Statistica Sinica 1(1991), 247-269

TWO CANONICAL FORMS FOR VECTOR ARMA PROCESSES

Ruey S. Tsay

University of Chicago

Abstract: This paper discusses canonical forms of vector ARMA representations for a linear, time-invariant system. Both the Kronecker index and the scalar-component model (SCM) approaches are presented and discussed. The Kronecker index approach results in an Echelon form. For the SCM approach, a canonical SCM representation is introduced. The relation between Kronecker indices and orders of scalar component models is also established.

Key words and phrases: Echelon form, Hankel matrix, Kronecker index, scalar component models, vector ARMA models.

1. Introduction

Consider a k-dimensional linear time-invariant process

$$\boldsymbol{z}_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{a}_{t-i} \qquad \text{with} \quad \sum_{i=0}^{\infty} \|\boldsymbol{\Psi}_i\|^2 < \infty, \tag{1.1}$$

where $\boldsymbol{\Psi}_0 = \boldsymbol{I}_k$, the $k \times k$ identity matrix, $\|\boldsymbol{A}\|$ is a matrix norm such as the largest singular value of the matrix $\boldsymbol{A}$, and $\{\boldsymbol{a}_t\}$ is a sequence of innovations satisfying

$$E(\boldsymbol{a}_t) = \boldsymbol{0}, \quad \mathrm{Cov}(\boldsymbol{a}_t) = \boldsymbol{\Sigma} \text{ positive definite}, \quad E(\boldsymbol{a}_t \boldsymbol{a}_v) = \boldsymbol{0} \quad t \neq v. \tag{1.2}$$

In application, one often assumes that $\boldsymbol{z}_t$ of (1.1) can be described by a vector autoregressive moving average, ARMA(p, q), model

$$\boldsymbol{\Phi}(B)\boldsymbol{z}_t = \boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{1.3}$$

where B is the usual backshift operator such that $B\boldsymbol{z}_t = \boldsymbol{z}_{t-1}$, $\boldsymbol{\Phi}(z) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p)$ and $\boldsymbol{\Theta}(z) = (\boldsymbol{I} - \boldsymbol{\Theta}_1 z - \cdots - \boldsymbol{\Theta}_q z^q)$ are two matrix polynomials in z of degree p and q, respectively. Alternatively, an ARMA model for $\boldsymbol{z}_t$ can be written as

$$\Xi_0 z_t + \sum_{v=1}^{p} \Xi_v z_{t-v} = \Xi_0 a_t + \sum_{v=1}^{q} \Omega_v a_{t-v}, \tag{1.4}$$

where Ξ_0 is lower triangular with unity on the diagonal. The ARMA representation, (1.3) or (1.4), is equivalent to assuming that $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ of (1.1) is rational. The ARMA parameterizations, however, may encounter several difficulties. For instance, the model may not be unique for a given process z_t or the model may contain an excessive number of parameters. A key problem therefore is to understand the structure and parameterization of the vector ARMA model for a given linear time-invariant process z_t.

When $k = 1$, the structure of an AMRA(p, q) is well understood. One simply assumes that (a) there are no common factors between $\Phi(z)$ and $\Theta(z)$, and (b) all of the zeros of the preceding two polynomials are outside the unit circle. For the multivariate case, the situation becomes much more complicated. The direct generalization of the univariate case, which assumes that $\Phi(z)$ and $\Theta(z)$ are left coprime and all of the zeros of the determinantal polynomials $|\Phi(z)|$ and $|\Theta(z)|$ are outside the unit circle, is insufficient. There remains the problem of identifiability (or exchangeable models). Two vector ARMA models are exchangeable if the probability distributions of z_t they imply are equivalent. This identifiability problem has been discussed by many authors, e.g. Akaike (1976) and Hannan (1969). Some simple yet informative examples of exchangeable models can be found in Tiao and Tsay (1989).

The goal of this paper, therefore, is to discuss two canonical forms of vector ARMA representation of a linear time-invariant process. By a canonical form, we mean a well-specified ARMA model satisfying (a) $\Phi(z)$ and $\Theta(z)$ are left coprime, (b) the model contains no redundant parameters, and (c) the orders of the polynomials involved are as small as possible so that the total number of parameters that need estimation is minimized. The discussion is centered on two methods: one from the engineering literature that uses Kronecker indices and results in an Echelon form, and the other from the statistical literature that uses the idea of model for a scalar component and ends up with a canonical SCM representation. The relations and differences between the two approaches are also given.

2. The Kronecker Index Approach

2.1. Hankel matrices and Kronecker indices

Let Γ_ℓ be the lag-ℓ autocovariance matrix of z_t, i.e. $\Gamma_\ell = E(z_t z'_{t-\ell})$. Postmultiplying model (1.3) by z'_{t-j} and taking expectation, we have

$$\Gamma_j - \Phi_1\Gamma_{j-1} - \cdots - \Phi_p\Gamma_{j-p} = \begin{cases} \Delta_j & \text{if } j \leq q \\ \mathbf{0} & \text{if } j \geq q+1, \end{cases} \tag{2.1}$$

where Δ_j is a $k \times k$ matrix depending on Θ_i's and Σ such that $\Delta_q \neq \mathbf{0}$. This moment equation plays an important role in describing the structure of a vector ARMA model.

A direct approach to introducing the Kronecker index is to consider the Hankel matrix of the autocovariance matrices of z_t defined by

$$H_\ell = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_\ell \\ \Gamma_2 & \Gamma_3 & \cdots & \Gamma_{\ell+1} \\ \vdots & \vdots & & \vdots \\ \Gamma_\ell & \Gamma_{\ell+1} & \cdots & \Gamma_{2\ell-1} \end{bmatrix}, \tag{2.2}$$

where $\ell = 1, 2, \ldots, \infty$. In Hannan and Deistler (1988), the Hankel matrix is defined in terms of the coefficient matrices Ψ_i of (1.1). However, autocovariance matrices are more convenient for the purpose of this paper. For a given vector ARMA(p,q) model, by using the moment equation (2.1), it is clear that Rank(H_∞) is finite and Rank(H_∞) = Rank(H_ℓ) for $\ell \geq \max\{p,q\}$. On the other hand, as will be seen later, if H_∞ of z_t is of finite rank then z_t follows a vector ARMA(p,q) model, i.e. $\Psi(B)$ is rational.

Let $h(i,j)$ be the $[(i-1)k+j]$th row of H_∞. We say that $h(u,v)$ is a predecessor of $h(i,j)$ if $(u-1)k+v < (i-1)k+j$, that is, the row $h(u,v)$ appears before $h(i,j)$. The Toeplitz form of (2.2) provides a nice property for H_ℓ, namely, if $h(i,j)$ is a linear combination of $h(i_1,j_1), \ldots, h(i_s,j_s)$, then $h(i+1,j)$ is a linear combination of $h(i_1+1,j_1), \ldots, h(i_s+1,j_s)$. For convenience in reference, we summarize some properties of H_∞ into a theorem (see Theorem 2.4.1 of Hannan and Deistler (1988)).

Theorem 1. *The Hankel matrix H_∞, defined in* (2.2), *of the linear time-invariant process z_t in* (1.1) *has the following properties*:

(i) *If $h(i,j)$ is a linear combination of $h(i_1,j_1), \ldots, h(i_s,j_s)$, then $h(i+1,j)$ is a linear combination of $h(i_1+1,j_1), \ldots, h(i_s+1,j_s)$.*
(ii) *The rank of H_∞ is finite if and only if $\Psi(B)$ is rational.*
(iii) *If* Rank$(H_\infty) = m < \infty$, *then* Rank$(H_m) = m$.

Since this paper is concerned with vector ARMA models, in what follows let us assume that Rank$(H_\infty) = m < \infty$. The problem then is how to obtain a canonical form of a vector ARMA process with Hankel matrix H_∞. Based on the rank assumption, a basis for the row-space of H_∞ consists of m linearly independent rows. Due to the property Theorem 1(i), we consider a particular basis for the row-space of $\boldsymbol{H}_\infty$ that has the following property: if $h(i+1,j)$ is in the basis then $h(i,j)$ is also in the basis for all i and $1 \le j \le k$. In other words, the particular basis considered consists of the first m linearly independent rows of $\boldsymbol{H}_\infty$. Denote by $\mathcal{B}$ this particular basis which can be constructed by checking the linear dependence of each row of $\boldsymbol{H}_\infty$ from a top-down fashion beginning with the first row. More specifically, $\mathcal{B}$ can be obtained as follows.

1. Begin with the first row $h(1,1)$ of $\boldsymbol{H}_\infty$. If $h(1,1) = \mathbf{0}$, then, by Theorem 1(i), $h(i,1) = \mathbf{0}$ for all $i \ge 1$. In this case, set $d_1 = 0$; otherwise, set $\mathcal{B} = \{h(1,1)\}$ and proceed to the next row.
2. For any other row $h(i,j)$ of $\boldsymbol{H}_\infty$, if it is linearly dependent of its predecessors, that is, if $h(i,j)$ is a linear combination of those rows prior to it, discard $h(i,j)$; otherwise, augment $h(i,j)$ to $\mathcal{B}$ and check the next row in $\boldsymbol{H}_\infty$.
3. The checking procedure stops when $\mathcal{B}$ contains m rows.

