those for the stationary case. Now

the forecast error MSE matrix is unbounded, however, because the Φ_j's in general do not converge to zero as $j \to \infty$. Despite this general result, there may be linear combinations of the variables which can be forecast with bounded precision if the forecast horizon gets large. This situation arises if there is cointegration. For cointegrated processes it is of course also possible to base the forecasts directly on the EC form. For instance, using (2.6),

$$\Delta y_{\tau+h|\tau} = A_0^{-1}(\Pi y_{\tau+h-1|\tau} + \Gamma_1 \Delta y_{\tau+h-1|\tau} + \cdots + \Gamma_{p-1}\Delta y_{\tau+h-p+1|\tau}) + A_0^{-1}\sum_{i=h}^{q} M_i u_{\tau+h-i}, \quad (2.17)$$

and $y_{\tau+h|\tau} = y_{\tau+h-1|\tau} + \Delta y_{\tau+h|\tau}$ can be used to get a forecast of the levels variables.

As an illustration of forecasting cointegrated processes consider the following bivariate VAR model which has cointegrating rank 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \quad (2.18)$$

For this process

$$A(z)^{-1} = (I_2 - A_1 z)^{-1} = \sum_{j=0}^{\infty} A_1^j z^j = \sum_{j=0}^{\infty} \Phi_j z^j$$

exists only for $|z| < 1$ because $\Phi_0 = I_2$ and

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

does not converge to zero for $j \to \infty$. The forecast MSE is

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

where σ_2^2 is the variance of u_{2t}. The conditional expectations are $y_{k,\tau+h|\tau} = y_{2,\tau}$ $(k = 1, 2)$. Assuming normality of the white noise process, $(1-\gamma)100\%$ forecast intervals are easily seen to be

$$y_{2,\tau} \pm c_{1-\gamma/2}\sqrt{\sigma_k^2 + (h-1)\sigma_2^2}, \qquad k = 1, 2,$$

where $c_{1-\gamma/2}$ is the $(1-\gamma/2)100$ percentage point of the standard normal distribution. The lengths of these intervals increase without bounds for $h \to \infty$.

The EC representation of (2.18) is easily seen to be

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

Thus, $\mathrm{rk}(\Pi) = 1$ so that the two variables are cointegrated and some linear combinations can be forecasted with bounded forecast intervals. For the present example, multiplying (2.18) by

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

gives

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

Obviously, the cointegration relation $z_t = y_{1t} - y_{2t} = u_{1t} - u_{2t}$ is zero mean white noise and the forecast intervals for z_t for any forecast horizon $h \geq 1$ are of constant length, $z_{\tau+h|\tau} \pm c_{1-\gamma/2}\sigma_z(h)$ or $[-c_{1-\gamma/2}\sigma_z,\ c_{1-\gamma/2}\sigma_z]$. Note that $z_{\tau+h|\tau} = 0$ for $h \geq 1$ and $\sigma_z^2 = \mathrm{Var}(u_{1t}) + \mathrm{Var}(u_{2t}) - 2\mathrm{Cov}(u_{1t}, u_{2t})$ is the variance of z_t.

As long as theoretical results are discussed one could consider the first differences of the process, Δy_t, which also have a VARMA representation. If there is genuine cointegration, then Δy_t is overdifferenced in the sense that its VARMA representation has MA unit roots even if the MA part of the levels y_t is invertible.

2.4.2 Forecasting Aggregated Processes

We have argued in Section 2.3 that linear transformations of VARMA processes are often of interest, for example, if aggregation is studied. Therefore forecasts of transformed processes are also of interest. Here we present some forecasting results for transformed and aggregated processes from Lütkepohl (1987) where also proofs and further references can be found. We begin with general results which have immediate implications for contemporaneous aggregation. Then we will also present some results for temporally aggregated processes which can be obtained via the process representation (2.9).

Linear Transformations and Contemporaneous Aggregation

Suppose y_t is a stationary VARMA process with pure, invertible Wold MA representation (2.4), that is, $y_t = \Phi(L)u_t$ with $\Phi_0 = I_K$, F is an $(M \times K)$ matrix with rank M and we are interested in forecasting the transformed process $z_t = Fy_t$. It was discussed in Section 2.3 that z_t also has a VARMA representation so that the previously considered techniques can

be used for forecasting. Suppose that the corresponding Wold MA representation is

$$z_t = \sum_{i=0}^{\infty} \Psi_i v_{t-i} = \Psi(L) v_t. \tag{2.19}$$

From (2.13) the optimal h-step predictor for z_t at origin τ, based on its own past, is then

$$z_{\tau+h|\tau} = \sum_{i=h}^{\infty} \Psi_i v_{\tau+h-i}. \tag{2.20}$$

In general this predictor will differ from one based on forecasting y_t and then transforming the forecast,

$$z^o_{\tau+h|\tau} \equiv F y_{\tau+h|\tau}. \tag{2.21}$$

Before we compare the two forecasts $z^o_{\tau+h|\tau}$ and $z_{\tau+h|\tau}$ it may be of interest to draw attention to yet another possible forecast. If the dimension K of the vector y_t is large, it may be difficult to construct a suitable VARMA model for the underlying process and one may consider forecasting the individual components of y_t by univariate methods and then transforming the univariate forecasts. Because the component series of y_t can be obtained by linear transformations, they also have ARMA representations. Denoting the corresponding Wold MA representations by

$$y_{kt} = \sum_{i=0}^{\infty} \theta_{ki} w_{k,t-i} = \theta_k(L) w_{kt}, \qquad k = 1, \ldots, K, \tag{2.22}$$

the optimal univariate h-step forecasts are

$$y^u_{k,\tau+h|\tau} = \sum_{i=h}^{\infty} \theta_{ki} w_{k,\tau+h-i}, \qquad k = 1, \ldots, K. \tag{2.23}$$

Defining $y^u_{\tau+h|\tau} = (y^u_{1,\tau+h|\tau}, \ldots, y^u_{K,\tau+h|\tau})'$, these forecasts can be used to obtain an h-step forecast

$$z^u_{\tau+h|\tau} \equiv F y^u_{\tau+h|\tau} \tag{2.24}$$

of the variables of interest.

We will now compare the three forecasts (2.20), (2.21) and (2.24) of the transformed process z_t. In this comparison we denote the MSE matrices corresponding to the three forecasts by $\Sigma_z(h)$, $\Sigma^o_z(h)$ and $\Sigma^u_z(h)$, respectively. Because $z^o_{\tau+h|\tau}$ uses the largest information set, it is not surprising that it has the smallest MSE matrix and is hence the best one out of the three forecasts,

$$\Sigma_z(h) \geq \Sigma^o_z(h) \qquad \text{and} \qquad \Sigma^u_z(h) \geq \Sigma^o_z(h), \tag{2.25}$$

where "$\geq$" means that the difference between the left-hand and right-hand matrices is positive semidefinite. Thus, forecasting the original process y_t and then transforming the forecasts is generally more efficient than forecasting the transformed process directly or transforming univariate forecasts. It is possible, however, that some or all of the forecasts are identical. Actually, for $I(0)$ processes, all three predictors always approach the same long-term forecast of zero. Consequently,

$$\Sigma_z(h), \Sigma_z^o(h), \Sigma_z^u(h) \to \Sigma_z \equiv E(z_t z_t') \quad \text{as} \quad h \to \infty. \tag{2.26}$$

Moreover, it can be shown that if the one-step forecasts are identical, then they will also be identical for larger forecast horizons. More precisely we have,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau} \Rightarrow z_{\tau+h|\tau}^o = z_{\tau+h|\tau} \quad h = 1, 2, \ldots, \tag{2.27}$$

$$z_{\tau+1|\tau}^u = z_{\tau+1|\tau} \Rightarrow z_{\tau+h|\tau}^u = z_{\tau+h|\tau} \quad h = 1, 2, \ldots, \tag{2.28}$$

and, if $\Phi(L)$ and $\Theta(L)$ are invertible,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau}^u \Rightarrow z_{\tau+h|\tau}^o = z_{\tau+h|\tau}^u \quad h = 1, 2, \ldots. \tag{2.29}$$

Thus, one may ask whether the one-step forecasts can be identical and it turns out that this is indeed possible. The following proposition which summarizes results of Tiao & Guttman (1980), Kohn (1982) and Lütkepohl (1984), gives conditions for this to happen.

