February 11, 1997

Estimating the Kronecker Indices of Cointegrated Echelon Form VARMA Models

by

Holger Bartel and Helmut Lütkepohl[1]

Institut für Statistik und Ökonometrie
Humboldt–Universität
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5718

Abstract

Cointegrated VARMA models can be parameterized by using the echelon form, which is characterized by the Kronecker indices. Three different methods for estimating the Kronecker indices of cointegrated echelon form VARMA models are discussed and compared. They have the common feature of estimating the individual equations of the system separately and using order selection criteria. The small sample performance of the methods is compared in a simulation study. It is found that the performance is better if all echelon form restrictions implied by the Kronecker indices found in preceeding steps are incorporated immediately.

[1]Financial support by the Deutsche Forschungsgemeinschaft, Berliner Graduiertenkolleg für Angewandte Mikroökonomik and SFB 373, is gratefully acknowledged.

Email address Holger Bartel: holger@wiwi.hu-berlin.de.

1 Introduction

In the multiple time series literature a number of books and articles deal with estimating, specifying and analyzing vector autoregressive moving average (ARMA) models. In fact, Quenouille (1957) in his early contribution to the subject presents them as a possible framework for multiple time series analysis. Hannan and Deistler (1988), Lütkepohl (1991), Reinsel (1993) and Claessen (1995) are more recent books where some of the earlier literature is summarized and the current state of the art of analyzing stationary vector ARMA processes is dicussed. Hannan and Kavalieris (1984), Poskitt (1992), and Nsiri and Roy (1992) are, for example, important contributions where practical specification and analysis tools for stationary processes are introduced. In Lütkepohl and Poskitt (1996) several specification strategies are surveyed and extensions to integrated and cointegrated processes are considered by Lütkepohl and Claessen (1997), Claessen (1995) and Poskitt and Lütkepohl (1996).

Despite a considerable amount of theoretical work and despite the fact that strategies and algorithms for specifying and estimating vector ARMA models are available, there are only very few applied studies using the vector ARMA methodology. Clearly, one reason for this state of affairs is that pure AR models, for instance, are more easily dealt with in practice and a bit more is known about the small sample properties of inference methods for these models. On the other hand, it has been pointed out that vector ARMA models have several advantages over their AR competitors (e.g., Lütkepohl and Poskitt (1996)). Among these advantages is the potential of greater parsimony and the implied increase in forecast precision. Therefore, in this paper we will investigate the small sample properties of some specification strategies for vector ARMA models that have been proposed in the recent literature. We will do so in the framework of the echelon form because this form is fairly easy to deal with and at the same time it has a good potential for describing even complicated autocovariance structures in a parsimonious way.

Claessen (1995) reports the results of a simulation study comparing different strategies mainly for stationary processes. Since in practice most macro variables are integrated we will focus on integrated and potentially cointegrated processes in this study. Hence, we will concentrate on specification strategies which have, in particular, potential for such processes. We will also suggest and explore the properties of a procedure inspired by Koreisha and Pukkila (1995) who check the residuals of a univariate model for whiteness to decide on the ARMA orders of

the underlying data generation process (DGP). It will be shown how such a strategy can be taylored to the case of specifying nonstationary echelon form ARMA (ARMA_E) processes.

The structure of the paper is as follows. In the following section the general framework is introduced. In particular, ARME_E structures are presented in such a form so as to allow for nonstationary variables. In Section 3 some possible estimation procedures for the Kronecker indices which define the precise structure of an ARMA_E model are discussed. In Section 4 these procedures are compared in a simulation study. Conclusions follow in Section 5.

2 The ARMA_E Form

In the following it is assumed that the data generating process (DGP) of the K-dimensional multiple time series $y_1, \ldots, y_T$ with $y_t = (y_{1t}, \ldots, y_{Kt})'$ is from the VARMA (vector autoregressive moving average) class,

$$A(L)y_t = \nu + M(L)u_t, \tag{2.1}$$

where u_t is an unobservable white noise process with zero mean and nonsingular, time invariant covariance matrix $E(u_t u_t') = \Sigma_u$,

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

and

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

are matrix polynomials in the lag or backshift operator L, which is defined as usual by $Ly_t = y_{t-1}$. The matrix polynomials are assumed to satisfy

$$\det A(z) \neq 0,\ |z| \leq 1, z \neq 1, \quad \text{and} \quad \det M(z) \neq 0,\ |z| \leq 1. \tag{2.2}$$

The second part of this condition is the usual invertibility condition for the MA operator. The possibility that the operator $A(z)$ can have zeros for $z = 1$ as assumed in the first part of (2.2) is of special interest since thereby the components of y_t are allowed to be integrated, nonstationary variables which become stationary upon differencing. We also assume that each component series is stationary after differencing once. Our assumptions also allow for possible cointegration between the variables (see Engle and Granger (1987)) so that linear combinations of the levels variables may be stationary. The fact that we do not make assumptions regarding

the number of zeros at $z = 1$ in the autoregressive operator means that we leave open the possibility that the process is stationary or that there are some integrated component series which do not cointegrate with other variables. For a more complete discussion of the possibilities covered here, see Lütkepohl (1991, Chapter 11).

In addition to the foregoing conditions it is assumed that $[A(z) : M(z)]$ is (left) coprime and in echelon canonical form. Denoting the klth elements of $A(z)$ and $M(z)$ by $\alpha_{kl}(z)$ and $m_{kl}(z)$, respectively, the polynomial operators can be uniquely defined by the requirements that

$$m_{kk}(L) = 1 + \sum_{i=1}^{p_k} m_{kk,i} L^i, \quad \text{for } k = 1, \ldots, K, \tag{2.3a}$$

$$m_{kl}(L) = \sum_{i=p_k - p_{kl}+1}^{p_k} m_{kl,i} L^i, \quad \text{for } k \neq l, \tag{2.3b}$$

$$\alpha_{kl}(L) = \sum_{i=0}^{p_k} \alpha_{kl,i} L^i, \quad \text{with } \alpha_{kl,0} = m_{kl,0} \quad \text{for } k, l = 1, \ldots, K. \tag{2.3c}$$

Here

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

The row degrees p_k in this representation are the Kronecker indices (see Hannan and Deistler (1988) and Lütkepohl (1991)). In (2.1) $p = \max(p_1, \ldots, p_K)$, that is, p is the maximum row degree or Kronecker index. We follow Poskitt (1992) and abbreviate this echelon representation of a VARMA process by ARMA_E and $\text{ARMA}_E(p_1, \ldots, p_K)$ denotes an echelon form with Kronecker indices $p_1, \ldots, p_K$.

Note that in the formulation of the echelon form in (2.3) 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. This representation of the echelon form was introduced by Lütkepohl and Claessen (1997). It differs from the ARMA_E form usually found in the literature where the restrictions on low order lags are imposed on the AR coefficient matrices. The form in (2.3) has the advantage of being conveniently combined with the error correction (EC) form for specifying cointegrated processes. This form is useful in analyzing integrated and cointegrated systems. Therefore we use it in the following although we do not consider the EC form in the present paper because we are mainly interested in estimating the Kronecker indices which may be specified in an initial stage of a more detailed ARMA_E

cointegration analysis. It should be noted, however, that there is a relationship between the Kronecker indices and the cointegration rank of a system (see Lütkepohl and Claessen (1997, Sec. 3.3)).

At this point it may be useful to remind readers of the advantages of the echelon form which have been pointed out by many authors before (e.g., Lütkepohl and Poskitt (1996), Lütkepohl and Claessen (1997)). First, every rational matrix operator has a unique echelon form representation. Hence, the ARMA_E form is a canonical form. Akaike (1974) introduced it to the statistics literature by setting up a *minimal* predictor representation which leads to a further advantage of this form, namely its parsimony in terms of the number of parameters involved. This is not to say that it is always the most parsimonious representation. In general, however, the number of free parameters in the ARMA_E form is relatively small compared to other representations. The Kronecker indices specify the maximum row degrees and imply a number of zero restrictions which are sufficient for identifying the VARMA operators. Of course, there may be further overidentifying restrictions. In particular, the AR and MA operators need not necessarily have identical orders although they are identified (unique) even with identical orders. Overidentifying restrictions may be imposed once the Kronecker indices have been specified. In the following we will focus on this first step of the specification procedure namely the determination of the Kronecker indices. The simplicity of the identification restrictions imposed on the ARMA_E form turns out to be a further important advantage over other representations which require cross-equation and/or nonlinear restrictions for identification whereas the constraints on the echelon form are simple linear zero/one restrictions.

