14

Cointegrated VARMA Processes

14.1 Introduction

So far, we have concentrated on stationary VARMA processes for $I(0)$ variables. In this chapter, the variables are allowed to be $I(1)$ and may be cointegrated. As we have seen in Chapter 12, one of the problems in dealing with VARMA models is the nonuniqueness of their parameterization. For inference purposes, it is necessary to focus on a unique representation of a DGP. For stationary VARMA processes, we have considered the echelon form to tackle the identification problem. In the next section, this representation of a VARMA process will be combined with the error correction (EC) form. Thereby it is again possible to separate the long-run cointegration relations from the short-term dynamics. The resulting representation turns out to be a convenient framework for modelling cointegrated variables.

The representation of a VARMA process considered in this chapter is characterized by the cointegrating rank and the Kronecker indices. When these quantities are given, the model can be estimated. Estimation procedures and their asymptotic properties are considered in Section 14.3. A procedure for specifying the Kronecker indices and the cointegrating rank from a given multiple time series will be discussed in Section 14.4. The forecasting aspects of our models will be addressed briefly in Section 14.5 and an example is given in Section 14.6.

In this chapter, an introductory treatment of cointegrated VARMA models is given. The chapter draws on material from Lütkepohl & Claessen (1997) who introduced the error correction echelon form of a VARMA process, Poskitt & Lütkepohl (1995) and Poskitt (2003) who presented estimation and specification procedures for such models, as well as Bartel & Lütkepohl (1998) who explored the small sample properties of some of the procedures. Further references to more advanced treatments of specific issues will be given throughout the chapter.

14.2 The VARMA Framework for $I(1)$ Variables

14.2.1 Levels VARMA Models

In this chapter, it is assumed that some or all of the variables of interest are $I(1)$ variables, whereas the remaining ones are again $I(0)$. Moreover, the variables may be cointegrated. Thus, we consider the situation that was discussed extensively in Part II. In contrast to the framework of that part, we now assume that the DGP of $y_t = (y_{1t}, \dots, y_{Kt})'$ is from the VARMA class,

$$A_0 y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + M_0 u_t + M_1 u_{t-1} + \cdots + M_p u_{t-p}, \ t = 1, 2, \dots, \tag{14.2.1}$$

or

$$A(L) y_t = M(L) u_t, \quad t = 1, 2, \dots, \tag{14.2.2}$$

where $u_t = y_t = 0$ for $t \leq 0$ is assumed for convenience and, as usual, u_t is a white noise process with zero mean and nonsingular, time invariant covariance matrix $E(u_t u_t') = \Sigma_u$. Moreover, in (14.2.2) the VAR operator is

$$A(L) := A_0 - A_1 L - \cdots - A_p L^p$$

and the MA operator is

$$M(L) := M_0 + M_1 L + \cdots + M_p L^p.$$

The zero order matrices A_0 and M_0 are assumed to be nonsingular and some of the coefficient matrices may be zero so that the AR or MA order may actually be less than p. The matrix polynomials are assumed to satisfy

$$\det A(z) \neq 0, \ |z| \leq 1, z \neq 1, \quad \text{and} \quad \det M(z) \neq 0, \ |z| \leq 1. \tag{14.2.3}$$

The second part of this condition is the usual invertibility condition for the MA operator. As in the pure VAR case, we allow the VAR operator $A(z)$ to have roots for $z = 1$ to account for integrated and cointegrated components of y_t. As mentioned previously, all component series are at most $I(1)$, that is, Δy_t is stationary or at least asymptotically stationary.

Notice that there are no deterministic terms in our model. For the introductory treatment of the present chapter, this setup is convenient. Of course, in applied work, deterministic terms will usually be required. Although adding such terms is formally straightforward, it is known from the discussion in Chapter 6, Section 6.4, and Chapter 7, Section 7.2.4, that the implications of such terms in models with integrated variables are more complicated than in the stationary case, in particular with respect to statistical inference.

In this context, it may also be worth emphasizing that the zero initial value assumption ($u_t = y_t = 0$ for $t \leq 0$) is not altogether innocent. Allowing for more general initial values will result in additional complications which we intend to avoid here. Further comments on these issues will be provided later.

