Matrix valued statistical investigations of the double measurements modell

By J. KORMOS and GY. TERDIK (Debrecen)

Introduction

The study of matrix valued stationary processes was begun by B. GYIRES and T. BALOGH [1], [2]. Our earlier studies aimed at the spectral and moving average representation and the prediction of the nonquadratic matrix valued stationary processes [3], [4]. Following these we come to the statistical questions of the autoregressive matrix valued process [5]. These last results of ours were made actual by a work of T .W. ANDERSON [6] concerned with the double measurement model of time series, as double measurement is the case when we increase the number of columns of the matrix valued process.

In this paper we first point out the relation between matrix and vector valued processes, then we deal with the moving average representation of the matrix valued process and give a statistical examination of the double measurement model. Concerning this last case we want to remark that our method requires a weak condition from the errors, but it does not allow the dependence of coefficients from time.

1. Representation of the matrix valued autoregressive process as an infinite moving average

In this part we deal with the MA (moving average) representation of the AR time series defined by the equation

$$\sum_{j=0}^{p} A(j)x(t-j) = \varepsilon(t) \tag{1}$$

where $A(j)$ is a $n\times n$ scalar matrix
$x(t)$ is a $n\times m$ random matrix
$\varepsilon(t)$ is a $n\times m$ random matrix, and

$$E\varepsilon(t)\varepsilon'(s) = \delta_s^t \Sigma. \tag{2}$$

With the following remarks we would like to point out the difference between the process (1) and the vector valued processes related to it, and the difference between the MA representations in accordance with these processes.

(i) Let us consider m vector valued AR (autoregressive) processes with equal coefficients

$$\sum_{j=0}^{p} A(j)\underline{x}^{(i)}(t-j) = \underline{\varepsilon}^{(i)}(t) \quad (i = 1, 2, \ldots, m) \tag{3}$$

where the $\varepsilon^{(i)}(t)$'s are uncorrelated, that is

$$E\underline{\varepsilon}^{(i)}(t) = \underline{0}, \tag{4}$$

$$E\underline{\varepsilon}^{(i)}(t)\underline{\varepsilon}^{(j)\prime}(s) = \delta_j^i \delta_s^t \Sigma_i \quad (i, j = 1, 2, \ldots, m). \tag{5}$$

Let us construct the matrix valued process $x(t)$ by the formula:

$$x(t) = \big(\underline{x}^{(1)}(t), \ldots, \underline{x}^{(m)}(t)\big). \tag{6}$$

Then the $x(t)$ will be a matrix valued AR process with $A(j)$ coefficients and with the residue of

$$\varepsilon(t) = \big(\underline{\varepsilon}^{(1)}(t), \ldots, \underline{\varepsilon}^{(m)}(t)\big)$$

with

$$E\varepsilon(t)\varepsilon'(s) = \delta_s^t \Sigma, \quad \text{where} \quad \Sigma = \sum_{i=1}^{m} \Sigma_i.$$

(ii) If in the above case we exchange the condition (5) for the weaker

$$E\underline{\varepsilon}^{(i)}(t)\underline{\varepsilon}^{(j)\prime}(s) = \delta_s^t \Sigma_{i,j} \tag{5$'$}$$

the process formed in the manner of (6) will be a matrix valued AR process as well.

(iii) A matrix valued AR process (1) usually cannot be decomposed into vector valued AR processes of type (3) because condition (2) is weaker than (5) or (5)$'$. (It cannot be deduced from the equations

$$\delta_s^t \Sigma = E\varepsilon(t)\varepsilon'(s) = E\Big(\sum_{i=1}^{m} \underline{\varepsilon}^{(i)}(t)\underline{\varepsilon}^{(i)\prime}(s)\Big) = \sum_{i=1}^{m} E\underline{\varepsilon}^{(i)}(t)\underline{\varepsilon}^{(i)\prime}(s)$$

that

$$E\underline{\varepsilon}^{(i)}(t)\underline{\varepsilon}^{(i)\prime}(s) = \delta_s^t \Sigma_i$$

for all i)

In the case of statistical investigations the question of the MA producibility of (1) is of capital importance. This is what we get a favourable answer in the next proposition.

