13

Specification and Checking the Adequacy of VARMA Models

13.1 Introduction

A great number of strategies has been suggested for specifying VARMA models. There is not a single one that has become a standard like the Box & Jenkins (1976) approach in the univariate case. None of the multivariate procedures is in widespread use for modelling moderate or high-dimensional economic time series. Some are mainly based on a subjective assessment of certain characteristics of a process such as the autocorrelations and partial autocorrelations. A decision on specific orders and constraints on the coefficient matrices is then based on these quantities. Other methods rely on a mixture of statistical testing, use of model selection criteria and personal judgement of the analyst. Again other procedures are based predominantly on statistical model selection criteria and, in principle, they could be performed automatically by a computer. Automatic procedures have the advantage that their statistical properties can possibly be derived rigorously. In actual applications, some kind of mixture of different approaches is often used. In other words, the expertise and prior knowledge of an analyst will usually not be abolished in favor of purely statistical procedures. Models suggested by different types of criteria and procedures will be judged and evaluated by an expert before one or more candidates are put to a specific use such as forecasting. The large amount of information in a number of moderately long time series makes it usually necessary to condense the information considerably before essential features of a system become visible.

In the following, we will outline procedures for specifying the final equations form and the echelon form of a VARMA process. We do not claim that these procedures are superior to other approaches. They are just meant to illustrate what is involved in the specification of VARMA models. The specification strategies for both forms could be turned into automatic algorithms. On the other hand, they also leave room for human intervention if desired. In Section 13.4, some references for other specification strategies are given

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*

and model checking is discussed briefly in Section 13.5. A critique of VARMA modelling is given in Section 13.6.

13.2 Specification of the Final Equations Form

13.2.1 A Specification Procedure

Historically, procedures for specifying final equations VARMA representations were among the earlier strategies for modelling systems of economic time series (see, for example, Zellner & Palm (1974), Wallis (1977)). The objective is to find the orders p and q of the representation

$$\alpha(L)y_t = M(L)u_t, \tag{13.2.1}$$

where

$$\alpha(L) := 1 - \alpha_1 L - \cdots - \alpha_p L^p$$

is a (1×1) scalar operator,

$$M(L) := I_K + M_1 L + \cdots + M_q L^q$$

is a $(K \times K)$ matrix operator and it is assumed that the process mean has been removed in a previous step of the analysis.

If a K-dimensional system $y_t = (y_{1t}, \ldots, y_{Kt})'$ has a VARMA representation of the form (13.2.1), then it follows that each component series has a univariate ARMA representation

$$\alpha(L)y_{kt} = \overline{m}_k(L)v_{kt}, \quad k = 1, \ldots, K,$$

where $\overline{m}_k(L)$ is an operator of degree at most q because the k-th row of $M(L)u_t$ is

$$m_{k1}(L)u_{1t} + \cdots + m_{kK}(L)u_{Kt}.$$

In other words, it is a sum of MA(q) processes which is known to have an MA representation of degree at most q (see Proposition 11.1). Thus, each component series of y_t has the same AR operator and an MA operator of degree at most q. In general, at least one of the component series will have MA degree q because a reduction of the MA order of all component series requires a very special set of parameters which is not regarded as likely in practice. This fact is used in specifying the final form VARMA representation by first determining univariate component models and then putting them together in a joint model. Specifically, the following specification strategy is used.

STAGE I: Specify univariate models

$$\alpha_k(L)y_{kt} = m_k(L)v_{kt}$$

for the components of y_t. Here

$$\alpha_k(L) := 1 - \alpha_{k1}L - \cdots - \alpha_{kp_k}L^{p_k}$$

is of order p_k,

$$m_k(L) := 1 + m_{k1}L + \cdots + m_{kq_k}L^{q_k}$$

is of order q_k, and v_{kt} is a univariate white noise process. ■

The Box & Jenkins (1976) strategy for specifying univariate ARMA models may be used at this stage. Alternatively, some automatic procedure or criterion such as the one proposed by Hannan & Rissanen (1982) or Poskitt (1987) may be applied.

STAGE II: Determine a common AR operator $\alpha(L)$ for all component processes, specify the corresponding MA orders and choose the degree q of the joint MA operator as the maximum of the individual MA degrees obtained in this way. ■

At this stage, a common AR operator may, for example, be obtained as the product of the individual operators, that is,

$$\alpha(L) = \alpha_1(L) \cdots \alpha_K(L).$$

In this case, the k-th component process is multiplied by

$$\prod_{i=1, i\neq k}^{K} \alpha_i(L)$$

and $\alpha(L)$ has degree $p = \sum_{i=1}^{K} p_i$, while the MA operator

$$\overline{m}_k(L) = m_k(L) \prod_{i=1, i\neq k}^{K} \alpha_i(L)$$

has degree

$$q_k + \sum_{i=1, i\neq k}^{K} p_i.$$

The joint MA operator of the VARMA representation (13.2.1) is then assumed to have degree

$$\max_k \left(q_k + \sum_{i=1, i\neq k}^{K} p_i \right). \tag{13.2.2}$$

Of course, the $\alpha_k(L)$, $k = 1, \ldots, K$, may have common factors. In that case, a joint AR operator $\alpha(L)$ with degree much lower than $\sum_{i=1}^{K} p_i$ may be possible. Correspondingly, the joint MA operator may have degree lower than (13.2.2). Suppose, for instance, that $K = 3$ and

$$\begin{aligned} &\alpha_1(L) = 1 - \alpha_{11}L, && m_1(L) = 1 + m_{11}L, \\ &\alpha_2(L) = 1 - \alpha_{21}L - \alpha_{22}L^2, && m_2(L) = 1 + m_{21}L, \\ &\alpha_3(L) = 1 - \alpha_{31}L, && m_3(L) = 1. \end{aligned}$$

Now a joint AR operator is

$$\alpha(L) = \alpha_1(L)\alpha_2(L)\alpha_3(L),$$

which has degree 4. However, if $\alpha_2(L)$ can be factored as

$$\alpha_2(L) = (1 - \alpha_{11}L)(1 - \alpha_{31}L) = \alpha_1(L)\alpha_3(L),$$

then a common AR operator $\alpha(L) = \alpha_2(L)$ may be chosen and we get univariate models

$$\begin{aligned} &\alpha(L)y_{1t} = \alpha_3(L)m_1(L)v_{1t} && [\text{ARMA}(2,2)], \\ &\alpha(L)y_{2t} = m_2(L)v_{2t} && [\text{ARMA}(2,1)], \\ &\alpha(L)y_{3t} = \alpha_1(L)m_3(L)v_{3t} && [\text{ARMA}(2,1)]. \end{aligned} \tag{13.2.3}$$

The maximum of the individual MA degrees is chosen as the joint MA degree, that is, $q = 2$ and, of course,

$$p = \text{degree}(\alpha(L)) = 2.$$

A problem that should be noticed from this discussion and example is that the degrees p and q determined in this way may be quite large. It is conceivable that $p = \sum_{i=1}^{K} p_i$ is the smallest possible AR order for the final equations form representation and the corresponding MA degree may be quite substantial too. This, clearly, can be a disadvantage as unduely many parameters can cause trouble in a final estimation algorithm and may lead to imprecise forecasts and impulse responses.

