International Statistical Review

International Statistical Review (2023) doi: 10.1111/insr.12558

Matrix-Variate Time Series Analysis: A Brief Review and Some New Developments

Ruey S. Tsay

Booth School of Business, University of Chicago, Chicago, 60637, IL, USA
Email: ruey.tsay@chicagobooth.edu

Summary

This paper briefly reviews the recent research in matrix-variate time series analysis, discusses some new developments, especially for seasonal time series, and demonstrates some applications. A general matrix autoregressive moving-average model is introduced. The paper narrates a simple approach for understanding the model, identifiability issues, and estimation. Real examples are used to illustrate the theory.

Key words: exponential smoothing; high dimension; multiplicative seasonal model; multivariate time series.

1 Introduction

The availability of large-scale and vast serially dependent data in recent years opens many new research topics in time series analysis, ranging from factor models for modelling high-dimensional data to tensor-variate models for exploring rich data structures. In this paper, we focus on matrix-variate time series, which consists of a sequence of two-dimensional arrays. The extension to general tensor-variate time series analysis is possible, but it would require more restrictions to address model identifiability and nonlinear parameter constraints. Our goal is to propose a general framework for modelling matrix-variate time series, to study some basic properties of the series, to discuss the relationship between vector and matrix time series, and to provide a simple approach to understanding and applying matrix-variate time series models. In particular, we address the issues of model identifiability, estimation, and model checking. A general matrix autoregessive moving-average (MARMA) model and multiplicative seasonal models are also discussed.

The scalar and vector time series analyses have been well studied in the literature and are widely used in applications. See, for example, the textbooks of Box *et al.* (2016), Lütkepohl (2006), Shumway & Stoffer (2017), and Tsay (2010, 2014) and the references therein. It is natural to think of the analysis of vector time series as an extension of that of the scalar one, but some extensions from scalar to vector time series are not straightforward. For instance, the generalisation of scalar autoregressive moving-average (ARMA) models to the vector autoregressive moving-average (VARMA) models introduces certain issues concerning model identifiablility. See, for instance, the block identifiability conditions in Dunsmuir & Hannan (1976), the scalar component model approach to modelling in Tiao & Tsay (1989), and the Kronecker indices for structural specification in Hannan & Deistler (1988). Empirical applications of vector time series models often encounter the issue of co-integration, whereas those of scalar time series only need to address the order of difference in achieving stationarity. The extension of vector time series analysis to matrix-variate time series seems easier in some ways, but some important issues arise in the extension. We shall discuss properties of matrix-variate time series that are easy to obtain from those of vector series and study issues that require further investigation.

To begin, we define some notations and introduce a reshaping operator of a square matrix, which is useful in understanding the identifiability problem of a matrix-variate time series model. The uppercase and lowercase bold-face letters denote a matrix and a vector, respectively. For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_2$ denotes its Frobenious norm, $tr(\boldsymbol{A})$ is its trace, and $\boldsymbol{a} = \text{vec}(\boldsymbol{A})$ denotes its column stacking vector, that is, $\boldsymbol{a} = (\boldsymbol{a}'_{.1}, \boldsymbol{a}'_{.2}, \ldots, \boldsymbol{a}'_{.h})'$, where $\boldsymbol{a}_{.j}$ is the jth column of $\boldsymbol{A}$ and $'$ denotes the transpose of a matrix or a vector. Also, let a_{ij} be the (i, j)th element of the matrix $\boldsymbol{A}$. For two matrices $\boldsymbol{D}_{g \times h}$ and $\boldsymbol{C}_{p \times q}$, let $\boldsymbol{D}\otimes\boldsymbol{C}$ be their Kronecker product, that is, $\boldsymbol{D}\otimes\boldsymbol{C} = [d_{ij}\boldsymbol{C}]_{gp \times hq}$. For a square matrix $\boldsymbol{A}$, let $\rho(\boldsymbol{A})$ be its largest eigenvalue, in modulus. Finally, we use $\boldsymbol{\Sigma} > 0$ to denote that $\boldsymbol{\Sigma}$ is a positive-definite matrix.

For a $gh \times gh$ matrix $\boldsymbol{M}$, we partition it into h^2 block sub-matrices, each of dimension $g \times g$. That is,

$$\boldsymbol{M}_{gh \times gh} = [\boldsymbol{M}_{ij}]_{i,j=1}^{h},$$

where $\boldsymbol{M}_{ij}$ is a $g \times g$ matrix, for $i, j = 1, \ldots, h$. Following Van Loan & Pitsianis (1993), we define a reshaping operator $\mathcal{R}(.)$ of the square matrix $\boldsymbol{M}$ as

$$\mathcal{R}(\boldsymbol{M}) = \begin{bmatrix} \text{vec}(\boldsymbol{M}_{11})' \\ \vdots \\ \text{vec}(\boldsymbol{M}_{h1})' \\ \text{vec}(\boldsymbol{M}_{12})' \\ \vdots \\ \text{vec}(\boldsymbol{M}_{1h})' \\ \vdots \\ \text{vec}(\boldsymbol{M}_{hh})' \end{bmatrix}_{h^2 \times g^2}. \tag{1}$$

Clearly, $\mathcal{R}(\boldsymbol{M})$ simply rearranges elements of $\boldsymbol{M}$ into a $h^2 \times g^2$ matrix. It has a nice property in conjunction with the Kronecker product. Specifically, suppose that $\boldsymbol{D}_{h \times h}$ and $\boldsymbol{C}_{g \times g}$ are two square matrices, then

$$\mathcal{R}(\boldsymbol{D}\otimes\boldsymbol{C}) = \text{vec}(\boldsymbol{D})\text{vec}(\boldsymbol{C})'. \tag{2}$$

This identity says that one can reshape the Kronecker product into a rank-1 matrix, which is useful in understanding matrix-variate time series models.

Returning to time series analysis, a VARMA(p, q) model for a k-dimensional vector time series $\{\boldsymbol{x}_t\}$ is

$$\boldsymbol{\phi}(B)\boldsymbol{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{3}$$

where $\boldsymbol{\phi}(B) = \boldsymbol{I} - \sum_{i=1}^{p}\boldsymbol{\phi}_i B^i$ and $\boldsymbol{\theta}(B) = \boldsymbol{I} - \sum_{j=1}^{q}\boldsymbol{\theta}_j B^j$ are the autoregressive (AR) and the moving-average (MA) matrix polynomials of degrees p and q, respectively, $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_j$ are $k \times k$ real-valued matrices, B is the backshift or lag operator such that $B\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$, and $\boldsymbol{\phi}_0$ is a k-dimensional constant vector. The innovation series $\{\boldsymbol{a}_t\}$ of Equation (3) is a sequence of serially uncorrelated random vectors with $E(\boldsymbol{a}_t) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{a}_t) = \boldsymbol{\Sigma}_a > 0$. The model reduces to a scalar ARMA model when $k = 1$. For model (3), we assume that all solutions to the determinant polynomial $|\boldsymbol{\phi}(B)| = 0$ are greater than or equal to one in modulus and those of $|\boldsymbol{\theta}(B)| = 0$ are greater than one in modulus. For the block identifiability of model (3), we further require that (a) $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are left co-prime and (b) the joint matrix $[\boldsymbol{\phi}_p, \boldsymbol{\theta}_q]$ is of rank k, where q is as small as possible and p is as small as possible for a given q. See Dunsmuir & Hannan (1976). The series $\{\boldsymbol{x}_t\}$ is weakly stationary if all solutions to $|\boldsymbol{\phi}(B)| = 0$ are outside the unit circle and is invertible if all solutions to $|\boldsymbol{\theta}(B)| = 0$ are outside the unit circle. For more properties and applications of VARMA models, readers are referred to Lütkepohl (2006) and Tsay (2014), among others.

Consider next a matrix-variate time series $\{\boldsymbol{X}_t\}$, which is of dimension $g \times h$. For example, consider six mid-western states of the United States; namely, Illinois, Indiana, Iowa, Wisconsin, Minnesota, and Michigan. Suppose that we are interested in analysing the employment and unemployment of these six states. The available variables include (a) civilian labour force (in persons) and (b) all employees: leisure and hospitality (in thousands of persons). In this particular instance, we have a monthly 6×2 matrix-variate time series with each row representing a state and each column representing an employment variable. Here one can employ a vector time series of dimension six to analyse each employment variable, or a bivariate series to analyse the labour market of each state, or a 12-dimensional series to modelling jointly the employment series of the six states. However, it would be of interest to study the mid-western regional labour market jointly while maintaining the state structure. This leads to the analysis of a 6×2 matrix-variate time series.

2 Matrix-Variate Time Series

A straightforward generalisation of VARMA models in Equation (3) to matrix-variate time series $\{\boldsymbol{X}_t\}$ of dimension $g \times h$ is

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_0 + \sum_{i=1}^{p} \boldsymbol{C}_i \boldsymbol{X}_{t-i} \boldsymbol{D}_i' + \boldsymbol{A}_t - \sum_{j=1}^{q} \boldsymbol{L}_j \boldsymbol{A}_{t-j} \boldsymbol{R}_j', \tag{4}$$

where $\boldsymbol{C}_i$ and $\boldsymbol{L}_j$ are $g \times g$ real-valued matrices, $\boldsymbol{D}_i$ and $\boldsymbol{R}_j$ are $h \times h$ real-valued matrices, $\boldsymbol{\Phi}_0$ is a $g \times h$ constant matrix, and $\{\boldsymbol{A}_t\}$ is a sequence of serially uncorrelated $g \times h$ random matrices with $E(\boldsymbol{A}_t) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{a}_t) = \boldsymbol{\Sigma}_a > 0$, where $\boldsymbol{a}_t = \mathrm{vec}(\boldsymbol{A}_t)$, which is a gh-dimensional series. If one further assumes that $\boldsymbol{A}_t$ follows a matrix normal distribution, then $\boldsymbol{\Sigma}_a = \boldsymbol{V} \otimes \boldsymbol{U}$, where $\boldsymbol{U} > 0$ is the left covariance matrix and $\boldsymbol{V} > 0$ is the right covariance matrix. See Section 2.3 below for further information. We shall refer to model (4) as a (rank-1) matrix autoregressive moving-average (MARMA) model.