Theorem 1. *Let the* AR *equation* (1) *be given and the condition* (2) *be satisfied. If the equation*

$$\det\Big(\sum_{j=0}^{p} A(j)\,z^{p-j}\Big) = 0 \tag{7}$$

has no root on the unit circle at all, then there exists a single-valued solution of equation (1) *for given initial values* — $x(-p), \ldots, x(-1)$ — *and this solution can be written in the form of*

$$x(t) = H(t)C + \sum_{r=-\infty}^{\infty} \Lambda(r)\varepsilon(t-r) \tag{8}$$

where $H(t)$ is a $n \times np$ polynomial matrix, C is a $np \times n$ scalar matrix, which is defined by the initial values uniquely, $\Lambda(r)$ is a $n \times n$ scalar matrix the elements of which converge to zero exponentially in case of $|r| \to \infty$.

We do not prove the theorem in full detail but we sketch the considerations leading to the verification of the statement.

The solution of the homogeneous equation belonging to (1)

$$\sum_{j=0}^{p} A(j)\,Y(t-j) = (0)_{n\times m}$$

is $Y(t)=H(t)C$ where $H(t)$ is a polynomial matrix built from the column vectors

$$\underline{h}^{(l)}(t) = z_u^t t^j \underline{h}(k, u) \quad \big(l = l(u, j, k)\big)$$

where z_u is the root of the equation (7) and $\underline{h}^{(l)}(t)$ solves the equation

$$\big(\Sigma A(j)\,z^{p-j}\big)\underline{h} = \underline{0}$$

($l=1, \ldots, np$; u, j, k, run through intervals defined uniquely during the construction.)

By the conditions of the theorem the inverse of the polynomial matrix $\sum_{j=0}^{p} A(j)z^j$ can be expanded on any annulus containing the unit circle into a Laurent series which will be denoted by

$$\sum_{t=-\infty}^{\infty} L(t)\,z^t. \tag{9}$$

On the basis of simple considerations it can be realized that all these matrices $L(t)$ $(t=\ldots, -1, 0, 1, \ldots)$ satisfy the condition

$$\sum_{j=0}^{p} A(j)L(t-j) = \delta_0^t I_n$$

and the matrix valued process formed with their help

$$\Psi(t) = \sum_{u=-\infty}^{\infty} L(t-u)\,\varepsilon(t)$$

satisfies the equation (1).

Uniqueness follows from a simple algebraical consideration. $\Lambda(r)$ can be gained by index transformation from $L(t)$. The point under discussion in Theorem 1. is the convergence in mean square of matrix sequences. Here we mean the adequate convergence interpreted in Quasi—Hilbert space B. GYIRES [1]. At the same time in our case one also has componentwise convergence in mean square because of the role of coefficient matrices played in (9), and it is of basic importance during the statistical investigations. If the roots of (7) are within the unit circle we gain the solution

$$x(t) = \sum_{r=0}^{\infty} \Lambda(r)\,\varepsilon(t-r). \tag{10}$$

This is what we call the MA representation of the process, which is the stationary solution of the equation (1) on the basis of (2).

2. Estimating the matrix valued AR process

Let us consider the equation (1) examined in the previous point for $p=1$, and let us deal with the first order matrix valued AR process

$$x(t)+Ax(t-1)=\varepsilon(t). \tag{11}$$

We can simply realize that this is no restriction of generality. Namely in case of a known order $p>1$ the equation (1) can be transformed into a first order equation of higher dimension depending on p so that the conditions of the MA representation do not change (T. W. ANDERSON [7]), and $\Lambda^p(r)$, $\Lambda^1(r)$ respectively are defined by the same matrices $A(j)$.

Our purpose is to show the statistics which can be constructed for A and for Σ and which can be considered as generalizations of the vector case to have good features.

Let the basic equation of the estimation $\hat{A}$ related to A be

$$\frac{1}{T}\sum_{s=1}^{T} x(s-1)x'(s-1)\hat{A}' = -\frac{1}{T}\sum_{s=1}^{T} x(s-1)\,x'(s), \tag{12}$$

while as a consequence of this the estimation $\hat{\Sigma}$ arises in a natural way in the form of

$$\hat{\Sigma}=\frac{1}{T}\sum_{s=1}^{T}\big(x(s)+Ax(s-1)\big)\big(x(s)+Ax(s-1)\big)'. \tag{13}$$

The previous estimations can be gained by the least squares method or, requiring the normality of $\varepsilon(t)$'s, by the help of the maximum likelihood principle.