Often it may be possible to impose restrictions on the AR and MA operators in (13.2.1). This modification may either be done in a third stage of the procedure or it may be incorporated in Stages I and/or II, depending on the type of information available. Restrictions may be obtained with the help of statistical tools such as testing the significance of single coefficients or groups of parameters. Alternatively, restrictions may be implied by subject matter theory. Zellner & Palm (1974) give a detailed example where both types of restrictions are used.

Perhaps because of the potentially great number of parameters, final form modelling has not become very popular. It can only be recommended if it results in a reasonably parsimonious parameterization.

13.2.2 An Example

For illustrative purposes, we consider a bivariate system consisting of first differences of logarithms of income (y_1) and consumption (y_2). We use again the data from File E1 up to the fourth quarter of 1978. If the 3-dimensional system involving investment in addition really were generated by a VAR(2) process, as assumed in Chapters 3 and 4, it is quite possible that the subprocess consisting of income and consumption only has a mixed VARMA generation process with nontrivial MA part (see Section 11.6.1). Moreover, the marginal univariate processes for y_1 and y_2 may be of a mixed ARMA type. However, we found that the subset AR(3) models (with standard errors in parentheses)

$$(1 - \underset{(.113)}{.245}\, L^3)y_{1t} = \underset{(.003)}{.015} + v_{1t} \tag{13.2.4}$$

and

$$(1 - \underset{(.111)}{.309}\, L^2 - \underset{(.111)}{.187}\, L^3)y_{2t} = \underset{(.004)}{.010} + v_{2t} \tag{13.2.5}$$

fit the data quite well. For illustrative purposes, we will therefore proceed from these models. The reader may try to find better models and repeat the analysis with them.

Generally, a (1×1) scalar operator

$$\gamma(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p$$

of degree p can be factored in p components,

$$\gamma(L) = (1 - \lambda_1 L) \cdots (1 - \lambda_p L),$$

where $\lambda_1, \ldots, \lambda_p$ are the reciprocals of the roots of $\gamma(z)$. Thus, the two AR operators from (13.2.4) and (13.2.5) can be factored as

$$\begin{aligned} \alpha_1(L) &= 1 - .245L^3 \\ &= (1 - .626L)(1 + (.313 + .542i)L)(1 + (.313 - .542i)L) \end{aligned} \tag{13.2.6}$$

and

$$\begin{aligned} \alpha_2(L) &= 1 - .309L^2 - .187L^3 \\ &= (1 - .747L)(1 + (.374 + .332i)L)(1 + (.374 - .332i)L), \end{aligned} \tag{13.2.7}$$

where i denotes the imaginary part of the complex numbers. None of the factors in (13.2.6) is very close to any of the factors in (13.2.7). Thus, models with common AR operator may be of the form

$$\alpha_1(L)\alpha_2(L)y_{1t} = \alpha_2(L)v_{1t}$$

and

$$\alpha_1(L)\alpha_2(L)y_{2t} = \alpha_1(L)v_{2t}.$$

With the arguments of the previous subsection, the resulting bivariate final equations model is a VARMA$(6, 3)$ process,

$$\alpha_1(L)\alpha_2(L)\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = (I_3 + M_1L + M_2L^2 + M_3L^3)\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{13.2.8}$$

Obviously, this model involves very many parameters and is therefore unattractive. In fact, such a heavily parameterized model may cause numerical problems when full maximum likelihood estimation is attempted. It is possible, if not likely, that some parameters turn out to be insignificant and could be set to zero. However, the significance of parameters is commonly judged on the basis of their standard errors or t-ratios. These quantities become available in the ML estimation round which, as we have argued, may be problematic in the present case.

Given the estimation uncertainty, one may argue that the real factors in the operators $\alpha_1(L)$ and $\alpha_2(L)$ may be identical. Proceeding under that assumption results in a VARMA$(5, 2)$ final equations form. Such a model is more parsimonious and has therefore more appeal than (13.2.8). Still it involves a considerable number of parameters. This example illustrates why final equations modelling, although relatively simple, does not enjoy much popularity. For higher-dimensional models, the problem of heavy parameterization becomes even more severe because the number of parameters is likely to increase rapidly with the dimension of the system. We will now present procedures for specifying echelon forms.

13.3 Specification of Echelon Forms

In specifying an echelon VARMA representation, the objective is to find the Kronecker indices and possibly impose some further restrictions on the parameters. For a K-dimensional process, there are K Kronecker indices. Different strategies have been proposed for their specification. We will discuss some of them in the following. Once the Kronecker indices are determined, further restrictions may be imposed, for instance, on the basis of significance tests for individual coefficients or groups of parameters.

In the first subsection below, we will discuss a procedure for specifying the Kronecker indices which is usually not feasible in practice. It is nevertheless useful to study that procedure because the feasible strategies considered in Subsections 13.3.2–13.3.4 may be regarded as approximations or short-cuts of that procedure with similar asymptotic properties. In Subsection 13.3.2, we present a procedure which is easy to carry out for systems with small Kronecker indices and low dimensions. It is quite costly for higher-dimensional systems, though. For such systems a specification strategy inspired by Hannan & Kavalieris (1984) or a procedure due to Poskitt (1992) may be more

appealing. These approaches are considered in Subsections 13.3.3 and 13.3.4, respectively. The material discussed in this section is covered in more depth and more rigorously in Hannan & Deistler (1988, Chapters 5, 6, 7) and Poskitt (1992).

13.3.1 A Procedure for Small Systems

If it is known that the generation process of a given K-dimensional multiple time series admits an echelon VARMA representation with Kronecker indices $p_k \leq p_{\max}$, $k = 1, \ldots, K$, where $p_{\max}$ is a prespecified number, then, in theory, it is possible to evaluate the maximum log-likelihood for all sets of Kronecker indices $\mathbf{p} = (p_1, \ldots, p_K)$ with $p_k \leq p_{\max}$ and choose the set $\widehat{\mathbf{p}}$ that optimizes a specific criterion. This approach is completely analogous to the specification of the VAR order in the finite order VAR case considered in Section 4.3. In that section, we have discussed the possibility to consistently estimate the VAR order with such an approach. It turns out that a similar result can be obtained for the present more general VARMA case.

