Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Autoregressive moving average model for matrix time series

Shujin Wu & Ping Bi

To cite this article: Shujin Wu & Ping Bi (2023) Autoregressive moving average model for matrix time series, Statistical Theory and Related Fields, 7:4, 318-335, DOI: 10.1080/24754269.2023.2262360

To link to this article: https://doi.org/10.1080/24754269.2023.2262360

Published online: 13 Dec 2023.

STATISTICAL THEORY AND RELATED FIELDS
2023, VOL. 7, NO. 4, 318–335
https://doi.org/10.1080/24754269.2023.2262360

Autoregressive moving average model for matrix time series

Shujin Wu [a,b] and Ping Bi [c]

[a]College of Liberal Arts and Sciences, China University of Petroleum-Beijing at Karamay, Karamay, People's Republic of China; [b]KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China; [c]School of Mathematical Sciences, Key Laboratory of Pure Mathematics and Mathematical Practice, East China Normal University, Shanghai, People's Republic of China

ABSTRACT

In the paper, the autoregressive moving average model for matrix time series (MARMA) is investigated. The properties of the MARMA model are investigated by using the conditional least square estimation, the conditional maximum likelihood estimation, the projection theorem in Hilbert space and the decomposition technique of time series, which include necessary and sufficient conditions for stationarity and invertibility, model parameter estimation, model testing and model forecasting.

ARTICLE HISTORY
Received 8 December 2022
Revised 10 August 2023
Accepted 18 September 2023

KEYWORDS
Matrix time series; autoregressive moving average model; bilinear model; statistical inference

1. Introduction

Matrix time series is a time series whose cross-sectional data are matrices, which can be found in a variety of fields such as economics, business, ecology, psychology, meteorology, biology and fMRI (Samadi, 2014). For example, consider two stocks, A_1 and A_2, as potential investment products, whose prices and volumes are selected as two analysis factors. Denote the price and volume of stock A_k at time t by $P_k(t)$ and $V_k(t)$, $k = 1, 2$, and then a 2×2-dimensional matrix time series can be constructed as follows:

$$\left\{ X_t \equiv \begin{bmatrix} P_1(t) & P_2(t) \\ V_1(t) & V_2(t) \end{bmatrix}, \quad t = 1, 2, \ldots \right\}.$$

Matrix time series has attracted a few scholars' attention and research at the beginning of the century. Walden and Serroukh (2002) studied the construction of matrix-valued filters for multi-resolution analysis of matrix time series. Samadi (2014) brought forward and investigated a p-order autoregressive model for matrix time series, which is essentially a VAR(p) model in matrix form. D. Wang et al. (2019) proposed a novel factor model

$$X_t = RF_tC^\top + \varepsilon_t, \quad t = 1, 2, \ldots,$$

where X_t and F_t are matrix time series. Chen et al. (2021) first proposed one-order autoregressive model for matrix time series in the bilinear form, denoted by MAR(1),

$$X_t = AX_{t-1}B^\top + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{1}$$

and investigated its stationarity, causality, method of parameter estimation, and asymptotics of statistic. Wu and Hua (2022) independently proposed the p-order autoregressive model for matrix time series in the bilinear form, denoted by MAR(p),

$$X_t = \sum_{k=1}^{p} A_k X_{t-k} B_k^\top + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{2}$$

and presented parameter estimation, model identification criterion and model checking. For more literature studies on matrix time series, one can refer to H. Wang and West (2009), Zhou et al. (2018), Getmanov et al. (2021) and their references.

It is widely known that the autoregressive moving average model of time series (ARMA) plays a very important role in the theory and the application of one-dimensional time series, and we will show later that a bilinear model has its unique advantages for matrix time series. In the paper, autoregressive moving average models for matrix

CONTACT Shujin Wu sjwu@stat.ecnu.edu.cn College of Liberal Arts and Sciences, China University of Petroleum-Beijing at Karamay, Karamay 834000, People's Republic of China; KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China

time series (MARMA) are first proposed and investigated. Necessary and sufficient conditions for stationarity of MARMA are provided, and parameter estimations are also considered by the conditional least squares method and the conditional maximum likelihood estimation method. At last, an example is presented to show the applications of the MARMA model.

2. Preliminaries

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a σ-filtration $\{\mathcal{F}_t, t \in \mathbb{N}\}$ in which the second moment of each variable exists, and $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Definition 2.1: For any given positive integers m and n, an $m \times n$-dimensional matrix time series refers to

$$X = \left\{ \left(X_{ij}(t)\right)_{m\times n}, \quad t \in \mathbb{N} \right\}, \tag{3}$$

where $\{X_{ij}(t), t \in \mathbb{N}\}$ is a one-dimensional time series on a probability space $(\Omega, \mathcal{F}, P)$ for any $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

Definition 2.2: Let $X = \{X(t), t \in \mathbb{N}\}$ be an $m \times n$-dimensional matrix time series defined by (3), and then its mean function follows as

$$\mu_X(t) \equiv E[X(t)] = \left(E[X_{ij}(t)]\right)_{m\times n}, \quad t \in \mathbb{N}. \tag{4}$$

Additionally, its autocovariance function follows as

$$\Gamma_X(t,s) \equiv \Gamma_{\mathrm{vec}(X)}(t,s) = \left(\sigma_{ij,k\ell}(t,s)\right)_{mn\times mn}, \tag{5}$$

where $\sigma_{ij,k\ell}(t,s) = \mathrm{cov}(X_{ij}(t), X_{k\ell}(s))$, $i,k = 1,2,\ldots,n$ and $j,\ell = 1,2,\ldots,n$; $t,s \in \mathbb{N}$, and $\mathrm{vec}(X(t))$ is the vectorization of $X(t)$ by columns, that is,

$$\mathrm{vec}(X(t)) = [X_{11}(t), X_{21}(t), \ldots, X_{m1}(t), X_{12}(t), X_{22}(t), \ldots, X_{mn}(t)]^\top. \tag{6}$$

Stationarity and matrix white noise play a very important role on time series analysis. Thus, we will introduce the concept of stationary matrix time series and matrix white noise in the following.

Definition 2.3: Let $\{X(t), t \in \mathbb{N}\}$ be a matrix time series defined by (3) and $\mathrm{vec}(X(t))$ be the vectorization of $X(t)$ defined by (6). Then $\{X(t), t \in \mathbb{N}\}$ is a stationary matrix time series if and only if $\{\mathrm{vec}(X(t)), t \in \mathbb{N}\}$ is stationary.

Definition 2.4: For any given positive integers m and n, denote an $m \times n$-dimensional matrix time series $\varepsilon = \{(\varepsilon_{ij}(t))_{m\times n}, \ t \in \mathbb{N}\}$, and then ε is called an $m \times n$-dimensional matrix white noise, if it satisfies the following conditions.

(1) Its mean function $\mu_\varepsilon(t) = O_{m\times n}$ for all $t \in \mathbb{N}$, where $O_{m\times n}$ is the $m \times n$-dimensional zero matrix.
(2) Its autocovariance function $\Gamma_\varepsilon(t,s)$ defined by Definition 2.3 satisfies that

$$\Gamma_\varepsilon(t,s) = \begin{cases} O_{mn}, & t \neq s, \\ \Sigma_{mn}, & t = s, \end{cases} \qquad \forall t,s \in \mathbb{N},$$

where O_{mn} is the $mn \times mn$-dimensional zero matrix, and

$$\Sigma_{mn} = \mathrm{diag}\left(\sigma_{11}^2, \sigma_{21}^2, \ldots, \sigma_{m1}^2, \sigma_{12}^2, \sigma_{22}^2, \ldots, \sigma_{(m-1)n}^2, \sigma_{mn}^2\right) \tag{7}$$

is the $mn \times mn$-dimensional diagonal matrix with diagonal entries $\sigma_{11}^2, \sigma_{21}^2, \ldots, \sigma_{m1}^2, \sigma_{12}^2, \sigma_{22}^2, \ldots, \sigma_{(m-1)n}^2, \sigma_{mn}^2$.

For any matrix white noise $\{\varepsilon(t), t \in \mathbb{N}\}$, if its vectorization by columns, $\{\mathrm{vec}(\varepsilon(t)), t \in \mathbb{N}\}$, is Gaussian, then $\{\varepsilon(t), t \in \mathbb{N}\}$ is called a matrix Gaussian white noise.

Property 2.1: For any $m \times n$-dimensional matrix time series $\{\varepsilon(t), t \in \mathbb{N}\}$, it is an $m \times n$-dimensional matrix white noise if and only if $\{\mathrm{vec}(\varepsilon(t)), t \in \mathbb{N}\}$ is an mn-dimensional vector white noise, where $\mathbb{N} = \{1, 2, 3, \ldots\}$.

The proof of Property 2.1 is not difficult, so we omit it.

When we investigate the autoregressive moving average model for matrix time series, we may use the Kronecker product, matrix reshape and derivative of matrix. Thus, we introduce them in the following.

Definition 2.5 (Graham, 2018): Assume matrices $A = (a_{ij})_{m\times n}$ and $C = (c_{ij})_{p\times q}$, and then the $m \times n$ block matrix $(a_{ij}C)_{m\times n}$ is called the Kronecker product of A and C, denoted by $A \otimes C$, that is,

$$A \otimes C = \begin{bmatrix} a_{11}C & a_{12}C & \cdots & a_{1n}C \\ a_{21}C & a_{22}C & \cdots & a_{2n}C \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}C & a_{m2}C & \cdots & a_{mn}C \end{bmatrix}.$$

Definition 2.6: For any $A = (a_{ij})_{m\times n}$ and positive integers p, q satisfying $pq = mn$, the (p, q)-order reshaped matrix of A, denoted by $\mathrm{Res}(A, p, q)$, is defined by

$$\mathrm{Res}(A, p, q) = \begin{bmatrix} a_1 & a_{p+1} & \cdots & a_{(p-1)q+1} \\ a_2 & a_{p+2} & \cdots & a_{(p-1)q+2} \\ \vdots & \vdots & \ddots & \vdots \\ a_p & a_{2p} & \cdots & a_{pq} \end{bmatrix},$$

where $a_k = a_{ij}$ for all $k = 1, 2, \ldots, pq$, $i = k - m[(k-1)/m]$ and $j = [(k-1)/m] + 1$, where $[\cdot]$ is the operator of taking the integer part.