We remark that in the sequel by the distribution of a matrix valued random variable $x=(\underline{x}^{(1)},\dots,\underline{x}^{(m)})$ we mean the distribution of the vector $x'=(\underline{x}^{(1)'},\dots,\underline{x}^{(m)'})$ obtainable by "stretching" the columns of the matrix. The interpretation is justified in accordance with the remarks (i)—(iii) made at the beginning of point 1.

First of all we demonstrate that under adequate conditions the estimations (12)—(13) are consistent. Let us consider the MA representation belonging to (11)

$$x(t)=\sum_{r=0}^{\infty}(-A)^r\,\varepsilon(t-r) \tag{14}$$

and let the subsequent conditions be satisfied

$$E\varepsilon(t)=(0)_{n\times m};\ \ E\varepsilon(t)\varepsilon'(t)=\Sigma. \tag{15}$$

Moreover let the $\varepsilon(t)$-s be independent and identically distributed.

In that case

$$\frac{1}{T}\sum_{r=1}^{T}\varepsilon(r)\varepsilon'(r)\xrightarrow[T\to\infty]{st}\Sigma, \tag{16}$$

$$\frac{1}{T}\sum_{r=1}^{T}\varepsilon(r)x'(r-1)\xrightarrow[T\to\infty]{st}(0)_{n\times n}, \tag{17}$$

$$\frac{1}{T}\sum_{r=1}^{T}x(r)\,x'(r)\xrightarrow[T\to\infty]{st}G \quad \text{where} \quad G=Ex(t)\,x'(t). \tag{18}$$

These statements can be verified by adapting the conditions neccessary to the representation (14) and theorems, inequalities which can be ranked among the classical laws of large numbers. The complete proof can be found in J. KORMOS [5].

Theorem 2. *Let the conditions* (14)—(15) *be satisfied and let* Σ *be positive definite. Then*

$$\hat{A} \xrightarrow[T\to\infty]{st} A, \tag{19}$$

$$\hat{\Sigma} \xrightarrow[T\to\infty]{st} \Sigma. \tag{20}$$

To establish the first part of the theorem let us consider the equation sequence

$$\hat{A}-A = \tag{21}$$

$$= \left(\frac{1}{T}\sum_{s=1}^{T} x(s-1)\,x'(s-1)\right)^{-1}\left(-\frac{1}{T}\sum_{s=1}^{T} x(s-1)x'(s)-\frac{1}{T}\sum_{s=1}^{T} x(s-1)\,x'(s-1)\,A'\right) =$$

$$= -\left(\frac{1}{T}\sum_{s=1}^{T} x(s-1)\,x'(s-1)\right)^{-1}\frac{1}{T}\sum_{s=1}^{T} x(s-1)\,\varepsilon'(s).$$

Following this we obtain (19) by using (16)—(18). The justification of the second part, that is of the consistency of the estimation $\hat{\Sigma}$, arises from the consistency of the estimation $\hat{A}$. Next we examine the asymptotic distribution of the estimation $\hat{A}$. Similarly to the vector case we can verify (J. KORMOS [5]):

Theorem 3. *Let condition* (14)—(15) *be satisfied and let* G *be positive definite. In this case the limit distribution of* $\sqrt{T}(\hat{A}'-A')$ *is normal with zero mean and with covariance matrix* $G^{-1}\otimes\Sigma$.

3. Estimation of the parameters of the double measurements model

Let us consider the first order vector valued AR process

$$\underline{x}(t)+A\underline{x}(t-1) = \underline{\varepsilon}(t) \tag{24}$$

the parameters of which we would like to estimate in case of so called double measurements.