Under our assumptions of zero initial values, the process has the pure VAR representation

$$y_t = \sum_{i=1}^{t-1} \Pi_i y_{t-i} + u_t, \tag{14.2.4}$$

where

$$\Pi(z) = \sum_{i=1}^{\infty} \Pi_i z^i = M(z)^{-1} A(z),$$

as in Section 11.3. Notice that the inverse of $M(z)$ exists under our invertibility assumption (14.2.3). The process also has a pure MA representation

$$y_t = \sum_{i=0}^{t-1} \Phi_i u_{t-i},$$

where

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

Here the inverse of $A(z)$ is defined only in a small neighborhood of zero and, in particular,

$$\sum_{i=1}^{n} \Phi_i$$

may diverge for $n \to \infty$. Our MA representation is still valid due to the zero initial value assumption. The VAR and MA representations of the process show that the uniqueness of the VARMA representation can be discussed in the same way as in Chapter 12. We have to find restrictions for $A(L)$ and $M(L)$ such that a unique relation between $[A(L) : M(L)]$ and $M(L)^{-1}A(L)$ is obtained. From Chapter 12, we know already that the echelon form restrictions can be used for that purpose. In the present situation, a slight modification turns out to be useful. We will present it in the next subsection.

If zero initial values are not assumed, the initial values may also help in identifying the model. A discussion of how initial values can contribute to uniquely identifying a VARMA process with $I(1)$ variables is provided by Poskitt (2004). The problem is, however, that the initial values of the u_t will usually be unknown in practice and may not be available for identification.

14.2.2 The Reverse Echelon Form

In order to obtain a unique representation, we use similar restrictions as in Definition 12.2. We will, however, reverse the roles of $A(L)$ and $M(L)$ in this case, as proposed by Lütkepohl & Claessen (1997). In other words, we now impose the restrictions placed on the VAR operator in Definition 12.2 on $M(L)$ and similarly, the restrictions for $M(L)$ in that definition will now be imposed on the VAR operator. This modification will turn out to be convenient in combining the restrictions with the error correction form. We denote the kl-th elements of $A(z)$ and $M(z)$ by $\alpha_{kl}(z)$ and $m_{kl}(z)$, respectively, and impose the constraints specified in the following definition.

Definition 14.1 (*Reverse Echelon Form*)
The VARMA representation (14.2.1) is in *reverse echelon form* if $A(L)$ and $M(L)$ satisfy the following restrictions: The operator $[A(z) : M(z)]$ is left-coprime,

$$m_{kk}(L) = 1 + \sum_{i=1}^{p_k} m_{kk,i}L^i, \quad \text{for } k = 1, \dots, K, \tag{14.2.5}$$

$$m_{kl}(L) = \sum_{i=p_k-p_{kl}+1}^{p_k} m_{kl,i}L^i, \quad \text{for } k \neq l, \tag{14.2.6}$$

and

$$\alpha_{kl}(L) = \alpha_{kl,0} - \sum_{i=1}^{p_k} \alpha_{kl,i}L^i, \quad \text{with } \alpha_{kl,0} = m_{kl,0} \quad \text{for } k, l = 1, \dots, K. \tag{14.2.7}$$

Here

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

The row degrees p_k in this representation are again called *Kronecker indices.* In (14.2.1), $p = \max(p_1, \dots, p_K)$, that is, p is the maximum row degree or Kronecker index. $\text{ARMA}_{RE}(p_1, \dots, p_K)$ denotes a reverse echelon form with Kronecker indices $p_1, \dots, p_K$. ■

It was argued by Poskitt (2004) that the initial conditions may contribute to a unique representation of an integrated VARMA process in such a way that $[A(z) : M(z)]$ does not have to be left-coprime. In that case, the mapping from the set of operators $[A(z) : M(z)]$ to the set of admissible transfer functions $\Phi(z)$ may not be one-to-one, whereas in the present formulation which requires left-coprimeness of $[A(z) : M(z)]$, a one-to-one mapping may be obtained using the reverse echelon form restrictions.

