Matrix Time Series

Analysis

by

Seyed Yaser Samadi

(Under the direction of Professor Lynne Billard)

Abstract

Many data sets in the sciences (broadly defined) deal with multiple sets of multivariate time series. The case of a single univariate time series is very well developed in the literature; and single multivariate series though less well studied have also been developed (under the rubric of vector time series). A class of matrix time series models is introduced for dealing with the situation where there are multiple sets of multivariate time series data. Explicit expressions for a matrix autoregressive model of order one and of order p along with its cross-autocorrelation functions are derived. This includes obtaining the infinite order moving average analogues of these matrix time series. Stationarity conditions are also provided. Parameters of the proposed matrix time series model are estimated by ordinary and generalized least squares method, and maximum likelihood estimation method.

Index words: Matrix variate, time series, autoregressive model, cross-autoregressive and cross-autocorrelation function.

Matrix Time Series Analysis

Seyed Yaser Samadi

Matrix Time Series Analysis

by

Seyed Yaser Samadi

M.S., Shahid Beheshti University, Tehran, Iran, 2006

B.S., Shahid Bahonar University of Kerman, Iran, 2003

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2014

Matrix Time Series Analysis

by

Seyed Yaser Samadi

Approved:

Major Professors: Lynne Billard

Committee: Nicole Lazar
William McCormick
T.N. Sriram
Xiangrong Yin

Electronic Version Approved:

Dean's Name Here
Dean of the Graduate School
The University of Georgia
May 2014

Dedicated to:

My parents, Seyed Habib Samadi, and Seyedeh Hawa Hashemi

Acknowledgments

I would like to express my deepest appreciation and special thanks to my advisor, University Professor Dr. Lynne Billard, who has been a tremendous help and support for me, ready with brilliant ideas, and has shown the attitude and the substance of a genius. Her advice on both research as well as on my career have been invaluable; and without her constant encouragement, guidance, and persistent help this dissertation would not have been possible.

I would also like to thank my committee members, Professor Nicole Lazar, Professor William McCormick, Professor T.N. Sriram , and Professor Xiangrong Yin for serving as my committee members, and for their thoughtful comments and suggestions. Professor Sriram has been like a second advisor to me, ready with good ideas, honest advice, and encouraging words whenever I needed them.

I am very grateful to other faculty members in the Department of Statistics at UGA who provided me education. Also, I am indebted to all my teachers in schools, and faculty members in my previous universities in Iran, in the Department of Statistics at Shahid Beheshti University, Tehran, and Shahid Bahonar University of Kerman, who have taught and guided me throughout the years. Especially, thank to my previous advisor, Professor Mohammad Reza Meshkani, who has been a great advisor for me at Shahid Beheshti University.

I would like to thank the staff of the Department of Statistics at UGA for their helpful assistance, especially the Business Manager, Tim Cheek, who has been very kind and helpful since I entered this country. Also, I would like to thank all my friends who have been supporting and helping me in various ways.

Last but not least, I want to thank my loving and caring family. Words cannot express how grateful I am to my parents, my brothers, and my sisters, for their endless and unconditional love and support, and for always being there for me.

Contents

List of Figures

Chapter 1

Introduction

Time series processes are ubiquitous, arising in a variety of fields, across all scientific disciplines including econometrics, finance, business, psychology, biometrics, ecology, meteorology, astronomy, engineering, genetics, physics, medicine, biology, social science, and the like. In this work, the focus is on data sets which consist of multiple sets of multivariate time series, where the number of sets is $S > 1$, the number of variables is $K > 1$, and the number of time points is N.

Like many other statistical procedures, time series analysis has been classified into univariate, multiple and multivariate time series analysis. Models started with univariate autoregressive moving average (ARMA) processes and thereafter extended to multiple and multivariate time series. However, in the time series literature, multivariate time series analysis come under the heading of vector-variate time series and is called vector autoregressive moving average (VARMA) processes. In this work, we will extend the theory and methodology of VARMA time series models to matrix-variate time series. That is, matrix autoregressive time series models (MAR) are proposed for the first time in this study.

Matrix variate time series can be found in a variety of fields such as economics, business, ecology, psychology, meteorology, biology, fMRI, etc. For example, in a macroeconomics setting, we may be interested in a study of simultaneous behavior over time of employment statistics for different US states across different industrial sectors (Wang and West, 2009). Therefore, consider the data of Employment Statistics for eight US states, which is explored across nine industrial sectors at time t as follows

$$
\begin{array}{r} \\ \text{New Jersey} \\ \text{New York} \\ \text{Massachusetts} \\ \text{Georgia} \\ \text{North Carolina} \\ \text{Virginia} \\ \text{Illinois} \\ \text{Ohio} \end{array}
\begin{array}{c} \begin{array}{cccc} \text{construction} & \text{manufacturing} & \dots & \text{business services} \end{array} \\
\left[\begin{array}{r|c|c|c}
y_{11t} & y_{12t} & \cdots & y_{19t} \\
y_{21t} & y_{22t} & & y_{29t} \\
y_{31t} & y_{32t} & \cdots & y_{39t} \\
y_{41t} & y_{42t} & & y_{49t} \\
y_{51t} & y_{52t} & \cdots & y_{59t} \\
y_{61t} & y_{62t} & \cdots & y_{69t} \\
y_{71t} & y_{72t} & \cdots & y_{79t} \\
y_{81t} & y_{82t} & \cdots & y_{89t}
\end{array}\right] \end{array}
$$

where y_{ijt} is the Employment Statistics at time t from industrial sector j in state i.

As an another example, in an fMRI study, the blood oxygenation level is measured at different brain locations (voxels) associated with different types of stimuli (Antognini et al., 1997). Therefore, consider an fMRI data set of the blood oxygenation level at seven brain locations for three types of stimuli (shock, heat, brush) at time t, viz.,

$$
\begin{array}{r} \\ \text{Shock} \\ \text{Heat} \\ \text{Brush} \end{array}
\begin{array}{c} \begin{array}{cccc} \text{Location1} & \text{Location2} & \dots & \text{Location7} \end{array} \\
\left[\begin{array}{r|c|c|c}
y_{11t} & y_{12t} & \cdots & y_{17t} \\
y_{21t} & y_{22t} & & y_{27t} \\
y_{31t} & y_{32t} & \cdots & y_{37t}
\end{array}\right] \end{array}
$$

where y_{ijt} is the blood oxygenation level at time t from stimuli i at location j of the brain.

In both of the two examples given above, at each time t, the data set has two components. In the first example, for each given industrial sector (say, "construction"), we have a vector time series, where the variables of the vector time series are US states. On the other hand, we have nine (number of industrial sectors) vector time series with dimension eight (number of states). Obviously, there are some kinds of dependencies between these vector time series (industrial sectors). Also, in the second example, for each brain location (voxel), we have a vector time series of dimension three (types of stimuli). Clearly, there are dependencies between the voxels (vector time series). Therefore, it turns out to be a matrix time series data by considering all dependent vector time series simultaneously over time.

Wang and West (2009) considered a matrix normal distribution for both observational and evolution errors of a dynamic linear model of a matrix-variate time series data to fit and explore dynamic graphical models. We will extend fundamental concepts and results for vector time series analysis to matrix time series. New problems and challenges arise in the theory and application due to the greater difficulty and complexity of model dimensions, and due to the parametrization in the matrix situation.

A comprehensive literature review of time series analysis is provided in chapter 2. This review introduces time series data and the class of autoregressive moving average (ARMA) models for analyzing time series data. This literature review follows a chronological order of the development of ARMA models for univariate and then for vector time series data. Furthermore, after introducing matrix-variate processes, the existing tools and works to analyze matrix time series data are given. Chapter 2 will finish by introducing and defining the matrix variate normal distribution.

One of the advantages of the matrix variate normal distribution, besides the most desirable aspect of being able to estimate within and between time series variations, is that it gives parameters-wise parsimonious models. Because of the Kronecker product structure, the number of parameters to be estimated decreases quickly by increasing the dimension of the matrix. This is so because when $\mathbf{Y}_t$ is a matrix time series of dimension $K \times S$, the number of variance-covariance parameters that is needed to be estimated while using the multivariate (vector) normal distribution, is $KS(KS+1)/2$. However, this number of parameters decreases to $K(K+1)/2 + S(S+1)/2$ by applying matrix normal distributions.

In chapter 3, we introduce a model for matrix time series and develop some theory for this class of models. In particular, we model the matrix time series to obtain expectations for the variance-covariances; see section 3.2. Then, after introducing matrix autoregressive series of order one in section 3.3, we consider and describe stationary matrix processes in general in section 3.4. In section 3.5, we propose and derive the corresponding matrix moving average representation process of order infinity for the matrix autoregressive series of order one defined in section 3.3. Then, in section 3.6, we derive the autocovariance and autocorrelation functions of the matrix autoregressive models of order one and its marginal vectors. In section 3.7, we introduce the matrix autoregressive time series of order p, and find its corresponding matrix moving average representation, and hence we derive the autocovariance and autocorrelation functions of the matrix autoregressive model of order p. In section 3.8, we study the matrix autoregressive processes with nonzero mean, and we find the intercept of such series by deriving its moving average representation. Finally, in section 3.9, we derive the Yule-Walker Equations for MAR processes.

In chapter 4, we estimate the parameters of the matrix autoregressive processes of order p (MAR(p)), proposed in chapter 3, based on a sample of matrix observations. This chapter will start with some preliminary material and basic results in section 4.2 that will be used in the rest of the chapter. We estimate the parameters of the matrix time series based on two main estimation methods, namely, least squares estimation, and maximum likelihood estimation in sections 4.3 and 4.4, respectively. In the least squares estimation method, we consider both ordinary least squares (OLS) estimation, and generalized least squares (GLS) estimation in sections 4.3.2 and 4.3.3, respectively, for the MAR(1) process. In section 4.3.4, the least squares estimators of parameters of the mean adjusted MAR(1) model will be derived. Finally, in section 4.4, we will use the maximum likelihood method to estimate the parameters of the MAR(p) model.

Eventually, in chapter 5, numerical and simulation studies are conducted to compare different matrix autoregressive of order one (MAR(1)) models when they have different coefficient matrices. In fact, like univariate and vector time series, the structure of the autocorrelation functions of MAR models is dependent on the configuration of the coefficient matrices.

The final chapter outlines some proposals for future work.

Chapter 2

Literature Review

2.1 Introduction

A time series is a sequence of observations measured at successive times or over consecutive periods of times. Unlike observations of a random sample, observations of a time series are statistically dependent. That is, an inherent feature of a time series is that, usually, the adjacent observations are not independent. There is a considerable practical interest in analyzing and determining the pattern of dependencies among observations of a time series. Time series analysis is concerned with probabilistic and statistical methods for analyzing the dependence among observations and making inferences based on sequential data. There are three main goals of time series analysis: identifying patterns for characterization, modeling the pattern of the process, and forecasting future values. This leads to development of analytical methods of stochastic processes for analyzing and predicting dependent data. Time series processes are ubiquitous in stochastic phenomenon, arising in a variety of fields, including econometrics, finance, business, biometrics, biology, ecology, meteorology, medicine, astronomy, engineering, genetics, physics, fMRI, social science, etc.

Let y_t be a stochastic process, where the index t takes integer values. In this case, y_t is a random variable at time t, and a time series, in fact, is a random sample from such a process. In general, for a given time series y_t the object of interest is given by

$$y_t = f(y_{t-1}, y_{t-2}, \ldots) + \varepsilon_t, \quad t = 1, 2, \ldots, N, \tag{2.1}$$

where $f(.)$ is a suitable function of past observations, and ε_t are independent, identically distributed (i.i.d.) errors with mean zero and finite variance σ^2, which is called white noise.

Determination of the function $f(.)$ is a major task in time series analysis. In most applications, $f(.)$ has been considered as a linear function of past observations. The autoregressive integrated moving average ($ARIMA$) models are the best examples and the most commonly used of these kind of linear functions.

Because of convenient mathematical properties of linear functions and because they are relatively easy to use in applications, the classic $ARIMA$ models are the most popular models for analyzing time series data. These models have used the information criteria for lag selection since 1990s. Autoregressive (AR) models were introduced by Yule (1927), moving average (MA) models proposed by Walker (1931) and Slutsky (1937), then AR and MA type models were combined into the mixed autoregressive moving average ($ARMA$) models. Later on, Box and Jenkins (1970) extended the $ARMA$ models so as to be suitable for particular types of nonstationary time series, so-called $ARIMA$ process.

Like many other statistical procedures, time series analysis has been classified into univariate, multiple and multivariate time series analysis. Models started with univariate $ARMA$ processes and thereafter extended to multiple and multivariate time series. However, in the time series literature, multivariate time series analysis use vector-variate time series and is frequently called vector autoregressive moving average ($VARMA$) processes. In this work, we deal with multiple sets of dependent single multivariate (vector) time series. This constitutes matrix time series, and we will extend the theory and methodology of $VARMA$ time series models to matrix-variate time series. That is, matrix autoregressive time series models (MAR) are proposed for the first time in this study.

2.2 Literature review

Time series analysis started in the early twentieth century with three influential papers by Yule (1921, 1926, 1927). Yule (1921) considered the time-correlation problem for correlations between unrelated quantities observed over time. Yule (1926) formulated the correlations between the time-variables, what he called nonsense-correlations, and he found the relationship between the serial correlations for the series, r_i, and the serial correlations for the difference series, ρ_i. Yule pioneered the idea of autoregressive series and applied a second order autoregressive process, $AR(2)$, for modeling Wolfer's sunspot data (successive annual sun-spot numbers). That is, he showed that past observations of a variable can explain and determine its motion in the present (Yule, 1927). Yule's second order autoregressive process $AR(2)$ is given by

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \tag{2.2}$$

where y_t is the observation at time t, a_1 and a_2 are the regression coefficients (autocorrelations), and ε_t are assumed to be independent random error terms with mean zero and variance one. Yule's work was extended to a general pth-order autoregressive, $AR(p)$, by Walker (1931). In that paper, it was shown that the relation between successive uninterrupted terms of the series y_t plus a error term ε_t is

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t. \tag{2.3}$$

Walker (1931) showed that, for a large enough number of observations, n, there is a similar equation, but without the random shock ε_t, between the successive correlation coefficient (autocorrelations) values of the series terms, r_i, i.e.,

$$r_k = a_1 r_{k-1} + a_2 r_{k-2} + \ldots + a_p r_{k-p}. \tag{2.4}$$

This equation for $k = 1, 2, \ldots, p$, gives a set of linear equations that are usually called the Yule-Walker equations. Slutsky (1937) proposed the moving average (MA) models, and studied how these models could lead to cyclical processes. That is, a coherent series y_t can be decomposed into a weighted moving summation of incoherent (random) series ε_t as follows

$$y_t = \varepsilon_t + m_1\varepsilon_{t-1} + \ldots + m_q\varepsilon_{t-q} \tag{2.5}$$

where $m_1, m_2, \ldots, m_q$ are weights of the random series ε_t at different lags.

Wold (1938) was the first to apply moving average processes to data and proved that any stationary time series can be decomposed into a moving average of independent random variables. Moving average models were not usually used because of the difficulty of finding an appropriate model and a lack of suitable methods for determining, fitting, and assessing these models. Therefore, to achieve better flexibility in fitting real time series data, moving average (MA) and autoregressive (AR) processes were merged into autoregressive moving average ($ARMA$) processes.

Since then, progress was made in the area of inference for time series models, dealing with estimating the parameters, properties of the estimators, identification, and assessing models. Mann and Wald (1943) studied the autoregressive model of Eq (2.3) and derived the asymptotic theory for ordinary least squares parameter estimation. Champernowne (1948) proposed a successive approximation of least squares estimates and maximum likelihood estimates for autoregressive models and autoregressive models with regression terms. However, he did not develop properties of the estimators. Cochrane and Orcutt (1949) introduced a new method to estimate the regression parameters when the error terms are autocorrelated. In fact, this procedure is dealing with the problem of correlated errors in the time series context. Bartlett (1946) and Moran (1947) developed some asymptotic properties of estimators.

The foundation of the mathematical and probabilistic formulation and properties of AR, MA, and $ARMA$ processes can be found in Box and Jenkins (1970, 1976), and Box et al. (1994, 2008). The $ARMA$ models can be applied just for stationary time series. Box and Jenkins (1970) used mathematical statistics and probability theory for extending the $ARMA$ models to include certain types of nonstationary time series, and they proposed a class of models called autoregressive integrated moving average ($ARIMA$) models. The $ARMA$ and $ARIMA$ models are very useful and can be applied to a wide range of time series data in many fields. They proposed an intelligible readily accomplished three-stage iterative modeling approach for time series, viz., model identification, parameter estimation, and model checking. These procedures are now known as the Box-Jenkins approach, and lead to estimating the parameters p, d, q, and other parameters in a suitable $ARIMA(p, d, q)$ model for a set of data, where p, d, and q are non-negative integer values that refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. Box and Jenkins (1970) also investigated other models such as transfer function noise (TF) models, and seasonal autoregressive integrated moving average ($SARIMA$) models. A good review of time series analysis is provided by De Gooijer and Hyndman (2006).

More details on the analysis, modeling, and forecasting of time series data can be found in several well-known books and references including: Box and Jenkins (1970, 1976), Box et al. (1994, 2008), Brillinger (1975), Priestley (1981), Shumway (1988), Brockwell and Davis (1991, 2002), Pourahmadi (2001), and Wei (2006).

Multivariate (vector) time series analysis was pioneered during the 1980s. However, Whittle (1953) had earlier stressed the need to develop methods for multivariate time series analysis also called multiple series, because the majority of practical problems required an analysis of a multiple series. He derived least squares estimators for a stationary multivariate time series, and showed that they are equivalent to those obtained through the maximum likelihood principle if the variates are assumed to the normally distributed. He also investigated the asymptotic properties of the parameter estimates. The multivariate (vector) generalization of the univariate $ARIMA$ model is called a vector $ARIMA$ ($VARIMA$) model. Quenouille (1957) was the first to determine the characteristics of vector autoregressive integrated moving average ($VARIMA$) processes. Parzen (1969) represented the probability structure of covariance-stationary multiple time series by considering multiple time series as a series of random vectors. Tunnicliffe Wilson (1973) presented a practical method for estimating parameters in multivariate stationary $ARMA$ time series models. His estimator was an asymptotic approximation of the exact maximum likelihood estimator which was proposed by Osborn (1977). Wallis (1977) proposed the joint estimation and model selection procedure with a likelihood ratio test for multiple time series models.

During the 1980s and 1990s, multivariate (vector) or multiple time series analysis came to be pioneered due to the progress and availability of suitable softwares to implement $VARMA$ models. Tiao and Box (1981) proposed an approach to analyze and model multiple time series, and explained the properties of a class of $VARMA$ models. They extended the univariate three-stage iterative modeling procedure of the Box and Jenkins's procedure to analyze multivariate (vector) time series, i.e., model identification, parameter estimation and diagnostic checking. Much work has been done on the problems of identifying, estimating, formulating, and explaining different kinds of relationships among several series of a multivariate time series; e.g., Quenouille (1957), Hannan (1970), Box and Jenkins (1970, 1976), Brillinger (1975), Brockwell and Davis (1991, 2002), Wei (2006), Lütkepohl (1991, 2006), and Box et al. (2008).

In this work, we will study the relationship between a matrix of time series variables $\{Y_{ijt} : i = 1, 2, \ldots, K; j = 1, 2, \ldots, S\}$. We can find such matrix-variate processes when we observe several related vector time series simultaneously over time, rather than observing just a single time series or several related single time series as is the case in univariate time series or vector-variate time series, respectively. Therefore, let $Y_{.jt}$ be a j^{th} $(j = 1, 2, \ldots, S)$ vector variate time series and let $\mathbf{Y}_t$ denote the matrix variate time series at time t given by

$$\mathbf{Y}_t = (Y_{.1t}, Y_{.2t}, \dots Y_{.St}) = \begin{bmatrix} y_{11t} & y_{12t} & \cdots & y_{1St} \\ y_{21t} & y_{22t} & \cdots & y_{2St} \\ \vdots & & & \vdots \\ y_{K1t} & y_{K2t} & \cdots & y_{KSt} \end{bmatrix}, \qquad t = 1, 2, \dots. \tag{2.6}$$

Matrix variate time series can be found in a variety of fields such as economics, business, ecology, psychology, meteorology, biology, fMRI, etc. For example, in a macroeconomics setting, we may be interested in a study of simultaneous behavior over time of employment statistics for different US states across different industrial sectors (Wang and West, 2009). As another example, for linguists it is important to know the lexical structure of a language rather than isolated words. Specifically, in the Australian Sign Language (Auslan) the number of variables on a fitted glove, as different "spoken" words, are measured over time; where variables are $Y_{1.}, Y_{2.}, Y_{3.}$ signed in sign language in three dimensional space (X, Y, Z) respectively, $Y_{4.}$ = roll (of palm as it moved), $Y_{5.}$ = thumb bend, $Y_{6.}$ = forefinger bend, $Y_{7.}$ = index finger bend, $Y_{8.}$ = ring finger bend (with the bends in $Y_{5.}, \dots, Y_{8.}$ going from straight to fully bent) across four words $1 \equiv$ know, $2 \equiv$ maybe, $3 \equiv$ shop, $4 \equiv$ yes, each word is spoken three times. Therefore in our model framework, we have $K = 8$, $S = 12$, $\mathbf{Y}_t = (Y_{1.t}, Y_{2.t}, \dots Y_{8.t})^T$, where $Y_{i.t}^T = (y_{i1t}, y_{i2t}, \dots, y_{i12t})$, $i = 1, 2, \dots, 8$ (8×12 matrix variate time series). See Kadous (1995, 1999).

We will extend most of the fundamental concepts and results for vector time series analysis to matrix time series. New problems and challenges arise in the theory and application due to the greater difficulty and complexity of model dimensions, and parametrization in the matrix situation.

In recent years, a few papers have dealt with matrix-variate time series. Quintana and West (1987) for the first time used multivariate (vector) time series models based on a general class of a dynamic matrix variate normal extensions of a dynamic linear model. They considered a q-vector time series Y_t with the following model

$$
\begin{aligned}
&\text{Observational equation:} && Y_t^T = F_t^T \Theta_t + e_t^T, && \mathbf{e}_t \sim N(0, v_t \Sigma) \\
&\text{Evolution equation:} && \Theta_t = G_t \Theta_{t-1} + \Omega_t, && \Omega_t \sim N(0, W_t, \Sigma) \\
&\text{Prior distributions:} && \Theta_{t-1} \sim N(M_{t-1}, C_{t-1}, \Sigma), && \Sigma \sim W^{-1}(S_{t-1}, d_{t-1})
\end{aligned}
$$

where $\mathbf{e_t}$ is a $q-$vector of observational error, Ω_t is a $p \times q$ evolution error matrix, and follows a matrix-variate normal distribution with mean 0, left covariance matrix W_t and right covariance matrix Σ. The left covariance matrix W_t contains the variance and covariance of the p variables, and the right covariance matrix Σ contains the variance and covariance between the q variables. Matrix normal notations and properties will be defined later. This method was further developed and applied in West and Harrison (1999), Carvalho and West (2007 a, b).

Note, however, that in their proposed models the evolution error term Ω_t corresponding to the matrix of states Θ_t is a matrix, the response variable Y_t is a vector. Wang and West (2009) developed these proposed models when the response variable Y_t is a matrix. They introduced and analyzed matrix normal graphical models, in which conditional independencies induced in the graphical model structure were characterized by the covariance matrix parameters. They considered a fully Bayesian analysis of the matrix normal model as a special case of the full graphs, and extended computational methods to evaluate marginal likelihood functions under an indicated graphical model. Then, they developed this graphical modeling to a matrix time series analysis and their associated graphical models. They assumed that a $q \times p$ matrix variate time series $\mathbf{Y}_t$ follows a dynamic linear model where both observational and evolution error terms change over time, and for given t they have a given matrix normal distribution with

$$\text{Observational equation:} \quad \mathbf{Y}_t = (I_q \otimes F_t^T)\Theta_t + v_t, \qquad v_t \sim N(0, U, \Sigma)$$

$$\text{Evolution equation:} \quad \Theta_t = (I_q \otimes G_t)\Theta_{t-1} + \Omega_t, \qquad \Omega_t \sim N(0, U \otimes W_t, \Sigma).$$

Wang and West (2009) applied this model and procedure to a macroeconomic time series data set. The data were monthly 8×9 matrix variates over several years for Employment Statistics for eight US states across nine industrial sectors, viz., construction, manufacturing, trade, transportation and utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality, and government.

Hoff (2011) considered a class of multi-indexed data arrays $Y = \{y_{i_1}, y_{i_2} \ldots, y_{i_k} : i_k \in \{1, 2, \ldots m_k\}, k = 1, 2, \ldots, K\}$, and introduced estimation methods and accommodated a construction for an array normal class of distributions. However in that paper, Hoff derived some properties of covariance structures of multidimensional data arrays and focused on an extension of the matrix normal model for a multidimensional array, but had nothing about time series data. However, if we consider the third dimension in a three-dimensional data array as an index of time, then the data array would be a matrix time series data. Wang (2011) proposed matrix variate Gaussian graphical models for correlated samples to analyze the effect of correlations based on matrix-variate normal distribution.

Triantafyllopoulos (2008) proposed and explored missing observations of any sub-vector or sub-matrix of observation time series matrix by developing Bayesian inference for matrix variate dynamic linear models.

Fox and West (2011) introduced and analyzed a class of stationary time series models, first order autoregressive process, for variance matrices by using the structure of conditional and marginal distributions in the inverse Wishart family.

2.3 Matrix Variate

Random matrix theory has found many of its applications in physics, mathematics, engineering, and statistics. It was first introduced in mathematical statistics by Wishart (1928). Wigner (1955, 1957, 1965, 1967) developed and applied the matrix theory in physics.

Sample observation matrices are very common to use in statistical methodologies and applications. These matrices basically are composed from independent multivariate observations, and first introduced by Roy (1957). When sampling from multivariate normal distributions, the columns (or rows) of such matrices are independently distributed as multivariate normal distributions with common mean vector and covariance matrix. In many data sets, the independence assumption of multivariate observations is not valid, i.e., time series, stochastic processes and repeated measurements on multivariate variables (Gupta and Nagar, 2000). Therefore, the study of matrix variate distributions and their properties began for analyzing such matrices of observations when all entries are dependent.

We usually make some assumptions on the model to make it simple and to enable the use of existing methodologies or computation methods. Especially, for analyzing matrix data, we often assume that columns (or rows) are independent (Allen and Tibshirani, 2012). With the assumption of independence between the columns of a matrix, the variables of one dimension, rows, can be considered as points of interest. Then, we just would be able to study the relationship among the rows, but not among the columns. However, these assumptions are not always met and sometimes we would like to find the structure between the variables of both rows and columns. Allen (2010) gave two real examples that this assumption is not met, i.e., gene-expression microarrays and the Netflix movie-rating data.

There is a different terminology for matrix data that both columns and rows are dependent, i.e., two-way data, and transposable data. Transposable data are a type of two-way data (matrix data) where both the rows and columns are correlated, and the data can be viewed as a row-model, meaning the rows are the features of interest, or as a column-model (Allen, 2010). Allen (2010) proposed a model for studying transposable data, by modification of the matrix variate normal distribution that has a covariance matrix for both columns and rows.

Viroli (2012) introduced and proposed matrix-variate regression models for analyzing three-way data where for each statistical unit, response observations are matrices composed of multivariate observations in different occasions. In fact, Viroli extended and represented multivariate regression analysis to matrix observations for dealing with a set of variables that are simultaneously observed on different occasions. The matrix-variate regression model is given by

$$\mathbf{Y}_i = \Theta \mathbf{X}_i + \mathbf{e}_i, \qquad i = 1, 2, \ldots, n, \tag{2.7}$$

where $\mathbf{Y}_i$ is a $p \times s$ observed matrix, $\mathbf{X}_i$ is a predictor matrix of dimension $m \times s$, Θ is a matrix of unknown parameters matrix of dimension $p \times m$, and $\mathbf{e}_i$ is the model errors of dimension $p \times s$ with matrix normal distribution. By these assumptions, the parameters of the model are estimated and the properties of the model are studied.

2.3.1 Matrix Variate Normal Distribution

Matrix variate normal distributions, like univariate and multivariate normal distributions, have a most important and effective role in applications with nice mathematical properties among the matrix variate distributions. A matrix variate normal distribution was first studied in the 1980s by Dawid (1981), Wall (1988), among others. Recently, this family of distributions is growing tremendously in many applications due to computational progress and advances.

The $K \times S$ random matrix $\mathbf{Y}$ has a matrix variate normal distribution with $K \times S$ mean matrix $\mathbf{M}$ and covariance matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}$, if $Vec(\mathbf{Y})$ has a multivariate normal distribution with mean $Vec(\mathbf{M})$ and covariance matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}$. That is, $Vec(\mathbf{Y}) \sim N_{KS}(Vec(\mathbf{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega})$, where $Vec(\mathbf{Y})$ is the vector of all columns $\mathbf{Y}$ from left to right, and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are positive definite matrices of dimension $K \times K$ and $S \times S$, respectively; see Gupta and Nagar (2000).