2.1 Identifiability

The MARMA model in Equation (4) is not identifiable without further restrictions. For instance, the pairs $(\boldsymbol{C}_i, \boldsymbol{D}_i)$ and $\left(\frac{1}{c}\boldsymbol{C}_i, c\boldsymbol{D}_i\right)$, for $c \neq 0$, provide the same dependence of $\boldsymbol{X}_t$ on $\boldsymbol{X}_{t-i}$. A common practice in the literature is to require $\|\boldsymbol{C}_i\|_2 = 1$, for $i = 1, \ldots, p$, and $\|\boldsymbol{L}_i\|_2 = 1$, for $j = 1, \ldots, q$. Even with such requirements, the pair $(\boldsymbol{C}_i, \boldsymbol{D}_i)$ remains un-identified, because $(\boldsymbol{C}_i, \boldsymbol{D}_i)$ and $(-\boldsymbol{C}_i, -\boldsymbol{D}_i)$ would produce the same results. This sign issue is not critical in some applications, but it creates problems for parameter estimation. A possible solution is to require that $tr(\boldsymbol{D}_i) > 0$, for $i = 1, \ldots, p$, and $tr(\boldsymbol{R}_j) > 0$, for $j = 1, \ldots, q$. See, for instance, Hsu *et al.* (2021). We consider another simple solution to this identifiability problem later.

The MARMA model in Equation (4) can be rewritten as

$$\boldsymbol{x}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} (\boldsymbol{D}_i \otimes \boldsymbol{C}_i)\boldsymbol{x}_{t-i} + \boldsymbol{a}_t - \sum_{j=1}^{q} (\boldsymbol{R}_j \otimes \boldsymbol{L}_j)\boldsymbol{a}_{t-j}, \tag{5}$$

where $\boldsymbol{x}_t = \mathrm{vec}(\boldsymbol{X}_t)$ and $\boldsymbol{\phi}_0 = \mathrm{vec}(\boldsymbol{\Phi}_0)$. This is a VARMA representation for the MARMA model in (4) with AR and MA matrix polynomials being $\boldsymbol{\phi}_{C,D}(B) = \boldsymbol{I} - \sum_{i=1}^{p} (\boldsymbol{D}_i \otimes \boldsymbol{C}_i) B^i$ and $\boldsymbol{\theta}_{L,R}(B) = \boldsymbol{I} - \sum_{j=1}^{q} (\boldsymbol{R}_j \otimes \boldsymbol{L}_j) B^j$, respectively, where the subscripts (C, D) and (L, R) signify that the coefficient matrices of the matrix polynomials are Kronecker products of $(\boldsymbol{D}_i, \boldsymbol{C}_i)$ and $(\boldsymbol{R}_j, \boldsymbol{L}_j)$.

Taking advantages of the block identifiability conditions of VARMA models, we obtain that the MARMA model in (4) is block identifiable if the following two conditions hold: (a) $\boldsymbol{\phi}_{C,D}(B)$ and $\boldsymbol{\theta}_{L,R}(B)$ are left co-prime and (b) the joint matrix $[\boldsymbol{D}_p \otimes \boldsymbol{C}_p, \boldsymbol{R}_q \otimes \boldsymbol{L}_q]$ is of rank gh, where q is as small as possible and p is as small as possible for a given q.

An intuitive approach to modelling the matrix time series $\{\boldsymbol{X}_t\}$ is to fit a VARMA model of Equation (3) to the gh-dimensional vector process $\{\boldsymbol{x}_t\}$. However, the VARMA(p, q) model for $\{\boldsymbol{x}_t\}$ fails to capture the matrix structure of $\{\boldsymbol{X}_t\}$ and often employs many more parameters than the MARMA model in (4) does. Comparing models in Equations (3) and (4), we see that, ignoring the covariance matrices of $\boldsymbol{A}_t$ and $\boldsymbol{a}_t$ and constant terms, the VARMA model uses $(p+q)g^2h^2$ parameters whereas the MARMA model only contains $(p+q)(g^2+h^2-1)$ parameters. We subtract one parameter from each pair of $(\boldsymbol{D}_i, \boldsymbol{C}_i)$ and $(\boldsymbol{R}_j, \boldsymbol{L}_j)$ due to the scaling issue mentioned in Section 2.1. The difference in the number of parameters between the two models can be substantial even for small g and h. For example, consider the case of $(g, h, p, q) = (6, 2, 2, 0)$. In this particular case, the VARMA(2,0) model of dimension 12 employs 288 coefficient parameters, but the matrix model in Equation (4) only uses 78 coefficient parameters. In general, advantages of using MARMA models over VARMA ones include (a) making use of the matrix data structure and (b) obtaining a parsimonious model, which in turns avoids the problem of over-parameterisation. Mathematically speaking, the extension from (3) to the model in (4) is not sufficiently general. We consider a general (rank-r) MARMA model in the next section.

2.2 General Matrix-Variate ARMA Models

In theory, the MARMA model in (4) can be considered as a reduced (or sub) model of the VARMA model in (3) with nonlinear parameter constraints. Alternatively, one can think of the MARMA model as an approximation to the VARMA model with $\boldsymbol{\phi}_i$ being approximated by $(\boldsymbol{D}_i \otimes \boldsymbol{C}_i)$, for $i = 1, \ldots, p$, and $\boldsymbol{\theta}_j$ being approximated by $(\boldsymbol{R}_j \otimes \boldsymbol{L}_j)$, for $j = 1, \ldots, q$. Here we use a single Kronecker product to approximate each coefficient matrix of the corresponding VARMA model. The accuracy of such an approximation might be questionable in real applications. In fact, using a single Kronecker product to approximate a coefficient matrix can be considered as using a rank-1 approximation to the coefficient matrix. See the reshaping identity in (2). In addition, using multiple Kronecker products to approximate a given matrix has been investigated recently by Cai *et al.* (2022). To increase model flexibility and to improve accuracy in matrix approximation, Hsu *et al.* (2023) proposed a rank-r MAR(p) model for modelling spatio-temporal data observed on a regular grid, in which each AR coefficient matrix is approximated by r Kronecker products, that is, $\boldsymbol{\phi}_i = \sum_{j=1}^{r}(\boldsymbol{D}_{i,j} \otimes \boldsymbol{C}_{i,j})$. We can generalise their model to a general rank-r MARMA(p, q) model as follows:

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_0 + \sum_{i=1}^{p}\left(\sum_{j=1}^{r}\gamma_{i,j}\boldsymbol{C}_{ij}\boldsymbol{X}_{t-i}\boldsymbol{D}'_{ij}\right) + \boldsymbol{A}_t - \sum_{i=1}^{q}\left(\sum_{j=1}^{r}\omega_{i,j}\boldsymbol{L}_{ij}\boldsymbol{A}_{t-i}\boldsymbol{R}'_{ij}\right), \tag{6}$$

where r is a positive integer, $\gamma_{i,1} \geq \ldots \geq \gamma_{i,r} > 0$, for $i = 1, \ldots, p$, and $\omega_{i,1} \geq \ldots \geq \omega_{i,r} > 0$, for $i = 1, \ldots, q$. For model identification of (6), we assume that $\{\text{vec}(\boldsymbol{C}_{i,1}), \ldots, \text{vec}(\boldsymbol{C}_{i,r})\}$ and $\{\text{vec}(\boldsymbol{D}_{i,1}), \ldots, \text{vec}(\boldsymbol{D}_{i,r})\}$ are two sets of orthonormal vectors, for $i = 1, \ldots, p$, and $\{\text{vec}(\boldsymbol{L}_{i,1}), \ldots, \text{vec}(\boldsymbol{L}_{i,r})\}$ and $\{\text{vec}(\boldsymbol{R}_{i,1}), \ldots, \text{vec}(\boldsymbol{R}_{i,r})\}$ are also two sets of orthonormal vectors, for $i = 1, \ldots, q$. These identifiability conditions of coefficient matrices can equivalently be written as: for $1 \leq u, v \leq r$,

$$tr(\boldsymbol{C}'_{i,u}\boldsymbol{C}_{i,v}) = tr(\boldsymbol{D}'_{i,u}\boldsymbol{D}_{i,v}) = \delta_{uv},\ \gamma_{i,1} \geq \ldots \geq \gamma_{i,r} > 0;\ i = 1, \ldots, p, \tag{7}$$

$$tr(\boldsymbol{L}'_{i,u}\boldsymbol{L}_{i,v}) = tr(\boldsymbol{R}'_{i,u}\boldsymbol{R}_{i,v}) = \delta_{uv},\ \omega_{i,1} \geq \ldots \geq \omega_{i,r} > 0;\ i = 1, \ldots, q, \tag{8}$$

where $\delta_{uv} = 1$ if $u = v$ and = 0, otherwise. The requirement of using orthonormal vectors is to distinguish each Kronecker product and to fix the scaling effect of the Kronecker products so that the quantities $\gamma_{i,j}$ and $\omega_{i,j}$ can quantify the contribution of each Kronecker product in their matrix approximation. Clearly, this general MARMA model uses r Kronecker products to approximate each AR and MA coefficient matrix of the VARMA representation for $\{\boldsymbol{X}_t\}$. Indeed, from model (6), the VARMA representation of $\boldsymbol{X}_t$ becomes Equation (3) with

$$\boldsymbol{\phi}_i = \sum_{j=1}^{r}\gamma_{i,j}(\boldsymbol{D}_{i,j} \otimes \boldsymbol{C}_{i,j}),\ i = 1, \ldots, p;\ \ \boldsymbol{\theta}_i = \sum_{j=1}^{r}\omega_{i,j}(\boldsymbol{R}_{i,j} \otimes \boldsymbol{L}_{i,j}),\ i = 1, \ldots, q. \tag{9}$$

Applying the identity (2), we have

$$\mathcal{R}(\boldsymbol{\phi}_i) = \sum_{j=1}^{r}\gamma_{i,j}\mathcal{R}(\boldsymbol{D}_{i,j} \otimes \boldsymbol{C}_{i,j}) = \sum_{j=1}^{r}\gamma_{i,j}\text{vec}(\boldsymbol{D}_{i,j})\text{vec}(\boldsymbol{C}_{i,j})'.$$

Consequently, one can reshape each AR coefficient matrix to perform a singular value decomposition to identity $\gamma_{i,j}$ and $(\boldsymbol{D}_{i,j}, \boldsymbol{C}_{i,j})$, for $j = 1, \ldots, r$ and $i = 1, \ldots, p$. The same operation also applies to each MA coefficient matrix $\boldsymbol{\theta}_i$, for $i = 1, \ldots, q$. The idea of using rank-r approximations for AR coefficient matrices is also used in Li & Han (2021) for tensor autoregressive models.