Before we give further details, it may be worth emphasizing, however, that in the VARMA case, such a specification strategy is generally not feasible in practice because the maximization of the log-likelihood is usually quite costly and, for systems with moderate or high dimensions, an enormous number of likelihood maximizations would be required. For instance, for a five-dimensional system, evaluating the maximum of the log-likelihood for all vectors of Kronecker indices $\mathbf{p} = (p_1, \ldots, p_5)$ with $p_k \leq 8$ requires $9^5 = 59,049$ likelihood maximizations. Despite this practical problem, we discuss the theoretical properties of this procedure to provide a basis for the following subsections.

Let us denote by $\widetilde{\Sigma}(\mathbf{p})$ the ML estimator of the white noise covariance matrix Σ_u obtained for a set of Kronecker indices $\mathbf{p}$. Furthermore, let

$$\mathrm{Cr}(\mathbf{p}) := \ln |\widetilde{\Sigma}_u(\mathbf{p})| + c_T d(\mathbf{p})/T \tag{13.3.1}$$

be a criterion to be minimized over all sets of Kronecker indices $\mathbf{p} = (p_1, \ldots, p_K)$, $p_k \leq p_{\max}$. Here $d(\mathbf{p})$ is the number of freely varying parameters in the $\mathrm{ARMA}_E(\mathbf{p})$ form. For example, for a bivariate system with Kronecker indices $\mathbf{p} = (p_1, p_2) = (1, 0)$, the $\mathrm{ARMA}_E(1, 0)$ form is

$$\begin{bmatrix} 1 & 0 \\ -\alpha_{21,0} & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_{11,1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -\alpha_{21,0} & 1 \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} + \begin{bmatrix} m_{11,1} & m_{12,1} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}.$$

Thus, $d(1, 0) = 4$. In (13.3.1), c_T is a sequence indexed by the sample size T.

In general, if models are included in the search procedure for which all Kronecker indices exceed the true ones, the estimation of unidentified models is required for which cancellation of the VAR and MA operators is possible.

This over-specification is not necessarily a problem for evaluating the criterion in (13.3.1) because we only need the maximum log-likelihood or rather $\ln|\widetilde{\Sigma}_u(\mathbf{p})|$ in that criterion. That quantity can be determined even if the corresponding VARMA coefficients are meaningless. The coefficients cannot and should not be interpreted, however.

Note that the criterion (13.3.1) is very similar to that considered in Proposition 4.2 of Chapter 4. In that proposition, the consistency or inconsistency of a criterion is seen to depend on the choice of the sequence c_T. Hannan (1981) and Hannan & Deistler (1988, Chapter 5, Section 5) showed that a criterion such as the one in (13.3.1) provides a consistent estimator of the true set of Kronecker indices if c_T is a nondecreasing function of T satisfying

$$c_T \to \infty \quad \text{and} \quad c_T/T \to 0 \quad \text{as} \quad T \to \infty, \tag{13.3.2}$$

and the true data generation process satisfies some weak conditions. If, in addition,

$$c_T/2\ln\ln T > 1 \tag{13.3.3}$$

eventually as $T \to \infty$, the procedure provides a strongly consistent estimator of the true Kronecker indices. The conditions for the VARMA process are, for instance, satisfied if the white noise process u_t is identically distributed standard white noise (see Definition 3.1) and the true data generation process admits a stable and invertible ARMA_E representation with Kronecker indices not greater than $p_{\max}$. This result extends Proposition 4.2 to the VARMA case.

Implications of this result are that the Schwarz criterion with $c_T = \ln T$,

$$\mathrm{SC}(\mathbf{p}) := \ln|\widetilde{\Sigma}_u(\mathbf{p})| + d(\mathbf{p})\ln T/T, \tag{13.3.4}$$

is strongly consistent and that the Hannan-Quinn criterion, using the borderline penalty term $c_T = 2\ln\ln T$,

$$\mathrm{HQ}(\mathbf{p}) := \ln|\widetilde{\Sigma}_u(\mathbf{p})| + 2d(\mathbf{p})\ln\ln T/T, \tag{13.3.5}$$

is consistent. Hannan & Deistler (1988) also showed that

$$\mathrm{AIC}(\mathbf{p}) := \ln|\widetilde{\Sigma}_u(\mathbf{p})| + 2d(\mathbf{p})/T \tag{13.3.6}$$

with $c_T = 2$ is not a consistent criterion. Again these results are similar to those for the finite order VAR case.

As in that case, it is worth emphasizing that these results do not necessarily imply the inferiority of AIC or HQ. In small samples, these criteria may be preferable. They may, in fact, provide superior models for a specific analysis of interest. Also, in practice, the actual data generation mechanism will usually not really admit a VARMA representation. Recall that the best we can hope for is that our model is a good approximation to the true data generation

process. In that case, the relevance of the consistency property is of course doubtful.

Again, the specification strategy presented in the foregoing is not likely to have much practical appeal as it is computationally too burdensome. In the next subsections, more practical modifications are discussed.

13.3.2 A Full Search Procedure Based on Linear Least Squares Computations

The Procedure

A major obstacle for using the procedure described in the previous subsection is the requirement to maximize the log-likelihood various times. This maximization is costly because for mixed VARMA models the log-likelihood function is nonlinear in the parameters and iterative optimization algorithms have to be employed. Because we just need an estimator of Σ_u for the evaluation of model selection criteria such as (13.3.1), an obvious modification of the procedure would use an estimator that avoids the nonlinear optimization problem. Such an estimator may be obtained from the preliminary estimation procedure described in Chapter 12, Section 12.3.4. Therefore, a specification of the Kronecker indices may proceed in the following stages.

STAGE I: Fit a long VAR process of order n, say, to the data and obtain the estimated residual vectors $\widehat{u}_t(n)$, $t = 1, \ldots, T$. ■

The choice of n could be based on an order selection criterion such as AIC. In any case, n has to be greater than the largest Kronecker index $p_{\max}$ to be considered in the next stage of the procedure.

STAGE II: Using the residuals $\widehat{u}_t(n)$ from Stage I, compute the preliminary estimator of Section 12.3.4 for all sets of Kronecker indices $\mathbf{p}$ with $p_k \leq p_{\max}$, where the latter number is a prespecified upper bound for the Kronecker indices. Determine all corresponding estimators $\widetilde{\Sigma}_u(\mathbf{p})$ based on the residuals of the preliminary estimations (see (12.3.24)). (Here we suppress the order n from the first stage for notational convenience because the same n is used for all $\widetilde{\Sigma}_u(\mathbf{p})$ at this stage.) Choose the estimator $\widehat{\mathbf{p}}$ which minimizes a prespecified criterion of the form (13.3.1). ■

The choice of the criterion $\mathrm{Cr}(\mathbf{p})$ is left to the researcher. SC, HQ, and AIC from (13.3.4)–(13.3.6) are possible candidates. Stage II could be iterated by using the residuals from a previous run through Stage II instead of the residuals from Stage I. Once an estimate $\widehat{\mathbf{p}}$ of the Kronecker indices is determined, the ML estimates conditional on $\widehat{\mathbf{p}}$ may be computed in a final stage.