Definition 2.7 (Graham, 2018): Let $F = (F_{ij})_{m\times n}$ and $X = (X_{ij})_{p\times q}$ be two matrices, where m, n, p and q are natural numbers. The derivative of matrix F with respect to matrix X is defined by

$$\frac{\partial F}{\partial X} = \begin{bmatrix} \frac{\partial F}{\partial X_{11}} & \frac{\partial F}{\partial X_{12}} & \cdots & \frac{\partial F}{\partial X_{1q}} \\ \frac{\partial F}{\partial X_{21}} & \frac{\partial F}{\partial X_{22}} & \cdots & \frac{\partial F}{\partial X_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F}{\partial X_{p1}} & \frac{\partial F}{\partial X_{p2}} & \cdots & \frac{\partial F}{\partial X_{pq}} \end{bmatrix},$$

where the derivative of matrix F with respect to scalar X_{ij} is defined by

$$\frac{\partial F}{\partial X_{ij}} = \begin{bmatrix} \frac{\partial F_{11}}{\partial X_{ij}} & \frac{\partial F_{12}}{\partial X_{ij}} & \cdots & \frac{\partial F_{1n}}{\partial X_{ij}} \\ \frac{\partial F_{21}}{\partial X_{ij}} & \frac{\partial F_{22}}{\partial X_{ij}} & \cdots & \frac{\partial F_{2n}}{\partial X_{ij}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_{m1}}{\partial X_{ij}} & \frac{\partial F_{m2}}{\partial X_{ij}} & \cdots & \frac{\partial F_{mn}}{\partial X_{ij}} \end{bmatrix}, \quad \begin{array}{l} i = 1, 2, \ldots, p, \\ j = 1, 2, \ldots, q. \end{array}$$

For the derivative of matrix with respect to matrix, its product rule and two common formulas follow as Properties 2.2 and 2.3.

Property 2.2 (Graham, 2018): For any $X = (x_{ij})_{m\times n}$, $Y = (y_{ij})_{n\times u}$ and $Z = (z_{ij})_{p\times q}$, it follows that

$$\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z}(I_q \otimes Y) + (I_p \otimes X)\frac{\partial Y}{\partial Z},$$

where I_q is the $q \times q$-dimensional identity matrix.

Taking $Y = (y_{ij})_{n\times 1}$ and $X = Y^\top$ into Property 2.2, we obtain Corollary 2.1.

Corollary 2.1: *For any $Y = (y_{ij})_{n\times 1}$ and $Z = (z_{ij})_{p\times q}$, it follows that*

$$\frac{\partial (Y^\top Y)}{\partial Z} = 2\frac{\partial Y^\top}{\partial Z}(I_q \otimes Y).$$

Property 2.3 (Graham, 2018): For any $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{n\times u}$ and invertible $X = (X_{ij})_{n\times n}$, it follows that

$$\frac{\partial \mathrm{vec}(AX^{-1}B)}{\partial \mathrm{vec}(X)} = -(X^{-1}B)\otimes(AX^{-1})^\top$$

and

$$\frac{\partial \ln(X)}{\partial X} = (X^{-1})^\top.$$

Taking $B = (b_{ij})_{n\times 1}$ and $A = B^\top$ into Property 2.3, we obtain Corollary 2.2.

Corollary 2.2: *For any $B = (b_{ij})_{n\times 1}$ and invertible $X = (X_{ij})_{n\times n}$, it follows that*

$$\frac{\partial B^\top X^{-1}B}{\partial X} = -\mathrm{Res}\left((X^{-1}B)\otimes(AX^{-1})^\top, n, n\right).$$

3. Autoregressive moving average model for matrix time series

The autoregressive moving average model for matrix time series is an extension of the vector autoregressive moving average model (VARMA) to matrix time series. However, we cannot build the autoregressive moving average model for matrix time series like the VARMA model as follows:

$$X(t) = \Phi_0 + \sum_{i=1}^{p}\Phi_i X(t-i) + \varepsilon(t) - \sum_{j=1}^{q}\Psi_j\varepsilon(t-j). \tag{8}$$

The reason is that the form of (8) cannot describe the dependent relation between the different columns of $X(t)$ according to the rule of matrix multiplication. That is, the ℓth column of $X(t)$ will not be affected by the sth column of $X(t-1), X(t-2), \ldots, X(t-p)$ as $s \neq \ell$.

3.1. MARMA (*p*, *q*) model

In this section, an autoregressive moving average model for matrix time series (MARMA) is first brought forward, whose degradation model, autoregressive model for matrix-valued time series (MAR), is just the model (2) proposed by Wu and Hua (2022) and the extension of model (1) proposed by Chen et al. (2021).

Definition 3.1: Let $\{X(t), t\in\mathbb{N}\}$ be an $m\times n$-dimensional matrix time series. If X is stationary and for each $t\in\mathbb{N}$ it follows that

$$X(t) = C + \sum_{k=1}^{p}\Phi_k X(t-k)\Psi_k + \varepsilon(t) - \sum_{j=1}^{q}\Theta_j\varepsilon(t-j)\Xi_j, \tag{9}$$

where C is an $m\times n$-dimensional matrix; Φ_k and Θ_j are $m\times m$-dimensional matrices, and Ψ_k and Ξ_j are $n\times n$-dimensional matrices for each $k = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$, where p and q are two nonnegative integers; $\{\varepsilon(t), t\in\mathbb{N}\}$ is an $m\times n$-dimensional matrix white noise satisfying that $\mathrm{vec}(\varepsilon(t))$ is independent with $\mathrm{vec}(X(s))$ for all $s < t$, and then $\{X(t), t\in\mathbb{N}\}$ is said to follow a (p, q)-order autoregressive moving average model for matrix time series, denoted by MARMA(p, q).

When $q = 0$, MARMA$(p, 0)$ model (9) degenerates into the form

$$X(t) = C + \sum_{k=1}^{p}\Phi_k X(t-k)\Psi_k + \varepsilon(t), \tag{10}$$

which is a p-order autoregressive model for matrix time series, MAR(p).

When $p = 0$, MARMA$(0, q)$ model (9) degenerates into the form

$$X(t) = C + \varepsilon(t) - \sum_{j=1}^{q}\Theta_j\varepsilon(t-j)\Xi_j, \tag{11}$$

which is called a q-order moving average model for matrix time series, denoted by MMA(q).

If $X = \{X(t), t \in \mathbb{N}\}$ is an $m \times n$-dimensional matrix time series defined by (3) and X is stationary, denote

$$\mu = E[X(t)], \quad \forall t \in \mathbb{N},$$

and then it follows from MARMA(p, q) model (9) that

$$\mu = C + \Phi_1 \mu \Psi_1 + \Phi_2 \mu \Psi_2 + \cdots + \Phi_p \mu \Psi_p. \tag{12}$$

Denote

$$Y(t) = X(t) - \mu.$$

It yields from (12) and MARMA(p, q) model (9) that

$$Y(t) = \sum_{k=1}^{p} \Phi_k Y(t-k) \Psi_k + \varepsilon(t) - \sum_{j=1}^{q} \Theta_j \varepsilon(t-j) \Xi_j \tag{13}$$

holds for all $t \in \mathbb{N}$, and then $Y = \{Y(t), t \in \mathbb{N}\}$ is said to follow a (p, q)-order centralized MARMA(p, q) model.

Because every MARMA(p, q) model can be changed into a centralized MARMA(p, q) model and they have the same coefficient parameters. Thus, while estimating coefficient parameters of MARMA(p, q) model (9) we will mainly study centralized MARMA(p, q) model (13).

For any MARMA(p, q) model (9), and for any $c_k \neq 0$ and $d_j \neq 0$, $k = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$, it follows that

$$X(t) = C + \sum_{k=1}^{p} (c_k \Phi_k) X(t-k) \left(\frac{1}{c_k} \Psi_k \right) + \varepsilon(t) - \sum_{j=1}^{q} (d_j \Theta_j) \varepsilon(t-j) \left(\frac{1}{d_j} \Xi_j \right),$$

that is, coefficient parameters of MARMA(p, q) model (9) are not unique! Thus, we present constraint conditions that

$$\Psi_k = (\psi_{uv})_{n \times n} \quad \text{satisfies} \quad \arg\max_{\psi_{ij}} \{|\psi_{ij}|,\ i, j = 1, 2, \ldots, n\} = 1 \tag{14}$$

and

$$\Xi_j = (\xi_{uv})_{n \times n} \quad \text{satisfies} \quad \arg\max_{\xi_{ij}} \{|\xi_{ij}|,\ i, j = 1, 2, \ldots, n\} = 1 \tag{15}$$

for all $k = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$.

3.2. Relationship between MARMA model and VARMA model

When the column number of matrix $X(t)$ equals one, i.e., $n = 1$, MARMA(p, q) model (9) degenerates into a (p, q)-order vector autoregressive moving average model, VARMA(p, q), as follows:

$$X(t) = C + \sum_{k=1}^{p} \Phi_k X(t-k) + \varepsilon(t) - \sum_{j=1}^{q} \Theta_j \varepsilon(t-j), \tag{16}$$

where $\{X(t), t \in \mathbb{N}\}$ is an m-dimensional vector time series, C is an m-dimensional vector, Φ_k and Θ_j are $m \times m$-dimensional matrices for all $k = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$, and $\{\varepsilon(t), t \in \mathbb{N}\}$ is a white noise of the m-dimensional vector time series satisfying that $\varepsilon(t)$ is independent with $X(s)$ for all $s < t$. Obviously, VARMA model (16) is a special case of MARMA model (9).