Proposition 1. Let y_t be a K-dimensional stochastic process with MA representation as in (2.4) with $\Phi_0 = I_K$ and F an $(M \times K)$ matrix with rank M. Then, using the notation introduced in the foregoing and defining $\Theta(L) = \text{diag}[\theta_1(L), \ldots, \theta_K(L)]$, the following relations hold:

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau} \iff F\Phi(L) = \Psi(L)F, \tag{2.30}$$

$$z_{\tau+1|\tau}^u = z_{\tau+1|\tau} \iff F\Theta(L) = \Psi(L)F \tag{2.31}$$

and, if $\Phi(L)$ and $\Theta(L)$ are invertible,

$$z_{\tau+1|\tau}^o = z_{\tau+1|\tau}^u \iff F\Phi(L)^{-1} = F\Theta(L)^{-1}. \tag{2.32}$$

□

There are several interesting implications of this proposition. First, if y_t consists of independent components ($\Phi(L) = \Theta(L)$) and z_t is just their sum, i.e., $F = (1, \ldots, 1)$, then

$$z^o_{\tau+1|\tau} = z_{\tau+1|\tau} \iff \theta_1(L) = \cdots = \theta_K(L). \tag{2.33}$$

In other words, forecasting the individual components and summing up the forecasts is strictly more efficient than forecasting the sum directly whenever the components are not generated by identical stochastic processes. Second, forecasting the univariate components of y_t individually can only be as efficient a forecast for y_t as forecasting on the basis of the multivariate process if and only if $\Phi(L)$ is a diagonal matrix operator. Related to this result is a well-known condition for Granger-noncausality. For a bivariate process $y_t = (y_{1t}, y_{2t})'$, y_{2t} is said to be Granger-causal for y_{1t} if it is helpful for improving the forecasts of the latter variable. In terms of the previous notation this may be stated by specifying $F = (1, 0)$ and defining y_{2t} as being Granger-causal for y_{1t} if $z^o_{\tau+1|\tau} = Fy_{\tau+1|\tau} = y^o_{1,\tau+1|\tau}$ is a better forecast than $z_{\tau+1|\tau} = y_{1,\tau+1|\tau}$. From (2.30) it then follows that y_{2t} is not Granger-causal for y_{1t} if and only if $\phi_{12}(L) = 0$, where $\phi_{12}(L)$ denotes the upper right hand element of $\Phi(L)$. This characterization of Granger-noncausality is well-known in the related literature (e.g., Lütkepohl (1991, Section 2.3.1)).

It may also be worth noting that in general there is no unique ranking of the forecasts $z_{\tau+1|\tau}$ and $z^u_{\tau+1|\tau}$. Depending on the structure of the underlying process y_t and the transformation matrix F, either $\Sigma_z(h) \geq \Sigma^u_z(h)$ or $\Sigma_z(h) \leq \Sigma^u_z(h)$ will hold and the relevant inequality may be strict in the sense that the left-hand and right-hand matrices are not identical.

Some but not all the results in this section carry over to nonstationary $I(1)$ processes. For example, the result (2.26) will not hold in general if some components of y_t are $I(1)$ because in this case the three forecasts do not necessarily converge to zero as the forecast horizon gets large. On the other hand, the conditions in (2.30) and (2.31) can be used for the differenced processes. For these results to hold, the MA operator may have roots on the unit circle and hence overdifferencing is not a problem.

The previous results on linearly transformed processes can also be used to compare different predictors for temporally aggregated processes by setting up the corresponding process (2.9). Some related results will be summarized next.

Temporal Aggregation

Different forms of temporal aggregation are of interest, depending on the types of variables involved. If y_t consists of stock variables, then temporal aggregation is usually associated with *systematic sampling*, sometimes called *skip-sampling* or *point-in-time sampling*. In other words, the process

$$\mathbf{s}_\vartheta = y_{m\vartheta} \tag{2.34}$$

is used as an aggregate over m periods. Here the aggregated process $\mathbf{s}_\vartheta$ has a new time index which refers to another observation frequency than the original subscript t. For example, if t refers to months and $m = 3$, then ϑ refers to quarters. In that case the process $\mathbf{s}_\vartheta$ consists of every third variable of the y_t process. This type of aggregation contrasts with temporal aggregation of flow variables where a temporal aggregate is typically obtained by summing up consecutive values. Thus, aggregation over m periods gives the aggregate

$$\mathbf{z}_\vartheta = y_{m\vartheta} + y_{m\vartheta-1} + \cdots + y_{m\vartheta-m+1}. \tag{2.35}$$

Now if, for example, t refers to months and $m = 3$, then three consecutive observations are added to obtain the quarterly value. In the following we again assume that the disaggregated process y_t is stationary and invertible and has a Wold MA representation as in (2.4), $y_t = \Phi(L)u_t$ with $\Phi_0 = I_K$. As we have seen in Section 2.3, this implies that $\mathbf{s}_\vartheta$ and $\mathbf{z}_\vartheta$ are also stationary and have Wold MA representations. We will now discuss forecasting stock and flow variables in turn. In other words, we consider forecasts for $\mathbf{s}_\vartheta$ and $\mathbf{z}_\vartheta$.

Suppose first that we wish to forecast $\mathbf{s}_\vartheta$. Then the past aggregated values $\{\mathbf{s}_\vartheta, \mathbf{s}_{\vartheta-1}, \dots\}$ may be used to obtain an h-step forecast $\mathbf{s}_{\vartheta+h|\vartheta}$ as in (2.13) on the basis of the MA representation of $\mathbf{s}_\vartheta$. If the disaggregate process y_t is available, another possible forecast results by systematically sampling forecasts of y_t which gives $\mathbf{s}^o_{\vartheta+h|\vartheta} = y_{m\vartheta+mh|m\vartheta}$. Using the results for linear transformations, the latter forecast generally has a lower MSE than $\mathbf{s}_{\vartheta+h|\vartheta}$ and the difference vanishes if the forecast horizon $h \to \infty$. For special processes the two predictors are identical, however. It follows from relation (2.30) of Proposition 1 that the two predictors are identical for $h = 1, 2, \dots$, if and only if

$$\Phi(L) = \left(\sum_{i=0}^{\infty} \Phi_{im} L^{im}\right)\left(\sum_{i=0}^{m-1} \Phi_i L^i\right) \tag{2.36}$$

(Lütkepohl (1987, Proposition 7.1)). Thus, there is no loss in forecast efficiency if the MA operator of the disaggregate process has the multiplicative structure in (2.36). This condition

is, for instance, satisfied if y_t is a purely seasonal process with seasonal period m such that

$$y_t = \sum_{i=0}^{\infty} \Phi_{im} u_{t-im}. \tag{2.37}$$

It also holds if y_t has a finite order MA structure with MA order less than m. Interestingly, it also follows that there is no loss in forecast efficiency if the disaggregate process y_t is a VAR(1) process, $y_t = A_1 y_{t-1} + u_t$. In that case, the MA operator can be written as

$$\Phi(L) = \left(\sum_{i=0}^{\infty} A_1^{im} L^{im}\right)\left(\sum_{i=0}^{m-1} A_1^i L^i\right)$$

and, hence, it has the required structure.

