

Model specification and selection for multivariate time series

Rajendra J. Bhansali

Department of Mathematics, Imperial College London, South Kensington Campus, Huxley Building, London, SW7 2AZ,
United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:

Received 16 July 2018

Received in revised form 16 August 2019

Accepted 16 August 2019

Available online 19 September 2019

AMS 2010 subject classifications:

primary 62M10

secondary 62M20

Keywords:

Canonical correlations

Reversed echelon form

State space representation

VARMA model

ABSTRACT

Three major difficulties are identified with an established echelon form approach (see Hannan (1987)) to specifying a Vector Autoregressive Moving Average, VARMA, model for an observed time series. A family of state space representations, valid for each integer, h , is introduced, and collectively referred to as multistep state space representations. This family includes as its special case, with $h = 0$, a state space representation introduced earlier by Akaike (1974), and, with $h = 1$, that introduced by Cooper and Wood (1982). Appropriate generalizations of the notions of minimality, McMillan degree, left matrix fraction description and Kronecker indices, as applicable individually to each member of this family, are presented. The reverse echelon form and state space representation corresponding to the Kronecker indices for each h are derived, and the former illustrated with three examples of standard VARMA processes. The question of how the presence of zero constraints on the coefficients of a reverse echelon form may be detected solely from an inspection of the Kronecker indices is examined. A canonical correlation procedure proposed originally by Akaike (1976) for $h = 0$ is considered for estimating the Kronecker indices with each h . The efficacy of the estimation procedure is investigated by a simulation study. A procedure is suggested for implementing the new approach introduced in this paper with an observed time series, and three different applications of this approach are outlined. This approach is also related to some of its alternatives, including the Kronecker invariants of Poskitt (1992) and the scalar component approach of Tiao and Tsay (1989).

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Let $y(t) = [y_1(t), \dots, y_s(t)]^\top$, $t \in \{1, \dots, T\}$ denote an observed multivariate time series of dimension $s \geq 1$ and length $T \geq 1$. Suppose that $y(t)$ is a part-realization of a discrete-time process, $\{y(t)\}$, satisfying the following assumption:

Assumption 1. That $\{y(t), t \in Z\}$ is a vector autoregressive moving average process of order (p, q) , VARMA(p, q) process,

$$\sum_{u=0}^p A(u)y(t-u) = \sum_{u=0}^q B(u)\epsilon(t-u), \quad (1)$$

where $Z \in \{0, \pm 1, \pm 2, \dots\}$ is a set of all integers, $A(0) = B(0) = I_s$, an $s \times s$ identity matrix, the $A(u)$ and $B(u)$ are matrices of coefficients such that if

$$a(z) = \sum_{u=0}^p A(u)z^u, \quad b(z) = \sum_{u=0}^q B(u)z^u,$$

E-mail addresses: sa17@liv.ac.uk, r.bhansali@ic.ac.uk.

<https://doi.org/10.1016/j.jmva.2019.104539>

0047-259X/© 2019 Elsevier Inc. All rights reserved.

$[a(z) \ b(z)]$ is left coprime, $\det[a(z)] \neq 0$, $|z| \leq 1$, $\det[b(z)] \neq 0$, $|z| < 1$, and $\{\epsilon(t)\}$ is a sequence of uncorrelated random variables, each with mean zero and a positive definite covariance matrix, Ψ .

As in Hannan [21], let

$$k(z) = a(z)^{-1}b(z) = \sum_{j=0}^{\infty} K(j)z^j, \quad K(0) = I_s,$$

define the corresponding rational transfer function, and $R(u) = E\{y(t+u)y(t)^T\}$, $t, u \in Z$, the covariance function. Then, $\{y(t)\}$ has the infinite moving average representation:

$$y(t) = \sum_{j=0}^{\infty} K(j)\epsilon(t-j), \quad (2)$$

and which is one of the several different solutions to the difference equation (1). It will also be convenient to set $K(j) = 0$, if $j < 0$. The matrices, $K(j)$, occurring on the right of this representation define the impulse response coefficients of $\{y(t)\}$.

As is well-known, a major stumbling block in modelling the observed multivariate time series by a VARMA process is that $k(z)$ remains unaltered if $a(z)$ and $b(z)$ are pre-multiplied by a non-singular matrix, $u(z)$, though the 'coprime' assumption means that $u(z)$ need only be unimodular, a matrix with a constant determinant (see Hannan [21]). The problem of specifying a unique VARMA model to fit an observed time series is consequently recognized as non-trivial.

A unique model may be specified if the degrees, (p, q) , of $[a(z) \ b(z)]$ are known a priori, and a model identification procedure given in Hannan [21] applies. Such prior information would, however, be rarely available. There are also (see Lütkepohl and Poskitt [30]) several important problems with this procedure. An alternative echelon form approach avoids many of these problems and provides a 'black box' representation for $k(z)$ together with a state space representation for $\{y(t)\}$ in which the observation error process coincides with the innovation process, $\{\epsilon(t)\}$. This approach, henceforth called the standard echelon form approach, or, more succinctly, the standard approach, has also been described as 'canonical' to the problem of VARMA model specification in Hannan and Deistler [22].

From the point of view of specifying an appropriate model for an observed time series, however, this approach suffers from the following three distinct difficulties, henceforth referred to as [Difficulties 1, 2, and 3](#), respectively:

Difficulty 1. *A lack of choice.* A 'one-size-fits-all' methodology to model specification is adopted and only a single model is prescribed. An analyst may instead prefer to gain a better understanding of the uncertainty present in choosing an appropriate 'black box' model by having a choice of specifying a range of different models, each uniquely specified, rather than (see Cox [15]) 'force unique answers' to questions involving uncertainty.

Difficulty 2. *Overparameterisation.* The standard approach classifies the parameters of the specified model in two categories: constrained and unconstrained. For an observed time series, the former category of parameters do not require estimation, as they are constrained to equal zero or one; those in the latter category, by contrast, require estimation since they are treated as unknown but freely varying. However, simple examples of rational transfer functions given in Section 4 show that the number of unconstrained parameters so specified can be much larger than necessary, and vanishing parameters can be classified as unconstrained. The consequent parameter redundancy would increase parameter estimation errors and, in turn, impact further statistical inferences drawn from the estimated model, including prediction errors. Lütkepohl and Poskitt [30] have earlier suggested a graphical procedure to get around this difficulty; this procedure is, however, rather ad hoc and involves visually identifying potentially vanishing coefficients from a full model estimated initially by the standard approach.

Difficulty 3. *Structural rigidity.* The standard approach imposes a straitjacket on the specified model by requiring that its AR and MA polynomials have identical row degrees. An analyst may perceive such a restriction as rather limiting and may instead prefer to have the flexibility of specifying a 'black box' representation in which the row degrees of AR and MA polynomials differ by a constant, c , say, where c could vary and it need not equal zero. A further variation would be to enable c to assume two or more values, for example, to let $c = c_1$ for a subset of these rows and $c = c_2$ for the remaining rows, where $c_1 \neq c_2$; in general c could be a vector.

A three-fold impact of the difficulties described above should be noted, and which, with minor modifications, apply also to the approach of Akaike [4]: First, a range of ad hoc procedures for fitting VARMA-like models have attracted much attention; see, for example, Peña and Box [31], Tiao and Tsay [43]. Although such procedures have a place in multivariate time series analysis, they cannot substitute for specification of a suitable 'black box' model. Secondly, some analysts find the task of fitting a VARMA model daunting and on occasion opt not to even pursue this possibility. Thirdly, VAR models are frequently used as a substitute for fitting a VARMA model. As stated in Lütkepohl and Poskitt [30], however, there are several reasons why the VARMA class may often be preferred; to save space, these reasons are not discussed here.

This paper seeks to boost and enhance the standard approach, and, also, the approach of Akaike [4], by addressing [Difficulties 1–3](#) discussed above.

Our results are based on the pioneering work of Akaike [2], who established that a given VARMA(p, q) model may be represented in a state space form, and vice versa. A family of h -step state space representations for a VARMA(p, q)

model, valid for each $h \in \mathbb{Z}$, is introduced in Section 2. For convenience, we refer to this family collectively as multistep state space representations, and its members as multistep state space forms. This family includes the representation given by Akaike [2] as a special case with $h = 0$, and that given by Cooper and Wood [14] with $h = 1$. However, the state vector in our representation has dimension $n(h) = \max(p, q - h + 1)s$. By contrast, the state vector in Akaike's representation has dimension $n(0) = \max(p, q + 1)s$ and that in Cooper and Wood's representation has dimension $n(1) = \max(p, q)s$. As the state space representation and the associated Kalman filter find applications in a variety of different problems (see Harvey [24]), the notion of a multistep representation may be of independent interest.

Sections 3 and 4 extend to each $h \in \mathbb{Z}$ an exposition of the standard approach given in Hannan and Deistler [22] for $h = 1$. Thus, the structure of multistep state space forms is discussed in Section 3, where h -step generalizations of some key concepts in Systems Theory are given for each $h \in \mathbb{Z}$. Also, Section 4 introduces a family of h -step echelon VARMA and state space representations for $k(z)$, valid for each $h \in \mathbb{Z}$. This family includes the representations given in the standard approach as a special case with $h = 1$, and that given in Akaike [4] with $h = 0$. However, by varying h , it is possible to specify a unique representation for each h in exactly the same way as with the standard approach. The number of possible VARMA models, and the associated state space forms, that could be specified is large, countably infinite. Difficulty 1 could thus be overcome by specifying a finite number of such 'multistep' models for an observed time series.

This last-mentioned possibility also enables Difficulties 2 and 3 to be addressed, as further explained below: The row degrees of AR and MA polynomials in the h -step echelon VARMA representations differ by a constant, c , say, for each $h \in \mathbb{Z}$, where $c = 0$, if $h = 1$, but $c \neq 0$, for all $h \neq 1$. The first part of Difficulty 3 could thus be overcome by considering multistep echelon VARMA models for a range of different values of h . An explicit mathematical expression for the number of unconstrained parameters, $d_\pi(h)$, say, requiring estimation in an h -step echelon representation, with each $h \in \mathbb{Z}$, is given in Section 4. The precise numerical value of $d_\pi(h)$ depends upon the nature of generating VARMA process, $\{y(t)\}$, and the value of h . Its minimum, however, need not necessarily occur for $h = 1$. Difficulty 2 could, in principle, be thus side-stepped by first evaluating $d_\pi(h)$ for several different values of h and then choosing the value of h which minimizes $d_\pi(h)$. Section 4 also illustrates with three specific examples Difficulties 1–3 with the standard approach, and presents results demonstrating how multistep echelon forms contribute towards overcoming these difficulties.

The coefficients in echelon VARMA model with two successive values of h are related to each other in Section 5 and theoretical results which can help determine the vanishing parameters in a systematic fashion are derived.

The estimation of echelon models from an observed time series is considered in Section 6, where the canonical correlation procedure proposed originally by Akaike [4] for $h = 0$ is generalized to each fixed $h \in \mathbb{Z}$. The effectiveness of the estimation procedure with observed time series of moderate length is examined in Section 7 by a simulation study. Section 6 also includes a discussion of why a related procedure suggested earlier in Tsay [45], see, also, Piccolo and Tunnicliffe Wilson [32], differs in important respects from that introduced here.

The question of how to specify and select an appropriate echelon model to fit an observed time series is discussed in Section 8, where three different applications of the approach suggested in this paper are also outlined, and, in addition, the multistep echelon forms are related to an alternative SCM approach suggested in Tiao and Tsay [43], and the Kronecker invariants of Poskitt [33].

A brief summary and conclusions of the paper are given in Section 9.

2. Multistep state space representation for a VARMA process

The following theorem extends to each $h \in \mathbb{Z}$ the state space forms introduced earlier by Akaike [2] for $h = 0$ and Cooper and Wood [14] for $h = 1$. In this theorem (see Hannan [19] and Box and Jenkins [12]), for each $u > 0$,

$$\hat{y}_{t-h}(u) = \sum_{j=u}^{\infty} K(j)\epsilon(t-h+u-j), \quad h \in \mathbb{Z}, \quad (3)$$

denotes an optimal linear predictor of $y(t-h+u)$ based on $\{y(t-h-j), j \geq 0\}$, and $\hat{y}_{t-h}(u) = y(t-h+u)$, if $u \leq 0$.