Let $\mathbf{Y}$ be a random matrix of dimension $K \times S$ with a matrix variate normal distribution. Then, the probability density of $\mathbf{Y}$ is given by

$$p(\mathbf{Y}|\mathbf{M}, \mathbf{\Omega}, \mathbf{\Sigma}) = (2\pi)^{-\frac{KS}{2}} |\mathbf{\Omega}|^{-\frac{K}{2}} |\mathbf{\Sigma}|^{-\frac{S}{2}} \exp\left\{ -\frac{1}{2} tr\Big(\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{M})\mathbf{\Omega}^{-1}(\mathbf{Y}-\mathbf{M})^T\Big)\right\} \quad (2.8)$$

where $M_{K\times S} \in \mathbb{R}^{K\times S}$ is the expected matrix value of $\mathbf{Y} \in \mathbb{R}^{K\times S}$, $\mathbf{\Sigma}$ is the left (row, or between) covariance matrix of dimension $K \times K$, and $\mathbf{\Omega}$ is the right (column, or within) covariance matrix of dimension $S \times S$. The matrix variate normal distribution is denoted by $\mathbf{Y} \sim N_{K\times S}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Omega})$.

One of the advantages of the matrix variate normal distribution, besides the most desirable aspect of being able to estimate within and between variations, is that it gives parameters-wise parsimonious models. Because of the Kronecker product structure, the number of parameters to be estimated decreases quickly by increasing the dimension of the matrix. This is so because when $\mathbf{Y}$ is a matrix of dimension $K \times S$, the number of parameters that is needed to be estimated, while using the multivariate normal distribution, is $KS(KS+1)/2$. However, this number of parameters decreases to $K(K+1)/2 + S(S+1)/2$ by applying matrix normal distributions.

Chapter 3

Matrix Time Series - Models

3.1 Introduction

Let Y_t be a stochastic process, where the index t takes integer values. In our case, Y_t is a random variable at time t and a time series is a random sample from such processes. In general, for a given Y_t, a time series model can be considered as

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots) + \varepsilon_t, \quad t = 1, 2, \ldots, N, \tag{3.1}$$

where $f(.)$ is a function of past observations, and ε_t, $t = 1, 2, \ldots, N$, are independent and identically distributed (i.i.d.) random errors with mean zero and finite variance σ^2, called white noise. For standard univariate time series, this $f(.)$ function is a real scalar function, and for vector time series, $f(.)$ is a real vector function.

Determination of the function $f(.)$ is a major task in time series analysis. In most applications, $f(.)$ is considered to be a linear function of past observations, of which the autoregressive integrated moving average (ARIMA) models are the most commonly used examples.

The time series variable Y_t can be univariate or multivariate. In the time series literature, a multivariate time series is often a vector of series; however, Y_t can be a matrix of time series.

There is a vast literature for univariate and for vector time series which study properties and features of univariate and vector time series as reviewed in chapter 2. With the one exception of Wang and West (2009), there seems to be no literature for the representation of matrix time series. Wang and West considered a matrix normal distribution for both observational and evolution errors of a dynamic linear model of a matrix-variate time series Y_t to fit and explore dynamic graphical models.

In this work, we introduce a model for matrix time series and develop some theory for this class of models. In particular, we model the matrix time series to obtain expectations for the variance-covariances; see section 3.2. Then, after introducing matrix autoregressive series of order one in section 3.3, we consider and describe stationary matrix processes in general in section 3.4. In section 3.5, we propose and derive the corresponding matrix moving average representation process of order infinity for the matrix autoregressive series of order one defined in section 3.3. Then, in section 3.6, we derive the autocovariance and autocorrelation functions of the matrix autoregressive models of order one and its marginal vectors. In section 3.7, we introduce the matrix autoregressive time series of order p, and find its corresponding matrix moving average representation, and hence we derive the autocovariance and autocorrelation functions of the matrix autoregressive model of order p. Finally, in section 3.8, we study the matrix autoregressive processes with nonzero mean, and we find the intercept of such series by deriving its moving average representation.

We note that vector time series are sometimes referred to as multiple time series, where "multiple" refers to multiple variables (i.e., number of variables $K > 1$). As for univariate series, the vector series is a single (but multivariate) series. Our work deals with multiple series of multivariate time series, i.e., number of series $S > 1$. To avoid confusion, we restrict the use of "multiple" to our matrix time series, i.e., multiple multivariate (or multiple vectors) time series.

3.2 Matrix Time Series

3.2.1 The Model

Consider S time series $Y_{.1t}, Y_{.2t}, \ldots, Y_{.St}$, such that each series $Y_{.jt},\ j = 1, 2, \ldots, S$, itself is a $K-$dimensional vector time series. That is, we have

$$Y_{.jt} = (y_{1jt}, y_{2jt}, \ldots, y_{Kjt})^T, \qquad j = 1, 2, \ldots, S.$$

Also, suppose that these S vector time series are not independent; i.e., $Cov(Y_{.jt}, Y_{.j't}) = \mathbf{\Sigma}_{jj'}(t) \neq 0,\ j \neq j' = 1, 2, \ldots, S$. Due to these dependencies and considering the contributions and effects of these underlying cross-correlations to the results of any analysis, it is necessary to put all S vector series into one model, so as to be able to analyze them simultaneously. To this end, we put all S vector series, $Y_{.jt},\ j = 1, 2, \ldots, S$, beside each other in a matrix and then we will have a matrix time series $\mathbf{Y}_t$. Therefore, suppose that the matrix time series variate $\mathbf{Y}_t$ is given as

$$\mathbf{Y}_t = (Y_{.1t}, Y_{.2t}, \ldots Y_{.St}) = \begin{bmatrix} y_{11t} & y_{12t} & \cdots & y_{1St} \\ y_{21t} & y_{22t} & \cdots & y_{2St} \\ \vdots & & & \vdots \\ y_{K1t} & y_{K2t} & \cdots & y_{KSt} \end{bmatrix}. \tag{3.2}$$

For the sake of convenience and simplicity in notation, the bold capital letter $\mathbf{Y}_t$ will be used for representing the matrix time series, the vector time series will be shown by capital letter, but not bold Y_t, and the uncapitalized letter y_t will be used for a univariate time series.

Now, as for univariate and vector time series, the same model of Eq (3.1) can be considered for the matrix time series variate $\mathbf{Y}_t$. In section 3.4, we will introduce and extend features of a function $f(.)$ for matrix time series corresponding to the function $f(.)$ for univariate and vector time series. We can rewrite the model in Eq (3.1) for the matrix time series $\mathbf{Y}_t$ as

$$\mathbf{Y}_t = f(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots) + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \ldots, N, \tag{3.3}$$

where $f(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots)$ and $\boldsymbol{\varepsilon}_t$ (as are $\mathbf{Y}_{t-i}, i = 0, 1, \ldots$) are matrices of dimension $K \times S$. Note that in contrast to the function $f(.)$ in Eq (3.1), where it is a real scalar function of a univariate time series y_{ijt}, or a real vector function of vector time series $Y_{.jt}$, here in Eq (3.3), $f(.)$ is a real matrix function of a matrix time series.

3.2.2 Variance-Covariance of Matrix Time Series

Knowing the structure of the variance-covariance matrix (across the random variables themselves, as distinct from the autocovariance function considered in section 3.6) of a matrix time series is necessary for building a statistical model. In this section, we will define the structure of the variance-covariance of a matrix time series. To this end, first we assume that the matrix time series $\mathbf{Y}_t$, like univariate and vector time series, has zero mean and finite variance. We use the Kronecker product of matrix white noise $\boldsymbol{\varepsilon}_t$ and its expectation in order to define the variance-covariance of white noise $\boldsymbol{\varepsilon}_t$. That is, first assume that $\boldsymbol{\Psi}(t)$ is defined as

$$\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}(t) = E(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T). \tag{3.4}$$

Note that $\boldsymbol{\varepsilon}_t$ is a $K \times S$ matrix; therefore, by definition of a Kronecker product, $\boldsymbol{\Psi}(t)$ is a $KS \times KS$ matrix. One of the purposes of introducing the matrix time series model is to find a way to be able to analyze each vector of the matrix time series such that the effect of all series of the matrix can be considered simultaneously. Therefore, the variance-covariance matrix of a time series matrix needs to be well defined such that the variance and covariance matrices of vectors can be part of the variance-covariance matrix of a time series matrix. That is, if we partition the variance-covariance matrix of the matrix time series, then each partition can be the variance of a vector or the covariance of two vectors. However, it easily can be seen that the matrix $\boldsymbol{\Psi}(t)$ defined in Eq (3.4) cannot be readily partitioned in a way that partitions represent the variance-covariance of the corresponding vectors. To overcome this obstacle, we define a transformation matrix $\mathbf{T}$. Then, the mean and variance-covariance of $\boldsymbol{\varepsilon}_t$ is defined as

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_{(K \times S)}, \qquad Var(\boldsymbol{\varepsilon}_t) = E(\mathbf{T}\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T) = \mathbf{T}\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}(t) = \boldsymbol{\Sigma}_{(KS \times KS)}(t) \tag{3.5}$$

where the transformation matrix $\mathbf{T}$ is defined as

$$\mathbf{T}_{(KS \times KS)} = (t_{ij}) = \begin{cases} 1, & \text{if } \begin{cases} i = lK+1, lK+2, \ldots, (l+1)K, \\ j = (l+1) + (i - lK - 1)S, \\ l = 0, 1, 2, \ldots, (S-1); \end{cases} \\ \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

This is equivalent to the following equation defined by Brewer (1978)

$$\mathbf{T}_{(KS \times KS)} = \sum_{i=1}^{K} \sum_{j=1}^{S} \mathbf{E}_{ji}^{S \times K} \otimes \mathbf{E}_{ij}^{K \times S} \tag{3.7}$$

where $\mathbf{E}_{ij}^{K \times S}$ is a $K \times S$ matrix of one's and zero's, in particular its lf^{th} element, $(e_{lf})_{ij}$, is given by

$$(e_{lf})_{ij} = \begin{cases} 1, & \text{if } l = i \text{ and } f = j, \\ 0, & \text{otherwise,} \end{cases}, \qquad l = 1, 2, \ldots, K, \; f = 1, 2, \ldots, S. \tag{3.8}$$

Now we can partition the matrix $\boldsymbol{\Sigma}(t)$ defined in Eq (3.5) into S^2 sub-matrices each of dimension $K \times K$. Then, the diagonal sub-matrices of the $S \times S$ dimensional block matrix $\boldsymbol{\Sigma}(t)$ are the variances of the vectors $Y_{.jt}$ and the off-diagonal sub-matrices of $\boldsymbol{\Sigma}(t)$ are the covariance of the vectors $Y_{.jt}$ and $Y_{.j't}$, $j \neq j'$. That is,

$$\boldsymbol{\Sigma}(t) = (\boldsymbol{\Sigma}_{jj'}(t)) = \begin{bmatrix} \boldsymbol{\Sigma}_{11}(t) & \boldsymbol{\Sigma}_{12}(t) & \dots & \boldsymbol{\Sigma}_{1S}(t) \\ \boldsymbol{\Sigma}_{21}(t) & \boldsymbol{\Sigma}_{22}(t) & \dots & \boldsymbol{\Sigma}_{2S}(t) \\ \vdots & & \ddots & \vdots \\ \boldsymbol{\Sigma}_{S1}(t) & \boldsymbol{\Sigma}_{S2}(t) & \dots & \boldsymbol{\Sigma}_{SS}(t) \end{bmatrix} \tag{3.9}$$

where

$$\boldsymbol{\Sigma}_{jj}(t) = Var(\boldsymbol{\varepsilon}_{.jt}) = E[\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.jt}], \qquad j = 1, 2, \dots, S,$$

$$\boldsymbol{\Sigma}^T_{jj'}(t) = Cov(\boldsymbol{\varepsilon}_{.jt}, \boldsymbol{\varepsilon}_{.j't}) = E[\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\mu}^T_{.j't}], \qquad j \neq j',\ j, j' = 1, 2, \dots, S.$$

The following example shows how to find the transformation matrix $\mathbf{T}$ and how it looks when $K = 4$ and $S = 3$.

Example 3.2.1 *We wish to find the elements $t_{ij}, i = 1, 2, 3, 4, j = 1, 2, 3,$ of matrix $\mathbf{T}$ when $K = 4$ and $S = 3$. First note that because $S = 3$, the l in Eq (3.6) can take values, $0, 1$ and 2 only. Also, for each value of l, index i and index j are different. That is, we can write*

$$l = 0 \Rightarrow \begin{cases} i = 1, 2, 3, 4 \\ j = 1 + (i-1)3 \end{cases}, \quad \Rightarrow \quad t_{1,1} = t_{2,4} = t_{3,7} = t_{4,10} = 1;$$

$$l = 1 \Rightarrow \begin{cases} i = 5, 6, 7, 8 \\ j = 2 + (i-5)3 \end{cases}, \quad \Rightarrow \quad t_{5,2} = t_{6,5} = t_{7,8} = t_{8,11} = 1;$$

$$l = 2 \Rightarrow \begin{cases} i = 9, 10, 11, 12 \\ j = 3 + (i-9)3 \end{cases}, \quad \Rightarrow \quad t_{9,3} = t_{10,6} = t_{11,9} = t_{12,12} = 1.$$

Therefore, the matrix $\mathbf{T}$*, for the given* K *and* S *of dimension* 12×12*, is given by*

$$\mathbf{T}_{12\times 12} = (t_{ij}) = \left[\begin{array}{ccc|ccc|ccc|ccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\hline
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]. \quad (3.10)$$

This matrix, also can be found by using the quantity in Eq (3.7)*. That is,*

$$\mathbf{T}_{(12\times 12)} = \sum_{i=1}^{4}\sum_{j=1}^{3} \mathbf{E}_{ji}^{3\times 4} \otimes \mathbf{E}_{ij}^{4\times 3}.$$

Theorem 3.2.1 *The square matrix* $\mathbf{T}$ *defined in Eq* (3.6) *is nonsingular and invertible, and* $\mathbf{T}^T = \mathbf{T}^{-1}$.

Proof: *Write* $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{KS})$ *where* $\mathbf{t}_m, m = 1, 2, \ldots, KS$*, is the* m^{th} *column of* $\mathbf{T}$*. Then, we can reorder the columns of the matrix* $\mathbf{T}$ *to give the identity matrix* I_{KS}*.*

Example 3.2.2 *(*Continuation of Example 3.2.1*) In the previous Example* 3.2.1*, write* $\mathbf{T}_{12\times 12} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{12})$*. Then, rewriting the columns of* $\mathbf{T}$ *as* $(\mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_7, \mathbf{t}_{10}, \mathbf{t}_2, \mathbf{t}_5, \mathbf{t}_8, \mathbf{t}_{11}, \mathbf{t}_3, \mathbf{t}_6, \mathbf{t}_9, \mathbf{t}_{12})$ *yields the identity matrix* $\mathbf{I}_{12}$*.*

It is helpful to look at the variance-covariance matrix of the white noise matrix $\boldsymbol{\varepsilon}_t$ in a different way and to compare it with $\mathbf{\Sigma}(t)$ defined in Eq (3.9). Consider the operator Vec of matrix $\boldsymbol{\varepsilon}_t$. Then, by the definition of Vec, we have

$$Vec(\boldsymbol{\varepsilon}_t) = Vec(\boldsymbol{\varepsilon}_{.1t}, \boldsymbol{\varepsilon}_{.2t}, \ldots, \boldsymbol{\varepsilon}_{.St}) = (\boldsymbol{\varepsilon}_{.1t}^T, \boldsymbol{\varepsilon}_{.2t}^T, \ldots, \boldsymbol{\varepsilon}_{.St}^T)^T. \tag{3.11}$$

Now, $Vec(\boldsymbol{\varepsilon}_t)$ is a vector with dimension KS, and its variance, $\boldsymbol{\Sigma}^*(t)$, is defined by

$$Var(Vec(\boldsymbol{\varepsilon}_t)) = E[Vec(\boldsymbol{\varepsilon}_t)Vec(\boldsymbol{\varepsilon}_t)^T] = \boldsymbol{\Sigma}^*_{KS\times KS}(t). \tag{3.12}$$

If we partition the covariance matrix $\boldsymbol{\Sigma}^*(t)$ into S^2 sub-matrices each with dimension $K \times K$, so as to have the same pattern as $\boldsymbol{\Sigma}(t)$ defined in Eq (3.9), then the diagonal terms of $\boldsymbol{\Sigma}(t)$ and $\boldsymbol{\Sigma}^*(t)$ will be the same; however, the off-diagonal terms of $\boldsymbol{\Sigma}(t)$ will be the transpose of the off-diagonal terms of $\boldsymbol{\Sigma}^*(t)$. That is:

$$\boldsymbol{\Sigma}_{jj}(t) = \boldsymbol{\Sigma}^*_{jj}(t), \qquad j = 1, 2, \ldots, S, \tag{3.13}$$

$$\boldsymbol{\Sigma}^T_{jj'}(t) = \boldsymbol{\Sigma}^*_{jj'}(t), \qquad j, j' = 1, 2, \ldots, S. \tag{3.14}$$

Since for a given j, $\boldsymbol{\Sigma}_{jj}(t)$ is the covariance matrix of the j^{th} vector time series $Y_{.jt}$ in the time series matrix $\mathbf{Y}_t$, and since $\boldsymbol{\Sigma}_{jj}(t)$ is symmetric, therefore, in general, we have

$$\boldsymbol{\Sigma}^T_{jj'}(t) = \boldsymbol{\Sigma}^*_{jj'}(t), \qquad j, j' = 1, 2, \ldots, S. \tag{3.15}$$

Example 3.2.3 *Let $K = 4$ and $S = 3$, then the expectation of the Kronecker product $\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T$, $\boldsymbol{\Psi}$, is given by*

$$E[\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T] = E\Big(\begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \\ \varepsilon_{41} & \varepsilon_{42} & \varepsilon_{43} \end{bmatrix} \otimes \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} & \varepsilon_{31} & \varepsilon_{41} \\ \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{32} & \varepsilon_{42} \\ \varepsilon_{13} & \varepsilon_{23} & \varepsilon_{33} & \varepsilon_{43} \end{bmatrix} \Big)$$

$$= \left(\begin{array}{cccc|cccc|cccc}
\sigma_{11,11} & \sigma_{11,21} & \sigma_{11,31} & \sigma_{11,41} & \sigma_{12,11} & \sigma_{12,21} & \sigma_{12,31} & \sigma_{12,41} & \sigma_{13,11} & \sigma_{13,21} & \sigma_{13,31} & \sigma_{13,41} \\
\sigma_{11,12} & \sigma_{11,22} & \sigma_{11,32} & \sigma_{11,42} & \sigma_{12,12} & \sigma_{12,22} & \sigma_{12,32} & \sigma_{12,42} & \sigma_{13,12} & \sigma_{13,22} & \sigma_{13,32} & \sigma_{13,42} \\
\sigma_{11,13} & \sigma_{11,23} & \sigma_{11,33} & \sigma_{11,43} & \sigma_{12,13} & \sigma_{12,23} & \sigma_{12,33} & \sigma_{12,43} & \sigma_{13,13} & \sigma_{13,23} & \sigma_{13,33} & \sigma_{13,43} \\
\hline
\sigma_{21,11} & \sigma_{21,21} & \sigma_{21,31} & \sigma_{21,41} & \sigma_{22,11} & \sigma_{22,21} & \sigma_{22,31} & \sigma_{22,41} & \sigma_{23,11} & \sigma_{23,21} & \sigma_{23,31} & \sigma_{23,41} \\
\sigma_{21,12} & \sigma_{21,22} & \sigma_{21,32} & \sigma_{21,42} & \sigma_{22,12} & \sigma_{22,22} & \sigma_{22,32} & \sigma_{22,42} & \sigma_{23,12} & \sigma_{23,22} & \sigma_{23,32} & \sigma_{23,42} \\
\sigma_{21,13} & \sigma_{21,23} & \sigma_{21,33} & \sigma_{21,43} & \sigma_{22,13} & \sigma_{22,23} & \sigma_{22,33} & \sigma_{22,43} & \sigma_{23,13} & \sigma_{23,23} & \sigma_{23,33} & \sigma_{23,43} \\
\hline
\sigma_{31,11} & \sigma_{31,21} & \sigma_{31,31} & \sigma_{31,41} & \sigma_{32,11} & \sigma_{32,21} & \sigma_{32,31} & \sigma_{32,41} & \sigma_{33,11} & \sigma_{33,21} & \sigma_{33,31} & \sigma_{33,41} \\
\sigma_{31,12} & \sigma_{31,22} & \sigma_{31,32} & \sigma_{31,42} & \sigma_{32,12} & \sigma_{32,22} & \sigma_{32,32} & \sigma_{32,42} & \sigma_{33,12} & \sigma_{33,22} & \sigma_{33,32} & \sigma_{33,42} \\
\sigma_{31,13} & \sigma_{31,23} & \sigma_{31,33} & \sigma_{31,43} & \sigma_{32,13} & \sigma_{32,23} & \sigma_{32,33} & \sigma_{32,43} & \sigma_{33,13} & \sigma_{33,23} & \sigma_{33,33} & \sigma_{33,43} \\
\hline
\sigma_{41,11} & \sigma_{41,21} & \sigma_{41,31} & \sigma_{41,41} & \sigma_{42,11} & \sigma_{42,21} & \sigma_{42,31} & \sigma_{42,41} & \sigma_{43,11} & \sigma_{43,21} & \sigma_{43,31} & \sigma_{43,41} \\
\sigma_{41,12} & \sigma_{41,22} & \sigma_{41,32} & \sigma_{41,42} & \sigma_{42,12} & \sigma_{42,22} & \sigma_{42,32} & \sigma_{42,42} & \sigma_{43,12} & \sigma_{43,22} & \sigma_{43,32} & \sigma_{43,42} \\
\sigma_{41,13} & \sigma_{41,23} & \sigma_{41,33} & \sigma_{41,43} & \sigma_{42,13} & \sigma_{42,23} & \sigma_{42,33} & \sigma_{42,43} & \sigma_{43,13} & \sigma_{43,23} & \sigma_{43,33} & \sigma_{43,43}
\end{array} \right)$$

where $\sigma_{ij,i'j'} = E[\varepsilon_{ij}\varepsilon_{i'j'}]$. *Then, the variance-covariance matrix of the* 4×3 *white noise matrix* $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\Sigma}_{12\times 12}$, *can be obtained by premultiplying* $\boldsymbol{\Psi}$ *by the transformation matrix* $\mathbf{T}$ *given in Eq* (3.10) *as*

$$\boldsymbol{\Sigma} = \mathbf{T}\boldsymbol{\Psi} = \left(\begin{array}{cccc|cccc|cccc}
\sigma_{11,11} & \sigma_{11,21} & \sigma_{11,31} & \sigma_{11,41} & \sigma_{12,11} & \sigma_{12,21} & \sigma_{12,31} & \sigma_{12,41} & \sigma_{13,11} & \sigma_{13,21} & \sigma_{13,31} & \sigma_{13,41} \\
\sigma_{21,11} & \sigma_{21,21} & \sigma_{21,31} & \sigma_{21,41} & \sigma_{22,11} & \sigma_{22,21} & \sigma_{22,31} & \sigma_{22,41} & \sigma_{23,11} & \sigma_{23,21} & \sigma_{23,31} & \sigma_{23,41} \\
\sigma_{31,11} & \sigma_{31,21} & \sigma_{31,31} & \sigma_{31,41} & \sigma_{32,11} & \sigma_{32,21} & \sigma_{32,31} & \sigma_{32,41} & \sigma_{33,11} & \sigma_{33,21} & \sigma_{33,31} & \sigma_{33,41} \\
\sigma_{41,11} & \sigma_{41,21} & \sigma_{41,31} & \sigma_{41,41} & \sigma_{42,11} & \sigma_{42,21} & \sigma_{42,31} & \sigma_{42,41} & \sigma_{43,11} & \sigma_{43,21} & \sigma_{43,31} & \sigma_{43,41} \\
\hline
\sigma_{11,12} & \sigma_{11,22} & \sigma_{11,32} & \sigma_{11,42} & \sigma_{12,12} & \sigma_{12,22} & \sigma_{12,32} & \sigma_{12,42} & \sigma_{13,12} & \sigma_{13,22} & \sigma_{13,32} & \sigma_{13,42} \\
\sigma_{21,12} & \sigma_{21,22} & \sigma_{21,32} & \sigma_{21,42} & \sigma_{22,12} & \sigma_{22,22} & \sigma_{22,32} & \sigma_{22,42} & \sigma_{23,12} & \sigma_{23,22} & \sigma_{23,32} & \sigma_{23,42} \\
\sigma_{31,12} & \sigma_{31,22} & \sigma_{31,32} & \sigma_{31,42} & \sigma_{32,12} & \sigma_{32,22} & \sigma_{32,32} & \sigma_{32,42} & \sigma_{33,12} & \sigma_{33,22} & \sigma_{33,32} & \sigma_{33,42} \\
\sigma_{41,12} & \sigma_{41,22} & \sigma_{41,32} & \sigma_{41,42} & \sigma_{42,12} & \sigma_{42,22} & \sigma_{42,32} & \sigma_{42,42} & \sigma_{43,12} & \sigma_{43,22} & \sigma_{43,32} & \sigma_{43,42} \\
\hline
\sigma_{11,13} & \sigma_{11,23} & \sigma_{11,33} & \sigma_{11,43} & \sigma_{12,13} & \sigma_{12,23} & \sigma_{12,33} & \sigma_{12,43} & \sigma_{13,13} & \sigma_{13,23} & \sigma_{13,33} & \sigma_{13,43} \\
\sigma_{21,13} & \sigma_{21,23} & \sigma_{21,33} & \sigma_{21,43} & \sigma_{22,13} & \sigma_{22,23} & \sigma_{22,33} & \sigma_{22,43} & \sigma_{23,13} & \sigma_{23,23} & \sigma_{23,33} & \sigma_{23,43} \\
\sigma_{31,13} & \sigma_{31,23} & \sigma_{31,33} & \sigma_{31,43} & \sigma_{32,13} & \sigma_{32,23} & \sigma_{32,33} & \sigma_{32,43} & \sigma_{33,13} & \sigma_{33,23} & \sigma_{33,33} & \sigma_{33,43} \\
\sigma_{41,13} & \sigma_{41,23} & \sigma_{41,33} & \sigma_{41,43} & \sigma_{42,13} & \sigma_{42,23} & \sigma_{42,33} & \sigma_{42,43} & \sigma_{43,13} & \sigma_{43,23} & \sigma_{43,33} & \sigma_{43,43}
\end{array} \right).$$

Therefore, the variance-covariance matrix $\mathbf{\Sigma}$ *in this example can be partitioned as*

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} & \mathbf{\Sigma}_{33} \end{pmatrix}$$

where

$$\mathbf{\Sigma}^T_{jj'} = E[\boldsymbol{\varepsilon}_{.j}\boldsymbol{\varepsilon}^T_{.j'}] = E\Big([\begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \end{bmatrix} \begin{bmatrix} \varepsilon_{1j'} & \varepsilon_{2j'} & \varepsilon_{3j'} & \varepsilon_{4j'} \end{bmatrix}\Big).$$

3.3 Matrix Autoregressive Process of order one

Linear models are easy to understand and interpret, are sufficiently accurate for most problems, and are also mathematically simple. Therefore, linear functions of past observations (lags) of $\mathbf{Y}_t$ can be good candidates for the function $f(.)$ in Eq (3.3). In time series terminology, these kinds of functions are called autoregressive functions. In particular, for autoregressive models, the current value of the process is a linear function of past values (lags) plus an observational error.

In this section, we expand the concepts behind univariate autoregressive models (AR) and vector autoregressive models (VAR) to define matrix autoregressive time series models (MAR). To this end, we start with the definition of a univariate autoregressive model (AR) and the definition of a vector autoregressive model (VAR). After that, we extend these two definitions to matrix autoregressive models. Let us look at these autoregressive models of order one, in turn. Let $\mathbf{\Gamma}$ be the covariance matrix of $\mathbf{Y}_t$.

- **AR(1):** If the covariance matrix of $\mathbf{Y}_t$, $\mathbf{\Gamma}$, is an identity matrix ($\mathbf{\Gamma} = \mathbf{I}_{KS}$), then all entries of the time series matrix $\mathbf{Y}_t$ are independent, and each y_{ijt} can be considered as a univariate autoregressive time series. There are many well known books for univariate time series analysis, e.g., Box et al. (1994, 2008), Brockwell and Davis (1991), William (2006). By definition of an AR(1) model, for a given time series y_{ijt}, we have

$$y_{ijt} = \mu_{ij} + a_{ij} y_{ij(t-1)} + \varepsilon_{ijt}, \qquad i = 1, 2, \ldots, K, \ j = 1, 2, \ldots, S, \tag{3.16}$$

where μ_{ij} is the intercept, a_{ij} is a coefficient parameter, and ε_{ijt} is white noise such that $E(\varepsilon_{ijt}) = 0$, $Var(\varepsilon_{ijt}) = E(\varepsilon^2_{ijt}) = \sigma_{ij}$, and $E(\varepsilon_{ijt}\varepsilon_{ijt'}) = 0, t \neq t'$.

