

SPECTRAL FACTORISATION AND PREDICTION OF MULTIVARIATE PROCESSES WITH TIME-DEPENDENT RATIONAL SPECTRAL DENSITY MATRICES

N. M. SPENCER¹ and V. V. ANH²

(Received 17 May 1990; revised 26 November 1990)

Abstract

This paper considers discrete multivariate processes with time-dependent rational spectral density matrices and gives a solution to the spectral factorisation problem. As a result, the corresponding state space representation for the process is obtained. The relationship between multivariate processes with time-dependent rational spectral density matrix functions and multivariate ARMA processes with time-dependent coefficients is discussed. Solutions for the prediction problem are given for the case when only finite data is available and the case when the whole history of the process is known.

1. Introduction

In both the univariate and multivariate cases, ARMA models are a popular choice for the analysis of stochastic processes; one reason being their distinctive rational form for the spectral density function. A discrete multivariate stationary ARMA(n, m) process

$$\Phi(B)\mathbf{X}(t) = \theta(B)\xi(t), \quad (1.1)$$

where $\Phi(z) = I + A_1z + \dots + A_nz^n$ and $\theta(z) = D_0 + D_1z + \dots + D_mz^m$, B is the backshift operator such that $B\mathbf{X}(t) = \mathbf{X}(t-1)$ and $\{\xi(t)\}$ is a white noise process, is known to have the spectral density matrix function with elements as rational functions of $e^{-i\lambda}$. This matrix can be factorised into

$$\mathbf{f}(\lambda) = \frac{1}{2\pi} \Phi^{-1}(e^{-i\lambda}) \theta(e^{-i\lambda}) \Xi \theta(e^{-i\lambda})^* \Phi(e^{-i\lambda})^{*-1}, \quad -\pi \leq \lambda \leq \pi, \quad (1.2)$$

^{1,2}School of Mathematics, Queensland Institute of Technology, Brisbane Q 4001, Australia.

© Copyright Australian Mathematical Society 1991, Serial-fee code 0334-2700/91

where

$$E\xi(t)\xi^*(s) = \begin{cases} \Xi, & t = s, \\ 0, & t \neq s, \end{cases}$$

and $*$ denotes the complex conjugate transpose (see [2] Section 11.8). Conversely, if $\{X(t)\}$ has the spectral density matrix of the form (1.2) where all the zeros of $\Phi(z)$ are on or inside the unit circle and those of $\theta(z)$ are inside the unit circle then $\{X(t)\}$ has the ARMA representation (1.1).

In the analysis of time series, stationarity is considered a desirable feature; however nonstationary processes are more realistic in many applications. By allowing the coefficients of the ARMA model or the rational spectrum to become time-dependent we introduce a form of nonstationarity. It should be noted that, for the time-dependent coefficient case, the equivalence of ARMA processes (1.1) and those with spectral density matrix (1.2) no longer holds.

Priestley [11, 12], using the concept of evolutionary spectra, showed for the discrete univariate case that an ARMA process with time-dependent coefficients has a rational spectrum if the coefficients vary *smoothly* with time. As a by-product of our solution to the factorisation problem in Section 2, we give another specific condition for a process with a time-dependent rational spectral density matrix to have a time-dependent multivariate ARMA form.

Since the assumption of an ARMA form for a process is more restrictive than that of a rational spectrum form, this paper will initiate from processes with a rational spectrum of the form (1.2), but now the coefficients in the matrix polynomials are allowed to be time-dependent. We shall consider the prediction problem for these processes.

Considering a univariate discrete-time process $\{X(t)\}$ with an ARMA(p, q) representation with time-dependent coefficients, Whittle [19] shows that the linear least squares (LS) predictor of $X(T+k)$ based upon $\{X(t), t \leq T\}$ is given by a recursion of type

$$\sum_{i=0}^q \beta_i(T) \hat{X}(T+k-i) = \sum_{j=0}^r c_j(T) X(T-j) + v(T) \quad (1.3)$$

where $r = \max(p-1, q-k)$, β_i, c_i and v are functions of the coefficients of the ARMA (p, q) process. It should be noted that, in Whittle's solution, the coefficients β_i must satisfy a linear system, which necessitates the assumption that the matrix of this system be nonsingular.

Another notable solution to the prediction problem for $\{X(t)\}$, an ARMA process with time-dependent coefficients observed on $(-\infty, T]$, is given by Rissanen and Barbosa [14, 15] (see also [13] and [3], pp. 147–155). In Rissanen and Barbosa's solution, the predictor is also given in a recursive form similar to (1.3), but the corresponding coefficients β_i and c_i are now given

in terms of the elements of the Cholesky factor of an appropriate covariance matrix. The Cholesky factor itself can be generated by a recursion.

In this paper, we consider multivariate discrete-time processes with a time-dependent rational spectral density matrix. The consideration of evolutionary spectra for nonstationary processes was first introduced by Priestley [10]. A description of these processes with more recent references is given in Section 2. Our approach to the prediction problem of these processes is to obtain, under some slowly-varying condition on the coefficients, a state-space representation of the process. This is equivalent to solving the spectral factorisation problem or, in the same direction, the Wold decomposition problem (see [1], Section 9.3). Our solution relies on the reproducing kernel Hilbert space (RKHS) techniques as explored by Parzen [8] and Hajek [4].

The state-space representation problem is considered in Section 2. Section 3 gives a solution to the prediction problem when only finite data is available. When the entire history of the process is available, we obtain in Section 4 a recursive form for the predictor. The recursion is explicit and requires no further condition.