STAGE III: Estimate the echelon form VARMA model with Kronecker indices $\widehat{\mathbf{p}}$ by maximizing the Gaussian log-likelihood function or by just one step

of the scoring algorithm (see Section 12.3.2). ■

Hannan & Deistler (1988, Chapter 6) discussed conditions under which this procedure provides a consistent estimator of the Kronecker indices and VARMA parameter estimators that have the same asymptotic properties as the estimators obtained for given, known, true Kronecker indices (see Proposition 12.1). In addition to our usual assumptions for the VARMA process such as stability and invertibility, the required assumptions relate to the criteria for choosing the VAR order in Stage I and the Kronecker indices in Stage II. These conditions are asymptotic conditions that leave some room for the actual choice in small samples. Any criterion from (13.3.4)–(13.3.6) may be a reasonable choice in practice.

The procedure still involves extensive computations, unless the dimension K of the underlying multiple time series and $p_{\max}$ are small. For example, for a five-dimensional system with $p_{\max} = 8$ we still have to perform $9^5 = 59,049$ estimations in order to compare all feasible models. Although these estimations involve linear least squares computations only, the computational costs may be substantial. Therefore, we outline two less costly procedures in the following subsections. For small systems, the present procedure is a reasonable choice. We give an example next.

An Example

We consider again the income/consumption example from Section 13.2.2. In the first stage of our procedure, we fit a VAR(8) model ($n = 8$) and we use the residuals, $\widehat{u}_t(8)$, at the next stage. The choice of $n = 8$ is to some extent arbitrary. We have chosen a fairly high order to gain flexibility for the Kronecker indices considered at Stage II. Recall that n must exceed all Kronecker indices to be considered subsequently. Using the procedure described as Stage II, we have estimated models with Kronecker indices $p_k \leq p_{\max} = 4$ and we have determined the corresponding values of the criteria AIC and HQ. They are given in Tables 13.1 and 13.2, respectively. Both criteria reach their minimum for $\mathbf{p} = (p_1, p_2) = (0, 2)$. The ARMA$_E(0, 2)$ form is precisely the model estimated in Chapter 12, Section 12.3.5. Replacing the parameters by their ML estimates with estimated standard errors in parentheses, we have

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \underset{(.252)}{.225} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \underset{(.166)}{.061} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} \widehat{u}_{1,t} \\ \widehat{u}_{2,t} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \underset{(.090)}{.313} & \underset{(.274)}{-.750} \end{bmatrix} \begin{bmatrix} \widehat{u}_{1,t-1} \\ \widehat{u}_{2,t-1} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 \\ \underset{(.141)}{.140} & \underset{(.233)}{.160} \end{bmatrix} \begin{bmatrix} \widehat{u}_{1,t-2} \\ \widehat{u}_{2,t-2} \end{bmatrix}.$$

Obviously, some of the parameters are quite small compared to their estimated standard errors. In such a situation, one may want to impose further zero constraints on the parameters. Because $\widehat{\alpha}_{22,1}, \widehat{\alpha}_{22,2}$, and $\widehat{m}_{22,2}$ have the smallest t-ratios in absolute terms, we restrict these estimates to zero and reestimate the model. The resulting system obtained by ML estimation is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \widehat{u}_{1,t} \\ \widehat{u}_{2,t} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .308 & -.475 \\ (.088) & (.104) \end{bmatrix} \begin{bmatrix} \widehat{u}_{1,t-1} \\ \widehat{u}_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .302 & 0 \\ (.076) & \end{bmatrix} \begin{bmatrix} \widehat{u}_{1,t-2} \\ \widehat{u}_{2,t-2} \end{bmatrix}.$$

Now all parameters are significant under a two-standard error criterion.

Table 13.1. AIC values of $\text{ARMA}_E(p_1, p_2)$ models for the income/consumption data

	p_1				
p_2	0	1	2	3	4
0	−16.83	−18.41	−18.30	−18.25	−18.15
1	−18.50	−18.42	−18.30	−18.23	−18.13
2	−18.64*	−18.55	−18.42	−18.29	−18.19
3	−18.57	−18.50	−18.37	−18.27	−18.19
4	−18.47	−18.38	−18.27	−18.20	−18.05

*Minimum

Table 13.2. HQ values of $\text{ARMA}_E(p_1, p_2)$ models for the income/consumption data

	p_1				
p_2	0	1	2	3	4
0	−16.83	−18.35	−18.21	−18.12	−17.98
1	−18.46	−18.31	−18.14	−18.03	−17.89
2	−18.56*	−18.41	−18.21	−18.03	−17.88
3	−18.45	−18.32	−18.12	−17.95	−17.82
4	−18.31	−18.16	−17.98	−17.84	−17.63

*Minimum

13.3.3 Hannan-Kavalieris Procedure

A full search procedure for the optimal Kronecker indices, as in Stage II of the previous subsection, involves a substantial amount of computation work if the

dimension K of the time series considered is large or if the upper bound $p_{\max}$ for the Kronecker indices is high. For instance, if monthly data are considered and lags of at least one year are deemed necessary, $p_{\max} \geq 12$ is required. Even if the system involves just three variables ($K = 3$) the number of models to be compared is vast, namely $13^3 = 2197$. Therefore, shortcuts for Stage II of the previous subsection were proposed. The first one we present here is inspired by discussions of Hannan & Kavalieris (1984). Therefore, we call it the Hannan-Kavalieris procedure although these authors proposed a more sophisticated approach. In particular, they discussed a number of computational simplifications (see also Hannan & Deistler (1988, Chapter 6)). The following modification of Stage II may be worth trying.

STAGE II (HK): Based on the covariance estimators obtained from the preliminary estimation procedure of Section 12.3.4, find the Kronecker indices, say $\mathbf{p}^{(1)} = p^{(1)}(1, \ldots, 1)$, that minimize a prespecified criterion of the type $\mathrm{Cr}(\mathbf{p})$ in (13.3.1) over $\mathbf{p} = p(1, \ldots, 1)$, $p = 0, \ldots, p_{\max}$, that is, all Kronecker indices are identical in this first step. Then the last index p_K is varied between 0 and $p^{(1)}$ while all other indices are fixed at $p^{(1)}$. We denote the optimal value of p_K by $\widehat{p}_K$, that is, $\widehat{p}_K$ minimizes the prespecified criterion. Then we proceed in the same way with p_{K-1} and so on. More generally, $\widehat{p}_k$ is chosen such that

$$\mathrm{Cr}(p^{(1)}, \ldots, p^{(1)}, \widehat{p}_k, \ldots, \widehat{p}_K) = \min\{\mathrm{Cr}(p^{(1)}, \ldots, p^{(1)}, p, \widehat{p}_{k+1}, \ldots, \widehat{p}_K) | p = 0, \ldots, p^{(1)}\}.$$

■

This modification reduces the computational burden considerably. Just to give an example, for $K = 5$ and $p_{\max} = 8$, at most $9 + 5 \cdot 9 = 54$ models have to be estimated. If $p^{(1)}$ is small, then the number may be substantially lower. For comparison purposes, we repeat that the number of estimations in a full search procedure would be $9^5 = 59,049$.