- **VAR(1):** Assume that the covariance matrix $\mathbf{\Gamma}$ is a block diagonal matrix. That is, for every $j \neq j'$, $\mathbf{\Gamma}_{jj'} = \mathbf{0}$ and for every j, $\mathbf{\Gamma}_{jj} \neq \mathbf{0}$. Then, all vectors $Y_{.1t}$, $Y_{.2t}$, $\ldots$, $Y_{.St}$ will be independent of each other and each of these vectors can be considered as a vector autoregressive time series (VAR). Lütkepohl (1991, 2006), Hannan (1970), and William (2006) are very good references for VAR models. From the definition of a VAR model of order one (VAR(1)), we have

$$Y_{.jt} = \boldsymbol{\mu}_{.j} + \mathbf{A}_j Y_{.j(t-1)} + \boldsymbol{\varepsilon}_{.jt}, \qquad j = 1, 2, \ldots, S, \tag{3.17}$$

where $Y_{.jt} = (y_{1jt}, y_{2jt}, \ldots, y_{Kjt})^T$ is the j^{th} $K-$dimensional vector time series, $\mathbf{A}_j$ is a $K \times K$ matrix of coefficient parameters $(a_{il}, i, l = 1, 2, \ldots, K)$, and $\boldsymbol{\varepsilon}_{.jt} = (\varepsilon_{1jt}, \varepsilon_{2jt}, \ldots, \varepsilon_{Kjt})^T$ is a $K-$dimensional white noise vector such that $E(\boldsymbol{\varepsilon}_{.jt}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.jt}) = \mathbf{\Sigma}_{jj}$, and $E(\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.jt'}) = \mathbf{0}$, $t \neq t'$.

- **MAR(1):** Now, assume that none of the previous assumptions are satisfied for the covariance matrix $\mathbf{\Gamma}$. Then, we will have a matrix of time series which, like those for the AR and VAR models, is needed to define a new and appropriate model for matrix time series.

We use the idea of the definition of VAR models in Eq (3.17) to define a new model for an MAR process and extend the properties of the VAR model to the MAR model. Note that in a vector autoregressive time series, each variable of the vector time series is a linear combination of its past observations and past observations of other variables. For instance, each univariate time series $y_{ijt}, i = 1, 2, \ldots, K$, at time t, in the vector time series $Y_{.jt}$ has the feature

$$y_{ijt} = a_{i1}y_{1j(t-1)} + a_{i2}y_{2j(t-1)} + \ldots + a_{iK}y_{Kj(t-1)} + \varepsilon_{ijt}, \qquad j = 1, 2 \ldots, S. \tag{3.18}$$

Therefore, each $y_{ijt}, i = 1, 2, \ldots, K$, in $Y_{.jt}$ of Eq (3.17) is a linear combination of its past values and past values of other variables. Now, we extend the idea of vector time series to a matrix time series $\mathbf{Y}_t$. First, like the AR(1) and VAR(1) models, the model of a matrix autoregressive time series of order one (MAR(1)) is considered as

$$\mathbf{Y}_t = \boldsymbol{\mu} + F(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \tag{3.19}$$

where $F(\mathbf{Y}_{t-1})$ needs to be defined such that, as for VAR processes, each y_{ijt} is a linear combination of its past values of its other variables, and of the past values of all other series in the matrix $\mathbf{Y}_t$. To this end, the linear matrix function $F(\mathbf{Y}_{t-1})$ is defined by

$$F(\mathbf{Y}_{t-1}) = \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1}\mathbf{E}_{rj} \tag{3.20}$$

where $\mathbf{A}_r^j, r, j = 1, 2, \ldots, S$, are matrices of parameters with dimension $K \times K$, with elements a_{lf}^{rj}, defined by

$$\mathbf{A}_r^j = (a_{lf}^{rj}) = \begin{bmatrix} a_{11}^{rj} & a_{12}^{rj} & \dots & a_{1K}^{rj} \\ a_{21}^{rj} & a_{22}^{rj} & \dots & a_{2K}^{rj} \\ \vdots & & \ddots & \vdots \\ a_{K1}^{rj} & a_{K2}^{rj} & \dots & a_{KK}^{rj} \end{bmatrix}, \qquad r, j = 1, 2, \dots, S, \tag{3.21}$$

and $\mathbf{E}_{rj}, r, j = 1, 2, \dots, S$, are the $S \times S$ matrices whose elements $(e_{lf})_{rj}$ are zero or one given by

$$(e_{lf})_{rj} = \begin{cases} 1, & \text{if } \ l = r \text{ and } f = j \\ 0, & \text{otherwise.} \end{cases}, \qquad l, f = 1, 2, \dots, S. \tag{3.22}$$

Therefore, the matrix autoregressive model of order one defined in Eq (3.19) can be rewritten as

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + F(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu} + \sum_{j=1}^{S} \sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1} \mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t, \qquad t = 0, \pm 1, \pm 2, \dots, \end{aligned} \tag{3.23}$$

where $\boldsymbol{\varepsilon}_t$ is a matrix white noise such that $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T) = \boldsymbol{\Psi}$, and $E(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_{t'}^T) = \mathbf{0}$, $t \neq t'$.

To illustrate the model MAR(1) defined in Eq (3.23) more fully, consider Example 3.3.1.

Example 3.3.1 *Let $\mathbf{Y}_t$ be a matrix time series with dimension $K = 4$ and $S = 3$. Then, Eq (3.23) can be expanded as, for $t = 1, 2, \dots$,*

$$\begin{aligned} \mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^{3} \sum_{r=1}^{3} \mathbf{A}_r^j \mathbf{Y}_{t-1} \mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t \qquad (3.24) \\ = \boldsymbol{\mu} + \mathbf{A}_1^1 \mathbf{Y}_{t-1} \mathbf{E}_{11} + \mathbf{A}_2^1 \mathbf{Y}_{t-1} \mathbf{E}_{21} + \mathbf{A}_3^1 \mathbf{Y}_{t-1} \mathbf{E}_{31} + \mathbf{A}_1^2 \mathbf{Y}_{t-1} \mathbf{E}_{12} + \mathbf{A}_2^2 \mathbf{Y}_{t-1} \mathbf{E}_{22} \\ + \mathbf{A}_3^2 \mathbf{Y}_{t-1} \mathbf{E}_{32} + \mathbf{A}_1^3 \mathbf{Y}_{t-1} \mathbf{E}_{13} + \mathbf{A}_2^3 \mathbf{Y}_{t-1} \mathbf{E}_{23} + \mathbf{A}_3^3 \mathbf{Y}_{t-1} \mathbf{E}_{33} + \boldsymbol{\varepsilon}_t. \end{aligned}$$

For instance, all the details of one of these terms, say $A_3^2\mathbf{Y}_{t-1}E_{32}$, are given by

$$\mathbf{A}_3^2\mathbf{Y}_{t-1}\mathbf{E}_{32} = \begin{bmatrix} a_{11}^{32} & a_{12}^{32} & a_{13}^{32} & a_{14}^{32} \\ a_{21}^{32} & a_{22}^{32} & a_{23}^{32} & a_{24}^{32} \\ a_{31}^{32} & a_{32}^{32} & a_{33}^{32} & a_{34}^{32} \\ a_{41}^{32} & a_{42}^{32} & a_{43}^{32} & a_{44}^{32} \end{bmatrix} \begin{bmatrix} y_{11(t-1)} & y_{12(t-1)} & y_{13(t-1)} \\ y_{21(t-1)} & y_{22(t-1)} & y_{23(t-1)} \\ y_{31(t-1)} & y_{32(t-1)} & y_{33(t-1)} \\ y_{41(t-1)} & y_{42(t-1)} & y_{43(t-1)} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{4} a_{1i}^{32} y_{i1} & \sum_{i=1}^{4} a_{1i}^{32} y_{i2(t-1)} & \sum_{i=1}^{4} a_{1i}^{32} y_{i3(t-1)} \\ \sum_{i=1}^{4} a_{2i}^{32} y_{i1(t-1)} & \sum_{i=1}^{4} a_{2i}^{32} y_{i2(t-1)} & \sum_{i=1}^{4} a_{2i}^{32} y_{i3(t-1)} \\ \sum_{i=1}^{4} a_{3i}^{32} y_{i1(t-1)} & \sum_{i=1}^{4} a_{3i}^{32} y_{i2(t-1)} & \sum_{i=1}^{4} a_{3i}^{32} y_{i3(t-1)} \\ \sum_{i=1}^{4} a_{4i}^{32} y_{i1(t-1)} & \sum_{i=1}^{4} a_{4i}^{32} y_{i2(t-1)} & \sum_{i=1}^{4} a_{4i}^{32} y_{i3(t-1)} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \sum_{i=1}^{4} a_{1i}^{32} y_{i3(t-1)} & 0 \\ 0 & \sum_{i=1}^{4} a_{2i}^{32} y_{i3(t-1)} & 0 \\ 0 & \sum_{i=1}^{4} a_{3i}^{32} y_{i3(t-1)} & 0 \\ 0 & \sum_{i=1}^{4} a_{4i}^{32} y_{i3(t-1)} & 0 \end{bmatrix}.$$

Now, for given $j = 2$, the sum over all r of the second term of the model given in Eq (3.24), $\sum_{r=1}^{3} \mathbf{A}_r^2\mathbf{Y}_{t-1}\mathbf{E}_{r2}$, *is equal to*

$$\sum_{r=1}^{3} \mathbf{A}_r^2\mathbf{Y}_{t-1}\mathbf{E}_{r2} = \begin{bmatrix} 0 & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{1i}^{r2} y_{ir(t-1)} & 0 \\ 0 & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{2i}^{r2} y_{ir(t-1)} & 0 \\ 0 & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{3i}^{r2} y_{ir(t-1)} & 0 \\ 0 & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{4i}^{r2} y_{ir(t-1)} & 0 \end{bmatrix}.$$

Eventually, summing over all $j = 1, 2, 3$, the matrix autoregressive process in Eq (3.24) *can be written as*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \begin{bmatrix} \sum_{r=1}^{3}\sum_{i=1}^{4} a_{1i}^{r1} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{1i}^{r2} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{1i}^{r3} y_{ir(t-1)} \\ \sum_{r=1}^{3}\sum_{i=1}^{4} a_{2i}^{r1} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{2i}^{r2} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{2i}^{r3} y_{ir(t-1)} \\ \sum_{r=1}^{3}\sum_{i=1}^{4} a_{3i}^{r1} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{3i}^{r2} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{3i}^{r3} y_{ir(t-1)} \\ \sum_{r=1}^{3}\sum_{i=1}^{4} a_{4i}^{r1} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{4i}^{r2} y_{ir(t-1)} & \sum_{r=1}^{3}\sum_{i=1}^{4} a_{4i}^{r3} y_{ir(t-1)} \end{bmatrix} + \boldsymbol{\varepsilon}_t. \tag{3.25}$$

Therefore, similar to Eq (3.18) *for the vector time series, each univariate time series of matrix time series* $\mathbf{Y}_t$ *is a linear combination of past values of its own and other series and all variables. For instance, for* y_{23t} *in Eq* (3.25)*, we have*

$$y_{23t} = \mu_{23} + \sum_{r=1}^{3}\sum_{i=1}^{4} a_{2i}^{r3} y_{ir(t-1)} + \varepsilon_{23t}.$$

3.3.1 A more suitable and appropriate definition for $F(\mathbf{Y}_{t-1})$

For the sake of brevity and convenience in notation, and also in order to find an appropriate linear model for a matrix time series of order one (MAR(1)) which is compatible with the standard linear model notation and also to be compatible with AR(1) and VAR(1) models, $F(\mathbf{Y}_{t-1})$ can be written as

$$F(\mathbf{Y}_{t-1}) = \mathbf{A} f(\mathbf{Y}_{t-1}). \tag{3.26}$$

Therefore, we can rewrite the matrix time series $\mathbf{Y}_t$ defined in Eq (3.23) as

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + F(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu} + \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1} \mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu} + \mathbf{A} f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t, \qquad\qquad t = 0, \pm 1, \pm 2, \ldots, \end{aligned} \tag{3.27}$$

where $\mathbf{A}$ is a $1 \times S^2-$dimensional block matrix of coefficient parameters. Each block, $\mathbf{A}_r^j$, is a $K \times K$ dimensional matrix. Thus, $\mathbf{A}$ is a matrix with dimension $K \times KS^2$ given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^1 & \mathbf{A}_2^1 & \dots & \mathbf{A}_S^1 & \mathbf{A}_1^2 & \mathbf{A}_2^2 & \dots & \mathbf{A}_S^2 & \dots & \mathbf{A}_1^j & \mathbf{A}_2^j & \dots & \mathbf{A}_S^j & \dots & \mathbf{A}_1^S & \mathbf{A}_2^S & \dots & \mathbf{A}_S^S \end{bmatrix} \tag{3.28}$$

and $f(\mathbf{Y}_{t-1})$ is a $S^2 \times 1-$dimensional block matrix where the l^{th} block, $l = 1, 2, \dots, S^2$, is the $K \times S-$dimensional matrix $\mathbf{Y}_{t-1}\mathbf{E}_{rj}$, where $\mathbf{E}_{rj}$ is given in Eq (3.22). Therefore, $f(\mathbf{Y}_{t-1})$ is a matrix with dimension $KS^2 \times S$ and defined by

$$f(\mathbf{Y}_{t-1}) = \begin{bmatrix} f_1(\mathbf{Y}_{t-1}) \\ f_2(\mathbf{Y}_{t-1}) \\ \vdots \\ f_j(\mathbf{Y}_{t-1}) \\ \vdots \\ f_S(\mathbf{Y}_{t-1}) \end{bmatrix} \tag{3.29}$$

where $f_j(\mathbf{Y}_{t-1})$ is defined as

$$f_j(\mathbf{Y}_{t-1}) = \begin{bmatrix} \mathbf{Y}_{t-1}\mathbf{E}_{1j} \\ \mathbf{Y}_{t-1}\mathbf{E}_{2j} \\ \vdots \\ \mathbf{Y}_{t-1}\mathbf{E}_{rj} \\ \vdots \\ \mathbf{Y}_{t-1}\mathbf{E}_{Sj} \end{bmatrix} = \begin{bmatrix} \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{j-1}, Vec(\mathbf{Y}_{t-1}), \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{S-j} \end{bmatrix}_{KS \times S} = \mathbf{e}_j^T \otimes Vec(\mathbf{Y}_{t-1}) \tag{3.30}$$

where $\mathbf{e}_j = (\overbrace{0, 0, \dots, 0}^{j-1}, 1, \overbrace{0, \dots, 0}^{S-j})^T$. Therefore, by substituting the right-hand side of Eq (3.30) for each $j = 1, 2, \dots, S$, into Eq (3.29), it can be shown that $f(\mathbf{Y}_{t-1})$ is a $S \times S$ block matrix function with block size $KS \times 1$. Moreover, the diagonal block entries of $f(\mathbf{Y}_{t-1})$ are $Vec(\mathbf{Y}_{t-1})$, and its off-diagonal elements are $KS \times 1$ zero vectors. That is,

$$f(\mathbf{Y}_{t-1}) = \begin{bmatrix} Vec(\mathbf{Y}_{t-1}) & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & Vec(\mathbf{Y}_{t-1}) & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & Vec(\mathbf{Y}_{t-1}) \end{bmatrix}_{KS^2 \times S} = \mathbf{I}_S \otimes Vec(\mathbf{Y}_{t-1}). \tag{3.31}$$

Let us define $\mathbf{A}_j^{\dagger}$ to be a block matrix with dimension $1 \times S$, and each block is a $K \times K$ coefficient matrix $\mathbf{A}_r^j$ which are the elements of matrix $\mathbf{A}$ in Eq (3.28). In fact, $\mathbf{A}_j^{\dagger}$ has dimension $K \times KS$ and is defined as

$$\mathbf{A}_j^{\dagger} = \begin{bmatrix} \mathbf{A}_1^j & \mathbf{A}_2^j & \dots, \mathbf{A}_S^j \end{bmatrix}, \qquad j = 1, 2, \dots, S. \tag{3.32}$$

Therefore, from the definition of $\mathbf{A}_j^{\dagger}$ in Eq (3.32), we see that the matrix $\mathbf{A}$ defined in Eq (3.28) can be rewritten as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^{\dagger} & \mathbf{A}_2^{\dagger} & \cdots & \mathbf{A}_j^{\dagger} & \cdots & \mathbf{A}_S^{\dagger} \end{bmatrix}. \tag{3.33}$$

In a manner similar to that used in section 3.2 where we compared the variance-covariance matrix of the white noise matrix $\boldsymbol{\varepsilon}_t$ with the variance-covariance matrix of Vec of $\boldsymbol{\varepsilon}_t$ (see Eqs (3.12)-(3.15)), it is worth while to compare the MAR(1) model with the corresponding vector autoregressive process of order one. Toward this end, let us take Vec on both sides of Eq (3.23), i.e.,

$$Vec(\mathbf{Y}_t) \quad = \quad Vec(\boldsymbol{\mu}) + \sum_{j=1}^{S}\sum_{r=1}^{S} Vec(\mathbf{A}_r^j \mathbf{Y}_{t-1}\mathbf{E}_{rj}) + Vec(\boldsymbol{\varepsilon}_t). \tag{3.34}$$

Applying the Vec operator rule $Vec(ABC) = (C^T \otimes A)Vec(B)$ for the middle term on the right-hand side of Eq (3.34), and using the fact that $E_{rj}^T = E_{jr}$, we have

$$Vec(\mathbf{Y}_t) \quad = \quad Vec(\boldsymbol{\mu}) + \sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{E}_{jr} \otimes \mathbf{A}_r^j)Vec(\mathbf{Y}_{t-1}) + Vec(\boldsymbol{\varepsilon}_t). \tag{3.35}$$

Therefore, the corresponding vector autoregressive model of a MAR(1) process is given by

$$Vec(\mathbf{Y}_t) \quad = \quad Vec(\boldsymbol{\mu}) + \mathbf{B}Vec(\mathbf{Y}_{t-1}) + Vec(\boldsymbol{\varepsilon}_t) \tag{3.36}$$

where $\mathbf{B} = \sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{E}_{jr} \otimes \mathbf{A}_r^j)$ is the $S \times S-$dimensional block matrix of $K \times K$ coefficient matrices $\mathbf{A}_r^j$ given in Eq (3.21), and given by

$$\mathbf{B} = \sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{E}_{jr} \otimes \mathbf{A}_r^j) = \begin{pmatrix} \mathbf{A}_1^1 & \mathbf{A}_2^1 & \dots & \mathbf{A}_S^1 \\ \mathbf{A}_1^2 & \mathbf{A}_2^2 & \dots & \mathbf{A}_S^2 \\ \vdots & & & \vdots \\ \mathbf{A}_1^S & \mathbf{A}_2^S & \dots & \mathbf{A}_S^S \end{pmatrix} = \begin{bmatrix} \mathbf{A}_1^\dagger \\ \mathbf{A}_2^\dagger \\ \vdots \\ \mathbf{A}_S^\dagger \end{bmatrix}. \tag{3.37}$$

Also, from the dimension of $\mathbf{A}_j^\dagger$, we see that the matrix $\mathbf{B}$ is the coefficient matrix of parameters with dimension $KS \times KS$. Note that, the components of the matrix $\mathbf{B}$, $\mathbf{A}_j^\dagger$, are the same as the components of the coefficient matrix $\mathbf{A}$ in Eq (3.33), except the dimension of $\mathbf{B}$ is $KS \times KS$ whereas the dimension of $\mathbf{A}$ is $K \times KS^2$. Indeed, this follows from the fact that $\mathbf{B}$ is the $Vecb^{KS}(.)$ of $\mathbf{A}$, or $\mathbf{B} = Vec(\mathbf{A})$ by assuming the $\mathbf{A}_j^\dagger, j = 1, 2, \dots, S$, to be scalar.

Notice, the operator $Vecb^{KS}(.)$ is a block Vec operation which executes in the same way as does the operation Vec, except that it takes each block entry with K columns as a scalar entry. For instance, let $\mathbf{W}$ be a block matrix with block entries $\mathbf{W}_{ij}$ such that each entry has dimension $r_i \times c$ given by

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}_{(r_1+r_2)\times 2c}.$$

Then, the operation $Vecb^c(\mathbf{W})$ is defined as

$$Vecb^c(\mathbf{W}) = \begin{pmatrix} \mathbf{W}_{11} \\ \mathbf{W}_{21} \\ \mathbf{W}_{12} \\ \mathbf{W}_{22} \end{pmatrix}_{2(r_1+r_2)\times c}. \tag{3.38}$$

Example 3.3.2 *(Continuation of Example* 3.3.1 *) The MAR*(1) *model when* $K = 4$*,* $S = 3$*, given in Eq* (3.24)*, can be rewritten as*

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}(\mathbf{I}_3 \otimes Vec(\mathbf{Y}_{t-1})) + \boldsymbol{\varepsilon}_t$$

where

$$Vec(\mathbf{Y}_{t-1}) = \begin{bmatrix} y_{11} & y_{21} & y_{31} & y_{41} & y_{12} & y_{22} & y_{32} & y_{42} & y_{13} & y_{23} & y_{33} & y_{43} \end{bmatrix}^T,$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^1 & \mathbf{A}_2^1 & \mathbf{A}_3^1 & \mathbf{A}_1^2 & \mathbf{A}_2^2 & \mathbf{A}_3^2 & \mathbf{A}_1^3 & \mathbf{A}_2^3 & \mathbf{A}_3^3 \end{bmatrix}, \quad \mathbf{A}_r^j = \begin{bmatrix} a_{11}^{rj} & a_{12}^{rj} & a_{13}^{rj} & a_{14}^{rj} \\ a_{21}^{rj} & a_{22}^{rj} & a_{23}^{rj} & a_{24}^{rj} \\ a_{31}^{rj} & a_{32}^{rj} & a_{33}^{rj} & a_{34}^{rj} \\ a_{41}^{rj} & a_{42}^{rj} & a_{43}^{rj} & a_{44}^{rj} \end{bmatrix}.$$

3.4 Stationary Processes

In this section, the stationarity (weak stationarity) and strict stationarity of a matrix time series process will be discussed. Studying stationary processes is fundamental and is an important concept for analyzing a time series.

Definition 3.4.1 *(Weak Stationarity) A stochastic matrix process $\{\mathbf{Y}_t\}$ is (weakly) stationary if it possesses finite first and second moments that are time invariant. In particular, a $K \times S$ matrix time series $\mathbf{Y}_t$ is stationary if, for all t and all $h \in \mathbb{N}$ (the natural numbers), it satisfies*

$$E[\mathbf{Y}_t] = \boldsymbol{\upsilon} < \infty, \tag{3.39}$$

$$E[\mathbf{Y}_{t+h} \otimes \mathbf{Y}_t^T] = \boldsymbol{\Psi}(h) < \infty; \tag{3.40}$$

$\boldsymbol{\Psi}(h)$ will be called the lag function.

The condition given in Eq (3.39) means that, for all t, the mean matrix $\boldsymbol{\upsilon}$ of the matrix time series $\mathbf{Y}_t$ is the same and finite, and the condition based on Eq (3.40) says that the autocovariances of the matrix process are finite and only depend on the lag h but do not depend on t. In the time series literature, the weakly stationary process is also referred to as a covariance or second-order stationary process.

Note that, in contrast to an univariate stationary time series y_t where we have $\gamma(h) = E[y_{(t+h)}y_t] = E[y_{(t-h)}y_t] = \gamma(-h)$, the same is not true for a matrix time series $\mathbf{Y}_t$ where now $\boldsymbol{\Psi}(h) \neq \boldsymbol{\Psi}(-h)$. Instead, the precise correspondence between $\boldsymbol{\Psi}(h)$ and $\boldsymbol{\Psi}(-h)$ can be derived by replacing t in Eq (3.40) with $t-h$, i.e.,

$$\boldsymbol{\Psi}(h) = E[\mathbf{Y}_{(t-h)+h} \otimes \mathbf{Y}_{t-h}^T] = E[\mathbf{Y}_t \otimes \mathbf{Y}_{t-h}^T]. \tag{3.41}$$

Taking the transpose of both sides of (3.41) leads to

$$\boldsymbol{\Psi}^T(h) = E[\mathbf{Y}_t^T \otimes \mathbf{Y}_{t-h}]. \tag{3.42}$$

Furthermore, it can be shown that (see Brewer, 1978)

$$(\mathbf{Y}_t^T \otimes \mathbf{Y}_{t-h}) = \mathbf{T}(\mathbf{Y}_{t-h} \otimes \mathbf{Y}_t^T)\mathbf{T} \tag{3.43}$$

where $\mathbf{T}$ is the transformation matrix defined in Eqs (3.6) and (3.7). Then, by taking the expectation on both sides of (3.43), and using Eqs (3.40)-(3.42), we have

$$\boldsymbol{\Psi}^T(h) = E[\mathbf{Y}_t^T \otimes \mathbf{Y}_{t-h}] = \mathbf{T}E[\mathbf{Y}_{t-h} \otimes \mathbf{Y}_t^T]\mathbf{T} = \mathbf{T}\boldsymbol{\Psi}(-h)\mathbf{T}. \tag{3.44}$$

Therefore, the precise relationship between the lag functions $\boldsymbol{\Psi}(h)$ and $\boldsymbol{\Psi}(-h)$ is given by

$$\boldsymbol{\Psi}^T(h) = \mathbf{T}\boldsymbol{\Psi}(-h)\mathbf{T}. \tag{3.45}$$

Definition 3.4.2 *(Strict Stationarity) A stochastic matrix process $\{\mathbf{Y}_t\}$ is strictly stationary if for given $t_1, t_2, \ldots, t_n$, the probability distributions of the random matrices $\mathbf{Y}_{t_1}$, $\mathbf{Y}_{t_2}$, …, $\mathbf{Y}_{t_n}$ and $\mathbf{Y}_{t_1+h}$, $\mathbf{Y}_{t_2+h}$, …, $\mathbf{Y}_{t_n+h}$ are time invariant for all n and h. That is,*

$$(\mathbf{Y}_{t_1}, \mathbf{Y}_{t_2}, \ldots, \mathbf{Y}_{t_n}) \overset{D}{\equiv} (\mathbf{Y}_{t_1+h}, \mathbf{Y}_{t_2+h}, \ldots, \mathbf{Y}_{t_n+h}) \tag{3.46}$$

where $\overset{D}{\equiv}$ means "equal in distribution".

The definition of strict stationarity means that all moments of a matrix time series are the same along the process. Note that strict stationarity implies weak stationarity as long as the first and second moments of the process exist. In contrast, in general weak stationarity does not imply strict stationarity. However, if the matrix time series $\mathbf{Y}_t$ is a Gaussian process, with the consequence that the probability distributions of $\mathbf{Y}_t$ for all t, follow the matrix normal distribution, then weak stationarity implies strict stationarity.

If a matrix process is (weak) stationary, then one consequence is that the matrix white noise $\boldsymbol{\varepsilon}_t$ elements eventually have to die out. The following proposition gives a condition for stationarity of the MAR(1) model.

Proposition 3.4.1 *(Stationarity Condition) A MAR(1) process is stationary if all eigenvalues of the coefficient matrix $\mathbf{B}$ have modulus less than one.*

In the time series literature, the stationary condition refers to the *stability condition*. The condition in Proposition 3.4.1 is equivalent to the condition that: if all roots of the polynomial $\Pi(x) = |\mathbf{I}_{KS} - \mathbf{B}x|$ lie outside of the unit circle, these conditions holds (see Lütkepohl, 2006, or Hamilton, 1994). That is, for $|x| \leq 1$,

$$|\mathbf{I}_{KS} - \mathbf{B}x| \neq 0. \tag{3.47}$$

3.4.1 Simulation Study

In this section, we will compare stationary and nonstationary MAR(1) models by simulating a MAR(1) process. To simulate a MAR(1) process, first we assume that the matrix white noise $\boldsymbol{\varepsilon}_t$ follows a matrix normal distribution.

The $K \times S$ random matrix $\boldsymbol{\varepsilon}$ is said to have a matrix normal distribution with mean matrix $\mathbf{M}$, row (within) covariance matrix $\mathbf{\Sigma}$ with dimension $K \times K$, and column (between) covariance matrix $\mathbf{\Omega}$ with dimension $S \times S$, if $Vec(\boldsymbol{\varepsilon})$ has a multivariate (vector) normal distribution with mean $Vec(\mathbf{M})$ and covariance matrix $\mathbf{\Omega} \otimes \mathbf{\Sigma}$ (Gupta and Nagar, 2000).