Again, using the VARMA representation, one can easily obtain the stationarity, invertibility, and block identiability conditions for a general rank-r MARMA model, provided that the parameter identification conditions in (6) are met. Furthermore, one can even relax the assumption of using the same rank r for all coefficient matrices by using rank p_i for $\boldsymbol{\phi}_i$ and rank q_j for $\boldsymbol{\theta}_j$. In real applications, one expects that the rank r is sufficiently small so that the resulting MARMA model does not employ too many parameters.

2.3 A Brief Survey

In this section, we briefly review the analysis of matrix-variate time series. To begin, a $g \times h$ random matrix $\boldsymbol{A}$ is said to follow a matrix normal distribution with mean $\boldsymbol{M}$, left (or row) covariance matrix $\boldsymbol{U} > 0$, and right (column) covariance matrix $\boldsymbol{V} > 0$ if the probability density function of $\boldsymbol{A}$ is

$$p(\boldsymbol{A}|\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V}) = \frac{\exp\left(-\frac{1}{2}tr[\boldsymbol{V}^{-1}(\boldsymbol{A} - \boldsymbol{M})'\boldsymbol{U}^{-1}(\boldsymbol{A} - \boldsymbol{M})]\right)}{(2\pi)^{gh/2}|\boldsymbol{V}|^{g/2}|\boldsymbol{U}|^{h/2}},$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are $g \times g$ and $h \times h$ matrices. We use the notation $\boldsymbol{A} \sim N(\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V})$ to denote such a matrix normal distribution. It is easy to see that $\boldsymbol{A} \sim N(\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V})$ if and only if $\boldsymbol{a} = \text{vec}(\boldsymbol{A}) \sim N(\text{vec}(\boldsymbol{M}), \boldsymbol{V} \otimes \boldsymbol{U})$.

Research in matrix-variate time series has a relatively short history and most of the articles focus on matrix-variate autoregressive (MAR) models. H. Wang & West (2009) consider a dynamic linear model for the matrix-variate time series $\{\boldsymbol{X}_t\}$ as follows:

$$\boldsymbol{X}_t = (\boldsymbol{I}_g \otimes \boldsymbol{F}_t')\boldsymbol{Q}_t + \boldsymbol{H}_t, \;\; \boldsymbol{H}_t \sim N(\boldsymbol{0}, \boldsymbol{U}, \boldsymbol{V}), \tag{10}$$

$$\boldsymbol{Q}_t = (\boldsymbol{I}_g \otimes \boldsymbol{G}_t)\boldsymbol{Q}_{t-1} + \boldsymbol{\Omega}_t, \;\; \boldsymbol{\Omega}_t \sim N(\boldsymbol{0}, \boldsymbol{U} \otimes \boldsymbol{W}_t, \boldsymbol{V}), \tag{11}$$

where $\boldsymbol{Q}_t$ denotes the state matrix, and $\boldsymbol{F}_t$, $\boldsymbol{G}_t$ and $\boldsymbol{W}_t$ are time-varying matrices of proper dimensions. The authors applied the model (10)–(11) to study the dynamic relationship of a monthly 8×9 matrix-variate of employment statistics of eight U.S. states and nine industrial sectors. Samadi (2014) studies MAR models for the matrix series $\{\boldsymbol{X}_t\}$ with the right matrix $\boldsymbol{D}_i$ assuming a location matrix. Specifically, he considers the MAR(1) model

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_0 + \sum_{j=1}^{h}\sum_{i=1}^{h} \boldsymbol{G}_i^j \boldsymbol{X}_{t-1}\boldsymbol{E}_{ij} + \boldsymbol{A}_t, \tag{12}$$

where $\boldsymbol{G}_i^j$, for $i, j = 1, \ldots, h$, are $g \times g$ real-valued matrices and $\boldsymbol{E}_{ij} = [(e_{uv})_{ij}]$, for $i, j = 1, \ldots, h$, are $h \times h$ location matrices such that

$$(e_{uv})_{ij} = \begin{cases} 1, & \text{if } u = i \text{ and } v = j, \\ 0, & \text{otherwise.} \end{cases}$$

The model states that each element $x_{ij,t}$ of $\boldsymbol{X}_t$ is a linear function of elements of $\boldsymbol{X}_{t-1}$ plus the innovation $a_{ij,t}$ of $\boldsymbol{A}_t$. This is different from the other MAR models discussed so far. As a matter of fact, it is not hard to see that model (12) is just an alternative way to write the VAR(1) model for the vector $\{\boldsymbol{x}_t\}$ series in a matrix form.

Chen *et al.* (2021) propose MAR models to maintain and utilise the matrix structure of the data and investigate probabilistic properties of the proposed model. The authors also consider estimation via iterated least squares and maximum likelihood methods. Hsu *et al.* (2021) consider structured MAR models to characterise the temporal dynamics of spatio-temporal matrix-variate time series. The model is in the form of Equation (4) with $q = 0$, but certain banded structures are allowed for the coefficient matrices to take care of the nearest neighbours of individual series $x_{ij,t}$ and special attention is paid to the covariance structure of $\boldsymbol{A}_t$ to allow for non-separable covariance structure, which is commonly seen in spatial data analysis. Specifically, the covariance matrix of $\boldsymbol{a}_t$ assumes the form

$$\boldsymbol{\Sigma}_a = \boldsymbol{B}\boldsymbol{G}\boldsymbol{B}' + \sigma^2\boldsymbol{I}, \tag{13}$$

where $\boldsymbol{B}$ is a known $gh \times k$ matrix of basis functions with rank $k \le gh$, $\boldsymbol{G}$ is an unknown $k \times k$ nonnegative-definite matrix and $\sigma^2 \ge 0$ is an unknown parameter. The basis functions used in

Hsu *et al.* (2021) is the multi-resolution spline basis functions of Tzeng & Huang (2018). Other commonly used basis functions include wavelet basis functions, bisquare functions and Gaussian functions. Both iterated least squares and maximum likelihood methods are used to estimate the model, including $\boldsymbol{G}$ and σ^2, and information criteria such as AIC and BIC are used to perform model selection. Li & Han (2021) study multi-linear tensor autoregressive models, including model identifiability conditions for the AR coefficient matrices.

When the dimension gh is large, D. Wang *et al.* (2019) propose a matrix factor model for $\{\boldsymbol{X}_t\}$, which can be written as

$$\boldsymbol{X}_t = \boldsymbol{L}\boldsymbol{F}_t\boldsymbol{R}' + \boldsymbol{E}_t, \tag{14}$$

where $\boldsymbol{F}_t$ is a $g_1 \times h_1$ matrix-variate factor process with $g_1 << g$ and $h_1 << h$, $\boldsymbol{L}$ is an $g \times g_1$ left loading matrix and $\boldsymbol{R}$ is a $h \times h_1$ right loading matrix. The model in Equation (14) is not identifiable, because for any two nonsingular matrices $\boldsymbol{U}_{g_1 \times g_1}$ and $\boldsymbol{V}_{h_1 \times h_1}$, we can rewrite the model as $\boldsymbol{X}_t = \boldsymbol{L}\boldsymbol{U}(\boldsymbol{U}^{-1}\boldsymbol{F}_t\boldsymbol{V}^{-1})(\boldsymbol{R}\boldsymbol{V}')' + \boldsymbol{E}_t$. To address this identifiability issue, one can rewrite $\boldsymbol{L} = \boldsymbol{Q}_1\boldsymbol{W}_1$ and $\boldsymbol{R} = \boldsymbol{Q}_2\boldsymbol{W}_2$, where $\boldsymbol{Q}_1$ is a $g \times g_1$ semi-orthonormal matrix such that $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_{g_1}$ and $\boldsymbol{Q}_2$ is a $h \times h_1$ semi-orthonormal matrix such that $\boldsymbol{Q}_2'\boldsymbol{Q}_2 = \boldsymbol{I}_{h_1}$, and $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are $g_1 \times g_1$ and $h_1 \times h_1$ nonsingular matrices. The model in Equation (14) then becomes $\boldsymbol{X}_t = \boldsymbol{Q}_1(\boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{W}_2')\boldsymbol{Q}_2' + \boldsymbol{E}_t = \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \boldsymbol{E}_t$, where $\boldsymbol{Z}_t = \boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{W}_2'$. The orthonormal condition of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ enables us to consistently estimate the column spaces of $\boldsymbol{L}$ and $\boldsymbol{R}$. The authors then estimate the dynamic part of $\{\boldsymbol{X}_t\}$, defined as $\boldsymbol{Y}_t = \boldsymbol{L}\boldsymbol{F}_t\boldsymbol{R}'$, by

$$\hat{\boldsymbol{Y}}_t = \hat{\boldsymbol{Q}}_1\hat{\boldsymbol{Q}}_1'\boldsymbol{X}_t\hat{\boldsymbol{Q}}_2\hat{\boldsymbol{Q}}_2'. \tag{15}$$

Sufficient conditions are also given to derive some asymptotic properties of the proposed estimates. E. Chen *et al.* (2020) extend the factor models one-step further by studying constrained factor models for high-dimensional matrix-variate time series. The constraints are known *a priori*, such as the prior knowledge concerning the categories of the variables involved, and are used to obtain a parsimonious factor model when the dimension of $\boldsymbol{X}_t$ is high.