To illustrate the procedure, consider the following panel of criterion values for Kronecker indices (p_1, p_2):

	p_1			
p_2	0	1	2	3
0	3.48	3.28	3.26	3.27
1	3.25	3.23	3.14	3.20
2	3.23	3.21	3.15	3.19
3	3.24	3.20	3.21	3.18

The minimum value on the main diagonal is obtained for $p^{(1)} = 2$ with $\mathrm{Cr}(p^{(1)}, p^{(1)}) = 3.15$. Going upward from $(p_1, p_2) = (2, 2)$, the minimum is seen to be $\mathrm{Cr}(2, 1) = 3.14$. Turning left from $(p_1, p_2) = (2, 1)$, a further reduction of the criterion value is not obtained. Therefore, the estimate for the Kronecker indices is $\widehat{\mathbf{p}} = (2, 1)$.

In this case, we have actually found the overall minimum of all criterion values in the panel, that is,

$$\mathrm{Cr}(2,1) = \min\{\mathrm{Cr}(p_1,p_2)|p_i = 0,1,2,3\}.$$

In general, the Hannan-Kavalieris procedure will not lead to the overall minimum. For instance, for our bivariate income/consumption example from Section 13.3.2, we can find the HK estimates $\widehat{\mathbf{p}}$ from Tables 13.1 and 13.2. For example, on the main diagonal of the panel in Table 13.2, the HQ criterion assumes its minimum for $(p_1,p_2) = (1,1)$ and $\widehat{\mathbf{p}}(\mathrm{HQ}) = (1,0)$. Clearly, this result differs from the estimate $\widehat{\mathbf{p}} = (0,2)$ that was obtained in the full search procedure.

Under suitable conditions for the model selection criteria, the HK procedure is consistent. Hannan & Deistler (1988) also discussed the consequences of the true data generation process being not in the class of considered processes.

13.3.4 Poskitt's Procedure

Another short-cut version of Stage II of the model specification procedure was suggested by Poskitt (1992). It capitalizes on the important property of echelon forms that the restrictions for the k-th equation implied by a set of Kronecker indices $\mathbf{p}$ are determined by the indices $p_i \leq p_k$. They do not depend on the specific values of the p_j which are greater than p_k. The proposed modification of Stage II is based on separate LS estimations of each of the K equations of the system. The estimation is similar to the preliminary estimation method outlined in Section 12.3.4, that is, it uses the residuals of a long autoregression from Stage I instead of the true u_t's. A model selection criterion of the form

$$\mathrm{Cr}_k(\mathbf{p}) := \ln \widetilde{\sigma}_k^2(\mathbf{p}) + c_T d_k(\mathbf{p})/T \tag{13.3.7}$$

is then evaluated for each of the K equations separately. Here $\widetilde{\sigma}_k^2(\mathbf{p})$ is the residual variance estimate such that $T\widetilde{\sigma}_k^2(\mathbf{p})$ is the residual sum of squares of the k-th equation in a system with Kronecker indices $\mathbf{p}, d_k(\mathbf{p})$ is the number of freely varying parameters in the k-th equation, and c_T is a number that depends on the sample size T. Of course, (13.3.7) is the single equation analogue of the systems criterion (13.3.1). Stage II of Poskitt's procedure proceeds then as follows:

STAGE II (P): Determine the required values $\mathrm{Cr}_k(\mathbf{p})$ and choose the estimates $\widehat{p}_k$ of the Kronecker indices according to the following rule:

If $\mathrm{Cr}_k(0,\ldots,0) \geq \mathrm{Cr}_k(1,\ldots,1)$ for all $k = 1,\ldots,K$, compute $\mathrm{Cr}_k(2,\ldots,2)$, $k = 1,\ldots,K$, and compare to $\mathrm{Cr}_k(1,\ldots,1)$. If the $\mathrm{Cr}_k(2,\ldots,2)$ are all not greater than the corresponding $\mathrm{Cr}_k(1,\ldots,1)$, proceed to $\mathrm{Cr}_k(3,\ldots,3)$ and so on.

If at some stage

$$\mathrm{Cr}_k(j-1,\ldots,j-1) \geq \mathrm{Cr}_k(j,\ldots,j)$$

does *not* hold *for all* k, choose $\widehat{p}_k = j-1$ for all k with

$$\mathrm{Cr}_k(j-1,\ldots,j-1) < \mathrm{Cr}_k(j,\ldots,j).$$

The $\widehat{p}_k$ obtained in this way are fixed in all the following steps. We continue by increasing the remaining indices and comparing the criteria for those equations for which the Kronecker indices are not yet fixed. Here it is important that the restrictions for the k-th equation do not depend on the Kronecker indices $p_i > p_k$ which are chosen in subsequent steps. ■

To make the procedure a bit more transparent, it may be helpful to consider an example. Suppose that interest centers on a three-dimensional system, that is, $K = 3$. First $\mathrm{Cr}_k(0,0,0)$ and $\mathrm{Cr}_k(1,1,1)$ are computed for $k = 1,2,3$. Suppose

$$\mathrm{Cr}_k(0,0,0) \geq \mathrm{Cr}_k(1,1,1), \quad \text{for } k = 1,2,3.$$

Then we evaluate $\mathrm{Cr}_k(2,2,2)$, $k = 1,2,3$. Suppose

$$\mathrm{Cr}_1(1,1,1) < \mathrm{Cr}_1(2,2,2)$$

and

$$\mathrm{Cr}_k(1,1,1) \geq \mathrm{Cr}_k(2,2,2), \quad \text{for } k = 2,3.$$

Then $\widehat{p}_1 = 1$ is fixed and $\mathrm{Cr}_k(1,2,2)$ is compared to $\mathrm{Cr}_k(1,3,3)$ for $k = 2,3$. Suppose

$$\mathrm{Cr}_2(1,2,2) \geq \mathrm{Cr}_2(1,3,3) \quad \text{and} \quad \mathrm{Cr}_3(1,2,2) < \mathrm{Cr}_3(1,3,3).$$

Then we fix $\widehat{p}_3 = 2$ and compare $\mathrm{Cr}_2(1,3,2)$ to $\mathrm{Cr}_2(1,4,2)$ and so on until p_2 can also be fixed because no further reduction of the criterion $\mathrm{Cr}_2(\cdot)$ is obtained in one step. It is important to note that for each index only the first local minimum of the corresponding criterion is searched for. We are not seeking a global minimum over all $\mathbf{p}$ with p_k less than some prespecified upper bound. For moderate or large systems, the present procedure has the advantage of involving a very reasonable amount of computation work only.