Note that, according to the definition of the matrix white noise in section 3.3, the matrix white noise $\boldsymbol{\varepsilon}_t$ of Eq (3.23) has to have mean zero. To simulate a matrix normal distribution $\boldsymbol{\varepsilon}$ with mean zero ($\mathbf{M} = \mathbf{0}$), and row and column covariance matrices $\mathbf{\Sigma}$ and $\mathbf{\Omega}$, respectively, let

$$\boldsymbol{\varepsilon} = \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{Z} \mathbf{\Omega}^{\frac{1}{2}} \tag{3.48}$$

where $\mathbf{Z}$ is a $K \times S$ matrix of independent standard normal distributions. Then, it is easy to see that the expectation of $\boldsymbol{\varepsilon}$ in Eq (3.48) is zero. Moreover, by taking Vec on both sides of Eq (3.48), and applying the Vec operator's rule $Vec(ABC) = (C^T \otimes A)Vec(B)$, we have

$$Vec(\boldsymbol{\varepsilon}) = (\mathbf{\Omega}^{\frac{1}{2}} \otimes \mathbf{\Sigma}^{\frac{1}{2}})Vec(\mathbf{Z}). \tag{3.49}$$

We know that any linear combination of independent normal random variables is normal. Since $\mathbf{Z}$ is a matrix of independent standard normals, $Vec(\boldsymbol{\varepsilon})$ in Eq (3.49) is normal. Hence, we just need to find the mean and variance of $Vec(\boldsymbol{\varepsilon})$. Obviously, the mean is zero, i.e., $E[Vec(\boldsymbol{\varepsilon})] = \mathbf{0}$. To obtain the variance, taking the variance on both sides of Eq (3.49) leads to

$$\begin{aligned} Var(Vec(\boldsymbol{\varepsilon})) &= (\boldsymbol{\Omega}^{\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{\frac{1}{2}}) Var(Vec(\mathbf{Z}))(\boldsymbol{\Omega}^{\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{\frac{1}{2}})^T \\ &= (\boldsymbol{\Omega}^{\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{\frac{1}{2}})\mathbf{I}(\boldsymbol{\Omega}^{\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{\frac{1}{2}})^T. \end{aligned} \tag{3.50}$$

Now, by using the Kronecker product rule $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ in Eq (3.50), the variance of the Vec of the random matrix $\boldsymbol{\varepsilon}$ is given by

$$Var(Vec(\boldsymbol{\varepsilon})) = (\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}). \tag{3.51}$$

Therefore, by the definition of a matrix normal distribution, $\boldsymbol{\varepsilon}$ in Eq (3.48) has a matrix normal distribution with mean zero and variance-covariance $\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$.

Hence, to generate a matrix normal distribution, first, we can generate a $K \times S$ matrix of independent standard normal distribution $\mathbf{Z}$; then, the simulated matrix normal $\boldsymbol{\varepsilon}$ will be obtained by using Eq (3.48).

Eventually, a MAR(1) process can be simulated by using the MAR(1) in Eqs (3.23) or (3.27).

Example 3.4.1 *Figure* 3.1 *shows a stationary* 3×2 *matrix time series. It can be seen that the mean and variance of the series are constant over the time, and the series fluctuate around the mean with a constant variation.*

Example 3.4.2 *Figure* 3.2 *shows a nonstationary* 3×2 *matrix time series. Note that, all four graphs in Figure* 3.2 *are for the same MAR(*1*) model but they are plotted for different*

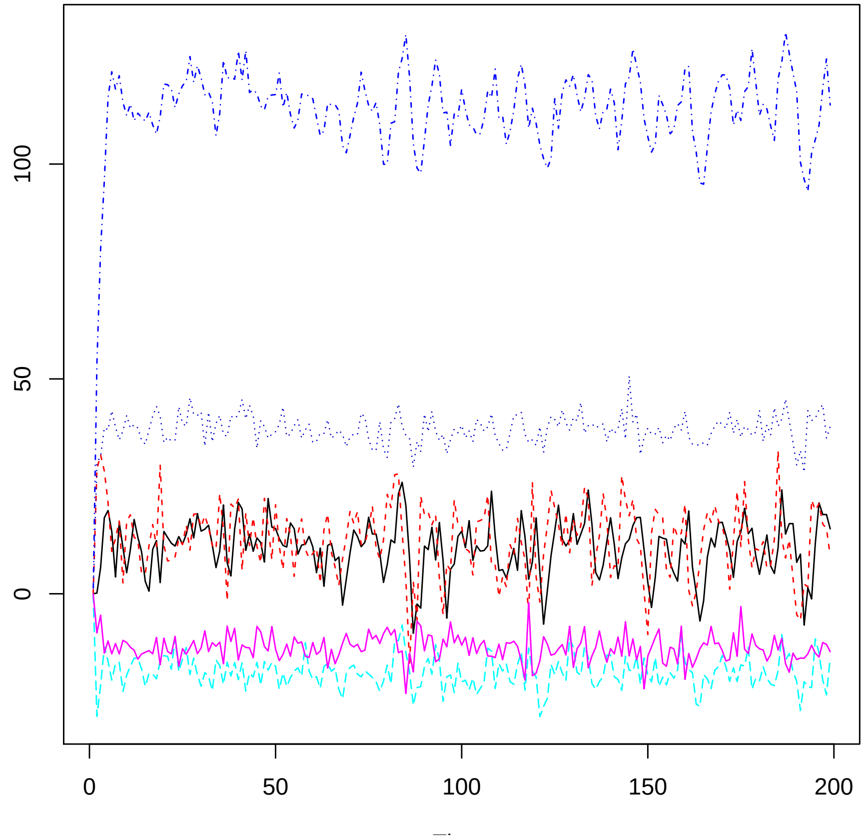

Figure 3.1: A Stationary MAR(1) model

lengths of time to see how they are exploding when increasing their length (time). The left-upper graph has length ten, the right-upper graph has length twenty, the left-lower plot has length eighty, and the right-lower plot has length two hundred. Therefore, in contrast to the stationary matrix series in Figure 3.1*, Figure* 3.2 *shows that the series of nonstationary matrix time series are exploding over time.*

In the sequel, we assume that the matrix time series $\mathbf{Y}_t$ is a stationary process. The computer codes are in Appendix A.

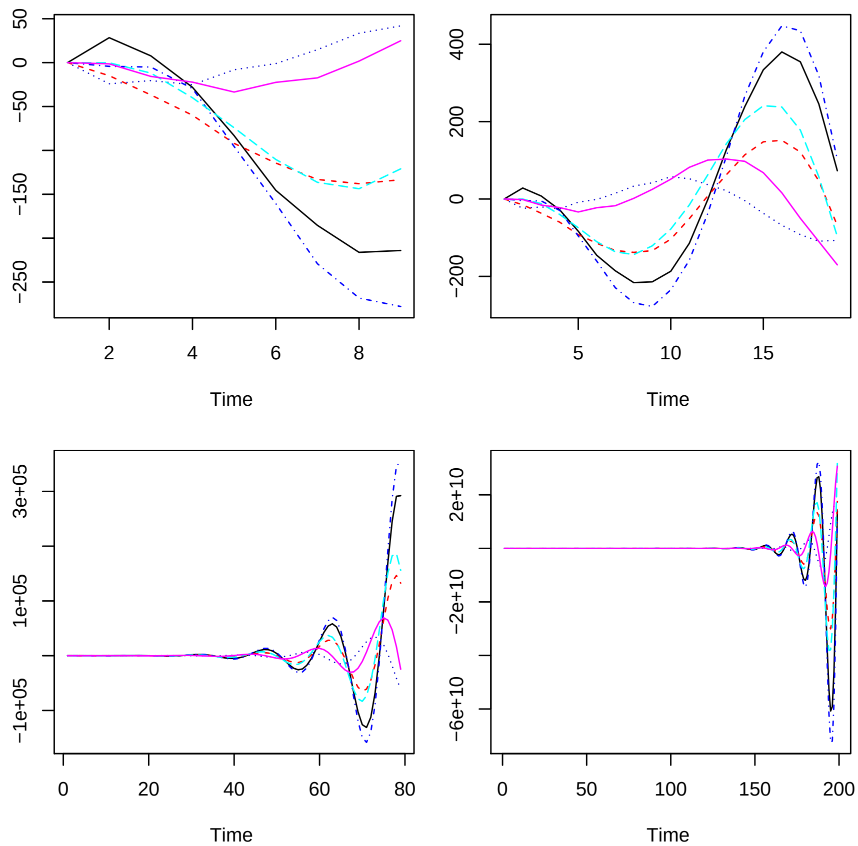

Figure 3.2: A Nonstationary MAR(1) model with different length (Time)

3.5 The Moving Average Representation of MAR(1)

The model defined in Eq (3.27) is a recursive representation of a MAR(1) model. There is an alternative representation based upon Wold's decomposition theorem (Wold, 1938), which states that every weakly stationary stochastic process with no deterministic component can be written as a linear combination of uncorrelated random variables. Therefore, any stationary autoregressive process can be represented as an infinite order moving average process (MA(∞)). In particular, the MAR(1) model can be written as the (infinite) sum of independent random matrices $\boldsymbol{\varepsilon}_t$, $t = 1, 2, \ldots, \infty$. The moving average representation is useful for many purposes, such as finding the autocovariance functions in section 6 below.

In this section, we use the recursive property of an autoregressive model of $\mathbf{Y}_t$ in Eq

(3.27) to find its moving average representation. To this end, we will find the moving average representation of $\mathbf{Y}_t$ in three steps. First of all, without loss of generality, we assume that the mean of the matrix time series $\mathbf{Y}_t$ is zero ($\boldsymbol{\mu} = 0$).

- **First Step:** By using the recursive property of a matrix time series $\mathbf{Y}_t$ as in Eq (3.27), the first step can be started with $f(\mathbf{Y}_{t-1})$. Since we are assuming the matrix time series $\mathbf{Y}_t$ has mean zero, $\mathbf{Y}_t$ in Eq (3.27) can be written as

$$\mathbf{Y}_t = \mathbf{A} f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t. \tag{3.52}$$

With this definition of $\mathbf{Y}_t$, the function $f(\mathbf{Y}_{t-1})$ can be extended as

$$\begin{aligned} f(\mathbf{Y}_{t-1}) &= f(\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1}) \\ &= f(\mathbf{A} f[\mathbf{A} f(\mathbf{Y}_{t-3}) + \boldsymbol{\varepsilon}_{t-2}] + \boldsymbol{\varepsilon}_{t-1}) \\ &= \dots \end{aligned}$$

Now, we can simplify this recursive equation according to the definition of $f(\mathbf{Y}_{t-1})$ in Eq (3.29). In other words, $f(\mathbf{Y}_{t-1})$ can be broken down as

$$f(\mathbf{Y}_{t-1}) = \left[\begin{array}{c} \mathbf{Y}_{t-1}\mathbf{E}_{11} \\ \mathbf{Y}_{t-1}\mathbf{E}_{21} \\ \vdots \\ \mathbf{Y}_{t-1}\mathbf{E}_{S1} \\ \hline \mathbf{Y}_{t-1}\mathbf{E}_{12} \\ \mathbf{Y}_{t-1}\mathbf{E}_{22} \\ \vdots \\ \mathbf{Y}_{t-1}\mathbf{E}_{S2} \\ \hline \vdots \\ \mathbf{Y}_{t-1}\mathbf{E}_{rj} \\ \vdots \\ \hline \mathbf{Y}_{t-1}\mathbf{E}_{1S} \\ \mathbf{Y}_{t-1}\mathbf{E}_{2S} \\ \dots \\ \mathbf{Y}_{t-1}\mathbf{E}_{SS} \end{array}\right] = \left[\begin{array}{c} (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{11} \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{21} \\ \vdots \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{S1} \\ \hline (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{12} \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{22} \\ \vdots \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{S2} \\ \hline \vdots \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{rj} \\ \vdots \\ \hline (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{1S} \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{2S} \\ \vdots \\ (\mathbf{A} f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{SS} \end{array}\right]. \tag{3.53}$$

Also, note that from Eq (3.27), we have

$$\mathbf{A}f(\mathbf{Y}_{t-2}) = \sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj}. \tag{3.54}$$

Hence, we can simplify $f(\mathbf{Y}_{t-1})$ by substituting this Eq (3.54) into each entry of the first term of the right hand side of Eq (3.53). Before that, we need to use two subscripts that are different from r and j to be able to distinguish between $f(\mathbf{Y}_{t-1})$ and $\mathbf{A}f(\mathbf{Y}_{t-2})$. To this end, we consider l and m to be the same as r and j, respectively, in Eq (3.53) for $f(\mathbf{Y}_{t-1})$. I.e., in Eq (3.53) the (r,j) subscripts became (l,m), $(r,j) \equiv (l,m) = 1,2,\ldots,S$. With respect to this convention, and by substituting Eq (3.54) into Eq (3.53), we have

$$f(\mathbf{Y}_{t-1}) = \left[\begin{array}{c}
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{11} \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{21} \\
\vdots \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{S1} \\
\hline
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{12} \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{22} \\
\vdots \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{S2} \\
\hline
\vdots \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{lm} \\
\vdots \\
\hline
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{1S} \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{2S} \\
\vdots \\
(\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j\mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{SS}
\end{array}\right] + \left[\begin{array}{c}
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{11} \\
\boldsymbol{\varepsilon}_{t-1}E_{21} \\
\vdots \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{S1} \\
\hline
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{12} \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{22} \\
\vdots \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{S2} \\
\hline
\vdots \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{lm} \\
\vdots \\
\hline
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{1S} \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{2S} \\
\vdots \\
\boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{SS}
\end{array}\right]. \tag{3.55}$$

We can simplify $f(\mathbf{Y}_{t-1})$ further since we can show from the definition of $\mathbf{E}_{rj}$ (see Eq (3.22)) that

$$\mathbf{E}_{rj}\mathbf{E}_{lm} = \begin{cases} \mathbf{E}_{rm}, & \text{if } \ j = l, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{3.56}$$

Therefore, for any m over all $l = 1, 2, \ldots, S$, we have

$$(\sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{lm} = \sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-2}\mathbf{E}_{rm}, \qquad \text{if } \ j = l. \tag{3.57}$$

Without loss of generality, we can assume that l and m are r and j, respectively, the same as they were before in Eq (3.53). Hence, Eq (3.57) can be rewritten as

$$(\sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-2}\mathbf{E}_{rj})\mathbf{E}_{lm} = \sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-2}\mathbf{E}_{rj}, \qquad \text{if } \ j = l. \tag{3.58}$$

Now, let us consider $f_j(\mathbf{Y}_{t-1})$, the j^{th} entry of the matrix function $f(\mathbf{Y}_{t-1})$, defined in Eq (3.30). By applying Eq (3.53) and Eq (3.55), and substituting Eq (3.57) into all entries of $f_j(\mathbf{Y}_{t-1})$, we have

$$\begin{aligned} f_j(\mathbf{Y}_{t-1}) &= \begin{bmatrix} (\mathbf{A}f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{1j} \\ (\mathbf{A}f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{2j} \\ \vdots \\ (\mathbf{A}f(\mathbf{Y}_{t-2}) + \boldsymbol{\varepsilon}_{t-1})\mathbf{E}_{Sj} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{1j} \\ \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{2j} \\ \vdots \\ \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{Sj} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{r=1}^{S} \mathbf{A}_r^1 \mathbf{Y}_{t-2}\mathbf{E}_{rj} \\ \sum_{r=1}^{S} \mathbf{A}_r^2 \mathbf{Y}_{t-2}\mathbf{E}_{rj} \\ \vdots \\ \sum_{r=1}^{S} \mathbf{A}_r^S \mathbf{Y}_{t-2}\mathbf{E}_{rj} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{1j} \\ \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{2j} \\ \vdots \\ \boldsymbol{\varepsilon}_{t-1}\mathbf{E}_{Sj} \end{bmatrix}. \end{aligned} \tag{3.59}$$

Furthermore, note that, for simplicity and also in order to be able to find a model that is compatible with the classical linear model, the right side of Eq (3.58) can be written as

$$\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-2} \mathbf{E}_{rj} = \mathbf{A}_j^{\dagger} f_m(\mathbf{Y}_{t-2}), \qquad j = 1, 2, \ldots, S, \tag{3.60}$$

where $\mathbf{A}_j^{\dagger}$ is defined in Eq (3.32). Now, we can substitute Eq (3.60) into the entities of the $f_j(\mathbf{Y}_{t-1})$ in Eq (3.59). This leads to

$$\begin{aligned} f_j(\mathbf{Y}_{t-1}) &= \begin{bmatrix} \mathbf{A}_1^{\dagger} f_j(\mathbf{Y}_{t-2}) \\ \mathbf{A}_2^{\dagger} f_j(\mathbf{Y}_{t-2}) \\ \vdots \\ \mathbf{A}_S^{\dagger} f_j(\mathbf{Y}_{t-2}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{t-1} E_{1j} \\ \boldsymbol{\varepsilon}_{t-1} E_{2j} \\ \vdots \\ \boldsymbol{\varepsilon}_{t-1} E_{Sj} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_1^{\dagger} \\ \mathbf{A}_2^{\dagger} \\ \vdots \\ \mathbf{A}_S^{\dagger} \end{bmatrix} f_j(\mathbf{Y}_{t-2}) + f_j(\boldsymbol{\varepsilon}_{t-1}). \end{aligned} \tag{3.61}$$

Therefore, $f_j(\mathbf{Y}_{t-1})$ can be written as

$$f_j(\mathbf{Y}_{t-1}) = \mathbf{B} f_j(\mathbf{Y}_{t-2}) + f_j(\boldsymbol{\varepsilon}_{t-1}) \tag{3.62}$$

where $\mathbf{B}$ is given in Eq (3.37).

Eventually, by applying Eq (3.61) to the entities of the function $f(\mathbf{Y}_{t-1})$ defined in Eq (3.29), we have

$$
f(\mathbf{Y}_{t-1}) = \begin{bmatrix} f_1(\mathbf{Y}_{t-1}) \\ f_2(\mathbf{Y}_{t-1}) \\ \vdots \\ f_j(\mathbf{Y}_{t-1}) \\ \vdots \\ f_S(\mathbf{Y}_{t-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{B} f_1(\mathbf{Y}_{t-2}) + f_1(\boldsymbol{\varepsilon}_{t-1}) \\ \mathbf{B} f_2(\mathbf{Y}_{t-2}) + f_2(\boldsymbol{\varepsilon}_{t-1}) \\ \vdots \\ \mathbf{B} f_j(\mathbf{Y}_{t-2}) + f_j(\boldsymbol{\varepsilon}_{t-1}) \\ \vdots \\ \mathbf{B} f_S(\mathbf{Y}_{t-2}) + f_S(\boldsymbol{\varepsilon}_{t-1}) \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} \end{bmatrix} \begin{bmatrix} f_1(\mathbf{Y}_{t-2}) \\ f_2(\mathbf{Y}_{t-2}) \\ \vdots \\ f_j(\mathbf{Y}_{t-2}) \\ \vdots \\ f_S(\mathbf{Y}_{t-2}) \end{bmatrix} + \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_{t-1}) \\ f_2(\boldsymbol{\varepsilon}_{t-1}) \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_{t-1}) \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_{t-1}) \end{bmatrix}
$$

$$
= (\mathbf{I}_S \otimes \mathbf{B}) \; f(\mathbf{Y}_{t-2}) + f(\boldsymbol{\varepsilon}_{t-1}).
$$

Therefore, $f(\mathbf{Y}_{t-1})$ can be written as the summation of a recursive function $f(\mathbf{Y}_{t-2})$ and the error function $f(\boldsymbol{\varepsilon}_{t-1})$ as follows

$$
f(\mathbf{Y}_{t-1}) = (\mathbf{I}_S \otimes \mathbf{B}) \; f(\mathbf{Y}_{t-2}) + f(\boldsymbol{\varepsilon}_{t-1}). \tag{3.63}
$$

- **Second Step:** Now, we can substitute $f(\mathbf{Y}_{t-1})$ of Eq (3.63) into Eq (3.52). This leads to

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{A} f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B}) f(\mathbf{Y}_{t-2}) + \mathbf{A} f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t, \qquad t = 0, \pm 1, \pm 2, \ldots
\end{aligned} \tag{3.64}
$$

- **Third Step:** Finally, with this new Eq (3.64) for $\mathbf{Y}_t$ and utilizing the recursive property of $f(\mathbf{Y}_{t-1})$ in Eq (3.63), we can find the moving average representation of MAR(1) as

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{A}f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})f(\mathbf{Y}_{t-2}) + \mathbf{A}f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})[\mathbf{I}_S \otimes (\mathbf{B})f(\mathbf{Y}_{t-3}) + f(\boldsymbol{\varepsilon}_{t-2})] + \mathbf{A}f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})^2 f(\mathbf{Y}_{t-3}) + \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})f(\boldsymbol{\varepsilon}_{t-2}) + \mathbf{A}f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})^3 f(\mathbf{Y}_{t-4}) + \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})^2 f(\boldsymbol{\varepsilon}_{t-3}) + \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})f(\boldsymbol{\varepsilon}_{t-2}) + \mathbf{A}f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&\vdots \\
&= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I}_S \otimes \mathbf{B})^n f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_t.
\end{aligned}
$$

For the sake of brevity and convenience in notation, hereinafter the identity matrix $\mathbf{I}$ without any index will be referred to as the identity matrix $\mathbf{I}_S$, while the dimension of other identity matrices will be specified with an index. Also note that, according to properties of the Kronecker product for given matrices $\mathbf{A}$ and $\mathbf{B}$ with appropriate dimension, we have

$$
(\mathbf{A} \otimes \mathbf{B})^n = \mathbf{A}^n \otimes \mathbf{B}^n. \tag{3.65}
$$

Therefore, we can represent the moving average form of the model MAR(1) with mean zero defined in Eq (3.52) by

$$
\mathbf{Y}_t = \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_t \tag{3.66}
$$

where for $n = 0$, $\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) = \mathbf{A}$.

3.6 The Autocovariance and Autocorrelation Functions of the MAR(1)

The concepts of autocovariance and autocorrelation functions are fundamental for analyzing stationary time series models. These functions provide a measure of the degree of dependencies between the values of a time series at different time steps. They are useful for determining an appropriate ARMA model.

In this section, we introduce autocovariance and autocorrelation functions for the matrix time series $\mathbf{Y}_t$. First, like the process of defining the variance-covariance of matrix white noise $\boldsymbol{\varepsilon}_t$ in section 3.2.2, we will find the expectation of $\mathbf{Y}_t \otimes \mathbf{Y}_t^T$ where the series $\mathbf{Y}_t$ follows a matrix autoregressive model of order one. This expectation will be the main term of the autocovariance function of the MAR(1) model. Then, the definition of the variance-covariance of white noise matrix $\boldsymbol{\varepsilon}_t$ in Eq (3.5) will be used to calculate the variance-covariance of $\mathbf{Y}_t$, i.e., the transformation matrix $\mathbf{T}$ defined in Eq (3.6) will be premultiplied to $E[\mathbf{Y}_t \otimes \mathbf{Y}_t^T]$ to obtain the autocovariance function of the MAR(1) process. The symbol $\boldsymbol{\Psi}(h)$ will be used for the expectation of the Kronecker product $\mathbf{Y}_{t+h}$ to $\mathbf{Y}_t^T$, i.e., $\boldsymbol{\Psi}(h) = E[\mathbf{Y}_{t+h} \otimes \mathbf{Y}_t^T]$, and it will be called the lag function at lag h. We will use the symbol $\boldsymbol{\Gamma}(h)$ for representing the autocovariance function of the MAR models at lag h. Eventually, similarly to Eq (3.5), the relationship between the lag function $\boldsymbol{\Psi}(h)$ and the autocovariance function of MAR models at lag h, $\boldsymbol{\Gamma}(h)$, is given by

$$\boldsymbol{\Gamma}(h) = \mathbf{T}\boldsymbol{\Psi}(h). \tag{3.67}$$

In section 3.6.1 the autocovariance function of the MAR(1) model at lag zero (variance-covariance of MAR(1)) will be derived. Later, in section 3.6.2, we will define an appropriate vector $\mathbf{e}_j$, such that $\mathbf{Y}_t\mathbf{e}_j = Y_{.jt}$, in order to obtain autocovariance and autocorrelation functions for the vector time series $Y_{.jt}$ and cross-autocovariance and cross-autocorrelation functions of vectors $Y_{.jt}, Y_{.j't}, j \neq j'$ where for given (j, j'), the elements of the cross-autocorrelation matrix function are

$$\rho_{ii'}(h) = \frac{\gamma_{ii'}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{i'i'}(0)}} \tag{3.68}$$

where $\gamma_{ii'}(h),\ i, i' = 1, 2, \ldots, K$, are the elements of the block matrix elements of $\mathbf{\Gamma}_{jj}^{T}(h)$ of $\mathbf{\Gamma}(h)$ in Eq (3.67). The column vector $Y_{.jt}$ of the matrix time series $\mathbf{Y}_t$ is called the marginal vector of matrix time series models. Finally, in section 3.6.3, we will calculate the autocovariance and autocorrelation functions at lag $h > 0$ of a matrix time series MAR(1). Moreover, in section 3.6.3, the autocovariance function at lag $h > 0$ for the marginal vectors of a MAR(1) will be derived.

To obtain the autocovariance functions for any time series model, it is much easier to use the moving average representation of an autoregressive process rather than use the autoregressive process directly. Recall from section 3.5 that the moving average representation of a MAR(1) model is given by

$$\mathbf{Y}_t = \sum_{n=0}^{\infty} \mathbf{A}\mathbf{D}^n f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_t \tag{3.69}$$

where $\mathbf{D} = (\mathbf{I} \otimes \mathbf{B})$. Also, recall from Eqs (3.28), (3.37), and (3.29) that $\mathbf{A}$, $\mathbf{D}$, and $f(\boldsymbol{\varepsilon}_t)$ are matrices with dimension $K \times KS^2$, $KS^2 \times KS^2$, and $KS^2 \times S$, respectively. We will use the same idea and definition that we used for defining the variance-covarаince $\boldsymbol{\varepsilon}_t$ in Eq (3.5) for finding (defining) the variance-covariance of a MAR(1) model and later for obtaining the autocovariance functions of a matrix time series $\mathbf{Y}_t$. To this end, first we need to determine the expectation of the Kronecker product $\mathbf{Y}_t \otimes \mathbf{Y}_t^T$. This will introduce the Kronecker product of the term $\mathbf{A}\mathbf{D}^n f(\boldsymbol{\varepsilon}_t)$ of Eq (3.69), i.e.,

$$\big(\mathbf{A}\mathbf{D}^n f(\boldsymbol{\varepsilon}_t)\big) \otimes \big(\mathbf{A}\mathbf{D}^n f(\boldsymbol{\varepsilon}_t)\big)^T. \tag{3.70}$$

The expectation of this quantity needs to be determined. The following proposition is useful for simplifying and calculating these expectation values.