2. State-space representation

Let $\{\mathbf{U}(t), t \in \mathbf{Z}\}$ be a nonstationary p -variate second-order purely non-deterministic process with constant maximum rank. Without loss of generality, we assume that $E\mathbf{U}(t) = \mathbf{0}$. As shown in Melard and Herteleer-de Schutter [6], $\mathbf{U}(t)$ can be represented as

$$\mathbf{U}(t) = \int_{-\pi}^{\pi} \phi_t(\lambda) \mathbf{Z}(d\lambda)$$

where

$$\phi_t(\lambda) = e^{it\lambda} \mathcal{A}(t, \lambda)$$

and

$$\mathcal{A}(t, \lambda) = \sum_{j=0}^{\infty} e^{-ij\lambda} \psi_j(t),$$

the $\psi_j(t)$ being the coefficients in the normalised Wold-Cramer decomposition of $\mathbf{U}(t)$ (see Equations (1) and (2) of [6]). Now the Wold-Cramer spectral density function is defined by

$$\mathbf{g}_t(\lambda) = (1/(2\pi)) \mathcal{A}(t, \lambda) \mathcal{A}(t, \lambda)^*.$$

In this paper we consider a process with the particular form for $\mathbf{g}_t(\lambda)$ defined by

$$\mathbf{g}(t, s, \lambda) = \frac{1}{2\pi} \left(\sum_{j=0}^q \mathbf{D}_j(t) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \times \Xi \left(\sum_{k=0}^n \mathbf{A}_k(s) e^{-ik\lambda} \right)^{* -1} \left(\sum_{j=0}^q \mathbf{D}_j(s) e^{-ij\lambda} \right)^* \tag{2.1}$$

$q < n$, $-\pi \leq \lambda \leq \pi$, where for each t , $|\mathbf{D}_0(t)| \neq 0$ and $\mathbf{A}_k(t)$ and $\mathbf{D}_j(t)$ are $p \times p$ matrices such that

$$\left| \sum_{k=0}^n \mathbf{A}_k(t) z^k \right| \neq 0 \quad \text{for } |z| \leq 1 \tag{2.2}$$

and

$$\left| \sum_{j=1}^q \mathbf{D}_j(t) z^j \right| \neq 0 \quad \text{for } |z| \leq 1. \tag{2.3}$$

and Ξ being a nonnegative definite $p \times p$ matrix. [The conditions (2.2) and (2.3) correspond to the AR and MA regularity conditions for ARMA models with time-varying coefficients. (See, for example, [17].) Another type of regularity conditions for these models involves

$$\sum_{s=-\infty}^t |H(t, s)| < \infty \quad \text{for all } t,$$

where $H(t, s)$ is the one-sided Green's function associated with the homogeneous equation of the AR or MA difference operator (see [7, 9]). However, conditions of the above type do not seem simpler than the conditions on the roots given in (2.2) and (2.3).]

We first consider a finite segment $\{\mathbf{X}'(t) = (X_1(t), \dots, X_p(t)), 0 \leq t \leq N, N \geq 2n\}$ of a zero mean second-order full rank process with the representation

$$\mathbf{X}(t) = \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda) \tag{2.4}$$

with

$$\langle \mathbf{Z}_i(d\lambda), \mathbf{Z}_j(d\mu) \rangle = \begin{cases} \frac{1}{2\pi} d\lambda & \text{if } \mu = \lambda, \\ 0 & \text{if } \mu \neq \lambda, \end{cases} \quad i, j = 1, \dots, p; \\ \Xi = \Delta \Delta^*,$$

$\mathbf{A}_k(t)$ being a $p \times p$ matrix defined above.

The spectral density matrix of $\{\mathbf{X}(t)\}$ is then

$$f(t, s, \lambda) = \frac{1}{2\pi} \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \Xi \left(\sum_{k=0}^n \mathbf{A}_k(s) e^{-ik\lambda} \right)^{* -1}, \quad -\pi \leq \lambda \leq \pi. \tag{2.5}$$

For this process define the $p \times 1$ vector $\mathbf{Y}(t)$ as

$$\mathbf{Y}(t) = \sum_{k=0}^n \mathbf{C}_k(t) \mathbf{X}(t - k), \quad n \leq t \leq N, \tag{2.6}$$

where the coefficient matrices $\mathbf{C}_k(t)$ are chosen according to

$$\sum_{k=0}^n \mathbf{C}_k(t) \left(\sum_{j=0}^n \mathbf{A}_j(t - k) e^{-ij\lambda} \right)^{-1} e^{-ik\lambda} = \mathbf{I}. \tag{2.7}$$

We shall consider only those processes for which (2.7) has a unique solution.

THEOREM 1. *The process $\{\mathbf{Y}(t), n \leq t \leq N\}$ under conditions (2.2) and (2.7) is mutually uncorrelated with covariance matrix Ξ and is uncorrelated with $\{\mathbf{X}(0), \dots, \mathbf{X}(n - 1)\}$. The closed linear manifold $L_2\{\mathbf{X}(0), \dots, \mathbf{X}(n - 1), \mathbf{Y}(t), n \leq t \leq N\}$ coincides with the closed linear manifold $L_2\{\mathbf{X}(t), 0 \leq t \leq N\}$.*

PROOF. From (2.4), $\mathbf{Y}(t)$ has the spectral representation

$$\begin{aligned} \mathbf{Y}(t) &= \sum_{k=0}^n \mathbf{C}_k(t) \int_{-\pi}^{\pi} e^{i(t-k)\lambda} \left(\sum_{j=0}^n \mathbf{A}_j(t - k) e^{-ij\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda) \\ &= \int_{-\pi}^{\pi} e^{it\lambda} \sum_{k=0}^n \mathbf{C}_k(t) e^{-ik\lambda} \left(\sum_{j=0}^n \mathbf{A}_j(t - k) e^{-ij\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda) \\ &= \int_{-\pi}^{\pi} e^{it\lambda} \Delta \mathbf{Z}(d\lambda), \end{aligned}$$

by choice of the coefficient matrices $\mathbf{C}_k(\cdot)$ as per (2.7).