Now consider the case of a vector of flow variables y_t for which the temporal aggregate is given in (2.35). For forecasting the aggregate $\mathbf{z}_\vartheta$ one may use the past aggregated values and compute an h-step forecast $\mathbf{z}_{\vartheta+h|\vartheta}$ as in (2.13) on the basis of the MA representation of $\mathbf{z}_\vartheta$. Alternatively, we may again forecast the disaggregate process y_t and aggregate the forecasts. This forecast is denoted by $\mathbf{z}^o_{\vartheta+h|\vartheta}$, that is,

$$\mathbf{z}^o_{\vartheta+h|\vartheta} = y_{m\vartheta+mh|m\vartheta} + y_{m\vartheta+mh-1|m\vartheta} + \cdots + y_{m\vartheta+mh-m+1|m\vartheta}. \tag{2.38}$$

Again the results for linear transformations imply that the latter forecast generally has a lower MSE than $\mathbf{z}_{\vartheta+h|\vartheta}$ and the difference vanishes if the forecast horizon $h \to \infty$. In this case equality of the two forecasts holds for small forecast horizons $h = 1, 2, \ldots$, if and only if

$$\begin{aligned}(1 + L + \cdots + L^{m-1})&\left(\sum_{i=0}^{\infty} \Phi_i L^i\right) \\ &= \left(\sum_{j=0}^{\infty} (\Phi_{jm} + \cdots + \Phi_{jm-m+1}) L^{jm}\right)\left(\sum_{i=0}^{m-1} (\Phi_0 + \Phi_1 + \cdots + \Phi_i) L^i\right),\end{aligned} \tag{2.39}$$

where $\Phi_j = 0$ for $j < 0$ (Lütkepohl (1987, Proposition 8.1)). In other words, the two forecasts are identical and there is no loss in forecast efficiency from using the aggregate directly if the MA operator of y_t has the specified multiplicative structure upon multiplication by $(1 + L + \cdots + L^{m-1})$. This condition is also satisfied if y_t has the purely seasonal structure (2.37). However, in contrast to what was observed for stock variables, the two predictors are generally not identical if the disaggregate process y_t is generated by an MA process of order less than m.

It is perhaps also interesting to note that if there are both stock and flow variables in one system, then even if the underlying disaggregate process y_t is the periodic process (2.37), a

forecast based on the disaggregate data may be better than directly forecasting the aggregate (Lütkepohl (1987, pp. 177-178)). This result is interesting because for the purely seasonal process (2.37) using the disaggregate process will not result in superior forecasts if a system consisting either of stock variables only or of flow variables only is considered.

So far we have considered temporal aggregation of stationary processes. Most of the results can be generalized to $I(1)$ processes by considering the stationary process Δy_t instead of the original process y_t. Recall that forecasts for y_t can then be obtained from those of Δy_t. Moreover, in this context it may be worth taking into account that in deriving some of the conditions for forecast equality, the MA operator of the considered disaggregate process may have unit roots resulting from overdifferencing. A result which does not carry over to the $I(1)$ case, however, is the equality of long horizon forecasts based on aggregate or disaggregate variables. The reason is again that optimal forecasts of $I(1)$ variables do not settle down at zero eventually when $h \to \infty$.

Clearly, so far we have just discussed forecasting of known processes. In practice, the DGPs have to be specified and estimated on the basis of limited sample information. In that case quite different results may be obtained and, in particular, forecasts based on disaggregate processes may be inferior to those based on the aggregate directly. This issue is taken up again in Section 4.2 when forecasting estimated processes is considered.

Forecasting temporally aggregated processes has been discussed extensively in the literature. Early examples of treatments of temporal aggregation of time series are Abraham (1982), Amemiya & Wu (1972), Brewer (1973), Lütkepohl (1986a, b), Stram & Wei (1986), Telser (1967), Tiao (1972), Wei (1978) and Weiss (1984) among many others. More recently, Breitung & Swanson (2002) have studied the implications of temporal aggregation when the number of aggregated time units goes to infinity.

2.5 Extensions

So far we have considered processes which are too simple in some respects to qualify as DGPs of most economic time series. This was mainly done to simplify the exposition. Some important extensions will now be considered. In particular, we will discuss deterministic terms, higher order integration and seasonal unit roots as well as non-Gaussian processes.

2.5.1 Deterministic Terms

An easy way to integrate deterministic terms in our framework is to simply add them to the stochastic part. In other words, we consider processes

$$y_t = \mu_t + x_t,$$

where μ_t is a deterministic term and x_t is the purely stochastic part which is assumed to have a VARMA representation of the type considered earlier. The deterministic part can, for example, be a constant, $\mu_t = \mu_0$, a linear trend, $\mu_t = \mu_0 + \mu_1 t$, or a higher order polynomial trend. Furthermore, seasonal dummy variables or other dummies may be included.

From a forecasting point of view, deterministic terms are easy to handle because by their very nature their future values are precisely known. Thus, in order to forecast y_t, we may forecast the purely stochastic process x_t as discussed earlier and then simply add the deterministic part corresponding to the forecast period. In this case, the forecast errors and MSE matrices are the same as for the purely stochastic process. Of course, in practice the deterministic part may contain unknown parameters which have to be estimated from data. For the moment this issue is ignored because we are considering known processes. It will become important, however, in Section 4, where forecasting estimated processes is discussed.

2.5.2 More Unit Roots

In practice the order of integration of some of the variables can be greater than one and $\det A(z)$ may have roots on the unit circle other than $z = 1$. For example, there may be seasonal unit roots. Considerable research has been done on these extensions of our basic models. See, for instance, Johansen (1995b, 1997), Gregoir & Laroque (1994) and Haldrup (1998) for discussions of the $I(2)$ and higher order integration frameworks, and Johansen & Schaumburg (1999) and Gregoir (1999a, b) for research on processes with roots elsewhere on the unit circle. Bauer & Wagner (2003) consider state space representations for VARMA models with roots at arbitrary points on the unit circle.

As long as the processes are assumed to be known these issues do not create additional problems for forecasting because we can still use the general forecasting formulas for VARMA processes. Extensions are important, however, when it comes to model specification and estimation. In these steps of the forecasting procedure taking into account extensions in the methodology may be useful.

2.5.3 Non-Gaussian Processes

If the DGP of a multiple time series is not normally distributed, point forecasts can be computed as before. They will generally still be best *linear* forecasts and may in fact be minimum MSE forecasts if u_t is independent white noise, as discussed in Section 2.4. In setting up forecast intervals the distribution has to be taken into account, however. If the distribution is unknown, bootstrap methods can be used to compute interval forecasts (e.g., Findley (1986), Masarotto (1990), Grigoletto (1998), Kabaila (1993), Pascual, Romo & Ruiz (2004)).

3 Specifying and Estimating VARMA Models

As we have seen in the previous section, for forecasting purposes the pure VAR or MA representations of a stochastic process are quite useful. These representations are in general of infinite order. In practice, they have to be replaced by finite dimensional parameterizations which can be specified and estimated from data. VARMA processes are such finite dimensional parameterizations. Therefore, in practice, a VARMA model such as (2.1) or even a pure finite order VAR as in (1.1) will be specified and estimated as a forecasting tool.

As mentioned earlier, the operators $A(L)$ and $M(L)$ of the VARMA model (2.2) are not unique or not identified, as econometricians sometimes say. This nonuniqueness is problematic if the process parameters have to be estimated because a unique representation is needed for consistent estimation. Before we discuss estimation and specification issues related to VARMA processes we will therefore present identifying restrictions. More precisely, the echelon form of VARMA and EC-VARMA models will be presented. Then estimation procedures, model specification and diagnostic checking will be discussed.