Theorem 1. Let $\{y(t)\}$ satisfy Assumption 1. Then, for each $h \in \mathbb{Z}$, $\{y(t)\}$ admits a state space representation defined as follows:

$$y(t) = \Omega(h)x_h(t) + e_h(t), \quad (4)$$

$$x_h(t+1) = F(h)x_h(t) + V(h)z(t), \quad (5)$$

where, with $m = \max(p, q - h + 1)$, $z(t) = \epsilon(t - h + 1)$, and $A(j) = 0, j > p$, or $j < 0$,

$$x_h(t) = [\hat{y}_{t-h}(h) \quad \hat{y}_{t-h}(h+1) \quad \dots \quad \hat{y}_{t-h}(h+m+1)]^\top \quad (6)$$

$$V(h)^\top = [K(h)^\top \quad K(h+1)^\top \quad \dots \quad K(h+m+1)^\top],$$

$$\Omega(h) = \Omega = [I_s \quad 0 \quad \dots \quad 0],$$

$$e_h(t) = \sum_{j=0}^{h-1} K(j)\epsilon(t-j), \quad h \geq 1, \quad (7)$$

$$e_h(t) = 0, \quad h \leq 0, \quad (8)$$

$$F(h) = F = \begin{bmatrix} 0 & I_s & 0 & \dots & 0 \\ 0 & 0 & I_s & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_s \\ -A(m) & -A(m-1) & -A(m-2) & \dots & -A(1) \end{bmatrix}.$$

Proof. The theorem follows by noting that $\hat{y}_{t-h}(u)$ satisfies the following difference and updating equations:

$$\hat{y}_{t-h}(u) = - \sum_{j=1}^p A(j) \hat{y}_{t-h}(u-j) + \sum_{j=u}^q B(j) \epsilon(t-h+u-j), \quad 1 \leq u \leq q, \quad (9)$$

where the second sum to the right of (9) vanishes if $u > q$, and

$$\hat{y}_{t-h+1}(u-1) = \hat{y}_{t-h}(u) + K(u-1) \epsilon(t-h+1), \quad u \geq 1. \quad \square$$

Although a strict definition of what is a state vector is currently not available, this term is used in Theorem 1 in the same sense as used in Akaike [2] and Caines [13]. For each $h > 1$, however, $x_h(t)$ shares the following two properties with the corresponding standard state vector with $h = 1$, albeit in a somewhat restricted sense: (i) it is a function of the past h -step removed, $\{y(t-h-j), j \geq 0\}$; (ii) it provides a sufficient statistic for prediction h -steps ahead and beyond of $\{y(t+j), j \geq 0\}$. In addition, the observation error process, $e_h(t)$, coincides with the h -step prediction error, $y(t+h-1) - \hat{y}_{t-1}(h)$, and it is serially correlated, though uncorrelated with the state vector, $x_h(t)$. While its special structure has implications for parameter estimation, it is not germane to applications discussed below.

By contrast, if $h \leq 0$, the state vector is 'anticipative', meaning it is a function not only of the past, $\{y(t-j), j \geq 1\}$, but also of the $1-h$ 'future' values, $\{y(t), y(t+1), \dots, y(t-h)\}$. Hence, property (i) above holds in a modified sense. However, since these future values are unobserved at time $(t-1)$, property (ii) fails and $x_h(t)$ is not sufficient for prediction of $\{y(t+j), j \geq 0\}$. This last property is reflected in Eq. (8), where the observation error process, $e_h(t)$, vanishes.

Let $\bar{x}_h(t) = U(h)x_h(t)$, where $U(h)$ is a conformal non-singular matrix. Theorem 1 then still holds if $\bar{x}_h(t)$ replaces $x_h(t)$, provided $\bar{V}(h) = U(h)V(h)$, $\bar{F}(h) = U(h)F(h)U(h)^{-1}$ and $\bar{\Omega}(h) = \Omega(h)U(h)^{-1}$ also replace $V(h)$, $F(h)$ and $\Omega(h)$. The state vector, $x_h(t)$, is thus not unique. However, Reinsel [38], Yamamoto [51] and [13] discuss different applications of this result with $h = 0$ and 1.

3. Structure of multistep state space forms

A formal definition of an h -step state space representation is given below in Definition 1 for a fixed $h \in Z$, together with that of the associated multistep state space representation:

Definition 1. A process, $\{y(t), t \in Z\}$ is said to admit an h -step state space representation for a fixed $h \in Z$ if the following conditions are satisfied:

- (a) $y(t)$, for all t , may be written as in (4), where $x_h(t)$, the h -step state vector, is of finite dimension, $n(h)$, say, and satisfies (5), though not necessarily given by (6);
- (b) the matrices $F(h)$, $\Omega(h)$ and $V(h)$ do not depend on t ;
- (c) the observation error process, $e_h(t)$, satisfies (7), if $h \geq 1$, or (8), $h \leq 0$, where $\{\epsilon(t)\}$ denotes a sequence of uncorrelated random variables, each with mean zero and a positive definite covariance matrix, Ψ .

If, in addition, the h -step representation holds for a collection of values of h , then $\{y(t)\}$ is said to admit a multistep state space representation.

If $h = 1$, the standard prediction error form of a state space system results from the above definition, and the h -step representation generalizes this basic system to each $h \in Z$.

For $h \geq 1$, the state vector, $x_h(t)$, depends on $\{\epsilon(t-h-j), j \geq 0\}$ and hence on the entire past history up to $t-h$ of $\{y(t)\}$. For $h \leq 0$, however, it depends on $\{\epsilon(t+j), j = 0, 1, \dots, -h\}$, the unrealized future and concurrent innovations, and, also, on the past innovations, $\{\epsilon(t-j), j \geq 1\}$. The resulting state vector may be postulated, for example, to reflect the effects of possible future 'shocks' in financial markets.

It may be noticed that Definition 1 does not invoke Assumption 1. The well-known equivalence between a state space form and a VARMA process, however, readily extends to all $h \in Z$, and we may show that a process, $\{y(t)\}$, satisfying Definition 1 also admits a VARMA($n(h)$, $n(h) + h - 1$) representation for each $h \in Z$. The proof is omitted as it is similar to that given in Akaike [2] for $h = 0$ and Caines [13] for $h = 1$.

The standard concepts of observability, reachability, minimality and McMillan degree readily extend to an h -step state space system by recognizing that these concepts are defined independently of the nature of $z(t)$, the state noise, but which equals $\epsilon(t)$ for a one-step system and $\epsilon(t-h+1)$ for an h -step system:

Thus, for each $h \in Z$, let

$$Q(h) = \begin{bmatrix} K(h) & K(h+1) & K(h+2) & \dots & \dots \\ K(h+1) & K(h+2) & K(h+3) & \dots & \dots \\ K(h+2) & K(h+3) & K(h+4) & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \\ \vdots & \vdots & \vdots & \ddots & \end{bmatrix} \quad (10)$$

be an infinite-dimensional block Hankel matrix with the block in its (i, j) th position being $K(h+i+j-2)$, $i, j \in \{1, 2, \dots\}$.

Also, let $f(h+i-1, j)$ be an infinite-dimensional row vector denoting the row in the j th position of the i th block of rows of $Q(h)$, $j \in \{1, \dots, s\}$, $i \in \{1, 2, \dots\}$. We may write

$$Q(h)^\top = [f(h, 1)^\top, f(h, 2)^\top, \dots, f(h, s)^\top, f(h+1, 1)^\top, f(h+1, 2)^\top, \dots, f(h+1, s)^\top, \dots].$$

Further let $Q_{n(h)n(h)}(h) = [K(h+i+j-2)]$, $i, j \in \{1, \dots, n(h)\}$, be a sub-matrix in the top left $n(h) \times n(h)$ corner of $Q(h)$, where $n(h)$ denotes the dimension of the h -step state vector, $x_h(t)$. Then, as with $h = 1$, an h -step state space system is said to be minimal if $Q_{n(h)n(h)}(h)$ is of rank $n(h)$ and the rank of $Q(h)$ for a minimal system also equals $n(h)$.

For a fixed $h \in Z$, let

$$k_h^*(z) = \sum_{j=1}^{\infty} K(j+h-1)z^{-j},$$

and let

$$k_h^*(z) = a_h^*(z)^{-1} b_h^*(z) \quad (11)$$

denote a left matrix fraction description of $k_h^*(z)$, where $a_h^*(z)$ and $b_h^*(z)$ are polynomial matrices.

The standard theory, see Hannan and Deistler [22], applies to $k_h^*(z)$ and the factorization (11) is said to be irreducible if $[a_h^*(z), b_h^*(z)]$ only have unimodular matrices as their left divisors, and now $a_h^*(z)$ and $b_h^*(z)$ are also said to be left coprime. Moreover, the degree of the determinant of $a_h^*(z)$, $\tau(\det[a_h^*(z)])$, is an invariant for all irreducible left matrix fraction descriptions of $k_h^*(z)$ and it equals the dimension, $n(h)$, say, of a minimal h -step state space system. We accordingly call $n(h)$ the h -step McMillan degree. To avoid redundancy, we assume that the factorization (11) is irreducible.

Next, let

$$\tilde{k}_h(z) = z^{1-h} k_h^*(z).$$

It readily follows from (11) that $\tilde{k}_h(z)$ admits a left matrix factorization as follows:

$$\tilde{k}_h(z) = a_h^*(z)^{-1} b_h^\#(z), \quad (12)$$

where $b_h^\#(z) = z^{1-h} b_h^*(z)$ is a Laurent polynomial matrix and, with $p(h)$ denoting the degree of $a_h^*(z)$, the coefficients, $A_h^*(u)$, say, of $a_h^*(z)$ satisfy the following equations for $u \in \{h, h+1, \dots\}$:

$$\sum_{j=0}^{p(h)} A_h^*(j) K(u+j) = 0.$$

Corresponding to factorization (12), it is also possible to express $y(t)$ as a 'forward' difference equation in $y(t+1), \dots, y(t+p(h))$, $\epsilon(t), \epsilon(t+1), \dots, \epsilon(t-p(h)+h-1)$. The details are omitted to save space. Suffice to say that such a representation generalizes to each $h \in Z$ that known for $h = 1$ (see Hannan and Deistler [22]).

If Definition 1 holds and $\|F(h)\| < 1$, where $\|A\|$ denotes the Euclidian norm of a matrix, A , an h -step state space representation for each $h \in Z$ also provides a realization of the transfer function, $k_h^*(z)$, as follows:

$$k_h^*(z^{-1}) = \Omega(h)[I_s - F(h)z]^{-1} V(h).$$

Hence, we may write, for each fixed $h \in Z$ and all $j \geq h$,

$$K(j) = \Omega(h)F(h)^{j-h}V(h). \quad (13)$$

A simple test of minimality of an h -step system may also be given (see Hannan and Deistler [22]). Let

$$C(h) = [\Omega(h)^\top, F(h)^\top \Omega(h)^\top, \dots]^\top, \quad D(h) = [V(h), F(h)V(h), \dots].$$

Also let, for $i \in \{1, \dots, n(h)\}$,

$$C_{n(h)}(h)^\top = [F^\top(h)^{i-1} \Omega(h)^\top], \quad D_{n(h)}(h) = [F(h)^{i-1} V(h)]$$

be the corresponding finite-dimensional block matrices, consisting of the first $n(h)$ blocks of $C(h)$ and $D(h)$, respectively. An h -step state space system is said to be observable if $C_{n(h)}(h)$ is of rank $n(h)$ and reachable if $D_{n(h)}(h)$ is of rank $n(h)$.

An h -step state space system is then minimal if and only if it is both observable and reachable. Moreover, we have, from (13),

$$\begin{aligned} Q(h) &= C(h)D(h), \\ Q_{n(h)n(h)}(h) &= C_{n(h)}(h)D_{n(h)}(h). \end{aligned} \quad (14)$$

The representation (14) enables analytical properties of $Q(h)$ to be studied from the corresponding properties of the 'system' matrices $C(h)$ and $D(h)$. It is applied in Section 5 for establishing a rank stability property of $Q(h)$ as $h \in \mathbb{Z}$ increases.

4. Multistep echelon VARMA and state space representations

The following result enables (see Tuan [48], Hannan [21]) an h -step echelon VARMA representation, and, also, an h -step echelon state space representation, to be derived for $\{y(t)\}$, individually with each h :

$$y^{(t)} = Q(h)\epsilon_{t-h} + L(h)\epsilon^{(t-h-1)}, \quad (15)$$

where $y^{(t)} = [y(t)^\top \ y(t+1)^\top \ \dots]^\top$, $y_t = [y(t)^\top \ y(t-1)^\top \ \dots]^\top$, $L(h) = [K(h+i-j-1)]$, $i, j \in \{1, 2, \dots\}$, is a block Toeplitz matrix, $Q(h)$ is given by (10) and ϵ_t and ϵ_t are defined analogously to $y^{(t)}$ and y_t , respectively.