On the other hand, for any $m \times n$-dimensional matrix time series $\{X(t), t \in \mathbb{N}\}$, its vectorization $\{\text{vec}(X(t)), t \in \mathbb{N}\}$ is an $mn \times 1$-dimensional time series, and the (p, q)-order vector autoregressive moving average model VARMA(p, q) for $\{\text{vec}(X(t)), t \in \mathbb{N}\}$ follows as

$$\text{vec}(X(t)) = A_0 + \sum_{k=1}^{p} A_k \text{vec}(X(t-k)) + \varepsilon(t) - \sum_{j=1}^{q} B_j \varepsilon(t-j), \tag{17}$$

where A_0 is an $mn \times 1$-dimensional vector; A_k and B_j are $mn \times mn$-dimensional matrices for $k = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$; and $\{\varepsilon(t), t \in \mathbb{N}\}$ is an $mn \times 1$-dimensional white noise satisfying that $\varepsilon(t)$ is independent with $\text{vec}(X(s))$ for all $s < t$.

A natural question is why the authors still bring forward MARMA(p,q) model (9) for $\{X(t), t \in \mathbb{N}\}$ but directly use VARMA(p,q) model (17) for $\{\mathrm{vec}(X(t)), t \in \mathbb{N}\}$.

In fact, there are two important reasons that the authors propose MARMA(p,q) model (9) for $\{X(t), t \in \mathbb{N}\}$. Firstly, MARMA$(p,q)$ model (9) for $\{X(t), t \in \mathbb{N}\}$ can reveal the information structure of $\{X(t), t \in \mathbb{N}\}$ very clearly. Secondly, MARMA(p,q) model (9) for $\{X(t), t \in \mathbb{N}\}$ can reduce model parameters more greatly than VARMA(p,q) model (17) for $\{\mathrm{vec}(X(t)), t \in \mathbb{N}\}$. In fact, the parameter number of MARMA(p,q) model (9) for $\{X(t), t \in \mathbb{N}\}$ is $2mn + (p+q)(m^2+n^2)$. However, the parameter number of VARMA(p,q) model (17) for $\{\mathrm{vec}(X(t)), t \in \mathbb{N}\}$ is $2mn + (p+q)m^2n^2$. Generally,

$$2mn + (p+q)(m^2+n^2) \ll 2mn + (p+q)m^2n^2.$$

For example, if $p = q = 1$ and $m = n = 10$, then

$$2mn + (p+q)(m^2+n^2) = 600 \ll 2mn + (p+q)m^2n^2 = 20200.$$

In today's big data era, m and n are often very large, taking $m = n = 100$ and $p = q = 1$ as an example, and then

$$2mn + (p+q)(m^2+n^2) = 60000 \ll 2mn + (p+q)m^2n^2 = 200020000.$$

Remark 3.1: MARMA(p,q) model (9) greatly reduces model parameters compared with VARMA(p,q) model (17).

Although it is not a good idea to replace MARMA(p,q) model (9) with VARMA(p,q) model (17), in the following we will show there exists a special VARMA(p,q) model equivalent to MARMA(p,q) model, which will play a very important role in theoretical analysis of MARMA(p,q) model (9).

Theorem 3.1: MARMA(p,q) *model* (9) *for* $\{X(t), t \in \mathbb{N}\}$ *is equivalent to* VARMA(p,q) *model* (18) *for* $\{\mathrm{vec}(X(t)), t \in \mathbb{N}\}$ *as follows:*

$$\begin{aligned}\mathrm{vec}(X(t)) = \mathrm{vec}(C) &+ \sum_{k=1}^{p} \Psi_k^\top \otimes \Phi_k \mathrm{vec}(X(t-k)) + \mathrm{vec}(\varepsilon(t)) \\ &- \sum_{j=1}^{q} \Xi_j^\top \otimes \Theta_j \mathrm{vec}(\varepsilon(t-j)),\end{aligned} \tag{18}$$

where $\mathrm{vec}(X(t))$ *and* $\mathrm{vec}(\varepsilon(t))$ *represent the vectorization of matrices* $X(t)$ *and* $\varepsilon(t)$ *by columns, and* $\otimes$ *is the Kronecker product.*

Theorem 3.1 can be proved by the following equivalence relation: for any matrices $Y_{m\times n}, A_{m\times m}, B_{m\times n}$ and $C_{n\times n}$, it follows that

$$Y_{m\times n} = A_{m\times m}B_{m\times n}C_{n\times n} \iff \mathrm{vec}(Y_{m\times n}) = (C_{n\times n}^\top \otimes A_{m\times m})\mathrm{vec}(B_{m\times n}).$$

The equivalence relation is not difficult to prove, so we omit the proof and that of Theorem 3.1.

3.3. Stationary and invertible conditions for MARMA model

According to Theorem 3.1, any MARMA(p,q) model (9) can be converted into its corresponding VARMA(p,q) model (18). Furthermore, VARMA(p,q) model (18) can be rewritten as

$$P(B)\mathrm{vec}(X(t)) = \mathrm{vec}(C) + Q(B)\mathrm{vec}(\varepsilon(t)), \quad t \in \mathbb{N}, \tag{19}$$

where

$$P(B) = I_{mn} - \sum_{k=1}^{p} \Psi_k^\top \otimes \Phi_k B^k, \tag{20}$$

$$Q(B) = I_{mn} - \sum_{j=1}^{q} \Xi_j^\top \otimes \Theta_j B^j \tag{21}$$

and B is the delay operator, i.e., $BX(t) = X(t-1)$ holds for all $t \in \mathbb{N}$.

Theorem 3.2: *For* MARMA(p,q) *model* (9), *the necessary and sufficient conditions for stationarity are that any root* λ *of* (22) *is in the unit circle, where*

$$\left|\lambda^p I_{mn} - \lambda^{p-1}\Psi_1^\top \otimes \Phi_1 - \lambda^{p-2}\Psi_2^\top \otimes \Phi_2 - \cdots - \lambda\Psi_{p-1}^\top \otimes \Phi_{p-1} - \Psi_p^\top \otimes \Phi_p\right| = 0. \tag{22}$$

The necessary and sufficient conditions for invertibility are that any root λ *of* (23) *is in the unit circle, where*

$$\left|\lambda^q I_{mn} - \lambda^{q-1}\Xi_1^\top \otimes \Theta_1 - \lambda^{q-2}\Xi_2^\top \otimes \Theta_2 - \cdots - \lambda\Xi_{q-1}^\top \otimes \Theta_{q-1} - \Xi_q^\top \otimes \Theta_q\right| = 0. \tag{23}$$

The proof of Theorem 3.2 is presented in Appendix 1.

Corollary 3.1: *For* MAR*(p) model* (10), *the necessary and sufficient conditions for stationarity are that any root* λ *of* (22) *is in the unit circle.*

Remark 3.2: Corollary 3.1 expands Proposition 1 in Chen et al. (2021).

Corollary 3.2: *For* MMA*(q) model* (11), *the necessary and sufficient conditions for invertibility are that any root* λ *of* (23) *is in the unit circle.*

3.4. Parameter estimation for MARMA model

In the section, we will present the conditional least square method and the conditional maximum likelihood estimation method for MARMA(p,q) model (9).

Let $x_1, x_2, \ldots, x_N$ be a series of samples of the centralized matrix time series $X = \{X(t), t \in \mathbb{N}\}$ defined by (3) with $C = O_{m\times n}$, where

$$x_t = \begin{bmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1n}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}(t) & x_{m2}(t) & \cdots & x_{mn}(t) \end{bmatrix}, \quad t = 1, 2, \ldots, N, \tag{24}$$

where the integer N is the sample length.

When the coefficient parameters of MARMA(p,q) model (9) have been obtained, it follows from (12) that

$$C = \mu - \Phi_1\mu\Psi_1 - \Phi_2\mu\Psi_2 - \cdots - \Phi_p\mu\Psi_p,$$

and then the constant matrix C of MARMA(p,q) model (9) can be estimated as follows:

$$\widehat{C} = \overline{X} - \Phi_1\overline{X}\Psi_1 - \Phi_2\overline{X}\Psi_2 - \cdots - \Phi_p\overline{X}\Psi_p,$$

where

$$\overline{X} = \frac{1}{N}\sum_{t=1}^{N} x_t.$$

Thus, in the following we always assume the samples come from a centralized MARMA(p,q) model (9), i.e., $C = O_{m\times n}$.

We use VARMA(p,q) model (19) with $C = O_{m\times n}$, equivalent to centralized MARMA(p,q) model (9), to estimate the coefficient parameters of MARMA(p,q) model (9) by the conditional least square method.

It yields from (19) with $C = O_{m\times n}$ that

$$\mathrm{vec}(\varepsilon(t)) = Q^{-1}(B)P(B)\mathrm{vec}(X(t)), \quad t \in \mathbb{N}, \tag{25}$$

where $Q^{-1}(B)$ is the inverse operator of $Q(B)$, and

$$P(B) = I_{mn} - \sum_{k=1}^{p} \Psi_k^\top \otimes \Phi_k B^k, \quad Q(B) = I_{mn} - \sum_{j=1}^{q} \Xi_j^\top \otimes \Theta_j B^j.$$

For the sake of briefness, denote

$$G(B) \triangleq \sum_{k=0}^{+\infty} G_k B^k = Q^{-1}(B)P(B) \tag{26}$$

and

$$P(B) = \sum_{i=0}^{+\infty} P_i B^i, \quad Q(B) = \sum_{j=0}^{+\infty} Q_j B^j,$$

where we stipulate that

$$P_i = \begin{cases} I_{mn}, & i = 0, \\ -\Psi_i^\top \otimes \Phi_i, & 1 \le i \le p, \\ O_{mn}, & i \ge p+1, \end{cases} \quad \text{and} \quad Q_j = \begin{cases} I_{mn}, & j = 0, \\ -\Xi_j^\top \otimes \Theta_j B^j, & 1 \le j \le q, \\ O_{mn}, & j \ge q+1. \end{cases} \tag{27}$$

It follows from (26) that $Q(B)G(B) = P(B)$, which means that

$$\sum_{i=0}^{k} Q_i G_{k-i} = P_k, \quad k = 0, 1, 2, \ldots. \tag{28}$$

It yields from (28) and (27) that

$$G_k = \begin{cases} I_{mn}, & k = 0, \\ -\Psi_k^\top \otimes \Phi_k + \sum_{i=1}^{k\wedge q} \left(\Xi_i^\top \otimes \Theta_i\right) G_{k-i}, & 1 \le k \le p, \\ \sum_{i=1}^{k\wedge q} \left(\Xi_i^\top \otimes \Theta_i\right) G_{k-i}, & k \ge p+1, \end{cases} \tag{29}$$

where $k \wedge q = \min\{k, q\}$.