Poskitt (1992) derived the properties of the Kronecker indices and the VARMA coefficients estimated by this procedure. He gave conditions under which the Kronecker indices are estimated consistently and the final VARMA parameter estimators have the asymptotic distribution given in Proposition 12.1. Assuming that the true data generation process can indeed be described by a stable, invertible ARMA_E representation with a finite set of Kronecker

indices, the conditions imposed by Poskitt relate to the distribution of the white noise process u_t, to the choice of n, and to the criteria $\mathrm{Cr}_k(\mathbf{p})$.

With respect to the process or white noise distribution, the assumptions are satisfied, for example, if u_t is Gaussian. In fact, for most results it suffices that u_t is standard white noise. An exception is the asymptotic distribution of the white noise covariance estimator $\widetilde{\Sigma}_u$. It may change if u_t has a nonnormal distribution.

The VAR order n at Stage I is assumed to go to infinity with the sample size at a certain rate. In practice, the order selection criteria AIC or HQ may be used at Stage I. It must be guaranteed, however, that n is greater than the Kronecker indices considered at Stage II(P).

Poskitt (1992) also discussed a modification of his algorithm that appears to have some practical advantages. We do not go into that procedure here but recommend that the interested reader examines the relevant literature. The message from the present discussion should be that consistent and feasible strategies for estimating the Kronecker indices exist. Poskitt also discussed the case where the true data generation process is not in the class of VARMA processes considered in the specification procedure. He derived some asymptotic results for this case as well.

In summary, a full search procedure is feasible for low-dimensional systems if the maximum for the Kronecker indices is small or moderate. For high-dimensional systems and/or large upper bounds of the Kronecker indices, the Hannan-Kavalieris procedure or Poskitt's specification strategy are preferable from a computational point of view. The relative performance in small samples is so far unknown in general. It is left to the individual researcher to decide on a specific specification procedure with his or her available resources and perhaps the objective of the analysis in mind. Of course, it is legitimate to try different strategies and criteria and compare the resulting models and the implications for the subsequent analysis.

13.4 Remarks on Other Specification Strategies for VARMA Models

A number of other specification strategies for VARMA processes were proposed and investigated in the literature based on representations other than the final equations and echelon forms. Examples are Quenouille (1957), Tiao & Box (1981), Jenkins & Alavi (1981), Aoki (1987), Cooper & Wood (1982), Granger & Newbold (1986), Akaike (1976), Tiao & Tsay (1989), Tsay (1989a, b), to list just a few. Some of these strategies are based on subjective criteria. As mentioned earlier, none of these procedures seems to be in common use for analyzing economic time series and none of them has become *the* standard procedure. So far, few VARMA analyses of higher-dimensional time series are reported in the literature. Given this state of affairs, it is difficult to give well-founded recommendations as to which strategy to use in any particular

situation. Those familiar with the Box-Jenkins approach for univariate time series modelling will be aware of the problems that can arise even in the univariate case if the investigator has to decide on a model on the basis of statistics such as the autocorrelations and partial autocorrelations. Therefore, it is an obvious advantage to have an automatic or semiautomatic procedure if one feels uncertain about the interpretation of statistical quantities related to specific characteristics of a process and if little or no prior information is available. On the other hand, if firmly based prior information about the data generation process is available, then it may be advantageous to use that at an early stage and depart from automatic procedures.

13.5 Model Checking

Prominent candidates in the model checking tool-kit are tests of statistical hypotheses. All three testing principles, LR (likelihood ratio), LM (Lagrange multiplier), and Wald tests (see Appendix C.7) can be applied in principle in the VARMA context. Because estimation requires iterative procedures, it is often desirable to estimate just one model. Hence, LR tests which require estimation under both the null and alternative hypotheses are often unattractive. In finite order VAR modelling, the unrestricted version is usually relatively easy to estimate and therefore it makes sense to use Wald tests in the pure VAR case because these tests are based on the unconstrained estimator. In contrast, the restricted estimator is often easier to obtain in the VARMA context when models with nontrivial MA part are considered. In this situation, LM tests have an obvious advantage because the LM statistic involves the restricted estimator only. Of course, the restricted estimator is especially easy to determine if the constrained model is a pure, finite order VAR process. We will briefly discuss LM tests in the following. For further discussions and proofs the reader is referred to Kohn (1979), Hosking (1981b), and Poskitt & Tremayne (1982).

13.5.1 LM Tests

Suppose we wish to test

$$H_0 : \varphi(\boldsymbol{\beta}) = 0 \quad \text{against} \quad H_1 : \varphi(\boldsymbol{\beta}) \neq 0, \tag{13.5.1}$$

where $\boldsymbol{\beta}$ is an M-dimensional parameter vector and $\varphi(\cdot)$ is a twice continuously differentiable function with values in the N-dimensional Euclidean space. In other words, $\varphi(\boldsymbol{\beta})$ is an $(N \times 1)$ vector and we assume that the matrix $\partial\varphi/\partial\boldsymbol{\beta}'$ of first order partial derivatives has rank N at the true parameter vector. In this setup, we consider the case where the restrictions relate to the VARMA coefficients only. Moreover, we assume that the conditions of Proposition 12.1 are satisfied.

For instance, in the bivariate zero mean VARMA$(1,1)$ model with Kronecker indices $(p_1, p_2) = (1, 1)$,

$$\begin{aligned}\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} &= \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \\ &\quad + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} + \begin{bmatrix} m_{11,1} & m_{12,1} \\ m_{21,1} & m_{22,1} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}, \end{aligned} \tag{13.5.2}$$

with $\boldsymbol{\beta}' = (\alpha_{11,1}, \alpha_{21,1}, \alpha_{12,1}, \alpha_{22,1}, m_{11,1}, m_{21,1}, m_{12,1}, m_{22,1})$, one may wish to test that the MA degree is zero, that is,

$$\varphi(\boldsymbol{\beta}) = \begin{bmatrix} m_{11,1} \\ m_{21,1} \\ m_{12,1} \\ m_{22,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The corresponding matrix of partial derivatives is

$$\frac{\partial \varphi}{\partial \boldsymbol{\beta}'} = [0 : I_4]$$

which obviously has rank $N = 4$.