The underlying idea is to express a linear function of the term on the left of Eq. (15) in terms of another linear function of the second term to its right. This is done by constructing a suitable basis for the row space of $Q(h)$, for each h , and applying this basis to eliminate the first term to the right of this equation. This suitable basis is henceforth referred to as a canonical basis. As $Q(h)$ is block Hankel, the canonical basis is selected by finding (see Akaike [2,4] and Solo [41]) the first maximum set of linearly independent rows which span the row space of $Q(h)$.

For a fixed $h \in \mathbb{Z}$, let

$$Q_\pi(h) = [f(h+i, j), \ i \in \{0, \dots, n_h(j)-1\}, \ j \in \{1, \dots, s\}]$$

define the canonical basis. Denote the collection of all $n_h(j)$ by the following vector:

$$\pi(h) = [n_h(1) \ \dots \ n_h(s)]^\top, \quad (16)$$

and the rank of $Q(h)$ by their sum, $n(h) = n_h(1) + \dots + n_h(s)$.

Clearly, $\pi(h)$ is a multiindex which describes the dependence structure of the row space of $Q(h)$. Moreover, the numerical value of $n_h(j)$, for each j , gives the exact number of rows in the j th position of each block of $Q(h)$ that are included in the canonical basis. If $n_h(j)$ vanishes for some j then no $f(h+i, j)$, $i \in \{0, 1, \dots\}$ are selected in the basis for that j . For $h = 1$, the $n_h(j)$ are known as the Kronecker indices (see Hannan and Deistler [22]), and $n(h)$ as the McMillan degree. We follow this standard terminology and refer to the $n_h(j)$, for each fixed $h \in \mathbb{Z}$, as h -step Kronecker indices of $Q(h)$, or of $k(z)$, and $n(h)$ as the h -step McMillan degree.

Let $S \in \{1, \dots, s\}$,

$$n(h, j, k) = \begin{cases} \min\{n_h(j) + 1, n_h(k)\}, & j > k, \\ \min\{n_h(j), n_h(k)\}, & j \leq k. \end{cases} \quad (17)$$

Also, let

$$S_h(j) = \{k : n(h, j, k) \geq 1, k \in S\}$$

denote the set of values of k for which $n(h, j, k)$ is strictly positive, and let

$$Q_\pi(j, h) = \{f(h+i, j), \ i \in \{0, \dots, n(h, j, k)-1\}, k \in S_h(j)\}, \quad (18)$$

denote a subset of $Q_\pi(h)$.

For a fixed $h \in \mathbb{Z}$ and each $j \in S$, $f(h+n_h(j), j)$ is readily seen to be linearly dependent on the rows of $Q_\pi(j, h)$. Hence, by a standard argument, unique constants, $\tilde{\alpha}_{hjk}(u)$, $u \in \{0, \dots, n(h, j, k)-1\}$, say, may be found such that

$$f(h+n_h(j), j) + \sum_{k \in S_h(j)} \sum_{u=1}^{n(h, j, k)} \tilde{\alpha}_{hjk}(u-1) f(h+u-1, k) = 0. \quad (19)$$

Let $p(h) = \max\{n_h(j), j \in \{1, \dots, s\}\}$, $\tilde{\alpha}_{hjj}(p(h)) = 1$, and $\tilde{A}_h(u) = [\tilde{\alpha}_{hjk}(u)]$, $j, k \in \{1, \dots, s\}$; $u \in \{0, \dots, p(h)\}$ be an $s \times s$ matrix which brings together the constants defined by (19) for all $j \in S$. We then have, for each $h \in \mathbb{Z}$,

$$\begin{aligned} [\tilde{A}_h(0) \ \tilde{A}_h(1) \ \dots \ \tilde{A}_h(p(h)) \ 0] Q(h) &= 0, \\ [\tilde{A}_h(0) \ \tilde{A}_h(1) \ \dots \ \tilde{A}_h(p(h)) \ 0] y^{(t)} &= [\tilde{B}_h(0) \ \tilde{B}_h(1) \ \dots \ \tilde{B}_h(h+p(h)-1) \ 0] \epsilon^{(t-h-1)}, \end{aligned} \quad (20)$$

where

$$[\tilde{B}_h(0) \quad \tilde{B}_h(1) \quad \dots \quad \tilde{B}_h(h+p(h)-1) \quad 0] = [\tilde{A}_h(0) \quad \tilde{A}_h(1) \quad \dots \quad \tilde{A}_h(p(h)) \quad 0] L(h). \quad (21)$$

From (19) and (21), it follows that $\tilde{A}_h(p(h))$ is lower triangular with 1's down the main diagonal and $\tilde{B}_h(p(h)+h-1) \equiv \tilde{A}_h(p(h))$.

An h -step echelon VARMA representation for $\{y(t)\}$ may now be found by setting $\tilde{t} = t + p(h)$, $q(h) = p(h) + h - 1$, $A_h(u) = A_h(p(h) - u)$, $u \in \{0, \dots, p(h)\}$, $B_h(u) = B_h(q(h) - u)$, $u \in \{0, \dots, q(h)\}$ and reversing Eq. (20). Let $\alpha_{hjk}(u)$ and $\beta_{hjk}(u)$ denote the entries in the (j, k) th positions of $A_h(u)$ and $B_h(u)$, respectively.

Theorem 2 gives the formal skeletal structure of an h -step echelon VARMA representation for $\{y(t)\}$, and **Corollary 1** gives the total number of unconstrained coefficients, $d_\pi(h)$, say, in this representation. For an observed time series, the latter can guide the choice of an appropriate h , and, also, help judge the extent to which **Difficulty 2** may be overcome by an h -step representation with $h \neq 1$:

Theorem 2. Let $\{y(t)\}$ satisfy **Assumption 1**. Corresponding to the h -step Kronecker indices given by (16), $\{y(t)\}$ admits the following unique (reversed) h -step echelon VARMA representation, valid for each $h \in Z$,

$$\sum_{u=0}^{p(h)} A_h(u) y(t-u) = \sum_{u=0}^{q(h)} B_h(u) \epsilon(t-u), \quad (22)$$

in which $A_h(u) = [\alpha_{hjk}(u)]$, $B_h(u) = [\beta_{hjk}(u)]$, $j, k \in \{1, \dots, s\}$, are matrices with coefficients satisfying the constraints given below:

- (i) $\alpha_{hjk}(0) = \beta_{hjk}(0)$,
- (ii) $\alpha_{hjj}(0) = 1$,
- (iii) $\alpha_{hjk}(u) = 0$, $u > n_h(j)$,
- (iv) $\alpha_{hjk}(u) = 0$, $0 \leq u \leq n_h(j) - n(h, j, k)$, $n(h, j, k) \leq n_h(j)$, $j \neq k$,
- (v) $\beta_{hjk}(u) = 0$, $u > q(h)$.

Corollary 1. Let $\{y(t)\}$ satisfy **Assumption 1**. For each $h \in Z$, the total number of unconstrained coefficients in the h -step echelon representation, (22), is given by

$$d_\pi(h) = \sum_{j=1}^s d_\pi(j, h),$$

where $d_\pi(j, h) = d_{A\pi}(j, h) + d_{B\pi}(j, h)$, and, for $j \in \{1, \dots, s\}$,

$$d_{A\pi}(j, h) = \sum_{k=1}^s n(h, j, k), \quad d_{B\pi}(j, h) = \{n_h(j) + h - 1\}s$$

denote the number of unconstrained autoregressive, AR, and moving average, MA, parameters, respectively, in the j th equation of (22).

A state space representation for $y(t)$ may also be specified for each h from the corresponding canonical basis for $Q(h)$. **Theorem 3** gives the actual representation. The proof is omitted as it is similar to that available with $h = 0$ in Akaike [4] and with $h = 1$ in Hannan and Deistler [22] and Solo [41]:

Theorem 3. Let $\{y(t)\}$ satisfy **Assumption 1**. For each $h \in Z$, $y(t)$ admits an h -step state space representation defined as follows:

- (a) the state vector is given by $x_h(t) = Q_\pi(h) \epsilon_{t-h}$,
- (b) $V(h)$ consists of the first s columns of $Q_\pi(h)$,
- (c) $\Omega(h)$ has 1 in $(u, v_h(u))$ th place and 0's everywhere else, $u \in \{1, \dots, s\}$, where $v_h(1) = 1$, $v_h(u) = v_h(u-1) + n_h(u-1)$, $u > 1$,
- (d) $F(h) = [F_{jk}(h)]$, $j, k \in \{1, \dots, s\}$, where, with $\alpha_{h1, n(h, j, k)} = [\alpha_{hjk}(n_h(j) - 1), \dots, \alpha_{hjk}(n_h(j) - n(h, j, k) + 1)]$,

$$F_{jj}(h) = \begin{bmatrix} 0 & I_{n_h(j)} \\ -\alpha_{hjj}(n_h(j)) & -\alpha_{h1, n(h, j, k)} \end{bmatrix},$$

$$F_{jk}(h) = \begin{bmatrix} 0 & 0 \\ -\alpha_{hjk}(n_h(j)) & -\alpha_{h1, n(h, j, k)} \end{bmatrix}, \quad j \neq k.$$

The (reversed) h -step echelon form, (22), retains some of the main properties of the standard echelon form with $h = 1$, and possesses some additional properties. Let

$$a_h(z) = \sum_{j=0}^{p(h)} A_h(j)z^j, \quad b_h(z) = \sum_{j=0}^{q(h)} B_h(j)z^j,$$

denote the respective characteristic polynomials of the $A_h(j)$ and $B_h(j)$, and put

$$k_h(z) = a_h(z)^{-1}b_h(z).$$

For each $h \in \mathbb{Z}$, $k_h(z)$ encodes $k(z)$, meaning $k_h(z) \equiv k(z)$, and a unique left matrix fraction description of $k(z)$ is obtained in which the degree of the j th row of $a_h(z)$ is $n_h(j)$ and that of $b_h(z)$ is $n_h(j) + h - 1$. For $h \neq 1$, the two could differ from each other. Also, a range of different echelon VARMA models may be fitted by allowing h to vary. Hence, [Difficulty 1](#) and the first part of [Difficulty 3](#) with the standard echelon form technology could be overcome. Also, in principle, [Difficulty 2](#) could be side-stepped by choosing the value of h that minimizes $d_\pi(h)$. The question of how to choose an appropriate h is, however, more complex as parameter estimation errors are also involved, and a detailed discussion is deferred to Section 8.

The term $(h-1)s$ occurring in $d_{B\pi}(j, h)$ reflects the fact that $K(1), \dots, K(h-1)$ are treated as unknown when $h > 1$ and would need to be parameterized and estimated, but, if $h < 1$, the values of $K(0), \dots, K(-h)$ are fixed and known.

Vaninskii and Yaglom [50] (see also Pourahmadi [36, p. 289]), give conditions under which the number of non-zero canonical correlations between the past, $\{y(t), t \leq 0\}$, and the future beyond h , $\{y(t), t \geq h\}$, $h \geq 1$, of a univariate stationary process is finite. A left matrix fraction description of $k(z)$ is, however, not discussed there, whose main focus differs from ours.

Next, three specific examples of standard VARMA processes, called [Examples 1–3](#), respectively are considered to demonstrate [Difficulties 1–3](#) with the standard approach, and to show also how multistep echelon forms contribute towards overcoming these difficulties.

Example 1 (VAR(1) Process). Suppose that

$$k(z) = (I_s - Az)^{-1} = \sum_{j=0}^{\infty} A^j z^j,$$

where A is a fixed matrix such that $\text{rank}(A) = s$, $\det(I_s - A) \neq 0$, $|z| \leq 1$.

Example 2 (VMA(1) Process). Suppose that

$$k(z) = I_s + Bz,$$

where B is a fixed matrix such that $\text{rank}(B) = s$, $\det(I_s + Bz) \neq 0$, $|z| < 1$.

Example 3 (VARMA(1, 1) Process). Suppose that

$$k(z) = (I_s - Az)^{-1}(I_s + Bz) = I_s + \sum_{j=1}^{\infty} A^{j-1}(A + B)z^j,$$

where A and B are fixed matrices such that $s = \text{rank}(A) = \text{rank}(B) = \text{rank}(A + B)$, and, with $a(z) = I_s - Az$, $b(z) = I_s + Bz$, $\det(a(z)) \neq 0$, $|z| \leq 1$; $\det(b(z)) \neq 0$, $|z| < 1$, $[a(z) b(z)]$ is left coprime.