In summary, centralized MARMA(p, q) model (9), i.e., $C = O_{m\times n}$, is equivalent to VARMA(p, q) model (30).

$$\mathrm{vec}(\varepsilon(t)) = \sum_{k=0}^{+\infty} G_k \mathrm{vec}(X(t-k)), \quad t \in \mathbb{N}, \tag{30}$$

where G_k, $k = 0, 1, 2, \ldots$, are given by (29).

Theorem 3.3: *According to the conditional least square method, the parameters of* MARMA(p, q) *model* (9) *satisfy the following matrix differential equations:*

$$\begin{cases} \sum_{t=p+1}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^\top}{\partial \Phi_i} \left(I_m \otimes \left(\tilde{x}_t + \sum_{\ell=1}^{t-1} G_\ell \tilde{x}_{t-\ell}\right)\right) = O_m, & i = 1, 2, \ldots, p, \\ \sum_{t=p+1}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^\top}{\partial \Psi_i} \left(I_n \otimes \left(\tilde{x}_t + \sum_{\ell=1}^{t-1} G_\ell \tilde{x}_{t-\ell}\right)\right) = O_n, & i = 1, 2, \ldots, p, \\ \sum_{t=p+1}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^\top}{\partial \Theta_j} \left(I_m \otimes \left(\tilde{x}_t + \sum_{\ell=1}^{t-1} G_\ell \tilde{x}_{t-\ell}\right)\right) = O_m, & j = 1, 2, \ldots, q, \\ \sum_{t=p+1}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^\top}{\partial \Xi_j} \left(I_n \otimes \left(\tilde{x}_t + \sum_{\ell=1}^{t-1} G_\ell \tilde{x}_{t-\ell}\right)\right) = O_n, & j = 1, 2, \ldots, q, \end{cases}$$

where G_k is given by (29).

The proof of Theorem 3.3 is presented in Appendix 2.

Corollary 3.3: *According to the conditional least square method, the parameters of* MAR*(p) model* (10) *satisfy the following matrix differential equations:*

$$\begin{cases} \sum_{t=p+1}^{N} \frac{\partial\left(\tilde{x}_{t-i}^{\top}(\Psi_i \otimes \Phi_i^{\top})\right)}{\partial \Phi_i} \left(I_m \otimes \left(\tilde{x}_t - \sum_{\ell=1}^{p}(\Psi_\ell^{\top} \otimes \Phi_\ell)\tilde{x}_{t-\ell}\right)\right) = O_m, \\ \sum_{t=p+1}^{N} \frac{\partial\left(\tilde{x}_{t-i}^{\top}(\Psi_i \otimes \Phi_i^{\top})\right)}{\partial \Psi_i} \left(I_n \otimes \left(\tilde{x}_t - \sum_{\ell=1}^{p}(\Psi_\ell^{\top} \otimes \Phi_\ell)\tilde{x}_{t-\ell}\right)\right) = O_n, \end{cases} \quad i = 1, 2, \ldots, p.$$

Theorem 3.4: *Assume the innovations are Gaussian with the mean $O_{m\times n}$ and covariance matrix Σ_{mn}. According to the conditional maximum likelihood estimation method, the parameters of centralized* MARMA(p, q) *model* (9) *satisfy the following matrix differential equations:*

$$\begin{cases} \sum_{t=2}^{N}\sum_{k=1}^{t-1} \frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial \Phi_i} \left(I_m \otimes \left[\Sigma_{mn}^{-1}H(t, G.)\right]\right) = O_m, \quad i = 1, 2, \ldots, p, \\ \sum_{t=2}^{N}\sum_{k=1}^{t-1} \frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial \Psi_i} \left(I_n \otimes \left[\Sigma_{mn}^{-1}H(t, G.)\right]\right) = O_n, \quad i = 1, 2, \ldots, p, \\ \sum_{t=2}^{N}\sum_{k=1}^{t-1} \frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial \Theta_j} \left(I_m \otimes \left[\Sigma_{mn}^{-1}H(t, G.)\right]\right) = O_m, \quad j = 1, 2, \ldots, q, \\ \sum_{t=2}^{N}\sum_{k=1}^{t-1} \frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial \Xi_j} \left(I_n \otimes \left[\Sigma_{mn}^{-1}H(t, G.)\right]\right) = O_n, \quad j = 1, 2, \ldots, q, \\ \frac{1}{N}\sum_{t=1}^{N} \mathrm{Res}\left(\left[\Sigma_{mn}^{-1}H(t, G.)\right] \otimes \left[\Sigma_{mn}^{-1}H(t, G.)\right], mn, mn\right) = \Sigma_{mn}^{-1}, \end{cases}$$

where $H(t, G.) = \tilde{x}_t + \sum_{k=1}^{t-1} G_k\tilde{x}_{t-k}$ and G_k is given by (29).

The proof of Theorem 3.4 is presented in Appendix 3.

Corollary 3.4: *Assume the innovations are Gaussian with the mean $O_{m\times n}$ and covariance matrix Σ_{mn}. According to the conditional maximum likelihood estimation method, the parameters of centralized* MAR*(p) model* (10) *satisfy the following matrix differential equations:*

$$\begin{cases} \sum_{t=i+1}^{N} \frac{\partial\left(\tilde{x}_{t-i}^{\top}(\Psi_i \otimes \Phi_i^{\top})\right)}{\partial \Phi_i} \left(I_m \otimes \left[\Sigma_{mn}^{-1}H(t, \Phi., \Psi.)\right]\right) = O_m, \quad i = 1, 2, \ldots, p, \\ \sum_{t=i+1}^{N} \frac{\partial\left(\tilde{x}_{t-i}^{\top}(\Psi_i \otimes \Phi_i^{\top})\right)}{\partial \Psi_i} \left(I_n \otimes \left[\Sigma_{mn}^{-1}H(t, \Phi., \Psi.)\right]\right) = O_n, \quad i = 1, 2, \ldots, p, \\ \frac{1}{N}\sum_{t=1}^{N} \mathrm{Res}\left(\left[\Sigma_{mn}^{-1}H(t, \Phi., \Psi.)\right] \otimes \left[\Sigma_{mn}^{-1}H(t, \Phi., \Psi.)\right], mn, mn\right) = \Sigma_{mn}^{-1}, \end{cases}$$

where $H(t, \Phi., \Psi.) = \tilde{x}_t - \sum_{s=1}^{(t-1)\wedge p}(\Psi_s^{\top} \otimes \Phi_s)\tilde{x}_{t-s}$.

Remark 3.3: The matrix differential equations in Theorems 3.3 and 3.4 are very complex. Especially, the coefficients G_k in (29), $k = 1, 2, \ldots$, are defined by a series of recursions, whose implied parameters are to be estimated. Thus, it is difficult to obtain its closed solution, but its approximate solutions can be obtained by the numerical computation method.

3.5. Hypothesis testing for the MARMA model

Let $x_1, x_2, \ldots, x_N$ be a series of samples of the centralized matrix time series $X = \{X(t), t \in \mathbb{N}\}$ defined by (3) with $C = O_{m\times n}$ and $x_t = (x_{ij})_{m\times n}$ for all $t = 1, 2, \ldots, N$. Additionally assume $x_1, x_2, \ldots, x_N$ are Gaussian. In the section, we will test whether $x_1, x_2, \ldots, x_N$ follow MARMA(p, q) model (9).

The null hypothesis and the alternative hypothesis follow as

H_0: $X = \{X(t), t \in \mathbb{N}\}$ follows MARMA(p, q) model (9);
H_1: $X = \{X(t), t \in \mathbb{N}\}$ does not follow MARMA(p, q) model (9).

When H_0 holds, denote

$$\tilde{\varepsilon}(t) = \sum_{k=0}^{t-1} G_k \tilde{x}_{t-k}, \quad t = 1, 2, \ldots, N, \tag{31}$$

where $\tilde{(\cdot)} = \mathrm{vec}(\cdot)$. It follows from Corollary 5.3 (Karl & Simar, 2015) that

$$T^2 = N(\bar{\tilde{\varepsilon}})^\top \left(S^2_{\tilde{\varepsilon}}\right)^{-1} \bar{\tilde{\varepsilon}} \sim T^2(mn, N-1),$$

where

$$\bar{\tilde{\varepsilon}} = \frac{1}{N}\sum_{t=1}^{N} \tilde{\varepsilon}_t \quad \text{and} \quad S^2_{\tilde{\varepsilon}} = \frac{1}{N-1}\sum_{t=1}^{N} \tilde{\varepsilon}_t \tilde{\varepsilon}_t^\top. \tag{32}$$

It follows from Theorem 5.9 (Karl & Simar, 2015) that

$$\frac{N-mn}{(N-1)mn} T^2(mn, N-1) \sim F(mn, N-mn),$$

that is,

$$F = \frac{N(N-mn)}{(N-1)mn} (\bar{\tilde{\varepsilon}})^\top \left(S^2_{\tilde{\varepsilon}}\right)^{-1} \bar{\tilde{\varepsilon}} \sim F(mn, N-mn).$$

Summarize the above deduction and we obtain Theorem 3.5 for the hypothesis testing on MARMA(p, q) model (9).

Theorem 3.5: *For any given significance level* $\alpha \in (0, 1)$, *if* $F < F_{\frac{\alpha}{2}}(mn, N-mn)$ *or* $F > F_{1-\frac{\alpha}{2}}(mn, N-mn)$, *then reject* $\{X(t), t \in \mathbb{N}\}$ *following* MARMA(p, q) *model* (9); *otherwise, accept* $\{X(t), t \in \mathbb{N}\}$ *following* MARMA(p, q) *model* (9), *where*

$$F = \frac{N(N-mn)}{(N-1)mn} (\bar{\tilde{\varepsilon}})^\top \left(S^2_{\tilde{\varepsilon}}\right)^{-1} \bar{\tilde{\varepsilon}},$$

and $\bar{\tilde{\varepsilon}}$, $S^2_{\tilde{\varepsilon}}$, $\tilde{\varepsilon}_t$ *are given by* (32) *and* (31).