Thus

$$\begin{aligned} \langle \mathbf{Y}(t), \mathbf{Y}(s) \rangle &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(t\lambda - s\mu)} \Delta \Delta^* \langle \mathbf{Z}(d\lambda), \mathbf{Z}(d\mu) \rangle \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \Xi d\lambda \\ &= \Xi \delta_{ts}. \end{aligned}$$

Thus $\{\mathbf{Y}(t), n < t \leq N\}$ is a white noise process. Also, for $n < t < N$ and $0 \leq s \leq n - 1$,

$$\begin{aligned} \langle \mathbf{Y}(t), \mathbf{X}(s) \rangle &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(t\lambda - s\mu)} \Xi \left(\sum_{j=0}^n \mathbf{A}_j(s) e^{ij\mu} \right)^{* - 1} \langle \mathbf{Z}(d\lambda), \mathbf{Z}(d\mu) \rangle \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \Xi \left(\sum_{j=0}^n \mathbf{A}_j(s) e^{-ij\lambda} \right)^{* - 1} d\lambda \\ &= \frac{1}{2\pi i} \int_{|z|=1} z^{t-s-1} \Xi \left(\sum_{j=0}^n \mathbf{A}_j(s) z^{-j} \right)^{* - 1} dz, \text{ putting } z = e^{i\lambda} \\ &= 0 \end{aligned}$$

for $t > s$ by Cauchy's Theorem as

$$\left| \left(\sum_{j=0}^n \mathbf{A}_j(s) e^{-ij\lambda} \right)^{* - 1} \right| = \left| \left(\sum_{j=0}^n \mathbf{A}_j(s) e^{-ij\lambda} \right)^* \right|^{-1}$$

and

$$\left| \sum_{j=0}^n \mathbf{A}_j(s) e^{ij\lambda} \right| \neq 0$$

in the closed unit disc according to the assumption given in (2.4).

The RKHS $H(\mathbf{f})$ of $\{\mathbf{X}(t)\}$ corresponding to \mathbf{f} of (2.5) contains elements \varnothing defined by $\varnothing(t) = \langle \mathbf{X}(t), \mathbf{v} \rangle = E[\mathbf{X}(t)\mathbf{v}^*]$ for some unique \mathbf{v} in $L_2\{\mathbf{X}(t), 0 \leq t \leq N\}$. The Hilbert space $H(\mathbf{f})$ is isomorphic to the space of random vectors of the form $\mathbf{v} = \sum_{t=n}^N \mathbf{h}(t)\mathbf{Y}(t)$ where $\mathbf{h}(t)$ is an element of the space of quadratically summable matrix valued functions, and which solve the equation $\langle \mathbf{Y}(t), \mathbf{v} \rangle = \mathbf{h}(t)$, $n \leq t \leq N$, (see [8]).

Now $L_2\{\mathbf{X}(t), 0 \leq t \leq N\} \supset L_2\{\mathbf{X}(0), \dots, \mathbf{X}(n - 1), \mathbf{Y}(t), n \leq t \leq N\}$.

Let $\mathbf{v} \in L_2\{\mathbf{X}(t), 0 \leq t \leq N\}$ be such that $\mathbf{v} \perp \mathbf{X}(t)$, $0 \leq t \leq n - 1$ and $\mathbf{v} \perp \mathbf{Y}(t)$, $n \leq t \leq N$. Then,

$$\begin{aligned} 0 &= \langle \mathbf{X}(t), \mathbf{v} \rangle \\ &= \varnothing(t), \quad 0 \leq t \leq n - 1 \end{aligned} \tag{2.8}$$

and

$$\begin{aligned}
 0 &= \langle \mathbf{Y}(t), \mathbf{v} \rangle \\
 &= \left\langle \sum_{k=0}^n \mathbf{C}_k(t) \mathbf{X}(t-k), \mathbf{v} \right\rangle \\
 &= \sum_{k=0}^n \mathbf{C}_k(t) \langle \mathbf{X}(t-k), \mathbf{v} \rangle \\
 &= \sum_{k=0}^n \mathbf{C}_k(t) \varnothing(t-k).
 \end{aligned}
 \tag{2.9}$$

This is a homogeneous matrix difference equation with n initial conditions given by (2.8). We need to show that the solution to (2.9) with the initial conditions (2.8) is $\varnothing(t) = 0, 0 \leq t \leq N$. Now, when $t = n$, (2.9) reads

$$\varnothing(n) + \mathbf{C}_1(n)\varnothing(n-1) + \dots + \mathbf{C}_n(n)\varnothing(0) = 0$$

which by the initial conditions, gives

$$\varnothing(n) = 0.
 \tag{2.10}$$

When $t = n + 1$, (2.9) reads

$$\varnothing(n+1) + \mathbf{C}_1(n+1)\varnothing(n) + \dots + \mathbf{C}_n(n+1)\varnothing(1) = 0$$

which implies $\varnothing(n+1) = 0$ due to (2.8) and (2.10). Similarly, $\varnothing(n+2) = \dots = \varnothing(N) = 0$. Therefore the solution to the homogeneous matrix difference equation (2.9) with the n initial conditions (2.8) is $\varnothing(t) = 0, 0 \leq t \leq N$, and consequently $L_2\{\mathbf{X}(t), 0 \leq t \leq N\} = L_2\{\mathbf{X}(0), \dots, \mathbf{X}(n-1), \mathbf{Y}(n), \dots, \mathbf{Y}(N)\}$.

We shall now derive the state space system which represents a process $\{\mathbf{U}(t)\}$ obtained from the above white noise process $\{\mathbf{Y}(t)\}$.

Suppose $\{\mathbf{X}(t), -q \leq t \leq N\}$ is as defined by (2.4) and (2.5), and consider a finite segment of a second order full rank process $\{\mathbf{U}'(t) = (U_1(t), \dots, U_p(t)), 0 \leq t \leq N\}$ having the representation

$$\mathbf{U}(t) = \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{j=0}^q \mathbf{D}_j(t) e^{ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda)
 \tag{2.11}$$

where $\mathbf{Z}(d\lambda)$ is an orthogonal process:

$$\langle \mathbf{Z}_i(d\lambda), \mathbf{Z}_j(d\mu) \rangle = \begin{cases} \frac{1}{2\pi} d\lambda & \text{if } \mu = \lambda, \\ 0 & \text{if } \mu \neq \lambda, \end{cases}
 \tag{2.12}$$

for $i, j = 1, \dots, p$. The coefficient matrices $\mathbf{A}_k(t)$ and $\mathbf{D}_j(t)$ are assumed to satisfy the conditions (2.2) and (2.3). The spectral density matrix $\mathbf{g}(t, s, \lambda)$ of $\{\mathbf{U}(t)\}$ is given by (2.1).