The following groups of results describe how [Theorem 2](#) may be applied to these examples:

(i) The process defined by [Assumption 1](#) may be associated with [Examples 1–3](#) by imposing the following restrictions:

Example 1: $p = 1$, $q = 0$, and $A(1) = -A$.

Example 2: $p = 0$, $q = 1$, and $B(1) = B$.

Example 3: $p = 1$, $q = 1$, $A(1) = -A$ and $B(1) = B$.

(ii) The multistep Kronecker indices, $n_h(j)$, for these three examples are as follows:

Example 1: For all $j \in \mathbb{S}$, $n_h(j) = 1 - h$, if $h \leq 0$, $n_h(j) = 1$, if $h \geq 0$.

Example 2: For all $j \in \mathbb{S}$, $n_h(j) = 2 - h$, if $h \leq 2$, $n_h(j) = 0$, if $h \geq 2$.

Example 3: For all $j \in \mathbb{S}$, $n_h(j) = 2 - h$, if $h \leq 1$, $n_h(j) = 1$, if $h \geq 1$.

(iii) If only the multistep indices, $n_h(j)$, are treated as known, and not the $K(j)$, the following h -step echelon VARMA representations may be specified for each $h \in \mathbb{Z}$ by [Theorem 2](#):

Example 1: A full VARMA($-h + 1, 0$) model for all $h \leq 0$, and a full VARMA($1, h$) model for all $h \geq 1$, giving $d_\pi(h) = (1 - h)s^2, h \leq 0, d_\pi(h) = (1 + h)s^2, h \geq 1$.

Example 2: A full VARMA($2 - h, 1$) model for all $h \leq 2$, and a full VARMA($0, h - 1$) model for all $h > 2$, giving $d_\pi(h) = (3 - h)s^2, h \leq 2, d_\pi(h) = (h - 1)s^2, h \geq 3$.

Example 3: A full VARMA($2 - h, 1$) model for all $h \leq 1$ and a full VARMA($1, h$) model for all $h > 1$, giving $d_\pi(h) = (3 - h)s^2, h \leq 1, d_\pi(h) = (h + 1)s^2, h \geq 2$.

(iv) If, on the other hand, the $K(j)$ are treated as known, then the following additional constraints may be realized, and we get, in each example, $k_h(z) \equiv k(z)$, for all $h \in \mathbb{Z}$:

Example 1: $K(j) = A^j, j \geq 0, [A_h(0), A_h(1)] \equiv [I_s, -A], \text{ all } h,$

$A_h(j) \equiv 0, 2 \leq j \leq -h + 1, h < 0; B_h(j) \equiv 0, 1 \leq j \leq h, h > 0.$

Example 2: $K(1) = B, K(j) = 0, j > 1, A_h(0) \equiv I_s, [B_h(0), B_h(1)] \equiv [I_s, B], \text{ all } h,$

$A_h(j) \equiv 0, j \geq 1, h < 2; B_h(j) \equiv 0, 2 \leq j \leq h - 1, h > 2.$

Example 3: $K(j) = A^{j-1}(A + B) j \geq 1, [A_h(0), A_h(1)] \equiv [I_s, -A], [B_h(0), B_h(1)] \equiv [I_s, B], \text{ all } h, A_h(j) \equiv 0, 2 \leq j \leq 2 - h, h < 1; B_h(j) \equiv 0, 2 \leq j \leq h + 1, h > 1.$

Difficulties 1–3 are palpably illustrated by the three examples given above.

Difficulty 1: Only a single model, VARMA($1, 1$), is specified in all three examples by the standard approach.

Difficulty 3: The row degrees of the AR and MA polynomials specified by the standard approach equal 1 in all three examples and do not differ from each other; moreover, the profile of Kronecker indices with $h = 1$ is identical in these three data generating processes and the standard approach also fails to discriminate between them.

Difficulty 2: The standard echelon form specifies that $d_\pi(h) = 2s^2$ in Examples 1 and 2. However, when the $K(j)$ are treated as known, $B_h(1)$ vanishes in Example 1 and $A_h(1)$ in Example 2, and the actual number of unconstrained coefficients in these two examples equals s^2 . Thus, in both these examples, s^2 vanishing coefficients are specified as unconstrained by the standard approach.

The multistep echelon models, by contrast, help overcome these difficulties:

A wide range of different VARMA models, as listed above in (iii), could be specified in all three examples by varying h , and Difficulty 1 could be overcome. The row degrees of AR and MA polynomials specified by Theorem 2 are unequal in all three examples when $h \neq 1$, and Difficulty 3 could also be overcome.

Difficulty 2 may also be overcome by observing that $d_\pi(h)$ takes its smallest value with $h = 0$ in Example 1 and with $h = 2$ in Example 2. By contrast, $d_\pi(h)$ is minimized in Example 3 with $h = 1$, and which now is an appropriate choice to make. This example thus shows that the multistep echelon forms enhance the standard approach rather than replace it and the latter approach, if indicated, may still be implemented within the overall framework of the former approach. A similar comment also applies to the approach of Akaike [4] which is suitable for Example 1, as $h = 0$ is an appropriate choice in this example.

5. Identifying vanishing coefficients in multistep echelon forms

We examine whether the Kronecker indices for two or more values of h could provide information about possibly vanishing coefficients in an echelon VARMA model specified with a fixed h by Theorem 2. To this end, explicit scenarios describing how the h -step Kronecker indices, $n_h(j)$, change as h varies are considered, initially for just two consecutive values of h , but subsequently also for all $h \in \mathbb{Z}$. First, however, some basic properties of these indices are examined.

A lower bound on the values of $n_h(j)$ when $h \leq 0$ is given below in Lemma 1, which follows readily by noting that $K(j) = 0, j < 0, K(0) = I_s$:

Lemma 1. Let $\{y(t)\}$ satisfy Assumption 1. For each $h \leq 0, j \in S, n_h(j) \geq -h + 1$.

The following lemma may be useful at parameter estimation stage, since its main result is often required in some standard software packages:

Lemma 2. Let $\{y(t)\}$ satisfy Assumption 1 and suppose that for a fixed $h \in \mathbb{Z}$, the $n_h(j)$ are such that $n_h(1) \leq n_h(2) \leq \dots \leq n_h(s)$. Then $A_h(0) = B_h(0) = I_s$, an identity matrix. Conversely, if for some $j < k, n_h(j) > n_h(k)$, then $\alpha_{hjk}(0)$ is not constrained to vanish for these j and k .

An important property of the multistep Kronecker indices is established in the following theorem, which shows that as h increases by one to $h + 1$, the corresponding values of $n_h(j)$ and $n(h, j, k)$, for all $j, k \in S$, either do not change or decrease by one.

Theorem 4. Let $\{y(t)\}$ satisfy Assumption 1 and suppose that $h \in \mathbb{Z}$ is a fixed integer. Then:

(a) for $j \in S$,

$$0 \leq n_h(j) - n_{h+1}(j) \leq 1,$$

(b) for $j, k \in S$,

$$0 \leq n(h, j, k) - n(h+1, j, k) \leq 1,$$

that is, $n_h(j) - n_{h+1}(j)$ and $n(h, j, k) - n(h+1, j, k)$ either equal 0 or 1, for $j, k \in S$.

Proof. Let

$$Q_\pi^*(h+1) = \{f(h+1+u, j); u \in \{0, \dots, n_h(j)-1\}, j \in S\}.$$

By the Hankel structure of $Q(h+1)$, $Q_\pi^*(h+1) \subset Q(h+1)$, and we have, for $j \in S$,

$$f(h+1+n_h(j), j) + \sum_{k \in S_h(j)} \sum_{u=1}^{n(h,j,k)} \tilde{\alpha}_{hjk}(u-1)f(h+u, k) = 0.$$

However, by our hypothesis, only the rows defined by $Q_\pi(h+1)$ are linearly independent within $Q(h+1)$, and we have, for $j \in S$,

$$f(h+1+n_{h+1}(j), j) + \sum_{k \in S_{h+1}(j)} \sum_{u=1}^{n(h+1,j,k)} \tilde{\alpha}_{h+1jk}(u-1)f(h+u, k) = 0.$$

Hence, we get $Q_\pi(h+1) \subseteq Q_\pi^*(h+1)$ and $n_{h+1}(j) \leq n_h(j)$, for $j \in S$. Next, let $\tilde{Q}_\pi(h) = \{f(h+u, j); u \in \{0, \dots, n_h(j)-2\}, j \in S\}$, where if $n_h(j) \leq 1$, for some $j \in S$, then no rows for that j are included in $\tilde{Q}_\pi(h)$. We have, $\tilde{Q}_\pi(h) \subset Q_\pi(h)$, since, if not, the values of $n_h(j)$ may be uniformly reduced by one and a contradiction would ensue. Similarly, let $\tilde{Q}_\pi(h+1) = \{f(h+u+1, j); u \in \{0, \dots, n_h(j)-2\}, j \in S\}$. Part (a) now follows by recognizing that $\tilde{Q}(h+1) \subseteq Q_\pi(h+1)$, and that this last result in turn implies that $n_{h+1}(j) \leq n_h(j) - 1$, $j \in S$. Part (b) now follows from Part (a) and (17). \square

In the light of Theorem 4, it would be useful to distinguish between the following three sets, for each fixed $j \in S$, and each fixed $h \in Z$:

$$S_E(j, h) = \{k : n(h+1, j, k) = n(h, j, k), k \in S\},$$

$$S_L(j, h) = \{k : n(h+1, j, k) = n(h, j, k) - 1, k \in S\},$$

$$\tilde{S}_E(j, h) = S_E(j, h) \cap S_h(j).$$

The number of elements of $S_{Eh}(j)$ and $S_{Lh}(j)$ is denoted by $d(S_E(j, h))$ and $d(S_L(j, h))$, respectively. We also let $Q_\pi^*(j, h+1, h) = \{f(h+i+1, j); i \in \{0, \dots, n(h, j, k)-1\}, k \in S_h(j)\}$ denote a submatrix of $Q_\pi(j, h)$, the latter being defined in (18).

Suppose that $j \in S_L(j, h)$, but $S_E(j, h) \neq \{\emptyset\}$ is non-null. Now, $f(h+n_h(j), j)$ is simultaneously linearly dependent on the rows of $Q_\pi(j, h)$ and $Q_\pi(j, h+1)$. The following lemma establishes an equivalence between these two linear dependency relations by showing that each row of $Q_\pi(j, h+1) \cap Q_\pi(j, h)^c$ is in turn linearly dependent on the rows of $Q_\pi(j, h)$, where $Q_\pi(j, h)^c$ denotes the complement of $Q_\pi(j, h)$. Also, Theorem 5 relates $\alpha_{h+1ji}(u)$ to $\alpha_{hji}(u)$, for each fixed u .

Lemma 3. Let $\{y(t)\}$ satisfy Assumption 1. Suppose that the following conditions are satisfied for a fixed $j \in S$ and a fixed $h \in Z$:

- (i) $j \in S_L(j, h)$,
- (ii) $\tilde{S}_E(j, h) \neq \{\emptyset\}$ is non-null.

Then:

- (a) $Q_\pi(j, h) \cap Q_\pi(j, h+1)^c = \{f(h, k), k \in S_h(j)\}$,
- (b) $Q_\pi(j, h+1) \cap Q_\pi(j, h)^c = \{f(h+n_h(k), k), k \in \tilde{S}_E(j, h)\}$,
- (c) for $k \in \tilde{S}_E(j, h)$, $f(h+n_h(k), k)$ is linearly dependent on the rows of $Q_\pi(k, h)$, where $Q_\pi(k, h)$ is a subset of $Q_\pi(j, h)$.

Proof. Parts (a) and (b) follow from the definition of $Q_\pi(j, h)$. Also, under (i) and (ii), $n_{h+1}(k) = n_h(k)$ and $n(h, j, k) = n_h(k)$, for $k \in \tilde{S}_E(j, h)$, and part (c) follows from this result since it implies that $f(h+n_h(k), k)$ precedes $f(h+n_h(j), j)$, for $k \in \tilde{S}_E(j, h)$, and hence $Q_\pi(k, h)$ is a subset of $Q_\pi(j, h)$. \square

Theorem 5. Suppose that the conditions stated in Lemma 3 are satisfied. Then, for all $i \in S_h(j)$ and $u \in \{n_h(j) - n(h, j, i) + 1, \dots, n_h(j)\}$,

$$\alpha_{h+1ji}(u) = \alpha_{hji}(u) + \sum_{k \in \tilde{S}_E(j, h)} \alpha_{h+1jk}(n_h(j) - n_h(k))\alpha_{hki}(n_h(k) - n_h(j) + u), \quad (23)$$

where $\alpha_{hki}(n_h(k) - n_h(j) + u) = 0$, if, for an $i \in S_h(j)$, $f(h+u, i) \notin Q_\pi(k, h)$.