Gao & Tsay (2023) propose a two-way transformed factor model for matrix-variate time series. Assuming that $E(\boldsymbol{X}_t) = \boldsymbol{0}$, one can write the proposed model as

$$\boldsymbol{X}_t = \boldsymbol{L}\begin{bmatrix} \boldsymbol{F}_t & \boldsymbol{Z}_{12,t} \\ \boldsymbol{Z}_{21,t} & \boldsymbol{Z}_{22,t} \end{bmatrix}\boldsymbol{R}' = \boldsymbol{L}_1\boldsymbol{F}_t\boldsymbol{R}_1' + \boldsymbol{L}_2\boldsymbol{Z}_{21,t}\boldsymbol{R}_1' + \boldsymbol{L}_1\boldsymbol{Z}_{12,t}\boldsymbol{R}_2' + \boldsymbol{L}_2\boldsymbol{Z}_{22,t}\boldsymbol{R}_2', \tag{16}$$

where $\boldsymbol{F}_t$ is a $g_1 \times h_1$ factor process with $g_1 < g$ and $h_1 < h$, $\boldsymbol{Z}_{12,t}, \boldsymbol{Z}_{21,t}$ and $\boldsymbol{Z}_{22,t}$ are $g_1 \times h_2$, $g_2 \times h_1$ and $g_2 \times h_2$ matrix-variate white noise processes with $g_1 + g_2 = g$ and $h_1 + h_2 = h$, $\boldsymbol{L} = [\boldsymbol{L}_1, \boldsymbol{L}_2]$ is a $g \times g$ left transformation matrix with $\boldsymbol{L}_1$ being of dimension $g \times g_1$, and $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2]$ is a $h \times h$ right transformation matrix with $\boldsymbol{R}_1$ being of dimension $h \times h_1$. In model (16), $\boldsymbol{F}_t$ is uncorrelated with the three matrix-variate white noise processes. Note that the model in (16) is different from the model in (14), because the number of noise components in model (14) is gh whereas that in model (16) is only $gh - g_1h_1$. Decompose the transformation matrices $\boldsymbol{L}$ and $\boldsymbol{R}$ as follows:

$$\boldsymbol{L}_1 = \boldsymbol{K}_1\boldsymbol{W}_1,\ \boldsymbol{L}_2 = \boldsymbol{K}_2\boldsymbol{W}_2,\ \boldsymbol{R}_1 = \boldsymbol{P}_1\boldsymbol{G}_1,\ \text{ and } \boldsymbol{R}_2 = \boldsymbol{P}_2\boldsymbol{G}_2, \tag{17}$$

where $\boldsymbol{K}_i$ and $\boldsymbol{P}_i$ are semi-orthonomal matrices, that is, $\boldsymbol{K}_i'\boldsymbol{K}_i = \boldsymbol{I}_{g_i}$ and $\boldsymbol{P}_i'\boldsymbol{P}_i = \boldsymbol{I}_{h_i}$, for $i = 1$ and 2. This can be done via QR or singular value decomposition. Letting $\boldsymbol{Y}_t = \boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{G}_1'$, $\boldsymbol{E}_{21,t} = \boldsymbol{W}_2\boldsymbol{Z}_{21,t}\boldsymbol{G}_1'$, $\boldsymbol{E}_{12,t} = \boldsymbol{W}_1\boldsymbol{Z}_{12,t}\boldsymbol{G}_2'$ and $\boldsymbol{E}_{22,t} = \boldsymbol{W}_2\boldsymbol{Z}_{22,t}\boldsymbol{G}_2'$, we can rewrite the model (16) as

$$\boldsymbol{X}_t = \boldsymbol{K}_1\boldsymbol{Y}_t\boldsymbol{P}_1' + \boldsymbol{K}_2\boldsymbol{E}_{21,t}\boldsymbol{P}_1' + \boldsymbol{K}_1\boldsymbol{E}_{12,t}\boldsymbol{P}_2' + \boldsymbol{K}_2\boldsymbol{E}_{22,t}\boldsymbol{P}_2'. \tag{18}$$

Note that the model in (18) is still not identifiable, because the triplets $(\boldsymbol{K}_1, \boldsymbol{Y}_t, \boldsymbol{P}_1)$ can be replaced by $(\boldsymbol{K}_1\boldsymbol{H}_1', \boldsymbol{H}_1\boldsymbol{Y}_t\boldsymbol{H}_2, \boldsymbol{P}_1\boldsymbol{H}_2)$ for any orthonormal matrices $\boldsymbol{H}_1 \in R^{g_1 \times g_1}$ and $\boldsymbol{H}_2 \in R^{g_2 \times g_2}$ without altering the data generating process. Similarly to that of D. Wang *et al.* (2019), the orthonormal features of $\boldsymbol{K}_i$ and $\boldsymbol{P}_i$ enable us to estimate the column spaces of transformation matrices $\boldsymbol{L}$ and $\boldsymbol{R}$. On the other hand, Gao & Tsay (2023) propose a two-way projected principal component analysis to improve the estimation of the factor process. For $i = 1$ and 2, let $\boldsymbol{J}_i$ and $\boldsymbol{Q}_i$ be the orthonormal complements of $\boldsymbol{K}_i$ and $\boldsymbol{P}_i$, respectively. That is, $\boldsymbol{J}_1 \in R^{g \times g_2}$, $\boldsymbol{J}_2 \in R^{g \times g_1}$, $\boldsymbol{Q}_1 \in R^{h \times h_2}$ and $\boldsymbol{Q}_2 \in R^{h \times h_1}$ are semi-orthonomal matrices with $\boldsymbol{J}_1'\boldsymbol{K}_1 = \boldsymbol{0}$, $\boldsymbol{J}_2'\boldsymbol{K}_2 = \boldsymbol{0}$, $\boldsymbol{Q}_1'\boldsymbol{P}_1 = \boldsymbol{0}$, and $\boldsymbol{Q}_2'\boldsymbol{P}_2 = \boldsymbol{0}$. From (18), we have

$$\boldsymbol{J}_1'\boldsymbol{X}_t\boldsymbol{Q}_1 = \boldsymbol{J}_1'\boldsymbol{K}_2\boldsymbol{E}_{22,t}\boldsymbol{P}_2'\boldsymbol{Q}_1, \tag{19}$$

which implies that $\boldsymbol{J}_1'\boldsymbol{X}_t\boldsymbol{Q}_1$ is a matrix-variate white noise process. Also, from Equation (18), we have

$$\begin{aligned} \boldsymbol{J}_2'\boldsymbol{X}_t &= \boldsymbol{J}_2'\boldsymbol{K}_1\boldsymbol{Y}_t\boldsymbol{P}_1' + \boldsymbol{J}_2'\boldsymbol{K}_1\boldsymbol{E}_{12,t}\boldsymbol{P}_2', \\ \boldsymbol{X}_t\boldsymbol{Q}_2 &= \boldsymbol{K}_1\boldsymbol{Y}_t\boldsymbol{P}_1'\boldsymbol{Q}_2 + \boldsymbol{K}_2\boldsymbol{E}_{21,t}\boldsymbol{P}_1'\boldsymbol{Q}_2. \end{aligned}$$

Therefore, $\boldsymbol{J}_2'\boldsymbol{X}_t$ and $\boldsymbol{X}_t\boldsymbol{Q}_2$ are uncorrelated with $\boldsymbol{J}_1'\boldsymbol{X}_t\boldsymbol{Q}_1$ of Equation (19). Let $\boldsymbol{\Omega}_i = \text{cov}(\boldsymbol{x}_{i,t}, \boldsymbol{x}_t)$, where $\boldsymbol{x}_{i,t}$ is the i-column of $\boldsymbol{X}_t$ and $\boldsymbol{x}_t = \text{vec}(\boldsymbol{X}_t)$, where $i = 1, \ldots, h$. Also, let $\boldsymbol{\Delta}_{e22,ip} = \text{cov}(\boldsymbol{E}_{22,t}\boldsymbol{p}_{2,i.}', \text{vec}(\boldsymbol{E}_{22,t}))$, where $\boldsymbol{p}_{2,i.}$ is the ith row $\boldsymbol{P}_2$. Then, it follows from (18) and (19) that

$$\text{cov}(\boldsymbol{x}_{i,t}, \text{vec}(\boldsymbol{J}_1'\boldsymbol{X}_t\boldsymbol{Q}_1)) = \boldsymbol{\Omega}_i(\boldsymbol{Q}_1 \otimes \boldsymbol{J}_1) = \boldsymbol{K}_2\boldsymbol{\Delta}_{e22,ip}(\boldsymbol{P}_2'\boldsymbol{Q}_1 \otimes \boldsymbol{K}_2'\boldsymbol{J}_1). \tag{20}$$

Note that $\boldsymbol{J}_2'\boldsymbol{x}_{i,t}$ is uncorrelated with $\text{vec}(\boldsymbol{J}_1'\boldsymbol{X}_t\boldsymbol{Q}_1)$, for $i = 1, \ldots, h$. Therefore, define

$$\boldsymbol{S}_1 = \sum_{i=1}^{h} [\boldsymbol{\Omega}_i(\boldsymbol{Q}_1 \otimes \boldsymbol{J}_1)][\boldsymbol{\Omega}_i(\boldsymbol{Q}_1 \otimes \boldsymbol{J}_1)]', \tag{21}$$

from which we see that, via Equation (20), $\boldsymbol{S}_1\boldsymbol{J}_2 = \boldsymbol{0}$. Furthermore, the rank of $\boldsymbol{S}_1 \in R^{g \times g}$ is g_2 so that $\boldsymbol{J}_2$ contains all the eigenvectors corresponding to the zero eigenvalues of $\boldsymbol{S}_1$. Similarly, we can construct $\boldsymbol{S}_2$ such that $\boldsymbol{S}_2\boldsymbol{Q}_2 = \boldsymbol{0}$ and $\boldsymbol{Q}_2$ contains all the eigenvectors associated with the zero eigenvalues of $\boldsymbol{S}_2$. Finally, by Equation (18), we have $\boldsymbol{J}_2'\boldsymbol{X}_t\boldsymbol{Q}_2 = \boldsymbol{J}_2'\boldsymbol{K}_1\boldsymbol{Y}_t\boldsymbol{P}_1'\boldsymbol{Q}_2$ provided that $\boldsymbol{K}_1$, $\boldsymbol{P}_1$, $\boldsymbol{J}_2$ and $\boldsymbol{Q}_2$ are known. Consequently, we have the factor process

$$\boldsymbol{Y}_t = (\boldsymbol{J}_2'\boldsymbol{K}_1)^{-1}\boldsymbol{J}_2'\boldsymbol{X}_t\boldsymbol{Q}_2(\boldsymbol{P}_1'\boldsymbol{Q}_2)^{-1}, \tag{22}$$

where $\boldsymbol{J}_2'\boldsymbol{K}_1 \in R^{g_1 \times g_1}$ and $\boldsymbol{P}_1'\boldsymbol{Q}_2 \in R^{h_1 \times h_1}$ are two invertible matrices; see Gao & Tsay (2023) for further details. Asymptotic properties, such as consistency of the proposed estimates, are also derived. A contribution of the proposed two-way projected PCA is as follows. Recall that $\boldsymbol{Y}_t = \boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{G}_1'$, which does not involve the white noise matrices $\boldsymbol{Z}_{12,t}$, $\boldsymbol{Z}_{21,t}$ and $\boldsymbol{Z}_{22,t}$. On the other hand, from Equation (15), the estimated common factor process in D. Wang *et al.* (2019) contains the noise process $\{\boldsymbol{E}_t\}$.