Proposition 3.6.1 *Let* $\mathbf{A}$*,* $\mathbf{B}$ *and* $\mathbf{X}$ *be matrices of size* $p \times q$*,* $q \times n$*, and* $n \times m$*, respectively. Then, we have*

$$(\mathbf{ABX}) \otimes (\mathbf{ABX})^T = (\mathbf{ABX}) \otimes (\mathbf{X}^T\mathbf{B}^T\mathbf{A}^T) = (\mathbf{AB} \otimes \mathbf{I}_m)(\mathbf{X} \otimes \mathbf{X}^T)(\mathbf{I}_m \otimes \mathbf{B}^T\mathbf{A}^T). \quad (3.71)$$

Proof: *By twice applying the Kronecker product rule* $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ *in the right-hand side of Eq* (3.71)*, we obtain the left-hand side term of Eq* (3.71)*.*

3.6.1 Variance-covariance of matrix time series $\mathbf{Y}_t$

We first obtain the expectation of the Kronecker product $\mathbf{Y}_t \otimes \mathbf{Y}_t^T$. Then, this expectation will be used to find the variance-covariance matrix (autocovariance function at lag zero) of the matrix time series $\mathbf{Y}_t$. To this end, first consider the matrices $\mathbf{A}$, $\mathbf{D}^n$, and the matrix function $f(\boldsymbol{\varepsilon}_t)$ in Eq (3.70); these correspond to the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{X}$, respectively, of Proposition (3.6.1), with dimensions $p = K,\ q = n = KS^2$, and $m = S$. Then, by applying Proposition (3.6.1), the Kronecker product $(\mathbf{AD}^n f(\boldsymbol{\varepsilon}_t)) \otimes (\mathbf{AD}^n f(\boldsymbol{\varepsilon}_t))^T$ can be written as

$$\big(\mathbf{AD}^n f(\boldsymbol{\varepsilon}_t)\big) \otimes \big(\mathbf{AD}^n f(\boldsymbol{\varepsilon}_t)\big)^T = [\mathbf{AD}^n \otimes \mathbf{I}_S][f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t)][\mathbf{I}_S \otimes \mathbf{D}^{nT}\mathbf{A}^T]. \quad (3.72)$$

Let us take the autocovariance function at lag zero, i.e., $\boldsymbol{\Psi}(0)$. Recall from Eq (3.4), that we used the symbol $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}$ for the variance-covariance of the stationary matrix process $\boldsymbol{\varepsilon}_t$ of MAR models. Note that according to the white noise definition, the error matrices $\boldsymbol{\varepsilon}_t$'s are uncorrelated. Therefore, the expectation of the Kronecker product $\mathbf{Y}_t \otimes \mathbf{Y}_t^T$, $\boldsymbol{\Psi}(0)$, given in Eq (3.69), can be written as follows

$$
\begin{aligned}
\boldsymbol{\Psi}(0) &= E(\mathbf{Y}_t \otimes \mathbf{Y}_t^T) \\
&= E\bigg(\big[\sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t-n-1}) \big] \otimes \big[\sum_{m=0}^{\infty} f^T(\boldsymbol{\varepsilon}_{t-m-1}) (\mathbf{I} \otimes \mathbf{B}^n)^T \mathbf{A}^T \big] \bigg) + E(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T) \\
&= \sum_{n=0}^{\infty} E\bigg(\big[\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t-n-1}) \big] \otimes \big[f^T(\boldsymbol{\varepsilon}_{t-n-1}) (\mathbf{I} \otimes \mathbf{B}^{nT}) \mathbf{A}^T \big] \bigg) + E(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T). \quad (3.73)
\end{aligned}
$$

Now, by using Eq (3.72) and Eq (3.4), and using the fact that of all the matrices in Eq (3.73), only the matrix function $f(\boldsymbol{\varepsilon}_t)$ is random, $\boldsymbol{\Psi}(0)$ in Eq (3.73) can be rewritten as

$$
\begin{aligned}
\boldsymbol{\Psi}(0) &= \sum_{n=0}^{\infty} E\bigg(\big[\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) \otimes \mathbf{I} \big] \; \big[f(\boldsymbol{\varepsilon}_{t-n-1}) \otimes f^T(\boldsymbol{\varepsilon}_{t-n-1}) \big] \; \big[\mathbf{I} \otimes (\mathbf{I} \otimes \mathbf{B}^{nT}) \mathbf{A}^T \big] \bigg) + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} \\
&= \sum_{n=0}^{\infty} \big[\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) \otimes \mathbf{I} \big] \; E\big[f(\boldsymbol{\varepsilon}_{t-n-1}) \otimes f^T(\boldsymbol{\varepsilon}_{t-n-1}) \big] \; \big[\mathbf{I} \otimes (\mathbf{I} \otimes \mathbf{B}^{nT}) \mathbf{A}^T \big] + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}. \quad (3.74)
\end{aligned}
$$

Therefore, we need to determine the expectation of the Kronecker product of the matrix function $f(\boldsymbol{\varepsilon}_t)$ of its transpose, $E\big[f(\boldsymbol{\varepsilon}_{t-n-1}) \otimes f^T(\boldsymbol{\varepsilon}_{t-n-1})\big]$. To this end, first note that according to Eq (3.29), we have

$$
f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) = \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_t) \\ f_2(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_t) \end{bmatrix} \otimes \begin{bmatrix} f_1^T(\boldsymbol{\varepsilon}_t) & f_2^T(\boldsymbol{\varepsilon}_t) & \cdots & f_j^T(\boldsymbol{\varepsilon}_t) & \cdots & f_S^T(\boldsymbol{\varepsilon}_t) \end{bmatrix}. \quad (3.75)
$$

With respect to the definition of the Kronecker product, this can be written as

$$f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) = \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_t) \\ f_2(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_t) \end{bmatrix} \otimes f^T(\boldsymbol{\varepsilon}_t) = \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \\ f_2(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \end{bmatrix}_{KS^3 \times KS^3} . \tag{3.76}$$

By substituting Eq (3.30) into each j^{th} term $f_j(\boldsymbol{\varepsilon}_t)$ of the right hand side of Eq (3.75), $j = 1, 2, \ldots, S$, and then taking the Kronecker product, we have

$$f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) = \begin{bmatrix} [Vec(\boldsymbol{\varepsilon}_t), \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}] \otimes f^T(\boldsymbol{\varepsilon}_t) \\ [\mathbf{0}, Vec(\boldsymbol{\varepsilon}_t), \mathbf{0}, \ldots, \mathbf{0}] \otimes f^T(\boldsymbol{\varepsilon}_t) \\ \vdots \\ [\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, Vec(\boldsymbol{\varepsilon}_t)] \otimes f^T(\boldsymbol{\varepsilon}_t) \end{bmatrix} = \begin{bmatrix} Vec(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t), \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0} \\ \mathbf{0}, Vec(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t), \mathbf{0}, \ldots, \mathbf{0} \\ \vdots \\ \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, Vec(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \end{bmatrix} . \tag{3.77}$$

Then, by using Eq (3.31) for $f(\boldsymbol{\varepsilon}_t)$, the expectation value of the j^{th} term of the matrix $f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t)$ given in Eq (3.77) can be obtained as follows, for $j = 1, 2, \ldots, S$,

$$\begin{aligned} E[f_j(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t)] &= E\left[\mathbf{0}, \ldots, \mathbf{0}, Vec(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t), \mathbf{0}, \ldots, \mathbf{0}\right] \\ &= E\left[\overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{j-1}, Vec(\boldsymbol{\varepsilon}_t) \otimes \begin{pmatrix} Vec^T(\boldsymbol{\varepsilon}_t), \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0} \\ \mathbf{0}, Vec^T(\boldsymbol{\varepsilon}_t), \mathbf{0}, \ldots, \mathbf{0} \\ \vdots \\ \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, Vec^T(\boldsymbol{\varepsilon}_t) \end{pmatrix}, \overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{S-j}\right]; \end{aligned} \tag{3.78}$$

from Eq (3.12), we have $E[Vec(\boldsymbol{\varepsilon}_t)Vec(\boldsymbol{\varepsilon}_t)^T] = \boldsymbol{\Sigma}^*$, so

$$
\begin{aligned}
E[f_j(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t)] &= \left[\mathbf{0}, \ldots, \mathbf{0}, \mathbf{T}^{*-1} \begin{pmatrix} \boldsymbol{\Sigma}^* & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}^* \end{pmatrix}, \mathbf{0}, \ldots, \mathbf{0} \right] \\
&= \left[\overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{j-1}, \mathbf{T}^{*-1}(\mathbf{I} \otimes \boldsymbol{\Sigma}^*), \overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{S-j} \right] = \mathbf{e}_j^T \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \boldsymbol{\Sigma}^*) \qquad (3.79)
\end{aligned}
$$

where $\boldsymbol{\Sigma}^*$ is the covariance matrix, and the transformation matrix $\mathbf{T}^*$ is defined as was the transformation matrix $\mathbf{T}$ in Eq (3.6), except that now instead of index K the index is KS. That is,

$$
\mathbf{T}^*_{(KS^2 \times KS^2)} = \mathbf{T}_{(KS)S \times (KS)S} = (t_{ij}) = \begin{cases} 1, & \text{if} \begin{cases} i = l(KS)+1, l(KS)+2, \ldots, (l+1)(KS), \\ j = (l+1) + (i - l(KS) - 1)S, \\ l = 0, 1, 2, \ldots, (S-1); \end{cases} \\ 0, & \text{otherwise.} \end{cases} \qquad (3.80)
$$

Analogously to the equivalent formula for the transformation matrix $\mathbf{T}$ defined in Eq (3.7), the transformation matrix $\mathbf{T}^*$ given in Eq (3.80) can be written as the following equivalent form

$$
\mathbf{T}^*_{(KS^2 \times KS^2)} = \sum_{i=1}^{KS} \sum_{j=1}^{S} \mathbf{E}_{ji}^{S \times KS} \otimes \mathbf{E}_{ij}^{KS \times S}. \qquad (3.81)
$$

Therefore, taking expectations on both sides of Eq (3.77), and using the results of Eq (3.78) for all $j = 1, 2, \ldots, S$, we obtain

$$E[f(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)] = \begin{bmatrix} E[Vec(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)], \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0} \\ \mathbf{0}, E[Vec(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)], \mathbf{0}, \ldots, \mathbf{0} \\ \vdots \quad \ddots \quad \vdots \\ \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, E[Vec(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)] \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*), \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0} \\ \mathbf{0}, \mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*), \mathbf{0}, \ldots, \mathbf{0} \\ \vdots \quad \ddots \quad \vdots \\ \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*) \end{bmatrix}. \tag{3.82}$$

Hence, the expectation value of the Kronecker product $f(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)$ is given by

$$E[f(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)] = \mathbf{I}\otimes\mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*). \tag{3.83}$$

Finally, by substituting Eq (3.83) into Eq (3.74), $\boldsymbol{\Psi}(0)$ can be obtained as

$$\begin{aligned}
\boldsymbol{\Psi}(0) &= \sum_{n=0}^{\infty}\big[\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^n)\otimes\mathbf{I}\big]\; E\big[f(\boldsymbol{\varepsilon}_{t-n-1})\otimes f^T(\boldsymbol{\varepsilon}_{t-n-1})\big]\;\big[\mathbf{I}\otimes(\mathbf{I}\otimes\mathbf{B}^{nT})\mathbf{A}^T\big] + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} \\
&= \sum_{n=0}^{\infty}\big[\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^n)\otimes\mathbf{I}\big]\;\big[\mathbf{I}\otimes\mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*)\big]\;\big[\mathbf{I}\otimes(\mathbf{I}\otimes\mathbf{B}^{nT})\mathbf{A}^T\big] + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} \\
&= \sum_{n=0}^{\infty}\big[\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^n)\otimes\mathbf{I}\big]\;\big[\mathbf{I}\otimes\mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*)(\mathbf{I}\otimes\mathbf{B}^{nT})\mathbf{A}^T\big] + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} \\
&= \sum_{n=0}^{\infty}\big[\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^n)\otimes\mathbf{I}\big]\;\big[\mathbf{I}\otimes\mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^T\big] + \boldsymbol{\Psi} \\
&= \sum_{n=0}^{\infty}\big[\mathbf{A}\mathbf{D}^n\otimes\mathbf{I}\big]\;\big[\mathbf{I}\otimes\mathbf{T}^{*-1}(\mathbf{I}\otimes\boldsymbol{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^T\big] + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}.
\end{aligned}$$

Note that the two last equalities are obtained by using the Kronecker product rule $(A\otimes B)(C\otimes D) = AC\otimes BD$. Therefore, the lag function at lag zero for the MAR(1) model is

$$\mathbf{\Psi}(0) = E(\mathbf{Y}_t \otimes \mathbf{Y}_t^T) = \sum_{n=0}^{\infty} \left[\mathbf{A}\mathbf{D}^n \otimes \mathbf{I}\right] \left[\mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \mathbf{\Sigma}^* \mathbf{B}^{nT})\mathbf{A}^T\right] + \mathbf{\Psi}_{\boldsymbol{\varepsilon}}. \tag{3.84}$$

Let $\mathbf{\Gamma}(0)$ be the variance-covariance of the matrix time series $\mathbf{Y}_t$ (or equivalently based on Eq (3.67), $\mathbf{\Gamma}(0)$ is the autocovariance function at lag zero). Then, according to the definition of the variance-covariance of a matrix variable given in Eq (3.5) (similarly to Eq (3.67) for autocovariance function at lag zero), we have

$$\mathbf{\Gamma}(0) = Var(\mathbf{Y}_t) = \mathbf{T}E(\mathbf{Y}_t \otimes \mathbf{Y}_t^T) = \mathbf{T}\mathbf{\Psi}(0) \tag{3.85}$$

where $\mathbf{T}$ and $\mathbf{\Psi}(0)$ are given in Eq (3.6) and Eq (3.84), respectively.

3.6.2 Marginal Vector $Y_{.jt}$ of $\mathbf{Y}_t$

Now, we can study the properties of each vector time series $Y_{.jt}, j = 1, 2, \ldots, S$, of the matrix time series $\mathbf{Y}_t$ separately. Furthermore, this study will help us to find the block entries, $\mathbf{\Gamma}_{jj'}(0), j, j' = 1, 2, \ldots, S$, of the variance-covariance matrix $\mathbf{\Gamma}(0)$. First, note that the time series vector $Y_{.jt}$ can be obtained from the matrix time series $\mathbf{Y}_t$ by multiplying $\mathbf{Y}_t$ by an appropriate vector $\mathbf{e}_j$. That is,

$$Y_{.jt} = \mathbf{Y}_t \mathbf{e}_j, \qquad j = 1, 2, \ldots, S, \tag{3.86}$$

where $\mathbf{e}_j = (\overbrace{0, 0, \ldots, 0}^{j-1}, 1, \overbrace{0, \ldots, 0}^{S-j})^T$. Similarly to Eq (3.86), we can define other terms of MAR(1) model of Eq (3.69) as

$$\boldsymbol{\varepsilon}_{.jt} = \boldsymbol{\varepsilon}_t \mathbf{e}_j, \quad V_j(\boldsymbol{\varepsilon}_t) = f(\boldsymbol{\varepsilon}_t)\mathbf{e}_j, \qquad j = 1, 2, \ldots, S, \tag{3.87}$$

where V_j is an $S \times 1$ block matrix function, such that the j^{th} block is the vector $Vec(\boldsymbol{\varepsilon}_t)$ with dimension $KS \times 1$ and the other blocks are zero. That is,

$$V_j(\boldsymbol{\varepsilon}_t) = \left[\overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{j-1}, Vec^T(\boldsymbol{\varepsilon}_t), \overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{S-j} \right]^T. \tag{3.88}$$

Note, $V_j(\boldsymbol{\varepsilon}_t)$ is the j^{th} column of the matrix function $f(\boldsymbol{\varepsilon}_t)$ defined in Eq (3.31), and therefore has dimension $KS^2 \times 1$.

Therefore, according to the moving average representation of $\mathbf{Y}_t$ in Eq (3.69), the moving average representation of the vector time series $Y_{.jt}$ is given by

$$\begin{aligned} Y_{.jt} &= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t-n-1})\mathbf{e}_j + \boldsymbol{\varepsilon}_t \mathbf{e}_j \\ &= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) V_j(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_{.jt}. \end{aligned} \tag{3.89}$$

Now, the moving average representation of the vector time series $Y_{.jt}$ in Eq (3.89) can be used to determine the block entries, $\boldsymbol{\Gamma}_{jj'}(0), j, j' = 1, 2, \ldots, S$, of the variance-covariance matrix $\boldsymbol{\Gamma}(0)$. The elements $\boldsymbol{\Gamma}_{jj'}(0), j, j' = 1, 2, \ldots, S$, are the variance-covariance matrices of $Y_{jj'}(0), j, j' = 1, 2, \ldots, S$; that is,

$$\boldsymbol{\Gamma}^T_{jj'}(0) = Cov(Y_{.jt}, Y_{.j't}), \qquad j, j' = 1, 2, \ldots, S. \tag{3.90}$$

Note that there is the same analogous pattern between $\boldsymbol{\Gamma}^*(0)$ and $\boldsymbol{\Gamma}(0)$ as exists between $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Sigma}$ in Eqs (3.13)-(3.15). That is, we have

$$\boldsymbol{\Gamma}^*(0) = Var(Vec(\mathbf{Y}_t)) = E[Vec(\mathbf{Y}_t)Vec(\mathbf{Y}_t)^T] \tag{3.91}$$

$$\boldsymbol{\Gamma}^T_{jj'}(0) = \boldsymbol{\Gamma}^*_{jj'}(0), \qquad j, j' = 1, 2, \ldots, S. \tag{3.92}$$

From Eq (3.89) for $Y_{.jt}$, and using the fact that the matrix white noise $\boldsymbol{\varepsilon}_t$'s and therefore the vector function $V_j(\boldsymbol{\varepsilon}_t)$'s are uncorrelated random variables (white noise), the variance-covariance matrix $\boldsymbol{\Gamma}_{jj'}(0)$ can be derived as follows. First, we write

$$\begin{aligned}
\boldsymbol{\Gamma}^T_{jj'}(0) &= Cov(Y_{.jt}, Y_{.j't}) = E(Y_{.jt}Y^T_{.j't}) \\
&= E\left([\sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) V_j(\boldsymbol{\varepsilon}_{t-n-1})][\sum_{m=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{mT}) V_{j'}(\boldsymbol{\varepsilon}_{t-m-1})]^T \right) + E[\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.j't}] \\
&= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n)\ E[V_j(\boldsymbol{\varepsilon}_{t-n-1}) {V_{j'}}^T(\boldsymbol{\varepsilon}_{t-n-1})](\mathbf{I} \otimes \mathbf{B}^{nT})\mathbf{A}^T + E[\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.j't}]. \qquad (3.93)
\end{aligned}$$

To derive the expectation $E[V_j(\boldsymbol{\varepsilon}_t){V_{j'}}^T(\boldsymbol{\varepsilon}_t)]$, first note that based on the definition of the vector function $V_j(.)$ in Eq (3.88), we have

$$V_j(\boldsymbol{\varepsilon}_t){V_{j'}}^T(\boldsymbol{\varepsilon}_t) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & Vec(\boldsymbol{\varepsilon}_t)Vec^T(\boldsymbol{\varepsilon}_t) & \ddots & \mathbf{0} \\ \vdots & \ddots & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} = \mathbf{E}_{jj'} \otimes Vec(\boldsymbol{\varepsilon}_t)Vec^T(\boldsymbol{\varepsilon}_t). \qquad (3.94)$$

Now, by taking the expectation on both sides of Eq (3.94) and using the result of Eq (3.12), we have

$$E[V_j(\boldsymbol{\varepsilon}_t){V_{j'}}^T(\boldsymbol{\varepsilon}_t)] = \mathbf{E}_{jj'} \otimes \boldsymbol{\Sigma}^*. \qquad (3.95)$$

Moreover, from Eqs (3.12)-(3.15), we know that $E[\boldsymbol{\varepsilon}_{.jt}\boldsymbol{\varepsilon}^T_{.j't}] = \boldsymbol{\Sigma}^*_{jj'}$. Therefore, by substituting this and Eq (3.95) into Eq (3.93), $\boldsymbol{\Gamma}^T_{jj'}(0)$ becomes

$$\boldsymbol{\Gamma}^T_{jj'}(0) = \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n)\ (\mathbf{E}_{jj'} \otimes \boldsymbol{\Sigma}^*)(\mathbf{I} \otimes \mathbf{B}^{nT})\mathbf{A}^T + \boldsymbol{\Sigma}^*_{jj'};$$

using the Kronecker product rule $(A \otimes B)(C \otimes D) = AC \otimes BD$ gives us

$$\begin{aligned}\mathbf{\Gamma}^T_{jj'}(0) &= \sum_{n=0}^{\infty}\mathbf{A}(\mathbf{E}_{jj'}\otimes\mathbf{B}^n\mathbf{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^T+\mathbf{\Sigma}^*_{jj'}\\ &= \sum_{n=0}^{\infty}\mathbf{A}^{\dagger}_j(\mathbf{B}^n\mathbf{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^{\dagger}_{j'}{}^T+\mathbf{\Sigma}^*_{jj'}, \qquad j,j'=1,2,\cdots,S,\end{aligned}$$

where $\mathbf{A}^{\dagger}_j$ and its relation with the coefficient matrix $\mathbf{A}$ are as given in Eq (3.32) and Eq (3.33), and $\mathbf{B}$ is defined as in Eq (3.37). Therefore, the variance-covariance matrix of the j^{th} marginal vector $Y_{.jt}$ of the matrix time series $\mathbf{Y}_t$ defined in Eq (3.88) is equal to

$$\mathbf{\Gamma}^T_{jj'}(0)=\sum_{n=0}^{\infty}\mathbf{A}^{\dagger}_j(\mathbf{B}^n\mathbf{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^{\dagger}_{j'}{}^T+\mathbf{\Sigma}^*_{jj'}, \qquad j,j'=1,2,\cdots,S. \tag{3.96}$$

Now, let us use the variance-covariance matrix of the vector time series $Y_{.jt}$ obtained in Eq (3.96), $\mathbf{\Gamma}_{jj'}(0)$, to simplify and rewrite the variance-covariance matrix of the matrix time series $\mathbf{Y}_t$, MAR(1), $\mathbf{\Gamma}(0)$, defined in Eq (3.85), with block entries. That is,

$$\mathbf{\Gamma}(0)=(\mathbf{\Gamma}_{jj'}(0))=Var(\mathbf{Y}_t)=\mathbf{T}E(\mathbf{Y}_t\otimes\mathbf{Y}'_t)=\begin{bmatrix}\mathbf{\Gamma}_{11}(0) & \mathbf{\Gamma}_{12}(0) & \ldots & \mathbf{\Gamma}_{1S}(0)\\ \mathbf{\Gamma}_{21}(0) & \mathbf{\Gamma}_{22}(0) & \ldots & \mathbf{\Gamma}_{2S}(0)\\ \vdots & & \ddots & \vdots\\ \mathbf{\Gamma}_{S1}(0) & \mathbf{\Gamma}_{S2}(0) & \ldots & \mathbf{\Gamma}_{SS}(0)\end{bmatrix}. \tag{3.97}$$

3.6.3 Autocovariance function at lag $h > 0$

In section 3.6.1, we found the variance-covariance matrix of the matrix time series $\mathbf{Y}_t$, $\mathbf{\Gamma}(0)$, which is equivalent to the autocovariance matrix function at lag zero. In this section, the autocovariance function of the matrix time series $\mathbf{Y}_t$ at lag $h > 0$, $\mathbf{\Gamma}(h)$, will be derived.

Like the derivation of the autocovariance function at lag zero, first the lag function $\mathbf{\Psi}(h)$ will be obtained, and then the autocovariance function $\mathbf{\Gamma}(h)$ can be found according to the Eq (3.67), $\mathbf{\Gamma}(h) = \mathbf{T}\mathbf{\Psi}(h)$, similar to the derivation of the variance-covariance matrix $\mathbf{\Gamma}(0)$ given in Eq (3.85).

First, note that the moving average representation of $\mathbf{Y}_t$ in Eq (3.66) for $\mathbf{Y}_{t+h}$ can be broken down and rewritten as

$$\begin{aligned}\mathbf{Y}_{t+h} &= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t+h-n-1}) + \boldsymbol{\varepsilon}_{t+h} \\ &= \sum_{n=0}^{h-1} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h}) f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_{t+h}. \qquad (3.98)\end{aligned}$$

Then, to derive the lag function $\mathbf{\Psi}(h)$ of $\mathbf{Y}_t$ in Eq (3.98), we take

$$\begin{aligned}\mathbf{\Psi}(h) &= E(\mathbf{Y}_{t+h} \otimes \mathbf{Y}_t^T) \\ &= E\Big(\big[\sum_{n=0}^{h-1} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n) f(\boldsymbol{\varepsilon}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h}) f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_{t+h} \big] \\ &\qquad \otimes \big[\sum_{m=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^m) f(\boldsymbol{\varepsilon}_{t-m-1}) + \boldsymbol{\varepsilon}_t \big]^T \Big). \qquad (3.99)\end{aligned}$$

Since the matrix white noise elements $\boldsymbol{\varepsilon}_t$ are uncorrelated (i.e., $E[\boldsymbol{\varepsilon}_{t+h} \otimes \boldsymbol{\varepsilon}_t^T] = 0$, for $h \neq 0$, hence $E[f(\boldsymbol{\varepsilon}_{t+h}) \otimes f^T(\boldsymbol{\varepsilon}_t)] = 0$; likewise, $E[f(\boldsymbol{\varepsilon}_{t+h}) \otimes \boldsymbol{\varepsilon}_t^T] = 0$), Eq (3.99) can be simplified as

$$\begin{aligned}\mathbf{\Psi}(h) = \sum_{n=0}^{\infty} E([\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h}) f(\boldsymbol{\varepsilon}_{t-n-1})] \otimes [f^T(\boldsymbol{\varepsilon}_{t-n-1})(\mathbf{I} \otimes \mathbf{B}^{n^T})\mathbf{A}^T]) \qquad (3.100) \\ + \; E([\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{h-1}) f(\boldsymbol{\varepsilon}_t)] \otimes [\boldsymbol{\varepsilon}_t^T]).\end{aligned}$$

Now, by applying Proposition (3.6.1) to both terms $(\mathbf{A}\mathbf{D}^{n+h} f(\boldsymbol{\varepsilon}_{t-n-1})) \otimes (f^T(\boldsymbol{\varepsilon}_{t-n-1})\mathbf{D}^{n^T}\mathbf{A}^T)$ and $(\mathbf{A}\mathbf{D}^{h-1} f(\boldsymbol{\varepsilon}_t)) \otimes \boldsymbol{\varepsilon}_t^T$ (where $\mathbf{D}^{n+h} = \mathbf{I} \otimes \mathbf{B}^{n+h}$) in Eq (3.100), and by using the fact that except for $\boldsymbol{\varepsilon}_t$ and $f(\boldsymbol{\varepsilon}_t)$, all matrices in Eq (3.100) are not random, $\mathbf{\Psi}(h)$ can be rewritten as

$$\begin{aligned}\boldsymbol{\Psi}(h) &= \sum_{n=0}^{\infty} E([\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^{n+h})\otimes\mathbf{I}]\ [f(\boldsymbol{\varepsilon}_{t-n-1})\otimes f^T(\boldsymbol{\varepsilon}_{t-n-1})]\ [\mathbf{I}\otimes(\mathbf{I}\otimes\mathbf{B}^{nT})\mathbf{A}^T]) \\ &\quad + E([\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^{h-1})\otimes\mathbf{I}][f(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T]) \\ &= \sum_{n=0}^{\infty}[\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^{n+h})\otimes\mathbf{I}]\ E[f(\boldsymbol{\varepsilon}_{t-n-1})\otimes f^T(\boldsymbol{\varepsilon}_{t-n-1})]\ [\mathbf{I}\otimes(\mathbf{I}\otimes\mathbf{B}^{nT})\mathbf{A}^T] \\ &\quad + [\mathbf{A}(\mathbf{I}\otimes\mathbf{B}^{h-1})\otimes\mathbf{I}]\ E[f(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T]. \end{aligned} \tag{3.101}$$

We already know $E[f(\boldsymbol{\varepsilon}_t)\otimes f^T(\boldsymbol{\varepsilon}_t)]$ from Eq (3.83); and we need to derive the $E[f(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T]$.

Toward this end, first note that according to Eq (3.29), we have

$$f(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T = \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_t) \\ f_2(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_t) \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_t) \end{bmatrix} \otimes\boldsymbol{\varepsilon}_t^T = \begin{bmatrix} f_1(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T \\ f_2(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T \\ \vdots \\ f_j(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T \\ \vdots \\ f_S(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T \end{bmatrix}. \tag{3.102}$$

Now, substituting Eq (3.30) (for $\boldsymbol{\varepsilon}_t$ instead of $\mathbf{Y}_{t-1}$) into the j^{th} element, $j = 1, 2, \ldots, S$, of Eq (3.102), we have

$$f(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T = \begin{bmatrix} [Vec(\boldsymbol{\varepsilon}_t), 0, 0, \ldots, 0]\otimes\boldsymbol{\varepsilon}_t^T \\ [0, Vec(\boldsymbol{\varepsilon}_t), 0, \ldots, 0]\otimes\boldsymbol{\varepsilon}_t^T \\ \vdots \\ [0, 0, \ldots, 0, Vec(\boldsymbol{\varepsilon}_t)]\otimes\boldsymbol{\varepsilon}_t^T \end{bmatrix} = \begin{bmatrix} Vec(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T, 0, 0, \ldots, 0 \\ 0, Vec(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T, 0, \ldots, 0 \\ \vdots \\ 0, 0, \ldots, 0, Vec(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T \end{bmatrix}. \tag{3.103}$$

Recall from Eq (3.4) that $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} = E[\boldsymbol{\varepsilon}_t\otimes\boldsymbol{\varepsilon}_t^T]$; hence, it is easy to show that

$$E[Vec(\boldsymbol{\varepsilon}_t)\otimes\boldsymbol{\varepsilon}_t^T] = Vecb^K(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}) \tag{3.104}$$

where $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}$ is defined in Eq (3.4), and $Vecb^K(.)$ is defined in Eq (3.38) (for K instead of c). That is,

$$Vecb^K(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}) = \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}1} \\ \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}2} \\ \vdots \\ \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}S} \end{pmatrix}_{KS^2 \times K} . \tag{3.105}$$

We prove Eq (3.104) as follows. First, partition $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon},KS\times KS}$ as $(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}1}, \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}2}, \ldots, \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}S})$ such that each $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}j}, j = 1, 2, \ldots, S$, is a matrix with dimension $KS \times K$; likewise, partition $\boldsymbol{\varepsilon}_t$ as $(\boldsymbol{\varepsilon}_{.1t}, \boldsymbol{\varepsilon}_{.2t}, \ldots, \boldsymbol{\varepsilon}_{.St})$ such that each $\boldsymbol{\varepsilon}_{.jt}$, $j = 1, 2, \ldots, S$, is a K-dimensional vector. Then, from Eq (3.4), it is easy to show that $E[\boldsymbol{\varepsilon}_{.jt} \otimes \boldsymbol{\varepsilon}_t^T] = \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}j}$. Finally, the proof follows from the definitions of $Vec(.)$ and $Vecb^K(.)$ operators.