Proof. The theorem follows directly from (17) and part (c) of Lemma 3. \square

Suppose instead that $j \in S_E(j, h)$, but $S_L(j, h) \neq \{\emptyset\}$ is non-null. Now, $n_{h+1}(j) = n_h(j)$ and $f(h + n_h(j), j)$ is linearly dependent on the rows of $Q_\pi(j, h)$, and $f(h + 1 + n_h(j), j)$ on those of $Q_\pi(j, h + 1)$. However, $Q_\pi(j, h + 1)$ is a proper subset of $Q_\pi^*(j, h + 1, h)$, $Q_\pi(j, h + 1) \subset Q_\pi^*(j, h + 1, h)$, and excludes some rows which belong to the latter. Lemma 4 shows that each of these excluded rows is, nevertheless, linearly dependent on the rows of $Q_\pi(j, h + 1)$.

Lemma 4. Let $\{y(t)\}$ satisfy Assumption 1. Suppose that the following conditions are satisfied for a fixed $j \in S$ and a fixed $h \in Z$:

- (i) $j \in S_E(j, h)$,
- (ii) $S_L(j, h) \neq \{\emptyset\}$ is non-null.

Then

- (a) $Q_\pi^*(j, h + 1, h) \cap Q_\pi(j, h + 1)^c = \{f(h + n(h, j, k), k), k \in S_L(j, h)\}$,
- (b) for $k \in S_L(j, h)$, $f(h + n(h, j, k), k)$ is linearly dependent on the rows of $Q_\pi(k, h + 1)$, where $Q_\pi(k, h + 1)$ is a subset of $Q_\pi(j, h + 1)$.

Proof. Part (a) follows from the definition of $Q_\pi(j, h + 1)$ and $Q_\pi^*(j, h + 1, h)$. The proof is completed by demonstrating that, under the hypothesis of the lemma, $n_{h+1}(k) = n_h(k) - 1$ and $n(h, j, k) = n_h(k)$, for $k \in S_L(j, h)$. \square

Theorem 6. Under the conditions stated in Lemma 4, for $i \in S_{h+1}(j)$ and $u \in \{n_{h+1}(j) - n(h + 1, j, i) + 1, \dots, n_{h+1}(j)\}$,

$$\alpha_{h+1ji}(u) - \alpha_{hji}(u) = - \sum_{k \in S_L(j, h)} \alpha_{hjk}(n_h(j) - n(h, j, k) + 1) \alpha_{h+1ki}(n_{h+1}(k) - u + 1), \quad (24)$$

where $\alpha_{h+1ki}(u) = 0$, if, for $i \in S_h(j)$, $f(h + n_{h+1}(j) - u + 1, i) \notin Q_\pi(k, h + 1)$.

Proof. The theorem follows directly from (19) and part (b) of Lemma 4. \square

A special case of Theorem 5 occurs when the following condition holds:

$$n(h + 1, j, k) = n(h, j, k) - 1, \quad k \in S. \quad (25)$$

Now, $Q_\pi(j, h + 1) \cap Q_\pi(j, h)^c = \{\emptyset\}$, a null set and second term to the right of (23) vanishes. Analogously, a special case of Theorem 6 occurs when the following condition holds:

$$n(h + 1, j, k) = n(h, j, k), \quad k \in S. \quad (26)$$

Now, $Q_\pi(j, h + 1) \cap Q_\pi(j, h)^c = \{\emptyset\}$, a null set, and second term to the right of Eq. (24) vanishes. Moreover, in both cases, $\beta_{h+1jk}(u)$ may also be related to $\beta_{hjk}(u)$, by invoking Eq. (21). Hence, Theorems 7 and 8 given below hold, where the former applies under condition (25) and the latter under condition (26):

Theorem 7. Suppose that conditions stated in Lemma 3 hold but $\tilde{S}_E(j, h) = \{\emptyset\}$ is null. Then:

- (a) $\alpha_{h+1jk}(u) = \alpha_{hjk}(u)$, $u \in \{n_{h+1}(j) - n(h + 1, j, k) + 1, \dots, n_{h+1}(j)\}$, $k \in S_{h+1}(j)$,
- (b) $\alpha_{h+1jk}(u) = \alpha_{hjk}(u) \equiv 0$, $u \in \{0, \dots, n_{h+1}(j) - n(h + 1, j, k)\}$, $k \in S$,
 $n_{h+1}(j) - n(h + 1, j, k) \geq 0$,
- (c) $\alpha_{hjk}(n_h(j)) \equiv 0$, $k \in S$,
- (d) $\beta_{h+1jk}(u) = \beta_{hjk}(u)$, $u \in \{0, \dots, n_h(j) + h - 1\}$, $k \in S$.

Theorem 8. Suppose that conditions stated in Lemma 4 hold but $S_L(j, h) = \{\emptyset\}$ is null. Then:

- (a) $\alpha_{h+1jk}(u) = \alpha_{hjk}(u)$, $u \in \{0, \dots, n_h(j)\}$, $k \in S$,
- (b) $\beta_{h+1jk}(u) = \beta_{hjk}(u)$, $u \in \{0, \dots, n_h(j) + h - 1\}$, $k \in S$,
- (c) $\beta_{h+1jk}(n_h(j) + h) \equiv 0$, $k \in S$.

An important situation where Theorem 7 applies occurs if for each $j \in S$, $j \in S_L(j, h)$. Analogously, if for each $j \in S$, $j \in S_E(j, h)$, then Theorem 8 applies for all $j \in S$. Corollaries 2 and 3 given below summarize the results:

Corollary 2. Suppose that $\{y(t)\}$ satisfies Assumption 1 and $n_{h+1}(j) = n_h(j) - 1$, $j \in S$ and a fixed $h \in Z$. Then, the results (a)–(d) given in Theorem 7 hold for $j \in S$.

Corollary 3. Suppose that $\{y(t)\}$ satisfies Assumption 1, and $n_{h+1}(j) = n_h(j)$, $j \in S$ and a fixed $h \in Z$. Then, the results (a)–(c) given in Theorem 8 hold for $j \in S$.

Corollaries 2 and **3** may be extended to three or more values of h by making specific assumptions concerning the behaviour of $n_h(j)$ as h changes. **Theorem 9**, considers a 'boundary' situation in which, for some $\bar{h} \in Z$, the multistep indices decrease monotonically for all $h < \bar{h}$ and remain stable for all $h \geq \bar{h}$.

Theorem 9. Let $\{y(t)\}$ satisfy **Assumption 1**. Suppose that there exists an $\bar{h} \in Z$ such that:

$$n_h(j) = n_{h-1}(j) - 1, j \in \{1, \dots, s\} \text{ and } h < \bar{h}, \quad (29)$$

$$n_{h+1}(j) = n_h(j), j \in \{1, \dots, s\} \text{ and } h \geq \bar{h}. \quad (30)$$

Then

- (a) the VARMA representation specified by **Theorem 2** with each $h \neq \bar{h}$ coincides with that specified for $h = \bar{h}$ and the results (a)–(d) given in **Theorem 7** hold for $j \in S$ and $h < \bar{h}$ and the results (a)–(c) given in **Theorem 8** hold for $j \in S$ and $h \geq \bar{h}$.
- (b) for each $h \neq \bar{h}$, the reduction in the number of unconstrained parameters as a result of adopting the procedure given in (a) above is as follows:

$$d_\pi(h) - d_\pi(\bar{h}) = |h - \bar{h}|s^2.$$

Under **Assumption 1**, the rank, $n(h)$, of $Q(h)$ is finite for each $h \in Z$ and it cannot increase as h increases. It may thus be expected to stabilize for some $h_0 \in Z$, meaning $n(h) = n(h_0)$, for $h \geq h_0$. **Theorem 10** shows that the aforementioned intuition could be correct and gives a criterion for recognizing h_0 from a record of $n(h)$, $h \in Z$. A referee for a previous version of this paper supplied the proof given below, which is based on the decomposition (14) of $Q(h)$. Thanks are, however, also due to Professors J. Partington (Leeds), V. Peller (Michigan State) and N. Young (Leeds) for their respective proofs of this theorem, which to save space are omitted.

Theorem 10. Let **Assumption 1** hold. Suppose that there is an $h_0 \in Z$, such that $n(h_0) = n(h_0 + 1)$. Then, $n(h) = n(h_0)$, $h \geq h_0$.

Proof. On noting that $Q(h_0 + 1)$ is obtained by omitting the first block of rows of $Q(h_0)$ and using the decomposition (14), we have $Q(h_0 + 1) = C(h_0)F(h_0)D(h_0)$. Now, the hypothesis that $n(h_0) = n(h_0 + 1)$ implies that both $C(h_0)$ and $C(h_0)F(h_0)$ have column rank $n(h_0)$ and hence that $F(h_0)$ is non-singular. The theorem now follows by writing $Q(h) = C(h_0)F(h_0)^{h-h_0}D(h_0)$, for all $h \geq h_0$. \square

Theorem 9 is readily seen to apply to the three examples given in Section 4, with $\bar{h} = 0$, **Example 1**, $\bar{h} = 2$, **Example 2**, $\bar{h} = 1$, **Example 3**, and yield the same zero constraints as are satisfied by the $A_h(u)$ and $B_h(u)$. Moreover, for each $h \neq \bar{h}$, the corresponding values of $d_\pi(h)$ given in these three examples are in agreement with that given in **Theorem 9**. **Theorem 10** applies also to these three examples with $h_0 = \bar{h}$; to save space the corresponding numerical values of $n(h)$ are not shown here. **Corollaries 2** and **3**, in addition, apply sequentially to these three examples; the former with $h = \bar{h} - i - 1$, and the latter with $h = \bar{h} + i$, $i \in \{0, 1, \dots\}$, and for the resulting pair of values of h and $h + 1$, yield the same constraints as may be derived from the impulse responses, $K(j)$, are treated as known. Moreover, when collected together, the resulting constraints give the same result as stated earlier by **Theorem 9**. The details are omitted to save space.

6. Estimation of multistep Kronecker indices

Akaike [4] earlier suggested a procedure for estimating the Kronecker indices with $h = 0$. This procedure is generalized below to each fixed $h \in Z$. There are four reasons for considering this generalization. First, instead of $Q(h)$, an equivalent Hankel matrix, $\Gamma(h) = [R(h + u + v - 2)]$, $u, v \in \{1, 2, \dots\}$ is considered. Secondly an examination of the dependence structure of the rows of $Q(h)$ is avoided and replaced by a canonical correlation analysis between vectors of 'past' and 'future' variables. Thirdly, multiple hypothesis tests are not required and a model selection criterion, called *DIC*, is introduced for checking whether a canonical correlation vanishes. Fourthly, the procedure is formulated in a 'non parametric' framework, that is, without assuming a parametric model for the observed time series. Despite the known criticisms (see Hannan [20] and Solo [42]), a non-parametric approach may be preferred in the present context because a main motivation for estimating the multistep Kronecker indices occurs at the VARMA model specification stage. By contrast, the alternative approach considered in Hannan and Kavalieris [23] is parametric, cumbersome to implement and involves the fitting of different VARMA models for estimating these indices.

The vector of 'past' variables, $\eta_t(h)$, 'future' variables, θ_t , and their block Hankel covariance matrix, $\Gamma(h)$, are defined as follows for each $t \in Z$ and an $h \in Z$:

$$\eta_t(h) = [y(t - h)^\top, y(t - h - 1)^\top, \dots]^\top, \quad \theta_t = [y(t)^\top, y(t + 1)^\top, \dots]^\top,$$

$$\Gamma(h) = E\{\theta_t \eta_t(h)^\top\} = [R(h + u + v - 2)], \quad u, v \in \{1, 2, \dots\}.$$

In practice, $\eta_t(h)$ and θ_t are truncated at a large integer, M , say, and the following finite-dimensional vectors,

$$\eta_t(h, M) = [y(t - h)^\top, \dots, y(t - M)^\top]^\top, \quad \theta_t(h, M) = [y(t)^\top, \dots, y(t + M - h)^\top]^\top,$$

are considered instead for each $h \in [-H, H]$ together with their block Hankel matrix covariance matrix, $\Gamma(h, M) = E\{\theta_t(h, M)\eta_t(h, M)^\top\} = [R(h+u+v-2)]$, $u, v \in \{1, \dots, M\}$, where H denotes another large integer. The choice of H and M is subjective. However, H should be large enough to ensure that the h -step McMillan degree, $n(h)$, would not change for all $h > H$. Also, M may be chosen to be sufficiently large and such that it does not exclude any rows of $\Gamma(h)$ that could be linearly independent of their antecedents. Following Akaike [4], M may be chosen by an initial application of the minimum AIC procedure (see Akaike [1]) for selecting the order, \tilde{m} , say, of an initial autoregression that may be fitted to the observed time series. Now, M should not be smaller than \tilde{m} .