We now define the coefficient matrices $\mathbf{D}_j(t)$, $j = 0, \dots, q$, by the condition

$$\begin{aligned} & \left(\sum_{j=0}^q \mathcal{D}_j(t) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t-j) e^{-ik\lambda} \right)^{-1} \\ &= \left(\sum_{j=0}^q \mathbf{D}_j(t) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \end{aligned} \tag{2.13}$$

THEOREM 2. *The Markovian representation of $\{\mathbf{U}(t)\}$ under conditions (2.2), (2.3), (2.7) and (2.13) is given by*

$$\begin{aligned} & \sum_{k=0}^n \mathbf{C}_k(t) \mathbf{X}(t-k) = \mathbf{Y}(t) \\ & \mathbf{U}(t) = \sum_{j=0}^q \mathcal{D}_j(t) \mathbf{X}(t-j) \\ & E[\mathbf{Y}(t) \mathbf{Y}(s)] = \Xi \delta_{ts}. \end{aligned}$$

PROOF. Let \mathbf{M} be the $p \times p$ matrix difference operator $\mathbf{M} = \sum_{j=0}^q \mathcal{D}_j(t) B^j$ where B is the backshift operator.

Then

$$\mathbf{U}(t) = \sum_{j=0}^q \mathcal{D}_j(t) \mathbf{X}(t-j) = \mathbf{M} \mathbf{X}(t) \tag{2.14}$$

because using the representation (2.11) for $\mathbf{U}(t)$ we obtain

$$\begin{aligned} & \langle \mathbf{U}(t), \mathbf{U}(s) \rangle = E[\mathbf{U}(t) \mathbf{U}(s)^*] \\ &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(t\lambda - s\mu)} \left(\sum_{j=0}^q \mathbf{D}_j(t) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t) e^{-ik\lambda} \right)^{-1} \\ & \quad \times \Xi \left(\sum_{k=0}^n \mathbf{A}_k(s) e^{ik\mu} \right)^{* - 1} \left(\sum_{j=0}^q \mathcal{D}_j(s) e^{ij\mu} \right)^* \langle \mathbf{Z}(d\lambda), \mathbf{Z}(d\mu) \rangle \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} g(t, s, \lambda) d\lambda \\ &= \int_{-\pi}^{\pi} \left(\sum_{j=0}^q \mathbf{D}_j(t) e^{i(t-j)\lambda} \right) \mathbf{f}(t, s, \lambda) \left(\sum_{j=0}^q \mathbf{D}_j(s) e^{i(s-j)\lambda} \right)^* d\lambda \\ &= \left\langle \sum_{j=0}^q \mathcal{D}_j(t) \mathbf{X}(t-j), \sum_{j=0}^q \mathcal{D}_j(s) \mathbf{X}(s-j) \right\rangle, \end{aligned} \tag{2.15}$$

by (2.13).

Conversely, given $U(t)$ defined by (2.14),

$$\begin{aligned} & \left\langle \sum_{j=0}^q \mathcal{D}_j(t) \mathbf{X}(t-j), \sum_{j=0}^q \mathcal{D}_j(s) \mathbf{X}(s-j) \right\rangle \\ &= \left\langle \int_{-\pi}^{\pi} e^{it\lambda} \left(\sum_{j=0}^q \mathcal{D}_j(t) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(t-j) e^{ik\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda), \right. \\ & \quad \left. \int_{-\pi}^{\pi} e^{is\lambda} \left(\sum_{j=0}^n \mathcal{D}_j(s) e^{-ij\lambda} \right) \left(\sum_{k=0}^n \mathbf{A}_k(s-j) e^{-ik\lambda} \right)^{-1} \Delta \mathbf{Z}(d\lambda) \right\rangle \end{aligned}$$

which is (2.15) in view of (2.13). Then by reversing the steps that lead to (2.15) we find that the covariance of $\{\mathbf{U}_t\}$, where $\{\mathbf{U}_t\}$ is defined by (2.14), is the same as that obtained from the spectral representation (2.11).

The Markovian representation of $\{\mathbf{U}(t)\}$ now follows directly.

REMARK 1. The system given in Theorem 2 with white noise input $\{\mathbf{Y}(t)\}$ solves the time-varying spectral factorisation problem for $\mathbf{U}(t)$, and can be written in the standard form for a state space model as illustrated below.

Define

$$\mathbf{z}(t) = \begin{pmatrix} \mathbf{X}(t) \\ \vdots \\ \mathbf{X}(t-n+1) \end{pmatrix}_{((n \times p) \times 1)}$$

From (2.7), we get $\mathbf{C}_0(t) = \mathbf{A}_0(t)$ for any t . From condition (2.2), it is necessary that $\det \mathbf{A}_0(t) \neq 0$, which implies the existence of $\mathbf{C}_0^{-1}(t)$.

Then by writing $\mathbf{C}_k(t) = \mathbf{C}_0(t)^{-1} \mathbf{C}_k(t)$,

$$\begin{aligned} \mathbf{z}(t) &= \begin{pmatrix} -\mathcal{E}_1(t) & -\mathcal{E}_2(t) & \cdots & -\mathcal{E}_{n-1}(t) & -\mathcal{E}_n(t) \\ \mathbf{I}_p & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_p & \mathbf{0} \end{pmatrix} \mathbf{z}(t-1) \\ &+ \begin{pmatrix} \mathcal{E}_0^{-1}(t) \\ \mathbf{0} \\ \mathbf{0} \\ \cdot \\ \mathbf{0} \end{pmatrix} \mathbf{Y}(t) \\ &= \mathcal{A}(t) \mathbf{z}(t-1) + \mathbf{a} \mathbf{Y}(t) \end{aligned}$$

and $\mathbf{X}(t) = (1 \ 0 \ \cdots \ 0) \mathbf{z}(t)$.