3.6. Forecasting for the MARMA model

Let $\{X_t, t \in \mathbb{N}\}$ be an $m \times n$-dimensional matrix time series defined by (3) following MARMA(p, q) model (9), equivalently, $\{\mathrm{vec}(X_t), t \in \mathbb{N}\}$ following VARMA(p, q) model (18), that is,

$$\mathrm{vec}(X_t) = \mathrm{vec}(C) + \sum_{k=1}^{p} \Psi_k^\top \otimes \Phi_k \mathrm{vec}(X_{t-k}) + \mathrm{vec}(\varepsilon_t) \\ - \sum_{j=1}^{q} \Xi_j^\top \otimes \Theta_j \mathrm{vec}(\varepsilon_{t-j}), \tag{33}$$

where $\{\varepsilon_t, t \in \mathbb{N}\}$ is an $m \times n$-dimensional matrix white noise.

Denote the forecasting for $X_{t+\ell}$ under the condition that $X_1, X_2, \ldots, X_t$ have been known by $\widehat{X}_t(\ell)$, which refers to the ℓth step forecasting. It follows from (33) and the projection theorem in Hilbert space that

$$\mathrm{vec}(\widehat{X}_t(\ell)) = \mathrm{vec}(C) + \sum_{k=1}^{p} \Psi_k^\top \otimes \Phi_k \mathrm{vec}(\widetilde{X}_t(\ell-k)) - \sum_{j=1}^{q} \Xi_j^\top \otimes \Theta_j \mathrm{vec}(\tilde{\varepsilon}_t(\ell-j)), \tag{34}$$

where

$$\mathrm{vec}(\widetilde{X}_t(k)) = \begin{cases} \mathrm{vec}(X_{t+k}), & k \le 0, \\ \mathrm{vec}(\widehat{X}_t(k)), & k \ge 1, \end{cases} \quad \text{and} \quad \mathrm{vec}(\tilde{\varepsilon}_t(k)) = \begin{cases} \mathrm{vec}(\varepsilon_{t+k}), & k \le 0, \\ \mathrm{vec}(O_{m\times n}), & k \ge 1. \end{cases}$$

It yields from the equivalence relation of MARMA(p,q) model (9) and VARMA(p,q) model (18) that

$$\widehat{X}_t(\ell) = C + \sum_{k=1}^{p} \Phi_k \widetilde{X}_t(\ell-k)\Psi_k - \sum_{j=1}^{q} \Theta_j \tilde{\varepsilon}_t(\ell-j)\Xi_j, \tag{35}$$

where

$$\widetilde{X}_t(k) = \begin{cases} X_{t+k}, & k \le 0, \\ \widehat{X}_t(k), & k \ge 1, \end{cases} \quad \text{and} \quad \tilde{\varepsilon}_t(k) = \begin{cases} \varepsilon_{t+k}, & k \le 0, \\ O_{m\times n}, & k \ge 1. \end{cases}$$

In the following, we will study the interval estimation of MARMA(p,q) model (9) and assume the innovations are Gaussian. Equivalently, $\{\mathrm{vec}(X_t), t \in \mathbb{N}\}$ follows VARMA(p,q) model (19), that is,

$$P(B)\mathrm{vec}(X_t) = \mathrm{vec}(C) + Q(B)\mathrm{vec}(\varepsilon_t),$$

where $P(B)$ and $Q(B)$ are defined by (20) and (21), and $\{\mathrm{vec}(\varepsilon_t), t \in \mathbb{N}\}$ is a vector white noise.

Denote

$$\Pi(B) \triangleq \sum_{k=0}^{+\infty} \Pi_k B^k = P^{-1}(B)Q(B),$$

and then

$$\mathrm{vec}(X_t) = P^{-1}(B)\mathrm{vec}(C) + \sum_{k=0}^{+\infty} \Pi_k \mathrm{vec}(\varepsilon_{t-k}), \tag{36}$$

where

$$\Pi_k = \begin{cases} I_{mn}, & k = 0, \\ -\Xi_k^\top \otimes \Theta_k + \sum_{i=1}^{k\wedge p} \left(\Psi_i^\top \otimes \Phi_i\right) \Pi_{k-i}, & 1 \le k \le q, \\ \sum_{i=1}^{k\wedge p} \left(\Psi_i^\top \otimes \Phi_i\right) \Pi_{k-i}, & k \ge q+1, \end{cases} \tag{37}$$

with $k \wedge p = \min\{k, p\}$.

For any $\ell > 0$, it follows from (36) and the estimation method of $\mathrm{vec}(\widehat{X}_t(\ell))$ that

$$\mathrm{vec}(X_{t+\ell}) - \mathrm{vec}(\widehat{X}_t(\ell)) = \sum_{k=0}^{\ell-1} \Pi_k \mathrm{vec}(\varepsilon_{t+\ell-k}), \tag{38}$$

and then

$$\mathrm{vec}(X_{t+\ell}) - \mathrm{vec}(\widehat{X}_t(\ell)) \sim N\left(O_{mn\times 1}, \sum_{k=0}^{\ell-1} \Pi_k \Sigma_{mn} \Pi_k^\top\right). \tag{39}$$

For any given $\alpha \in (0,1)$, it yields from (39) that the confidence interval of $\mathrm{vec}(X_{t+\ell})$ with confidence level $1-\alpha$ follows as

$$\left(\mathrm{vec}(\widehat{X}_t(\ell)) - U_{1-\frac{\alpha}{2}}\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1} \Pi_k \Sigma_{mn} \Pi_k^\top\right)}, \mathrm{vec}(\widehat{X}_t(\ell)) + U_{1-\frac{\alpha}{2}}\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1} \Pi_k \Sigma_{mn} \Pi_k^\top\right)}\right),$$

where $\mathrm{diag}(\cdot)$ refers to the vector composed by all main diagonal elements, and $\sqrt{\cdot}$ means taking the square roots of every elements. It yields from the equivalence relation of MARMA(p,q) model (9) and VARMA(p,q) model (19) that the confidence interval of $X_{t+\ell}$ with confidence level $1-\alpha$ follows as

$$\left(\widehat{X}_t(\ell) - U_{1-\frac{\alpha}{2}}\mathrm{Res}\left(\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1}\Pi_k\Sigma_{mn}\Pi_k^\top\right)}, m, n\right), \widehat{X}_t(\ell) + U_{1-\frac{\alpha}{2}}\mathrm{Res}\left(\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1}\Pi_k\Sigma_{mn}\Pi_k^\top\right)}, m, n\right)\right).$$

In summary, we can obtain the following results.

Theorem 3.6: *Assume* $\{X_t, t \in \mathbb{N}\}$ *follows* MARMA(p,q) *model* (9).

(1) *For any* $\ell > 0$, *the* ℓ*-step point estimation follows as*

$$\widehat{X}_t(\ell) = C + \sum_{k=1}^{p}\Phi_k\widetilde{X}_t(\ell-k)\Psi_k - \sum_{j=1}^{q}\Theta_j\tilde{\varepsilon}_t(\ell-j)\Xi_j,$$

where

$$\widetilde{X}_t(k) = \begin{cases} X_{t+k}, & k \le 0, \\ \widehat{X}_t(k), & k \ge 1, \end{cases} \quad \text{and} \quad \tilde{\varepsilon}_t(k) = \begin{cases} \varepsilon_{t+k}, & k \le 0, \\ O_{m\times n}, & k \ge 1. \end{cases}$$

(2) *For any* $\ell > 0$ *and* $\alpha \in (0,1)$, *the* ℓ*-step interval estimation with confidence level* $1-\alpha$ *follows as*

$$\left(\widehat{X}_t(\ell) - U_{1-\frac{\alpha}{2}}\mathrm{Res}\left(\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1}\Pi_k\Sigma_{mn}\Pi_k^\top\right)}, m, n\right), \widehat{X}_t(\ell) + U_{1-\frac{\alpha}{2}}\mathrm{Res}\left(\sqrt{\mathrm{diag}\left(\sum_{k=0}^{\ell-1}\Pi_k\Sigma_{mn}\Pi_k^\top\right)}, m, n\right)\right),$$

where $U_{1-\frac{\alpha}{2}}$ *is the* $1-\frac{\alpha}{2}$ *level lower quantile of standard normal distribution,* $\mathrm{Res}(\cdot)$ *the reshape function by Definition 2.6,* $\mathrm{diag}(\cdot)$ *the vector composed by all main diagonal elements,* $\sqrt{\cdot}$ *takes the square roots of every elements, and*

$$\Pi_k = \begin{cases} I_{mn}, & k = 0, \\ -\Xi_k^\top \otimes \Theta_k + \sum_{i=1}^{k\wedge p}\left(\Psi_i^\top \otimes \Phi_i\right)\Pi_{k-i}, & 1 \le k \le q, \\ \sum_{i=1}^{k\wedge p}\left(\Psi_i^\top \otimes \Phi_i\right)\Pi_{k-i}, & k \ge q+1. \end{cases}$$

3.7. Supplementary notes for the MARMA model

3.7.1. Model identification for the MARMA model

According to Theorem 3.1, MARMA(p,q) model (9) is equivalent to VARMA(p,q) model (18). Thus, we can use the model identification method for the VARMA model to identify the order of MARMA model, such as

$$\mathrm{AIC}(p,q) = \ln(|\Sigma_{mn}(p,q)|) + \frac{2}{N}(p+q)(m^2+n^2),$$
$$\mathrm{BIC}(p,q) = \ln(|\Sigma_{mn}(p,q)|) + \frac{\ln(N)}{N}(p+q)(m^2+n^2),$$

or alternatively,

$$\mathrm{AIC}(p,q) = -\ln(L) + (p+q)(m^2+n^2),$$
$$\mathrm{BIC}(p,q) = -2\ln(L) + \ln(N)(p+q)(m^2+n^2),$$

where N is the length of observation sequence and $\ln(L)$ is the logarithm likelihood function.