Note that Theorem 1(i) is useful in checking the linear dependence in Step 2. We shall refer to $\mathcal{B}$ as the *fundamental basis* for the row-space of $\boldsymbol{H}_\infty$. Write

$$\mathcal{B} = \{h(1,1),\dots,h(d_1,1);h(1,2),\dots,h(d_2,2);\\ \dots;h(1,k),\dots,h(d_k,k)\}, \qquad (2.3)$$

where it is understood that if $d_j = 0$ then no row in the form of $h(\ell,j)$ appears in $\mathcal{B}$. The d_j's of (2.3) are pivotal quantities in understanding the structure of vector ARMA models of $\boldsymbol{z}_t$.

Definition 1. For a linear time-invariant process $\boldsymbol{z}_t$ with Hankel matrix $\boldsymbol{H}_\infty$ defined in (2.2) and the fundamental basis $\mathcal{B}$ in (2.3), the nonnegative integer d_j is called the jth *Kronecker index* of $\boldsymbol{z}_t$.

For an alternative definition of Kronecker indices, see Solo (1986). Denote by $K = \{d_j | j = 1,\dots,k\}$ the collection of Kronecker indices of $\boldsymbol{z}_t$. Obviously, $m = \sum_{i=1}^k d_i$, which is called the *McMillan degree* of the process $\boldsymbol{z}_t$ in the engineering literature. From the construction of $\mathcal{B}$, the Kronecker index d_j implies that $h(d_j,j)$ is linearly independent of its predecessors whereas $h(d_j+1,j)$ is a linear combination of its predecessors. Thus the Kronecker index d_j can be interpreted as the smallest nonnegative integer v such that $h(v+1,j)$ is a linear combination of its predecessors. Furthermore, using the above property of the Kronecker indices d_ℓ jointly, $h(d_j+1,j)$ can be rewritten as

$$h(d_j+1,j) = \sum_{i=1}^{j-1}\sum_{v=1}^{d_j+1} \beta_{v,i,j}h(v,i) + \sum_{i=j}^{k}\sum_{v=1}^{d_j} \beta_{v,i,j}h(v,i)$$

$$= \sum_{i=1}^{j-1} \sum_{v=1}^{\min(d_j+1,d_i)} \alpha_{v,i,j} h(v,i) + \sum_{i=j}^{k} \sum_{v=1}^{\min(d_j,d_i)} \alpha_{v,i,j} h(v,i), \tag{2.4}$$

where the $\alpha_{v,i,j}$'s and $\beta_{v,i,j}$'s are real numbers. As will be seen later, this equation provides a means by which one can determine the structure of the autoregressive matrix polynomial $\boldsymbol{\Xi}(z) = \sum_{i=0}^{p} \boldsymbol{\Xi}_i z^i$ of (1.4). It suffices now to note that the number of coefficients $\alpha_{v,i,j}$ in (2.4) is

$$\delta_j = \sum_{i=1}^{j-1} \min\{d_j + 1, d_i\} + d_j + \sum_{i=j+1}^{k} \min\{d_j, d_i\}. \tag{2.5}$$

2.2. A predictive interpretation

For a given time index t, define the past vector $\boldsymbol{P}_{t-1}$ and the future vector $\boldsymbol{F}_t$ of the process $\boldsymbol{z}_t$ by

$$\boldsymbol{P}_{t-1} = (\boldsymbol{z}'_{t-1}, \boldsymbol{z}'_{t-2}, \ldots)' \text{ and } \boldsymbol{F}_t = (\boldsymbol{z}'_t, \boldsymbol{z}'_{t+1}, \ldots)'. \tag{2.6}$$

The Hankel matrix $\boldsymbol{H}_\infty$ can then be written as

$$\boldsymbol{H}_\infty = E(\boldsymbol{F}_t \boldsymbol{P}'_{t-1}), \tag{2.7}$$

which implies that $\boldsymbol{H}_\infty$ is the covariance matrix between the past and the future vectors of $\boldsymbol{z}_t$.

Let $\boldsymbol{\alpha}_j$ be an infinitely dimensional, real-valued vector with $\alpha_{v,i}^{(j)}$ as its $[(v-1)k+i]$th element. For each Kronecker index d_j, from (2.4), we can define a vector $\boldsymbol{\alpha}_j$ as follows:

1. Let $\alpha_{d_j+1,j}^{(j)} = 1$.
2. For each $h(v,i)$ appearing in the right hand side of (2.4), set $\alpha_{v,i}^{(j)} = -\alpha_{v,i,j}$.
3. For all other (v,i)'s, set $\alpha_{v,i}^{(j)} = 0$.

By (2.4), we have

$$\boldsymbol{\alpha}'_j \boldsymbol{H}_\infty = \mathbf{0}. \tag{2.8}$$

To see the implication of this result, let $u_{j,t+d_j} = \boldsymbol{\alpha}'_j \boldsymbol{F}_t$. Note that, from (2.4) and (2.6), $t + d_j$ is the time index corresponding to the last non-zero element $\alpha_{d_j+1,j}^{(j)}$ of $\boldsymbol{\alpha}_j$. This explains the subscript $t+d_j$ of $u_{j,t+d_j}$. From (2.8) and (2.7), $u_{j,t+d_j}$ is uncorrelated with the past vector $\boldsymbol{P}_{t-1}$ of $\boldsymbol{z}_t$. Thus, corresponding to each Kronecker index d_j, there is a linear combination of the future vector $\boldsymbol{F}_t$ that is uncorrelated with the past $\boldsymbol{P}_{t-1}$.

On the other hand, by using the innovational representation (1.1) and the definition of $u_{j,t+d_j}$, it is easily seen that

$$u_{j,t+d_j} = \sum_{i=0}^{\infty} \boldsymbol{u}_i^{(j)} \boldsymbol{a}_{t+d_j-i}, \tag{2.9}$$

where $\boldsymbol{u}_i^{(j)}$'s are k-dimensional row vectors such that

$$\boldsymbol{u}_0^{(j)} = [\alpha_{d_j+1,1}^{(j)}, \alpha_{d_j+1,2}^{(j)}, \dots, 1, 0, \dots, 0]$$

with 1 being in the jth position, and all the other $\boldsymbol{u}_i^{(j)}$'s are linear functions of elements of $\boldsymbol{\Psi}_v$'s and non-zero elements of $\boldsymbol{\alpha}_j$. However, since $u_{j,t+d_j}$ is uncorrelated with $\boldsymbol{P}_{t-1}$, it follows that

$$u_{j,t+d_j} = \sum_{i=0}^{d_j} \boldsymbol{u}_i^{(j)} \boldsymbol{a}_{t+d_j-i}. \tag{2.10}$$

This says that the scalar process $u_{j,t+d_j}$ is at most an MA(d_j) series.

By (2.10) and the definitions of $u_{j,t+d_j}$ and $\boldsymbol{\alpha}_j$, we have for each Kronecker index d_j that

$$\begin{aligned} u_{j,t+d_j} &= z_{j,t+d_j} + \sum_{i=1}^{j-1} \sum_{v=1}^{\min\{d_j+1,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1} + \sum_{i=j}^{k} \sum_{v=1}^{\min\{d_j,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1} \\ &= a_{j,t+d_j} + \sum_{(i<j)\cap(d_j+1\le d_i)} \alpha_{d_j+1,i}^{(j)} a_{i,t+d_j} + \sum_{i=1}^{d_j} \boldsymbol{u}_i^{(j)} \boldsymbol{a}_{t+d_j-i}. \end{aligned} \tag{2.11}$$

By taking conditional expectation based on $\boldsymbol{P}_{t-1}$, (2.11) implies

$$\begin{aligned} z_{j,t+d_j|t-1} &+ \sum_{i=1}^{j-1} \sum_{v=1}^{\min\{d_j+1,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1|t-1} \\ &+ \sum_{i=j}^{k} \sum_{v=1}^{\min\{d_j,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1|t-1} = 0, \end{aligned} \tag{2.12}$$

where $z_{i,t+\ell|t-1} = E(z_{i,t+\ell}|\boldsymbol{P}_{t-1})$ is the conditional expectation of $z_{i,t+\ell}$ given $\boldsymbol{P}_{t-1}$. Thus, for each Kronecker index d_j, there exists a linear relationship among the forecasts in $\boldsymbol{F}_{t|t-1} = E(\boldsymbol{F}_t|\boldsymbol{P}_{t-1})$. Since d_j is the smallest integer for (2.12) to hold, one can interpret d_j as the number of forecasts $z_{j,t|t-1}, \dots, z_{j,t+d_j-1|t-1}$ needed to compute all the forecasts $z_{j,t+\ell|t-1}$ for any ℓ. Of course, to compute $z_{j,t+\ell|t-1}$, one also needs forecasts $z_{i,t+v|t-1}$ with $i \neq j$. However, these quantities are taken care of by the Kronecker index d_i with $i \neq j$. In view of this, the

McMillan degree m is the minimum number of quantities needed to compute all of the elements in $\boldsymbol{F}_{t|t-1}$ and the Kronecker index d_j is the minimum number of those quantities that the component $z_{j,t}$ must contribute. This is the approach used by Akaike (1976) in determining the ARMA structure. See also Tsay (1989).