3.1 The Echelon Form

Any pair of operators $A(L)$ and $M(L)$ that gives rise to the same VAR operator $\Xi(L) = I_K - \sum_{i=1}^{\infty} \Xi_i L^i = M(L)^{-1}A(L)$ or MA operator $\Phi(L) = A(L)^{-1}M(L)$ defines an equivalent VARMA process for y_t. Here $A_0 = M_0$ is assumed. Clearly, if we premultiply $A(L)$ and $M(L)$ by some invertible operator $D(L) = D_0 + D_1 L + \cdots + D_q L^q$ satisfying $\det(D_0) \neq 0$ and $\det D(z) \neq 0$ for $|z| \leq 1$, an equivalent VARMA representation is obtained. Thus, a first step towards finding a unique representation is to cancel common factors in $A(L)$ and $M(L)$. We therefore assume that the operator $[A(L) : M(L)]$ is *left-coprime*. To define this

property, recall that a matrix polynomial $D(z)$ and the corresponding operator $D(L)$ are *unimodular* if $\det D(z)$ is a constant which does not depend on z. Examples of unimodular operators are

$$D(L) = D_0 \quad \text{or} \quad D(L) = \begin{bmatrix} 1 & \delta L \\ 0 & 1 \end{bmatrix} \tag{3.1}$$

(see Lütkepohl (1996) for definitions and properties of matrix polynomials). A matrix operator $[A(L) : M(L)]$ is called left-coprime if only unimodular operators $D(L)$ can be factored. In other words, $[A(L) : M(L)]$ is left-coprime if for operators $\bar{A}(L)$, $\bar{M}(L)$ an operator $D(L)$ exists such that $[A(L) : M(L)] = D(L)[\bar{A}(L) : \bar{M}(L)]$ holds, then $D(L)$ must be unimodular.

Although considering only left-coprime operators $[A(L) : M(L)]$ does not fully solve the nonuniqueness problem of VARMA representations it is a first step in the right direction because it excludes many possible redundancies. It does not rule out premultiplication by some nonsingular matrix, for example, and thus, there is still room for improvement. Even if $A_0 = M_0 = I_K$ is assumed, uniqueness of the operators is not achieved because there are unimodular operators $D(L)$ with zero-order matrix I_K, as seen in (3.1). Premultiplying $[A(L) : M(L)]$ by such an operator maintains left-coprimeness. Therefore more restrictions are needed for uniqueness. The echelon form discussed in the next subsections provides sufficiently many restrictions in order to ensure uniqueness of the operators. We will first consider stationary processes and then turn to EC-VARMA models.

3.1.1 Stationary Processes

We assume that $[A(L) : M(L)]$ is left-coprime and we denote the kl-th elements of $A(L)$ and $M(L)$ by $\alpha_{kl}(L)$ and $m_{kl}(L)$, respectively. Let p_k be the maximum polynomial degree in the k-th row of $[A(L) : M(L)]$, $k = 1, \ldots, K$, and define

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

The VARMA process is said to be in *echelon form* or, briefly, ARMA_E form if the operators $A(L)$ and $M(L)$ satisfy the following restrictions (Lütkepohl & Claessen (1997), Lütkepohl (2002)):

$$m_{kk}(L) = 1 + \sum_{i=1}^{p_k} m_{kk,i} L^i, \quad \text{for } k = 1, \ldots, K, \tag{3.2}$$

$$m_{kl}(L) = \sum_{i=p_k-p_{kl}+1}^{p_k} m_{kl,i}L^i, \quad \text{for } k \neq l, \tag{3.3}$$

and

$$\alpha_{kl}(L) = \alpha_{kl,0} - \sum_{i=1}^{p_k} \alpha_{kl,i}L^i, \quad \text{with } \alpha_{kl,0} = m_{kl,0} \quad \text{for } k, l = 1, \ldots, K. \tag{3.4}$$

Here the row degrees p_k $(k = 1, \ldots, K)$ are called the *Kronecker indices* (see Hannan & Deistler (1988), Lütkepohl (1991)). In the following we denote an echelon form with Kronecker indices $p_1, \ldots, p_K$ by $\text{ARMA}_E(p_1, \ldots, p_K)$. Notice that it corresponds to a VARMA(p, p) representation in (2.1) with $p = \max(p_1, \ldots, p_K)$. An ARMA_E form may have more zero coefficients than those specified by the restrictions from (3.2)-(3.4). In particular, there may be models where the AR and MA orders are not identical due to further zero restrictions. Such over-identifying constraints are not ruled out by the echelon form. It does not need them to ensure uniqueness of the operator $[A(L) : M(L)]$ for a given VAR operator $\Xi(L)$ or MA operator $\Phi(L)$, however. Note also that every VARMA process can be written in echelon form. Thus, the echelon form does not exclude any VARMA processes.

The present specification of the echelon form does not restrict the autoregressive operator except for the maximum row degrees imposed by the Kronecker indices and the zero order matrix ($A_0 = M_0$). Additional identifying zero restrictions are placed on the moving average coefficient matrices attached to low lags of the error process u_t. This form of the echelon form was proposed by Lütkepohl & Claessen (1997) because it can be combined conveniently with the EC representation of a VARMA process, as we will see shortly. Thus, it is particularly useful for processes with cointegrated variables. A slightly different echelon form is usually used in the literature on stationary processes. Typically the restrictions on low order lags are imposed on the VAR coefficient matrices (e.g., Hannan & Deistler (1988), Lütkepohl (1991)).

To illustrate the present ARMA_E form we consider the following three-dimensional process from Lütkepohl (2002) with Kronecker indices $(p_1, p_2, p_3) = (1, 2, 1)$. It is easy to derive the p_{kl},

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

Hence, we get the $\text{ARMA}_E(1,2,1)$ form:

$$\begin{aligned}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} y_t &= \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} & \alpha_{13,1} \\ \alpha_{21,1} & \alpha_{22,1} & \alpha_{23,1} \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ \alpha_{21,2} & \alpha_{22,2} & \alpha_{23,2} \\ 0 & 0 & 0 \end{bmatrix} y_{t-2} \\ &+ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & m_{12,1} & m_{13,1} \\ 0 & m_{22,1} & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1} \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ m_{21,2} & m_{22,2} & m_{23,2} \\ 0 & 0 & 0 \end{bmatrix} u_{t-2}. \end{aligned} \tag{3.5}$$

Although in this example the zero order matrix is lower triangular, it will often be an identity matrix in an echelon form. In fact, it will always be an identity matrix if the Kronecker indices are ordered from smallest to largest.

3.1.2 $I(1)$ Processes

If the EC form of the ARMA_E model is set up as in (2.6), the autoregressive short-run coefficient matrices Γ_i $(i = 1, \ldots, p-1)$ satisfy similar identifying constraints as the A_i's $(i = 1, \ldots, p)$. More precisely, Γ_i obeys the same zero restrictions as A_{i+1} for $i = 1, \ldots, p-1$. This structure follows from the specific form of the zero restrictions on the A_i's. If $\alpha_{kl,i}$ is restricted to zero by the echelon form this implies that the corresponding element $\alpha_{kl,j}$ of A_j is also zero for $j > i$. Similarly, the zero restrictions on Π are the same as those on $A_0 - A_1$. Because the echelon form does not impose zero restrictions on A_1 if all Kronecker indices $p_k \geq 1$ $(k = 1, \ldots, K)$, there are no echelon form zero restrictions on Π if all Kronecker indices are greater than zero. On the other hand, if there are zero Kronecker indices, this has consequences for the rank of Π and, hence, for the integration and cointegration structure of the variables. In fact, denoting by ϱ the number of zero Kronecker indices, it can be shown that

$$\text{rk}(\Pi) \geq \varrho. \tag{3.6}$$

This result is useful to remember when procedures for specifying the cointegrating rank of a VARMA system are considered.