3.7.2. MARIMA model

For any matrix time series $\{X(t) = (X_{ij}(t))_{m\times n}, t \in \mathbb{N}\}$ defined by (3), the difference operator Δ for matrix time series follows as

$$\begin{aligned} \Delta X(t) &= X(t) - X(t-1), \\ \Delta^k X(t) &= \Delta^{k-1} X(t) - \Delta^{k-1} X(t-1), \quad k = 2, 3, \ldots, \end{aligned} \tag{40}$$

and Δ defined by (40) has the same effect as the difference operator for vector time series. That is, if $\{X(t) = (X_{ij}(t))_{m\times n}, t \in \mathbb{N}\}$ is nonstationary, then we can try to eliminate nonstationarity by Δ defined by (40). If there exists a positive integer d such that $\{\Delta^d X(t), t \in \mathbb{N}\}$ is stationary but $\{\Delta^{d-1} X(t), t \in \mathbb{N}\}$ is nonstationary, and $\{\Delta^d X(t), t \in \mathbb{N}\}$ follows a MARMA(p, q) model (9), then $\{X(t) = (X_{ij}(t))_{m\times n}, t \in \mathbb{N}\}$ is called to follow a (p, d, q)-order autoregressive integrated moving average for matrix time series, and denoted by MARIMA(p, d, q).

4. An application of the MARMA model

In this section, we will try to model the time series of daily closing prices and daily volumes of Haitong Securities Company Limited (Abbreviated as Haitong Securities; Stock code: 600837) and Ping An Insurance (Group) Company of China, Ltd. (Abbreviated as Ping An; Stock code: 601318). The data are downloaded from the China Stock Market & Accounting Research Database (CSMAR), and the time window is from January 2, 2018 to December 31, 2021, which includes 973 records every stock.

For the sake of clarity, we denote the time series by

$$\left\{ \begin{bmatrix} P_1(t) & P_2(t) \\ V_1(t) & V_2(t) \end{bmatrix}, \quad t = 1, 2, 3, \ldots \right\},$$

where $P_1(t)$ and $V_1(t)$ are the daily closing price and daily volume of Haitong Securities, and $P_2(t)$ and $V_2(t)$ are the daily closing price and daily volume of Ping An.

4.1. Data preprocessing

We first conduct the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test, i.e., 'kpsstest' function in the software MATLAB R2020b, to test the stationarity of the daily closing prices and daily volumes of Haitong Securities and Ping An, and the results show that the daily closing prices and daily volumes of Haitong Securities and Ping An are nonstationary.

In the following we will consider the logarithmic rates (log rate) of daily closing prices and daily volumes of Haitong Securities and Ping An. Denote

$$\left\{ R(t) = \begin{bmatrix} R_{11}(t) & R_{12}(t) \\ R_{21}(t) & R_{22}(t) \end{bmatrix}, \quad t = 2, 3, 4, \ldots \right\}, \tag{41}$$

where

$$R_{1k}(t) = \ln\left(\frac{P_k(t)}{P_k(t-1)}\right) \quad \text{and} \quad R_{2k}(t) = \ln\left(\frac{V_k(t)}{V_k(t-1)}\right), \quad k = 1, 2.$$

That is, $R_{11}(t)$ is the logarithmic rate of daily closing price of Haitong Securities, $R_{21}(t)$ the logarithmic rate of daily volume of Haitong Securities, $R_{12}(t)$ the logarithmic rate of daily closing price of Ping An and $R_{22}(t)$ the logarithmic rate of daily volume of Ping An.

We conduct the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test, i.e., 'kpsstest' function in the software MATLAB R2020b, to test the stationarity of the logarithmic rates of daily closing prices and daily volumes of Haitong Securities and Ping An, and the results show that the logarithmic rates of daily closing prices and daily volumes of Haitong Securities and Ping An are stationary.

Additionally, we conduct a Ljung-Box Q test, i.e., 'lbqtest' function in the software MATLAB R2020b, to test the pure randomness of the logarithmic rates of daily closing prices and daily volumes of Haitong Securities and Ping An, and the results show that the logarithmic rates of daily closing prices or daily volumes of Haitong Securities and Ping An are not purely random.

In conclusion, for the stocks of Haitong Securities and Ping An, their daily closing prices and daily volumes are nonstationary, but their logarithmic rates of daily closing prices and daily volumes are stationary, and their logarithmic rates of daily closing prices or daily volumes are not purely random.

4.2. Modelling of MARMA (p, q)

We use the Bayesian information criterion (BIC) to select the model, and the results show that MARMA(4,0) is the best. Using the conditional least square method and MATLAB R2020b program, we establish MARMA(4,0) model for $\{R(t), t = 1, 2, 3, \ldots\}$ by (41) as follows:

$$R(t) = \widehat{C} + \widehat{\Phi}_1 R(t-1)\widehat{\Psi}_1 + \widehat{\Phi}_2 R(t-2)\widehat{\Psi}_2 + \widehat{\Phi}_3 R(t-3)\widehat{\Psi}_3 + \widehat{\Phi}_4 R(t-4)\widehat{\Psi}_4 + \varepsilon(t), \tag{42}$$

where

$$\widehat{\Phi}_1 = \begin{bmatrix} 2.11\times10^{-2} & 1.32\times10^{-3} \\ 3.4371 & -0.3527 \end{bmatrix}, \quad \widehat{\Psi}_1 = \begin{bmatrix} 1 & 0.3708 \\ -0.4014 & 0.0354 \end{bmatrix},$$

$$\widehat{\Phi}_2 = \begin{bmatrix} 0.0339 & -2.21\times10^{-3} \\ 1.1721 & -0.1271 \end{bmatrix}, \quad \widehat{\Psi}_2 = \begin{bmatrix} 0.4132 & 0.4468 \\ 1 & 0.2409 \end{bmatrix},$$

$$\widehat{\Phi}_3 = \begin{bmatrix} 0.0101 & -9.82\times10^{-4} \\ 0.3575 & -0.1739 \end{bmatrix}, \quad \widehat{\Psi}_3 = \begin{bmatrix} 1 & -0.4715 \\ -0.1625 & 0.2516 \end{bmatrix},$$

$$\widehat{\Phi}_4 = \begin{bmatrix} -0.0440 & 8.75\times10^{-4} \\ -0.4732 & -0.1012 \end{bmatrix}, \quad \widehat{\Psi}_4 = \begin{bmatrix} 0.9537 & 0.6853 \\ 0.4385 & 1 \end{bmatrix},$$

$$\widehat{\mu} = \begin{bmatrix} -6.66\times10^{-5} & -3.75\times10^{-4} \\ 4.86\times10^{-4} & -1.64\times10^{-3} \end{bmatrix},$$

and

$$\widehat{C} = \begin{bmatrix} -3.27\times10^{-5} & -3.48\times10^{-4} \\ 1.27\times10^{-3} & -3.40\times10^{-3} \end{bmatrix},$$

and then the covariance matrix of residuals $\{\varepsilon(t), t \in \mathbb{N}\}$ follows as

$$\Sigma_\varepsilon = \begin{bmatrix} 4.40\times10^{-4} & 2.74\times10^{-3} & 2.17\times10^{-4} & 1.30\times10^{-3} \\ 2.74\times10^{-3} & 0.1386 & 1.31\times10^{-3} & 5.60\times10^{-2} \\ 2.17\times10^{-4} & 1.31\times10^{-3} & 3.15\times10^{-4} & 1.31\times10^{-3} \\ 1.30\times10^{-3} & 5.60\times10^{-2} & 1.31\times10^{-3} & 0.1077 \end{bmatrix}. \tag{43}$$

4.3. Evaluation on MARMA (p, q)

For the sake of saving space, we will not show the model test, model optimization or forecasting of MARMA(4,0) model (42), but present a comparison of the MARMA model and ARMA model in this subsection. We first establish ARMA(p, q) model for $R_{11}(t), R_{21}(t), R_{12}(t)$ and $R_{22}(t)$, respectively, and obtain their models as follows:

$$\begin{aligned} R_{11}(t) &= 1.53\times10^{-8} + 0.0204R_{11}(t-1) + 0.0058R_{11}(t-2) \\ &\quad + 0.0625R_{11}(t-3) - 0.0393R_{11}(t-4) + e_{11}(t), \\ R_{21}(t) &= -2.98\times10^{-4} - 0.3994R_{21}(t-1) - 0.2598R_{21}(t-2) \\ &\quad - 0.1918R_{21}(t-3) - 0.1096R_{21}(t-4) + e_{21}(t), \\ R_{12}(t) &= 5.17\times10^{-7} - 0.0002R_{12}(t-1) - 0.0041R_{12}(t-2) \\ &\quad + 0.0476R_{12}(t-3) - 0.0699R_{12}(t-4) + e_{12}(t), \\ R_{22}(t) &= 2.52\times10^{-4} - 0.4842R_{22}(t-1) - 0.3860R_{22}(t-2) \\ &\quad - 0.2660R_{22}(t-3) - 0.1562R_{22}(t-4) + e_{22}(t), \end{aligned} \tag{44}$$

where the covariance matrix of residuals $\{e(t) \triangleq (e_{11}(t), e_{21}(t), e_{12}(t), e_{22}(t))^\top, t \in \mathbb{N}\}$ follows as

$$\Sigma_e = \begin{bmatrix} 4.49\times10^{-4} & 2.69\times10^{-3} & 2.22\times10^{-4} & 1.24\times10^{-3} \\ 2.69\times10^{-3} & 0.1692 & 1.40\times10^{-3} & 7.22\times10^{-2} \\ 2.22\times10^{-4} & 1.40\times10^{-3} & 3.22\times10^{-4} & 1.32\times10^{-3} \\ 1.24\times10^{-3} & 7.22\times10^{-2} & 1.32\times10^{-3} & 0.1369 \end{bmatrix}. \tag{45}$$

It follows from (43) and (45) that the residuals of MARMA(4,0) model (42) are almost consistently less than those of ARMA(4,0) model (44).