3 Strategies for Estimating the Kronecker Indices

In this section we summarize the specification procedures for Kronecker indices which will be considered in the simulations in Sec. 4. There are many other procedures which have been proposed for stationary processes (see, e.g., Claessen (1995)) and which are partly not suitable for nonstationary processes. Since the latter are of primary interest to us, we only consider procedures which are potentially suitable for that case. The first stage is the same in all the procedures. It consists of fitting a long autoregression by least squares in order to provide estimates of the unobservable innovations $u_t,\ t = 1, \ldots, T$.

Stage I:

Use multivariate least squares (LS) estimation (i.e., use LS for each equation separately) to fit a long VAR(h_T) process

$$y_t = \nu + \sum_{i=1}^{h_T} \Pi_{i,h_T} y_{t-i} + u_t(h_T) \tag{3.1}$$

to the data to obtain residuals $\hat{u}_t(h_T)$. □

The order h_T has to be chosen as a function of the sample size T in order to obtain favourable asymptotic properties of the procedures discussed next. More precisely, if h_T approaches infinity at a suitable rate as T goes to infinity, Poskitt and Lütkepohl (1995), Guo, Huang and Hannan (1990) and Huang and Guo (1990) show that the estimation residuals $\hat{u}_t(h_T)$ are "good" estimates of the true residuals (see Lemma 3.1 of Poskitt and Lütkepohl (1995) for details). These residuals are then used in estimating different structures which are compared to make a choice of the Kronecker indices based on a prespecified criterion.

The methods to be compared in the following differ in the way they choose the Kronecker indices in the next step. The first variant of Step II was proposed by Poskitt and Lütkepohl (1995). It 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 which includes the residual variance as a measure of goodness of fit. Formally this procedure can be described as follows.

Stage II(PL1):

Proceed in the following steps.

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

(ib) For $k = 1, \ldots, K$, compute a selection criterion of the 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 function of T which will be specified later.

(ii) Set the estimate of the kth Kronecker index equal to

$$\hat{p}_k = \arg \min_{0 \le n \le P_T} \Lambda_{k,T}(n), \quad k = 1, \ldots, K. \qquad \square$$

In the regressions in Step (ia) the echelon structure is not explicitly estimated, because for each value of n the algorithm is implicitly assuming that the current index under consideration is the smallest and thus no restrictions are imported from other equations. Still, it is clear that the kth equation will be misspecified whenever n is less than the true Kronecker index since one or more lagged values required for a correct specification will be omitted. On the other hand, if n is greater than the true Kronecker index, the kth equation will be correctly specified but may include redundant parameters and variables. Therefore the criterion function $\Lambda_{k,T}(n)$ asymptotically will possess a global minimum when n is equal to the true Kronecker index if C_T is specified appropriately. In practice possible choices of this function of T are $C_T = h_T \log T$ or $C_T = h_T^2$.

Poskitt and Lütkepohl (1995) also propose a modification of Stage II which permits to take into account coefficient restrictions derived from those equations in the system that 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 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. The following formal description of this stage is taken from Poskitt and Lütkepohl (1995).

Stage II(PL2):

Complete the following steps.

(i) Set $\hat{p}'_{k(K)} = \min(\hat{p}_k), \quad k(K) = \arg\min_k(\hat{p}_k)$

(iia) For some q, assume that $\hat{p}'_{k(q+1)} \ge \cdots \ge \hat{p}'_{k(K)}$ are given. For $k \notin \{k(q+1), \ldots, k(K)\}$, regress y_{kt} on a constant and $(y_{jt} - \hat{u}_{jt})$, $j \ne k, j \notin \{k(q+1), \ldots, k(K)\}$, plus $\hat{u}_{k(j)(t-s)}$, $s = n - \hat{p}'_{k(j)} + 1, \ldots, n$, $j = q+1, \ldots, K$, and y_{t-s} and $\hat{u}_{j(t-s)}$, $j \notin \{k(q+1), \ldots, k(K)\}$, $s = 1, \ldots, n$, and compute the residual sum of squares $T\hat{\sigma}^2_{k,T}(n)$ for $n = \hat{p}'_{k(q+1)}, \ldots, P_T$.

(iib) Determine the values of the selection criterion

$$\Lambda_{k,T}(n) = \log \hat{\sigma}^2_{k,T}(n) + C_T n/T, \quad n = \hat{p}'_{k(q+1)}, \ldots, P_T,$$

for those $k \notin \{k(q+1), \ldots, k(K)\}$.

(iii) Set the estimate of the $k(q)$th Kronecker index equal to

$$\hat{p}'_{k(q)} = \min_k \{\arg\min_n \Lambda_{k,T}(n)\}$$

where $k(q) = \arg\min_k \{\arg\min_n \Lambda_{k,T}(n)\}$.

(iv) Repeat Steps (ii) and (iii) for $q = K-1, \ldots, 1$. □

Poskitt and Lütkepohl (1995) show that for a suitable choice of C_T the procedure results in consistent estimators of the Kronecker indices. In this version of Stage II the coefficient restrictions derived from the echelon canonical form are directly incorporated into the identification stage which may result in a superior performance of the selection procedure. On the other hand, the computational burden is increased substantially which may be problematic for high dimensional systems.

In Stages II(PL1) and (PL2) we have to assign values for h_T, P_T and C_T. The theoretical consistency results are quite general and provide an asymptotic justification for many different values of these quantities. Poskitt and Lütkepohl (1995) propose the following choices:

(1) Choose h_T by AIC or use $h_T = \max\{(\log T)^a, h(AIC)\}$ where $a > 1$.

(2) Choose $P_T = \frac{1}{2} h_T$.

(3) Choose $C_T = h_T \log T$ or $C_T = h_T^2$.

We will explore different combinations of these rules in the simulation study reported in the next section.

Another variant of Stage II is inspired by results of Koreisha and Pukkila (1993, 1995), Koreisha and Yoshimoto (1991) and Pukkila, Koreisha and Kallinin (1990) who propose to fit a model and then check, via some model selection criterion, whether the residuals are white noise. Such a procedure can also be used in the present context. Hence, we suggest to fit models of increasing degrees to each equation of our system and for each degree the residuals are checked for being white noise. If they are found to be white the Kronecker index of

that equation is fixed and the corresponding dependent variable is placed last in the vector of variables. In the next steps its Kronecker index remains fixed and its implied restrictions are observed in the remaining equations for which the row degrees are increased one by one until the residuals are white noise. Whenever a residual series is found to be white the variable is placed last in the list of remaining variables, its Kronecker index is fixed and the implied restrictions are taken into account in the further steps. In this way we end up with a set of nonincreasing Kronecker indices $\hat{p}_1 \geq \hat{p}_2 \geq \cdots \geq \hat{p}_K$. Formally this procedure may be described as follows:

Stage II(WN):

Set $n = 0$ and $l = 0$.

(*) If $n = P_T$ estimate $\hat{p}_k = P_T$ for $k = 1, \ldots, K - l$ and stop.