Eventually, by taking the expectation on both sides of Eq (3.103), and substituting from Eq (3.104), we have

$$E[f(\boldsymbol{\varepsilon}_t) \otimes \boldsymbol{\varepsilon}_t^T] = \mathbf{I} \otimes Vecb^K(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}). \tag{3.106}$$

Therefore, by substituting Eq (3.106) and Eq (3.83) into Eq (3.101), the lag function $\boldsymbol{\Psi}(h)$ becomes

$$\begin{aligned} \boldsymbol{\Psi}(h) &= \sum_{n=0}^{\infty} [\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h}) \otimes \mathbf{I}] \left[\mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \boldsymbol{\Sigma}^*)\right] \left[\mathbf{I} \otimes (\mathbf{I} \otimes \mathbf{B}^{nT})\mathbf{A}^T\right] \\ &\qquad + [\mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{h-1}) \otimes \mathbf{I}][\mathbf{I} \otimes Vecb^K(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}})] \\ &= \sum_{n=0}^{\infty} \left[\mathbf{A}\mathbf{D}^{n+h} \otimes \mathbf{I})\right] \left[\mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \boldsymbol{\Sigma}^* \mathbf{B}^{nT})\mathbf{A}^T\right] + [\mathbf{A}\mathbf{D}^{h-1} \otimes \mathbf{I}][\mathbf{I} \otimes Vecb^K(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}})]. \end{aligned} \tag{3.107}$$

Finally, the autocovariance function at lag $h > 0$ for the matrix time series $\mathbf{Y}_t$ as defined in Eq (3.67) is obtained by multiplying the transformation matrix $\mathbf{T}$ into the lag function $\boldsymbol{\Psi}(h)$ as follows

$$\mathbf{\Gamma}(h) = (\mathbf{\Gamma}_{jj'}(h)) = Cov(\mathbf{Y}_{t+h}, \mathbf{Y}_t^T) = \mathbf{T\Psi}(h). \tag{3.108}$$

Similarly to the derivation of the autocovariance function, $\mathbf{\Gamma}(0)$, at lag $h = 0$ for the MAR(1) model in section 3.6.2, finding the block entries $\mathbf{\Gamma}_{jj'}(h)$ of the autocovariance function at lag $h > 0$ will be helpful in simplifying the autocovariance function $\mathbf{\Gamma}(h)$ of the MAR(1) model. In particular, it will be necessary to study the features of a specific marginal vector time series $Y_{.jt}$ of the matrix time series $\mathbf{Y}_t$. To this end, first note that according to the moving average representation of $Y_{.jt}$ in Eq (3.89), and with a similar process to derive $\mathbf{\Psi}(h)$ in Eq (3.99) and in Eq (3.100), $\mathbf{\Gamma}^T_{jj'}(h)$ can be written as

$$\begin{aligned}
\mathbf{\Gamma}^T_{jj'}(h) &= Cov(Y_{.jt+h}, Y_{.j't}) = E(Y_{.jt+h}Y^T_{.j't}) \\
&= E\Bigg(\big[\sum_{n=0}^{h-1} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^n)V_j(\boldsymbol{\varepsilon}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h})V_{j'}(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_{.jt+h}\big] \\
&\qquad \big[\sum_{m=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^m)V_{j'}(\boldsymbol{\varepsilon}_{t-m-1}) + \boldsymbol{\varepsilon}_{.j't}\big]^T \Bigg) \\
&= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h}) \ E[V_j(\boldsymbol{\varepsilon}_{t-n-1}){V_{j'}}^T(\boldsymbol{\varepsilon}_{t-n-1})](\mathbf{I} \otimes \mathbf{B}^{nT})\mathbf{A}^T \\
&\qquad + \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{h-1})E[V_j(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.j't}].
\end{aligned} \tag{3.109}$$

We know $E[V_j(\boldsymbol{\varepsilon}_t){V_{j'}}^T(\boldsymbol{\varepsilon}_t)]$ from Eq (3.95). Furthermore, we need to determine $E[V_j(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.j't}]$. Toward this end, first use the definition of the vector function $V_j(\boldsymbol{\varepsilon}_t)$ in Eq (3.88), and with respect to the fact that $\mathbf{\Sigma}^* = E[Vec(\boldsymbol{\varepsilon}_t)Vec(\boldsymbol{\varepsilon}_t)^T]$ given in Eq (3.12), it is easy to show that

$$E\big[Vec(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.jt}\big] = \mathbf{\Sigma}^*_j, \qquad j = 1, 2, \ldots, S, \tag{3.110}$$

where $\mathbf{\Sigma}^*_j$, with dimension $KS \times K$, is the j^{th} block column of the block covariance matrix $\mathbf{\Sigma}^*$. Then, from the definition of $V_j(\boldsymbol{\varepsilon}_t)$ in Eq (3.88), we have $V_j(\boldsymbol{\varepsilon}_t) = \mathbf{e}_j \otimes Vec(\boldsymbol{\varepsilon}_t)$; hence, we have

$$E[V_j(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.j't}] = \mathbf{e}_j \otimes \boldsymbol{\Sigma}^*_{j'}, \qquad j, j' = 1, 2, \ldots, S. \tag{3.111}$$

Therefore, by substituting Eqs (3.95) and (3.111) into Eq (3.109), and using the Kronecker product rule, we have

$$\begin{aligned}
\boldsymbol{\Gamma}^T_{jj'}(h) &= \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{n+h})\ (\mathbf{E}_{jj'} \otimes \boldsymbol{\Sigma}^*)(\mathbf{I} \otimes \mathbf{B}^{nT})\mathbf{A}^T + \mathbf{A}(\mathbf{I} \otimes \mathbf{B}^{h-1})(\mathbf{e}_j \otimes \boldsymbol{\Sigma}^*_{j'}) \\
&= \sum_{n=0}^{\infty} \mathbf{A}(E_{jj'} \otimes \mathbf{B}^{n+h}\boldsymbol{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^T + \mathbf{A}(\mathbf{e}_j \otimes \mathbf{B}^{h-1}\boldsymbol{\Sigma}^*_{j'}) \\
&= \sum_{n=0}^{\infty} \mathbf{A}^{\dagger}_j(\mathbf{B}^{n+h}\boldsymbol{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^{\dagger\,T}_{j'} + \mathbf{A}^{\dagger}_j\mathbf{B}^{h-1}\boldsymbol{\Sigma}^*_{j'}, \qquad j, j' = 1, 2, \cdots, S.
\end{aligned}$$

Therefore, for $h > 0$, we have

$$\boldsymbol{\Gamma}^T_{jj'}(h) = \sum_{n=0}^{\infty} \mathbf{A}^{\dagger}_j(\mathbf{B}^{n+h}\boldsymbol{\Sigma}^*\mathbf{B}^{nT})\mathbf{A}^{\dagger\,T}_{j'} + \mathbf{A}^{\dagger}_j\mathbf{B}^{h-1}\boldsymbol{\Sigma}^*_{j'}, \qquad j, j' = 1, 2, \cdots, S, \tag{3.112}$$

where $\mathbf{A}^{\dagger}_j$, $\mathbf{B}$, and $\boldsymbol{\Sigma}^*_j$ are as given in Eqs (3.32), (3.12), and (3.110), respectively.

3.7 Matrix Autoregressive Process of order p ($MAR(p)$)

3.7.1 The model

So far, a matrix autoregressive process of order one ($MAR(1)$) has been introduced, and its autocovariance and autocorrelation functions have been found by applying the moving average representation of the $MAR(1)$. As we have seen, the moving average representation has some advantages that help us to find the autocovariance and autocorrelation functions of an autoregressive process easily. In this section, the matrix autoregressive process of order $p > 1$ ($MAR(p)$) will be introduced. After introducing the $MAR(p)$ model, a reparametrization technique will be used to rewrite the $MAR(p)$ as a $MAR(1)$ model. Then, all results of the $MAR(1)$ model derived in the previous sections can be generalized to properties of the $MAR(p)$ model.

Let $\mathbf{Y}_t$ be a matrix time series given in Eq (3.2). Then, the matrix autoregressive time series $\mathbf{Y}_t$ of order p is given by

$$\mathbf{Y}_t = \mathbf{A}_1 f(\mathbf{Y}_{t-1}) + \mathbf{A}_2 f(\mathbf{Y}_{t-2}) + \ldots + \mathbf{A}_p f(\mathbf{Y}_{t-p}) + \boldsymbol{\varepsilon}_t \tag{3.113}$$

where $\mathbf{A}_\nu, \nu = 1, 2, \ldots, p$, are the coefficient matrices with dimension $K \times KS^2$ of different lags (ν) that correspond to the coefficient matrix $\mathbf{A}$ given in Eq (3.28) for the MAR(1) process of Eq (3.27) but now with p lags, and $\boldsymbol{\varepsilon}_t$ is the $K \times S$ matrix error term process. Furthermore, the matrix functions $f(\mathbf{Y}_{t-\nu})$, $\nu = 1, 2, \ldots, p$, are the same as for the matrix function $f(\mathbf{Y}_t)$ in the first order matrix autoregressive process of Eq (3.29) according to

$$f(\mathbf{Y}_{t-\nu}) = \begin{bmatrix} f_1(\mathbf{Y}_{t-\nu}) \\ f_2(\mathbf{Y}_{t-\nu}) \\ \vdots \\ f_j(\mathbf{Y}_{t-\nu}) \\ \vdots \\ f_S(\mathbf{Y}_{t-\nu}) \end{bmatrix}, \quad f_j(\mathbf{Y}_{t-\nu}) = \begin{bmatrix} \mathbf{Y}_{t-\nu} E_{1j} \\ \mathbf{Y}_{t-\nu} E_{2j} \\ \vdots \\ \mathbf{Y}_{t-\nu} E_{Sj} \end{bmatrix}, \qquad \begin{array}{r} r, j = 1, 2, \ldots, S, \\ \nu = 1, 2, \ldots, p; \end{array}$$

likewise, $\mathbf{A}_\nu$ with elements $(\mathbf{A}_r^j)_\nu$, $r, j = 1, 2, \ldots, S$, $\nu = 1, 2, \ldots, p$, is defined analogously to $\mathbf{A}_r^j$ of Eq (3.21). In order to be able to represent and to rewrite the matrix autoregressive model of order $p > 1$ into a matrix autoregressive model of order one, we need to reparameterize the terms of the model given in Eq (3.113). To do this reparametrization, let us define new variables and parameters as follows

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p+1} \end{bmatrix}, \qquad C(\mathbf{Y}_{t-1}) = \begin{bmatrix} f(\mathbf{Y}_{t-1}) \\ f(\mathbf{Y}_{t-2}) \\ \vdots \\ f(\mathbf{Y}_{t-p}) \end{bmatrix}, \qquad \mathbf{u}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tag{3.114}$$

$$\mathbf{A}_{p\times p} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{J} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{J} & \mathbf{0} \end{bmatrix}. \tag{3.115}$$

Remember that both $\mathbf{Y}_t$'s and $\boldsymbol{\varepsilon}_t$'s are matrices with dimension $K \times S$. Therefore, the dimension of both $\mathbf{X}_t$ and $\mathbf{u}_t$ is $Kp \times S$. Also, $f(\mathbf{Y}_t)$ is a matrix with dimension $KS^2 \times S$. Hence, $C(\mathbf{Y}_{t-1})$ is a matrix with dimension $KS^2p \times S$. Furthermore, $\mathbf{A}_{p\times p}$ is a block matrix that consists of p^2 matrices each of dimension $K \times KS^2$. The coefficient matrices $\mathbf{A}_\nu, \nu = 1, 2, \dots, p$, in the first row of the block matrix $\mathbf{A}_{p\times p}$ are the same as the coefficient matrices of the $MAR(p)$ model given in Eq (3.113). We define a matrix $\mathbf{J}$, like $\mathbf{A}_\nu$'s, as a $K \times KS^2$ matrix that has S identity matrices $\mathbf{I}_K$ and $S(S-1)$ matrices zero with dimension $K \times K$ in the following order

$$\mathbf{J}_{K\times KS^2} = [\mathbf{I}_K, \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{S}, \mathbf{I}_K, \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{S}, \mathbf{I}_K, \mathbf{0}, \dots, \mathbf{I}_K]. \tag{3.116}$$

This $\mathbf{J}$ matrix has matrix elements $\mathbf{J}_{1j}$, $j = 1, 2, \ldots, S^2$, which can be written as

$$\mathbf{J}_{1j} = \begin{cases} \mathbf{J}_{1j} = \mathbf{I}_K, & \text{if} \begin{cases} j = rS + r + 1, \\ r = 0, 1, \ldots, (S-1); \end{cases} \\ \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{3.117}$$

Note, by defining the block matrix $\mathbf{J}$ in this way, we have $\mathbf{J}f(\mathbf{Y}_{t-\nu}) = \mathbf{Y}_{t-\nu}$. Now, with respect to the new variables and parameters in Eq (3.114) and Eq (3.115), the matrix autoregressive model of order p in Eq (3.113) can be rewritten as a matrix autoregressive model of order one that is given by

$$\mathbf{X}_t = \mathbf{A}C(\mathbf{Y}_{t-1}) + \mathbf{u}_t. \tag{3.118}$$

Thus, all results of the $MAR(1)$ model can be generalized to the $MAR(p)$ model by using the reparameterized model given in Eq (3.118). Therefore, to do this, we first obtain the moving average representation in section 3.7.2, and the autocovariance and autocorrelation functions of the $MAR(p)$ are derived in section 3.7.3.

3.7.2 Moving average representation of $MAR(p)$

In this section, a moving average representation of the matrix autoregressive process of order p defined in Eq (3.113) will be derived. First, because any $MAR(p)$ process can be rewritten in the $MAR(1)$ form of Eq (3.118), and secondly, by using the $K \times Kp$ matrix $\mathbf{J}^*$

$$\mathbf{J}^* = [\mathbf{I}_K, \overbrace{\mathbf{0}, \ldots, \mathbf{0}}^{p-1}], \tag{3.119}$$

the $MAR(p)$ process $\mathbf{Y}_t$ defined in Eq (3.113) can be found by $\mathbf{Y}_t = \mathbf{J}^*\mathbf{X}_t$, where $\mathbf{X}_t$ is the new response variable defined in Eq (3.114).

However, first we need to obtain the moving average representation of the stochastic process $\mathbf{X}_t$. Then, in the next section 3.7.3, we will use the moving average representation of $\mathbf{X}_t$ to obtain its lag functions and corresponding covariance and correlation matrices. Toward this end, first, by following the same process as was used in section 3.5 to obtain Eq (3.62), it can be shown that for each $f_j(\mathbf{Y}_{t-1}), j = 1, 2, \ldots, S$, when $\mathbf{Y}_t$ follows the MAR(p) model given in Eq (3.113), we have

$$f_j(\mathbf{Y}_{t-1}) = \mathbf{B}_1 \ f_j(\mathbf{Y}_{t-2}) + \mathbf{B}_2 \ f_j(\mathbf{Y}_{t-3}) + \ldots + \mathbf{B}_p \ f_j(\mathbf{Y}_{t-p-1}) + f_j(\boldsymbol{\varepsilon}_{t-1}) \tag{3.120}$$

where the $KS \times KS$ matrices $\mathbf{B}_\nu, \nu = 1, 2, \ldots, p$, have the same structure as has the matrix $\mathbf{B}$ in Eq (3.37). That is,

$$\mathbf{B}_\nu = \begin{bmatrix} \mathbf{A}^\dagger_{\nu 1} \\ \mathbf{A}^\dagger_{\nu 2} \\ \vdots \\ \mathbf{A}^\dagger_{\nu S} \end{bmatrix} = \begin{pmatrix} \mathbf{A}^1_{\nu 1} & \mathbf{A}^1_{\nu 2} & \ldots & \mathbf{A}^1_{\nu S} \\ \mathbf{A}^2_{\nu 1} & \mathbf{A}^2_{\nu 2} & \ldots & \mathbf{A}^2_{\nu S} \\ \vdots & & & \vdots \\ \mathbf{A}^S_{\nu 1} & \mathbf{A}^S_{\nu 2} & \ldots & \mathbf{A}^S_{\nu S} \end{pmatrix}, \qquad \nu = 1, 2, \ldots, p. \tag{3.121}$$

Then, with respect to the relationship between the $f_j(\mathbf{Y}_{t-1})$ and $\mathbf{Y}_t$ in Eqs (3.29)-(3.31), it is easy to show that

$$f(\mathbf{Y}_{t-1}) = (\mathbf{I} \otimes \mathbf{B}_1) \ f(\mathbf{Y}_{t-2}) + (\mathbf{I} \otimes \mathbf{B}_2) \ f(\mathbf{Y}_{t-3}) + \ldots + (\mathbf{I} \otimes \mathbf{B}_p) \ f(\mathbf{Y}_{t-p-1}) + f(\boldsymbol{\varepsilon}_{t-1}). \tag{3.122}$$

Therefore, following the same idea for defining the autoregressive process $\mathbf{X}_t$ in Eq (3.118) which is a congruence equation of the matrix autoregressive process of order p, $\mathbf{Y}_t$, in Eq (3.113), we can define the congruence equation of $f(\mathbf{Y}_{t-1})$ defined in Eq (3.122). To this end, the matrix function $C(\mathbf{Y}_{t-1})$ given in Eq (3.114) can be written as follows

$$C(\mathbf{Y}_{t-1}) = \begin{bmatrix} f(\mathbf{Y}_{t-1}) \\ f(\mathbf{Y}_{t-2}) \\ \vdots \\ f(\mathbf{Y}_{t-p}) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \dots & \mathbf{D}_{p-1} & \mathbf{D}_p \\ \mathbf{I}_{KS^2} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{KS^2} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{KS^2} & \mathbf{0} \end{bmatrix} \begin{bmatrix} f(\mathbf{Y}_{t-2}) \\ f(\mathbf{Y}_{t-3}) \\ \vdots \\ f(\mathbf{Y}_{t-p-1}) \end{bmatrix} + \begin{bmatrix} f(\boldsymbol{\varepsilon}_{t-1}) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

$$= \mathbf{G}_{p\times p} C(\mathbf{Y}_{t-2}) + f(\mathbf{u}_{t-1}) \tag{3.123}$$

where $\mathbf{D}_\nu = \mathbf{I} \otimes \mathbf{B}_\nu$, $\nu = 1, 2, \dots, p$. Hence, the matrix function $C(\mathbf{Y}_{t-1})$ has a recursive property and its given by

$$C(\mathbf{Y}_{t-1}) = \mathbf{G}_{p\times p} C(\mathbf{Y}_{t-2}) + f(\mathbf{u}_{t-1}) \tag{3.124}$$

where $\mathbf{G}_{p\times p}$ is a $p \times p$ block matrix such that each block is a $KS^2 \times KS^2$ matrix, and $f(\mathbf{u}_{t-1})$ is a $p \times 1$ matrix function such that the first block is the $KS^2 \times S$ matrix function $f(\boldsymbol{\varepsilon}_{t-1})$ and the other blocks are zero. That is,

$$\mathbf{G}_{p\times p} = \begin{bmatrix} \mathbf{I}\otimes\mathbf{B}_1 & \mathbf{I}\otimes\mathbf{B}_2 & \dots & \mathbf{I}\otimes\mathbf{B}_{p-1} & \mathbf{I}\otimes\mathbf{B}_p \\ \mathbf{I}_{KS^2} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{KS^2} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{KS^2} & \mathbf{0} \end{bmatrix}, \quad f(\mathbf{u}_{t-1}) = \begin{bmatrix} f(\boldsymbol{\varepsilon}_{t-1}) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{3.125}$$

Now, by applying the recursive property of the matrix function $C(\mathbf{Y}_{t-1})$ of Eq (3.124) to Eq (3.118), it can be shown that the moving average representation of $\mathbf{X}_t$ is given by the following geometric sum of the past matrix error terms

$$\mathbf{X}_t = \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1}) + \mathbf{u}_t \tag{3.126}$$

where $\mathbf{A}$ and $\mathbf{G}$ are defined in Eq (3.115) and Eq (3.125), respectively. To verify Eq (3.126), note that we can start with the MAR(1) model in Eq (3.118), and substitute the recursive formula of $C(\mathbf{Y}_{t-1})$ in Eq (3.124) into Eq (3.118) as follows

$$\begin{aligned}
\mathbf{X}_t &= \mathbf{A}C(\mathbf{Y}_{t-1}) + \mathbf{u}_t \\
&= \mathbf{A}[\mathbf{G}C(\mathbf{Y}_{t-2}) + f(\mathbf{u}_{t-1})] + \mathbf{u}_t \\
&= \mathbf{A}\mathbf{G}C(\mathbf{Y}_{t-2}) + \mathbf{A}f(\mathbf{u}_{t-1}) + \mathbf{u}_t \\
&= \mathbf{A}\mathbf{G}^2C(\mathbf{Y}_{t-3}) + \mathbf{A}\mathbf{G}f(\mathbf{u}_{t-2}) + \mathbf{A}f(\mathbf{u}_{t-1}) + \mathbf{u}_t \\
&\vdots \\
&= \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1}) + \mathbf{u}_t.
\end{aligned}$$

With this moving average representation of the MAR(p) model, we can now find the autocovariance and autocorrelation functions of the MAR(p) model.

3.7.3 Autocovariance and autocorrelation functions of $MAR(p)$

In this section, the autocovariance and autocorrelation functions of a matrix autoregressive process of order p will be derived. Since we have the moving average form of the matrix autoregressive process of order p in Eq (3.126), it is easy to find the autocovariance functions of a MAR(p) model following analogous arguments used in section 3.6 for the MAR(1) process.

Let $\boldsymbol{\Psi}_{\mathbf{Y}}(h)$ be the lag function of the MAR(p) model of Eq (3.113) at lag h, and let $\boldsymbol{\Psi}_{\mathbf{X}}(h)$ represent the lag function of the MAR(1) process defined in Eq (3.118) at lag h. Therefore, the stochastic process $\mathbf{X}_t$ can be used to obtain properties of the $MAR(p)$ process $\mathbf{Y}_t$. That is, the mean of the the MAR(p) model is

$$E[\mathbf{Y}_t] = \mathbf{J}^* E[\mathbf{X}_t] \tag{3.127}$$

and we can apply Proposition (3.6.1) to obtain the lag function at lag h of the MAR(p) model from quantity $\mathbf{Y}_t = \mathbf{J}^*\mathbf{X}_t$. That is,

$$\begin{aligned}\mathbf{\Psi}_{\mathbf{Y}}(h) &= E[(\mathbf{J}^*\mathbf{X}_t) \otimes (\mathbf{J}^*\mathbf{X}_t)^T] = (\mathbf{J}^* \otimes \mathbf{I})E[\mathbf{X}_t \otimes \mathbf{X}_t^T](\mathbf{I} \otimes \mathbf{J}^{*T}) \\ &= (\mathbf{J}^* \otimes \mathbf{I})\mathbf{\Psi}_{\mathbf{X}}(h)(\mathbf{I} \otimes \mathbf{J}^{*T}). \end{aligned} \tag{3.128}$$

After finding the lag function $\mathbf{\Psi}_{\mathbf{Y}}(h)$, the autocovariance function at lag h, $\mathbf{\Gamma}_{\mathbf{Y}}(h)$, can be found similarly to the derivation of Eq (3.67) by premultiplying the transformation matrix $\mathbf{T}$ given in Eq (3.6) by the lag function. That is,

$$\mathbf{\Gamma}_{\mathbf{Y}}(h) = \mathbf{T}\mathbf{\Psi}_{\mathbf{Y}}(h). \tag{3.129}$$

3.7.3.1 Autocovariance function at lag zero

Similarly to the derivation of the autocovariance and autocorrelation functions for the matrix autoregressive model of order one in section 3.6, we start with the variance-covariance of the random process $\mathbf{X}_t$ which is also the autocovariance function at lag zero. To this end, first let $\mathbf{\Psi}_X(0)$ be the expectation of the Kronecker product matrix autoregressive process $\mathbf{X}_t$ into its transpose. That is,

$$\mathbf{\Psi}_X(0) = E[\mathbf{X}_t \otimes \mathbf{X}_t^T]. \tag{3.130}$$

Then, since $\mathbf{X}_t$ involves $f(\mathbf{u}_t)$ (from Eq (3.126)) we need to find the expectation of the Kronecker product $f(\mathbf{u}_t) \otimes f^T(\mathbf{u}_t)$. First, note that in section 3.6, it was shown that $E[f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t)] = \mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \mathbf{\Sigma}^*)$ (see Eq (3.83)). From Eq (3.125), we have

$$f(\mathbf{u}_t) \otimes f^T(\mathbf{u}_t) = \begin{bmatrix} f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} = \mathbf{E}_{11}^p \otimes f(\boldsymbol{\varepsilon}_t) \otimes f^T(\boldsymbol{\varepsilon}_t) \tag{3.131}$$

where $\mathbf{E}_{11}^p$ is a $p \times p$ matrix where all of its entries are zero except the first entry. Indeed, the matrix $\mathbf{E}_{11}^p$ is defined analogously as the $\mathbf{E}_{ij}$'s in Eq (3.22), except that the $\mathbf{E}_{ij}$ matrices had dimension $S \times S$ whereas now $\mathbf{E}_{11}^p$ has dimension $p \times p$. Then, taking expectations on both sides of Eq (3.131), we have

$$E[f(\mathbf{u}_t) \otimes f^T(\mathbf{u}_t)] = \mathbf{E}_{11}^p \otimes \mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \boldsymbol{\Sigma}^*). \tag{3.132}$$

Furthermore, similarly to this process, and with respect to the definition of $\mathbf{u}_t$ in Eq (3.114), and noting that $E[\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T] = \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}$ (see Eq (3.4)), it can be shown that

$$\boldsymbol{\Psi}_u = E(\mathbf{u}_t \otimes \mathbf{u}_t^T) = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} = \mathbf{E}_{11}^p \otimes \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}. \tag{3.133}$$

Now, substituting the moving average representation of the autoregressive process $\mathbf{X}_t$ given in Eq (3.126), and using the fact that the random matrix error $\boldsymbol{\varepsilon}_t$'s and therefore the linear matrix function $f(\mathbf{u}_t)$'s are uncorrelated, we can show that the expectation of the Kronecker product of the $\mathbf{X}_t$, $\boldsymbol{\Psi}_X(0)$, of Eq (3.130) becomes

$$\begin{aligned}\mathbf{\Psi}_X(0) &= E(\mathbf{X}_t \otimes \mathbf{X}_t^T) \\ &= E([\sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1})] \otimes [\sum_{m=0}^{\infty} \mathbf{A}\mathbf{G}^m f(\mathbf{u}_{t-m-1})]^T) + E(\mathbf{u}_t \otimes \mathbf{u}_t^T) \\ &= \sum_{n=0}^{\infty} E([\mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1})] \otimes [f^T(\mathbf{u}_{t-n-1})\mathbf{G}^{nT}\mathbf{A}^T]) + E(\mathbf{u}_t \otimes \mathbf{u}_t^T). \end{aligned} \tag{3.134}$$

Then, by applying Proposition 3.6.1 and using the results of Eq (3.132) and Eq (3.133), Eq (3.134) can be simplified as

$$\begin{aligned}\mathbf{\Psi}_X(0) &= \sum_{n=0}^{\infty} [\mathbf{A}\mathbf{G}^n \otimes \mathbf{I}] \; E[f(\mathbf{u}_t) \otimes f^T(\mathbf{u}_t)][\mathbf{I} \otimes \mathbf{G}^{nT}\mathbf{A}^T] + E(\mathbf{u}_t \otimes \mathbf{u}_t^T) \\ &= \sum_{n=0}^{\infty} [\mathbf{A}\mathbf{G}^n \otimes \mathbf{I}][\mathbf{E}_{11}^p \otimes \mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \mathbf{\Sigma}^*)][\mathbf{I} \otimes \mathbf{G}^{nT}\mathbf{A}^T] + \mathbf{\Psi}_u. \end{aligned} \tag{3.135}$$

Then, by using Eq (3.128), we can find the lag function of the MAR(p) model defined in Eq (3.113), $\mathbf{\Psi}_{\mathbf{Y}}(0)$. That is, we obtain

$$\mathbf{\Psi}_{\mathbf{Y}}(0) = (\mathbf{J}^* \otimes \mathbf{I})\mathbf{\Psi}_{\mathbf{X}}(0)(\mathbf{I} \otimes \mathbf{J}^{*T}). \tag{3.136}$$

Eventually, from Eq (3.129), we can find the autocovariance function of the $MAR(p)$ model at lag zero, $\mathbf{\Gamma}_{\mathbf{Y}}(0)$, by premultiplying the $\mathbf{\Psi}_{\mathbf{Y}}(0)$ in Eq (3.136) by the transformation matrix $\mathbf{T}$ as follows

$$\mathbf{\Gamma}_{\mathbf{Y}}(0) = \mathbf{T}\mathbf{\Psi}_{\mathbf{Y}}(0). \tag{3.137}$$

3.7.3.2 Marginal vector $X_{.jt}$ of $\mathbf{X}_t$

In this section, we will study features of the j^{th} marginal vector of a matrix time series $\mathbf{Y}_t$ of order p for $j = 1, 2, \ldots, S$. This study is important both for finding the block entries of the autocovariance matrix $\mathbf{\Gamma}_X(0)$, and also for perusing and analyzing a single vector of the MAR(p) model itself. First, note that similarly to the marginal vector $Y_{.jt}$ in Eq (3.86) for the MAR(1) process, the marginal vector $X_{.jt}$ of the MAR(p) process can be obtained from the matrix process $\mathbf{X}_t$ by multiplying $\mathbf{X}_t$ and the vector $\mathbf{e}_j$, that is,

$$X_{.jt} = \mathbf{X}_t \mathbf{e}_j, \qquad j = 1, 2, \ldots, S. \tag{3.138}$$

Let $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$ be the jj'^{th} block entry of the autocovariance matrix $\mathbf{\Gamma}_{\mathbf{X}}(0)$. Based upon the definition of the random process $\mathbf{X}_t$ in Eq (3.114), $X_{.jt}$ is a vector with dimension Kp given by $X_{.jt}^T = (Y_{.jt}^T, Y_{.j(t-1)}^T, \ldots, Y_{.j(t-p+1)}^T)$. Hence, the variance-covariance matrix $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$ has dimension $Kp \times Kp$. To calculate the variance-covariance matrix $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$, first we need to find the moving average representation of the random vector $X_{.jt}$. To this end, from the moving average representation of the matrix process $\mathbf{X}_t$ in Eq (3.126) and its connection to the vector $X_{.jt}$ in Eq (3.138), we have

$$X_{.jt} = \mathbf{X}_t \mathbf{e}_j = \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1})\mathbf{e}_j + \mathbf{u}_t \mathbf{e}_j = \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n W_j(\mathbf{u}_{t-n-1}) + \mathbf{u}_{.jt} \tag{3.139}$$

where $W_j(\mathbf{u}_{t-n-1})$ is a vector function and $\mathbf{u}_{.jt}$ is a vector of white noise, with dimension KS^2p and Kp, respectively; $W_j(\mathbf{u}_{t-n-1})$ contains elements $V_j(\boldsymbol{\varepsilon}_{t-n-1})$ which are as defined in Eq (3.88), that is,

$$W_j(\mathbf{u}_{t-n-1}) = \begin{pmatrix} V_j(\boldsymbol{\varepsilon}_{t-n-1}) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \qquad \mathbf{u}_{.jt} = \begin{pmatrix} \boldsymbol{\varepsilon}_{.jt} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}. \tag{3.140}$$