Xiao *et al.* (2022) study reduced-rank MAR models. To highlight the idea, we consider the simple MAR(1) model in Equation (4) with $\boldsymbol{\Phi}_0 = \boldsymbol{0}$, that is,

$$\boldsymbol{X}_t = \boldsymbol{C}\boldsymbol{X}_{t-1}\boldsymbol{D}' + \boldsymbol{A}_t. \tag{23}$$

Suppose that $\text{Rank}(\boldsymbol{C}) = k_1 < g$ and $\text{Rank}(\boldsymbol{D}) = k_2 < h$. The authors proposed two methods for estimation and derived some asymptotic properties of the estimates. The first estimation method is the alternating reduced rank regression. Suppose that $\boldsymbol{D}$ is given, one can rewrite the model as

$$\boldsymbol{x}_{j,\,t} = \boldsymbol{C}[\boldsymbol{X}_{t-1}\boldsymbol{d}'_{j.}] + \boldsymbol{a}_{j,\,t},\ \ j = 1,\ \dots,\ h, \tag{24}$$

where $\boldsymbol{x}_{j,\,t}$ and $\boldsymbol{a}_{j,\,t}$ are the j-th column of $\boldsymbol{X}_t$ and $\boldsymbol{A}_t$, respectively, and $\boldsymbol{d}_{j.}$ denotes the j-th row of $\boldsymbol{D}$. In this equation, $[\boldsymbol{X}_{t-1}\boldsymbol{d}'_{j.}]$ are the predictors and $\boldsymbol{x}_{j,\,t}$ is the response vector so that the problem becomes a reduced-rank regression. One can then consider the h equations in (24) jointly to estimate $\boldsymbol{C}$ with rank k_1. The same idea applies in estimating $\boldsymbol{D}$ if $\boldsymbol{C}$ is given. The second estimation method is the maximum likelihood method for which $\boldsymbol{A}_t$ is assumed to follow a matrix normal distribution $N(\boldsymbol{0},\ \boldsymbol{U},\ \boldsymbol{V})$. Again, updating procedures are available to compute the estimates of $\boldsymbol{C}$ and $\boldsymbol{D}$ for given ranks k_1 and k_2. The authors further proposed an extended Bayesian information criterion to select the ranks k_1 and k_2.

Han *et al.* (2022) propose a bilinear transformation for matrix-variate time series so that the transformed series assumes certain block structures such that the dynamic dependence of the transformed series only occurs within individual blocks. In this way, one can simplify the complexity in modelling matrix-variate time series. The authors further demonstrated that the proposed segmentation technique can improve the accuracy in forecasting. Recently, Hsu *et al.* (2023) extend the spatio-temporal model of Hsu *et al.* (2021) to a rank-r MAR(p) model, which assumes the form

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_0 + \sum_{i=1}^{p}\sum_{j=1}^{r}\sigma_{i,j}\boldsymbol{C}_{i,j}\boldsymbol{X}_{t-i}\boldsymbol{D}'_{i,j} + \boldsymbol{A}_t, \tag{25}$$

where $r \geq 1$ is a positive integer, $\boldsymbol{C}_{i,j}$ and $\boldsymbol{D}_{i,j}$ are $g \times g$ and $h \times h$ matrices, characterising the row and column effects of $\boldsymbol{X}_{t-i}$ on $\boldsymbol{X}_t$ for the j-th component of Kronecker product approximation, and $\sigma_{i,j}$ is a scalar, for $i = 1,\ \dots,\ p$ and $j = 1,\ \dots,\ r$. As mentioned before, for model identifiability, they assume that $[\mathrm{vec}(\boldsymbol{C}_{i,\,1}),\ \dots,\ \mathrm{vec}(\boldsymbol{C}_{i,\,r})]$ and $[\mathrm{vec}(\boldsymbol{D}_{i,\,j}),\ \dots,\ \mathrm{vec}(\boldsymbol{D}_{i,\,r})]$ are two sets of orthonormal vectors and $\sigma_{i,\,1} \geq \dots \geq \sigma_{i,\,r} > 0$, for $i = 1,\ \dots,\ p$. A modified alternating direction method of multipliers (ADMM) algorithm is proposed for parameter estimation via maximum likelihood. Note that this model is in the framework of (6) with $q = 0$, except that the covariance matrix of $\boldsymbol{a}_t$ is given in (13). Other related works of matrix-variate time series analysis include Walden & Serroukh (2017), which considers wavelet analysis of the series.

2.4 Properties of MARMA Models

From Section 2.3, we see that most of the works on matrix-variate time series focus either on MAR models or on dimension reduction. In this section, we leverage the connection between the MARMA model in (4) and the VARMA model in (3) to obtain some basic properties of MARMA processes. Properties for the general MARMA models in (6) can be similarly obtained. We assume that the identifiability conditions of MARMA models discussed in Section 2.1 hold in this section.

Property 1: Stationarity and invertibility. The MARMA($p,\ q$) process of (4) is weakly stationary if all solutions to $|\boldsymbol{\phi}_{C,\,D}(B)| = |\boldsymbol{I} - \sum_{i=1}^{p}(\boldsymbol{D}_i \otimes \boldsymbol{C}_i)B| = 0$ are outside the unit circle. The process is invertible if all solutions to $|\boldsymbol{\theta}_{L,\,R}(B)| = |\boldsymbol{I} - \sum_{j=1}^{q}(\boldsymbol{R}_j \otimes \boldsymbol{L}_j)B^j| = 0$ are outside the unit circle.

For an MAR(1) model, the stationarity condition is equivalent to that all eigenvalues of $(\boldsymbol{D}_1 \otimes \boldsymbol{C}_1)$ are less than 1 in modulus, which hold if $\rho(\boldsymbol{D}_1) < 1$ and $\rho(\boldsymbol{C}_1) < 1$. See proposition 1 in Chen *et al.* (2021). In this case, the MAR(1) model has the matrix moving-average (MMA) representation

$$\tilde{\boldsymbol{X}}_t = \boldsymbol{A}_t + \sum_{j=1}^{\infty} \boldsymbol{C}_1^j \boldsymbol{A}_{t-j} (\boldsymbol{D}_1')^j,$$

where $\tilde{\boldsymbol{X}}_t = \boldsymbol{X}_t - E(\boldsymbol{X}_t)$ is the mean-adjusted series. Similarly, an MMA(1) model is invertible if all eigenvalues of $(\boldsymbol{R}_1 \otimes \boldsymbol{L}_1)$ are less than 1 in modulus. In this case, the process has the MAR representation

$$\tilde{\boldsymbol{X}}_t + \sum_{i=1}^{\infty} \boldsymbol{L}_1^i \tilde{\boldsymbol{X}}_{t-i} (\boldsymbol{R}_1')^i = \boldsymbol{A}_t.$$

For the general MARMA(p, q) model, there is no simple matrix expression for MAR(∞) or MMA(∞) representations. However, one can use its VARMA representation to obtain an expression that links $\boldsymbol{X}_t$ to its past values $\{\boldsymbol{X}_{t-j} | j = 1, \ldots\}$ or an expression that links $\boldsymbol{X}_t$ to its past innovations $\{\boldsymbol{A}_{t-j} | j = 1, \ldots\}$.

For moment properties, taking expectations in Equation (4), we have

$$\boldsymbol{\Xi} \equiv E(\boldsymbol{X}_t) = \boldsymbol{\Phi}_0 + \sum_{i=1}^{p} \boldsymbol{C}_i \boldsymbol{\Xi} \boldsymbol{D}_i'. \tag{26}$$

Therefore,

$$\boldsymbol{\xi} = \left[\boldsymbol{I} - \sum_{i=1}^{p} (\boldsymbol{D}_i \otimes \boldsymbol{C}_i) \right]^{-1} \boldsymbol{\phi}_0,$$

where $\boldsymbol{\xi} = \mathrm{vec}(\boldsymbol{\Xi}) = E(\boldsymbol{x}_t)$ and $\boldsymbol{\phi}_0 = \mathrm{vec}(\boldsymbol{\Phi}_0)$. This is precisely the result of taking expectations in the VARMA representation in Equation (5). Using Equation (26), we can rewrite the MARMA model as

$$\tilde{\boldsymbol{X}}_t = \sum_{i=1}^{p} \boldsymbol{C}_i \tilde{\boldsymbol{X}}_{t-i} \boldsymbol{D}_i' + \boldsymbol{A}_t - \sum_{j=1}^{q} \boldsymbol{L}_j \boldsymbol{A}_{t-j} \boldsymbol{R}_j'.$$

For simplicity, we assume that $E(\boldsymbol{X}_t) = \boldsymbol{0}$ if $\{\boldsymbol{X}_t\}$ is stationary. There are two ways to define the autocovariance matrices of a stationary $\{\boldsymbol{X}_t\}$. The first approach is to use the vector process $\{\boldsymbol{x}_t\}$ so that

$$\boldsymbol{\Gamma}_j = \mathrm{cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t-j}) = E(\boldsymbol{x}_t \boldsymbol{x}_{t-j}'), \; j = 0, \pm 1, \ldots. \tag{27}$$