If $n < P_T$ perform the next steps for $k = 1, \ldots, K - l$. Fit the following models by LS:

$$\begin{aligned} y_{kt} = \nu_k \quad & + \textstyle\sum_{\substack{j=1 \\ j \neq k}}^{K-l} \alpha_{kj,0}(y_{jt} - \hat{u}_{jt}) + \sum_{j=1}^{K} \sum_{i=1}^{n} \alpha_{kj,i} y_{j,t-i} \\ & + \textstyle\sum_{j=1}^{K-l} \sum_{i=1}^{n} m_{kj,i} \hat{u}_{j,t-i} + \sum_{j=K-l+1}^{K} \sum_{i=n-\hat{p}_j+1}^{n} m_{kj,i} \hat{u}_{j,t-i} + u_{kt}. \end{aligned}$$

Denote the residuals by $\tilde{u}_{kt}$ and compute the residual variance $\tilde{\sigma}_k^2(n) = T^{-1} \sum_{t=1}^{T} \tilde{u}_{kt}^2$. Fit AR($q$) models for $q = 1, \ldots, Q_T$,

$$\tilde{u}_{kt} = a_1 \tilde{u}_{k,t-1} + \cdots + a_q \tilde{u}_{k,t-q} + \varepsilon_{kt}$$

by LS, determine the residual variances $\tilde{\sigma}_{k,\varepsilon}^2(q)$ and compute the values of the criterion

$$\delta_{k,T}(q) = T \log \tilde{\sigma}_{k,\varepsilon}^2(q) + q c_T$$

where c_T is a suitable function of the sample size T which will be specified below.

If for some k, $\delta_{k,T}(q) \geq T \log \tilde{\sigma}_k^2(n)$ for all $q = 1, \ldots, Q_T$, replace $y_{K-l,t}$ with $y_{k,t}$, choose $\hat{p}_{k-l} = n$, increase l by one and return to (*). Note that $\tilde{\sigma}_k^2(n) = \tilde{\sigma}_{k,\varepsilon}^2(0)$.

If for all $k = 1, \ldots, K - l$, $\delta_{k,T}(q) < T \log \tilde{\sigma}_k^2(n)$ for some q, increase n by one and return to (*). □

Two possible choices of c_T are $c_T = \log T$ which corresponds to the Schwarz Criterion (SC) and $c_T = 2$ which corresponds to Akaikes Information Criterion (AIC). The maximum order of the Kronecker indices for Stage II(WN) is again proposed to be $P_T = \frac{1}{2} h_T$. The maximum order of the AR process fitted to the estimated residuals $\tilde{u}_{kt}$ is given by $Q_T = h_T - P_T - 1$ in order to guarantee $P_T + Q_T < h_T$.

These different variants of Stage II will be compared in a Monte Carlo experiment in the following section.

4 Monte Carlo Comparison

4.1 Data Generation Processes

Eight different data generating processes are used in the Monte Carlo study. They are presented in Table 1. All processes have dimension $K = 3$ and the error covariance matrix is $\Sigma_u = I_K$. The error distribution is normal $N(0, \Sigma_u)$. $\bar{p} = (p_1, \ldots, p_K)$ denotes the Kronecker indices and ϱ is the cointegrating rank. The first DGP is a white noise process with $\bar{p} = (0, 0, 0)$, $\varrho = 3$, $\nu = 0$, and $A_0 = I_3$. The second process consists of independent random walks. Hence, $\bar{p} = (1, 1, 1)$, $\varrho = 0$, $\nu = 0$, $A_0 = I_3$, $A_1 = -I_3$, and $M_1 = 0$.

The remaining DGPs 3 to 8 all have Kronecker indices $\bar{p} = (2, 1, 1)$ and cointegrating rank $\varrho = 1$. Their intercept vector is $\nu = 0$ except for DGP 4. In order for the DGPs to have a realistic structure a process estimated by Lütkepohl and Claessen (1997) was taken as a basis and their estimated coefficient matrices were modified to obtain simple DGPs. The process considered in their study is based on a four dimensional system of U.S. economic variables. They use time series consisting of 136 quarterly observations for the years 1954/1 to 1987/4 to fit a cointegrated VARMA model. The variables are real money stock M1, Gross National Product (GNP) in billions of 1982 dollars, the discount interest rate on new issues of 91–day treasury bills (r^s), and the yield on long term (20 years) treasury bonds (r^l). Lütkepohl and Claessen (1997) found an estimate $\hat{\bar{p}} = (2, 1, 1, 1)$ of the Kronecker indices and determined a cointegrating rank $\hat{\varrho} = 1$. Using roughly the coefficients corresponding to the three variables GNP, M1 and r^s results in the following coefficient matrices of the VARMA model $A_0 y_t + A_1 y_{t-1} + A_2 y_{t-2} = \nu + A_0 u_t + M_1 u_{t-1} + M_2 u_{t-2}$:

$$A_0 = M_0 = \begin{bmatrix} 1 & - & - \\ -0.5 & 1 & - \\ 0.0 & - & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.8 & 0.0 & 0.8 \\ - & - & - \\ - & - & - \end{bmatrix}, \quad A_1 = BC - A_0 - A_2,$$

Table 1: Data Generation Processes Used in the Simulations

Data generation process	Kronecker indices $\bar{p}$	cointegration rank ϱ	other characteristics
DGP 1 (white noise)	(0,0,0)	3	$\nu = 0$
DGP 2 (independent random walks)	(1,1,1)	0	$\nu = 0$
DGP 3 (medium eigenvalues)	(2,1,1)	1	$\nu = 0$ $\lambda_1^{ar} = 0.7,\ \lambda_2^{ar} = 0.4$ $\lambda_1^{ma} = 0.6,\ \lambda_2^{ma} = -0.5$
DGP 4 (medium eigenvalues, nonzero intercept)	(2,1,1)	1	$\nu = (0.1, 0.2, 0.2)'$ $\lambda_1^{ar} = 0.7,\ \lambda_2^{ar} = 0.4$ $\lambda_1^{ma} = 0.6,\ \lambda_2^{ma} = -0.5$
DGP 5 (large negative MA eigenvalues)	(2,1,1)	1	$\nu = 0$ $\lambda_1^{ar} = 0.7,\ \lambda_2^{ar} = 0.4$ $\lambda_1^{ma} = -0.95,\ \lambda_2^{ma} = -0.7$
DGP 6 (large positive MA eigenvalues)	(2,1,1)	1	$\nu = 0$ $\lambda_1^{ar} = 0.7,\ \lambda_2^{ar} = 0.4$ $\lambda_1^{ma} = 0.95,\ \lambda_2^{ma} = 0.7$
DGP 7 (large negative AR eigenvalues)	(2,1,1)	1	$\nu = 0$ $\lambda_1^{ar} = -0.95,\ \lambda_2^{ar} = -0.7$ $\lambda_1^{ma} = 0.6,\ \lambda_2^{ma} = -0.5$
DGP 8 (large positive AR eigenvalues)	(2,1,1)	1	$\nu = 0$ $\lambda_1^{ar} = 0.95,\ \lambda_2^{ar} = 0.7$ $\lambda_1^{ma} = 0.6,\ \lambda_2^{ma} = -0.5$

where

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -0.6 & 0.3 \end{bmatrix}$$

and

$$[M_1 : M_2] = \begin{bmatrix} -0.6 & - & - & : & m_2 & 0.0 & m_2 \\ 0.0 & 0.0 & 0.0 & : & - & - & - \\ m_1 & 0.0 & m_1 & : & - & - & - \end{bmatrix}.$$

where some of the parameters are left unspecified to gain flexibility. The coefficients b_2, m_1 and m_2 appear twice in the coefficient matrices and thereby imply some coefficients to be equal. The corresponding coefficients of the original restricted coefficient matrices from Lütkepohl and Claessen (1997) are similar in size, too. The echelon form zero restrictions given by (2.3) are denoted by a bar in order to distinguish them from the freely varying coefficients which have been set to zero and which are denoted by 0.0 in contrast. The restriction $A_0 = M_0$ is also part of the echelon form restrictions.