To see the difference to the ARMA_E form discussed in Section 12.1, consider a three-dimensional process with Kronecker indices $(p_1, p_2, p_3) = (1, 2, 1)$ as in (12.1.21)/(12.1.22). In this case,

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

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

$$\begin{aligned} &\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} y_t \\ &= \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} & \alpha_{13,1} \\ \alpha_{21,1} & \alpha_{22,1} & \alpha_{23,1} \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 & 0 \\ \alpha_{21,2} & \alpha_{22,2} & \alpha_{23,2} \\ 0 & 0 & 0 \end{bmatrix} y_{t-2} \\ &\quad + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & m_{12,1} & m_{13,1} \\ 0 & m_{22,1} & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1} \\ &\quad + \begin{bmatrix} 0 & 0 & 0 \\ m_{21,2} & m_{22,2} & m_{23,2} \\ 0 & 0 & 0 \end{bmatrix} u_{t-2}. \end{aligned} \tag{14.2.8}$$

Clearly, in this representation the autoregressive operator is unrestricted except for the constraints imposed by the maximum row degrees or Kronecker indices and the zero order matrix ($A_0 = M_0$), whereas zero restrictions are placed on the moving average coefficient matrices attached to low lags of the u_t. For example, in (14.2.8), there are two zero restrictions on M_1. A comparison with the representation in (12.1.21)/(12.1.22) shows that the restrictions imposed on A_1 in (12.1.21) correspond to those imposed on M_1 in (14.2.8).

14.2.3 The Error Correction Echelon Form

The EC form may be obtained from (14.2.1) by subtracting $A_0 y_{t-1}$ on both sides and rearranging terms, as for the VECM form of a VAR model in Section 6.3:

$$\begin{aligned} A_0 \Delta y_t &= \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p+1} \\ &\quad + M_0 u_t + M_1 u_{t-1} + \cdots + M_p u_{t-p} \end{aligned} \tag{14.2.9}$$

where

$$\mathbf{\Pi} = -(A_0 - A_1 - \cdots - A_p)$$

and

$$\mathbf{\Gamma}_i = -(A_{i+1} + \cdots + A_p), \quad i = 1, \ldots, p-1.$$

Again, $\mathbf{\Pi} y_{t-1}$ is the error correction term and $r = \text{rk}(\mathbf{\Pi})$ is the cointegrating rank of the system.

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

$$\text{rk}(\mathbf{\Pi}) \geq \varrho \tag{14.2.10}$$

(see Problem 14.1). This result has to be taken into account in the procedure for specifying the cointegrating rank of a VARMA system, as discussed in Section 14.4.

An EC model which satisfies the reverse echelon from restrictions will be called an EC-ARMA$_{RE}$ form in the following. As an example, consider again the system (14.2.8). Its EC-ARMA$_{RE}$ form is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} \Delta y_t$$

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

$$+ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha_{32,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & m_{12,1} & m_{13,1} \\ 0 & m_{22,1} & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1}$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ m_{21,2} & m_{22,2} & m_{23,2} \\ 0 & 0 & 0 \end{bmatrix} u_{t-2}.$$

As a further example, consider the three-dimensional ARMA$_{RE}(0,0,1)$ model

$$y_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} y_{t-1}$$

$$+ u_t + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ m_{31,1} & m_{32,1} & m_{33,1} \end{bmatrix} u_{t-1}. \qquad (14.2.11)$$

Its EC-ARMA$_{RE}$ form is

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

Obviously, the rank of

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

is at least 2 and, thus, the cointegrating rank in this case is also at least 2.

Specifying an EC-ARMA$_{RE}$ model requires that the cointegrating rank r is determined, the Kronecker indices $p_1, \dots, p_K$ are obtained and possibly further over identifying zero restrictions are placed on the coefficient matrices $\mathbf{\Gamma}_i$ and M_j. Before we consider strategies for these tasks, we discuss the estimation of EC-ARMA$_{RE}$ models for given cointegrating rank and Kronecker indices in the next section.