The DIC criterion suggested in Akaike [4] applies only with $h = 0$. The following generalization of DIC, called $DIC\alpha$, applies with each fixed $h \in Z$, and obtained by replacing a constant of 2 occurring in the definition of DIC by a real number, α (see Bhansali and Downham [11]):

$$DIC\alpha(v) = -T \log\{1 - r_h(v+1)^2\} - \alpha\{(M-h+1)s - v\}, \quad (13)$$

where the constant, $(M-h+1)s$, in the definition of $DIC\alpha$ above represents the dimension of $\eta_t(h, M)$. and $r_h(v+1)$, $v \geq 0$, denotes the next canonical correlation between $\eta_t(h, M)$ and $\theta_t(h, M)$ to be judged to be positive, given that $r_h(1), \dots, r_h(v)$, the first v canonical correlations, are already found to be positive. Akaike [4] gives a step-by-step algorithm for estimating the Kronecker indices with $h = 0$, and which readily extends to each $h \in Z$. For brevity, only a brief outline of the procedure to be adopted is given below:

Let $\theta^*(h, u-1)$ denote some $(u-1)$ components of $\theta_t(h, M)$ such that the canonical correlations between $\theta^*(h, u-1)$ and $\eta_t(h, M)$ are positive, where $h \in [-H, H]$ is a fixed integer. Also, let θ_u denote the next 'untested' component of $\theta_t(h, M)$, and set $\theta^\#(h, u) = [\theta^*(h, u-1)^\top, \theta_u]^\top$. Initially, $u = 1$, $\theta^*(h, u-1) = \{\emptyset\}$, the null vector, and $\theta^\#(h, u) = [y_1(t)]^\top$. Now check whether or not the canonical correlation associated with θ_u is positive by a canonical correlation analysis between $\theta^\#(h, u)$ and $\eta_t(h, M)$. This will entail inspecting the sign of $DIC\alpha(u-1)$. If $DIC\alpha(u-1)$ is negative, the optimal linear predictor, $\hat{\theta}_u$, say, of θ_u based on $\eta_t(h, M)$ is judged to be linearly dependent on the corresponding predictors of the components of $\theta^*(h, u-1)$, the antecedents of θ_u . For appropriate values of i and $\hat{n}_h(i)$, write $\theta_u = y_i(t + \hat{n}_h(i))$, adopt $\hat{n}_h(i)$ as an estimate of $n_h(i)$ and discard $y_i(t+j)$, $j \in \{\hat{n}_h(i), \hat{n}_h(i)+1, \dots\}$ from $\theta_t(h, M)$. If some components of $\theta_t(h, M)$ still remain, put $\theta^\#(h, u) = [\theta^*(h, u-1)^\top, \theta_{u+1}]^\top$, where θ_{u+1} denotes the next unchecked component of $\theta_t(h, M)$, and repeat the procedure described in the last sentence with this new $\theta^\#(h, u)$. Conversely, if $DIC\alpha(u-1)$ is non-negative, accept θ_u as the u th element of $\theta^*(h, u)$, set $\theta^*(h, u) = [\theta^*(h, u-1)^\top, \theta_u]^\top$, and continue with $\theta^\#(h, u+1) = [\theta^*(h, u)^\top, \theta_{u+1}]^\top$ in place of $\theta^\#(h, u)$. For the chosen value of h , the procedure is completed once $\theta_t(h, M)$ is exhausted, that is, after no components of $\theta_t(h, M)$ remain unchecked or discarded. Now, increase the value of h by one to $h+1$ and repeat the procedure with this new value, and continue until $h = H$, when the indices for all the chosen values of h have been estimated.

Akaike et al. [6] give a software, called Program CANOCA, for implementing the original algorithm of Akaike [4]. This software was modified and extended for use with each fixed $h \in Z$.

Cooper and Wood [14] (see also Fuchs [18], Tsay [45]) have suggested a test of whether $r_h(v+1)$ vanishes. The arguments given there for a χ^2 approximation to the asymptotic distribution of the test statistic are, however, heuristic and lack a rigorous justification. There is also a problem of simultaneous inference (see Bhansali [9]) with the hypothesis-testing approach introduced there. The $DIC\alpha$ criterion by-passes the aforementioned simultaneous inference problem, and, also, possesses a relatively simple structure.

Tsay [45] (see also Toscano and Reisen [44]) has earlier suggested a related extension to that given above of Akaike's original algorithm. The notion of multistep Kronecker indices and the associated multistep echelon forms are, however, not discussed there. In addition, a problem of simultaneous inference mentioned above is also present with the procedure suggested there, which seeks to find a vanishing canonical correlation by increasing and decreasing the value of h in steps of one.

Identification of system matrices and state dimension by computationally-effective procedures has attracted attention lately (see Van Overschee and De Moor [49]). Deistler et al. [16] and Bauer et al. [7] consider an 'autoregressive' estimate of the state by a regression of $\theta_t(h, k)$ on $\eta_t(h, k)$, with $h = 1$ and k chosen by an order selection criterion. The main application, however, of the method suggested there, which readily extends to any $h \in Z$, is likely to be for prediction and interpolation and not for estimation of h -step Kronecker indices, which is the main focus of the procedure proposed above.

7. Plan and results of a simulation study

An extensive simulation study was carried out for examining how effective the estimation procedure given in Section 6 is with an observed time series. Although a total of 17 different models were considered, including the models simulated in Hannan and Kavalieris [23], the detailed results reported in Section 7.2 are confined to only three processes described below in Section 7.1.

7.1. Computational setup

The three processes considered are henceforth called [Experiments 1–3](#) respectively. Their parameters, $A(1), \dots, A(p); B(1), \dots, B(q)$ and Ψ , are given below, row by row, with symbol ϑ used to denote this operation; for example, $\vartheta[I_2] = [1, 0; 0, 1]$:

Experiment 1.

$$p = 1, q = 0, s = 2, \vartheta[A(1)] = [-0.79, 0.68; -0.29, -0.51], \vartheta[\Psi] = [0.061, 0.022; 0.022, 0.058].$$

Experiment 2.

$$\begin{aligned} p = 1, q = 1, s = 3, \vartheta[B(1)] &= [-0.55, 0.0, 0.0; 0.0, 0.0, 0.0; 0.0, 0.0, 0.0], \\ \vartheta[A(1)] &= [-0.36, 0.0, 0.0; -0.22, -0.73, -0.12; -0.16, 0.0, -0.42], \\ \vartheta[\Psi] &= [0.51, -0.11, 0.11; -0.11, 0.27, 0.14; 0.11, 0.14, 0.77]. \end{aligned}$$

Experiment 3.

$$p = 0, q = 1, s = 2, \vartheta[B(1)] = [0.8, 0.7, -0.4, 0.6], \vartheta[\Psi] = [4.0, 1.0; 1.0, 2.0].$$

[Experiment 1](#) simulates a simplified model for the Mink–Muskrat data suggested in Harvey [24, p. 251]; [Experiment 2](#) is similar to that considered by Tsay [45]; [Experiment 3](#) is Example 2.1 of Reinsel [38].

The subroutine G05HDF of the NAG library was invoked for simulating a set of T observations, with $T = 200, 500$ and 1000 , from each experiment, and the multistep Kronecker indices were estimated by the h -step generalization of Program CANOCA of Akaike et al. [6]. The $DIC_{h\alpha}$ criterion was applied with $\alpha = 2, 3, 4$ and $\ln T$ for determining the number of non-vanishing canonical correlations. We set $H = 2$ and $M = \tilde{m} + h + 1$, if $h \geq 1$, and $M = \tilde{m} + 1$, if $h \leq 0$, where \tilde{m} denotes the order of an initial autoregression selected by AIC. This choice helped to ensure that the respective dimensions of the vectors $\eta_t(h, M)$ and $\theta_t(h, M)$ equal $\tilde{m} + 2$, for all $h \geq 0$, and $\tilde{m} + 2 - h$, for all $h \leq 0$. The latter choice helped to ensure that the common dimension does not decrease as h decreases, which would have been the case if M had been held fixed instead for all h . For brevity, however, we denote the estimated indices simply by $\hat{\pi}(h)$, and suppress their dependence on α , \tilde{m} and M . The total number of simulations for each (Experiment, T , h , α) configuration was 1000.

Three principal aims of the simulation study are as follows: first, to examine how effective the canonical correlation analysis procedure is for estimating the multistep Kronecker indices; second to investigate how useful the resulting estimates are for specifying an appropriate echelon VARMA model to fit an observed time series, and for overcoming [Difficulties 1–3](#) with the standard approach; third to examine the extent to which a simultaneous interpretation of these indices is feasible for applying the theoretical results of Sections 4 and 5 to an observed time series. Although the relative efficacy of the models specified by the standard approach and those specified from the estimated multistep Kronecker indices was also investigated, to save space, we do not report the detailed simulation results here.

7.2. Simulation results

[Fig. 1](#) displays histograms of the observed frequencies of selecting the various multistep Kronecker indices for [Experiments 1–3](#); the corresponding theoretical frequencies are also shown by a bold line. To save space, however, [Fig. 1](#) only shows the results with $h = 0, 1$ for [Experiment 1](#), $h = 1$ for [Experiment 2](#) and $h = 1, 2$ for [Experiment 3](#). More detailed results, namely those with all $h \in [-2, 2]$ and individually for each of [Experiments 1–3](#), are, nevertheless, available in the supplementary material related to this article.

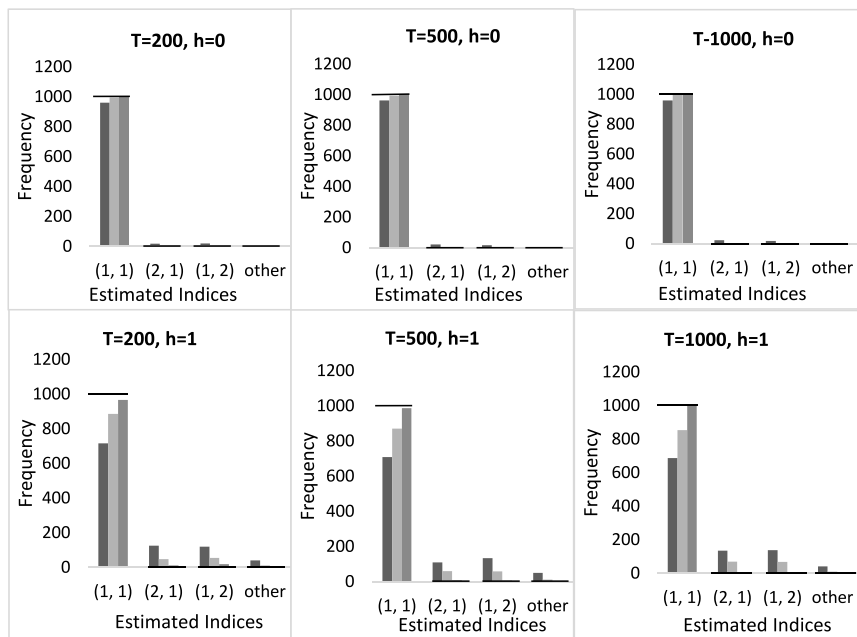
The degree of agreement between the estimated and theoretical indices may be seen to depend upon the following three main factors, and, also, on an interaction between them: (1) the nature of the generated process, especially the strength of dependence between the corresponding rows of $Q(h)$; (2) the value of T ; (3) the chosen value of α .

[Theorem 9](#) applies with $\bar{h} = 0$ to [Experiment 1](#) and $\bar{h} = 2$ to [Experiment 3](#). There is evidence to suggest that the estimated indices in these two experiments are closer to the theoretical when h is adjacent to \bar{h} than when it is away from it. Although this theorem does not apply to [Experiment 2](#), the theoretical indices for $h \leq 0$ accord with condition (29), but not condition (30).