Note that the cointegrating rank ϱ is the rank of the matrix $BC = A_0 + A_1 + A_2$. Hence, choosing $A_1 = BC - A_0 - A_2$ ensures a cointegrating rank of $\varrho = 1$. Such a cointegrating rank in a system of dimension $K = 3$ requires to have two unit roots in the autoregressive part. That is, the polynomial $\det(A(z))$ must have two roots at unity. Generally this polynomial has degree $\delta(\det(A(z))) = \sum_{k=1}^{K} p_k =: m$, see Poskitt (1996), where m is called the McMillan degree. For the present case we have $\delta(\det(A(z))) = m = 2 + 1 + 1 = 4$. Dividing $\det(A(z))$ by the two unit roots $(z - 1)$ and $(z - 1)$ using polynomial division, gives a polynomial of order two, the roots of which can be computed easily. Since the eigenvalues λ_i^{ar}, $i = 1, \ldots, m$, of the autoregressive part are the reciprocals of the roots of the reverse characteristic polynomial $\det(A(z)) = \det(A_0 + A_1 z + A_2 z^2)$, see Lütkepohl (1991, pp. 12 and 455), the resulting scalar polynomial of order 2, $z^2 + c_1 z + c_2$ say, may be written as

$$\begin{aligned} (z - \tfrac{1}{\lambda_1^{ar}})(z - \tfrac{1}{\lambda_2^{ar}}) &= z^2 + (-\tfrac{1}{\lambda_1^{ar}} - \tfrac{1}{\lambda_2^{ar}})z + \tfrac{1}{\lambda_1^{ar}\lambda_1^{ar}} \\ &=: z^2 + c_1 z + c_2, \end{aligned}$$

if it is assumed that two real valued zeros of the polynomial exist. Thus, equating $c_1 = (-\frac{1}{\lambda_1^{ar}} - \frac{1}{\lambda_2^{ar}})$ and $c_2 = \frac{1}{\lambda_1^{ar}\lambda_1^{ar}}$ and solving for the variables b_1 and b_2 of the autoregressive part we obtain

$$b_1 = \tfrac{15}{7} + \tfrac{10}{7}(-\lambda_1^{ar} - \lambda_2^{ar} + \tfrac{3}{8}\lambda_1^{ar}\lambda_2^{ar})$$

and

$$b_2 \quad = \quad -1 + \tfrac{5}{4}\lambda_1^{ar}\lambda_2^{ar} \ .$$

We may choose λ_1^{ar} and λ_2^{ar} to take any real value strictly between 1 and -1. The other roots are given by $\lambda_3^{ar} = \lambda_4^{ar} = 1$. The corresponding values of b_1 and b_2 then lead to the autoregressive part used in the simulations.

A similar computation is done for the moving average part. The polynomial $\det(M(z))$ $= \det(A_0 + M_1 z + M_2 z^2)$ in this special case is of order 2 which is due to the additional zero restrictions of the moving average coefficient matrices. The resulting expressions for the moving average parameters m_1 and m_2 are

$$m_1 \quad = \quad \tfrac{3}{5} - \lambda_1^{ma} - \lambda_2^{ma}$$

and

$$m_2 \quad = \quad \tfrac{3}{5}(\tfrac{3}{5} - \lambda_1^{ma} - \lambda_2^{ma}) + \lambda_1^{ma}\lambda_2^{ma} \ .$$

Again we may choose λ_1^{ma} and λ_2^{ma} to take any real value between 1 and -1. The corresponding values of m_1 and m_2 then lead to the desired moving average part.

The eigenvalues of the DGPs 3 to 8 which are not unity have the following characteristics:

DGP 3 : medium AR and medium MA eigenvalues.

DGP 4 : medium AR and medium MA eigenvalues, nonzero $\nu = (0.1, 0.2, 0.2)'$.

DGP 5 : medium AR and large negative MA eigenvalues.

DGP 6 : medium AR and large positive MA eigenvalues.

DGP 7 : large negative AR and medium MA eigenvalues.

DGP 8 : large positive AR and medium MA eigenvalues.

These processes which are characterized by the magnitude of their eigenvalues have the following real eigenvalues and corresponding coefficients b_1, b_2 and m_1, m_2 of the autoregressive and moving average part.

AR part:

Large positive AR eigenvalues: $\lambda_1^{ar} = 0.95$, $\lambda_2^{ar} = 0.7 \Longrightarrow b_1 = \frac{159}{1120} \approx 0.14$, $b_2 = -\frac{27}{160} \approx 0.17$.

Medium AR eigenvalues: $\lambda_1^{ar} = 0.7$, $\lambda_2^{ar} = 0.4 \Longrightarrow b_1 = \frac{101}{140} \approx 0.72$, $b_2 = -\frac{13}{20} = 0.65$.

Large negative AR eigenvalues: $\lambda_1^{ar} = -0.95$, $\lambda_2^{ar} = -0.7 \Longrightarrow b_1 = \frac{777}{160} \approx 4.86$, $b_2 = -\frac{27}{160} \approx 0.17$.

MA part:

Large positive MA eigenvalues: $\lambda_1^{ma} = 0.95$, $\lambda_2^{ma} = 0.7 \Longrightarrow m_1 = -\frac{21}{20} = 1.05$, $m_2 = \frac{7}{200} \approx 0.04$.

Medium MA eigenvalues: $\lambda_1^{ma} = 0.6$, $\lambda_2^{ma} = -0.5 \Longrightarrow m_1 = \frac{1}{2} = 0.5$, $m_2 = 0$.

Large negative MA eigenvalues: $\lambda_1^{ma} = -0.95$, $\lambda_2^{ma} = -0.7 \Longrightarrow m_1 = \frac{9}{4} = 2.25$, $m_2 = \frac{403}{200} \approx 2.02$.

The intercept term $\nu = (0.1, 0.2, 0.2)'$ has roughly the same size of the original estimation by Lütkepohl and Claessen (1997) with one element deleted and the remaining elements rounded to one digit precision. As will be seen in the simulation study, this intercept of DGP 4 has only little influence on the estimation of the Kronecker indices.

4.2 Simulation Design

In the following Monte Carlo simulation Stage I is combined with Stages II(PL1), II(PL2), and II(WN), respectively. These three methods all have in common that they do not condition on the cointegrating rank ϱ which would have to be estimated in advance otherwise. The methods PL1, PL2 and WN estimate the individual equations of the system separately and they use order selection criteria. Since the methods require estimation of the individual equations only, the computational burden of the order search procedures is reduced dramatically relative to procedures working on the full system simultaneously. The reason is that the multidimensional search problem is split into K separate one dimensional search procedures.

In Stage I a choice of the order h_T of the long VAR process fitted to the VARMA realizations has to be made. Moreover, in the different versions of Stage II the weighting functions C_T or c_T must be chosen. For C_T the proposals of Poskitt and Lütkepohl (1995) mentioned earlier will be used. In total the 6 combinations of different long VAR orders h_T and penalty terms C_T and c_T given in Table 2 are considered.

Strictly speaking for consistency results to hold, a has to be greater than one in $(\log T)^a$. Nevertheless it is of interest to check the borderline case $a = 1$ in Simulation Designs 3 and 4.

Table 2: Design Characteristics of the Simulations

Design	h_T	C_T	c_T
1	$h(AIC)$	$h_T \log T$	2 (WN–AIC)
2	$h(AIC)$	h_T^2	$\log T$ (WN–SC)
3	$\max\{\log T, h(AIC)\}$	$h_T \log T$	2 (WN–AIC)
4	$\max\{\log T, h(AIC)\}$	h_T^2	$\log T$ (WN–SC)
5	$\max\{(\log T)^{1.5}, h(AIC)\}$	$h_T \log T$	2 (WN–AIC)
6	$\max\{(\log T)^{1.5}, h(AIC)\}$	h_T^2	$\log T$ (WN–SC)

Of course, $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1.5$ is greater or equal to $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1$. So the former has a tendancy to fit higher order VARs to the data. It can be seen in Table 3 that the latter quantity in most cases is strictly greater than the order chosen by AIC. Exceptions are the DGPs 5 and 6, where the moving average part takes on extreme (negative or positive) eigenvalues.