In practice, we are more concerned about the residual variance, i.e., the variance of every element of residual. Using (43) and (45), we compute the relative change of the residual variance of MARMA(4,0) model (42) to the residual variance of ARMA(4,0) model (44) as follows:

$$\begin{bmatrix} \dfrac{\mathrm{var}(\varepsilon_{11}(t)) - \mathrm{var}(e_{11}(t))}{\mathrm{var}(e_{11}(t))} & \dfrac{\mathrm{var}(\varepsilon_{12}(t)) - \mathrm{var}(e_{12}(t))}{\mathrm{var}(e_{12}(t))} \\ \dfrac{\mathrm{var}(\varepsilon_{21}(t)) - \mathrm{var}(e_{21}(t))}{\mathrm{var}(e_{21}(t))} & \dfrac{\mathrm{var}(\varepsilon_{22}(t)) - \mathrm{var}(e_{22}(t))}{\mathrm{var}(e_{22}(t))} \end{bmatrix} = \begin{bmatrix} -1.93\% & -2.24\% \\ -18.10\% & -21.31\% \end{bmatrix}.$$

That is, MARMA(4,0) model (42) reduces all residual variance relative to ARMA(4,0) model (44). Especially, the relative change of volume's residual variance exceeds 10% by MARMA(4,0) model (42) relative to ARMA(4,0) model (44), which means the MARMA model could really improve the prediction accuracy.

5. Conclusion

We proposed an autoregressive moving average model for matrix time series (MARMA), which is an extension of the autoregressive model for matrix time series (MAR). Like the MAR model, the MARMA model retains the original matrix structure, and provides a much more parsimonious model, compared with the approach of the vector autoregressive model for vectorizing the matrix into a long vector. Compared with MAR model, MARMA models are capable of modelling the unknown process with the minimum number of parameters.

As for MARMA model, the necessary and sufficient conditions for stationarity and invertibility are established. Parameter estimation methods are investigated for the conditional least square method and the conditional maximum likelihood estimation method. Point forecasting and interval forecasting are presented by using the projection theorem in the Hilbert space and the decomposition technique of time series. Additionally, model identification, model testing and possible extensions are discussed.

There are many directions to extend the scope of the MARMA model. Random environment such as the Markov environment might be imposed on the MARMA model to depict the impact of environmental change. Additionally, sparsity or group sparsity might be imposed on coefficient matrices to reach a further dimension reduction. Furthermore, the idea of MARMA can be applied for yield modelling, volatility modelling, weather forecast modelling and animal migration modelling.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper is partially supported by the basic scientific research business expenses of Universities in Xinjiang, China [Grant Number XQZX20230057] and the National Natural Science Foundation of China [Grant Number 11671142].

ORCID

Shujin Wu http://orcid.org/0000-0002-5274-3450

References

Chen, R., Xiao, H., & Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, *222*(1), 539–560. https://doi.org/10.1016/j.jeconom.2020.07.015

Getmanov, V. G., Chinkin, V. E., Sidorov, R. V., Gvishiani, A. D., Dobrovolsky, M. N., Soloviev, A. A., Dmitrieva, A. N., Kovylyaeva, A. A., & Yashin, I. I. (2021). Methods for recognition of local anisotropy in muon fluxes in the URAGAN hodoscope matrix data time series. *Physics of Atomic Nuclei*, *84*(6), 1080–1086. https://doi.org/10.1134/S106377882113010X

Graham, A. (2018). *Kronecker products and matrix calculus with applications*. Dover Publications Inc.

Karl, W., & Simar, L. (2015). *Applied multivariate statistical analysis* (4th ed.). Springer-Verlag.

Samadi, S. (2014). *Matrix time series analysis* [PhD Dissertation]. The University of Georgia.

Walden, A., & Serroukh, A. (2002). Wavelet analysis of matrix-valued time series. *Proceedings: Mathematical, Physical and Engineering Sciences*, *458*(2017), 157–179.

Wang, D., Liu, X., & Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, *208*(1), 231–248. https://doi.org/10.1016/j.jeconom.2018.09.013

Wang, H., & West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, *96*(4), 821–834. https://doi.org/10.1093/biomet/asp049

Wu, S. J., & Hua, N. (2022). Autoregressive model of time series with matrix cross-section data (in Chinese). *Acta Mathematica Sinica*, *65*(6), 1093–1104.

Zhou, L. H., Du, G. W., Tao, D. P., Chen, H. M., Cheng, J., & Gong, L. B. (2018). Clustering multivariate time series data via multi-nonnegative matrix factorization in multi-relational networks. *IEEE Access*, *2018*(6), 74747–74761. https://doi.org/10.1109/Access.6287639 doi: 10.1109/ACCESS.2018.2882798.

Appendices

Appendix 1. Proof of Theorem 3.2

In order to obtain stationary conditions and invertible conditions for MARMA(p,q) model (9), we first give a lemma as follows.

Lemma A.1: *For any square matrices $A_1, A_2, \ldots, A_k$, the operator*

$$G(B) \triangleq I - A_1 B - \cdots - A_{k-1}B^{k-1} - A_k B^k$$

is invertible if and only if any root λ of (A1) *satisfies $|\lambda| < 1$, where k is a natural number and B is the delay operator.*

$$\left|\lambda^k I - \lambda^{k-1}A_1 - \cdots - \lambda A_{k-1} - A_k\right| = 0. \tag{A1}$$

Proof: For the polynomial with k degree and matrix coefficients

$$G(z) = I - A_1 z - \cdots - A_{k-1}z^{k-1} - A_k z^k,$$

it can be factorized into k linear polynomials with the matrix coefficient in the complex field as follows:

$$G(z) = (I - C_1 z)(I - C_2 z)\cdots(I - C_k z),$$

where $C_1, C_2, \ldots, C_k$ are determined by

$$\sum_{1\le j_1<j_2<\cdots<j_u\le k} C_{j_1}C_{j_2}\cdots C_{j_u} = (-1)^{u-1}A_u, \quad u = 1,2,\ldots,k. \tag{A2}$$

Thus,

$$G(B) = (I - C_1 B)(I - C_2 B)\cdots(I - C_k B). \tag{A3}$$

For any $i = 1,2,\ldots,k$, it is easy to prove that $I - C_i B$ is invertible if and only if $\rho(C_i) < 1$, that is, all roots of $|\lambda I - C_i| = 0$ are in the unit circle. It follows from (A3) that $G(B)$ is invertible if and only if all $I - C_i B$, $i = 1,2,\ldots,k$, are invertible. Thus, $G(B)$ is invertible if and only if all roots of $|\lambda I - C_i| = 0$ are in the unit circle for all $i = 1,2,\ldots,k$. According to determinant properties, $G(B)$ is invertible if and only if all roots of

$$|(\lambda I - C_1)(\lambda I - C_2)\cdots(\lambda I - C_k)| = 0 \tag{A4}$$

are in the unit circle. It yields from (A2) that

$$(\lambda I - C_1)(\lambda I - C_2)\cdots(\lambda I - C_k) = \lambda^k I - \lambda^{k-1}A_1 - \cdots - \lambda A_{k-1} - A_k.$$

Thus, $G(B)$ is invertible if and only if all roots of

$$\left|\lambda^k I - \lambda^{k-1}A_1 - \cdots - \lambda A_{k-1} - A_k\right| = 0$$

are in the unit circle. ■

Proof of Theorem 3.2: For VARMA(p,q) model (19),

$$P(B)\mathrm{vec}(X(t)) = \mathrm{vec}(C) + Q(B)\mathrm{vec}(\varepsilon(t)), \quad t \in \mathbb{N}.$$

It follows from the concept of stationarity that the necessary and sufficient conditions of stationarity are that the operator $P(B)$ is invertible. According to Lemma A.1, the operator $P(B)$ is invertible if and only if any root λ of (22) satisfies $|\lambda| < 1$. Thus, VARMA(p,q) model (19) is stationary if and only if any root λ of (22) satisfies $|\lambda| < 1$. Note that VARMA(p,q) model (19) is equivalent to MARMA(p,q) model (9), so MARMA(p,q) model (9) is stationary if and only if any root λ of (22) satisfies $|\lambda| < 1$.

The necessary and sufficient conditions for invertibility can be obtained by the similar method to obtain the necessary and sufficient conditions for stationarity, so we omit it. ■

Appendix 2. Proof of Theorem 3.3

Noting that $\{\mathrm{vec}(\varepsilon(t)), t \in \mathbb{N}\}$ is an $mn \times 1$-dimensional white noise, and the objective function of VARMA(p,q) model (30) using the conditional least square method follows as

$$\begin{aligned} &J\left(\Phi_1,\ldots,\Phi_p,\Psi_1,\ldots,\Psi_p,\Theta_1,\ldots,\Theta_q,\Xi_1,\ldots,\Xi_q\right) \\ &= \sum_{t=p+1}^{N}\left(\mathrm{vec}(x_t) + \sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right)^{\top}\left(\mathrm{vec}(x_t) + \sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right), \end{aligned} \tag{A5}$$

where we take $x_t = O_{m\times n}$ for all $t \le 0$.