Therefore, by using the moving average representation of $X_{.jt}$ in Eq (3.139), and with respect to the fact that the $\boldsymbol{\varepsilon}_t$'s and hence the $W_j(\mathbf{u}_t)$'s are white noise, the autocovariance matrix (at lag zero) $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$ can be written as

$$\begin{aligned}
\mathbf{\Gamma}^T_{jj',\mathbf{X}}(0) &= Cov(X_{.jt}, X_{.j't}) = E[X_{.jt}X^T_{.j't}] \\
&= E\Big[\big(\sum_{n=0}^{\infty}\mathbf{A}\mathbf{G}^n W_j(\mathbf{u}_{t-n-1})\big)\big(\sum_{m=0}^{\infty}\mathbf{A}\mathbf{G}^m W_{j'}(\mathbf{u}_{t-m-1})\big)^T\Big] + E[\mathbf{u}_{.jt}\mathbf{u}^T_{.j't}] \\
&= \sum_{n=0}^{\infty}\mathbf{A}\mathbf{G}^n E\Big[W_j(\mathbf{u}_{t-n-1})W^T_{j'}(\mathbf{u}_{t-n-1})\Big]\mathbf{G}^{nT}\mathbf{A}^T + E[\mathbf{u}_{.jt}\mathbf{u}^T_{.j't}]. \qquad (3.141)
\end{aligned}$$

Recall from Eq (3.95) that we found $E(V_j(\boldsymbol{\varepsilon}_t){V_{j'}}^T(\boldsymbol{\varepsilon}_t)) = \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^*$; therefore, according to the definition of the $W_j(\mathbf{u}_{t-n-1})$ in Eq (3.140), it is easy to show that

$$E\big[W_j(\mathbf{u}_{t-n-1})W^T_{j'}(\mathbf{u}_{t-n-1})\big] = \mathbf{E}^p_{11} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^*, \qquad (3.142)$$

and $E[\mathbf{u}_{.jt}\mathbf{u}^T_{.j't}] = \mathbf{E}^p_{11} \otimes \mathbf{\Sigma}^*_{jj'}$, where $\mathbf{E}^p_{11}$ is similar to that used in Eq (3.133). Therefore, by substituting $E[\mathbf{u}_{.jt}\mathbf{u}^T_{.j't}] = \mathbf{E}^p_{11} \otimes \mathbf{\Sigma}^*_{jj'}$ and Eq (3.142) into Eq (3.141), and then by applying the Kronecker product rule $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, the variance-covariance matrix $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$ can be simplified to, for $j, j' = 1, 2, \dots, S$,

$$\mathbf{\Gamma}^T_{jj',\mathbf{X}}(0) = \sum_{n=0}^{\infty}\mathbf{A}\mathbf{G}^n(\mathbf{E}^p_{11} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^*)\mathbf{G}^{nT}\mathbf{A}^T + \mathbf{E}^p_{11} \otimes \mathbf{\Sigma}^*_{jj'}. \qquad (3.143)$$

To obtain the block entries of the autocovariance function at lag zero of the matrix autoregressive precess of order p, $\mathbf{\Gamma}_{jj',\mathbf{Y}}(0)$, defined in Eq (3.137), first note that $Y_{.jt}$ can be found by premultiplying $X_{.jt}$ by the $K \times Kp$ matrix $\mathbf{J}^*$ defined in Eq (3.119), that is,

$$Y_{.jt} = \mathbf{J}^* X_{.jt}. \qquad (3.144)$$

Therefore, based on Eq (3.144) and using the result in Eq (3.143) for $\mathbf{\Gamma}_{jj',\mathbf{X}}(0)$, we can derive the autocovariance function $\mathbf{\Gamma}_{jj'\mathbf{Y}}(0)$ as follows, for $j, j' = 1, 2, \dots, S$,

$$
\begin{aligned}
\mathbf{\Gamma}^T_{jj',\mathbf{Y}}(0) &= E[Y_{.jt}Y^T_{.j't}] = \mathbf{J}^* E[X_{.jt}X^T_{.j't}]\mathbf{J}^{*T} = \mathbf{J}^*\mathbf{\Gamma}_{jj',\mathbf{X}}(0)\mathbf{J}^{*T} \\
&= \sum_{n=0}^{\infty} \mathbf{J}^*\mathbf{A}\mathbf{G}^n(\mathbf{E}^p_{11} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^*)\mathbf{G}^{nT}\mathbf{A}^T\mathbf{J}^{*T} + \mathbf{J}^*(\mathbf{E}^p_{11} \otimes \mathbf{\Sigma}^*_{jj'})\mathbf{J}^{*T} \\
&= \sum_{n=0}^{\infty} \mathbf{J}^*\mathbf{A}\mathbf{G}^n(\mathbf{E}^p_{11} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^*)\mathbf{G}^{nT}\mathbf{A}^T\mathbf{J}^{*T} + \mathbf{\Sigma}^*_{jj'}. \qquad (3.145)
\end{aligned}
$$

3.7.3.3 Autocovariance function at lag $h > 0$

Let $\mathbf{\Psi}_X(h)$ be the lag function defined to be the expectation of the Kronecker product $\mathbf{X}_{t+h} \otimes \mathbf{X}^T_t$. Similar to Eq (3.98) for the matrix time series $\mathbf{Y}_{t+h}$, the moving average representation of the random process $\mathbf{X}_{t+h}$ defined in Eq (3.126) can be rewritten as

$$
\begin{aligned}
\mathbf{X}_{t+h} &= \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t+h-n-1}) + \mathbf{u}_{t+h} \\
&= \sum_{n=0}^{h-1} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} f(\mathbf{u}_{t-n-1}) + \mathbf{u}_{t+h}. \qquad (3.146)
\end{aligned}
$$

From Eq (3.146), and using the fact that the matrix function $f(\mathbf{u}_t)$'s are a function of matrix white noise $\boldsymbol{\varepsilon}_t$ and hence are uncorrelated, the lag function $\mathbf{\Psi}_X(h)$ can be written as follows

$$
\begin{aligned}
\mathbf{\Psi}_X(h) &= E[\mathbf{X}_{t+h} \otimes \mathbf{X}^T_t] \\
&= E\Big(\sum_{n=0}^{h-1} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} f(\mathbf{u}_{t-n-1}) + \mathbf{u}_{t+h} \Big) \\
&\qquad \otimes \Big(\sum_{m=0}^{\infty} \mathbf{A}\mathbf{G}^m f(\mathbf{u}_{t-m-1}) + \mathbf{u}_t \Big)^T \\
&= \sum_{n=0}^{\infty} E\big([\mathbf{A}\mathbf{G}^{n+h} f(\mathbf{u}_{t-n-1})] \otimes [f^T(\mathbf{u}_{t-n-1})\mathbf{G}^{nT}\mathbf{A}^T]\big) + E\big([\mathbf{A}\mathbf{G}^{h-1} f(\mathbf{u}_t)] \otimes [\mathbf{u}^T_t]\big).
\end{aligned}
$$

(3.147)

Now by applying Proposition 3.6.1, the lag function $\mathbf{\Psi}_X(h)$ in Eq (3.147) becomes

$$\mathbf{\Psi}_X(h) = \sum_{n=0}^{\infty}(\mathbf{AG}^{n+h} \otimes \mathbf{I})E\big(f(\mathbf{u}_{t-n-1}) \otimes f^T(\mathbf{u}_{t-n-1})\big)(\mathbf{I} \otimes \mathbf{G}^{nT}\mathbf{A}^T)$$
$$+ (\mathbf{AG}^{h-1} \otimes \mathbf{I})E\big(f(\mathbf{u}_t) \otimes \mathbf{u}_t^T\big). \qquad (3.148)$$

The expectation of the Kronecker product $f(\mathbf{u}_{t-n-1}) \otimes f^T(\mathbf{u}_{t-n-1})$ was given in Eq (3.132). According to the definition of $f(\mathbf{u}_t)$ and $\mathbf{u}_t$ in Eq (3.125) and Eq (3.114), respectively, and with respect to Eq (3.106), it can be shown that

$$E\big(f(\mathbf{u}_t) \otimes \mathbf{u}_t\big) = \mathbf{e}_1^p \otimes \mathbf{I} \otimes Vecb(\mathbf{\Psi}_{\boldsymbol{\varepsilon}}) \qquad (3.149)$$

where $\mathbf{e}_1^p$ is a p-dimensional vector given by $\mathbf{e}_1^p = (1, 0, \ldots, 0)^T$, and $Vecb(\mathbf{\Psi}_{\boldsymbol{\varepsilon}})$ is as defined in Eq (3.105). Therefore, by substituting these results into Eq (3.148), $\mathbf{\Psi}_X(h)$ can be obtained as

$$\mathbf{\Psi}_X(h) = \sum_{n=0}^{\infty}(\mathbf{AG}^{n+h} \otimes \mathbf{I})\big(\mathbf{E}_{11}^p \otimes \mathbf{I} \otimes \mathbf{T}^{*-1}(\mathbf{I} \otimes \mathbf{\Sigma}^*)\big)(\mathbf{I} \otimes \mathbf{G}^{nT}\mathbf{A}^T)$$
$$+ (\mathbf{AG}^{h-1} \otimes \mathbf{I})\big(\mathbf{e}_1^p \otimes \mathbf{I} \otimes Vecb(\mathbf{\Psi}_{\boldsymbol{\varepsilon}})\big). \qquad (3.150)$$

Analogous to the derivation of the autocovariance function at lag zero in Eqs (3.136) and (3.137), the autocovariance function of the MAR(p) model $\mathbf{Y}_t$ at lag $h > 0$, $\mathbf{\Gamma}_{\mathbf{Y}}(h)$, can be obtained by finding the corresponding lag function $\mathbf{\Psi}_{\mathbf{Y}}(h)$, and then by premultiplying it by the transformation matrix $\mathbf{T}$. That is,

$$\mathbf{\Psi}_{\mathbf{Y}}(h) = (\mathbf{J}^* \otimes \mathbf{I})\mathbf{\Psi}_{\mathbf{X}}(h)(\mathbf{I} \otimes \mathbf{J}^{*T}), \qquad \mathbf{\Gamma}_{\mathbf{Y}}(h) = \mathbf{T}\mathbf{\Psi}_{\mathbf{Y}}(h). \qquad (3.151)$$

Block entries of this new autocovariance matrix $\mathbf{\Gamma}_{\mathbf{Y}}(h)$ can be obtained in the same way as for the autocovariance matrices $\mathbf{\Gamma}(0)$ and $\mathbf{\Gamma}(h)$ in sections 3.6.2 and 3.6.3.

Similarly to the case of $h = 0$ in Section 3.7.3.2, we can derive the variance-covariance matrix $\mathbf{\Gamma}_{jj',\mathbf{Y}}(h)$ for $h > 0$. Toward this end, first note that, from Eqs (3.138) and (3.146), $X_{.jt+h}$ can be written as, for $j, j' = 1, 2, \ldots, S$,

$$\begin{aligned} X_{.jt+h} = \mathbf{X}_{t+h}\mathbf{e}_j &= \sum_{n=0}^{h-1} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t+h-n-1})\mathbf{e}_j + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} f(\mathbf{u}_{t-n-1})\mathbf{e}_j + \mathbf{u}_{t+h}\mathbf{e}_j \\ &= \sum_{n=0}^{h-1} \mathbf{A}\mathbf{G}^n W_j(\mathbf{u}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} W_j(\mathbf{u}_{t-n-1}) + \mathbf{u}_{.jt+h}. \end{aligned} \quad (3.152)$$

By definition, the variance-autocovariance matrix (at lag $h > 0$), $\mathbf{\Gamma}_{jj',\mathbf{X}}(h)$, can be obtained as

$$\begin{aligned} \mathbf{\Gamma}^T_{jj',\mathbf{X}}(h) &= Cov(X_{.jt+h}, X_{.j't}) = E[X_{.jt+h}X^T_{.j't}] \\ &= E\Big[\Big(\sum_{n=0}^{h-1} \mathbf{A}\mathbf{G}^n W_j(\mathbf{u}_{t+h-n-1}) + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} W_j(\mathbf{u}_{t-n-1}) \\ + & \qquad \mathbf{u}_{.jt+h})\Big)\Big(\sum_{m=0}^{\infty} \mathbf{A}\mathbf{G}^m W_{j'}(\mathbf{u}_{t-m-1}) + \mathbf{u}_{.j't}\Big)^T\Big] \\ &= \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h} E\Big[W_j(\mathbf{u}_{t-n-1})W^T_{j'}(\mathbf{u}_{t-n-1})\Big]\mathbf{G}^{nT}\mathbf{A}^T + \mathbf{A}\mathbf{G}^{h-1}E[W_j(\mathbf{u}_t)\mathbf{u}^T_{.j't}]. \end{aligned} \quad (3.153)$$

From Eq (3.142), we know $E\Big[W_j(\mathbf{u}_{t-n-1})W^T_{j'}(\mathbf{u}_{t-n-1})\Big]$, but we need to determine the $E[W_j(\mathbf{u}_t)\mathbf{u}^T_{.j't}]$. To this end, first note that from Eq (3.140) $W_j(\mathbf{u}_t)$ and $\mathbf{u}_{.j't}$ can be written as

$$W_j(\mathbf{u}_t) = \mathbf{e}^p_1 \otimes V_j(\boldsymbol{\varepsilon}_t), \qquad \mathbf{u}_{.j't} = \mathbf{e}^p_1 \otimes \boldsymbol{\varepsilon}_{.j't}. \quad (3.154)$$

Hence, we have

$$W_j(\mathbf{u}_t)\mathbf{u}^T_{.j't} = \big(\mathbf{e}^p_1 \otimes V_j(\boldsymbol{\varepsilon}_t)\big)\big(\mathbf{e}^p_1 \otimes \boldsymbol{\varepsilon}_{.j't}\big)^T = (\mathbf{e}^p_1\mathbf{e}^{pT}_1 \otimes V_j(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.j't}) = \mathbf{E}^p_{11} \otimes V_j(\boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}^T_{.j't} \quad (3.155)$$

where $\mathbf{E}_{11}^{p}$ is the same as in Eqs (3.131), (3.132), and (3.133). Now, by taking expectations on both sides of Eq (3.155), and using the result of Eq (3.111), we obtain

$$E[W_j(\mathbf{u}_t)\mathbf{u}_{.j't}^T] = \mathbf{E}_{11}^{p} \otimes \mathbf{e}_j \otimes \mathbf{\Sigma}_{j'}^{*}, \qquad j, j' = 1, 2, \ldots, S. \tag{3.156}$$

By substituting Eqs (3.142) and (3.156) into Eq (3.153), the variance-covariance matrix $\mathbf{\Gamma}_{jj',\mathbf{X}}(h)$ is obtained as

$$\mathbf{\Gamma}_{jj',\mathbf{X}}^T(h) = \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^{n+h}\big(\mathbf{E}_{11}^{p} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^{*}\big)\mathbf{G}^{nT}\mathbf{A}^T + \mathbf{A}\mathbf{G}^{h-1}\big(\mathbf{E}_{11}^{p} \otimes \mathbf{e}_j \otimes \mathbf{\Sigma}_{j'}^{*}\big). \tag{3.157}$$

Eventually, by using the relationship between the vectors $X_{.jt}$ and $Y_{.jt}$ in Eq (3.144), the variance-covariance function $\mathbf{\Gamma}_{jj',\mathbf{Y}}^T(h)$ is derived as follows, for $j, j' = 1, 2, \ldots, S$,

$$\begin{aligned}\mathbf{\Gamma}_{jj',\mathbf{Y}}^T(h) &= E[Y_{.jt+h}Y_{.j't}^T] = \mathbf{J}^{*}E[X_{.jt+h}X_{.j't}^T]\mathbf{J}^{*T} = \mathbf{J}^{*}\mathbf{\Gamma}_{jj',\mathbf{X}}(h)\mathbf{J}^{*T}\\ &= \sum_{n=0}^{\infty} \mathbf{J}^{*}\mathbf{A}\mathbf{G}^{n+h}(\mathbf{E}_{11}^{p} \otimes \mathbf{E}_{jj'} \otimes \mathbf{\Sigma}^{*})\mathbf{G}^{nT}\mathbf{A}^T\mathbf{J}^{*T} + \mathbf{J}^{*}\mathbf{A}\mathbf{G}^{h-1}\big(\mathbf{E}_{11}^{p} \otimes \mathbf{e}_j \otimes \mathbf{\Sigma}_{j'}^{*}\big)\mathbf{J}^{*T}.\end{aligned} \tag{3.158}$$

3.8 Matrix Autoregressive Process with Nonzero Mean

So far, we assumed that the matrix time series $\mathbf{Y}_t$, given in Eq (3.2), has mean zero; based on this assumption, we introduced and explored the autoregressive process of the matrix time series $\mathbf{Y}_t$. In practice, this assumption may not be satisfied and it is worth investigating and analyzing models with nonzero mean. To this end, let $\boldsymbol{\mu}$ be the intercept matrix of the matrix time series $\mathbf{Y}_t$ each with dimension $K \times S$. Then, the matrix autoregressive process of order one (MAR(1)) of the matrix time series $\mathbf{Y}_t$ as first introduced in Eq (3.27) is

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A} f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1S} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2S} \\ \vdots & & \ddots & \vdots \\ \mu_{K1} & \mu_{K2} & \dots & \mu_{KS} \end{bmatrix}. \tag{3.159}$$

Now we can find the moving average representation of the matrix time series with nonzero mean in Eq (3.159). First, note that it is easy to show that the matrix function $f(\mathbf{Y}_{t-1})$ in Eq (3.159) has the following recursive property

$$f(\mathbf{Y}_{t-1}) = f(\boldsymbol{\mu}) + (\mathbf{I} \otimes \mathbf{B})\ f(\mathbf{Y}_{t-2}) + f(\boldsymbol{\varepsilon}_{t-1}). \tag{3.160}$$

Based on the recursive model given in Eq (3.160), the autoregressive process $\mathbf{Y}_t$ in Eq (3.159) can be simplified as

$$\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu} + \mathbf{A} f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\mu} + \mathbf{A}[f(\boldsymbol{\mu}) + (\mathbf{I} \otimes \mathbf{B})\ f(\mathbf{Y}_{t-2}) + f(\boldsymbol{\varepsilon}_{t-1})] + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\mu} + \mathbf{A} f(\boldsymbol{\mu}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})\ f(\mathbf{Y}_{t-2}) + \mathbf{A} f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\mu} + \mathbf{A} f(\boldsymbol{\mu}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})\ [f(\boldsymbol{\mu}) + (\mathbf{I} \otimes \mathbf{B})\ f(\mathbf{Y}_{t-3}) + f(\boldsymbol{\varepsilon}_{t-2})] + \mathbf{A} f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\mu} + \mathbf{A} f(\boldsymbol{\mu}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})\ f(\boldsymbol{\mu}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^2\ f(\mathbf{Y}_{t-3}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B}) f(\boldsymbol{\varepsilon}_{t-2}) \\
&\qquad + \mathbf{A} f(\boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&\ \ \vdots \\
&= \boldsymbol{\mu} + \Big(\mathbf{A} + \mathbf{A}(\mathbf{I} \otimes \mathbf{B}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^2 + \ldots + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^r\Big) f(\boldsymbol{\mu}) \\
&\qquad + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^{r+1}\ f(\mathbf{Y}_{t-r-2}) + \sum_{i=0}^{r} \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^i\ f(\boldsymbol{\varepsilon}_{t-i-1}) + \boldsymbol{\varepsilon}_t.
\end{aligned} \tag{3.161}$$

If all eigenvalues of the coefficient matrix $\mathbf{B}$ (hence all eigenvalues of $\mathbf{I} \otimes \mathbf{B}$) have modulus less than one, then it can be shown that the sequence $\mathbf{A}(\mathbf{I} \otimes \mathbf{B})^i$, $i = 0, 1, 2, \ldots$, is absolutely summable (see Lütkepohl, 2006). Therefore, the infinite sum $\sum_{i=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^i f(\boldsymbol{\varepsilon}_{t-i-1})$ exists in mean square. Furthermore, it can be shown that

$$\big(\mathbf{A} + \mathbf{A}(\mathbf{I} \otimes \mathbf{B}) + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^2 + \ldots + \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^r\big) f(\boldsymbol{\mu}) \longrightarrow^{r \uparrow \infty} \frac{\mathbf{A}}{\mathbf{I}_{KS^2} - (\mathbf{I} \otimes \mathbf{B})} f(\boldsymbol{\mu}). \quad (3.162)$$

Also, because it is assumed that all eigenvalues of $\mathbf{B}$, and therefore also of $\mathbf{I} \otimes \mathbf{B}$, have modulus less than one, then $(\mathbf{I} \otimes \mathbf{B})^{r+1}$ converges to zero rapidly as $r \to \infty$; hence, the term $\mathbf{A}(\mathbf{I} \otimes \mathbf{B})^{r+1}\ f(\mathbf{Y}_{t-r-2})$ can be ignored in the limit. Therefore, if all eigenvalues of $\mathbf{B}$ have modulus less than one, then the moving average representation of the matrix autoregressive process of $\mathbf{Y}_t$ in Eq (3.159) is

$$\mathbf{Y}_t = \boldsymbol{\upsilon} + \sum_{n=0}^{\infty} \mathbf{A}(\mathbf{I} \otimes \mathbf{B})^n\ f(\boldsymbol{\varepsilon}_{t-n-1}) + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\upsilon} = \boldsymbol{\mu} + \frac{\mathbf{A}}{\mathbf{I}_{KS^2} - (\mathbf{I} \otimes \mathbf{B})} f(\boldsymbol{\mu}). \qquad (3.163)$$

Then, according to the properties of the moving average representation, it is easy to show that the expectation of the matrix time series $\mathbf{Y}_t$ in Eq (3.163) is equal to $\boldsymbol{\upsilon}$. That is,

$$E[\mathbf{Y}_t] = \boldsymbol{\upsilon}. \qquad (3.164)$$

The matrix autoregressive process of order p (MAR(p)) defined in Eq (3.118), also can be found when the matrix time series $\mathbf{Y}_t$ has intercept matrix $\boldsymbol{\mu} \neq 0$. In order to convert a MAR(p) model, with nonzero intercept $\boldsymbol{\mu}$, to a MAR(1) model, we can use exactly the same model as we used in Eq (3.118) except that we add one extra term $\boldsymbol{\mu}_X$. That is,

$$\mathbf{X}_t = \boldsymbol{\mu}_X + \mathbf{A}C(\mathbf{Y}_{t-1}) + \mathbf{u}_t \qquad (3.165)$$

where the reparameterized variables and parameters $\mathbf{X}_t$, $C(\mathbf{Y}_{t-1})$, $\mathbf{u}_t$, $\mathbf{A}$ are defined as in Eq (3.114) and Eq (3.115). Also, $\boldsymbol{\mu}_X$ like $\mathbf{X}_t$ and $\mathbf{u}_t$, is a $p \times 1$ block matrix where each block matrix is a $K \times S$ matrix. The first block entry of the block matrix $\boldsymbol{\mu}_X$ is the $K \times S$

intercept matrix $\boldsymbol{\mu}$, the intercept of the matrix time series $\mathbf{Y}_t$, and other block entries are zero. That is,

$$\boldsymbol{\mu}_X = \begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}_{(p\times 1)} . \tag{3.166}$$

With some algebra similar to that for the nonzero mean MAR(1) model through Eqs (3.161)-(3.163), it can be shown that the moving average representation of autoregressive process given in Eq (3.165) is equal to

$$\mathbf{X}_t = \boldsymbol{\upsilon}_X + \sum_{n=0}^{\infty} \mathbf{A}\mathbf{G}^n f(\mathbf{u}_{t-n-1}) + \mathbf{u}_t \tag{3.167}$$

where $\mathbf{A}$ and $\mathbf{G}$ is defined in Eqs (3.115) and (3.125), respectively, and $\boldsymbol{\upsilon}_X$ is equal to

$$\boldsymbol{\upsilon}_X = \boldsymbol{\mu}_X + \frac{\mathbf{A}}{(\mathbf{I}_{KS^2p} - \mathbf{G})} C(\boldsymbol{\mu}). \tag{3.168}$$

Moreover, from the moving average representation of $\mathbf{X}_t$ in Eq (3.167), we have

$$E(\mathbf{X}_t) = \boldsymbol{\upsilon}_X. \tag{3.169}$$

3.9 Yule-Walker Equations for MAR Processes

Let $\mathbf{Y}_t$ be a stationary MAR(1) process, with matrix white noise variance-covariance $\boldsymbol{\Gamma}(0) = \mathbf{T}\boldsymbol{\Psi}(0) = \mathbf{T}E[(\mathbf{Y}_t - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_t - \boldsymbol{\upsilon})^T]$, given by

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + \mathbf{A}f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu} + \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1}\mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t. \end{aligned} \tag{3.170}$$

The mean-adjusted form of Eq (3.170) can be written as

$$\mathbf{Y}_t - \boldsymbol{\upsilon} = \sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j(\mathbf{Y}_{t-1} - \boldsymbol{\upsilon})\mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t. \tag{3.171}$$

where $\boldsymbol{\upsilon} = E(\mathbf{Y})$ is defined in Eqs (3.163) and (3.164). Kronecker multiplication of Eq (3.171) by $(\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T$ gives us

$$(\mathbf{Y}_t - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T = \sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j(\mathbf{Y}_{t-1} - \boldsymbol{\upsilon})\mathbf{E}_{rj} \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T + \boldsymbol{\varepsilon}_t \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T. \tag{3.172}$$

Alternatively, this quantity, by using the Kronecker product rule $(AC \otimes BD) = (A \otimes B)(C \otimes D)$, can be written as

$$\begin{aligned}(\mathbf{Y}_t - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T = \sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{A}_r^j \otimes \mathbf{I})\big((\mathbf{Y}_{t-1} - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T\big)(\mathbf{E}_{rj} \otimes \mathbf{I}_K)\\ + \boldsymbol{\varepsilon}_t \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T.\end{aligned} \tag{3.173}$$

Now, by taking expectations on both sides of Eq (3.173), for $h > 0$, and recalling that $\boldsymbol{\Psi}(h) = E[(\mathbf{Y}_t - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t+h} - \boldsymbol{\upsilon})^T]$ (see section 3.6, Eq (3.99) and section 3.8, Eq (3.164)), we have

$$\boldsymbol{\Psi}(h) = \sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{A}_r^j \otimes \mathbf{I})\boldsymbol{\Psi}(h-1)(\mathbf{E}_{rj} \otimes \mathbf{I}_K). \tag{3.174}$$

For $h = 0$, let us use Eq (3.171) for $(\mathbf{Y}_t - \boldsymbol{\upsilon})^T$ and substitute it in the first term on the right side of Eq (3.172). Then, Eq (3.172) can be rewritten as

$$(\mathbf{Y}_t-\boldsymbol{\upsilon})\otimes(\mathbf{Y}_t-\boldsymbol{\upsilon})^T=\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j(\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})\mathbf{E}_{rj}\otimes\sum_{j'=1}^{S}\sum_{r'=1}^{S}\mathbf{E}_{j'r'}(\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})^T\mathbf{A}_{r'}^{j'T}$$
$$+\sum_{j=1}^{S}\sum_{r=1}^{S}\mathbf{A}_r^j(\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})\mathbf{E}_{rj}\otimes\boldsymbol{\varepsilon}_{t-1}^T+\boldsymbol{\varepsilon}_t\otimes(\mathbf{Y}_t-\boldsymbol{\upsilon})^T. \quad (3.175)$$

Here, by using the same Kronecker product rule as was used to obtain Eq (3.173), Eq (3.175) can be written as

$$(\mathbf{Y}_t-\boldsymbol{\upsilon})\otimes(\mathbf{Y}_t-\boldsymbol{\upsilon})^T=\sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S}(\mathbf{A}_r^j\otimes\mathbf{E}_{j'r'})\big((\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})\otimes(\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})^T\big)(\mathbf{E}_{rj}\otimes\mathbf{A}_{r'}^{j'T})$$
$$+\sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{A}_r^j\otimes\mathbf{I})\big((\mathbf{Y}_{t-1}-\boldsymbol{\upsilon})\otimes\boldsymbol{\varepsilon}_{t-1}^T\big)(\mathbf{I}\otimes\mathbf{E}_{rj})+\boldsymbol{\varepsilon}_t\otimes(\mathbf{Y}_t-\boldsymbol{\upsilon})^T.$$
$$(3.176)$$

Taking expectations on both sides of Eq (3.176) leads to

$$\boldsymbol{\Psi}(0)=\sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S}(\mathbf{A}_r^j\otimes\mathbf{E}_{j'r'})\boldsymbol{\Psi}(0)(\mathbf{E}_{rj}\otimes\mathbf{A}_{r'}^{j'T})+\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} \quad (3.177)$$

where $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}$ is given in Eq (3.4). To solve this equation for $\boldsymbol{\Psi}(0)$, Vec can be taken on both sides of Eq (3.177). Then, by applying the Vec operator rule $Vec(\mathbf{ABC})=(\mathbf{C}^T\otimes\mathbf{A})Vec(\mathbf{B})$, we have

$$Vec(\boldsymbol{\Psi}(0))=\sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S}(\mathbf{E}_{jr}\otimes\mathbf{A}_{r'}^{j'})\otimes(\mathbf{A}_r^j\otimes\mathbf{E}_{j'r'})Vec(\boldsymbol{\Psi}(0))+Vec(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}) \quad (3.178)$$

or

$$Vec(\boldsymbol{\Psi}(0))=\Big(\mathbf{I}_{K^2S^2}-\sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S}(\mathbf{E}_{jr}\otimes\mathbf{A}_{r'}^{j'})\otimes(\mathbf{A}_r^j\otimes\mathbf{E}_{j'r'})\Big)^{-1}Vec(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}). \quad (3.179)$$

The invertibility of $\Big(\mathbf{I}_{K^2S^2} - \sum\limits_{j=1}^{S}\sum\limits_{r=1}^{S}\sum\limits_{j'=1}^{S}\sum\limits_{r'=1}^{S}(\mathbf{E}_{jr} \otimes \mathbf{A}_{r'}^{j'}) \otimes (\mathbf{A}_{r}^{j} \otimes \mathbf{E}_{j'r'})\Big)$ follows from the stationarity of $\mathbf{Y}_t$. Therefore, the Yule-Walker equations for the MAR(1) stationary processes are given in Eqs (3.179) and (3.174).