2.3. An ARMA representation

By the stationarity of $\boldsymbol{z}_t$, (2.11) can be rewritten as

$$z_{j,t} + \sum_{i=1}^{j-1} \sum_{v=1}^{\min\{d_j+1,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1-d_j} + \sum_{i=j}^{k} \sum_{v=1}^{\min\{d_j,d_i\}} \alpha_{v,i}^{(j)} z_{i,t+v-1-d_j}$$
$$= a_{j,t} + \sum_{(i<j)\cap(d_j+1\leq d_i)} \alpha_{d_j+1,i}^{(j)} a_{i,t} + \sum_{i=1}^{d_j} \boldsymbol{u}_i^{(j)} \boldsymbol{a}_{t-i}, \tag{2.13}$$

where the number of coefficients $\alpha_{v,i}^{(j)}$ in the left hand side is δ_j given by (2.5) and the number of elements of $\boldsymbol{u}_i^{(j)}$'s in the right hand side is $d_j \times k$.

Next, for the Kronecker indices $K = \{d_j\}$, by considering the equation (2.13) jointly for $j = 1, \dots, k$, we have an ARMA model for the process $\boldsymbol{z}_t$

$$\Xi_0 \boldsymbol{z}_t + \sum_{v=1}^{p} \Xi_v \boldsymbol{z}_{t-v} = \Xi_0 \boldsymbol{a}_t + \sum_{v=1}^{p} \boldsymbol{\Omega}_v \boldsymbol{a}_{t-v}, \tag{2.14}$$

where $p = \max_j\{d_j\}$, Ξ_0 is a lower triangular matrix, the (j,i)th element of which, where $i < j$, is unknown only if $d_j + 1 \leq d_i$, and the coefficient matrices Ξ_v and $\boldsymbol{\Omega}_v$ are specified by (2.13). More specifically, we have the following:

1. For $\boldsymbol{\Omega}_v$ with $v > 0$: (a) the jth row is zero if $d_j < v \leq p$; (b) all the other rows are unknown.
2. For Ξ_v with $v > 0$: (a) the jth row is zero if $d_j < v \leq p$; (b) the (j,j)th element is unknown if $v \leq d_j$; and (c) the (j,i)th element with $j \neq i$ is unknown only if $d_i + v > d_j$.

The preceding equation gives rise to an ARMA representation for $\boldsymbol{z}_t$, the jth row of which contains δ_j unknown parameters in the AR polynomials and $k \times d_j$ unknown parameters in the MA polynomials, where δ_j is defined in (2.5). In sum, for a linear, time-invariant process $\boldsymbol{z}_t$ of (1.1) with Kronecker indices $K = \{d_j\}$ such that $m = \sum_{j=1}^{k} d_j < \infty$, one can specify an ARMA representation to describe the process. Such a representation is given by (2.14) which contains

$$N = m(1+k) + \sum_{j=1}^{k} \left[\sum_{i<j} \min\{d_j+1, d_i\} + \sum_{i>j} \min\{d_j, d_i\} \right] \tag{2.15}$$

unknown parameters in the AR and MA matrix polynomials.

2.4. An illustrative example

To better understand the preceding results, let us consider a simple example. Suppose that $\boldsymbol{z}_t$ is 3-dimensional with Kronecker indices $K = \{d_j = 3, 1, 2$ for $j = 1, 2, 3$, respectively$\}$. Here the fundamental basis of the corresponding Hankel matrix $\boldsymbol{H}_\infty$ defined in (2.3) is

$$\mathcal{B} = \{h(1,1), h(2,1), h(3,1); h(1,2); h(1,3), h(2,3)\},$$

the three equations in (2.13) are

$$z_{1,t} + \sum_{v=1}^{3} \alpha_{v,1}^{(1)} z_{1,t+v-4} + \alpha_{1,2}^{(1)} z_{2,t-3} + \sum_{v=1}^{2} \alpha_{v,3}^{(1)} z_{3,t+v-4} = a_{1,t} + \sum_{i=1}^{3} \boldsymbol{u}_i^{(1)\prime} \boldsymbol{a}_{t-i}$$

$$z_{2,t} + \sum_{v=1}^{2} \alpha_{v,1}^{(2)} z_{1,t+v-2} + \alpha_{1,2}^{(2)} z_{2,t-1} + \alpha_{1,3}^{(2)} z_{3,t-1} = a_{2,t} + \boldsymbol{u}_1^{(2)\prime} \boldsymbol{a}_{t-1}$$

$$z_{3,t} + \sum_{v=1}^{3} \alpha_{v,1}^{(3)} z_{1,t+v-3} + \alpha_{1,2}^{(3)} z_{2,t-2} + \sum_{v=1}^{2} \alpha_{v,3}^{(3)} z_{3,t+v-3} = a_{3,t} + \sum_{i=1}^{2} \boldsymbol{u}_i^{(3)\prime} \boldsymbol{a}_{t-i}$$

and the corresponding ARMA representation of (2.14) is

$$\begin{bmatrix} 1 & 0 & 0 \\ X & 1 & 0 \\ X & 0 & 1 \end{bmatrix} \boldsymbol{z}_t + \begin{bmatrix} X & 0 & 0 \\ X & X & X \\ X & 0 & X \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} X & 0 & X \\ 0 & 0 & 0 \\ X & X & X \end{bmatrix} \boldsymbol{z}_{t-2} + \begin{bmatrix} X & X & X \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{z}_{t-3}$$

$$= \boldsymbol{\Xi}_0 \boldsymbol{a}_t + \sum_{v=1}^{3} \boldsymbol{\Omega}_v \boldsymbol{a}_{t-v},$$

where "X" denotes an unknown parameter, $\boldsymbol{\Omega}_1$ is a full matrix, the second row of $\boldsymbol{\Omega}_2$ is zero, the second and the third rows of $\boldsymbol{\Omega}_3$ are zero, and all the other rows of $\boldsymbol{\Omega}_2$ and $\boldsymbol{\Omega}_3$ are unknown resulting in a total number of unknown parameters $N = 6 \times (1 + 3) + 10 = 34$.

2.5. The Echelon form

In what follows, we discuss some further information concerning the ARMA representation (2.14) of $\boldsymbol{z}_t$. This provides a canonical form for the process $\boldsymbol{z}_t$.

Degree of the individual polynomial. Let $A_{jv}(B)$ be the (j, v)th element of the matrix polynomial $\boldsymbol{A}(B)$. Then, the degree of each individual polynomial of $\Xi(B) = \Xi_0 + \sum_{i=1}^{p} \Xi_i B^i$ of (2.14) is that $\deg[\Xi_{jv}(B)] = d_j$ for all $v = 1, \ldots, k$. In other words, the Kronecker index d_j is the degree of all the polynomials in the jth row of $\boldsymbol{\Xi}(B)$. The same result holds for the individual polynomials in $\boldsymbol{\Omega}(B)$. We remark that d_j is the maximum order of $\Xi_{jv}(B)$ and $\Omega_{jv}(B)$. The actual order might be smaller after further estimation or analysis.

Number of unknown coefficients of the individual polynomial. Let n_{jv} be the number of unknown coefficients of $\Xi_{jv}(B)$ in $\boldsymbol{\Xi}(B)$ of (2.14). Then, from the structure of $\boldsymbol{\Xi}(B)$, we have

$$n_{jv} = \begin{cases} \min\{d_j, d_v\} & \text{if } j \leq v \\ \min\{d_j + 1, d_v\} & \text{if } j > v. \end{cases} \tag{2.16}$$

Similarly, let m_{jv} be the number of unknown coefficients of the (j, v)th polynomial in $\boldsymbol{\Omega}(B) = \sum_{v=0}^{p} \boldsymbol{\Omega}_v B^v$ of (2.14), where $\boldsymbol{\Omega}_0 = \boldsymbol{\Xi}_0$. Then, we have

$$m_{jv} = \begin{cases} d_j & \text{if } j \leq v \\ \min\{d_j + 1, d_v\} & \text{if } j > v. \end{cases} \tag{2.17}$$

Form of the individual polynomial. Denote by $A_{jv}^{(i)}$ the (j, v)th element of the matrix $\boldsymbol{A}_i$. From the degree and the number of unknown coefficients of each individual polynomial, one can easily specify the form of each individual polynomial in $\boldsymbol{\Xi}(B)$ of (2.14). More specifically, we have

$$\Xi_{jj}(B) = 1 + \sum_{i=1}^{d_j} \Xi_{jj}^{(i)} B^i \quad \text{for} \quad j = 1, \ldots, k \tag{2.18}$$

$$\Xi_{jv}(B) = \sum_{i=d_j - n_{jv}+1}^{d_j} \Xi_{jv}^{(i)} B^i \quad \text{if} \quad j \neq v, \tag{2.19}$$

where n_{jv} is defined in (2.16). For the polynomials in $\boldsymbol{\Omega}(B)$, the result is as follows:

$$\Omega_{jj}(B) = 1 + \sum_{i=1}^{d_j} \Omega_{jj}^{(i)} B^i \quad \text{for} \quad j = 1, \ldots, k \tag{2.20}$$

$$\Omega_{jv}(B) = \sum_{i=d_j - m_{jv}+1}^{d_j} \Omega_{jv}^{(i)} B^i \quad \text{if} \quad j \neq v, \tag{2.21}$$

where m_{jv} is defined in (2.17).