14.3 Estimation

14.3.1 Estimation of ARMA$_{RE}$ Models

For given Kronecker indices, an ARMA$_{RE}$ model can be estimated even if the cointegrating rank is unknown. Under Gaussian assumptions, ML estimation can be used. The estimators may be determined by maximizing a log-likelihood function as in (12.2.24),

$$\ln l_0(\gamma, \Sigma_u) = -\frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} u_t(\gamma)' \Sigma_u^{-1} u_t(\gamma), \qquad (14.3.1)$$

where an additive constant is dropped and zero initial conditions are assumed so that

$$u_t(\gamma) = y_t - \sum_{i=1}^{t-1} \Pi_i(\gamma) y_{t-i}.$$

If the initial values are nonzero, $\ln l_0$ is just an approximate log-likelihood. Here $\boldsymbol{\gamma}$ contains all unrestricted autoregressive and moving average parameters, as in Section 12.2, and maximization may proceed by an iterative procedure, as in Section 12.3. Starting values are required for such an algorithm.

The preliminary estimator presented in Chapter 12, Section 12.3.4, can be used for that purpose (e.g., Poskitt (2003)).

As in the case of a cointegrated VAR model, the ML estimators have asymptotic properties which are in some respects different from those obtained in the stationary case. Roughly speaking, they are the same that would be obtained if the true cointegration matrix were known. Thus, if $0 < r < K$, generally the ML estimator $\widetilde{\gamma}$ is consistent and

$$\sqrt{T}(\widetilde{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widetilde{\gamma}}),$$

where the covariance matrix $\Sigma_{\widetilde{\gamma}}$ is singular. These results follow from Yap & Reinsel (1995) and also hold under suitable alternative conditions if y_t is not Gaussian.

If the cointegrating rank is known, it is often desirable to estimate the EC-ARMA$_{RE}$ form of the process because it also provides estimates of the cointegration relations which may well be of major interest. Therefore, estimation of these models will be considered next.

14.3.2 Estimation of EC-ARMA$_{RE}$ Models

If identifying restrictions are imposed on the cointegration matrix, then estimation of the EC-ARMA$_{RE}$ form can also be done by Gaussian ML based on a log-likelihood function similar to (14.3.1), where γ now contains the free parameters of the EC-ARMA$_{RE}$ form. An alternative approach would be to estimate the cointegration matrix β first by reduced rank regression or an EGLS procedure based on a long VAR(n) model, as in Section 7.2. The properties of this estimator will be discussed further in Chapter 15, where fitting approximate VAR models is discussed. For the present purposes, it is sufficient to note that this estimator, say $\widehat{\beta}$, may be used in an ML procedure which estimates the other parameters by maximizing the log-likelihood function conditionally on $\widehat{\beta}$. In other words, the cointegration parameters are fixed at the first stage estimator $\widehat{\beta}$ of the cointegration matrix β and then the log-likelihood is maximized with respect to the other parameters. The resulting estimators have the same asymptotic properties as the full ML estimators (see Yap & Reinsel (1995)).

Starting values for the other parameters that may be used as initial values for an iterative procedure to maximize the log-likelihood function, may be determined in an analogous way as in Section 12.3.4. The short-run and loading parameter estimators have an asymptotic normal distribution which is the same as if the cointegration matrix β were known. This result, of course, is analogous to the pure VAR case considered in Chapter 7 (see also Phillips (1991, Remark (n)) and Yap & Reinsel (1995)).

The previous discussion assumes given Kronecker indices and possibly a known cointegrating rank. Statistical procedures for specifying these quantities will be discussed next.

14.4 Specification of EC-ARMA$_{RE}$ Models

14.4.1 Specification of Kronecker Indices

For stationary processes, proposals for specifying the Kronecker indices of an ARMA$_E$ model were discussed in Section 13.3. The strategies for specifying the Kronecker indices of cointegrated ARMA$_{RE}$ forms presented in this section were proposed by Poskitt & Lütkepohl (1995) and Poskitt (2003). In the latter article, it is also argued that they result in consistent estimators of the Kronecker indices under suitable conditions. In a simulation study, Bartel & Lütkepohl (1998) found that they worked reasonably well in small samples, at least for the processes explored in their Monte Carlo study.