The results of methods PL1 and PL2 are influenced by the choice of the penalty function C_T. For Designs 3 and 4 we have $h_T \log T \leq h_T^2$ so that in the former there is a tendency to choose higher row orders and, hence, Kronecker indices than in Design 4. For Designs 5 and 6 the strict inequality $h_T \log T < h_T^2$ holds, whereas for Designs 1 and 2 we have $h_T \log T > h_T^2$ in more than 80 percent of all cases except for DGPs 5 and 6, where $h_T \log T \leq h_T^2$ in more than 85 percent of all cases (see Table 3). Method WN on the other hand is influenced by the specific form of the penalty function c_T. The term $c_T = 2$ corresponds to Akaikes Information Criterion AIC and chooses at least as large orders as $c_T = \log T$ which corresponds to the penalty term of the Schwarz Criterion SC. Depending on the penalty function c_T chosen, the white noise procedures are denoted by WN–AIC or WN–SC, respectively.

For PL1 and PL2 the maximum Kronecker index was chosen to be $P_T = \text{ceil}(\frac{1}{2}h_T)$, where ceil is the ceiling function which rounds up to the nearest larger integer. Poskitt and Lütkepohl (1995) note that equating the number of freely varying coefficients in each equation of the ARMA_E system obtained when $p_k = P_T$, $k = 1, \ldots, K$, with that in the autoregressive approximation gives the rule that $P_T = \frac{1}{2}h_T$ should not be exceeded. The value Q_T which is the maximum order of the autoregressive process fitted to the residuals in Stage II(WN) is determined by $Q_T = h_T - P_T - 1$. Thus, the condition $P_T + Q_T < h_T$ is fulfilled which is

Table 3: Percentage of Cases in Designs 1 and 2 where $h(AIC)$ is Greater or equal to the Deterministic Criterion $(\log T)^{1.0}$: $P^*(h(AIC) \geq (\log T)^{1.0})$.

DGP	$T = 150$	$T = 500$
1	0.05	0.01
2	0.05	0.00
3	0.12	0.04
4	0.16	0.03
5	1.00	1.00
6	0.87	1.00
7	0.14	0.05
8	0.16	0.05

The maximum order for the search by AIC was set to $h_{max}^{AIC} = 1.5(\log T)^{1.0}$ for Designs 1 to 4. This maximum has not been chosen in any of the replications of Designs 1 and 2. Each percentage tabulated here has been calculated using all $2 \cdot 200 = 400$ replications from Designs 1 and 2.

necessary to avoid zero residual variances $\tilde{\sigma}_{k,\varepsilon}^2(q)$.

For some replications, choosing h_T by AIC results in a numerical collinearity problem for DGP 2 which consists of independet random walks. To overcome this problem a lower bound $h_T \geq 2$ should be introduced for methods PL1 and PL2. This excludes the case $P_T = h_T$ when using $P_T = \text{ceil}(\frac{1}{2}h_T)$. Thus near collinearity is avoided which occurs between the colums of the LS regressor matrix if the DGP consits of random walks as in DGP 2. In this simulation study an even larger lower bound $h_T \geq 4$ was chosen in order to guarantee $Q_T = h_T - P_T - 1 \geq 1$. The restriction $h_T \geq 4$ is acceptable since in practice a long VAR approximation would have at least this order.

In Stage I an upper bound, h_{max}^{AIC} say, for the order of the fitted long VAR process has to be specified. In Designs 1, 2, 3 and 4 the AIC criterion searches up to a maximum order of $h_{max}^{AIC} = 1.5(\log T)^{1.0}$, whereas the maximum order for AIC in Simulation Designs 5 and 6 is $h_{max}^{AIC} = 1.5(\log T)^{1.5}$. That is, AIC is computed for orders which are up to 50 percent higher than the values of the deterministic order criteria $(\log T)^a$ with $a = 1$ and $a = 1.5$ respectively. Since in Designs 1 and 2 the maximum h_{max}^{AIC} was never chosen this bound for the order of the long VAR process seems to be sensible.

In Stages II(PL2) and II(WN) it is important to avoid a bias introduced by introducing an a proiri ordering of the variables. Therefore, from all variables not yet fixed which have the same smallest estimated Kronecker index one variable is chosen *randomly* and fixed for the following steps. Another rule to be checked in future simulations is, of course, to choose among all variables with the same smallest estimated Kronecker index the one which has the smallest value of the order selection criterion. This alternative rule also avoids an influence of a specific given order of the variables.

In the simulation study the number of replications was set to 200. When a new time series is generated its presample values are set to zero and 50 observations are discarded at the beginning of the time series. Only the last $T_1 = 150$ or $T_2 = 500$ observations are kept.

4.3 Simulation Results

In Tables 4 to 9 all sets of estimated Kronecker indices and their relative frequencies are presented. A set $\hat{\bar{p}}$ of Kronecker indices is listed if it has been chosen by at least one of the three procedures PL1, PL2 or WN in more than 10 percent of the replications. For each DGP the rows are sorted by the mean percentage. The true set $\bar{p}$ of Kronecker indices is always presented even if it has been chosen in less than 10 percent of the replications by all the procedures. The true set of Kronecker indices is marked by an asterisk. Figures 1 to 3 show an extract of the information from Tables 4 to 9 to provide a quick overview of some important results. For each of the three methods, PL1, PL2 and WN, the relative frequency of replications is shown for estimating the Kronecker indices correctly. These three figures give a good impression of the performace of the three methods since the true set of Kronecker indices is often identified in more than 50 percent of the replications and so the majority of replications is included in Figures 1 to 3. Before interpreting the results, it should be mentioned that the methods PL1 and PL2 are consistent under suitable conditions as shown in Poskitt and Lütkepohl (1995), whereas no such results exist for the WN method.

When looking at Figures 1 to 3 it becomes evident that the WN method cannot be recommended because, for instance, the processes DGP 3 and DGP 4 are identified correctly in less than 20 percent of the replications irrespective of the choice of h_T or c_T. However, it must be admitted that the method WN is not working terribly bad, since in the case of medium moving average eigenvalues as for example in DGPs 3 and 4 we have $m_2 = 0$ and therefore $M_2 = 0$.

Thus the maximum moving average order is only 1. So it is not surprising that the Kronecker indices are often estimated to be $\hat{\bar{p}} = (1, 1, 1)$. This is exactly what we can see from Table 9, for example, where for DGPs 3 and 4 the Kronecker indices $\hat{\bar{p}} = (1, 1, 1)$ have been estimated with relative frequencies ranging from 71 up to 91 percent. However, as can be seen there as well, these estimates are also found frequently by the PL1 and PL2 methods. Overall, based on the frequency of correctly estimating the Kronecker indices the result is that the WN method cannot compete with the PL1 and PL2 methods.

It can be seen in Tables 4 to 9 that the estimated Kronecker indices presented, in some cases sum up to less than 70 percent in total for the WN method. This is because there is a relatively high dispersion over the whole range of orders up to the maximum order P_T even for the large sample size $T_2 = 500$. Of course, the dispersion of WN is reduced a bit in Designs 1 and 2 where $h_T = h(AIC)$ often is relatively small and therefore $P_T = \text{ceil}(\frac{1}{2}h_T)$ is relatively small, too. But even in these cases the performance of WN is not improved. It is still not quite clear why WN does not work well given that it was found to perform well in other studies. One possible explanation is that the residual white noise test is applied here to residuals from a *non*stationary time series whereas Koreisha and Pukkila proved a good performance only for multivariate stationary series (Koreisha and Pukkila (1993)) or for univariate series which have been made stationary by differencing (Koreisha and Pukkila (1995)). If WN is used at all, the method WN–SC, see Designs 2, 4 and 6, should be used since it has a better performance than WN–AIC which is used in Simulation Designs 1, 3 and 5. This result is in line with Koreisha and Pukkila (1993, 1995) who found that their residual white noise test in conjunction with the AIC penalty function often overestimates the true orders.