In the following we use the acronym EC-ARMA_E for an EC-VARMA model which satisfies the echelon form restrictions. To illustrate its structure, we follow again Lütkepohl

(2002) and use the model (3.5). The corresponding EC-ARMA$_E$ form is

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

The following three-dimensional ARMA$_E(0,0,1)$ model is another example from Lütkepohl (2002):

$$y_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1} + u_t + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1}. \tag{3.7}$$

Note that in this case $A_0 = I_3$. Now two of the Kronecker indices are zero and, hence, according to (3.5), the cointegrating rank of this system must be at least 2. The EC-ARMA$_E$ form is

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

The rank of

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

is clearly at least two.

Because we now have a unique representation of VARMA models we can discuss estimation of such models. Of course, to estimate an ARMA$_E$ or EC-ARMA$_E$ form we need to specify the Kronecker indices and possibly the cointegrating rank. We will discuss parameter estimation first and then consider model specification issues.

Before we go on with these topics, we mention that there are other ways to achieve uniqueness or identification of a VARMA representation. For example, Zellner & Palm (1974) and Wallis (1977) consider a *final form* representation which also solves the identification problem. It often results in rather heavily parameterized models and has therefore not gained much popularity. Tiao & Tsay (1989) propose so-called *scalar component models* to overcome the identification problem. The idea is to consider linear combinations of the variables which can reveal simplifications of the general VARMA structure. The interested reader is referred to the aforementioned article. We have presented the echelon form here in some detail because it often results in parsimonious representations.

3.2 Estimation of VARMA Models for Given Lag Orders and Cointegrating Rank

For given Kronecker indices the ARMA$_E$ form of a VARMA DGP can be set up and estimated. We will consider this case first and then study estimation of EC-ARMA$_E$ models for which the cointegrating rank is given in addition to the Kronecker indices. Specification of the Kronecker indices and the cointegrating rank will be discussed in Sections 3.4 and 3.3, respectively.

3.2.1 ARMA$_E$ Models

Suppose the white noise process u_t is normally distributed (Gaussian), $u_t \sim N(0, \Sigma_u)$. Given a sample $y_1, \ldots, y_T$ and presample values $y_0, \ldots, y_{p-1}, u_0, \ldots, u_{q-1}$, the log-likelihood function of the VARMA model (2.1) is

$$l(\theta) = \sum_{t=1}^{T} l_t(\theta). \tag{3.8}$$

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

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

where

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

It is assumed that the uniqueness restrictions of the ARMA$_E$ form are imposed and θ contains the freely varying parameters only. The initial values are assumed to be fixed and if the u_t

($t \leq 0$) are not available, they may be replaced by zero without affecting the asymptotic properties of the estimators.

Maximization of $l(\theta)$ is a nonlinear optimization problem which is complicated by the inequality constraints that ensure invertibility of the MA operator. Iterative optimization algorithms may be used here. Start-up values for such algorithms may be obtained as follows: An unrestricted long VAR model of order h_T, say, is fitted by OLS in a first step. Denoting the estimated residuals by $\hat{u}_t$, the ARMA$_E$ form can be estimated when all lagged u_t's are replaced by $\hat{u}_t$'s. If $A_0 \neq I_K$, then unlagged u_{jt} in equation k ($k \neq j$) may also be replaced by estimated residuals from the long VAR. The resulting parameter estimates can be used as starting values for an iterative algorithm.

If the DGP is stable and invertible and the parameters are identified, the ML estimator $\hat{\theta}$ has standard limiting properties, that is, $\hat{\theta}$ is consistent and

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

where $\xrightarrow{d}$ signifies convergence in distribution and $\Sigma_{\hat{\theta}}$ is the inverse asymptotic information matrix. This result holds even if the true distribution of the u_t's is not normal but satisfies suitable moment conditions. In that case the estimators are just quasi ML estimators, of course.

There has been some discussion of the likelihood function of VARMA models and its maximization (Tunnicliffe Wilson (1973), Nicholls & Hall (1979), Hillmer & Tiao (1979)). Unfortunately, optimization of the Gaussian log-likelihood is not a trivial exercise. Therefore other estimation methods have been proposed in the literature (e.g., Koreisha & Pukkila (1987), Kapetanios (2003), Poskitt (2003)). Of course, it is also straightforward to add deterministic terms to the model and estimate the associated parameters along with the VARMA coefficients.

3.2.2 EC-ARMA$_E$ Models

If the cointegrating rank r is given and the DGP is a pure, finite order VAR(p) process, the corresponding VECM,

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t, \tag{3.9}$$

can be estimated conveniently by RR regression, as shown in Johansen (1995a). Concentrating out the short-run dynamics by regressing Δy_t and y_{t-1} on $\Delta Y'_{t-1} = [\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}]$

and denoting the residuals by R_{0t} and R_{1t}, respectively, the EC term can be estimated by RR regression from

$$R_{0t} = \alpha\beta' R_{1t} + u_t^c. \tag{3.10}$$

Because the decomposition $\Pi = \alpha\beta'$ is not unique, the estimators for α and β are not consistent whereas the resulting ML estimator for Π is consistent. However, because the matrices α and β have rank r, one way to make them unique is to choose

$$\beta' = [I_r : \beta'_{(K-r)}], \tag{3.11}$$

where $\beta_{(K-r)}$ is a $((K-r)\times r)$ matrix. This normalization is always possible upon a suitable ordering of the variables. The ML estimator of $\beta_{(K-r)}$ can be obtained by post-multiplying the RR estimator $\tilde{\beta}$ of β by the inverse of its first r rows and using the last $K-r$ rows as the estimator $\tilde{\beta}_{(K-r)}$ of $\beta_{(K-r)}$. This estimator is not only consistent but even superconsistent meaning that it converges at a faster rate than the usual $\sqrt{T}$ to the true parameter matrix $\beta_{(K-r)}$. In fact, it turns out that $T(\tilde{\beta}_{(K-r)} - \beta_{(K-r)})$ converges in distribution. As a result inference for the other parameters can be done as if the cointegration matrix β were known.

Other estimation procedures that can be used here as well were proposed by Ahn & Reinsel (1990) and Saikkonen (1992). In fact, in the latter article it was shown that the procedure can even be justified if the true DGP is an infinite order VAR process and only a finite order model is fitted, as long as the order goes to infinity with growing sample size. This result is convenient in the present situation where we are interested in VARMA processes, because we can estimate the cointegration relations in a first step on the basis of a finite order VECM without MA part. Then the estimated cointegration matrix can be used in estimating the remaining VARMA parameters. That is, the short-run parameters including the loading coefficients α and MA parameters of the EC-ARMA$_E$ form can then be estimated by ML conditional on the estimator for β. Because of the superconsistency of the estimator for the cointegration parameters this procedure maintains the asymptotic efficiency of the Gaussian ML estimator. Except for the cointegration parameters, the parameter estimators have standard asymptotic properties which are equivalent to those of the full ML estimators (Yap & Reinsel (1995)). If the Kronecker indices are given, the echelon VARMA structure can also be taken into account in estimating the cointegration matrix.

As mentioned earlier, before a model can be estimated, the Kronecker indices and possibly the cointegrating rank have to be specified. These issues are discussed next.

3.3 Testing for the Cointegrating Rank

A wide range of proposals exists for determining the cointegrating ranks of pure VAR processes (see Hubrich, Lütkepohl & Saikkonen (2001) for a recent survey). The most popular approach is due to Johansen (1995a) who derives likelihood ratio (LR) tests for the cointegrating rank of a pure VAR process. Because ML estimation of unrestricted VECMs with a specific cointegrating rank r is straightforward for Gaussian processes, the LR statistic for testing the pair of hypotheses $H_0 : r = r_0$ versus $H_1 : r > r_0$ is readily available by comparing the likelihood maxima for $r = r_0$ and $r = K$. The asymptotic distributions of the LR statistics are nonstandard and depend on the deterministic terms included in the model. Tables with critical values for various different cases are available in Johansen (1995a, Chapter 15). The cointegrating rank can be determined by checking sequentially the null hypotheses

$$H_0 : r = 0, H_0 : r = 1, \ldots, H_0 : r = K - 1$$

and choosing the cointegrating rank for which the first null hypothesis cannot be rejected in this sequence.