As another example, suppose we wish to test for Granger-causality from y_{2t} to y_{1t} in the model (13.5.2). In that case,

$$\varphi(\boldsymbol{\beta}) = \begin{bmatrix} \alpha_{12,1} + m_{12,1} \\ \alpha_{22,1} m_{12,1} - \alpha_{12,1} m_{22,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{13.5.3}$$

(see Remark 1 of Section 11.7.1). The corresponding matrix of partial derivatives is

$$\frac{\partial \varphi}{\partial \boldsymbol{\beta}'} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -m_{22,1} & m_{12,1} & 0 & 0 & \alpha_{22,1} & -\alpha_{12,1} \end{bmatrix}.$$

This matrix may have rank 1 under special conditions. In particular, this occurs if $\alpha_{12,1} = m_{12,1} = 0$ and $\alpha_{22,1} = -m_{22,1}$.

The LM statistic for testing (13.5.1) is

$$\lambda_{LM} := s(\widetilde{\boldsymbol{\beta}}_r)' \widetilde{\mathcal{I}}_a(\widetilde{\boldsymbol{\beta}}_r, \widetilde{\Sigma}_u^r)^{-1} s(\widetilde{\boldsymbol{\beta}}_r)/T, \tag{13.5.4}$$

where

$$s(\widetilde{\boldsymbol{\beta}}_r) = \left. \frac{\partial \ln l_0}{\partial \boldsymbol{\beta}} \right|_{\widetilde{\boldsymbol{\beta}}_r} = \sum_{t=1}^{T} \left[\left. \frac{\partial u_t(\overline{y}, \boldsymbol{\beta})'}{\partial \boldsymbol{\beta}} \right|_{\widetilde{\boldsymbol{\beta}}_r} \right] (\widetilde{\Sigma}_u^r)^{-1} \widetilde{u}_t(\overline{y}, \widetilde{\boldsymbol{\beta}}_r) \tag{13.5.5}$$

is the score vector evaluated at the restricted estimator $\widetilde{\boldsymbol{\beta}}_r$ and

$$\widetilde{\mathcal{I}}_a(\widetilde{\boldsymbol{\beta}}_r, \widetilde{\Sigma}_u^r) = \frac{1}{T} \sum_{t=1}^{T} \left[\left. \frac{\partial u_t(\overline{y}, \boldsymbol{\beta})'}{\partial \boldsymbol{\beta}} \right|_{\widetilde{\boldsymbol{\beta}}_r} \right] (\widetilde{\Sigma}_u^r)^{-1} \left[\left. \frac{\partial u_t(\overline{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right|_{\widetilde{\boldsymbol{\beta}}_r} \right] \tag{13.5.6}$$

is an estimator of the asymptotic information matrix based on the restricted estimator $\widetilde{\boldsymbol{\beta}}_r$. Here

$$\widetilde{\Sigma}_u^r = \frac{1}{T}\sum_{t=1}^{T} \widetilde{u}_t(\overline{y}, \widetilde{\boldsymbol{\beta}}_r)\widetilde{u}_t(\overline{y}, \widetilde{\boldsymbol{\beta}}_r)'.$$

Note that in contrast to Appendix C, Section C.7, an estimator of the *asymptotic* information matrix rather than the information matrix is used in (13.5.4). Therefore T appears in the denominator. If H_0 is true, the statistic λ_{LM} has an asymptotic $\chi^2(N)$-distribution under general conditions.

The LM test is especially suitable for model checking because testing larger VAR or MA orders against a maintained model is particularly easy. A new estimation is not required as long as the null hypothesis does not change. For instance, if we wish to test a given VARMA(p, q) specification against a VARMA$(p+s, q)$ or a VARMA$(p, q+s)$ model, we just need an estimator of the coefficients of the VARMA(p, q) process. Note, however, that a VARMA(p, q) cannot be tested against a VARMA$(p + s, q + s)$, that is, we cannot increase both the VAR and MA orders simultaneously because the VARMA$(p+s, q+s)$ model will not be identified (cancellation is possible!) if the null hypothesis is true. In that case, the LM statistic will not have its usual asymptotic χ^2-distribution.

13.5.2 Residual Autocorrelations and Portmanteau Tests

Alternative tools for model checking are the residual autocorrelations and portmanteau tests. The asymptotic distributions of the residual autocorrelations of estimated VARMA models were discussed by Hosking (1980), Li & McLeod (1981), and Poskitt & Tremayne (1982), among others. We do not give the details here but just mention that the resulting standard errors of autocorrelations at large lags obtained from asymptotic considerations are approximately $1/\sqrt{T}$, while they may be much smaller for low lags, just as for pure finite order VAR processes.

The modified portmanteau statistic is

$$\bar{Q}_h := T^2 \sum_{i=1}^{h} (T-i)^{-1} \operatorname{tr}(\widehat{C}_i' \widehat{C}_0^{-1} \widehat{C}_i \widehat{C}_0^{-1}), \tag{13.5.7}$$

where

$$\widehat{C}_i := \frac{1}{T}\sum_{t=i+1}^{T} \widetilde{u}_t(\overline{y}, \widetilde{\boldsymbol{\beta}})\widetilde{u}_{t-i}(\overline{y}, \widetilde{\boldsymbol{\beta}})'$$

and the $\widetilde{u}_t(\overline{y}, \widetilde{\boldsymbol{\beta}})$'s are the residuals of an estimated VARMA model, as before. Under general conditions, $\bar{Q}_h$ has an *approximate* asymptotic χ^2-distribution. The degrees of freedom are obtained by subtracting the number of freely estimated VARMA coefficients from $K^2 h$.

13.5.3 Prediction Tests for Structural Change

In the pure VAR case, we have considered prediction tests for structural change as model checking devices. If the data generation process is Gaussian, the two tests introduced in Chapter 4, Section 4.6.2, may be applied in the VARMA case as well with minor modifications.

The statistics based on h-step ahead forecasts only are of the form

$$\overline{\tau}_h := \widehat{e}_T(h)' \widehat{\Sigma}_{\widehat{y}}(h)^{-1} \widehat{e}_T(h), \tag{13.5.8}$$

where $\widehat{e}_T(h) = y_{T+h} - \widehat{y}_T(h)$ is the error vector of an h-step forecast based on an estimated VARMA(p, q) process and $\widehat{\Sigma}_{\widehat{y}}(h)$ is an estimator of the corresponding MSE matrix (see Section 12.5). The statistic may be applied in conjunction with an $F(K, T-K(p+q)-1)$-distribution. The denominator degrees of freedom may be used even if constraints are imposed on the VARMA coefficients because the F-distribution is just chosen as a small sample approximation to a $\chi^2(K)/K$ distribution. Its justification comes from the fact that $F(K, T-s)$ converges to $\chi^2(K)/K$ for any fixed constant s, as T approaches infinity. Thus, any constant that is subtracted from T in the denominator degrees of freedom of the F-distribution, is justified on the same asymptotic grounds. It is not clear which choice is best from a small sample point of view.