When comparing Figures 1 and 2 it can be seen that the method PL2 is always more successful in finding the true model structure than PL1. The corresponding figures from Table 7, e.g., show that there is an increase of approximately 20 percentage points for estimating the true Kronecker indices when using PL2. Thus, incorporating the echelon form restrictions already during the sequential specification procedure as in PL2 helps in estimating the Kronecker indices correctly. The method PL2 is the best of all three methods compared within this study. The percentage of correctly estimated Kronecker indices is quite high given that each single index has to be specified correctly to be counted here. The performace of PL2 in Simulation Designs 1 to 4 is very similar. Using the penalty function $C_T = h_T^2$ as in Designs 2 and 4 is slightly

preferable.

In Stage I the long VAR order should be chosen as $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1$ when used in conjunction with PL2 since this h_T leads to a performance similar to that in Simulation Designs 1 and 2 and it also satisfies the requirement that h_T should increase at least with rate $\log T$. This requirement is used for deriving asymptotic results as pointed out by Poskitt and Lütkepohl (1995, p. 13). For DGP 5 the choice $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1.5$ is preferable, however. In summary, we have a recommendation for all processes except for processes with strongly negative eigenvalues of the moving average part (DGP 5): it is recommended to use the method PL2 in conjunction with long VAR order $h_T = \max\{(\log T)^a, h(AIC)\}$ and $a = 1$ for Stage I and penalty function $C_T = h_T^2$ for Stage II, as in Simulation Design 4. Of course, an argument in favor of $C_T = h_T^2$ is parameter parsimony since this choice of C_T has a tendancy to result in lower orders. As can be seen from Table 7 if the set of Kronecker indices is not estimated exactly correct the deviations are only small ones. In the majority of these cases only one Kronecker index differs slightly from the true one. In this sense the PL2 method provides reliable estimates of the Kronecker indices.

There are some observations which can be made throughout Simulation Designs 1 to 6. If the eigenvalues of the MA part are strongly negative (DGP 5) none of the procedures is working well. As can be seen from Table 9, for example, there is a strong tendency to overestimate the third Kronecker index. This is plausible because in Stage I the autoregressive order h_T of the VAR process presumably is not high enough in order to approximate a moving average part with large negative eigenvalues. As can be seen from Figure 2, Simulation Design 6 with a large value of $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1.5$ is most succesful in estimating the true Kronecker indices of DGP 5. For DGP 5 and DGP 6 with extreme (positive or negative) moving average eigenvalues increasing the sample size from $T_1 = 150$ to $T_2 = 500$ does not help in estimating the true Kronecker indices with a higher probability. For the DGPs 3 and 4 all three methods lead to very similar estimated Kronecker indices. Obviously the chosen intercept term ν has no substantial impact on the results.

A property all methods have in common is that they are able to estimate the very simple processes DGP 1 (white noise) and DGP 2 (independent random walks) reliably. Of course when the structures become more complicated the method PL2 dominated as already mentioned.

In summary, the PL2 method is preferred over PL1 and WN. It should be used together

with $h_T = \max\{(\log T)^a, h(AIC)\}$ and $a = 1$ and $C_T = h_T^2$. Although the methods PL1 and PL2 often behave similarly, PL2 is usually more successful in finding the true Kronecker indices. Hence, the additional computational burden for PL2 seems to be justified at least for processes of moderate dimension.

5 Conclusions

The echelon form can be used to parameterize cointegrated VARMA models. The main advantage of cointegrated VARMA models against standard cointegrated VAR models is their parameter parsimony together with the implied potential improvement in forecast precision. The Kronecker indices which characterize the echelon form have to be estimated at the specification stage before conducting a detailed VARMA cointegration analysis. In this paper two stage procedures are investigated. In Stage I, a long VAR(h_T) approximation is fitted to the data. Stage I is followed by one of three alternative versions of Stage II. These three different methods for estimating the Kronecker indices of cointegrated echelon form VARMA models are discussed and compared in a simulation study.

The methods discussed here estimate the equations of the system separately and selection criteria are applied to these equations or to their residuals as in the case of the method WN. Due to this setup, the computer intensive multidimensional full search procedures (see e.g. Lütkepohl (1991, section 8.3.2)) known from the stationary case are split into one dimensional search procedures. The computational complexity is very moderate because all necessary calculations are exclusively based on linear least squares methods.

The Monte Carlo simulations show that a reliable estimate of the Kronecker indices is possible with a sample size of $T = 150$. It is suggested that the method PL2 should be preferred over PL1 and WN. PL2 should be used in combination with a long VAR order $h_T = \max\{(\log T)^a, h(AIC)\}$ with $a = 1$ and penalty function $C_T = h_T^2$. Although PL1 and PL2 often behave similarily, PL2 is more successful in finding the true Kronecker indices since the echelon restrictions found in preceeding steps of the sequential specification procedure are immediately incorporated into the estimation process. The additional computational burden of PL2 seems to be justified at least for processes of moderate dimension since the probability of estimating the Kronecker indices correctly increases a bit when using PL2 instead of PL1. The method WN cannot be recommended for at least two reasons. First, in some cases it estimates

the true Kronecker indices much less frequently than the other methods. Second, for most cases considered here, its performance does not improve much when the sample size increases and, thus, more sample information becomes available.

An interesting extension of this simulation study might be to include a procedure proposed by Poskitt (1996). This method fits into the framework considered here because the equations are estimated separately and a selection criterion is used as well. This method is similar to Stage II(PL2) but the time series is not analyzed in levels but the error correction form of the ARMA_E system is used. This, of course, requires the estimation of the cointegrating rank and the cointegrating basis in a prior step.

References

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

Engle, R.F. and Granger, C.W.J. (1987), Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55, 251-276.

Guo, L., Huang, D.W. and Hannan, E.J. (1990), On ARX(∞) Approximation. *Journal of Multivariate Analysis*, 32, 17-47.

Hannan, E.J. and Kavalieris, L. (1984), Multivariate Time Series Models. *Advances in Applied Probability*, 16, 492 - 561.

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

Huang, D. and Guo, L. (1990), Estimation of Nonstationary ARMAX Models Based on the Hannan-Rissanen Method. *Annals of Statistics*, 18, 1729 - 1756.

Koreisha, S.G. and Pukkila, T. (1993), Determining the Order of a Vector Autoregression when the Number of Component Series is Large. *Journal of Time Series Analysis*, 14, 47 - 69.

Koreisha, S.G. and Pukkila, T. (1995), A Comparison Between Different Order-Determination Criteria for Identification of ARIMA Models. *Journal of Business & Economic Statistics*

13, 127 - 131.

Koreisha, S.G. and Yoshimoto, G. (1991), A Comparison among Identification Procedures for Autoregressive Moving Average Models. *International Statistical Review*, 59, 37 - 57.

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

Lütkepohl, H. and Claessen, H. (1997), Analysis of Cointegrated VARMA Processes. *Journal of Econometrics,* forthcoming.

Lütkepohl, H. and Poskitt, D.S. (1996), Specification of Echelon Form VARMA Models. *Journal of Business and Economic Statistics*, 14, 69 - 79.

Nsiri, S. and Roy, R. (1992), On the Identification of ARMA Echelon–Form Models, *Canadian Journal of Statistics*, 20, 369 - 386.

Poskitt, D.S. (1992), Identification of Echelon Canonical Forms for Vector Linear Processes Using Least Squares. *Annals of Statistics*, 20, 195 - 215.

Poskitt, D.S. (1996), The Analysis of Cointegrated Autoregressive Moving-Average Systems. Institut für Statistik und Ökonometrie, Humboldt–Universität zu Berlin, Discussion Paper 58, 1996.

Poskitt, D.S. and Lütkepohl, H. (1995), Consistent Specification of Cointegrated Autoregressive Moving-Average Systems. Institut für Statistik und Ökonometrie, Humboldt–Universität zu Berlin, Discussion Paper 54, 1995.

Pukkila, T., Koreisha, S. and Kallinin, A. (1990), The Identification of ARMA Models. *Biometrika*, 77, 537 - 548.

Quenouille, M.A. (1957), *The Analysis of Multiple Time–Series.* Charles Griffin, London.