Let the sample of elements

$$x_m = \left(\underline{x}_m(1), \dots \underline{x}_m(T)\right) \quad m = 1, \dots M$$

be given with respect to the process (24) and let us assume that we can increase the number of measurements M in one moment over all limits instead of the usual increasing of the number of sample elements T. In this way we examine simultaneously samples of numbers M having elements of numbers T individually related to the process (24), or in another way with respect to the process

$$x(t)+Ax(t-1) = \varepsilon(t),$$

where

$$x(t) = (\underline{x}_1(t), \ldots, \underline{x}_m(t)).$$

Suppose that

$$E\underline{\varepsilon}_m(t) = \underline{0}, \tag{25}$$

$$E\underline{\varepsilon}_{m_1}(t_1)\,\underline{\varepsilon}_{m_2}(t_2) = \begin{cases} \delta_{t_2}^{t_1}\,\Sigma_{|m_1-m_2|}, & |m_1-m_2| \leqq l, \\ (0)_p, & |m_1-m_2| > l \end{cases} \tag{26}$$

for all m, m_1, m_2, t, t_1, t_2. The latter essentially means that the errors of the observations are "l-correlated".

This implies

$$E\underline{x}_{m_1}(t_1)\,\underline{x}'_{m_2}(t_2) = \begin{cases} C_{|m_1-m_2|,\,|t_1-t_2|} = \sum_{s=0}^{\infty} (-A)^s\,\Sigma_{|m_1-m_2|}\,(-A')^{|t_1-t_2|+s}, & |m_1-m_2| \leqq l \\ (0)_p, & |m_1-m_2| > l. \end{cases}$$

As a consequence the basic equations of the estimations concerning the matrices Σ_0 and A are

$$\frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} \underline{x}_m(t-1)\,\underline{x}'_m(t-1)\,\hat{A}' = -\frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} \underline{x}_m(t-1)\,\underline{x}'_m(t), \tag{27}$$

$$\hat{\Sigma} = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} \big(\underline{x}_m(t)+\hat{A}\underline{x}_m(t-1)\big)\big(\underline{x}_m(t)+\hat{A}\underline{x}_m(t-1)\big)'. \tag{28}$$

The good features of these estimations for fixed m are known (T. W. ANDERSON [6], [7]). First we show that they are consistent for fixed t also. To realize this it is enough to justify the following auxiliary statement.

Auxiliary statement. *Let the condition*

$$\frac{1}{M}\sum_{m=1}^{M} \underline{\varepsilon}_m(t)\,\underline{\varepsilon}'_m(t) \xrightarrow[M\to\infty]{st} \Sigma_0 \tag{29}$$

be satisfied. Then

$$\frac{1}{M}\sum_{m=1}^{M} \underline{x}_m(t)\,\underline{x}'_m(t) \xrightarrow[M\to\infty]{st} C_{0,0}, \tag{30}$$

$$\frac{1}{M}\sum_{m=1}^{M} \underline{x}_m(t)\,\underline{x}'_m(t-1) \xrightarrow[M\to\infty]{st} C_{0,1}. \tag{31}$$

PROOF. Let

$$\underline{x}_{m,R}(t) = \sum_{k=0}^{R} A^k\,\underline{\varepsilon}_m(t-k),$$

so

$$\frac{1}{M}\sum_{m=1}^{M}\Big(\underline{x}_m(t)\,\underline{x}'_m(t) - \sum_{k=0}^{\infty} A^k\Sigma_0 A'^k\Big) = \frac{1}{M}\sum_{m=1}^{M} \underline{x}_m(t)\,\underline{x}'_m(t) - C_{0,0} =$$

$$= \frac{1}{M}\sum_{m=1}^{M}\Big(\underline{x}_{m,R}(t)\,\underline{x}'_{m,R}(t) - \sum_{k=0}^{\infty} A^k\Sigma_0 A'^k\Big) + \frac{1}{M}\sum_{m=1}^{M}\big(\underline{x}_m(t)\,\underline{x}'_m(t) - \underline{x}_{m,R}(t)\,\underline{x}'_{m,R}(t)\big).$$

The first member is

$$\sum_{k,l=0}^{R} A^k \left(\frac{1}{M}\sum_{m=1}^{M} \underline{\varepsilon}_m(t-h)\underline{\varepsilon}'_m(t-l)-\delta_l^k\Sigma_0\right)A'^l - \sum_{k=R+1}^{\infty} A^k\Sigma_0 A'^k. \tag{32}$$