The preceding results show that, for a linear time-invariant process $\boldsymbol{z}_t$, the Kronecker indices $K = \{d_j | j = 1, \ldots, k\}$ specify an ARMA representation (2.14) for $\boldsymbol{z}_t$. This ARMA specification is complete in the sense that (a) all the unknown parameters in the AR and MA matrix polynomials are identified and (b) each individual polynomial is specifically given. In the literature, this ARMA representation is called a (reversed) Echelon form (see Deistler (1985) and Hannan and Deistler (1988)) and has the following nice properties:

Theorem 2. *Suppose that $\boldsymbol{z}_t$ is a k-dimensional linear, time-invariant process of* (1.1) *with Kronecker indices $K = \{d_j | j = 1, \ldots, k\}$ such that $m = \sum_{j=1}^{k} d_j < \infty$. Then, $\boldsymbol{z}_t$ follows the vector* ARMA *model* (2.14) *with $\boldsymbol{\Xi}(B)$ and $\boldsymbol{\Omega}(B)$ specified by* (2.16)-(2.21). *Furthermore, $\boldsymbol{\Xi}(B)$ and $\boldsymbol{\Omega}(B)$ are left coprime, and* $\deg[|\boldsymbol{\Xi}(B)|] + \deg[|\boldsymbol{\Omega}(B)|] \leq 2m$.

2.6. The example continued

For the 3-dimensional example of Subsection 2.4, the number of unknown coefficients in the individual polynomials is as follows:

$$[n_{jv}] = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 1 & 1 \\ 3 & 1 & 2 \end{bmatrix} \quad \text{and} \quad [m_{jv}] = \begin{bmatrix} 3 & 3 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 2 \end{bmatrix}.$$

Since $\boldsymbol{\Xi}_0 = \boldsymbol{\Omega}_0$, the total number of unknown coefficients N of (2.15) is different from the sum of all n_{jv} and m_{jv}, as the latter counts the unknown coefficients in $\boldsymbol{\Xi}_0$ twice.

3. The Scalar-Component Approach

3.1. Scalar component models

In Tiao and Tsay (1989), an alternative approach to specifying the structure of a vector ARMA model was introduced. This approach attempts to generalize the model-structure of each component z_{jt} in (1.3) to obtain simplifying structures of the system so that a simple vector ARMA model can be identified. It makes use of the idea that certain linear combinations of $\boldsymbol{z}_t$ might shed some light on the *skeleton* of a given vetor ARMA model.

Let $\boldsymbol{\Delta}_j^{(i)}$ be the jth row of the matrix $\boldsymbol{\Delta}_i$. The vector ARMA model (1.3) says that, for the first component z_{1t} of $\boldsymbol{z}_t$, there exist p k-dimensional row vectors $\boldsymbol{\Phi}_1^{(1)}, \ldots, \boldsymbol{\Phi}_1^{(p)}$ such that the process

$$u_{1t} = z_{1t} - \sum_{i=1}^{p} \boldsymbol{\Phi}_1^{(i)} \boldsymbol{z}_{t-i}$$

is uncorrelated with $\boldsymbol{a}_{t-j}$ for $j > q$, because u_{1t} can also be written as

$$u_{1t} = a_{1t} - \sum_{i=1}^{q} \boldsymbol{\Theta}_1^{(i)} \boldsymbol{a}_{t-i}.$$

This description, however, may fail to pinpoint the actual structure of z_{1t}, for the values of p and q may be too large. Motivated by such a consideration, Tiao and Tsay (1989) define a scalar component model (SCM) of order (r, s) of $\boldsymbol{z}_t$ as follows:

Definition 2. Suppose that $\boldsymbol{z}_t$ is a linear, time-invariant process of (1.1) with $\boldsymbol{\Psi}(B)$ rational. A non-zero linear combination of $\boldsymbol{z}_t$, denoted by $y_t = \boldsymbol{v}_0'\boldsymbol{z}_t$, is a *scalar component of order* (r, s) if there exist r k-dimensional vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ such that (a) $\boldsymbol{v}_r \neq \boldsymbol{0}$ if $r > 0$; (b) the scalar process

$$u_t = y_t + \sum_{i=1}^{r} \boldsymbol{v}_i'\boldsymbol{z}_{t-i} \quad \text{satisfies} \quad E(\boldsymbol{a}_{t-j}u_t)\begin{cases} = \boldsymbol{0} & \text{if } j > s \\ \neq \boldsymbol{0} & \text{if } j = s. \end{cases}$$

In other words, $y_t = \boldsymbol{v}_0'\boldsymbol{z}_t$ is an SCM of order (r, s) of $\boldsymbol{z}_t$ if the (transformed) scalar process u_t is uncorrelated with the past vector $\boldsymbol{P}_{t-j}$ for each $j > s$, but correlated with $\boldsymbol{P}_{t-s}$. Obviously, the requirements of $E(\boldsymbol{a}_{t-s}u_t) \neq \boldsymbol{0}$ and $\boldsymbol{v}_r \neq \boldsymbol{0}$ if $r > 0$ are used to reduce the order (r, s). We shall refer to y_t as an SCM(r, s) of $\boldsymbol{z}_t$.

By substituting (1.1) for $\boldsymbol{z}_{t-\ell}$ and collecting the coefficient vectors of $\boldsymbol{a}_{t-i}$, we can alternatively write

$$u_t = \boldsymbol{v}_0'\boldsymbol{a}_t + \sum_{i=1}^{s} \boldsymbol{h}_i'\boldsymbol{a}_{t-i}, \tag{3.1}$$

where $\boldsymbol{h}_s \neq \boldsymbol{0}$. Thus, an SCM of order (r, s) implies that there exists a non-zero linear combination of $\boldsymbol{z}_t, \ldots, \boldsymbol{z}_{t-r}$ which is also a linear function of $\boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-s}$. With this interpretation and (2.11), it is seen that a Kronecker index d_j of $\boldsymbol{z}_t$ implies the existence of an SCM(d_j, d_j) of $\boldsymbol{z}_t$.

Note that y_t being an SCM of order (r, s) does not necessarily imply that y_t follows a univariate ARMA(r, s) model. The SCM model is a concept within the vector framework and it uses all the components of $\boldsymbol{z}_t$ in describing a model. One should interpret an SCM from the vector framework. On the other hand, a univariate ARMA model of y_t only depends on the history of y_t, i.e. y_{t-j} with $j > 0$.

From the definition, the order (r, s) of an SCM y_t is not unique. For example, multiplying u_{t-m} with $m > 0$ by a non-zero constant d, then adding it to u_t, we obtain, from (3.1), a new scalar process

$$u_t^* = u_t + du_{t-m} = \boldsymbol{v}_0'\boldsymbol{a}_t + \sum_{\ell=1}^{s} \boldsymbol{h}_\ell'\boldsymbol{a}_{t-\ell} + d\left(\boldsymbol{v}_0'\boldsymbol{a}_{t-m} + \sum_{\ell=1}^{s} \boldsymbol{h}_\ell'\boldsymbol{a}_{t-m-\ell}\right),$$

which is uncorrelated with a_{t-j} for $j > s+m$. This type of redundancies should be eliminated, so we define an SCM of minimal order as follows:

Definition 3. Suppose that y_t is an SCM(r,s) of z_t. The order (r,s) is *minimal* if the sum $r+s$ is as small as possible.

Even with the minimal order requirement, the order of a given SCM is still not unique. For example, suppose that z_t follows a bivariate AR(1) or MA(1) model

$$z_t - \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} z_{t-1} = a_t \Longleftrightarrow z_t = a_t - \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix} a_{t-1}.$$

Then, it is easily seen that the second component z_{2t} is an SCM of order (1,0) or (0,1). Here both of the orders are minimal. In general, however, the sum $r+s$ is fixed when the order (r,s) is minimal. This sum has a special meaning which we shall discuss shortly. For further properties of SCM's, see Tiao and Tsay (1989).