It is easy to see that the v-th column of $\boldsymbol{\Gamma}_j$ shows the linear dependence of $\boldsymbol{x}_t$ on the v-th element of $\boldsymbol{x}_{t-j}$. The second approach is

$$\tilde{\boldsymbol{\Gamma}}_j = \mathrm{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-j}) = E(\boldsymbol{X}_t \otimes \boldsymbol{X}_{t-j}'), \tag{28}$$

which is a $g \times h$ block-matrix with each block of size $h \times g$. The (u, v)-block of $\tilde{\boldsymbol{\Gamma}}_j$ shows the linear dependence of $x_{uv,t}$ on the past lagged matrix $\boldsymbol{X}_{t-j}'$. Clearly, $\boldsymbol{\Gamma}_j$ and $\tilde{\boldsymbol{\Gamma}}_j$ contain the same elements so that there is a one-to-one mapping from $\tilde{\boldsymbol{\Gamma}}_j$ to $\boldsymbol{\Gamma}_j$. Specifically, writing the u-th row of $\boldsymbol{\Gamma}_j$ as $u = i + (v-1)g$, for $i = 1, \ldots, g$ and $v = 1, \ldots, h$, then the u-th row of $\boldsymbol{\Gamma}_j$ is $\mathrm{vec}(E(x_{iv,t} \boldsymbol{X}_{t-j}'))$ with $E(x_{iv,t} \boldsymbol{X}_{t-j}')$ being the (i, v)th block of $\tilde{\boldsymbol{\Gamma}}_j$. A nice feature of $\boldsymbol{\Gamma}_j$ is that its diagonal elements are the lag-j autocovariances of elements of $\boldsymbol{X}_t$ and this feature does not hold for $\tilde{\boldsymbol{\Gamma}}_j$. We define the lag-$j$ cross-correlation matrix (CCM) of $\{\boldsymbol{X}_t\}$ as

$$\boldsymbol{\rho}_j = \boldsymbol{G}^{-1/2}\boldsymbol{\Gamma}_j\boldsymbol{G}^{-1/2}, \tag{29}$$

where $\boldsymbol{G}$ is a diagonal matrix consisting of the diagonal elements of $\boldsymbol{\Gamma}_0$. In this way, $\boldsymbol{\rho}_j$ is the lag-j CCM of $\{\boldsymbol{x}_t\}$.

Property 2: Generalised Yule–Walker equation. Using $\boldsymbol{\Gamma}_j$, the generalised Yule–Walker equations for $\{\boldsymbol{X}_t\}$ can be written as

$$\boldsymbol{\Gamma}_\ell = \sum_{i=1}^{p}(\boldsymbol{D}_i \otimes \boldsymbol{C}_i)\boldsymbol{\Gamma}_{\ell - i}, \;\; \ell = q+1, \ldots, q+p.$$

Other properties such as the π-weights, ψ-weights and impulse response functions of $\{\boldsymbol{X}_t\}$ can also be defined as those of $\{\boldsymbol{x}_t\}$. In particular, the ψ-weights can be used to compute the standard errors of predictions. Predictions of an MARMA model can be computed recursively either using the matrix model or its corresponding vector model. The variances of forecast errors are easier to compute using the VARMA representation.

2.5 Estimation

For simplicity, we only consider the estimation of the MARMA model in (4) in this section. The estimation of the general MARMA model in (6) is much more involved and is left for future research. For MAR models in Equation (4), the estimation can be carried out by either an iterated least squares method or the maximum likelihood method. See the review in Section 2.3. For the MARMA(p, q) model in (4), we consider a conditional maximum likelihood method. Let $t_o = \max\{p, q\}$ and suppose the data available are $\{\boldsymbol{X}_t\}_{t=1}^{T}$. Conditional on $\{\boldsymbol{X}_t\}_{t=1}^{t_o}$ and $\boldsymbol{A}_t = \boldsymbol{0}$, for $t = 1, \ldots, t_o$, one can compute $\boldsymbol{A}_t$ recursively from model (4). If one assumes that $\boldsymbol{A}_t$ follows $N(\boldsymbol{0}, \boldsymbol{U}, \boldsymbol{V})$, then the log likelihood function becomes

$$\ell(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) \propto -\frac{1}{2}(T - t_o)(g\ln|\boldsymbol{V}| + h\ln|\boldsymbol{U}|) - \sum_{t=t_o+1}^{T} tr(\boldsymbol{V}^{-1}\boldsymbol{A}_t'\boldsymbol{U}^{-1}\boldsymbol{A}_t), \tag{30}$$

where $\boldsymbol{Y}$ denotes the collection of parameters $\boldsymbol{\Phi}_0$, $\{\boldsymbol{D}_i, \boldsymbol{C}_i\}_{i=1}^{p}$, $\{\boldsymbol{R}_j, \boldsymbol{L}_j\}_{j=1}^{q}$. The maximum likelihood estimates of $\boldsymbol{Y}$ can then be obtained by maximising $\ell(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V})$ subject to the identification constraints $||\boldsymbol{C}_i||_2 = 1$, for $i = 1, \ldots, p$, and $||\boldsymbol{L}_j||_2 = 1$, for $j = 1, \ldots, q$. To fix the issue of sign identification, we set the maximum element in each pair $(\hat{\boldsymbol{D}}_i, \hat{\boldsymbol{C}}_i)$ and $(\hat{\boldsymbol{L}}_j, \hat{\boldsymbol{R}}_j)$ as being fixed and compute the standard errors of the other estimates. The estimates of $\boldsymbol{U}$ and $\boldsymbol{V}$ can be calculated by the residual series as $\boldsymbol{U}$ being the row covariance matrix and $\boldsymbol{V}$ the column covariance matrix of $\boldsymbol{A}_t$.

Instead of assuming that $\boldsymbol{A}_t$ follows a matrix normal distribution, one can assume that $\boldsymbol{a}_t$ is normally distributed and estimates a general covariance matrix of $\boldsymbol{a}_t$. Here $\boldsymbol{a}_t$ can also be recursively calculated using the VARMA representation in Equation (5). The likelihood function is then similar to that for the VARMA process $\{\boldsymbol{x}_t\}$, even though, except for the constant term $\boldsymbol{\Phi}_0$, the likelihood function is nonlinear for all coefficient parameters. The limiting distributions of the maximum likelihood estimates can be established using the results of VARMA models. See, for instance, Tsay (2014, section 3.11).

Note that one can estimate a VARMA(p, q) model for the $\{\boldsymbol{x}_t\}$ process to obtain (unconstrained) estimates of the AR and MA coefficient matrices. Those estimates can then be used, via the Kronecker product approximation, to obtain initial estimates of $\boldsymbol{D}_i$, $\boldsymbol{C}_i$, $\boldsymbol{L}_j$, and $\boldsymbol{R}_j$. However, if gh is sufficiently large, then a penalised (or regularised) estimation is needed to estimate the VARMA model.

3 Seasonal Matrix-Variate Time Series

As shown in Section 2.3, a majority of literature on matrix-variate time series analysis focus on MAR models, but there are situations in which MMA models are needed. One of such applications is modelling matrix-variate seasonal time series. Figure 1 shows time plots of twelve monthly seasonal time series, in log scale, concerning two labour-market data of the six US mid-western states mentioned in Section 1. The two employment series are (a) the civilian labour force (in persons) and (b) all employments in leisure and hospitality (in thousands of persons). The six states are, in row order, Illinois, Indiana, Iowa, Wisconsin, Minnesota and Michigan. The data span is from January 1990 to January 2020, for $T = 361$. The data are available from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of St Louis and are not seasonally adjusted. They naturally form a 6×2 matrix-variate time series. All series show an upward trend and strong seasonality. Following the traditional time series analysis, one can take the regular and seasonal differences and consider the process $\{\boldsymbol{W}_t = (1 - B)(1 - B^{12})\boldsymbol{X}_t\}$, where $\boldsymbol{X}_t$ denotes the 6×2 matrix time series of log employment data. As expected, the sample cross-correlation matrices of $\{\boldsymbol{W}_t\}$ show the existence of high cross-correlations in lags 1, 11, 12, and 13, indicating that the seasonal features of scalar monthly time series continue to hold for the matrix-variate seasonal time series. Motivated by such an application, we discuss some useful seasonal models for matrix-variate time series.

3.1 Exponential Smoothing Models

One of the widely used scalar or vector models in forecasting is the exponential smoothing model. For matrix-variate series, the model can be written as

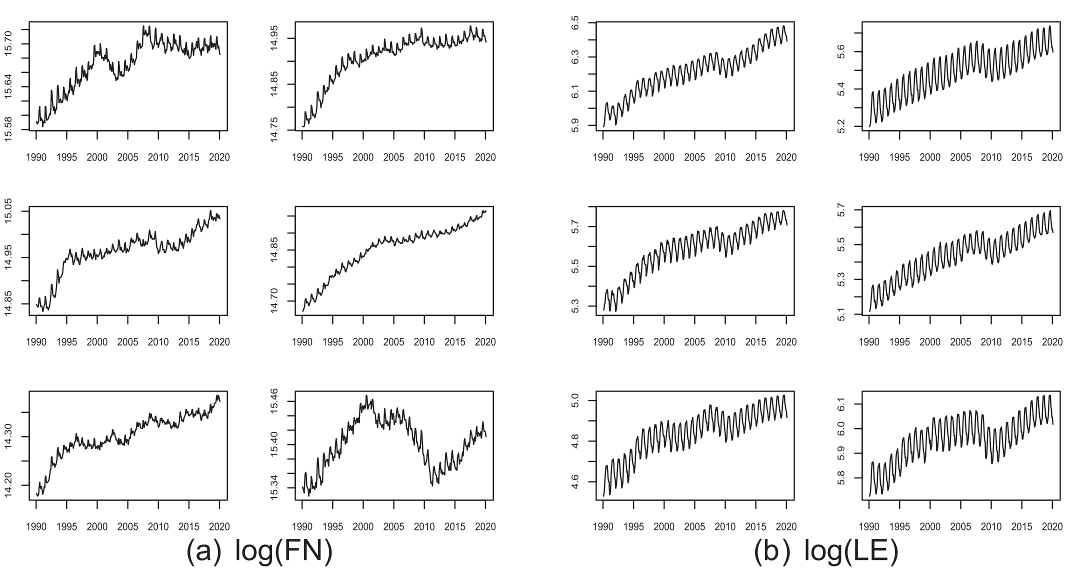


Figure 1. *Time plots of log series of (a) civilian labour force and (b) all employees in leisure and hospitality of six US mid-western states from January 1990 to January 2020*

$$(1 - B)\boldsymbol{X}_t = \boldsymbol{A}_t - \boldsymbol{L}_1\boldsymbol{A}_{t-1}\boldsymbol{R}_1', \tag{31}$$

where, for simplicity, we use r=1 in Equation (6) and assume that $\rho(\boldsymbol{L}_1) < 1$ and $\rho(\boldsymbol{R}_1) < 1$. By repeated substitutions, it is easy to see that

$$\boldsymbol{X}_t = \boldsymbol{A}_t + \sum_{j=1}^{\infty} \boldsymbol{L}_1^{j-1}\left(\boldsymbol{X}_{t-j} - \boldsymbol{L}_1\boldsymbol{X}_{t-j}\boldsymbol{R}_1'\right)\left(\boldsymbol{R}_1^{j-1}\right)'.$$

This is a matrix version of the exponential smoothing model and is equivalent to

$$\boldsymbol{x}_t = \boldsymbol{a}_t + \sum_{j=1}^{\infty} \boldsymbol{\theta}_1^{j-1}(\boldsymbol{I} - \boldsymbol{\theta}_1)\boldsymbol{x}_{t-j},$$

for the vectorised process $\{\boldsymbol{x}_t\}$ with discounting matrix $\boldsymbol{\theta}_1 = \boldsymbol{R}_1 \otimes \boldsymbol{L}_1$.