The specification procedures may be partitioned in two stages. The first stage is the same as for the procedures for stationary processes discussed in Section 13.3.2 and consists of fitting a long autoregression by least squares in order to obtain estimates of the unobservable innovations u_t, $t = 1, \dots, T$.

STAGE I: Use multivariate LS estimation to fit a long VAR(n) process to the data to obtain residuals $\widehat{u}_t(n)$. ■

These residuals are then substituted for the unknown lagged u_t's in the individual equations of an ARMA$_{RE}$ form which may then be estimated by linear LS procedures. Based on the equations estimated in this way, a choice of the Kronecker indices is made using model selection criteria. Poskitt & Lütkepohl (1995), Guo, Huang & Hannan (1990), and Huang & Guo (1990) showed that the estimated residuals $\widehat{u}_t(n)$ are "good" estimates of the true residuals if n approaches infinity at a suitable rate, as T goes to infinity (see Lemma 3.1 of Poskitt & Lütkepohl (1995) for details).

The methods presented in the following differ in the way they choose the Kronecker indices in the next step. An obvious idea may be to search over all models associated with Kronecker indices

$$\{(p_1, \dots, p_K) | 0 \le p_k \le p_{\max}, k = 1, \dots, K\}$$

for some prespecified upper bound $p_{\max}$ and choose the set of Kronecker indices which optimizes some model selection criterion, as in Section 13.3.2 for the stationary case. The two procedures presented in the following are more efficient computationally and they are similar to Poskitt's procedure presented in Section 13.3.4. The first variant uses linear regressions to estimate the individual equations separately for different lag lengths. A choice of the optimal lag length is then based on some prespecified criterion similar to those considered for the stationary case. The following formal description of the procedure is taken from Poskitt & Lütkepohl (1995).

STAGE II: Proceed in the following steps.

(ia) For $m = 0$, set $T\widetilde{\sigma}_k^2(m)$ equal to the residual sum of squares from the regression of y_{kt} on a constant and $(y_{jt} - \widehat{u}_{jt}(n))$, $j = 1, \dots, K$, $j \neq k$. For $m = 1, \dots, p_{\max} \leq n$, regress y_{kt} on a constant, $(y_{jt} - \widehat{u}_{jt}(n))$, $j = 1, \dots, K$, $j \neq k$, and y_{t-s} and $\widehat{u}_{t-s}(n)$, $s = 1, \dots, m$, and determine the residual sums of squares, $T\widetilde{\sigma}_k^2(m)$, for $k = 1, \dots, K$.

(ib) For $k = 1, \dots, K$, compute a selection criterion of the form

$$\mathrm{Cr}_k(m) = \ln \widetilde{\sigma}_k^2(m) + c_T m/T, \quad m = 0, 1, \dots, p_{\max},$$

where c_T is a function of T which will be specified later.

(ii) Set the estimate of the k-th Kronecker index equal to

$$\widehat{p}_k = \underset{0 \leq m \leq p_{\max}}{\arg\min}\ \mathrm{Cr}_k(m), \quad k = 1, \dots, K.$$

■

In the regressions in Step (ia), restrictions from the echelon structure are not explicitly taken into account, because for each value of m, the algorithm implicitly assumes that the current index under consideration is the smallest and, thus, no restrictions are imported from other equations. Still, the k-th equation will be misspecified whenever m is less than the true Kronecker index because in that case, lagged values required for a correct specification are omitted. On the other hand, if m is greater than the true Kronecker index, the k-th equation will be correctly specified but may include redundant parameters and variables. Therefore, it is intuitively plausible that for an appropriate choice of c_T, the criterion function $\mathrm{Cr}_k(m)$ will be minimized asymptotically when m is equal to the true Kronecker index. For practical purposes, possible choices of c_T are $c_T = n \ln T$ or $c_T = n^2$.

At Stage II, values for n, $p_{\max}$, and c_T have to be chosen. The theoretical consistency results stated in Poskitt (2003) are quite general and provide an asymptotic justification for many different values of these quantities. The following choices may be considered in practice:

- Choose n by AIC or use $n = \max\{(\ln T)^a, \widehat{p}(\mathrm{AIC})\}$, where $a > 1$.
- Choose $p_{\max} = \frac{1}{2}n$.
- Choose $c_T = n \ln T$ or $c_T = n^2$.