3.2. Model specification via scalar component models

Suppose that $y_{it} = v_0^{(i)'} z_t$ is an SCM(p_i, q_i) of z_t, where $i = 1, \ldots, k$. We say that these k SCM's are linearly independent if the $k \times k$ matrix $T' = [v_0^{(1)}, \ldots, v_0^{(k)}]$ is non-singular. Then, a vector ARMA model for z_t is specified once k linearly independent SCM's are found. From the Definition 2, for each y_{it} there exist p_i k-dimensional vectors $v_1^{(i)}, \ldots, v_{p_i}^{(i)}$ such that the scalar process $u_{it} = \sum_{\ell=0}^{p_i} v_\ell^{(i)'} z_{t-\ell}$ is uncorrelated with a_{t-j} for $j > q_i$. Let $u_t = (u_{1t}, \ldots, u_{kt})'$, $r = \max\{p_i\}$, and $s = \max\{q_i\}$. We have

$$u_t = T z_t + \sum_{\ell=1}^{r} G_\ell z_{t-\ell}, \tag{3.2}$$

where $G'_\ell = [v_\ell^{(1)}, \ldots, v_\ell^{(k)}]$ with $v_\ell^{(i)} = 0$ for $p_i < \ell \le r$. Furthermore, from (3.1), u_t can also be written as

$$u_t = T a_t + \sum_{\ell=1}^{s} H_\ell a_{t-\ell}, \tag{3.3}$$

where H_ℓ is a $k \times k$ matrix whose ith row is zero if $q_i < \ell \le s$. Combining (3.2) and (3.3), it is seen that a vector ARMA(r,s) model with detailed row structure of the coefficient matrices is specified for the transformed process $y_t = T z_t$. More specifically, we have

$$T z_t + \sum_{\ell=1}^{r} G_\ell z_{t-\ell} = T a_t + \sum_{\ell=1}^{s} H_\ell a_{t-\ell} \tag{3.4}$$

such that

1. the ith row of G_ℓ is zero if $p_i < \ell \le r$,
2. the ith row of H_ℓ is zero if $q_i < \ell \le s$, and
3. some further reduction in parameterization is possible under certain circumstances.

The last result is due to certain redundant parameters between AR and MA components in (3.4), which we shall discuss in the next subsection. From (3.4), a vector ARMA(r, s) model for z_t is obtained.

Notice that by inserting $T^{-1}T$ in the front of $z_{t-\ell}$'s and $a_{t-\ell}$'s in (3.4), one obtains a vector ARMA(r, s) model for the transformed process y_t

$$(I - \Delta_1 B - \cdots - \Delta_r B^r)y_t = (I - \Lambda_1 B - \cdots - \Lambda_s B^s)b_t, \tag{3.5}$$

where $b_t = Ta_t$ and the coefficient matrices Δ_i and Λ_j have the same *row-structure* as those of G_i and H_j, respectively, for $i = 1, \ldots, r$ and $j = 1, \ldots, s$. From this model, it is clear that the order (p_i, q_i) of an SCM signifies that one needs $p_i + q_i$ unknown rows to describe the structure of y_{it} in the vector ARMA model (3.5) of y_t. Here by an "unknown" row we mean that its parameters require estimation. This terminology is used in contrast with the other rows that are known to be zero. The requirement that the order (p_i, q_i) be minimal in Definition 3 becomes evident.

Next, to achieve the greatest reduction in parameterization (i.e., to have as small as possible unknown parameters), the requirement of each SCM being of minimal order is still insufficient. One also needs the condition that the k linearly independent SCM's used jointly have the lowest possible orders among all plausible sets of k linearly independent SCM's of the system. To describe this, let $o_i = p_i + q_i$ be the number of unknown rows corresponding to the SCM y_{it} and define

$$NR(y_t) = \sum_{i=1}^{k} o_i = \sum_{i=1}^{k} (p_i + q_i) \tag{3.6}$$

as the total number of unknown rows in the ARMA model (3.5).

Definition 4. For the vector linear process z_t of (1.1), suppose that $\{y_{it}\}_{i=1}^k$ and $\{y_{it}^*\}_{i=1}^k$ are two sets of k linearly independent SCM's of orders $\{(p_i, q_i)\}_{i=1}^k$ and $\{(p_i^*, q_i^*)\}_{i=1}^k$, respectively. Let $y_t = (y_{1t}, \ldots, y_{kt})'$ and $y_t^* = (y_{1t}^*, \ldots, y_{kt}^*)'$, and define the total numbers of unknown rows in ARMA representations of $NR(y_t)$ and $NR(y_t^*)$ by (3.6). Then, we say that (a) y_t is *at least as parsimonious in row as* y_t^* if $NR(y_t) \le NR(y_t^*)$, and (b) y_t is *more parsimonious in row than* y_t^* if $NR(y_t) < NR(y_t^*)$. Further, y_t is referred to as a *parsimonious set* of SCM's if no other set of k linearly independent SCM's is more parsimonious in row than y_t.

Notice that the term 'parsimony' is associated with the number of unknown rows. This, of course, is different from parsimony in parameterization, even though the two concepts are highly related. Obviously, only parsimonious sets of SCM's are of main interest in practical model building.

3.3 Redundant parameters

In this subsection, we consider the possible redundant parameters in the vector ARMA representation of (3.5) and discuss a method that can easily identify such parameters when they exist. It is worth mentioning that redundant parameters can occur even without overspecifying the overall order (r,s) of (3.5).

Suppose that the orders (p_i,q_i) of the first two SCM's y_{1t} and y_{2t} satisfy $p_2>p_1$ and $q_2>q_1$. In this case, we can write the model structure for y_{1t} and y_{2t} as

$$y_{it}-\{\boldsymbol{\Delta}_i^{(1)}B+\cdots+\boldsymbol{\Delta}_i^{(p_i)}B^{p_i}\}\boldsymbol{y}_t=b_{it}-\{\boldsymbol{\Lambda}_i^{(1)}B+\cdots+\boldsymbol{\Lambda}_i^{(q_i)}B^{q_i}\}\boldsymbol{b}_t, \quad (3.7)$$

where $i=1,2$ and $\boldsymbol{\Delta}_i^{(v)}$ and $\boldsymbol{\Lambda}_i^{(v)}$ are the ith rows of the matrices $\boldsymbol{\Delta}_v$ and $\boldsymbol{\Lambda}_v$, respectively. Now for $i=2$ we see from (3.7) that y_{2t} is related to $y_{1,t-1},\ldots,$ $y_{1,t-p_2}$ and $b_{1,t-1},\ldots,b_{1,t-q_2}$ via

$$(\Delta_{21}^{(1)}B+\cdots+\Delta_{21}^{(p_2)}B^{p_2})y_{1t}-(\Lambda_{21}^{(1)}B+\cdots+\Lambda_{21}^{(q_2)}B^{q_2})b_{1t}, \quad (3.8)$$

where $A_{ij}^{(v)}$ denotes the (i,j)th element of the matrix $\boldsymbol{A}_v$. Since, by (3.7) with $i=1$,

$$\begin{aligned} B^{\ell}(y_{1t}-b_{1t})=\{&\boldsymbol{\Delta}_1^{(1)}B+\cdots+\boldsymbol{\Delta}_1^{(p_1)}B^{p_1}\}\boldsymbol{y}_{t-\ell}\\ &-\{\boldsymbol{\Lambda}_1^{(1)}B+\cdots+\boldsymbol{\Lambda}_1^{(q_1)}B^{q_1}\}\boldsymbol{b}_{t-\ell}, \end{aligned} \quad (3.9)$$

it is clear that if all the $\boldsymbol{y}$'s and the $\boldsymbol{b}$'s on the right hand side of (3.9) are in the component model for y_{2t}, then either the coefficient of $y_{1,t-\ell}$ or that of $b_{1,t-\ell}$ is redundant given that the other is in the model. Therefore, if $p_2>p_1$ and $q_2>q_1$, then for each pair of parameters $(\Delta_{21}^{(\ell)},\Lambda_{21}^{(\ell)})$ in (3.7), $\ell=1,\ldots,\min\{p_2-p_1,q_2-q_1\}$, only one of them is needed.

The preceding method of spotting redundant parameters in a vector ARMA model of (3.5) is referred to as the *rule of elimination* in Tiao and Tsay (1989). In general, by considering an ARMA model constructed from SCM's and applying the rule of elimination in a pairwise fashion, all redundant AR or MA parameters can be eliminated. For instance, let η_i be the number of redundant parameters of the model structure for y_{it} in (3.5). By applying the rule of elimination to each pair of SCM's, we obtain

$$\eta_i = \sum_{v=1}^{k} \max[0, \min\{p_i - p_v, q_i - q_v\}]. \tag{3.10}$$

Consequently, the total number of unknown parameters in the coefficient matrices of (3.5) is

$$P = k \times \sum_{i=1}^{k}(p_i + q_i) - \sum_{i=1}^{k} \eta_i \tag{3.11}$$

which can be much smaller than $k^2(r + s)$.