For seasonal matrix time series, the seasonal exponential smoothing model can be written as

$$(1 - B^s)\boldsymbol{X}_t = \boldsymbol{A}_t - \boldsymbol{L}_s\boldsymbol{A}_{t-s}\boldsymbol{R}_s', \tag{32}$$

where s is the periodicity of the annual cycles. In this case, we have

$$\boldsymbol{X}_t = \boldsymbol{A}_t + \sum_{j=1}^{\infty} \boldsymbol{L}_s^{j-1}\left(\boldsymbol{X}_{t-js} - \boldsymbol{L}_s\boldsymbol{X}_{t-js}\boldsymbol{R}_s'\right)\left(\boldsymbol{R}_s^{j-1}\right)',$$

for which the annual discount matrix is $\boldsymbol{\theta}_s = \boldsymbol{R}_s \otimes \boldsymbol{L}_s$.

3.2 Multiplicative Seasonal Models

Analogously to the well-known *Airline model* for scalar seasonal time series, we have the matrix version of the multiplicative seasonal model as

$$(1 - B)(1 - B^s)\boldsymbol{X}_t = \boldsymbol{A}_t - \boldsymbol{L}_1\boldsymbol{A}_{t-1}\boldsymbol{R}_1' - \boldsymbol{L}_s\boldsymbol{A}_{t-s}\boldsymbol{R}_s' + \boldsymbol{L}_1\boldsymbol{L}_s\boldsymbol{A}_{t-s-1}\boldsymbol{R}_s'\boldsymbol{R}_1', \tag{33}$$

which corresponds to the model

$$(1 - B)(1 - B^s)\boldsymbol{x}_t = (\boldsymbol{I} - \boldsymbol{\theta}_1 B)(\boldsymbol{I} - \boldsymbol{\theta}_s B^s)\boldsymbol{a}_t,$$

where $\boldsymbol{\theta}_1 = \boldsymbol{R}_1 \otimes \boldsymbol{L}_1$ and $\boldsymbol{\theta}_s = \boldsymbol{R}_s \otimes \boldsymbol{L}_s$, where the identity $\boldsymbol{R}_1\boldsymbol{R}_s \otimes \boldsymbol{L}_1\boldsymbol{L}_s = (\boldsymbol{R}_1 \otimes \boldsymbol{L}_1)(\boldsymbol{R}_s \otimes \boldsymbol{L}_s)$ is used. Similarly to the vector case, another multiplicative seasonal model is

$$(1 - B)(1 - B^s)\boldsymbol{X}_t = \boldsymbol{A}_t - \boldsymbol{L}_1\boldsymbol{A}_{t-1}\boldsymbol{R}_1' - \boldsymbol{L}_s\boldsymbol{A}_{t-s}\boldsymbol{R}_s' + \boldsymbol{L}_s\boldsymbol{L}_1\boldsymbol{A}_{t-s-1}\boldsymbol{R}_1'\boldsymbol{R}_s',$$

which corresponds to

$$(1 - B)(1 - B^s)\boldsymbol{x}_t = (\boldsymbol{I} - \boldsymbol{\theta}_s B^s)(\boldsymbol{I} - \boldsymbol{\theta}_1 B)\boldsymbol{a}_t.$$

In applications, some multiplicative AR polynomial matrices might be needed, in addition to the MMA polynomial matrices mentioned above, to effectively model seasonal matrix time series. Details are omitted.

4 Empirical Examples

We demonstrate the analysis of matrix-variate time series by two empirical examples in this section. More complicated models can be found in the literature. For instance, Hsu et al. (2023) fitted a rank-2 MAR(6) model to a 17-by-17 matrix series of wind speed in the north-western Pacific Ocean.

Example *Consider the* 6×2 *matrix-variate time series of Figure 1. The upward trend and strong seasonality lead us to consider the differenced process* $\{\boldsymbol{W}_t = (1 - B)(1 - B^{12})\boldsymbol{X}_t\}$, *where* $\boldsymbol{X}_t$ *denotes the log series of labour market data, in which each row represents one state and each column a labour market series. The cross-correlation matrices of* $\{\boldsymbol{w}_t\}$ *show relatively strong cross-correlations at lags 1, 11, 12, and 13 so that the* airline model *of Equation (33) is entertained. The estimates of the model are*

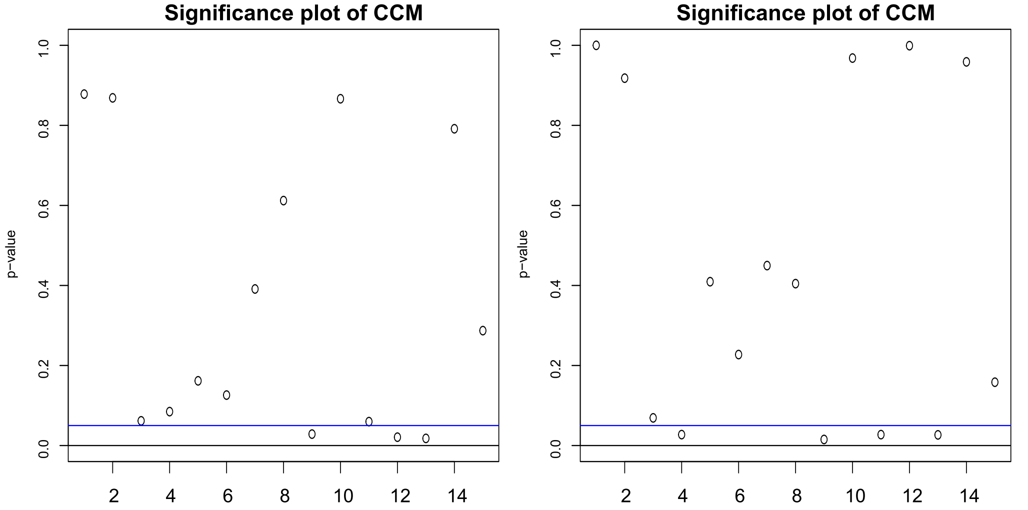
$$\hat{\boldsymbol{L}}_1 = \begin{bmatrix} \boldsymbol{-0.35} & 0.10 & -0.13 & 0.19 & 0.07 & 0.06 \\ -0.02 & \boldsymbol{-0.29} & 0.05 & -0.00 & 0.09 & 0.06 \\ -0.10 & -0.05 & \boldsymbol{-0.17} & -0.07 & 0.07 & \boldsymbol{0.25} \\ 0.15 & \boldsymbol{0.19} & -0.03 & \boldsymbol{-0.49} & 0.01 & -0.00 \\ 0.07 & 0.14 & \boldsymbol{-0.23} & 0.12 & -0.21 & 0.08 \\ \boldsymbol{-0.22} & \boldsymbol{0.21} & 0.06 & 0.06 & 0.12 & \boldsymbol{-0.20} \end{bmatrix}, \quad \hat{\boldsymbol{R}}_1 = \begin{bmatrix} \boldsymbol{0.34} & -0.03 \\ 0.21 & \boldsymbol{-0.52} \end{bmatrix}$$

$$\hat{\boldsymbol{L}}_s = \begin{bmatrix} \boldsymbol{0.46} & -0.01 & -0.01 & \boldsymbol{-0.03} & 0.02 & -0.02 \\ \boldsymbol{0.05} & \boldsymbol{0.42} & \boldsymbol{-0.05} & -0.03 & 0.02 & -0.02 \\ -0.02 & 0.00 & \boldsymbol{0.42} & 0.00 & -0.05 & 0.02 \\ \boldsymbol{0.04} & 0.02 & \boldsymbol{-0.04} & \boldsymbol{0.39} & -0.02 & -0.00 \\ 0.02 & 0.02 & \boldsymbol{-0.05} & -0.01 & \boldsymbol{0.40} & 0.02 \\ 0.04 & 0.02 & \boldsymbol{-0.04} & 0.04 & 0.04 & \boldsymbol{0.32} \end{bmatrix}, \quad \hat{\boldsymbol{R}}_s = \begin{bmatrix} \boldsymbol{1.92} & -0.00 \\ \boldsymbol{0.21} & \boldsymbol{1.54} \end{bmatrix},$$

where the condition that $\|\boldsymbol{L}_1\|_2 = \|\boldsymbol{L}_s\|_2 = 1$ is imposed and the boldfaced estimates are asymptotically significant at the 10% level. It is clear that larger coefficient estimates appear mainly in the diagonal elements of coefficient matrices, especially of the seasonal lag. Figure 2a shows the p -values of multivariate Ljung–Box statistics for residuals being serially uncorrelated. The plot shows that the matrix-airline model seems to fit the data reasonably well, even though some minor residual cross-correlations exist at the seasonal lag. One possible improvement is to use a rank-2 Kronecker product approximation for the coefficient matrix of the seasonal lag. The AIC and BIC statistics of the fitted model are -135.04 and -134.15, respectively.

From the estimate $\hat{\boldsymbol{L}}_s$, the seasonal parts of the model seem to show the effects of adjacent states. For instance, the first row of $\hat{\boldsymbol{L}}_s$ shows the significant dependence of Illinois on Wisconsin, the fourth row indicates Wisconsin depends significantly on Illinois and Iowa, and the fifth row shows Minnesota depends significantly on Iowa. It is also interesting to see that, from $\hat{\boldsymbol{L}}_s$, Iowa does not depend on other states, yet it has negative impacts on other states except Illinois. On the other hand, from $\hat{\boldsymbol{R}}_1$, the insignificant off-diagonal estimates indicate that the two employment variables used seem to have weak lag-1 cross-dependence between them.