Now define

$$\mathbf{z}_*(t) = \begin{pmatrix} \mathbf{X}(t) \\ \vdots \\ \mathbf{X}(t - q + 1) \end{pmatrix}_{((q \times p) \times 1)}, \quad q < n,$$

then as $\mathbf{D}_0(t)$ is nonsingular for each t

$$\mathbf{z}_*(t) = \begin{pmatrix} -\mathcal{D}_0(t)^{-1}\mathcal{D}_1(t) & -\mathcal{D}_0(t)^{-1}\mathcal{D}_2(t) & \cdots & -\mathcal{D}_0(t)^{-1}\mathcal{D}_{q-1}(t) & -\mathcal{D}_0(t)^{-1}\mathcal{D}_q(t) \\ \mathbf{I}_p & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{I}_p & 0 \end{pmatrix} \mathbf{z}_*(t-1) + \begin{pmatrix} \mathcal{D}_0(t)^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \mathbf{U}(t).$$

This can be rewritten as

$$\begin{pmatrix} \mathcal{D}_0(t)^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \mathbf{U}(t) = \begin{pmatrix} \mathbf{I}_p & \mathcal{D}_0(t)^{-1}\mathcal{D}_1(t) & \cdots & \mathcal{D}_0(t)^{-1}\mathcal{D}_q(t) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \mathbf{z}_{**}(t) \tag{2.16}$$

where

$$\mathbf{z}_{**}(t) = \begin{pmatrix} \mathbf{X}(t) \\ \vdots \\ \mathbf{X}(t - q) \end{pmatrix}_{(((q+1) \times p) \times 1)}.$$

Extending $\mathbf{z}_{**}(t)$ to $\mathbf{z}(t)$ converts (2.16) to

$$\begin{pmatrix} \mathcal{D}_0(t)^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \mathbf{U}(t) = \begin{pmatrix} \mathbf{I}_p & \mathcal{D}_0(t)^{-1}\mathcal{D}_1(t) & \cdots & \mathcal{D}_0(t)^{-1}\mathcal{D}_q(t) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \mathbf{z}(t)$$

which can be expressed as $\mathbf{u}(t) = \mathcal{B}(t)\mathbf{z}(t)$. Thus the system (2.15) can be written as

$$\begin{aligned} \mathbf{z}(t) &= \mathcal{A}(t)\mathbf{z}(t-1) + \mathbf{a}\mathbf{Y}(t) \\ \mathbf{u}(t) &= \mathcal{B}(t)\mathbf{z}(t) \quad n \leq t \leq N. \end{aligned} \tag{2.17}$$

REMARK 2. The processes $\{\mathbf{U}(t)\}$ considered in this section are restricted by conditions (2.7) and (2.13). These are a type of ‘slowly-varying’ conditions which are almost always required in the analysis of time-varying systems. There does not appear to be any convenient way to simplify these conditions further, unless specific forms for the matrices $\mathbf{A}_k(t)$ and $\mathbf{D}_j(t)$ are known. However, for the univariate case, the coefficients in the corresponding equations can be obtained explicitly (see [18]).

REMARK 3. The system given in Theorem 2 can be written as

$$\mathbf{M}^{-1}\mathbf{U}(t) = \mathbf{L}^{-1}\mathbf{Y}(t) \tag{2.18}$$

or $\mathbf{U}(t) = \mathbf{M}(\mathbf{L}^{-1}\mathbf{Y}(t))$ where \mathbf{L} is the matrix difference operator given by $\mathbf{L} = \sum_{k=0}^n \mathbf{C}_k(t)B^k$, B being the backshift operator. (2.18) will describe a time-dependent ARMA process if

$$\mathbf{M}(\mathbf{L}^{-1}\mathbf{Y}(t)) = \mathbf{L}^{-1}(\mathbf{M}\mathbf{Y}(t))$$

i.e., if the operators \mathbf{M} and \mathbf{L}^{-1} commute. However,

$$\begin{aligned} \mathbf{L}(\mathbf{M}(\mathbf{L}^{-1}\mathbf{Y}(t))) &= \mathbf{L}(\mathbf{M}\mathbf{X}(t)) \\ &= \mathbf{L}\mathbf{U}(t) \\ &= \sum_{k=0}^n \mathbf{C}_k(t)\mathbf{U}(t-k) \\ &= \sum_{k=0}^n \mathbf{C}_k(t) \sum_{j=0}^q \mathcal{D}_j(t-k)\mathbf{X}(t-k-j), \end{aligned}$$

while

$$\begin{aligned} \mathbf{M}\mathbf{Y}(t) &= \sum_{j=0}^q \mathcal{D}_j(t)\mathbf{Y}(t-j) \\ &= \sum_{j=0}^q \mathcal{D}_j(t) \sum_{k=0}^n \mathbf{C}_k(t-j)\mathbf{X}(t-j-k). \end{aligned}$$

So a sufficient condition for commutativity is

$$\mathbf{C}_k(t)\mathcal{D}_j(t-k) = \mathcal{D}_j(t)\mathbf{C}_k(t-j),$$

or equivalently,

$$\sum_{r=1}^p \mathbf{C}_{k,ir}(t)\mathcal{D}_{j,rl}(t-k) = \sum_{r=1}^p \mathcal{D}_{j,ir}(t)\mathbf{C}_{k,rl}(t-j), \quad i, r, l = 1, \dots, p.$$

If the matrix coefficients are constants, however, then as $\mathbf{D}_j(t) = \mathcal{D}_j(t)$ and $\mathbf{A}_k(t) = \mathbf{C}_k(t)$ for all t ,

$$\begin{aligned} \mathbf{L}(\mathbf{M}(\mathbf{L}^{-1}\mathbf{Y}(t))) &= \mathbf{L}(\mathbf{M}\mathbf{X}(t)) \\ &= \sum_{j=0}^q \mathbf{D}_j\mathbf{X}(t-j) \\ &= \sum_{k=0}^n \mathbf{A}_k \sum_{j=0}^q \mathbf{D}_j\mathbf{X}(t-j-k), \end{aligned}$$

while

$$MY(t) = \sum_{j=0}^q D_j \sum_{k=0}^n A_k X(t - k - j);$$

so the sufficient condition for commutativity becomes $A_k D_j = D_j A_k$. If this condition is satisfied then the system (2.18) describes a multivariate ARMA(n, q) process.