3.4. A canonical form for ARMA representation

The results of the preceding subsections outline the SCM approach to specifying a vector ARMA model. However, the condition that the k linearly independent SCM(p_i, q_i) y_{it} are jointly the most parsimonious ones in number of unknown rows does not necessarily imply a unique ARMA representation for $\boldsymbol{z}_t$. For instance, there is the possibility of fixing the redundant parameters in various ways. Thus, to select a "canonical" ARMA representation for $\boldsymbol{z}_t$ via the SCM approach, some further considerations are needed.

Since the transformation matrix $\boldsymbol{T}$ between $\boldsymbol{z}_t$ and $\boldsymbol{y}_t$ is non-singular, Rank $(\boldsymbol{H}^*_\infty)$ = Rank$(\boldsymbol{H}_\infty)$, where $\boldsymbol{H}^*_\infty$ is the corresponding Hankel matrix of the process $\boldsymbol{y}_t$. Thus, the transformed process $\boldsymbol{y}_t$ has the same McMillan degree as the observed process $\boldsymbol{z}_t$. With this in mind, it seems appropriate to select an ARMA model for $\boldsymbol{y}_t$ that directly reflects this McMillan degree. From the ith row of (3.5) and taking conditional expectation with respect to $\boldsymbol{P}_{t-1}$ of (2.6),we have

$$y_{i,t+\ell|t-1} - \sum_{v=1}^{s_i} \boldsymbol{\Delta}_i^{(v)} \boldsymbol{y}_{t+\ell-v|t-1} = 0 \quad \text{if } \ell \geq s_i, \tag{3.12}$$

where $s_i = \max\{p_i, q_i\}$. This shows that, to compute the forecasts $y_{i,t+\ell|t-1}$ for all ℓ, one needs $y_{i,t|t-1}, \dots, y_{i,t+s_i-1|t-1}$. In view of this, it is reasonable to require that s_i to be as small as possible. For the k linearly independent SCM's y_{it}, let

$$s(\boldsymbol{y}_t) = \sum_{i=1}^{k} \max\{p_i, q_i\} \qquad \text{with} \qquad \boldsymbol{y}_t = (y_{1t}, \dots, y_{kt})'. \tag{3.13}$$

Then, define a canonical SCM representation for $\boldsymbol{z}_t$ of (1.1) as follows:

Definition 5. Suppose that $\boldsymbol{z}_t$ is a linear, time-invariant process of (1.1) with $\boldsymbol{\Psi}(B)$ rational. The ARMA(r, s) model (3.5) for $\boldsymbol{y}_t$ is called a *canonical* SCM *representation* for $\boldsymbol{z}_t$ if (a) $\boldsymbol{y}_t$ is a parsimonious set of SCM's according to the

Definition 4, (b) the quantity $s(\boldsymbol{y}_t)$ of (3.13) is not greater than that of any other set of k linearly independent SCM's of $\boldsymbol{z}_t$, and (c) all the redundant parameters in the moving average matrix polynomial $\Lambda(B)$ of (3.5), as discussed in Subsection 3.3, have been set to zero.

The first condition of the above definition ensures that the first term on the right hand side of (3.11) is as small as possible whereas the last condition eliminates any redundancy in parameterization, i.e. maximizing the second term in the right hand side of (3.11). The second condition is imposed so that $s(\boldsymbol{y}_t)$ of (3.13) is as small as possible. From the previous discussion, the minimum of $s(\boldsymbol{y}_t)$, as a function of all possible $\boldsymbol{y}_t$, is the McMillan degree of $\boldsymbol{z}_t$. The third condition is feasible because the SCM approach provides a simple way to identify redundant parameters (see Subsection 3.3). However, this condition appears to be rather arbitrary; one can require instead that all the redundant parameters in the AR polynomials be zero. From the definition, a canonical SCM representation requires that (a) the number of unknown rows be minimum and (b) the orders $\max\{p_i, q_i\}$ of SCM's used be as small as possible.

By (3.5) and the results concerning redundant parameters, the total number of unknown parameters of a canonical SCM representation of $\boldsymbol{z}_t$ is given by (3.11). This parameter enumeration does not consider those in the transformation matrix $\boldsymbol{T}$ of (3.4). For a k-dimensional process, $\boldsymbol{T}$ may contain as many as $k(k-1)$ parameters upon normalization. In some cases, $\boldsymbol{T}$ can be reduced to an upper triangular matrix without changing the row-structure of the specified ARMA model. For instance, suppose that $k = 2$, $(p_1, q_1) = (1, 0)$ and $(p_2, q_2) = (1, 1)$. Then, the model specified by SCM approach for the original process $\boldsymbol{z}_t$ is

$$\boldsymbol{T}\boldsymbol{z}_t - \begin{bmatrix} X & X \\ X & X \end{bmatrix} \boldsymbol{T}\boldsymbol{z}_{t-1} = \boldsymbol{T}\boldsymbol{a}_t - \begin{bmatrix} 0 & 0 \\ X & X \end{bmatrix} \boldsymbol{T}\boldsymbol{a}_{t-1}, \tag{3.14}$$

where $\boldsymbol{T}' = [\boldsymbol{v}_0^{(1)}, \boldsymbol{v}_0^{(2)}]$. By rearranging the order of the components of $\boldsymbol{z}_t$ if necessary, we may assume that the first element of $\boldsymbol{v}_0^{(1)}$ is nonzero. Let $\boldsymbol{G}$ be a lower triangular 2×2 matrix with unity on the diagonal and its (2,1)th element the negative of the first element of $\boldsymbol{v}_0^{(2)}$ divided by that of $\boldsymbol{v}_0^{(1)}$. Pre-multiplying (3.14) by $\boldsymbol{G}$ and inserting $\boldsymbol{G}^{-1}\boldsymbol{G}$ in the front of $\boldsymbol{T}\boldsymbol{z}_{t-1}$ and $\boldsymbol{T}\boldsymbol{a}_{t-1}$, we have

$$\boldsymbol{T}^*\boldsymbol{z}_t - \begin{bmatrix} X & X \\ X & X \end{bmatrix} \boldsymbol{T}^*\boldsymbol{z}_{t-1} = \boldsymbol{T}^*\boldsymbol{a}_t - \begin{bmatrix} 0 & 0 \\ X & X \end{bmatrix} \boldsymbol{T}^*\boldsymbol{a}_{t-1}, \tag{3.15}$$

where $\boldsymbol{T}^* = \boldsymbol{G}\boldsymbol{T}$ is an upper triangular matrix. Thus, in this particular instance, one can make the transformation matrix upper triangular without changing the row-structure of the ARMA model of the transformed process. Furthermore, from (3.15), the orders of two SCM's are not altered. In general, whenever the orders of any two SCM's are nested, namely $p_i \le p_j$ and $q_i \le q_j$, one can simplify the transformation matrix $\boldsymbol{T}$, by eliminating a non-zero parameter, without altering the row-structure of SCM specification. More specifically, suppose that the orders of SCM's $\boldsymbol{y}_t$ of $\boldsymbol{z}_t$ are (p_i, q_i) for $i = 1, \dots, k$. Then, to obtain further simplification in the transformation matrix $\boldsymbol{T}$, one can simply examine the $\binom{k}{2}$ pairs of SCM's. For any nested pair, by using the technique illustrated in (3.14)-(3.15), one can identify a zero parameter of $\boldsymbol{T}$. Mathematically, the total number of zero parameters identified by such a procedure is

$$\tau = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \mathrm{Ind}(\min\{p_j - p_i, q_j - q_i\} \ge 0),$$

where $\mathrm{Ind}(\cdot)$ is an indicator operator which assumes the value 1 if its argument is true and the value 0, otherwise.

3.5. An illustration

Suppose that the observed process $\boldsymbol{z}_t$ follows a 4-dimensional ARMA(2,1) model that contains 4 SCM's of orders (0,0), (0,1), (1,0) and (2,1), and these orders are the most parsimonious ones. Then, the above results show that the transformed process $\boldsymbol{y}_t = \boldsymbol{T}\boldsymbol{z}_t$ follows an ARMA(2,1) model with coefficient matrices $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2$ and $\boldsymbol{\Lambda}_1$ given by

$$\boldsymbol{\Delta}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ X & X & X & X \\ X & X & X & X \end{bmatrix}, \qquad \boldsymbol{\Delta}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ X & X & X & X \end{bmatrix},$$
$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ X & X & X & X \\ 0 & 0 & 0 & 0 \\ 0 & X & 0 & X \end{bmatrix}, \tag{3.16}$$

where 0 denotes a zero parameter and X denotes an unknown parameter. In (3.16) the (4,1) and the (4,3) elements of $\boldsymbol{\Lambda}_1$ are set to zero, because by applying the rule of elimination they are redundant once the (4,1) and the (4,3) elements of $\boldsymbol{\Delta}_1$ are in the model. Consequently, in this particular instance modeling the transformed series $\boldsymbol{y}_t$ would involve 18 parameters in the coefficient matrices instead of 48, a saving of some 30 parameters in estimation. The reduction could be even more substantial when the dimension k is relatively large. In view of the possibility of high correlations among the unconstrained 48 parameter estimates, the reduction could drastically simplify the complexity in estimation.

3.6. Some properties of SCM model representation

We briefly mention in this subsection some properties of a canonical SCM representation.