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

Yap, S.F. and 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.

Table 4: Relative Frequencies of Estimated Kronecker Indices. **Design 1**: h_T chosen by AIC, $C_T = h_T \log T$, $c_T = 2$ (WN–AIC). The true indices are marked by an asterisk.

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	0.99	1.00	0.55	(0,0,0)*	1.00	1.00	0.54
(1,0,0)	0.00	0.00	0.13	(1,0,0)	0.00	0.00	0.14
(0,0,1)	0.01	0.00	0.12	(0,0,1)	0.00	0.00	0.12
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	0.99	0.99	0.73	(1,1,1)*	1.00	1.00	0.80
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.55	0.31	0.69	(1,1,1)	0.59	0.10	0.66
(2,1,1)*	0.40	0.63	0.10	(2,1,1)*	0.41	0.70	0.14
				(1,2,1)	0.00	0.20	0.07
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.57	0.30	0.69	(1,1,1)	0.61	0.07	0.67
(2,1,1)*	0.33	0.59	0.07	(2,1,1)*	0.37	0.69	0.12
				(1,2,1)	0.00	0.23	0.09
DGP 5 $\bar{p} = (2,1,1)$							
(2,1,3)	0.37	0.62	0.09	(2,1,2)	0.55	0.23	0.13
(2,1,2)	0.47	0.17	0.14	(2,1,3)	0.20	0.56	0.11
(2,1,4)	0.05	0.13	0.20	(2,1,4)	0.01	0.04	0.14
(2,1,1)*	0.01	0.03	0.10	(1,1,2)	0.14	0.00	0.00
				(2,1,1)*	0.01	0.04	0.05
DGP 6 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.61	0.76	0.24	(2,1,1)*	0.57	0.74	0.20
(2,1,2)	0.15	0.15	0.16	(2,1,2)	0.12	0.10	0.16
(1,1,1)	0.17	0.04	0.09	(1,1,1)	0.29	0.00	0.01
				(1,2,1)	0.00	0.15	0.03
				(2,1,3)	0.00	0.00	0.14
DGP 7 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.53	0.72	0.33	(2,1,1)*	0.56	0.72	0.36
(1,2,1)	0.00	0.23	0.40	(1,2,1)	0.00	0.27	0.29
(1,1,1)	0.41	0.00	0.01	(1,1,1)	0.42	0.00	0.00
DGP 8 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.65	0.83	0.38	(2,1,1)*	0.61	0.80	0.39
(1,1,1)	0.26	0.01	0.17	(1,2,1)	0.00	0.20	0.41
(1,2,1)	0.00	0.10	0.23	(1,1,1)	0.39	0.00	0.01

Table 5: Relative Frequencies of Estimated Kronecker Indices. **Design 2**: h_T chosen by AIC, $C_T = h_T^2$, $c_T = \log T$ (WN–SC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	0.97	0.99	0.89	(0,0,0)*	0.93	0.98	0.94
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	0.97	0.97	0.97	(1,1,1)*	0.97	1.00	0.97
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.49	0.31	0.93	(2,1,1)*	0.63	0.84	0.03
(2,1,1)*	0.39	0.59	0.01	(1,1,1)	0.33	0.04	0.95
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.49	0.26	0.97	(1,1,1)	0.41	0.04	0.95
(2,1,1)*	0.39	0.59	0.01	(2,1,1)*	0.53	0.77	0.03
				(1,2,1)	0.00	0.14	0.01
DGP 5 $\bar{p} = (2,1,1)$							
(2,1,2)	0.42	0.25	0.12	(2,1,2)	0.30	0.24	0.15
(2,1,3)	0.14	0.38	0.12	(2,1,3)	0.10	0.32	0.06
(2,1,1)*	0.10	0.20	0.23	(1,1,2)	0.41	0.00	0.00
(1,2,1)	0.00	0.09	0.23	(1,2,1)	0.00	0.20	0.21
(1,1,2)	0.20	0.00	0.00	(2,1,1)*	0.04	0.12	0.21
(1,2,2)	0.00	0.00	0.16	(1,2,2)	0.00	0.00	0.21
DGP 6 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.39	0.62	0.42	(2,1,1)*	0.37	0.66	0.52
(1,1,1)	0.47	0.25	0.38	(1,1,1)	0.61	0.07	0.23
				(1,2,1)	0.00	0.26	0.04
DGP 7 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.62	0.81	0.42	(2,1,1)*	0.68	0.84	0.40
(1,2,1)	0.00	0.12	0.42	(1,2,1)	0.00	0.12	0.43
(1,1,1)	0.25	0.00	0.03	(1,1,1)	0.23	0.00	0.00
DGP 8 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.57	0.72	0.51	(2,1,1)*	0.69	0.85	0.49
(1,1,1)	0.28	0.07	0.39	(1,2,1)	0.00	0.10	0.41
				(1,1,1)	0.23	0.00	0.07

Table 6: Relative Frequencies of Estimated Kronecker Indices. **Design 3**: $h_T = \max\{(\log T)^a, h(AIC)\}$ where $a = 1.0$, $C_T = h_T \log T$, $c_T = 2$ (WN–AIC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	0.99	1.00	0.51	(0,0,0)*	1.00	1.00	0.34
(0,1,0)	0.00	0.00	0.13	(0,0,1)	0.00	0.00	0.14
(0,0,1)	0.01	0.00	0.11				
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	0.99	1.00	0.73	(1,1,1)*	1.00	1.00	0.65
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.56	0.35	0.69	(2,1,1)*	0.45	0.73	0.18
(2,1,1)*	0.39	0.60	0.10	(1,1,1)	0.55	0.13	0.40
				(1,2,1)	0.00	0.14	0.15
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.57	0.29	0.77	(2,1,1)*	0.39	0.69	0.17
(2,1,1)*	0.34	0.63	0.04	(1,1,1)	0.61	0.16	0.42
				(1,2,1)	0.00	0.14	0.13
DGP 5 $\bar{p} = (2,1,1)$							
(2,1,3)	0.36	0.51	0.09	(2,1,3)	0.32	0.59	0.09
(2,1,2)	0.42	0.22	0.14	(2,1,2)	0.48	0.20	0.08
(2,1,4)	0.07	0.16	0.16	(2,1,1)*	0.01	0.05	0.07
(2,1,1)*	0.01	0.03	0.07				
DGP 6 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.53	0.71	0.23	(2,1,1)*	0.58	0.75	0.18
(2,1,2)	0.14	0.14	0.18	(1,1,1)	0.33	0.01	0.01
(1,1,1)	0.26	0.10	0.07	(1,2,1)	0.00	0.19	0.04
				(2,1,2)	0.04	0.06	0.11
				(2,1,3)	0.00	0.00	0.15
DGP 7 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.68	0.85	0.41	(2,1,1)*	0.58	0.79	0.39
(1,2,1)	0.00	0.10	0.32	(1,2,1)	0.00	0.21	0.27
(1,1,1)	0.24	0.00	0.00	(1,1,1)	0.41	0.00	0.00
DGP 8 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.71	0.89	0.46	(2,1,1)*	0.56	0.80	0.38
(1,1,1)	0.24	0.04	0.17	(1,1,1)	0.44	0.00	0.01
(1,2,1)	0.00	0.04	0.17	(1,2,1)	0.00	0.20	0.21