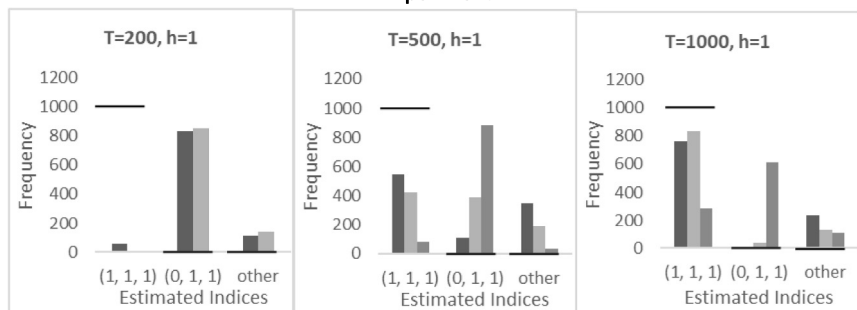
Thus, for [Experiment 1](#), the observed frequencies are in close agreement with the theoretical, especially when ($T = 1000, \alpha = \ln T$). Moreover, although the degree of agreement is less good with ($T = 200, \alpha = \ln T$), it is still quite high for all $h \leq 1$, and the discrepancy occurs mainly with $h = 2$.

For [Experiment 3](#), the estimated indices with $h = 1$ and $h = 2$ agree with the theoretical for all three values of α and $T = 1000$, or 500 ; in addition, the same holds with $T = 200$, but only when $h = 2$. For each T , however, the frequency of estimating the correct theoretical indices decreases as h takes smaller values, especially if $\alpha = \ln T$, and, even with $T = 1000$, the McMillan degree is underestimated. The agreement between these two sets of indices improves somewhat if $\alpha = 3$, but this occurs only for $T = 1000$; the frequency of underestimating the McMillan degree is noticeably high for $T = 200$ and $h = -2$, or $h = -1$.

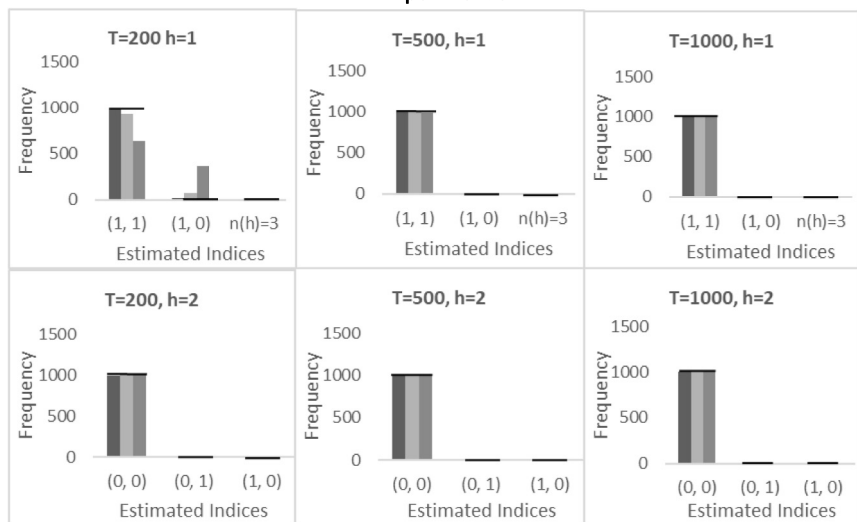
Experiment 1



Experiment 2



Experiment 3



Legend: $\alpha=3$ $\alpha=4$ $\alpha=\ln T$ Theoretical

Fig. 1. Histograms of the estimated h -step Kronecker indices in 1000 realizations of Experiments 1–3.

The estimated indices for [Experiment 2](#), by contrast, largely do not accord with the theoretical. This somewhat negative finding could, however, be attributed to the nature of the simulated VARMA(1, 1) process: the MA coefficient, $B(1)$, is not of full rank, and $\det[A(1)] = 0.11$, implying that the AR coefficient is close to being singular.

If the results for $h = 2$ are excluded, it is possible to specify an echelon VAR(1) model for [Experiment 1](#) by an inspection of the estimated multistep Kronecker indices. The standard approach, with $h = 1$, by contrast, specifies a full echelon VARMA(1, 1) model. A main reason for discounting the results for $h = 2$ is that the number of unconstrained parameters, $d_\pi(h)$, is much higher for this value of h than $h = 0$.

The simulation results for [Experiment 3](#) caution against a universal use of $\alpha = \ln T$ when estimating the multistep Kronecker indices, and support the use of a range of different values for this purpose. In addition, the estimated indices with $h = 1$ and $h = 2$ specify an echelon VMA(1) model for all three values of T . A sole use of $h = 1$, by contrast, leads to a full echelon VARMA(1, 1) model with 8 unconstrained coefficients. Although different echelon VARMA models could also be specified for $h \leq 0$, the number of unconstrained coefficients for each is much larger than that for the echelon VMA(1) model and they could be excluded on this ground.

For [Experiment 2](#), an inspection of the estimated indices for $h = -2, -1$ and 0 also enables a full echelon VAR(1) model with 9 free coefficients to be specified with all three values of T . By contrast, a variety of different patterns of estimated indices were observed with $h = 1$, and a unique model to fit was troublesome to specify. The most frequently selected pattern with all three values of T was $\hat{\pi}(h) = [0, 1, 1]^\top$ and it specified an echelon VARMA(1, 1) model with 10 free coefficients.

The simulation results, nevertheless, do not provide evidence to suggest that the estimated indices for any fixed value of h , say $h = 1$, are consistently closer to the theoretical than other values. The estimated indices for different values of h may be expected to be correlated with each other from general statistical considerations, and which, in turn, can obscure the appearance of actual indices. Care will thus need to be exercised when interpreting these indices simultaneously. On occasion, the estimated indices for a particular value of h may need to be deemphasized, especially if the corresponding estimates do not seem to conform with those for the remaining values. As with [Experiments 1](#) and [3](#), the number of unconstrained parameters, $d_\pi(h)$, requiring estimation for that h may be examined for making this judgement.

8. Implementation and applications of multistep echelon forms

The multistep echelon forms and state space representations introduced in this paper offer a two-fold advantage over the existing approaches to specifying a 'black box' VARMA model for an observed time series: First, enhanced information about the underlying data generating process is provided by the family of transfer functions, $k_h(z)$, introduced in [Section 4](#), especially about how the unknown transfer function, $k(z)$, is embedded in this family. Second, the problem of model specification can be approached systematically, and in a manner comparable to that used for univariate time series and VAR models (see [Lütkepohl \[28\]](#)). In particular, freedom to use a model selection criterion for choosing an appropriate model for the observed time series is available. Although the latter possibility does exist at present, the choice is confined only to models with $h = 0$ and $h = 1$, and equivalent to (see [Akaike \[5\]](#), [Söderström \[40\]](#)) a hypothesis test with a pre-determined level of significance. The new approach introduced in this paper, by contrast, enables an effective application of such criteria, as further discussed below.

As a first step, the canonical correlation procedure given in [Section 6](#) should be applied and estimated multistep Kronecker indices, $\hat{n}_h(j)$, say, computed, together with the associated estimates, $\hat{n}(h, j, k)$, say, of $n(h, j, k)$ and $\hat{d}_\pi(h)$, say, of $d_\pi(h)$. The estimated indices should then be plotted, grouped according to the values of h , and supplemented with the values of $\hat{d}_\pi(h)$. A table displaying the values of $\hat{n}(h, j, k)$, should also be constructed and examined. An aim here is to explore the possibility of detecting vanishing coefficients in the underlying echelon forms by applying the theoretical results of [Section 5](#), but with the estimated indices. Care will need to be exercised, however, when interpreting the estimated Kronecker indices. As discussed in [Section 7](#), a too detailed an interpretation of their plot should be avoided. Nevertheless, the hope is that this initial inspection will help in reducing the number of echelon VARMA models that should be considered to a relatively small value, H_R , say. For relatively small time series, however, H will be small, and H_R could equal $2H + 1$. Some guidance on how to choose the constants, H and M , has already been given in [Section 6](#). In practice, this will be a matter of judgement, which could be amended, as necessary, after an initial inspection of the estimated Kronecker indices by repeating the estimation procedure with new values.

Let δ_i denote the vector of unknown parameters in the i th retained model, $i \in \{1, \dots, M_R\}$, including the residual error covariance matrix, Ψ_i . Also, let $\hat{\delta}_i$ denote the corresponding vector of a maximum likelihood estimate of δ_i , and, in particular, $\hat{\Psi}_i$ that of Ψ_i . The actual model to be adopted for the observed time series should be that model which minimizes the following modified Akaike information criterion, for $i \in \{1, \dots, H_R\}$,

$$AIC_\alpha(i) = T \ln\{\det(\hat{\Psi}_i)\} + \alpha\{d(\hat{\delta}_i)\}, \quad (32)$$

where $\ln\{\det(\hat{\Psi}_i)\}$ denotes the natural logarithm of the determinant of $\hat{\Psi}_i$, $d(\hat{\delta}_i)$ denotes the number of estimated parameters in the i th retained model, that is, the number of elements in $\hat{\delta}_i$ and, as in [\(31\)](#), $\alpha > 1$ is a constant. In simulation results, not shown here, the generated model was selected with a very high frequency by the procedure described above. Furthermore, in comparison with the model specified by the standard approach, the model specified according to this

approach was also selected far more frequently. We have, in addition, also applied this procedure to three well-known data sets given in Reinsel [38, pp. 299–314]. The detailed results are, however, omitted to save space.

Instead of specifying a single ‘black box’ VARMA model for the observed time series, as discussed above, the H_R models that are retained after an inspection of the estimated multistep indices could also be employed for the following alternative applications:

(a) *Model Averaging*: This alternative was recommended by a referee. It seeks to combine, in a suitable fashion, the respective information content of the retained H_R models. It is particularly relevant for multistep prediction by forecast combination. A procedure discussed in Yamamoto [51] may be applied for constructing the multistep forecasts from each model. The final forecast, at each lead time, would then be a simple average of these H_R forecasts. A further variation would be to construct a weighted average of these forecasts by specifying a suitable weight function.

(b) *Model Uncertainty*: A Bayesian approach to the general problem of model uncertainty (see Draper [17])) may be applied to the family of ‘black box’ VARMA models introduced in this paper. A main difficulty, however, would be how to specify not only the prior distributions for the vector of unknown parameters of each model, but also the prior plausibility of the models themselves. One possibility would be to elicit both these priors after the ‘black box’ models have been specified. We note also that although Ravishankar and Ray [37] consider a Bayesian approach to a fixed degrees model, there is currently little work on a Bayesian approach to ‘black box’ models.

A ‘frequentist’ alternative (see Potscher [35]) could, in principle, be also applied to this problem. However, it would need to be recognized that the ‘black box’ VARMA model that is selected by an application of the AIC_α criterion is obtained from an inspection of the estimated multistep Kronecker indices.

A third alternative (see Poskitt and Tremayne [34]), is to build a portfolio of ‘closely competing’ black box VARMA models introduced in this paper. To save space, however, we defer a detailed discussion of these three approaches for future research.

(c) *VARMA Model Fitting*: If $\{y(t)\}$, does not satisfy Assumption 1, but it is a linear process defined in (2), then (see Hannan [21]) a finite rank approximation to the Hankel matrix, $Q(h)$, appearing in (10) is provided by an h -step echelon form with each $h \in \mathbb{Z}$. A potential statistical application of such an approximation is for estimating the impulse response coefficients of a vector time series. Sims and Zha [39], Killian [26] advocate the use of a VAR model for this purpose. However, the new approach introduced in this paper may be preferable when the first few impulse responses, $K(j)$, appear to be discrepant from the remaining values. This ability, which is not available with a standard VAR model (see Lütkepohl [29]), may be required when a transitory ‘bed-in’ period is needed before the long-term effects of a ‘shock’ applied to a system are visible. The transitory effects may be modelled by the moving average polynomial, $b_h(z)$, and the long-term effects by the autoregressive polynomial, $a_h(z)$. Moreover, the unequal row degrees for these two polynomials, which are admitted in a multistep echelon form, offer the flexibility of modelling the potential unequal impact of these two effects.

If $k(z)$ is not rational, then, instead of applying (9) iteratively for constructing multistep forecasts, a separate h -step state space form could be fitted for each prediction lead time, $h \geq 1$. Such a procedure has the advantage (see (7)) of producing the corresponding h -step prediction error, $e_h(t)$, as observation error, which now defines (see Bhansali [10]) a relevant modelling objective. Moreover, the corresponding state space representation for $h \leq 0$ may also be considered (see Section 3) for modelling the potential influence of the unrealized $h+1$ concurrent and future innovations, $\epsilon(t+j)$, $j \in \{0, 1, \dots, -h\}$. A procedure of this type should provide an alternative to a ‘Direct’ method of fitting a new autoregression for each step of prediction (see Ing and Wei [25]), and known to enjoy an asymptotic optimality property. Although the underlying idea was earlier suggested in Bhansali [8], the procedure given there is rather *ad hoc*. The procedure suggested here, by contrast, is justified within the unified framework of multistep state space representations.