3. Prediction based on finite data

Using classical methods for the solution of difference equations, prediction of processes with time-dependent rational spectral density matrices will be discussed in this section. From the knowledge of this spectral density matrix function for $\{U(t), 0 \leq t \leq N\}$ we have established the following state space representation for $U(T)$ under conditions (2.2), (2.3), (2.7) and (2.13):

$$\sum_{k=0}^n C_k(t) X(t - k) = Y(t) \quad n - m \leq t \leq N \tag{3.1}$$

$$U(t) = \sum_{j=0}^q \mathcal{D}_j(t) X(t - j) \quad 0 \leq t \leq N. \tag{3.2}$$

THEOREM 3. *The linear least squares predictor for $U(t)$ based on $\{U(t), t = 0, \dots, N\}$ is*

$$\begin{aligned} \hat{U}(t+r) = & \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{\rho=r-j-1}^{\tau-1} \Psi_\rho(\tau) \sum_{i=0}^n C_i(\tau-\rho) \sum_{s=0}^{\tau-\rho-i} d_s(\tau-\rho-i) U(s) \\ & - \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{\rho=0}^{\tau-1} \psi_\rho(\tau) \sum_{k=\tau-\rho}^n F_k(\tau-\rho) c(\tau-\rho-k) \end{aligned}$$

where $\tau = t + r - j$, and the coefficients $\Psi_\rho(t)$, $\psi_\rho(t)$, and $F_k(t)$ are as defined below.

PROOF. The solution to the system of matrix difference equations given by (4.1) is

$$X(t) = \sum_{\rho=0}^{t-1} \Psi_\rho(t) Y(t-\rho) - \sum_{\rho=0}^{t-1} \psi_\rho(t) \sum_{k=t-\rho}^n F_k(t-\rho) c(t-\rho-k) \tag{3.3}$$

where $\mathbf{c}(i) = \mathbf{X}(i)$, $i = -n + 1, \dots, 0$ are constant initial vectors,

$$\mathbf{G}(t) = \mathbf{C}_0^{-1}(t)$$

$$\mathbf{F}_k(t) = \mathbf{G}(t)\mathbf{C}_k(t), \quad k = 0, \dots, n$$

and

$$\Psi_\rho(t) = \psi_\rho(t)\mathbf{G}(t - \rho)$$

$$\psi_\rho(t) = \left(\prod_{j=1}^r -\mathbf{F}_{\rho_j} \left(t - \sum_{i=1}^{j-1} \rho_i \right) \right);$$

where the sum

$$\sum_{j=1}^r \rho_j = \rho \quad \text{with } \rho_0 = 0 \text{ for } \rho = 0, \text{ and } \rho_j > 0$$

is over all the permutations of the integers ρ_j that sum to ρ .

This result is a direct extension of the work by Jordan [5], Section 178. Hence for r steps ahead

$$\mathbf{U}(t + r) = \sum_{j=0}^q \mathcal{D}_j(t + r)\mathbf{X}(t + r - j)$$

where the $\mathbf{X}(t + r - j)$, $j = 0, \dots, q$, are given by (3.3). The linear least squares predictor for $\mathbf{U}(t + r)$ is obtained by setting the unobserved values of $\mathbf{Y}(t)$, i.e. $\mathbf{Y}(t + 1), \dots, \mathbf{Y}(t + r)$, equal to zero; thus

$$\hat{\mathbf{U}}(t + r) = \sum_{j=0}^q \mathcal{D}_j(t + r) \left[\sum_{\rho=r-j-1}^{\tau-1} \Psi_\rho(\tau)\mathbf{Y}(\tau - \rho) - \sum_{\rho=0}^{\tau-1} \psi_\rho(\tau) \sum_{k=\tau-\rho}^n \mathbf{F}_k(\tau - \rho)\mathbf{c}(\tau - \rho - k) \right] \quad (3.4)$$

where $\tau = t + r - j$. To express $\hat{\mathbf{U}}(t + r)$ in terms of known past values of $\{\mathbf{U}(t)\}$ we convert (3.2) into the series

$$\mathbf{X}(t) = \sum_{s=-\infty}^t \mathbf{d}_s(t)\mathbf{U}(s). \quad (3.5)$$

Note that this series is actually finite as we are only considering the segment $\{\mathbf{U}(s), 0 \leq s \leq N\}$. The $p \times p$ matrix coefficients $\mathbf{d}_s(t)$, $-\infty < s \leq t$, are

determined by substituting (3.5) into (3.2) and equating coefficients of $U(t)$:

$$\begin{aligned}
 U(t) &= \sum_{j=0}^q \mathcal{D}_j(t) \sum_{s=-\infty}^{t-j} \mathbf{d}(t-j)U(s) \\
 &= \sum_{j=0}^q \sum_{s=-\infty}^t I(s \leq t-j) \mathcal{D}_j(t) \mathbf{d}_s(t-j)U(s) \\
 &= \sum_{s=-\infty}^t \sum_{j=0}^q I(j \leq t-s) \mathcal{D}_j(t) \mathbf{d}_s(t-j)U(s) \\
 &= \sum_{s=-\infty}^t \sum_{j=0}^{q \wedge (t-s)} \mathcal{D}_j(t) \mathbf{d}_s(t-j)U(s)
 \end{aligned}$$

where I is the indicator function, that is, $I(r \leq s) = 1$ if $r \leq s$ and 0 otherwise, and $q \wedge (t - s)$ means $\min\{q, t - s\}$. Hence the relations determining $\mathbf{d}_s(t)$ are given by

$$\sum_{j=0}^{q \wedge (t-s)} \mathcal{D}_j(t) \mathbf{d}_s(t-j) = \mathbf{I}_p \delta_{st} \tag{3.6}$$

where

$$\delta_{st} = \begin{cases} 1, & s = t, \\ 0, & s \neq t. \end{cases}$$

Alternatively substituting (3.2) into (3.5) we obtain

$$\sum_{j=0}^{q \wedge (t-s)} \mathbf{d}_{s+j}(t) \mathcal{D}_j(s+j) = \mathbf{I}_p \delta_{st}.$$

Now using (3.5), (3.1) becomes

$$Y(t) = \sum_{k=0}^n \mathbf{C}_k(t) \sum_{s=-\infty}^{t-k} \mathbf{d}_s(t-k)U(s)$$

so that the least squares predictor (3.4) becomes

$$\begin{aligned}
 \hat{U}(t+r) &= \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{\rho=r-j-1}^{\tau-1} \Psi_\rho(\tau) \sum_{i=0}^n \mathbf{C}_i(\tau-\rho) \sum_{s=0}^{\tau-\rho-i} \mathbf{d}_s(\tau-\rho-i)U(s) \\
 &\quad - \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{\rho=0}^{\tau-1} \psi_\rho(\tau) \sum_{k=\tau-\rho}^n \mathbf{F}_k(\tau-\rho) \mathbf{c}(\tau-\rho-k)
 \end{aligned}$$

where $\tau = t + r - j$.