Lemma A.2: *$J(\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q)$ defined by* (A5) *has the minimum value about Φ_k, Ψ_k, Θ_j and Ξ_j for all $k=1,2,\dots,p$ and $j=1,2,\dots,q$.*

Proof: It yields from analysing (29) that $J(\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q)$ by (A5) is a multivariate polynomial of Φ_k, Ψ_k, Θ_j and Ξ_j for all $k=1,2,\dots,p$ and $j=1,2,\dots,q$. And it is obvious that $J(\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q)$ by (A5) is greater than or equal to zero, which means that $J(\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q)$ by (A5) has lower bound. Thus, $J(\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q)$ by (A5) has the minimum value about Φ_k, Ψ_k, Θ_j and Ξ_j for all $k=1,2,\dots,p$ and $j=1,2,\dots,q$. ■

Proof of Theorem 3.3.: It follows from Lemma A.2 that, according to the conditional least square method, the parameters of MARMA(p,q) model (9) satisfy the following matrix differential equations:

$$\begin{cases}
\dfrac{\partial}{\partial\Phi_i}\displaystyle\sum_{t=p+1}^{N}\left(\mathrm{vec}(x_t)+\sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right)^{\top}\left(\mathrm{vec}(x_t)+\sum_{\ell=1}^{t-1}G_\ell\mathrm{vec}(x_{t-\ell})\right)=O_m, & i=1,2,\dots,p,\\
\dfrac{\partial}{\partial\Psi_i}\displaystyle\sum_{t=p+1}^{N}\left(\mathrm{vec}(x_t)+\sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right)^{\top}\left(\mathrm{vec}(x_t)+\sum_{\ell=1}^{t-1}G_\ell\mathrm{vec}(x_{t-\ell})\right)=O_n, & i=1,2,\dots,p,\\
\dfrac{\partial}{\partial\Theta_j}\displaystyle\sum_{t=p+1}^{N}\left(\mathrm{vec}(x_t)+\sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right)^{\top}\left(\mathrm{vec}(x_t)+\sum_{\ell=1}^{t-1}G_\ell\mathrm{vec}(x_{t-\ell})\right)=O_m, & j=1,2,\dots,q,\\
\dfrac{\partial}{\partial\Xi_j}\displaystyle\sum_{t=p+1}^{N}\left(\mathrm{vec}(x_t)+\sum_{k=1}^{t-1}G_k\mathrm{vec}(x_{t-k})\right)^{\top}\left(\mathrm{vec}(x_t)+\sum_{\ell=1}^{t-1}G_\ell\mathrm{vec}(x_{t-\ell})\right)=O_n, & j=1,2,\dots,q.
\end{cases}$$

Using the derivative of scalar by matrix, it yields from Corollary 2.1 that

$$\begin{cases}
\displaystyle\sum_{t=p+1}^{N}\sum_{k=1}^{t-1}\frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial\Phi_i}\left(I_m\otimes\left(\tilde{x}_t+\sum_{\ell=1}^{t-1}G_\ell\tilde{x}_{t-\ell}\right)\right)=O_m, & i=1,2,\dots,p,\\
\displaystyle\sum_{t=p+1}^{N}\sum_{k=1}^{t-1}\frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial\Psi_i}\left(I_n\otimes\left(\tilde{x}_t+\sum_{\ell=1}^{t-1}G_\ell\tilde{x}_{t-\ell}\right)\right)=O_n, & i=1,2,\dots,p,\\
\displaystyle\sum_{t=p+1}^{N}\sum_{k=1}^{t-1}\frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial\Theta_j}\left(I_m\otimes\left(\tilde{x}_t+\sum_{\ell=1}^{t-1}G_\ell\tilde{x}_{t-\ell}\right)\right)=O_m, & j=1,2,\dots,q,\\
\displaystyle\sum_{t=p+1}^{N}\sum_{k=1}^{t-1}\frac{\partial\left(G_k\tilde{x}_{t-k}\right)^{\top}}{\partial\Xi_j}\left(I_n\otimes\left(\tilde{x}_t+\sum_{\ell=1}^{t-1}G_\ell\tilde{x}_{t-\ell}\right)\right)=O_n, & j=1,2,\dots,q.
\end{cases}$$ ■

Appendix 3. Proof of Theorem 3.4

It yields from (30) that

$$\mathrm{vec}(X(t))=-\sum_{k=1}^{+\infty}G_k\mathrm{vec}(X(t-k))+\mathrm{vec}(\varepsilon(t)),\quad t\in\mathbb{N}. \tag{A6}$$

For the sake of briefness, we denote

$$\widetilde{X}_t=\mathrm{vec}(X(t)),\quad t\in\mathbb{N}\quad\text{and}\quad\tilde{x}_k=\mathrm{vec}(x_k),\quad k=1,2,\dots,N.$$

It yields from (A6) that

$$\widetilde{X}_t\mid\left\{\widetilde{X}_{t-1},\widetilde{X}_{t-2},\dots\right\}\sim N\left(-\sum_{k=1}^{+\infty}G_k\widetilde{X}_{t-k},\Sigma_{mn}\right),\quad t\in\mathbb{N}, \tag{A7}$$

where Σ_{mn} is defined by (7).

Let $X(t)=O_{m\times n}$ for all $t\le 0$. It follows from (A7) that

$$\widetilde{X}_1\sim N\left(O_{mn\times 1},\Sigma_{mn}\right) \tag{A8}$$

and

$$\widetilde{X}_t\mid\left\{\widetilde{X}_{t-1},\widetilde{X}_{t-2},\dots,\widetilde{X}_1\right\}\sim N\left(-\sum_{k=1}^{t-1}G_k\widetilde{X}_{t-k},\Sigma_{mn}\right),\quad t\in\mathbb{N}. \tag{A9}$$

Thus, the maximum likelihood function of $x_1,x_2,\dots,x_V$ follows as

$$L\left(x_1,x_2,\dots,x_N;\Phi_1,\dots,\Phi_p,\Psi_1,\dots,\Psi_p,\Theta_1,\dots,\Theta_q,\Xi_1,\dots,\Xi_q\right)$$

$$
\begin{aligned}
&= L\left(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N; \Phi_1, \ldots, \Phi_p, \Psi_1, \ldots, \Psi_p, \Theta_1, \ldots, \Theta_q, \Xi_1, \ldots, \Xi_q\right) \\
&= f(\tilde{x}_1) f\left(\tilde{x}_2 \,\middle|\, \{\tilde{x}_1\}\right) f\left(\tilde{x}_3 \,\middle|\, \{\tilde{x}_2, \tilde{x}_1\}\right) \cdots f\left(\tilde{x}_N \mid \{\tilde{x}_{N-1}, \tilde{x}_{N-2}, \ldots, \tilde{x}_1\}\right) \\
&= (2\pi)^{-\frac{Nmn}{2}} |\Sigma_{mn}|^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N} \left(\tilde{x}_t + \sum_{k=1}^{t-1} G_k \tilde{x}_{t-k} \right)^{\top} \Sigma_{mn}^{-1} \left(\tilde{x}_t + \sum_{k=1}^{t-1} G_k \tilde{x}_{t-k} \right) \right\},
\end{aligned}
$$

where $f(\cdot)$ means the probability density function, and we stipulate $\sum_{k=1}^{0}(\cdot)$ equals zero vector or zero matrix as needed. Therefore, the logarithm maximum likelihood function of $x_1, x_2, \ldots, x_N$ follows as

$$
\begin{aligned}
&\ell\left(x_1, x_2, \ldots, x_N; \Phi_1, \ldots, \Phi_p, \Psi_1, \ldots, \Psi_p, \Theta_1, \ldots, \Theta_q, \Xi_1, \ldots, \Xi_q\right) \\
&\quad = \ln\left(L\left(x_1, x_2, \ldots, x_N; \Phi_1, \ldots, \Phi_p, \Psi_1, \ldots, \Psi_p, \Theta_1, \ldots, \Theta_q, \Xi_1, \ldots, \Xi_q\right)\right) \\
&\quad = -\frac{Nmn}{2} \ln(2\pi) - \frac{N}{2} \ln(|\Sigma_{mn}|) - \frac{1}{2} \sum_{t=1}^{N} \left(\tilde{x}_t + \sum_{k=1}^{t-1} G_k \tilde{x}_{t-k} \right)^{\top} \Sigma_{mn}^{-1} \left(\tilde{x}_t + \sum_{k=1}^{t-1} G_k \tilde{x}_{t-k} \right).
\end{aligned} \tag{A10}
$$

Using the derivative of scalar by matrix, it yields from (A10) that

$$
\begin{cases}
\sum_{t=1}^{N} \frac{\partial}{\partial \Phi_i} \left(H(t, G.)^{\top} \Sigma_{mn}^{-1} H(t, G.) \right) = O_m, \quad i = 1, 2, \ldots, p, \\
\sum_{t=1}^{N} \frac{\partial}{\partial \Psi_i} \left(H(t, G.)^{\top} \Sigma_{mn}^{-1} H(t, G.) \right) = O_n, \quad i = 1, 2, \ldots, p, \\
\sum_{t=1}^{N} \frac{\partial}{\partial \Theta_j} \left(H(t, G.)^{\top} \Sigma_{mn}^{-1} H(t, G.) \right) = O_m, \quad j = 1, 2, \ldots, q, \\
\sum_{t=1}^{N} \frac{\partial}{\partial \Xi_j} \left(H(t, G.)^{\top} \Sigma_{mn}^{-1} H(t, G.) \right) = O_n, \quad j = 1, 2, \ldots, q, \\
N \frac{\partial \ln(|\Sigma_{mn}|)}{\partial \Sigma_{mn}} + \sum_{t=1}^{N} \frac{\partial}{\partial \Sigma_{mn}} \left(H(t, G.)^{\top} \Sigma_{mn}^{-1} H(t, G.) \right) = O_{mn},
\end{cases} \tag{A11}
$$

where $H(t, G.) = \tilde{x}_t + \sum_{k=1}^{t-1} G_k \tilde{x}_{t-k}$. It yields from Corollary 2.2 and Property 2.3 that

$$
\begin{cases}
\sum_{t=2}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^{\top}}{\partial \Phi_i} \left(I_m \otimes \left[\Sigma_{mn}^{-1} H(t, G.) \right] \right) = O_m, \quad i = 1, 2, \ldots, p, \\
\sum_{t=2}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^{\top}}{\partial \Psi_i} \left(I_n \otimes \left[\Sigma_{mn}^{-1} H(t, G.) \right] \right) = O_n, \quad i = 1, 2, \ldots, p, \\
\sum_{t=2}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^{\top}}{\partial \Theta_j} \left(I_m \otimes \left[\Sigma_{mn}^{-1} H(t, G.) \right] \right) = O_m, \quad j = 1, 2, \ldots, q, \\
\sum_{t=2}^{N} \sum_{k=1}^{t-1} \frac{\partial \left(G_k \tilde{x}_{t-k}\right)^{\top}}{\partial \Xi_j} \left(I_n \otimes \left[\Sigma_{mn}^{-1} H(t, G.) \right] \right) = O_n, \quad j = 1, 2, \ldots, q, \\
\frac{1}{N} \sum_{t=1}^{N} \mathrm{Res}\left(\left[\Sigma_{mn}^{-1} H(t, G.) \right] \otimes \left[\left(\Sigma_{mn}^{-1} \right)^{\top} H(t, G.) \right], mn, mn \right) = \left(\Sigma_{mn}^{-1} \right)^{\top}.
\end{cases}
$$