The other statistic considered in Section 4.6.2 is based on 1- to h-step forecasts and, for the present case, it may be modified as

$$\overline{\lambda}_h := T \sum_{i=1}^{h} \widehat{u}'_{T+i} \widetilde{\Sigma}_u^{-1} \widehat{u}_{T+i} / [(T + K(p+q) + 1)Kh] \tag{13.5.9}$$

and its approximate distribution for a structurally stable Gaussian process is $F(Kh, T - K(p+q) - 1)$. Here $\widehat{u}_{T+i} = y_{T+i} - \widehat{y}_{T+i-1}(1)$ and $\widetilde{\Sigma}_u$ is the ML estimator of Σ_u. Note that the LS estimator of Σ_u was used in Section 4.6.2 instead. Again, there is not much theoretical justification for the choice of the denominator in (13.5.9) and for the denominator degrees of freedom in the approximating F-distribution. More detailed investigations of the small sample distribution of $\overline{\lambda}_h$ are required before firmly based recommendations regarding modifications of the statistic are possible. Here we have just used the direct analogue of the finite order pure VAR case.

It is also possible to fit a finite order VAR process to data generated by a mixed VARMA process and base the prediction tests on forecasts from that model. In Chapter 15, it will be shown that such an approach is theoretically sound under general conditions.

13.6 Critique of VARMA Model Fitting

In this and the previous two chapters, much of the analysis is based on the assumption that the true data generation mechanism is from the VARMA(p, q)

class. In practice, any such model is just an approximation to the actual data generation process. Therefore, the model selection task is not really the problem of finding the true structure but of finding a good or useful approximation to the real life mechanism. Despite this fact, it is sometimes helpful to assume a specific true process or process class to be able to derive, under ideal conditions, the statistical properties of the procedures used. One then hopes that the actual properties of a procedure in a particular practical situation are at least similar to those obtained under ideal conditions.

Against this background, one may wonder whether it is sufficient or even preferable to approximate the generation process of a given multiple time series by a finite order VAR(p) process rather than go through the painstaking specification and estimation of a mixed VARMA model. Clearly, the estimation of VARMA models is in general more complicated than that of finite order VAR models. Moreover, the specification of VAR models by statistical methods is much simpler than that of VARMA models. Are there still situations where it is reasonable to consider the more complicated VARMA models? The answer to this question is in the affirmative. For instance, if subject matter theory suggests a VARMA model with nontrivial MA part, it is often necessary to work with such a specification to answer the questions of interest or derive the relevant results. Also, in some cases, a VARMA approximation may be more parsimonious in terms of the number of parameters involved than an appropriate finite order VAR approximation. In such cases, the VARMA approximation may, for instance, result in more efficient forecasts that justify the costly specification and estimation procedures. The future attractiveness of VARMA models will depend on the easy availability of efficient and robust estimation and specification procedures that reduce the costs to an acceptable level.

In Chapter 15, we will follow another road and explicitly assume that just an approximating and not a true VAR(p) model is fitted. Assumptions will be provided that allow the derivation of statistical properties in that case. So far, we have considered stable, stationary VARMA processes. In the next chapter, extensions for integrated and cointegrated variables will be considered.

13.7 Exercises

Problem 13.1
At the first stage of a final equations form specification procedure, the following two univariate models were obtained:

$$\begin{aligned}(1 + 0.3L - 0.4L^2)y_{1t} &= (1 + 0.6L)v_{1t},\\ (1 - 0.5L)y_{2t} &= (1 + 0.6L)v_{2t}.\end{aligned}$$

Which orders do you choose for the bivariate final equations VARMA representation of $(y_{1t}, y_{2t})'$?

Problem 13.2
At Stage II of a specification procedure for an echelon form of a bivariate system, the following values of the HQ criterion are obtained:

	p_1				
p_2	0	1	2	3	4
0	2.1	1.9	1.5	1.5	1.6
1	1.8	1.7	1.4	1.2	1.3
2	1.7	1.4	1.3	1.4	1.4
3	1.7	1.4	1.3	1.4	1.5
4	1.8	1.7	1.6	1.5	1.5

Choose an estimate $(\widehat{p}_1, \widehat{p}_2)'$ by the Hannan-Kavalieris procedure. Interpret the estimate in the light of a full search procedure.

Problem 13.3
At the second stage of the Poskitt procedure for a bivariate model, the specification criteria $\mathrm{Cr}_1(p_1, p_2)$ and $\mathrm{Cr}_2(p_1, p_2)$ assume the following values:

$\mathrm{Cr}_1(p_1, p_2)$

	p_1			
p_2	0	1	2	3
0	3.5	2.5	1.7	1.8
1	3.5	1.5	1.8	1.7
2	3.5	1.5	1.8	1.9
3	3.5	1.5	1.8	1.4

$\mathrm{Cr}_2(p_1, p_2)$

	p_1			
p_2	0	1	2	3
0	4.2	3.2	3.2	3.2
1	3.5	1.8	1.9	1.9
2	3.1	1.9	1.7	1.6
3	3.4	2.1	1.8	1.9

Use the Poskitt strategy to find an estimate $(\widehat{p}_1, \widehat{p}_2)$ of the Kronecker indices.

The following problems require the use of a computer. They are based on the *first differences* of the U.S. investment data given in File E2.

Problem 13.4
Determine a final equations form VARMA model for the U.S. investment data for the years 1947–1968 using the specification strategy described in Section 13.2.1.

Problem 13.5
Determine an ARMA_E model for the U.S. investment data using the specification strategy described in Section 13.3.2 with $n = 6$ and based on the HQ criterion. Compare the model to the final equations form model from Problem 13.4.

Problem 13.6
Compute forecasts for the investment series for the years 1969 and 1970 based on (i) the final equations form VARMA model, (ii) the ARMA_E model, and (iii) a bivariate VAR(1) model. Compare the forecasts to the true values and interpret.

Problem 13.7
Compute Φ_i and Θ_i impulse responses from the two models obtained in Problems 13.4 and 13.5, compare and interpret them.

Problem 13.8
Specify a univariate ARMA model for the sum $z_t = y_{1t} + y_{2t}$ of the two investment series for the years 1947–1968. Is the univariate ARMA model compatible with the bivariate echelon form model specified in Problem 13.5? (Hint: Use the results of Section 11.6.1.)

Problem 13.9
Evaluate forecasts for the z_t series of the previous problem for the years 1969–1970 and compare them to forecasts obtained by aggregating the bivariate forecasts from the ARMA_E model of Problem 13.5.