For comparison, we also fit an unrestricted airline model to the corresponding vector process $\{\boldsymbol{x}_t\}$, which is of dimension 12. This model employs 288 coefficient parameters and Figure 2b shows the p-values of multivariate Ljung–Box statistics of its residuals. From the plot, it seems that the un-restricted model does not improve the fit over the matrix-variate model. The AIC and BIC statistics of the vector model are -135.01 and -131.82, respectively. Therefore, the information criteria confirm the choice of the multiplicative matrix seasonal model.



(a) Matrix model (b) Vector model

Figure 2. *Plots of p-values for multivariate Ljung–Box test statistics for residuals being serially uncorrelated. The x-axis denotes the number of lags used.*

Example *Consider the monthly imports and exports of goods of four Scandinavian countries, namely, Norway, Sweden, Denmark, and Finland, from January 1960 to December 2019 for 720 observations. The series are in US dollars, seasonally adjusted, and, again, available from FRED at https://fred.stlouisfed.org/categories/32264. We employ the growth rate series in our analysis, that is, the first difference of the log series. In addition, we remove the sample mean from each growth rate series so that no constant terms are needed in the modelling exercise. Figure 3 shows the time plots of the eight mean-adjusted growth rates of imports and exports of the four countries. The left panel shows the growth rates of monthly imports and the right panel those of monthly exports. For ease in viewing the plots, we add 3, 2, 1 to the top three plots so that the series can be separated. The plots are, from top to bottom, for Norway, Sweden, Denmark, and Finland, respectively. Since the number of series is only eight, one can easily treat the data as an 8-dimensional vector time series or as a* 4×2 *matrix-variate time series. This enables us to compare between VARMA and MARMA models.*

For VARMA modelling, if VAR models are entertained, then AIC and BIC statistics select VAR(5) and VAR(2), respectively. If VARMA models are entertained, then the VARMA(1,1) is selected by both AIC and BIC. Indeed, the two criterion functions prefer the VARMA(1,1) model over VAR models with the minimum AIC of -44.86 and the minimum BIC of -44.05 for the VARMA(1,1) model. For MARMA modelling, we entertained MAR(d) for $d = 1, \ldots, 5$, and MARMA(1,1), MARMA(2,1), MARMA(1,2), and MARMA(5,1) models. Both AIC and BIC statistics select the MARMA(1,1) model with AIC and BIC being -44.87 and -44.61, respectively. These values are slightly lower than those of the VARMA(1,1) model. Thus, in this particular instance, the MARMA model is preferred over the VARMA models. Note that the VARMA(1,1) model contains 128 coefficient parameters whereas the MARMA (1,1) model only uses 40. Model checking indicates that there exist some residual cross-correlations at lags 12 and 24, indicating that some seasonality remains in the data even

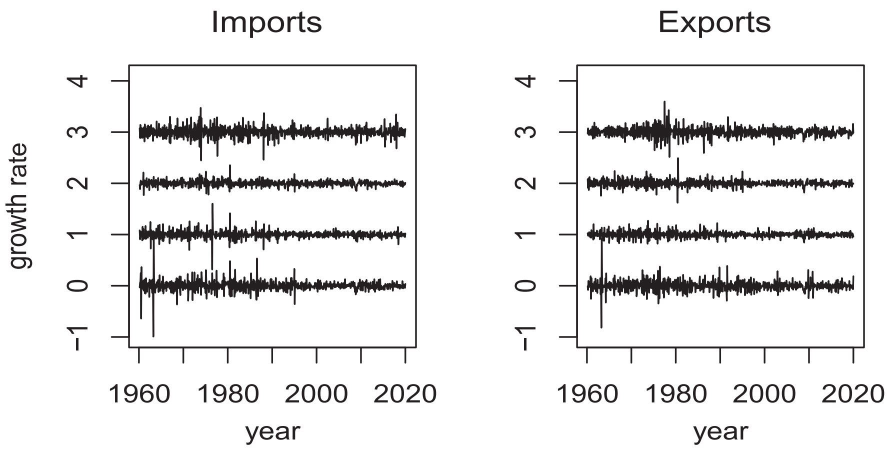


Figure 3. *The mean-adjusted growth rates of monthly imports and exports of Norway, Sweden, Denmark, and Finland from February 1960 to December 2019. The original data are in US dollars and are seasonally adjusted. We added 3, 2, and 1 to the top three series to separate the series in the plots.*

though they are seasonally adjusted. The fitted coefficient matrices of the MARMA(1,1) model are given below, where the boldfaced estimates are asymptotically significant at the 10% level.

$$\hat{\boldsymbol{C}}_1 = \begin{bmatrix} -0.085 & \boldsymbol{\mathit{0.315}} & -0.094 & 0.013 \\ \mathbf{0.130} & \mathbf{0.576} & 0.083 & 0.011 \\ \mathbf{0.151} & \mathbf{0.347} & -0.000 & 0.054 \\ \mathbf{0.220} & \mathbf{0.574} & 0.039 & 0.064 \end{bmatrix}, \quad \hat{\boldsymbol{D}}_1 = \begin{bmatrix} \mathbf{0.481} & 0.073 \\ \boldsymbol{\mathit{0.605}} & -0.077 \end{bmatrix},$$

$$\hat{\boldsymbol{L}}_1 = \begin{bmatrix} \boldsymbol{\mathit{0.455}} & \boldsymbol{\mathit{-0.107}} & \boldsymbol{\mathit{-0.081}} & -0.021 \\ 0.001 & \boldsymbol{\mathit{0.473}} & -0.022 & -0.015 \\ \boldsymbol{\mathit{-0.025}} & -0.027 & \boldsymbol{\mathit{0.481}} & -0.002 \\ -0.011 & \boldsymbol{\mathit{-0.112}} & -0.024 & \boldsymbol{\mathit{0.551}} \end{bmatrix}, \quad \hat{\boldsymbol{R}}_1 = \begin{bmatrix} \boldsymbol{\mathit{1.446}} & \boldsymbol{\mathit{-0.079}} \\ -0.015 & \boldsymbol{\mathit{1.384}} \end{bmatrix}.$$

It is interesting to see that for the MA coefficient matrices, most large and significant estimates are the diagonal elements of $\boldsymbol{L}_1$ and $\boldsymbol{R}_1$, indicating that the serial dependence is relatively strong within each series. The AR coefficient matrix $\hat{\boldsymbol{C}}_1$ suggests that Norway and Sweden play a more dominant role in the international trade of the four countries. The insignificance of estimates in column 2 of $\hat{\boldsymbol{D}}_1$ in conjunction with significant estimates in $\hat{\boldsymbol{R}}_1$ indicates that the contributions of lag-1 export growth rates to the series are mainly short-term impacts.

5 Concluding Remarks

We proposed a general rank-r matrix autoregressive moving-average model for matrix-variate time series and discussed its model identifiability. This enables us to leverage the relationship between matrix time series and its VARMA representation to obtain properties of the matrix series. We also considered seasonal matrix series and generalised the multiplicative seasonal models to the matrix case. The use of MARMA models is demonstrated by two empirical examples. On the other hand, it remains open to study the estimation of the general rank-r

MARMA model. It is also of interest to further investigate the difference between VARMA and MARMA models in applications, especially when the dimension of the matrix series under study is not large.

Data Availability Statement

Data are available online.

References

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. (2016). *Time Series Analysis: Forecasting and Control*, 5. Wiley and Sons: Hoboken, NJ.

Cai, C., Chen, R. & Xiao, H. (2022). KoPA: Automated kronecker product approximation. *Journal of Machine Learning Research*, **23**, 1–44.

Chen, E., Chen, R. & Tsay, R.S. (2020). Constrained factor model for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, **115**, 775–793.

Chen, R., Xiao, H. & Yang, D. (2021). Autoregressive models for matrix-variate time series. *Journal of Econometrics*, **222**, 539–560.

Dunsmuir, W. & Hannan, E.J. (1976). Vector linear time series models. *Advances in Applied Probability*, **8**, 339–364.

Gao, Z. & Tsay, R.S (2023). A two-way transformed factor model for matrix-variate time series. *Econometrics and Statistics*, **27**, 83–101.

Han, Y., Chen, R. & Zhang, C.H. (2022). Simultaneous decorrelation of matrix time series. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2022.2151448

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

Hsu, N.J., Huang, H.C. & Tsay, R.S. (2021). Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics*, **30**, 1143–1155. https://doi.org/10.1080/10618600.2021.1938587

Hsu, N.J., Huang, H.C., Tsay, R.S. & Kao, T.C (2023). *Rank-R bilinear autoregressive models for spatio-temporal matrix data*. to appear.

Li, Z. & Han, X. 2021. Multi-linear tensor autoregressive models. arXiv:2110.00928 (stat.ME).

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer: Berlin, Germany.

Samadi, S.Y. (2014). Matrix Time Series Analysis. PhD thesis, Department of Statistics University of Georgia, Athens, GA.

Shumway, R.H. & Stoffer, D.S. (2017). *Time Series Analysis and its Applications: with R examples*, 4. Springer: New York.

Tiao, G.C. & Tsay, R.S. (1989). Model specification in multivariate time series (with discussion). *Journal of the Royal Statistical Society. Series B*, **51**, 157–213.

Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3. Wiley and Sons: Hoboken, NJ.

Tsay, R.S. (2014). *Multivariate Time Series Analysis with R and Financial Applications*. Wiley and Sons: Hoboken, NJ.

Tzeng, S. & Huang, H.C. (2018). Resolution adaptive fixed rank kriging. *Technometrics*, **60**, 198–208.

Van Loan, C.F. & Pitsianis, N. 1993. Approximation with Kronecker products. In *Linear Algebra for Large Scale and Real-Time Applications*, Springer, pp. 293-314.

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

Wang, D., Liu, X. & Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, **208**, 231–248.

Wang, H. & West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, **96**, 821–834.

Xiao, H., Han, Y., Chen, R. & Liu, C. (2022). Reduced rank autoregressive models for matrix time series. *Journal of Business and Economic Statistics*. to appear.

[*Received November 2022; accepted October 2023*]