By conditions (26) and (29) the first part of (32) converges stochastically to zero in case of $M\to\infty$ for every R. So in case of $R\to\infty$ and $M\to\infty$ (32) converges stochastically to $(0)_p$. In case of the second member let us take into account consider that

$$\lim_{R\to\infty}\sup_{m,t} E\left|\underline{x}'_m(t)\underline{x}_m(t)-\underline{x}'_{m,R}(t)\underline{x}_{m,R}(t)\right| = 0.$$

Utilizing this

$$E\left|\underline{x}_m(t)\underline{x}'_m(t)-\underline{x}_{m,R}(t)\underline{x}'_{m,R}(t)\right| =$$

$$= E\left|\left(\underline{x}_m(t)-\underline{x}_{m,R}(t)\right)\underline{x}'_{m,R}(t)+\underline{x}_m(t)\left(\underline{x}_m(t)-\underline{x}_{m,R}(t)\right)'\right| \leqq$$

$$\leqq (C_{0,0})^{1/2}\Big(\sum_{k=R+1}^{\infty} A^k\Sigma_0 A'^k\Big)^{1/2}+\Big(\sum_{k=R+1}^{\infty} A^k\Sigma_0 A'^k\Big)^{1/2}(C_{0,0})^{1/2}.$$

After adapting so the Tchebicheff inequality the second member converges to $(0)_p$ in case of $R\to\infty$, $M\to\infty$. The justification of (31) can be performed in a similar way.

Theorem 4. *Let the characteristic roots of $-A$ be within the unit circle, and let (29) be sufficed. If C_0 is positive definite then*

$$\hat{A}\xrightarrow[M\to\infty]{st} A, \tag{33}$$

$$\hat{\Sigma}\xrightarrow[M\to\infty]{st} \Sigma_0. \tag{34}$$

Proof. Adapting the transformation corresponding to (21) the equation

$$C_{0,1}+AC_{0,0}=(0)_p \tag{35}$$

which follows from the Yule—Walker equations belonging to (24), and the auxiliary statement lead to (33). The justification of (34) can be performed simply by using (33) and (29). Now it is not our purpose to examine to which simpler more elementary conditions (29) can be traced back, we just remark that under the conditions (25), (26), (29) is ensured by the existence of the fourth moments of the elements $\varepsilon_m^i(t)$ $(i=1,\dots,p)$ and by the stationary characteristic in "nearly" fourth order moments of these elements.

Next we show that the limit distribution of the random matrix $\sqrt{M}(\hat{A}'-A')$ is normal. In the sense of the aforesaid it is enough to show that the limit distribution of an arbitrary linear combination of the elements of the matrix

$$z(m)=\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\underline{x}_m(t-1)\underline{\varepsilon}'_m(t) \tag{36}$$

is normal. Let (a) the condition (25) be satisfied, (b) the sequence

$$\underline{\varepsilon}_1(t),\ \underline{\varepsilon}_2(t),\ \dots,\ \underline{\varepsilon}_m(t),\ \dots \tag{37}$$

be l-dependent for every fixed t.

(c) $$\underline{\varepsilon}_m(1), \ldots, \underline{\varepsilon}_m(t), \ldots$$

independent for every fixed m (d) the fourth moment of $\varepsilon_m^i(t)$ $(i=1, \ldots, p)$ exist. Let us remark that

$$\mu(m) = \underline{\varepsilon}'_m(t) Bx_m(t-1). \tag{38}$$

It can be shown that

$$E\mu(m) = 0, \quad \text{that is the mean of} \quad \mu(m) \text{ is} \tag{39}$$

independent of m. According to (37) $\operatorname{cov}(\mu(m_1), \mu(m_2))$ does exist, and

$$\operatorname{cov}(\mu(m_1), \mu(m_2)) = E\underline{\varepsilon}'_{m_1}(t) B\underline{x}_{m_1}(t-1)\, \underline{\varepsilon}'_{m_2}(t) B\underline{x}_{m_2}(t-1). \tag{40}$$

Because of the previous conditions (37)

$$\operatorname{cov}(\mu(m_1), \mu(m_2)) = \operatorname{tr}(C_{|m_1-m_2|} B' \Sigma_{|m_1-m_2|} B). \tag{41}$$

The subsequent limit distribution theorem concerning the dependent random variables is of primary importance. (P. Billingsley [9].)