Poskitt & Lütkepohl (1995) also proposed a modification of Stage II which permits to take into account coefficient restrictions derived from those equations in the system which have smaller Kronecker indices. In that modification, after running through Stage II for the first time, we fix the smallest Kronecker index and repeat Stage II, but search only those equations which are found to have indices larger than the smallest. In this second application of Stage II, the restrictions implied by the smallest Kronecker index found in the first round are taken into account when the second smallest index is determined. We proceed in this way by fixing the smallest Kronecker index found in each successive round until all the Kronecker indices have been specified. In this procedure, the variables are ordered in such a way that the Kronecker indices of the final system are ordered from largest to smallest. That is, the variable whose equation is associated with the smallest Kronecker index is placed last in the list of variables. The one with the second smallest Kronecker index is assigned the next to the last place and so on. For details, see Poskitt & Lütkepohl (1995) and Poskitt (2003).

It should be understood that the Kronecker indices found in such a procedure for a given time series of finite length can only be expected to be a reasonable starting point for a more refined analysis of the system under consideration. Based on the specified Kronecker indices, a more efficient procedure for estimating the parameters may be applied and the model may be modified subsequently.

So far, we have not discussed the choice of the cointegrating rank. In practice, of course, this quantity is unlikely to be known. Comments on the estimation of r will be given in the following subsection.

14.4.2 Specification of the Cointegrating Rank

Saikkonen & Luukkonen (1997) and Lütkepohl & Saikkonen (1999b) showed that Johansen's LR tests for the cointegrating rank (see Section 8.2) maintain their asymptotic properties even if a finite order VAR process is fitted although the true underlying process has an infinite order VAR structure. Consequently, these tests may be applied at Stage I of the present specification procedure. The cointegrating rank is then determined independently of the Kronecker indices. Alternatively, Yap & Reinsel (1995) extended the likelihood ratio principle to VARMA processes and developed cointegration rank tests under the assumption that identified versions of $A(z)$ and $M(z)$ are used. Thus, these tests may be applied once the Kronecker indices have been specified. Whatever approach is adopted, for our purposes the following modification is noteworthy.

If a Kronecker index $p_k = 0$, the variable y_{kt} inherits all of its dynamics from other variables in the system and it is known from (14.2.10) that the cointegrating rank $r \geq \varrho$, the number of zero Kronecker indices. Hence, the testing procedure proceeds by considering only null hypotheses where r is greater than or equal to ϱ. In other words, the following sequence of null hypotheses is tested: $H_0 : r = \varrho$, $H_0 : r = \varrho + 1$, …, $H_0 : r = K - 1$. The estimator of r is chosen such that it is the smallest value for which H_0 cannot be rejected.

Once a model has been estimated, some checks for model adequacy are in order and possible further model reductions or modifications may be called for. For instance, insignificant parameter estimates may be restricted to zero. Here it is convenient that the t-ratios of the short-run parameters have their usual asymptotic standard normal distributions under the null hypothesis, due to the asymptotic normal distribution of the ML estimators. Thus, they can be used for significance tests in the usual way and may help to place over identifying restrictions on the parameters. Moreover, a detailed analysis of the residual properties should be performed to reveal possible model deficiencies. The checks for model adequacy described in Chapters 4 and 8 can be used here as well with appropriate modifications.

14.5 Forecasting Cointegrated VARMA Processes

Forecasting cointegrated VARMA processes proceeds completely analogously to forecasting stationary VARMA processes. The same formulas can be used. Like for pure VAR models, the properties of the forecasts will be different, however. In particular, the forecast error covariance matrices will be unbounded for increasing forecast horizon. Hence, also forecast intervals will be unbounded in length. In this respect, the properties of the forecasts are analogous to those of cointegrated pure VAR processes. The reader is referred to Section 6.5 for details.