Table 7: Relative Frequencies of Estimated Kronecker Indices. **Design 4**: $h_T = \max\{(\log T)^a, h(AIC)\}$ where $a = 1.0$, $C_T = h_T^2$, $c_T = \log T$ (WN–SC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	0.99	1.00	0.89	(0,0,0)*	1.00	1.00	0.92
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	0.99	1.00	0.96	(1,1,1)*	1.00	1.00	0.97
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.62	0.44	0.93	(1,1,1)	0.56	0.22	0.97
(2,1,1)*	0.33	0.55	0.02	(2,1,1)*	0.44	0.73	0.01
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.60	0.33	0.93	(1,1,1)	0.51	0.14	0.94
(2,1,1)*	0.36	0.61	0.01	(2,1,1)*	0.48	0.76	0.02
DGP 5 $\bar{p} = (2,1,1)$							
(2,1,2)	0.41	0.26	0.14	(2,1,2)	0.35	0.23	0.18
(2,1,3)	0.12	0.34	0.04	(2,1,1)*	0.08	0.17	0.27
(2,1,1)*	0.08	0.15	0.26	(2,1,3)	0.06	0.30	0.07
(1,2,1)	0.00	0.14	0.22	(1,1,2)	0.34	0.00	0.00
(1,1,2)	0.24	0.00	0.01	(1,2,1)	0.00	0.14	0.18
(1,2,2)	0.00	0.00	0.20	(1,2,2)	0.00	0.00	0.21
				(1,2,3)	0.00	0.14	0.05
DGP 6 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.43	0.68	0.51	(2,1,1)*	0.29	0.66	0.52
(1,1,1)	0.45	0.17	0.30	(1,1,1)	0.69	0.07	0.28
				(1,2,1)	0.00	0.27	0.04
DGP 7 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.63	0.76	0.41	(2,1,1)*	0.59	0.84	0.58
(1,2,1)	0.00	0.21	0.37	(1,2,1)	0.00	0.16	0.40
(1,1,1)	0.34	0.00	0.02	(1,1,1)	0.41	0.00	0.00
DGP 8 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.67	0.87	0.53	(2,1,1)*	0.62	0.78	0.48
(1,1,1)	0.28	0.03	0.33	(2,2,1)	0.01	0.21	0.46
				(1,1,1)	0.37	0.00	0.01

Table 8: Relative Frequencies of Estimated Kronecker Indices. **Design 5**: $h_T = \max\{(\log T)^a, h(AIC)\}$ where $a = 1.5$, $C_T = h_T \log T$, $c_T = 2$ (WN–AIC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	1.00	1.00	0.20	(0,0,0)*	1.00	1.00	0.18
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	1.00	1.00	0.38	(1,1,1)*	1.00	1.00	0.41
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.89	0.84	0.20	(1,1,1)	0.86	0.48	0.10
(2,1,1)*	0.11	0.16	0.07	(2,1,1)*	0.14	0.52	0.14
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.93	0.84	0.11	(1,1,1)	0.83	0.45	0.17
(2,1,1)*	0.07	0.15	0.04	(2,1,1)*	0.17	0.56	0.14
DGP 5 $\bar{p} = (2,1,1)$							
(2,1,3)	0.47	0.72	0.02	(2,1,3)	0.43	0.79	0.07
(2,1,2)	0.45	0.15	0.10	(2,1,2)	0.51	0.15	0.04
(2,1,5)	0.00	0.00	0.14	(2,1,4)	0.00	0.01	0.17
(2,1,1)*	0.04	0.07	0.01	(2,1,5)	0.00	0.00	0.13
				(2,1,1)*	0.00	0.02	0.01
DGP 6 $\bar{p} = (2,1,1)$							
(0,1,1)	0.46	0.45	0.00	(2,1,1)*	0.68	0.83	0.12
(2,1,1)*	0.28	0.40	0.11	(1,1,1)	0.33	0.09	0.00
(1,1,1)	0.24	0.13	0.00				
(2,1,2)	0.01	0.01	0.14				
DGP 7 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.85	0.94	0.12	(2,1,1)*	0.88	0.95	0.18
(1,1,1)	0.14	0.00	0.00	(1,1,1)	0.12	0.00	0.00
(2,1,2)	0.00	0.00	0.14	(2,1,2)	0.00	0.00	0.11
DGP 8 $\bar{p} = (2,1,1)$							
(2,1,1)*	0.57	0.76	0.12	(2,1,1)*	0.80	0.89	0.24
(1,1,1)	0.42	0.23	0.01	(1,1,1)	0.20	0.00	0.00
				(1,2,1)	0.00	0.11	0.04

Table 9: Relative Frequencies of Estimated Kronecker Indices. **Simulation 6**: $h_T = \max\{(\log T)^a, h(AIC)\}$ where $a = 1.5$, $C_T = h_T^2$, $c_T = \log T$ (WN–SC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 1 $\bar{p} = (0,0,0)$							
(0,0,0)*	1.00	1.00	0.80	(0,0,0)*	1.00	1.00	0.94
DGP 2 $\bar{p} = (1,1,1)$							
(1,1,1)*	0.79	0.79	0.91	(1,1,1)*	1.00	1.00	0.98
DGP 3 $\bar{p} = (2,1,1)$							
(1,1,1)	0.64	0.64	0.73	(1,1,1)	1.00	1.00	0.88
(0,1,1)	0.33	0.34	0.00	(2,1,1)*	0.00	0.00	0.05
(1,1,2)	0.00	0.00	0.12				
(2,1,1)*	0.00	0.00	0.04				
DGP 4 $\bar{p} = (2,1,1)$							
(1,1,1)	0.61	0.61	0.71	(1,1,1)	1.00	1.00	0.91
(0,1,1)	0.36	0.38	0.00	(2,1,1)*	0.00	0.00	0.04
(1,1,2)	0.00	0.00	0.13				
(2,1,1)*	0.00	0.00	0.03				
DGP 5 $\bar{p} = (2,1,1)$							
(1,1,1)	0.79	0.26	0.01	(1,1,1)	0.89	0.00	0.00
(2,1,1)*	0.03	0.42	0.16	(2,1,1)*	0.01	0.54	0.26
(1,2,1)	0.00	0.16	0.14	(1,2,1)	0.00	0.45	0.10
(1,1,0)	0.14	0.12	0.00	(2,1,2)	0.01	0.01	0.14
(2,1,2)	0.01	0.02	0.21	(1,2,2)	0.00	0.00	0.15
(1,2,2)	0.00	0.00	0.17	(2,1,3)	0.00	0.01	0.12
(2,1,4)	0.00	0.00	0.12				
DGP 6 $\bar{p} = (2,1,1)$							
(0,1,1)	0.55	0.69	0.00	(1,1,1)	0.76	0.35	0.05
(0,0,1)	0.26	0.14	0.00	(2,1,1)*	0.00	0.40	0.69
(2,1,1)*	0.00	0.00	0.40	(0,0,1)	0.24	0.24	0.00
(2,1,2)	0.00	0.00	0.20				
(1,1,1)	0.00	0.00	0.20				
(0,0,0)	0.18	0.00	0.00				

Table 9 (continued): Relative Frequencies of Estimated Kronecker Indices. **Design 6**: $h_T = \max\{(\log T)^a,\ h(AIC)\}$ where $a = 1.5$, $C_T = h_T^2$, $c_T = \log T$ (WN–SC).

	T=150				**T=500**		
$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**	$(\hat{p}_1, \hat{p}_2, \hat{p}_3)$	**PL1**	**PL2**	**WN**
DGP 7 $\bar{p} = (2,1,1)$							
(1,1,1)	0.96	0.37	0.02	(2,1,1)*	0.00	0.53	0.61
(2,1,1)*	0.02	0.43	0.45	(1,1,1)	1.00	0.00	0.00
(1,2,1)	0.00	0.20	0.20	(1,2,1)	0.00	0.47	0.34
(2,1,2)	0.00	0.00	0.14				
DGP 8 $\bar{p} = (2,1,1)$							
(1,1,1)	0.98	0.86	0.12	(1,1,1)	1.00	0.47	0.01
(2,1,1)*	0.00	0.12	0.47	(2,1,1)*	0.00	0.54	0.58
(1,2,1)	0.00	0.00	0.15	(1,2,1)	0.00	0.00	0.36

Figure 1: Method PL1. Relative Frequencies for Estimating the True Set of Kronecker Indices Correctly.

Figure 2: Method PL2. Relative Frequencies for Estimating the True Set of Kronecker Indices Correctly.

Figure 3: Method WN. Relative Frequencies for Estimating the True Set of Kronecker Indices Correctly.