For our present purposes it is of interest that Johansen's LR tests can be justified even if a finite-order VAR process is fitted to an infinite order DGP, as shown by Lütkepohl & Saikkonen (1999). It is assumed in this case that the order of the fitted VAR process goes to infinity with the sample size and Lütkepohl & Saikkonen (1999) discuss the choice of the VAR order in this approach. Because the Kronecker indices are usually also unknown, choosing the cointegrating rank of a VARMA process by fitting a long VAR process is an attractive approach which avoids knowledge of the VARMA structure at the stage where the cointegrating rank is determined. So far the theory for this procedure seems to be available for processes with nonzero mean term only and not for other deterministic terms such as linear trends. It seems likely, however, that extensions to more general processes are possible.

An alternative way to proceed in determining the cointegrating rank of a VARMA process was proposed by Yap & Reinsel (1995). They extended the likelihood ratio tests to VARMA processes under the assumption that an identified structure of $A(L)$ and $M(L)$ is known. For these tests the Kronecker indices or some other identifying structure has to be specified first. If the Kronecker indices are known already, a lower bound for the cointegrating rank is also known (see (3.6)). Hence, in testing for the cointegrating rank, only the sequence of null hypotheses $H_0 : r = \varrho, H_0 : r = \varrho + 1, \ldots, H_0 : r = K - 1$, is of interest. Again, the rank may be chosen as the smallest value for which H_0 cannot be rejected.

3.4 Specifying the Lag Orders and Kronecker Indices

A number of proposals for choosing the Kronecker indices of ARMA$_E$ models were made, see, for example, Hannan & Kavalieris (1984), Poskitt (1992), Nsiri & Roy (1992) and Lütkepohl & Poskitt (1996) for stationary processes and Lütkepohl & Claessen (1997), Claessen (1995), Poskitt & Lütkepohl (1995) and Poskitt (2003) for cointegrated processes. The strategies for specifying the Kronecker indices of cointegrated ARMA$_E$ processes presented in this section are proposed in the latter two papers. Poskitt (2003, Proposition 3.3) presents a result regarding the consistency of the estimators of the Kronecker indices. A simulation study of the small sample properties of the procedures was performed by Bartel & Lütkepohl (1998). They found that the methods work reasonably well in small samples for the processes considered in their study. This section draws partly on Lütkepohl (2002, Section 8.4.1).

The specification method proceeds in two stages. In the first stage a long reduced-form VAR process of order h_T, say, is fitted by OLS giving estimates of the unobservable innovations u_t as in the previously described estimation procedure. In a second stage the estimated residuals are substituted for the unknown lagged u_t's in the ARMA$_E$ form. A range of different models is estimated and the Kronecker indices are chosen by model selection criteria.

There are different possibilities for doing so within this general procedure. For example, one may search over all models associated with Kronecker indices which are smaller than some prespecified upper bound $p_{\max}$, $\{(p_1, \ldots, p_K) | 0 \leq p_k \leq p_{\max}, k = 1, \ldots, K\}$. The set of Kronecker indices is then chosen which minimizes the preferred model selection criterion. For systems of moderate or large dimensions this procedure is rather computer intensive and computationally more efficient search procedures have been suggested. One idea is to estimate the individual equations separately by OLS for different lag lengths. The lag length is then chosen so as to minimize a criterion of the general form

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

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

In this procedure the echelon structure is not explicitly taken into account because the equations are treated separately. The k-th equation will still be misspecified if the lag order is less than the true Kronecker index. Moreover, the k-th equation will be correctly specified but may include redundant parameters and variables if the lag order is greater than the

true Kronecker index. This explains why the criterion function $\Lambda_{k,T}(n)$ will possess a global minimum asymptotically when n is equal to the true Kronecker index, provided C_T is chosen appropriately. In practice, possible choices of C_T are $C_T = h_T \log T$ or $C_T = h_T^2$ (see Poskitt (2003) for more details on the procedure). Poskitt & Lütkepohl (1995) and Poskitt (2003) also consider a modification of this procedure where coefficient restrictions derived from those equations in the system which have smaller Kronecker indices are taken into account. The important point to make here is that procedures exist which can be applied in a fully computerized model choice. Thus, model selection is feasible from a practical point of view although the small sample properties of these procedures are not clear in general, despite some encouraging but limited small sample evidence by Bartel & Lütkepohl (1998). Other procedures for specifying the Kronecker indices for stationary processes were proposed by Akaike (1976), Cooper & Wood (1982), Tsay (1989b) and Nsiri & Roy (1992), for example.

The Kronecker indices found in a computer automated procedure for a given time series should only be viewed as a starting point for a further analysis of the system under consideration. Based on the specified Kronecker indices a more efficient procedure for estimating the parameters may be applied (see Section 3.2) and the model may be subjected to a range of diagnostic tests. If such tests produce unsatisfactory results, modifications are called for. Tools for checking the model adequacy will be briefly summarized in the following section.

3.5 Diagnostic Checking

As noted in Section 3.2, the estimators of an identified version of a VARMA model have standard asymptotic properties. Therefore the usual t- and F-tests can be used to decide on possible overidentifying restrictions. When a parsimonious model without redundant parameters has been found, the residuals can be checked. According to our assumptions they should be white noise and a number of model-checking tools are tailored to check this assumption. For this purpose one may consider individual residual series or one may check the full residual vector at once. The tools range from visual inspection of the plots of the residuals and their autocorrelations to formal tests for residual autocorrelation and autocorrelation of the squared residuals to tests for nonnormality and nonlinearity (see, e.g., Lütkepohl (1991), Doornik & Hendry (1997)). It is also advisable to check for structural shifts during the sample period. Possible tests based on prediction errors are considered in Lütkepohl (1991). Moreover, when new data becomes available, out-of-sample forecasts may be checked. Model defects detected at the checking stage should lead to modifications of the

original specification.

4 Forecasting with Estimated Processes

4.1 General Results

To simplify matters suppose that the generation process of a multiple time series of interest admits a VARMA representation with zero order matrices equal to I_K,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \tag{4.1}$$

that is, $A_0 = M_0 = I_K$. Recall that in the echelon form framework this representation can always be obtained by premultiplying by A_0^{-1} if $A_0 \neq I_K$. We denote by $\hat{y}_{\tau+h|\tau}$ the h-step forecast at origin τ given in Section 2.4, based on estimated rather than known coefficients. For instance, using the pure VAR representation of the process,

$$\hat{y}_{\tau+h|\tau} = \sum_{i=1}^{h-1} \hat{\Xi}_i \hat{y}_{\tau+h-i|\tau} + \sum_{i=h}^{\infty} \hat{\Xi}_i y_{\tau+h-i}. \tag{4.2}$$

For practical purposes one may truncate the infinite sum at $i = \tau$. For the moment we will, however, consider the infinite sum. For this predictor the forecast error is

$$y_{\tau+h} - \hat{y}_{\tau+h|\tau} = (y_{\tau+h} - y_{\tau+h|\tau}) + (y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau}),$$

where $y_{\tau+h|\tau}$ is the optimal forecast based on known coefficients and the two terms on the right-hand side are uncorrelated if only data up to period τ are used for estimation. In that case the first term can be written in terms of u_t's with $t > \tau$ and the second one contains only y_t's with $t \leq \tau$. Thus, the forecast MSE becomes

$$\begin{aligned} \Sigma_{\hat{y}}(h) &= \text{MSE}(y_{\tau+h|\tau}) + \text{MSE}(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau}) \\ &= \Sigma_y(h) + E[(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})']. \end{aligned} \tag{4.3}$$