4. Prediction with infinite past

When the whole history of the process is available we can extend the results due to Whittle [19] for the prediction of a scalar discrete-parameter ARMA process to the multivariate case. Our solution, however, initiates from a multivariate process with a time dependent rational spectral density matrix of the form (2.1) rather than a multivariate time-dependent ARMA process. The solution is explicit (compared with Whittle’s equation (32) and condition (33)).

For a process with a spectral density matrix of the form (2.1) under conditions (2.2), (2.3), (2.7) and (2.13) we have already established the corresponding state space representation which is given by (3.1) and (3.2).

We shall further assume that

$$\left| \sum_{k=0}^n C_k(t)z^k \right| \neq 0 \quad \text{for } |z| \leq 1. \tag{4.1}$$

From (3.1) we can write

$$X(t) = \sum_{v=-\infty}^t a_v(t)Y(v) \tag{4.2}$$

where the convergence of this series is assured by the assumption (4.1) for each t (see [2] Proposition 3.1.1 and Theorem 3.1.1).

The coefficients $a_v(t)$ are determined by substituting (4.2) into (3.1) and equating coefficients of $Y(t)$. This yields the relations

$$\sum_{k=0}^{n \wedge (t-v)} C_k(t)a_v(t-k) = I_p \delta_{vt}. \tag{4.3}$$

Alternatively if (3.1) is substituted into (4.2) we obtain the relations

$$\sum_{k=0}^{n \wedge (t-v)} a_{s+k}(t)C_k(s+k) = I_p \delta_{st}. \tag{4.4}$$

Now from (4.2) we obtain, for r steps ahead, $r \geq 1$,

$$\begin{aligned} X(t+r) &= \sum_{v=-\infty}^t a_v(t+r)Y(v) + \sum_{v=t+1}^{t+r} a_v(t+r)Y(v) \\ &= \sum_{v=-\infty}^t a_v(t+r) \sum_{k=0}^n C_k(v)X(v-k) + \sum_{v=t+1}^{t+r} a_v(t+r)Y(v) \end{aligned} \tag{4.5}$$

from (3.1). The least squares predictor of $\mathbf{X}(t+r)$ is achieved when the unobserved innovations $\mathbf{Y}(t+1), \dots, \mathbf{Y}(t+r)$ are set equal to zero; then (4.5) becomes

$$\widehat{\mathbf{X}}(t+r) = \sum_{v=-\infty}^t \mathbf{a}_v(t+r) \sum_{k=0}^n \mathbf{C}_k(v) \mathbf{X}(v-k).$$

Using the change of variable $v = t - j + k$, $\widehat{\mathbf{X}}(t+r)$ can be written

$$\widehat{\mathbf{X}}(t+r) = \sum_{j=0}^{\infty} \sum_{k=0}^{n \wedge j} \mathbf{a}_{t-j+k}(t+r) \mathbf{C}_k(t+k-j) \mathbf{X}(t-j). \tag{4.6}$$

The relations (4.4) yield that (4.6) is zero for $j \geq n$; therefore

$$\mathbf{X}(t+r) = \sum_{j=0}^{n-1} \sum_{k=0}^j \mathbf{a}_{t-j+k}(t+r) \mathbf{C}_k(t+k-j) \mathbf{X}(t-j). \tag{4.7}$$

A recursion for the predictor $\widehat{\mathbf{U}}(t+r)$ is now given by (4.7), (3.2), and (3.5). Note that only the last n values, $\mathbf{X}(t-n+1), \dots, \mathbf{X}(t)$, from (4.5), are required for the recursion.

The predictor $\widehat{\mathbf{X}}(t+r)$, and hence $\widehat{\mathbf{U}}(t+r)$, is given in recursive form in (4.7). We now show that $\widehat{\mathbf{U}}(t+r)$ can be written in terms of $\{\mathbf{U}(s), -\infty < s \leq t\}$.

THEOREM 4. *The linear least squares predictor of $\mathbf{U}(t+r)$ for r steps ahead in terms of $\{\mathbf{U}(s), -\infty < s \leq t\}$ is*

$$\widehat{\mathbf{U}}(t+r) = \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s),$$

if $q < r$,

$$\begin{aligned} \widehat{\mathbf{U}}(t+r) = & \sum_{j=0}^{r-1} \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s) \\ & + \sum_{j=0}^m \mathcal{D}_j(t+r) \sum_{v=-\infty}^{t+r-j} \mathbf{a}_v(t_r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s), \end{aligned}$$

if $q \geq r$,

and the mean square error is

$$\varepsilon(r)^2 = \sum_{j=0}^{q \wedge (r-1)} \sum_{k=0}^{q \wedge (r-1)} \sum_{v=t+1}^{(t+r-j) \wedge (t+r-k)} \text{tr}(\mathcal{D}_j(t+r) \mathbf{a}_v(t+r-j) \mathbf{\Xi} \mathbf{a}_v^*(t+r-k) \mathcal{D}_k^*(t+r)).$$

PROOF. Using (4.2) in (3.2)

$$U(t+r) = \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^{t+r-j} \mathbf{a}_v(t+r-j) \mathbf{Y}(v). \tag{4.8}$$

If $q < r$, (4.8) becomes

$$\begin{aligned} U(t+r) &= \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \mathbf{Y}(v) \\ &\quad + \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r} \mathbf{a}_v(t+r-j) \mathbf{Y}(v) \end{aligned}$$

and on setting the unobserved values of $Y(t)$ to zero we obtain the least squares predictor

$$\hat{U}(t+r) = \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \mathbf{Y}(v). \tag{4.9}$$

Now using (3.5) in (3.1), $\mathbf{Y}(v)$ is given by

$$\begin{aligned} \mathbf{Y}(v) &= \sum_{k=0}^n \mathbf{C}_k(v) \sum_{s=-\infty}^{v-k} \mathbf{d}_s(v-k) \mathbf{U}(s) \\ &= \sum_{k=0}^n \mathbf{C}_k(v) \sum_{s=-\infty}^v I(s \leq v-k) \mathbf{d}_s(v-k) \mathbf{U}(s) \\ &= \sum_{k=0}^n \mathbf{C}_k(v) \sum_{s=-\infty}^v I(k \leq v-s) \mathbf{d}_s(v-k) \mathbf{U}(s) \\ &= \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{D}_s(v-k) \mathbf{U}(s), \end{aligned} \tag{4.10}$$