Theorem 3. *For a k-dimensional linear, time-invariant process z_t of* (1.1), *suppose that $\{y_{it}\}$ is a parsimonious set of k linearly independent* SCM*'s of order (p_i, q_i). Then the autoregressive and the moving average matrix polynomials, $\Delta(B)$ and $\Lambda(B)$, of* (3.5) *are left coprime. Furthermore,* $\deg[|\Delta(B)|] \leq \sum_{i=1}^{k} p_i$ *and* $\deg[|\Lambda(B)|] \leq \sum_{i=1}^{k} q_i$.

Theorem 4. *Suppose that z_t is a linear, time-invariant process of* (1.1) *with $\Psi(B)$ rational. Then* (i) *there exists a canonical* SCM *representation for z_t;* (ii) *any two canonical* SCM *representations of z_t are related by a unimodular matrix.*

A proof of Theorem 3 is given in the Appendix. Part (i) of Theorem 4 follows directly the modeling procedure of Tiao and Tsay (1989) with (i) a minor modification in the searching path to ensure that the requirement $\max\{p_i, q_i\}$ be minimal is satisfied, and (ii) the sample autocovariances be replaced by the theoretical autocovariances of the process z_t. Part (ii) of Theorem 4 follows from Theorem 2.2.1 of Hannan and Deistler (1988) and Theorem 3 above.

Some remarks are in order. First, the transformation matrix $\boldsymbol{T}$ discussed so far is based on the theoretical autocovariance matrices of z_t. In practice, sample autocovariances are used, and hence $\boldsymbol{T}$ may be data-dependent. Second, since autocovariances of z_t are determined by the $\Psi(B)$ matrix polynomial of (1.1), any transfer function satisfying $[\Phi(B)]^{-1}\Theta(B) = \Psi(B)$ would give rise to the same canonical SCM representation.

4. Some Relationships

Based on the results of Kronecker indices and orders of SCM, we now discuss some relationships between them. Again, let $K = \{d_j | j = 1, \ldots, k\}$ be the set of Kronecker indices of z_t, and y_{it} be an SCM(p_i, q_i) of z_t such that $\{y_{it} | i = 1, \ldots, k\}$ jointly give rise to a canonical SCM representation for z_t.

By (2.11), each Kronecker index d_j introduces an SCM for z_t, and the $\boldsymbol{v}_0$ vector of which is the $\boldsymbol{u}_0^{(j)}$ defined in (2.9). Furthermore, this SCM has an order (p_j^*, q_j^*) such that $p_j^* \leq d_j$ and $q_j^* \leq d_j$. However, since d_j is the smallest nonnegative integer for (2.11) to hold, it follows that we have either $p_j^* = d_j$ or $q_j^* = d_j$, or both. In other words, at least one of the AR or MA order of the SCM introduced by the Kronecker index d_j is equal to d_j. Thus, $\max\{p_j^*, q_j^*\} = d_j$. Consequently, corresponding to the Kronecker indices $K = \{d_j\}$, there exists a set of k linearly independent SCM(p_j^*, q_j^*) such that $K = \{\max[p_j^*, q_j^*]\}$. Since $\sum_{j=1}^{k} d_j = \sum_{j=1}^{k} \max\{p_j^*, q_j^*\} = m$, the McMillan degree of z_t, it is seen that for a given linear time-invariant process $\boldsymbol{z}_t$, one can find at least a set of k linearly independent SCM's that achieves the minimum of the quantity $s(\boldsymbol{y}_t)$ defined in (3.13).

On the other hand, consider the SCM(p_i, q_i) y_{it} of $\boldsymbol{z}_t$. Define $s_i = \max\{p_i, q_i\}$. By (3.12), one needs the s_i quantities $y_{i,t|t-1}, \ldots, y_{i,t+s_i-1|t-1}$ to compute all the forecasts $y_{i,t+\ell|t-1}$ for any ℓ. Thus, for model (3.5), the total number of quantities needed in order to compute all of the elements of $\boldsymbol{F}_{t|t-1}$ is $s = \sum_{i=1}^{k} s_i$. Since $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$ have the same McMillan degree, we have $m \le s$ for $\boldsymbol{y}_t$. However, since $\boldsymbol{y}_t$ is canonical, $m = s$; otherwise, the second condition of Definition 5 is violated.

Next, we briefly explain that the two sets $K = \{d_j\}$ and $S = \{s_i\}$ are equivalent. Without loss of generality, assume that the k SCM's have been rearranged so that $s_1 \le s_2 \le \cdots \le s_k$. In other words, rearrange the SCM's according to increasing order of $s_i = \max\{p_i, q_i\}$. Denote the associated vectors of y_{it}'s by $\boldsymbol{v}_\ell^{(i)}$ for $\ell = 0, \ldots, p_i$; $i = 1, \ldots, k$. By Definition 2 and Equation (3.1), we have

$$\boldsymbol{v}_0^{(i)\prime} \boldsymbol{z}_t + \sum_{\ell=1}^{s_i} \boldsymbol{v}_\ell^{(i)\prime} \boldsymbol{z}_{t-\ell} = \boldsymbol{v}_0^{(i)\prime} \boldsymbol{a}_t + \sum_{\ell=1}^{s_i} \boldsymbol{h}_\ell^{(i)\prime} \boldsymbol{a}_{t-\ell}, \tag{4.1}$$

where $\boldsymbol{v}_\ell^{(i)} = \boldsymbol{0}$ if $\ell > p_i$ and $\boldsymbol{h}_\ell^{(i)} = \boldsymbol{0}$ if $\ell > q_i$. Consider $i = 1$ and denote by j_1 the index of the last non-zero element of $\boldsymbol{v}_0^{(1)}$. Then, by comparing (4.1) for $i = 1$ with (2.13), (4.1) implies that the j_1th Kronecker index of $\boldsymbol{z}_t$ satisfies $d_{j_1} \le s_1$. Suppose that $d_{j_1} < s_1$. Then, by the result stated in the second paragraph of this subsection, there exists an SCM with order $(p_{j_1}^*, q_{j_1}^*)$ such that $d_{j_1} = \max\{p_{j_1}^*, q_{j_1}^*\}$. Denote by $\boldsymbol{u}_0$ the vector associated with this SCM. If $\boldsymbol{u}_0$ is linearly independent of $\boldsymbol{v}_0^{(j)}$ for $j > 1$, then by replacing $\boldsymbol{v}_0^{(1)}$ with $\boldsymbol{u}_0$, we obtain a new set of k linearly independent SCM's, say $\boldsymbol{y}_t^*$, such that $s(\boldsymbol{y}_t^*) = d_{j_1} + \sum_{j=2}^{k} s_j < m$, which is impossible as m is the McMillan degree of $\boldsymbol{z}_t$. If $\boldsymbol{u}_0$ is linearly dependent of $\boldsymbol{v}_0^{(j)}$ for $j > 1$, then by replacing one of the $\boldsymbol{v}_0^{(j)}$'s with $j > 1$ by $\boldsymbol{u}_0$, we again obtain a set of k linearly independent SCM's, say $\boldsymbol{z}_t^*$, such that $s(\boldsymbol{z}_t^*) < m$, resulting in yet another contradiction. Consequently, $d_{j_1} = s_1$.

Consider next (4.1) with $i = 2$. Since $s_1 \le s_2$, we may replace $z_{j_1,t}$ in $\boldsymbol{v}_0^{(2)\prime} \boldsymbol{z}_t$ by (4.1) with $i = 1$ and obtain the following equation:

$$\boldsymbol{w}_0^{(2)\prime} \boldsymbol{z}_t + \sum_{\ell=1}^{s_2} \boldsymbol{w}_\ell^{(2)\prime} \boldsymbol{z}_{t-\ell} = \boldsymbol{w}_0^{(2)\prime} \boldsymbol{a}_t + \sum_{\ell=1}^{s_2} \boldsymbol{g}_\ell^{(2)\prime} \boldsymbol{a}_{t-\ell}, \tag{4.2}$$

where $\boldsymbol{w}_0^{(2)} \ne \boldsymbol{0}$ but its j_1th element is zero. The result of $\boldsymbol{w}_0^{(2)}$ is possible, because the k SCM's are linearly independent. Denote by j_2 the index of the last non-zero element of $w_0^{(2)}$. Again, compared with (2.11), (4.2) implies that the j_2th Kronecker index of z_t satisfies $d_{j_2} \leq s_2$. If $d_{j_2} < s_2$, then there exists an SCM of order $(p^*_{j_2}, q^*_{j_2})$ with $d_{j_2} = \max\{p^*_{j_2}, q^*_{j_2}\}$. Denote by $u_0^{(2)}$ the vector that produces this new SCM. From the result stated in the second paragraph of this subsection, the j_2th element of $u_0^{(2)}$ is unity. Also, since $j_2 \neq j_1$, it follows that $u_0^{(2)}$ is linearly independent of $v_0^{(1)}$. Now, if $u_0^{(2)}$ is linearly independent of $v_0^{(j)}$ for $j > 2$, then the k vectors $v_0^{(1)}, u_0^{(2)}$, and $v_0^{(j)}$'s for $j > 2$ give rise to a set of k linearly independent SCM's, say y_t^*, such that $s(y_t^*) < m$, a contradiction. If $u_0^{(2)}$ is linearly dependent of $v_0^{(j)}$'s with $j > 2$, then one can use $u_0^{(2)}$ in the place of one of $v_0^{(j)}$'s with $j > 2$ and obtains a set of k linearly independent SCM's, say x_t, such that $s(x_t) < m$, which again results in a contradiction. Thus, $d_{j_2} = s_2$.