The $\text{MSE}(y_{\tau+h|\tau} - \hat{y}_{\tau+h|\tau})$ can be approximated by $\Omega(h)/T$, where

$$\Omega(h) = E\left[\frac{\partial y_{\tau+h|\tau}}{\partial \theta'} \Sigma_{\tilde{\theta}} \frac{\partial y'_{\tau+h|\tau}}{\partial \theta}\right], \tag{4.4}$$

θ is the vector of estimated coefficients, and $\Sigma_{\tilde{\theta}}$ is its asymptotic covariance matrix (see Yamamoto (1980), Baillie (1981) and Lütkepohl (1991) for more detailed expressions for $\Omega(h)$ and Hogue, Magnus & Pesaran (1988) for an exact treatment of the AR(1) special

case). If ML estimation is used, the covariance matrix $\Sigma_{\tilde{\theta}}$ is just the inverse asymptotic information matrix. Clearly, $\Omega(h)$ is positive semidefinite and the forecast MSE,

$$\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h), \tag{4.5}$$

for estimated processes is larger (or at least not smaller) than the corresponding quantity for known processes, as one would expect. The additional term depends on the estimation efficiency because it includes the asymptotic covariance matrix of the parameter estimators. Therefore, estimating the parameters of a given process well is also important for forecasting. On the other hand, for large sample sizes T, the additional term will be small or even negligible.

It may be worth noting that deterministic terms can be accommodated easily, as discussed in Section 2.5. In the present situation the uncertainty in the estimators related to such terms can also be taken into account like that of the other parameters. If the deterministic terms are specified such that the corresponding parameter estimators are asymptotically independent of the other estimators, an additional term for the estimation uncertainty stemming from the deterministic terms has to be added to the forecast MSE matrix (4.5). For deterministic linear trends in univariate models more details are presented in Kim, Leybourne & Newbold (2004).

Various extensions of the previous results have been discussed in the literature. For example, Lewis & Reinsel (1985) and Lütkepohl (1985) consider the forecast MSE for the case where the true process is approximated by a finite order VAR, thereby extending earlier univariate results by Bhansali (1978). Reinsel & Lewis (1987), Basu & Sen Roy (1987), Engle & Yoo (1987), Sampson (1991) and Reinsel & Ahn (1992) present results for processes with unit roots. Stock (1996) and Kemp (1999) assume that the forecast horizon h and the sample size T both go to infinity simultaneously. Clements & Hendry (1998, 2001) consider various other sources of possible forecast errors. Taking into account the specification and estimation uncertainty in multi-step forecasts, it makes also sense to construct a separate model for each specific forecast horizon h. This approach is discussed in detail by Bhansali (2002).

4.2 Aggregated Processes

In Section 2.4 we have compared different forecasts for aggregated time series. It was found that generally forecasting the disaggregate process and aggregating the forecasts ($z^o_{\tau+h|\tau}$) is

more efficient than forecasting the aggregate directly ($z_{\tau+h|\tau}$). In this case, if the sample size is large enough, the part of the forecast MSE due to estimation uncertainty will eventually be so small that the estimated $\hat{z}^o_{\tau+h|\tau}$ is again superior to the corresponding $\hat{z}_{\tau+h|\tau}$. There are cases, however, where the two forecasts are identical for known processes. Now the question arises whether in these cases the MSE term due to estimation errors will make one forecast preferable to its competitors. Indeed if estimated instead of known processes are used, it is possible that $\hat{z}^o_{\tau+h|\tau}$ looses its optimality relative to $\hat{z}_{\tau+h|\tau}$ because the MSE part due to estimation may be larger for the former than for the latter. Consider the case, where a number of series are simply added to obtain a univariate aggregate. Then it is possible that a simple parsimonious univariate ARMA model describes the aggregate well, whereas a large multivariate model is required for an adequate description of the multivariate disaggregate process. Clearly, it is conceivable that the estimation uncertainty in the multivariate case becomes considerably more important than for the univariate model for the aggregate. Lütkepohl (1987) shows that this may indeed happen in small samples. In fact, similar situations can not only arise for contemporaneous aggregation but also for temporal aggregation. Generally, if two predictors based on known processes are nearly identical, the estimation part of the MSE becomes important and generally the predictor based on the smaller model is then to be preferred.

There is also another aspect which is important for comparing forecasts. So far we have only taken into account the effect of estimation uncertainty on the forecast MSE. This analysis still assumes a known model structure and only allows for estimated parameters. In practice, model specification usually precedes estimation and usually there is additional uncertainty attached to this step in the forecasting procedure. It is also possible to explicitly take into account the fact that in practice models are only approximations to the true DGP by considering finite order VAR and AR approximations to infinite order processes. This has also been done by Lütkepohl (1987). Under these assumptions it is again found that the forecast $\hat{z}^o_{\tau+h|\tau}$ looses its optimality and forecasting the aggregate directly or forecasting the disaggregate series with univariate methods and aggregating univariate forecasts may become preferable.

5 Conclusions

VARMA models are a powerful tool for producing linear forecasts for a set of time series variables. They utilize the information not only in the past values of a particular variable

of interest but also allow for information in other, related variables. We have mentioned conditions under which the forecasts from these models are optimal under an MSE criterion for forecast performance. Even if the conditions for minimizing the forecast MSE in the class of all functions are not satisfied the forecasts will be best linear forecasts under general assumptions. These appealing theoretical features of VARMA models make them attractive tools for forecasting.

Special attention has been paid to forecasting linearly transformed and aggregated processes. Both contemporaneous as well as temporal aggregation have been studied. It was found that generally forecasting the disaggregated process and aggregating the forecasts is more efficient than forecasting the aggregate directly and thereby ignoring the disaggregate information. Moreover, for contemporaneous aggregation, forecasting the individual components with univariate methods and aggregating these forecasts was compared to the other two possible forecasts. Forecasting univariate components separately may lead to better forecasts than forecasting the aggregate directly. It will be inferior to aggregating forecasts of the fully disaggregated process, however. These results hold if the DGPs are known.

In practice the relevant model for forecasting a particular set of time series will not be known, however, and it is necessary to use sample information to specify and estimate a suitable candidate model from the VARMA class. We have discussed estimation methods and specification algorithms which are suitable at this stage of the forecasting process for stationary as well as integrated processes. The nonuniqueness or lack of identification of general VARMA representations turned out to be a major problem at this stage. We have focussed on the echelon form as one possible parameterization that allows to overcome the identification problem. The echelon form has the advantage of providing a relatively parsimonious VARMA representation in many cases. Moreover, it can be extended conveniently to cointegrated processes by including an EC term. It is described by a set of integers called Kronecker indices. Statistical procedures were presented for specifying these quantities. We have also presented methods for determining the cointegrating rank of a process if some or all of the variables are integrated. This can be done by applying standard cointegrating rank tests for pure VAR processes because these tests maintain their usual asymptotic properties even if they are performed on the basis of an approximating VAR process rather than the true DGP. We have also briefly discussed issues related to checking the adequacy of a particular model. Overall a coherent strategy for specifying, estimating and checking VARMA models has been presented. Finally, the implications of using estimated rather than known processes for forecasting have been discussed.

If estimation and specification uncertainty are taken into account it turns out that forecasts based on a disaggregated multiple time series may not be better and may in fact be inferior to forecasting an aggregate directly. This situation is in particular likely to occur if the DGPs are such that efficiency gains from disaggregation do not exist or are small and the aggregated process has a simple structure which can be captured with a parsimonious model.

Clearly, VARMA models also have some drawbacks as forecasting tools. First of all, linear forecasts may not always be the best choice (see Chapter ??? for a discussion of forecasting with nonlinear models). Second, adding more variables in a system does not necessarily increase the forecast precision. Higher dimensional systems are typically more difficult to specify than smaller ones. Thus, considering as many series as possible in one system is clearly not a good strategy. The increase in estimation and specification uncertainty may offset the advantages of using additional information. VARMA models appear to be most useful for analyzing small sets of time series. Choosing the best set of variables for a particular forecasting exercise may not be an easy task. In conclusion, although VARMA models are an important forecasting tool, the actual success very much depends on the skills of the user of these tools.