Therefore (4.9) becomes

$$\hat{U}(t+r) = \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s). \tag{4.11}$$

Similarly if $q \geq r$

$$\begin{aligned} \hat{U}(t+r) &= \sum_{j=0}^{r-1} \mathcal{D}_j(t+r) \sum_{v=-\infty}^t \mathbf{a}_v(t+r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s) \\ &\quad + \sum_{j=0}^m \mathcal{D}_j(t+r) \sum_{v=-\infty}^{t+r-j} \mathbf{a}_v(t+r-j) \sum_{s=-\infty}^v \sum_{k=0}^{n \wedge (v-s)} \mathbf{C}_k(v) \mathbf{d}_s(v-k) \mathbf{U}(s). \end{aligned} \tag{4.12}$$

(4.11) and (4.12) give the predictor $\hat{U}(t+r)$ in terms of the entire past of $\{U(s), -\infty < s \leq t\}$.

Alternatively (4.9) can be expressed as

$$\begin{aligned} \hat{U}(t+r) &= \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=-\infty}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \\ &\quad - \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \\ &= \sum_{j=0}^q \mathcal{D}_j(t+r)\mathbf{X}(t+r-j) - \sum_{j=0}^q \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \\ &= \mathbf{U}(t+r) - \sum_{j=0}^m \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \quad \text{for } q < r; \end{aligned} \tag{4.13}$$

and (4.12) can be expressed as

$$\hat{U}(t+r) = \mathbf{U}(t+r) - \sum_{j=0}^{r-1} \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \quad \text{for } q \geq r. \tag{4.14}$$

On combining (4.13) and (4.14) we have

$$\hat{U}(t+r) = \mathbf{U}(t+r) - \sum_{j=0}^{q \wedge (r-1)} \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-j} \mathbf{a}_v(t+r-j)\mathbf{Y}(v), \quad r \geq 1,$$

so that the mean square error is

$$\begin{aligned} \varepsilon^2(r) &= tr(E(\mathbf{U}(t+r) - \hat{U}(t+r))(\mathbf{U}(t+r) - \hat{U}(t+r))^*) \\ &= tr \left(E \left(\sum_{j=0}^{q \wedge (r-1)} \mathcal{D}_j(t+r) \sum_{v=t+1}^{t+r-1} \mathbf{a}_v(t+r-j)\mathbf{Y}(v) \right) \right. \\ &\quad \left. \times \left(\sum_{k=0}^{q \wedge (p-1)} \mathcal{D}_k(t+r) \sum_{s=t+1}^{t+r-k} \mathbf{a}_s(t+r-k)\mathbf{Y}(s) \right)^* \right) \\ &= \sum_{j=0}^{q \wedge (r-1)} \sum_{k=0}^{q \wedge (r-1)} \sum_{v=t+1}^{(t+r-j) \wedge (t+r-k)} tr(\mathcal{D}_j(t+r)\mathbf{a}_v(t+r-j)\Xi\mathbf{a}_v^*(t+r-k)\mathcal{D}_k^*(t+r)). \end{aligned}$$

Acknowledgement

The authors are very grateful to the referees for their suggestions for improvement of the paper.

References

- [1] B. D. O. Anderson and J. B. Moore, *Optimal filtering* (Prentice Hall, 1979).
- [2] P. J. Brockwell and R. A. Davis, *Time series: theory and methods* (Springer-Verlag, 1987).
- [3] P. Caines, *Linear stochastic systems* (Wiley, 1988).
- [4] J. Hajek, "On linear statistical problems in stochastic processes", *Czechoslovak Math. J.* **12** (1962) 404–444.
- [5] C. Jordan, *The calculus of finite differences* (Chelsea, 1965).
- [6] G. Melard and A. Herteleer-de-Schutter, "Contributions to evolutionary spectral theory", *J. Time Series Analysis* **10** (1989) 41–63.
- [7] K. S. Miller, "Nonstationary autoregressive processes", *IEEE Trans. Form. Theory* **15** (1969) 315–316.
- [8] E. Parzen, "Statistical inference on time series by Hilbert space methods I", *Technical Report No. 23* (1959). App. Math. & Stat. Laboratory, Stanford University; reprinted in *Time Series Analysis Papers* (Holden Day, 1976).
- [9] M. S. Peiris, "On the prediction of multivariate ARMA processes with a time-dependent covariance structure", *Commun. Statist. Theory Meth.* **17** (1988) 27–37.
- [10] M. B. Priestley, "Evolutionary spectra and non stationary processes", *J. R. Statist. Soc. B.* **27** (1965) 204–237.
- [11] M. B. Priestley, "Parameter adaptive prediction and control", *Technical Report No. 7*, U.M.I.S.T. (1969).
- [12] M. B. Priestley, "Control systems with time-dependent parameters", *Bull. Inst. Internat. Statist.* **37** (1969) 359–361.
- [13] J. Rissanen, "A fast algorithm for optimum linear predictors", *IEEE Trans. Autom. Control* **AC-18** (1973) 555.
- [14] J. Rissanen and L. Barbosa, "Properties of infinite covariance matrices and stability of optimum predictors", *Information Sciences* **1** (1969) 221–236.
- [15] J. Rissanen and L. Barbosa, "A factorization on problem and the problem of predicting nonstationary vector-valued stochastic processes", *Z. Wahrsch. Verw. Geb.* **12** (1969) 255–266.
- [16] Y. A. Rozanov, *Stationary random processes* (Holden-Day, 1967).
- [17] N. Singh and M. S. Peiris, "A note on the properties of some nonstationary ARMA processes", *Stochastic Processes and their Applications* **24** (1987) 151–155.
- [18] N. M. Spencer and V. V. Anh, "Markovian representation and prediction of random sequences with time-dependent rational spectrum", *QUT Working Paper in Mathematics* 12/89.
- [19] P. Whittle, "Recursive relations for predictors", *J. R. Statist. Soc., Ser B* **28** (1965) 523–532.