By repeating the same exercise, we obtain that, for each s_i, there exists a Kronecker index d_{j_i} of z_t such that $d_{j_i} = s_i$. Since the vectors $v_0^{(i)}$'s are linearly independent, the set $\{j_i | i = 1, \ldots, k\}$ is the same as $\{i | i = 1, \ldots, k\}$. Consequently, we have $K = \{d_j\} = \{d_{j_i}\}$ where $d_{j_i} = s_i$ if the SCM's y_{it} corresponds to a canonical SCM representation for z_t. This establishes a relationship between Kronecker indices and the orders of SCM's in a canonical SCM representation:

Theorem 5. *Suppose that z_t is a linear, time-invariant process of* (1.1) *with Kronecker indices $K = \{d_j\}$ such that $m = \sum_{j=1}^k d_j < \infty$. Suppose also that $y_t = (y_{1t}, \ldots, y_{kt})'$, where y_{it} is an* SCM(p_i, q_i), *corresponds to a canonical* SCM *representation of z_t. Then, the set K is equivalent to the set* $\{\max[p_i, q_i]\}$ *and , hence, $m = \sum_{i=1}^k \max\{p_i, q_i\}$.*

5. Discussions

In this section, we point out some differences between the Kronecker index and the scalar-component model (SCM) approaches to specifying the ARMA structure of a linear, time-invariant process z_t. First of all, the Kronecker index approach specifies an Echelon form directly for z_t. On the other hand, the SCM approach identifies an ARMA model for a transformed process y_t. Of course, an ARMA(p, q) model for z_t can be obtained from that for y_t by a simple linear transformation, but the resulting model for z_t may not be in a simple form. Also, there are situations in which one may prefer to have a direct model for z_t. One can argue, however, that the transformed process y_t could be substantively meaningful in some cases. Thus, the choice between modelling z_t or y_t appears to be problem-dependent.

In terms of parameterization, Equations (2.15) and (3.11) give respectively the numbers of unknown parameters in an Echelon form for z_t and in a canonical SCM representation for z_t, i.e., a vector ARMA model for y_t. Two extreme cases may occur. The first extreme case is $q_i = 0$ or $p_i = 0$ for all k SCM's. In this case, (3.11) reduces to $P = ks$ which is smaller than the result of (2.15). The second extreme case is $p_i = q_i = d$ for all k SCM's. Here (3.11) becomes $P = 2ks$ which is exactly the same as N of (2.15). Of course, in the later case, there is no need to do the transformation in the SCM approach, because the transformation fails to produce any simplification in parameterization.

Another point that is worth mentioning is that the SCM approach requires specification of $2k$ integer-valued parameters, namely p_i and q_i, whereas the Kronecker index approach only needs k integer-valued parameters. In this regard, one may treat the SCM approach as a refinement over the Kronecker indices so that the AR and the MA orders of each scalar component can be separated. This refinement can sometimes produce further simplification in the number of real-valued parameters in the matrix polynomials.

Appendix: A Proof of Theorem 3

Proof. Write the vector ARMA model of $\boldsymbol{y}_t$ by

$$\boldsymbol{\Delta}(B)\boldsymbol{y}_t = \boldsymbol{\Lambda}(B)\boldsymbol{b}_t. \tag{A.1}$$

If $\boldsymbol{\Delta}(B)$ and $\boldsymbol{\Lambda}(B)$ are not left coprime, then there exists a matrix polynomial $\boldsymbol{L}(B)$ such that (a) deg[det $\boldsymbol{L}(B)$] > 0 and (b) $\boldsymbol{\Delta}(B) = \boldsymbol{L}(B)\boldsymbol{\Delta}^*(B)$ and $\boldsymbol{\Lambda}(B) = \boldsymbol{L}(B)\boldsymbol{\Lambda}^*(B)$. Since deg[det $\boldsymbol{L}(B)$] > 0, by the *Smith Form*, we can write

$$\boldsymbol{L}(B) = \boldsymbol{U}(B)\boldsymbol{\Xi}(B)\boldsymbol{V}(B), \tag{A.2}$$

where $\boldsymbol{U}(B)$ and $\boldsymbol{V}(B)$ are unimodular matrix polynomials and

$$\boldsymbol{\Xi}(B) = \mathrm{Diag}\{\alpha_1(B), \ldots, \alpha_k(B)\}$$

such that $\alpha_i(B)$ divides $\alpha_{i+1}(B)$ for $i = 1, \ldots, k-1$. Obviously, deg[$\alpha_\ell(B)$] > 0 for some ℓ; otherwise, deg[det $\boldsymbol{L}(B)$] $= 0$, a contradiction.

Now from (A.1) and (A.2) and by cancelling $\boldsymbol{U}(B)$, we have

$$\boldsymbol{\Xi}(B)\boldsymbol{V}(B)\boldsymbol{\Delta}^*(B)\boldsymbol{y}_t = \boldsymbol{\Xi}(B)\boldsymbol{V}(B)\boldsymbol{\Lambda}^*(B)\boldsymbol{b}_t. \tag{A.3}$$

Let (p_i^*, q_i^*) be the maximum orders of the AR and MA polynomials in the ith row of (A.3). Then, Equation (A.3) says that there exists an alternative ARMA representation for the linear system $\boldsymbol{y}_t$ (or equivalently $\boldsymbol{z}_t$), and this representation consists of k linearly independent SCM's of orders (p_i^*, q_i^*). From the cancellation of $\boldsymbol{U}(B)$, we have $\sum_{i=1}^k (p_i^* + q_i^*) \le \sum_{i=1}^k (p_i + q_i)$. On the other hand, by the condition of the theorem, $\sum_{i=1}^k (p_i + q_i) \le \sum_{i=1}^k (p_i^* + q_i^*)$. Thus the ARMA representations (A.3) and (A.1) have the same number of nonzero rows.

Next, write $M(B) = V(B)\Delta^*(B)$ and $N(B) = V(B)\Lambda^*(B)$ and let $A_{ij}(B)$ be the (i,j)th element of the matrix polynomial $A(B)$. From (A.3), since $\Xi(B)$ is diagonal, we have

$$p_i^* = \deg[\alpha_i(B)] + \max_j\{\deg[M_{ij}(B)]\}, \tag{A.4}$$

$$q_i^* = \deg[\alpha_i(B)] + \max_j\{\deg[N_{ij}(B)]\}. \tag{A.5}$$

However, from (A.3), we also have

$$M(B)y_t = N(B)b_t. \tag{A.6}$$

This model says that we can find a set of k linearly independent SCM's of orders (r_i, s_i) where $r_i = \max_j\{\deg[M_{ij}(B)]\}$ and $s_i = \max_j\{\deg[N_{ij}(B)]\}$. From (A.4) and (A.5), we have

$$r_i + s_i \leq p_i^* + q_i^* \quad \text{for all } i.$$

Furthermore, for any ℓ with $\deg[\alpha_\ell(B)] > 0$, we have $r_\ell + s_\ell < p_\ell^* + q_\ell^*$. Since there exists at least one such ℓ, we have

$$\sum_{i=1}^{k}(r_i + s_i) < \sum_{i=1}^{k}(p_i^* + q_i^*) = \sum_{i=1}^{k}(p_i + q_i),$$

implying that the set of newly found SCM's is more parsimonious in row than $\{y_{it}\}$. This contradicts the condition of the theorem. Finally, the results concerning degrees of the determinantal polynomials are easy to obtain, and the proof is complete.

Acknowledgement

This research is supported in part by the National Science Foundation and by the William S. Fishman Faculty Research Fund at the Graduate School of Business, the University of Chicago.

References

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

Deistler, M. (1985). General structure and parameterization of ARMA and state-space systems and its relation to statistical problems. *Handbook of Statistics 5* (Edited by E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 257-277. North-Holland, Amsterdam.

Hannan, E. J. (1969). The identification of vector mixed autoregressive-moving average systems. *Biometrika* **56**, 223-225.

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

Solo, V. (1986). Topics in advanced time series analysis. *Lecture Notes in Math.* Springer Verlag, New York.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser.B* **51**, 157-213.

Tsay, R. S. (1989). Parsimonious parameterization of vector ARMA models. *J. Bus. & Econom. Statist.* **7**, 327-341.

Graduate School of Business, University of Chicago, Chicago, IL 60637, U.S.A.

(Received June 1989; accepted June 1990)