References

Abraham, B. (1982). Temporal aggregation and time series, *International Statistical Review* **50**: 285–291.

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

Akaike, H. (1974). Stochastic theory of minimal realization, *IEEE Transactions on Automatic Control* **AC-19**: 667–674.

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion, *in* R. K. Mehra & D. G. Lainiotis (eds), *Systems Identification: Advances and Case Studies*, Academic Press, New York, pp. 27–96.

Amemiya, T. & Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association* **67**: 628–632.

Aoki, M. (1987). *State Space Modeling of Time Series*, Springer-Verlag, Berlin.

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

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

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

Bauer, D. & Wagner, M. (2003). A canonical form for unit root processes in the state space framework, Diskussionsschriften 03-12, Universität Bern.

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

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

Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Breitung, J. & Swanson, N. R. (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models, *Journal of Time Series Analysis* **23**: 651–665.

Brewer, K. R. W. (1973). Some consequences of temporal aggregation and ssystematic sampling for ARMA and ARMAX models, *Journal of Econometrics* **1**: 133–154.

Brockwell, P. J. & Davis, R. A. (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

Claessen, H. (1995). *Spezifikation und Schätzung von VARMA-Prozessen unter besonderer Berücksichtigung der Echelon Form*, Verlag Joseph Eul, Bergisch-Gladbach.

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

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

Cooper, D. M. & Wood, E. F. (1982). Identifying multivariate time series models, *Journal of Time Series Analysis* **3**: 153–164.

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

Dunsmuir, W. T. M. & Hannan, E. J. (1976). Vector linear time series models, *Advances in Applied Probability* **8**: 339–364.

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

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

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.

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. & Newbold, P. (1977). *Forecasting Economic Time Series*, Academic Press, New York.

Gregoir, S. (1999a). Multivariate time series with various hidden unit roots, part I: Integral operator algebra and representation theorem, *Econometric Theory* **15**: 435–468.

Gregoir, S. (1999b). Multivariate time series with various hidden unit roots, part II: Estimation and test, *Econometric Theory* **15**: 469–518.

Gregoir, S. & Laroque, G. (1994). Polynomial cointegration: Estimation and test, *Journal of Econometrics* **63**: 183–214.

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

Haldrup, N. (1998). An econometric analysis of I(2) variables, *Journal of Economic Surveys* **12**: 595–650.

Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.

Hannan, E. J. (1976). The identification and parameterization of ARMAX and state space forms, *Econometrica* **44**: 713–723.

Hannan, E. J. (1979). The statistical theory of linear systems, *in* P. R. Krishnaiah (ed.), *Developments in Statistics*, Academic Press, New York, pp. 83–121.

Hannan, E. J. (1981). Estimating the dimension of a linear system, *Journal of Multivariate Analysis* **11**: 459–473.

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

Hannan, E. J. & Kavalieris, L. (1984). Multivariate linear time series models, *Advances in Applied Probability* **16**: 492–561.

Hillmer, S. C. & Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models, *Journal of the American Statistical Association* **74**: 652–660.

Hogue, A., Magnus, J. & Pesaran, B. (1988). The exact multi-period mean-square forecast error for the first-order autoregressive model, *Journal of Econometrics* **39**: 327–346.

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

Jenkins, G. M. & Alavi, A. S. (1981). Some aspects of modelling and forecasting multivariate time series, *Journal of Time Series Analysis* **2**: 1–47.

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

Johansen, S. (1995b). A statistical analysis of cointegration for I(2) variables, *Econometric Theory* **11**: 25–59.

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

Johansen, S. & Schaumburg, E. (1999). Likelihood analysis of seasonal cointegration, *Journal of Econometrics* **88**: 301–339.

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

Kapetanios, G. (2003). A note on an iterative least-squares estiamtion method for ARMA and VARMA models, *Economics Letters* **79**: 305–312.

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

Kim, T. H., Leybourne, S. J. & Newbold, P. (2004). Asymptotic mean-squared forecast error when an autoregression with linear trend is fitted to data generated by an I(0) or I(1) process, *Journal of Time Series Analysis* **25**: 583–602.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past?, *Journal of Econometrics* **18**: 337–349.

Koreisha, S. G. & Pukkila, T. M. (1987). Identification of nonzero elements in the polynomial matrices of mixed VARMA processes, *Journal of the Royal Statistical Society* **B49**: 112–126.

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

Lütkepohl, H. (1984). Linear transformations of vector arma processes, *Journal of Econometrics* **26**: 283–293.

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

Lütkepohl, H. (1986a). Forecasting temporally aggregated vector ARMA processes, *Journal of Forecasting* **5**: 85–95.

Lütkepohl, H. (1986b). Forecasting vector ARMA processes with systematically missing observations, *Journal of Business & Economic Statistics* **4**: 375–390.

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.

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

Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons, Chichester.

Lütkepohl, H. (2002). Forecasting cointegrated VARMA processes, *in* M. P. Clements & D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 179–205.

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

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

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

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis, *Journal of Business & Economic Statistics* **17**: 129–136.

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

Nicholls, D. F. & Hall, A. D. (1979). The exact likelihood of multivariate autoregressive moving average models, *Biometrika* **66**: 259–264.

Nsiri, S. & Roy, R. (1992). On the identification of ARMA echelon-form models, *Canadian Journal of Statistics* **20**: 369–386.

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

Poskitt, D. S. (1992). Identification of echelon canonical forms for vector linear processes using least squares, *Annals of Statistics* **20**: 196–215.

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

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

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

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

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

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

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.

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

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

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

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

Stram, D. O. & Wei, W. W. S. (1986). Temporal aggregation in the ARIMA process, *Journal of Time Series Analysis* **7**: 279–292.

Telser, L. G. (1967). Discrete samples and moving sums in stationary stochastic processes, *Journal of the American Statistical Association* **62**: 484–499.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series, *Biometrika* **59**: 525–531.

Tiao, G. C. & Box, G. E. P. (1981). Modeling multiple time series with applications, *Journal of the American Statistical Association* **76**: 802–816.

Tiao, G. C. & Guttman, I. (1980). Forecasting contemporal aggregates of multiple time series, *Journal of Econometrics* **12**: 219–230.

Tiao, G. C. & Tsay, R. S. (1983). Multiple time series modeling and extended sample cross-correlations, *Journal of Business & Economic Statistics* **1**: 43–56.

Tiao, G. C. & Tsay, R. S. (1989). Model specification in multivariate time series (with discussion), *Journal of the Royal Statistical Society* **B51**: 157–213.

Tsay, R. S. (1989a). Identifying multivariate time series models, *Journal of Time Series Analysis* **10**: 357–372.

Tsay, R. S. (1989b). Parsimonious parameterization of vector autoregressive moving average models, *Journal of Business & Economic Statistics* **7**: 327–341.

Tunnicliffe Wilson, G. (1973). Estimation of parameters in multivariate time series models, *Journal of the Royal Statistical Society* **B35**: 76–85.

Wallis, K. F. (1977). Multiple time series analysis and the final form of econometric models, *Econometrica* **45**: 1481–1497.

Wei, W. W. S. (1978). Some consequences of temporal aggregation in seasonal time series models, *in* A. Zellner (ed.), *Seasonal Analysis of Economic Time Series*, U.S. Department of Commerce, Bureau of the Census, pp. 433–444.

Wei, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Redwood City, Ca.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models, *Journal of Econometrics* **26**: 271–281.

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

Yap, S. F. & Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association* **90**: 253–267.

Zellner, A. & Palm, F. (1974). Time series analysis and simultaneous equation econometric models, *Journal of Econometrics* **2**: 17–54.