14.6 An Example

For illustrative purposes, we use an example from Lütkepohl & Claessen (1997), based on the U.S. macroeconomic data which were also considered in Section 7.4.3. The data are available in File E3. It consists of 136 quarterly observations for the years 1954.1 to 1983.4 of the real money stock M1 (y_{1t}), GNP in billions of 1982 dollars (y_{2t}), the discount interest rate on new issues of 91-day treasury bills (y_{3t}), and the yield on long term (20 years) treasury bonds (y_{4t}). Logarithms of seasonally adjusted GNP and M1 data are used. Thus, y_t is a four-dimensional vector. Notice that we do not use the full sample period covered in File E3 but truncate the data for the last four years. The reason is that in the exercises readers are asked to perform a forecast comparison based on the model presented in the following. The data for the years 1984–1987 are set aside for this comparison.

Following the procedure outlined in Section 14.4.2, the cointegrating rank may be determined with LR type tests applied to a long VAR model. After running through an extensive specification procedure, Lütkepohl & Claessen (1997) finally specified an EC-ARMA$_{RE}(2,1,1,1)$ model with cointegrating rank 1 for the data generation process of this system and obtained the following estimated model

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \underset{(.117)}{-.509} & 1 & 0 & 0 \\ \underset{(.105)}{-.099} & 0 & 1 & 0 \\ \underset{(.043)}{.084} & 0 & 0 & 1 \end{bmatrix} \Delta y_t$$

$$
= \begin{bmatrix} \underset{(.035)}{.091} \\ \underset{(.060)}{.216} \\ \underset{(.056)}{.190} \\ \underset{(.022)}{.055} \end{bmatrix} + \begin{bmatrix} \underset{(.015)}{-.039} \\ \underset{(.026)}{-.090} \\ \underset{(.024)}{-.082} \\ \underset{(.010)}{-.023} \end{bmatrix} [1, -.343, -16.72, 19.35] y_{t-1}
$$

$$
+ \begin{bmatrix} \underset{(.084)}{.810} & \underset{(.069)}{.074} & \underset{(.101)}{-.682} & \underset{(.192)}{-.507} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Delta y_{t-1}
$$

$$
+ \begin{bmatrix} 1 & 0 & 0 & 0 \\ \underset{(.117)}{-.509} & 1 & 0 & 0 \\ \underset{(.105)}{-.099} & 0 & 1 & 0 \\ \underset{(.043)}{.084} & 0 & 0 & 1 \end{bmatrix} u_t
$$

$$
+ \begin{bmatrix} \underset{(.109)}{-.478} & 0 & 0 & 0 \\ \underset{(.113)}{-.101} & \underset{(.091)}{.006} & \underset{(.144)}{.339} & \underset{(.258)}{.898} \\ \underset{(.084)}{.160} & \underset{(.070)}{.123} & \underset{(.106)}{.377} & \underset{(.202)}{.154} \\ \underset{(.045)}{.037} & \underset{(.036)}{.043} & \underset{(.057)}{.093} & \underset{(.103)}{-.070} \end{bmatrix} u_{t-1}
$$

$$
+ \begin{bmatrix} \underset{(.091)}{.082} & \underset{(.073)}{-.022} & \underset{(.123)}{.205} & \underset{(.220)}{.646} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} u_{t-2}. \qquad (14.6.1)
$$

Estimated standard errors are given in parentheses. The cointegration vector $\widehat{\beta}' = [1, -.343, -16.72, 19.35]$ was obtained by estimating a VECM with one lagged difference of y_t and with cointegrating rank 1, using the ML procedure presented in Section 7.2.3 and normalizing the first element of $\widehat{\beta}$ to be 1.

Some of the parameter values in (14.6.1) are quite small compared to their estimated standard errors. In particular, some of them are not significant under a two-standard error criterion. Therefore, zero restrictions were placed on the coefficients and the following estimated model was obtained:

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ \underset{(.107)}{-.476} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \underset{(.032)}{.145} & 0 & 0 & 1 \end{bmatrix} \Delta y_t
$$

$$
= \begin{bmatrix} \underset{(.035)}{.094} \\ \underset{(.056)}{.219} \\ \underset{(.057)}{.207} \\ \underset{(.020)}{.069} \end{bmatrix} + \begin{bmatrix} \underset{(.016)}{-.042} \\ \underset{(.026)}{-.096} \\ \underset{(.026)}{-.094} \\ \underset{(.009)}{-.031} \end{bmatrix} [1, -.343, -16.72, 19.35] y_{t-1}
$$

$$
+ \begin{bmatrix} \underset{(.063)}{.772} & \underset{(.067)}{.087} & \underset{(.100)}{.788} & \underset{(.195)}{.198} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Delta y_{t-1}
$$

$$
+ \begin{bmatrix} 1 & 0 & 0 & 0 \\ \underset{(.107)}{-.476} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \underset{(.032)}{.145} & 0 & 0 & 1 \end{bmatrix} u_t
$$

$$
+ \begin{bmatrix} \underset{(.082)}{-.640} & 0 & 0 & \\ 0 & 0 & 0 & \underset{(.297)}{1.105} \\ \underset{(.104)}{.331} & 0 & \underset{(.083)}{.339} & 0 \\ \underset{(.042)}{.162} & 0 & \underset{(.054)}{.110} & \underset{(.095)}{-.323} \end{bmatrix} u_{t-1}
$$

$$
+ \begin{bmatrix} \underset{(.081)}{.107} & 0 & \underset{(.114)}{.233} & \underset{(.196)}{.984} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} u_{t-2}. \qquad (14.6.2)
$$

This example is just meant to illustrate that the procedures presented in this chapter are indeed feasible in practice. The reader is encouraged to perform a forecast comparison of the model presented here with pure VAR models and VECMs for the data (see Problem 14.6).

14.7 Exercises

14.7.1 Algebraic Exercises

Problem 14.1

Show that in the model (14.2.9), $\text{rk}(\mathbf{\Pi}) \geq \varrho$, where ϱ is the number of Kronecker indices which are zero. (Hint: Consider the matrix $A_0 - A_1$).

Problem 14.2
Write down an ARMA$_{RE}(2,1,2)$ model explicitly in matrix form and also write down the corresponding EC-ARMA$_{RE}$ form.

Problem 14.3
Consider the following EC-ARMA$_{RE}$ model:

$$\begin{aligned}\begin{bmatrix} 1 & 0 \\ \alpha_{21,0} & 1 \end{bmatrix} \Delta y_t &= \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} y_{t-1} + \begin{bmatrix} \gamma_{11,1} & \gamma_{12,1} \\ 0 & 0 \end{bmatrix} \Delta y_{t-1} \\ &\quad + \begin{bmatrix} 1 & 0 \\ \alpha_{21,0} & 1 \end{bmatrix} u_t + \begin{bmatrix} m_{11,1} & 0 \\ m_{21,1} & m_{22,1} \end{bmatrix} u_{t-1}.\end{aligned}$$

(a) Write the model in ARMA$_{RE}$ form.
(b) Specify the Kronecker indices.
(c) How many over-identifying restrictions are present in this model?
(d) Write the model in pure VAR form.

14.7.2 Numerical Exercises

The following problems are based on the U.S. data given in File E3, as described in Section 14.6. The variables are defined in the same way as in that section. Thus, a system of dimension four is considered.

Problem 14.4
Fit a pure VAR model to the four-dimensional data set without considering integration and cointegration properties of the variables. Use only the data for 1954.1–1983.4 for modelling and estimation. Compute forecasts from the model for the period 1984.1–1987.4.

Problem 14.5
Use the following steps in constructing VECMs for the period 1954.1–1983.4 and computing forecasts of the four variables for the period 1984.1–1987.4.

(a) Determine the cointegrating rank of the system.
(b) Estimate the cointegration relation(s) with the reduced rank ML and the EGLS methods discussed in Chapter 7.
(c) Construct subset VECMs based on the estimated cointegration relations from the previous step.
(d) Confirm that the models obtained in the previous steps are adequate representations of the data generation process.
(e) Compute forecasts from your model for the period 1984.1–1987.4.

Problem 14.6
Compare the forecasts obtained in Problems 14.4 and 14.5 with those from the EC-ARMA$_{RE}$ model (14.6.2) on the basis of the MSEs. Discuss the results. (Hint: See Lütkepohl & Claessen (1997).)