It should be noted that the $n_h(j)$, for each fixed h , are not invariant to a change in the order of individual components of $y(t)$, and their numerical values can change if this order is permuted. Following Poskitt [33], it is possible to define the h -step Kronecker invariants, $\mathcal{E}(h)\pi(h) = [\pi_h(r(1), \dots, \pi_h(r(s)))]^T$, for each h , such that $\pi_h(r(1)) \leq \dots \leq \pi_h(r(s))$, where $r(1), \dots, r(s)$ denotes a permutation of the first s integers, $\{1, \dots, s\}$, and $\mathcal{E}(h)$ the corresponding permutation matrix. For the particular value of h being considered, these do not change under any rearrangement of the component time series. Nevertheless, Theorem 4 shows that the permutation matrix, $\mathcal{E}(h)$, could itself change as h changes, and the corresponding h -step and h' -step invariants for a pair of values, $h \neq h'$, could differ. A special case, however, occurs when conditions (29) and (30) hold; now $\mathcal{E}(h) = \mathcal{E}(h')$, for all $h \in \mathbb{Z}$. Also, if Theorem 10 applies for some $h_0 \in \mathbb{Z}$, and, in addition, $\pi(h) = \pi(h_0)$, for all $h \geq h_0$ then $\mathcal{E}(h) = \mathcal{E}(h_0)$, for all $h \geq h_0$.

A related SCM approach of Tiao and Tsay [43] (see also Li and Tsay [27], Tsay [46]) works with a transformed series, $z(t) = Uy(t)$, say, where U is an unknown non-singular matrix; the implied VARMA model for $\{y(t)\}$ after U has been estimated may not, however, be in a simple form. By contrast, SCM’s of order $(p_h(j), q_h(j))$ are specified directly by Theorem 2 for all $j \in S$, where $p_h(j) = n_h(j)$, $q_h(j) = n_h(j) + h - 1$, and a main criticism of the standard approach (see Tsay [47]), that it only admits SCM’s with equal AR and MA orders is overcome by this result for all $h \neq 1$.

A main advantage of the multistep echelon form representation is that it is justified in the predictor space framework of Akaike [3]. The SCM approach, by contrast, is rather *ad hoc*, and if a proper account is taken of the parameters required for specifying the transformation matrix, U , the SCM’s may not be any less parsimonious than the standard echelon form (see Lütkepohl and Poskitt [30]). The latter is a special case of the approach proposed in this paper, and, at the same time, it also enables the main aim of introducing the SCM approach to be accomplished.

9. Summary and conclusions

A principal aim of the new approach introduced in this paper is to strengthen the existing echelon form approaches to specifying a 'black box' VARMA model for an observed time series, as they suffer from the three difficulties described in Section 1. The hope is that the new approach would also enhance practical applications of VARMA processes for modelling multivariate time series and reach a level somewhat closer to that enjoyed by their univariate counterpart.

The main suggestion is that instead of specifying a single model by an existing approach, a range of different multistep echelon forms are specified and fitted, including, as required, the models specified by the existing approaches. If a single VARMA model is to be finally chosen, then a model selection criterion, supplemented, as necessary, by a graphical procedure, may be used for achieving this goal. The former is known to balance the risk of increased parameter estimation errors due to choosing a model with too many parameters against the risk of an increased bias due to a model with too few parameters, and this advantage can be potentially useful for multivariate time series.

Alternatively, as discussed in Section 8, the information contained in the different multistep echelon forms may be combined by applying a 'model averaging' procedure, or to account for the uncertainty in specifying an appropriate VARMA model. In addition, it is also possible to recognize that there may not be a 'true' VARMA process generating the observed time series, and apply a VARMA model fitting procedure for estimating the impulse response function or multistep prediction.

A first step in implementing the new approach is to estimate the multistep Kronecker indices, introduced in Section 4. Simulations show that the multistep canonical correlation procedure may be applied for estimating these indices for a range of different values of an integer, h . Care will, however, need to be exercised when inspecting a plot of the estimated indices, as the estimates with adjacent values of h could be mutually correlated, and a too detailed an interpretation of this plot should be avoided. Nevertheless, an inspection of these indices is expected to be useful for the various applications described above.

Moreover, in comparison with the SCM approach of Tiao and Tsay [43], the proposed approach has an advantage of being theoretically justified in the framework of Wiener–Kolmogorov prediction theory, and it also permits the main aim of introducing the former approach to be achieved.

The family of multistep echelon forms introduced in this paper thus provide a natural extension of the existing approaches to fitting a 'black box' VARMA model. However, by augmenting the range of possibilities currently available for specifying and selecting such models, it should serve as a useful addition to the modelling toolbox for data analysis.

Acknowledgments

Thanks are due to the referees, an Associate Editor, the current Editor-in-Chief, Dietrich von Rosen, and the previous Editor-in-Chief, Christian Genest, for their many constructive comments. Professors Emma McCoy and Hugh Morton made useful comments on various aspects of this paper, and their help is sincerely appreciated. Computational assistance from Dilbag Sokhi, David Natsios, Garima Priyadarshini and Scott Ward is also gratefully acknowledged.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2019.104539>.

References

- [1] H. Akaike, Autoregressive model fitting for control, *Ann. Inst. Statist. Math.* 23 (1971) 163–180.
- [2] H. Akaike, Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average process, *Ann. Inst. Statist. Math.* 26 (1974) 363–387.
- [3] H. Akaike, Stochastic theory of minimal realization, *IEEE Trans. Automat. Control* AC-19 (1974) 667–674.
- [4] H. Akaike, Canonical correlation analysis of time series and the use of an information criterion, in: R.K.M. Ehra, D.G.L. Ainiotis (Eds.), *System Identification: Advances and Case Studies*, Academic Press, New York, 1976, pp. 27–96.
- [5] H. Akaike, Comments on 'on model structure testing in system identification', *Internat. J. Control* 27 (1978) 323–324.
- [6] H. Akaike, E. Arahata, T. Ozaki, *TIMSAC-84, A Time Series Analysis and Control Program Package*, The Institute of Statistical Mathematics, Tokyo, 1975.
- [7] D. Bauer, M. Deistler, W. Scherrer, Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs, *Automatica* 35 (1999) 1243–1254.
- [8] R.J. Bhansali, Comment on Hannan, *Statist. Sci.* 2 (1987) 151–152.
- [9] R.J. Bhansali, Discussion of the paper by Tiao and Tsay, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 51 (1989) 201.
- [10] R.J. Bhansali, Asymptotically efficient autoregressive model selection for multistep prediction, *Ann. Inst. Statist. Math.* 48 (1996) 577–602.
- [11] R.J. Bhansali, D.Y. Downham, Some properties of the order of an autoregressive model selected by a generalization of the Akaike's FPE criterion, *Biometrika* 64 (1977) 547–551.
- [12] G.E.P. Box, G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden Day, 1970.
- [13] P.E. Caines, *Linear Stochastic Systems*, Wiley, New York, 1988.
- [14] D.M. Cooper, E.F. Wood, Identifying multivariate times series models, *J. Time Ser. Anal.* 3 (1982) 153–164.
- [15] D.R. Cox, Discussion of the papers by Mr. Campbell et al., *J. R. Stat. Soc. Ser. A Stat. Soc.* 140 (1977) 453–454.
- [16] M. Deistler, K. Peternell, W. Scherrer, Consistency and relative efficiency of subspace methods, *Automatica* 31 (1995) 1865–1975.
- [17] D. Draper, Assessment and propagation of model uncertainty, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57 (1995) 45–97.
- [18] J.-J. Fuchs, Structure and order estimation of multivariate stochastic processes, *IEEE Trans. Autom. Control* 35 (1990) 1338–1341.

- [19] E.J. Hannan, Multiple Time Series, Wiley, New York, 1970.
- [20] E.J. Hannan, The statistical theory of linear systems, in: P.R.K. Ishnaiah (Ed.), Developments in Statistics, Academic Press, New York, 1979, pp. 1–82.
- [21] E.J. Hannan, Rational transfer function approximation, *Statist. Sci.* 2 (1987) 135–151.
- [22] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Wiley, New York, 1988.
- [23] E.J. Hannan, L. Kavalieris, Multivariate linear times series models, *Adv. Appl. Probab.* 16 (1984) 492–561.
- [24] A.C. Harvey, Time Series Models, second ed., Harvester Wheatsheaf, New York, 1993.
- [25] C.-K. Ing, C.-K. Wei, Order selection for same realization prediction in autoregressive processes, *Ann. Statist.* 33 (2005) 2423–2474.
- [26] L. Killian, Small sample confidence intervals for impulse response functions, *Rev. Econ. Stat.* 80 (1998) 218–230.
- [27] H. Li, R.S. Tsay, A unified approach to identifying multivariate time series models, *J. Amer. Statist. Assoc.* 93 (1998) 770–782.
- [28] H. Lütkepohl, Comparison of criteria for estimating the order of a vector autoregressive process, *J. Time Ser. Anal.* 6 (1985) 35–52.
- [29] H. Lütkepohl, Introduction to Multiple Time Series Analysis, Springer, Berlin, 1991.
- [30] H. Lütkepohl, D.S. Poskitt, Specification of the echelon-form VARMA models, *J. Bus. Econ. Stat.* 14 (1996) 69–79.
- [31] D. Peña, G.E.P. Box, Identifying a simplifying structure in time series, *J. Amer. Statist. Assoc.* 82 (2001) 836–843.
- [32] D. Piccolo, G. Tunnicliffe Wilson, A unified approach to ARMA model identification and preliminary estimation, *J. Time Ser. Anal.* 5 (1984) 183–204.
- [33] D.S. Poskitt, Identification of echelon canonical forms for vecotr linear processes using least-squares, *Ann. Statist.* 20 (1992) 195–215.
- [34] D.S. Poskitt, A.R. Tremayne, Determining a portfolio of lienar times series models, *Biometrika* 74 (1987) 125–137.
- [35] B.M. Pötscher, Discussion of the paper by Draper, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57 (1995) 86.
- [36] M. Pourahmadi, Foundations of Times Series Analysis and Prediction Theory, Wiley, New York, 2001.
- [37] N. Ravishanker, B.K. Ray, Bayesian analysis of vector ARMA models using Gibbs sampling, *J. Forecast.* 16 (1997) 177–194.
- [38] G.C. Reinsel, Elements of Multivariate Time Series Analysis, second ed., Springer, New York, 1997.
- [39] C.A. Sims, T. Zha, Error bands for impulse responses, *Econometrica* 67 (1999) 1113–1155.
- [40] T. Söderström, On model structure testing in system identification, *Int. J. Control* 26 (1977) 1–18.
- [41] V. Solo, Topics in advanced times series analysis, in: G.D. Pino, R.R. Eboledo (Eds.), Lectures in Probability and Statistics, Springer, New York, 1986, pp. 165–328.
- [42] V. Solo, Comments on hannan, *Statist. Sci.* 2 (1987) 158–159.
- [43] G.C. Tiao, R.S. Tsay, Model specification in multivariate times series, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 51 (1989) 157–213.
- [44] E.M.M. Toscano, V.A. Reisen, The use of canonical correlation analysis to identify the order of multivariate ARMA models: simulation and application, *J. Forecast.* 19 (2000) 441–456.
- [45] R.S. Tsay, Identifying multivariate time series models, *J. Time Ser. Anal.* 10 (1989) 357–372.
- [46] R.S. Tsay, Parsimonious parametrization of vector autoregressive moving average models, *J. Bus. Econ. Stat.* 7 (1989) 327–341.
- [47] R.S. Tsay, Two canonical forms for vector ARMA processes, *Statist. Sinica* 1 (1991) 247–269.
- [48] P.-D. Tuan, On the fitting of multivariate proesses of the autoregressive-moving average type, *Biometrika* 65 (1978) 99–107.
- [49] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems, Kluwer, Boston, 1996.
- [50] K.L. Vaninskii, A.M. Yaglom, Stationary processes with a finite number of non-zero correlations, *J. Time Ser. Anal.* 11 (1990) 361–375.
- [51] T. Yamamoto, Predictions of multivariate autoregressive-moving average models, *Biometrika* 68 (1981) 485–492.