Auxiliary statement. If ξ_n *is a* φ*-mixing sequence and* $\sum_n \varphi_n^{1/2} < \infty$; $E\xi_0 = 0$ *and moreover* ξ_0 *has finite variance, then the series*

$$\sigma^2 = E\xi_0^2 + 2\sum_{k=1}^{\infty} E\xi_0 \xi_k$$

is absolutely convergent, and if $\sigma^2 > 0$ *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \xrightarrow[n\to\infty]{D} N(0, \sigma^2).$$

We prove that the sequence $\mu(m)$ satisfies the conditions of the above-mentioned statement. As a consequence of (37) the sequence $\mu(m)$ is l-dependent and so φ-mixing. Following from the interpretation of $\mu(m)$ and from (42) $E\mu(0)$ and $E\mu^2(0)$ do exist. We have proved that the limiting distribution of $\frac{1}{\sqrt{M}} \sum_{m=1}^{M} \mu(m)$ is $\mathcal{N}(0, \sigma^2)$ where

$$\sigma^2 = \sum_{k=-l}^{l} \operatorname{tr}(C_{|k|} B' \Sigma_{|k|} B).$$

That is

$$\sigma^2 = \operatorname{tr}(C_0 B' \Sigma_0 B) + 2\sum_{k=1}^{l} \operatorname{tr}(C_k B' \Sigma_k B).$$

On the basis of all this the following theorem is true:

Theorem 5. *Let us consider the first order vector valued* AR *process* (24) *and let the conditions* (37) (a)—(d) *be satisfied. Then the estimation* (27) *concerning the coefficient matrix* A *is asymptotically normal, that is the limit distribution of* $\sqrt{M}(\hat{A}' - A')$ *is normal.*

Taking into consideration that

$$\tilde{A}=\frac{1}{T}\sum_{t=1}^{T}\hat{A}(t)$$

where $\hat{A}(t)$ satisfies the equation

$$\sum_{m=1}^{M}\underset{\sim}{x}_m(t-1)\,\underset{\sim}{x}'_m(t-1)\hat{A}'(t)=\sum_{m=1}^{M}\underset{\sim}{x}_m(t)\,\underset{\sim}{x}'_m(t)$$

the following theorem is true:

Theorem 6. *Under the assumptions of theorem 5.* $\sqrt{MT}(\tilde{A}'-A')$ *is asymptotically normal with zero mean and covariance matrix*

$$C_0\otimes\Sigma_0+2\sum_{k=1}^{l}C_0^{-1}C_kC_0^{-1}\otimes\Sigma_k.$$

For the proof of Theorem 6. we point out only that $\hat{A}(t)$ and $\hat{A}(s)$ are asymptotically independent when $s\neq t$. We remark that the last theorem is very useful for statistical investigations concerning the double measurements model.

Bibliography

[1] B. Gyires, Über die Spuren der verallgemeinerten Toeplitzschen Matrizes, *Publ. Math. (Debrecen)* **8** (1961), p. 93—116.
[2] T. Balogh, Matrix-Valued Stochastical Processes, *Publ. Math. (Debrecen)* **8** (1961), p. 369—377.
[3] J. Kormos, G. Terdik, Notes on matrix-valued stationary stochastic processes I., *Publ. Math. (Debrecen)*, **25** (1978), p. 281—283.
[4] G. Terdik, Notes on matrix-valued stochastic processes II., **26** (1979), p. 123—126.
[5] J. Kormos, Remarks on multiple autoregressive time series (1980), *(unpublished dissertation)*.
[6] T. W. Anderson, The statistical analysis of time series, *John Wiley and Sons, New York* (1971).
[7] T. W. Anderson, Repeated measurements on autoregressive processes, *J. Amer. Statist. Ass.* **73**. (1978), p. 371—378.
[8] E. J. Hannan, Multiple time series, *John Wiley and Sons, New York* (1970).
[9] P. Billingsley, Convergence of probability measures, *John Wiley and Sons, New York* (1968).

(Received May 5, 1983)