To obtain the Yule-Walker equations for a matrix autoregressive model of order $p > 1$ (MAR(p)), first note that from Eq (3.113), we have

$$(\mathbf{Y}_t - \boldsymbol{\upsilon}) = \sum_{\nu=1}^{p} \mathbf{A}_\nu f(\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon}) + \boldsymbol{\varepsilon}_t, \tag{3.180}$$

and by using Eqs (3.26) and (3.27), for each ν, $\mathbf{A}_\nu f(\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon})$ can be written as

$$\mathbf{A}_\nu f(\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon}) = \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_{\nu r}^{j}(\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon})\mathbf{E}_{rj}. \tag{3.181}$$

Hence, by applying Eq (3.181) in Eq (3.180) we have

$$(\mathbf{Y}_t - \boldsymbol{\upsilon}) = \sum_{\nu=1}^{p}\sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_{\nu r}^{j}(\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon})\mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t. \tag{3.182}$$

Note that $\mathbf{A}_{\nu r}^{j} \in \mathbf{A}_\nu$, $\nu = 1, 2, \ldots, p$. Then, by Kronecker multiplying this quantity on the right side with $(\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T$, and then by using the same Kronecker product rule as was used to obtain Eq (3.173), we have

$$\begin{aligned}(\mathbf{Y}_t - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T = \sum_{\nu=1}^{p}\sum_{j=1}^{S}\sum_{r=1}^{S} \big(\mathbf{A}_{\nu r}^{j} \otimes \mathbf{I}\big)\big((\mathbf{Y}_{t-\nu} - \boldsymbol{\upsilon}) \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T\big)\big(\mathbf{E}_{rj} \otimes \mathbf{I}_K\big) \\ + \boldsymbol{\varepsilon}_t \otimes (\mathbf{Y}_{t-h} - \boldsymbol{\upsilon})^T. \end{aligned} \tag{3.183}$$

Now, taking expectations on both sides of Eq (3.183), for $h > 0$, and using the independency of the matrix error terms, leads to

$$\boldsymbol{\Psi}(h) = \sum_{\nu=1}^{p}\sum_{j=1}^{S}\sum_{r=1}^{S}(\mathbf{A}_{\nu r}^{j} \otimes \mathbf{I})\boldsymbol{\Psi}(h-\nu)(\mathbf{E}_{rj} \otimes \mathbf{I}_K). \tag{3.184}$$

If $\mathbf{A}^j_{\nu r}, r, j = 1, 2, \ldots, S, \nu = 1, 2, \ldots, p$, and lag functions $\mathbf{\Psi}(0), \mathbf{\Psi}(1), \ldots, \mathbf{\Psi}(p-1)$ in Eq (3.184) are known, then the recursive equation given in Eq (3.184) can be used to obtain lag functions $\mathbf{\Psi}(h)$ for $h \geq p$. Therefore, we need to derive the initial lag functions $\mathbf{\Psi}(h)$ for $|h| < p$. The MAR(1) process $\mathbf{X}_t$ given in Eq (3.118) can be applied to determine the initial lag functions.

First, note that we have $\mathbf{J}f(\mathbf{Y}_t) = \mathbf{Y}_t$ where $\mathbf{J}$ is defined in Eqs (3.116) and (3.117). Hence, $\mathbf{X}_t$ defined in Eq (3.114) can be rewritten as a function of the $C(\mathbf{Y}_t)$ defined in Eq (3.114) as follows

$$\mathbf{X}_t = (\mathbf{I}_p \otimes \mathbf{J})C(\mathbf{Y}_t). \tag{3.185}$$

Then, using the Tracy-Singh product for multiplying Eq (3.185) on the right side by $\mathbf{X}_t^T$, we have

$$\begin{aligned} \mathbf{X}_t \bowtie \mathbf{X}_t^T &= \big((\mathbf{I}_p \otimes \mathbf{J})C(\mathbf{Y}_t)\big) \bowtie \big(C^T(\mathbf{Y}_t)(\mathbf{I}_p \otimes \mathbf{J})^T\big) \\ &= \big((\mathbf{I}_p \otimes \mathbf{J}) \bowtie \mathbf{I}\big)\big(C(\mathbf{Y}_t) \bowtie C^T(\mathbf{Y}_t)\big)\big(\mathbf{I} \bowtie (\mathbf{I}_p \otimes \mathbf{J})^T\big). \end{aligned} \tag{3.186}$$

The Tracy-Singh product ($\bowtie$) is a block Kronecker product introduced by Tracy and Singh (1972), and is defined as follows.

Definition 3.9.1 *(Tracy-Singh product) Suppose the $m \times n$ matrix $\mathbf{A}$ is partitioned into block matrices $\mathbf{A}_{ij}$ with dimensions $m_i \times n_j$, $i = 1, 2, \ldots, r_a$, $j = 1, 2, \ldots, c_a$, and the $p \times q$ matrix $\mathbf{B}$ is partitioned into block matrices $\mathbf{B}_{kl}$ with dimensions $p_k \times q_l$, $k = 1, 2, \ldots, r_b$, $l = 1, 2, \ldots, c_b$, such that $\sum_i m_i = m$, $\sum_j n_j = n$, $\sum_k p_k = p$, and $\sum_l q_l = q$. Then, the Tracy-Singh product $\mathbf{A} \bowtie \mathbf{B}$ is defined as*

$$\mathbf{A} \bowtie \mathbf{B} = (\mathbf{A}_{ij} \bowtie \mathbf{B})_{ij} = \big((\mathbf{A}_{ij} \otimes \mathbf{B}_{kl})_{kl}\big)_{ij} \tag{3.187}$$

where $\mathbf{A}_{ij} \otimes \mathbf{B}_{kl}$ is of order $m_i p_k \times n_j q_l$, $\mathbf{A}_{ij} \bowtie \mathbf{B}$ is of order $m_i p \times n_j q$ and $\mathbf{A} \bowtie \mathbf{B}$ is of order $mp \times nq$.

The second equality in Eq (3.186) follows from the same property as we have for the Kronecker product, that is (see Liu and Trenkler, 2008)

$$(\mathbf{A} \bowtie \mathbf{B})(\mathbf{C} \bowtie \mathbf{D}) = (\mathbf{AC} \bowtie \mathbf{BD}). \tag{3.188}$$

On the other hand, by using the MAR(1) representation of $\mathbf{X}_t$ given in Eq (3.118) and Tracy-Singh multiplying on the right side of Eq (3.118) by $\mathbf{X}_t^T$, we have

$$\begin{aligned}\mathbf{X}_t \bowtie \mathbf{X}_t^T &= \big(\mathbf{A}C(\mathbf{Y}_{t-1})\big) \bowtie \big(C^T(\mathbf{Y}_{t-1})\mathbf{A}^T + \mathbf{u}_t^T\big) + \mathbf{u}_t \bowtie \mathbf{X}_t^T \\ &= \big(\mathbf{A} \bowtie \mathbf{I}\big)\big(C(\mathbf{Y}_{t-1}) \bowtie C^T(\mathbf{Y}_{t-1})\big)\big(\mathbf{I} \bowtie \mathbf{A}^T\big) + \mathbf{A}C(\mathbf{Y}_{t-1}) \bowtie \mathbf{u}_t^T + \mathbf{u}_t \bowtie \mathbf{X}_t^T.\end{aligned} \tag{3.189}$$

The second equality in Eq (3.189) follows from the Tracy-Singh product rule of Eq (3.188).

Example 3.9.1 *Let* $\mathbf{A} = (\mathbf{A}_{ij})$ *and* $\mathbf{B} = (\mathbf{B}_{kl})$ *be two partitioned matrices given by*

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}.$$

Then, according to the Tracy-Singh product defined in Eq (3.187) *we have*

$$\mathbf{A} \bowtie \mathbf{B} = \left[\begin{array}{c:c} \mathbf{A}_{11} \bowtie \mathbf{B} & \mathbf{A}_{12} \bowtie \mathbf{B} \\ \hdashline \mathbf{A}_{21} \bowtie \mathbf{B} & \mathbf{A}_{22} \bowtie \mathbf{B} \end{array}\right] = \left[\begin{array}{cc:cc} \mathbf{A}_{11} \otimes \mathbf{B}_{11} & \mathbf{A}_{11} \otimes \mathbf{B}_{12} & \mathbf{A}_{12} \otimes \mathbf{B}_{11} & \mathbf{A}_{12} \otimes \mathbf{B}_{12} \\ \mathbf{A}_{11} \otimes \mathbf{B}_{21} & \mathbf{A}_{11} \otimes \mathbf{B}_{22} & \mathbf{A}_{12} \otimes \mathbf{B}_{21} & \mathbf{A}_{12} \otimes \mathbf{B}_{22} \\ \hdashline \mathbf{A}_{21} \otimes \mathbf{B}_{11} & \mathbf{A}_{21} \otimes \mathbf{B}_{12} & \mathbf{A}_{22} \otimes \mathbf{B}_{11} & \mathbf{A}_{22} \otimes \mathbf{B}_{12} \\ \mathbf{A}_{21} \otimes \mathbf{B}_{21} & \mathbf{A}_{21} \otimes \mathbf{B}_{22} & \mathbf{A}_{22} \otimes \mathbf{B}_{21} & \mathbf{A}_{22} \otimes \mathbf{B}_{22} \end{array}\right].$$

Let $\mathbf{\Phi}_{\mathbf{X}}(0) = E(\mathbf{X}_t \bowtie \mathbf{X}_t^T)$ and $\mathbf{\Phi}_{C(\mathbf{Y})}(0) = E\big(C(\mathbf{Y}_t) \bowtie C^T(\mathbf{Y}_t)\big)$. Then, by taking the expectation on both sides of Eq (3.186), we have

$$\mathbf{\Phi}_{\mathbf{X}}(0) = \big((\mathbf{I}_p \otimes \mathbf{J}) \bowtie \mathbf{I}\big)\mathbf{\Phi}_{C(\mathbf{Y})}(0)\big(\mathbf{I} \bowtie (\mathbf{I}_p \otimes \mathbf{J})^T\big), \tag{3.190}$$

and, similarly for Eq (3.189), and using the fact that the matrix error terms are independent, we obtain

$$\mathbf{\Phi}_{\mathbf{X}}(0) = \big(\mathbf{A} \bowtie \mathbf{I}\big)\mathbf{\Phi}_{C(\mathbf{Y})}(0)\big(\mathbf{I} \bowtie \mathbf{A}^T\big) + \mathbf{E}_{11}^p \otimes \mathbf{\Psi}_{\boldsymbol{\varepsilon}} \tag{3.191}$$

where $\mathbf{E}_{11}^p$ is similar to that used in Eqs (3.133) and (3.142), and $\mathbf{\Psi}_{\boldsymbol{\varepsilon}}$ is defined in Eq (3.4). The $\mathbf{\Phi}_{C(\mathbf{Y})}(0)$ can be obtained first by combining Eqs (3.190) and (3.191), which leads to

$$\big((\mathbf{I}_p \otimes \mathbf{J}) \bowtie \mathbf{I}\big)\mathbf{\Phi}_{C(\mathbf{Y})}(0)\big(\mathbf{I} \bowtie (\mathbf{I}_p \otimes \mathbf{J})^T\big) = \big(\mathbf{A} \bowtie \mathbf{I}\big)\mathbf{\Phi}_{C(\mathbf{Y})}(0)\big(\mathbf{I} \bowtie \mathbf{A}^T\big) + \mathbf{E}_{11}^p \otimes \mathbf{\Psi}_{\boldsymbol{\varepsilon}}. \tag{3.192}$$

Then, by taking Vec on both sides of Eq (3.192), we have

$$\mathbf{Z}Vec(\mathbf{\Phi}_{C(\mathbf{Y})}(0)) = Vec\big(\mathbf{E}_{11}^p \otimes \mathbf{\Psi}_{\boldsymbol{\varepsilon}}\big) \tag{3.193}$$

where

$$\mathbf{Z} = \Big(\big(\mathbf{I} \bowtie (\mathbf{I}_p \otimes \mathbf{J})^T\big)^T \otimes \big((\mathbf{I}_p \otimes \mathbf{J}) \bowtie \mathbf{I}\big) - \big(\mathbf{I} \bowtie \mathbf{A}^T\big)^T \otimes \big(\mathbf{A} \bowtie \mathbf{I}\big)\Big). \tag{3.194}$$

Multiplying $\mathbf{Z}^T$ and then $(\mathbf{Z}^T\mathbf{Z})^{-1}$ on both sides of Eq (3.193) leads to

$$Vec(\mathbf{\Phi}_{C(\mathbf{Y})}(0)) = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T Vec\big(\mathbf{E}_{11}^p \otimes \mathbf{\Psi}_{\boldsymbol{\varepsilon}}\big). \tag{3.195}$$

The invertibility of the first term on the right side of Eq (3.195) follows from the stationarity of the matrix time series $\mathbf{X}_t$. Eventually, by obtaining the $\mathbf{\Phi}_{C(\mathbf{Y})}(0)$ from Eq (3.195), $\mathbf{\Phi}_{\mathbf{X}}(0)$ can be found by utilizing Eq (3.190).

Note that, by using the definition of the Tracy-Singh product in Eq (3.187) for $\mathbf{X}_t$ and $\mathbf{X}_t^T$ instead of $\mathbf{A}$ and $\mathbf{B}$, respectively, and with respect to the definition of $\mathbf{X}_t$ in Eq (3.114), we have

$$\mathbf{X}_t \bowtie \mathbf{X}_t^T = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p+1} \end{bmatrix} \bowtie \begin{bmatrix} \mathbf{Y}_t^T & \mathbf{Y}_{t-1}^T & \dots & \mathbf{Y}_{t-p+1}^T \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Y}_t \otimes \mathbf{Y}_t^T & \mathbf{Y}_t \otimes \mathbf{Y}_{t-1}^T & \dots & \mathbf{Y}_t \otimes \mathbf{Y}_{t-p+1}^T \\ \mathbf{Y}_{t-1} \otimes \mathbf{Y}_t^T & \mathbf{Y}_{t-1} \otimes \mathbf{Y}_{t-1}^T & & \mathbf{Y}_{t-1} \otimes \mathbf{Y}_{t-p+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}_{t-p+1} \otimes \mathbf{Y}_t^T & \mathbf{Y}_{t-p+1} \otimes \mathbf{Y}_{t-1}^T & \dots & \mathbf{Y}_{t-p+1} \otimes \mathbf{Y}_{t-p+1}^T \end{bmatrix}. \tag{3.196}$$

Therefore, by taking expectations on both sides of Eq (3.196), $\mathbf{\Phi}_{\mathbf{X}}(0)$ can be partitioned by the lag functions $\mathbf{\Psi}(0), \mathbf{\Psi}(1), \dots, \mathbf{\Psi}(p-1)$, as follows

$$\mathbf{\Phi}_{\mathbf{X}}(0) = \begin{pmatrix} \mathbf{\Psi}(0) & \mathbf{\Psi}(1) & \dots & \mathbf{\Psi}(p-1) \\ \mathbf{\Psi}(-1) & \mathbf{\Psi}(0) & \dots & \mathbf{\Psi}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Psi}(-p+1) & \mathbf{\Psi}(-p+2) & \dots & \mathbf{\Psi}(0) \end{pmatrix}. \tag{3.197}$$

Hence, the Yule-Walker equations for the stationary MAR(p) processes are given in Eqs (3.197) and (3.184).

Chapter 4

Matrix Time Series - Estimation

4.1 Introduction

In this chapter, we will estimate the parameters of the matrix autoregressive processes of order p (MAR(p)), proposed in chapter 3, based on a sample of matrix observations $\mathbf{Y}_1$, $\mathbf{Y}_2$, ..., $\mathbf{Y}_N$. There are several different procedures that can be used to estimate time series parameters. All of these methods may have the same or almost the same answer, but may be more or less efficient for any model. We estimate the parameters of the matrix time series based on two main estimation methods, namely, least squares estimation, and maximum likelihood estimation in sections 4.3 and 4.4, respectively. In the least squares estimation method, we consider both ordinary least squares (OLS) estimation, and generalized least squares (GLS) estimation in sections 4.3.2 and 4.3.3, respectively, for the MAR(1) processes. In section 4.3.4, the least squares estimators of parameters of the mean adjusted MAR(1) model will be derived; and in section 4.3.5, the least squares estimators are given for MAR(p), $p > 1$, processes. Finally, in section 4.4, we will use the maximum likelihood method to estimate the parameters of the MAR(p) model. We start with some preliminary material and basic results in section 4.2 that will be used in the rest of this chapter.

4.2 Basic Results

Before starting any analysis to estimate the parameters, let us introduce some important properties and rules of matrix operations, matrix multiplication, and matrix derivatives that will be used through this chapter. For given matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ with appropriate dimensions, we have the following rules and relationships

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC}\otimes\mathbf{BD}) \tag{4.1}$$

$$Vec(\mathbf{AB}) = (\mathbf{B}^T\otimes\mathbf{I})Vec(\mathbf{A}) \tag{4.2}$$

$$(Vec(\mathbf{A}))^T(\mathbf{D}\otimes\mathbf{B})Vec(\mathbf{C}) = tr(\mathbf{A}^T\mathbf{BCD}^T) \tag{4.3}$$

$$tr(\mathbf{AB}) = tr(\mathbf{BA}) \tag{4.4}$$

$$\frac{\partial tr(\mathbf{AB})}{\partial \mathbf{A}} = \mathbf{B}^T \tag{4.5}$$

$$\frac{\partial tr(\mathbf{ABA}^T\mathbf{C})}{\partial \mathbf{A}} = \frac{\partial tr(\mathbf{A}^T\mathbf{CAB})}{\partial \mathbf{A}} = \mathbf{C}^T\mathbf{AB}^T + \mathbf{CAB} \tag{4.6}$$

$$\frac{\partial tr(\mathbf{BA}^{-1}\mathbf{C})}{\partial \mathbf{A}} = -(\mathbf{A}^{-1}\mathbf{CBA}^{-1})^T \tag{4.7}$$

$$\frac{\partial \ln|\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}^T)^{-1} \tag{4.8}$$

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}^T} = (\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}})^T \tag{4.9}$$

$$(\mathbf{A}\otimes\mathbf{B})^{-1} = (\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}) \tag{4.10}$$

$$(\mathbf{A}\otimes\mathbf{B})^T = (\mathbf{A}^T\otimes\mathbf{B}^T) \tag{4.11}$$

$$tr(\mathbf{A}+\mathbf{B}) = tr\mathbf{A} + tr\mathbf{B} \tag{4.12}$$

where T is the transpose operator, $\otimes$ is the Kronecker product, Vec is a matrix operator defined in Eq (3.11), and tr is the trace operation which gives the sum of the diagonal entries of a square matrix. Also, for given matrix $\mathbf{A}$ and vectors $\mathbf{b}$ and $\boldsymbol{\beta}$ with appropriate dimensions, we have

$$\frac{\partial(\mathbf{b}^T\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \frac{\partial(\boldsymbol{\beta}^T\mathbf{b})}{\partial\boldsymbol{\beta}} = \mathbf{b} \tag{4.13}$$

$$\frac{\partial(\mathbf{A}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T} = \mathbf{A}, \qquad \frac{\partial(\boldsymbol{\beta}^T\mathbf{A}^T)}{\partial\boldsymbol{\beta}} = \mathbf{A}^T \tag{4.14}$$

$$\frac{\partial(\boldsymbol{\beta}^T\mathbf{A}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = (\mathbf{A} + \mathbf{A}^T)\boldsymbol{\beta} \tag{4.15}$$

These can be found from any of the many basic texts on matrix algebra, e.g., Seber (2008), Harville (1997).

4.3 Least Squares Estimation

The ordinary least squares (OLS) estimators are derived in section 4.3.2 for the MAR(1) model. In section 4.3.3, we derive the generalized least squares (GLS) estimators for the MAR(1) model. The corresponding estimators for the adjusted MAR(1) model are obtained in section 4.3.4. Finally, in section 4.3.5, we will extend the least squares estimators for the MAR(1) model to MAR(p) processes. We begin, in section 4.3.1, with some basic results.

4.3.1 The Model

We will start with the matrix autoregressive process of order one (MAR(1)). Let $\mathbf{Y}_t$ be a matrix time series of dimension $K \times S$, and assume that it has the same structure as the matrix autoregressive processes of order one that is given in Eq (3.27). That is,

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{A}f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t \tag{4.16}$$

where all terms have their typical meaning, that is, $\boldsymbol{\mu}$ is the intercept matrix allowing for the possibility of a nonzero mean $E[\mathbf{Y}_t]$, $\boldsymbol{\varepsilon}_t$ is the $K \times S$ observational error matrix with properties $E[\boldsymbol{\varepsilon}_t] = 0$, $E[\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t^T] = \boldsymbol{\Psi}$, and $E[\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_r^T] = 0$ for $t \neq r$, which is the white noise or innovation process. The matrix $\mathbf{A}$ is the matrix of coefficient parameters with dimension $K \times KS^2$ given in Eq (3.28), and $f(\mathbf{Y}_{t-1})$ is a matrix function of the matrix $\mathbf{Y}_{t-1}$ with dimension $KS^2 \times S$ defined in Eqs (3.29)-(3.31).

Assume that $\mathbf{Y}_1$, $\mathbf{Y}_2$, …, $\mathbf{Y}_N$ are matrix observations from a stationary matrix time series $\mathbf{Y}_t$ at equal intervals of time. That is, we have a sample of size N for each of the KS variables of the matrix time series $\mathbf{Y}_t$ for the same sample period. Moreover, it is assumed that presample values for the matrix variable $f(\mathbf{Y}_0)$ are available. In order to simplify the notation and put all matrix observations in one compact model, we partition a matrix time series into sample and presample values; that is,

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_N) \quad \text{with dimenstion} \quad (K \times SN) \tag{4.17}$$

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{A}) \quad '' \quad (K \times (KS^2 + S)) \tag{4.18}$$

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{I} \\ f(\mathbf{Y}_t) \end{bmatrix} \quad '' \quad ((KS^2 + S) \times S) \tag{4.19}$$

$$\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_{N-1}) \quad '' \quad ((KS^2 + S) \times SN \tag{4.20}$$

$$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_N) \quad '' \quad (K \times SN) \tag{4.21}$$

$$\mathbf{y} = Vec(\mathbf{Y}) \quad '' \quad (KSN \times 1) \tag{4.22}$$

$$\boldsymbol{\beta} = Vec(\boldsymbol{\theta}) \quad '' \quad ((K^2S^2 + KS) \times 1) \tag{4.23}$$

$$\boldsymbol{\omega} = Vec(\boldsymbol{\varepsilon}) \quad '' \quad (KSN \times 1) \tag{4.24}$$

By using the notations of Eqs (4.17)-(4.21), the MAR(1) model in Eq (4.16) for $t = 1, 2, \ldots, N$, can be written as the following compact linear model

$$\mathbf{Y} = \boldsymbol{\theta}\mathbf{X} + \boldsymbol{\varepsilon}. \tag{4.25}$$

By taking $Vec(.)$ on both sides of Eq (4.25), and using the Vec operation rule given in Eq (4.2), we have

$$
\begin{aligned}
Vec(\mathbf{Y}) &= Vec(\boldsymbol{\theta}\mathbf{X}) + Vec(\boldsymbol{\varepsilon}) \\
&= (\mathbf{X}^T \otimes \mathbf{I}_K)Vec(\boldsymbol{\theta}) + Vec(\boldsymbol{\varepsilon}).
\end{aligned} \tag{4.26}
$$

Therefore, with respect to the notations in Eqs (4.22)-(4.24), Eq (4.26) can be written as

$$
\mathbf{y} = (\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta} + \boldsymbol{\omega}. \tag{4.27}
$$

This model will be used to find the least squares estimator of the matrix parameter $\boldsymbol{\theta}$ by obtaining the least squares estimator for the parameter vector $\boldsymbol{\beta}$.

4.3.2 Ordinary Least Squares (OLS) Estimation

The ordinary least squares (OLS) method is generally applied when estimating the unknown parameters of linear models. The OLS estimators are the best linear unbiased estimators (BLUE) when errors of the linear model are assumed to be uncorrelated. In this section, we use the OLS estimation method to estimate the coefficient matrix $\mathbf{A}$ of the autoregressive matrix process defined in Eq (4.16) and the intercept matrix $\boldsymbol{\mu}$, based on N matrix sample observations. In the compact model defined in Eq (4.25), we put all these parameters together in the matrix $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{A})$ (see Eq (4.18)). Our goal is to estimate the coefficient matrix $\boldsymbol{\theta}$ in the general linear matrix model given in Eq (4.25). To this end, first assume that the matrix error terms $\boldsymbol{\varepsilon}_t$ in Eq (4.16) have an identity covariance matrix, i.e.,

$$
Var(\boldsymbol{\varepsilon}_t) = \sigma^2 \mathbf{I}_{KS}, \qquad t = 1, 2, \ldots, N. \tag{4.28}
$$

The OLS method minimizes the sum of squares of the errors. Therefore, to apply the OLS method, we use the model defined in Eq (4.27) to find the estimator of the vector parameter $\boldsymbol{\beta}$. Then, based on the relationship between $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in Eq (4.18), the estimator of $\boldsymbol{\theta}$ can be obtained. To this end, we need to minimize the sum of squares of the errors of the model. Thence, first let $S_1(\boldsymbol{\beta})$ be the function of the sum of squares of the errors of the model (4.27) as follows

$$\begin{aligned} S_1(\boldsymbol{\beta}) &= \boldsymbol{\omega}^T\boldsymbol{\omega} \\ &= (\mathbf{y} - (\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta})^T(\mathbf{y} - (\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T(\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta} - \boldsymbol{\beta}^T(\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y} + \boldsymbol{\beta}^T(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta}. \end{aligned} \tag{4.29}$$

By taking the derivative on both sides of Eq (4.29) with respect to the vector parameter $\boldsymbol{\beta}$ and using the matrix derivative rules in Eqs (4.13) and (4.15), we have

$$\frac{\partial S_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2(\mathbf{X} \otimes I_K)\mathbf{y} + 2(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta}. \tag{4.30}$$

Then, by setting Eq (4.30) equal to zero, the following normal equation can be obtained

$$(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K)\widehat{\boldsymbol{\beta}} = (\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y}, \tag{4.31}$$

or, equivalently,

$$\widehat{\boldsymbol{\beta}} = [(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K)]^{-1}(\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y}. \tag{4.32}$$

Recall from Eq (4.18) that $\boldsymbol{\beta} = Vec(\boldsymbol{\theta})$. By applying the rules given in Eqs (4.10) and (4.1), Eq (4.32) can be written as

$$\begin{aligned} Vec(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\beta}} &= [(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K)]^{-1}(\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y} \\ &= ((\mathbf{X}\mathbf{X}^T)^{-1} \otimes \mathbf{I}_K)(\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y} \\ &= ((\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X} \otimes \mathbf{I}_K)\mathbf{y}. \end{aligned} \tag{4.33}$$

Then, by using the Vec operation rule in Eq (4.2) and with respect to Eq (4.22) that $\mathbf{y} = Vec(\mathbf{Y})$, we have

$$Vec(\widehat{\boldsymbol{\theta}}) = Vec(\mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}). \tag{4.34}$$

Therefore, the OLS estimator of $\boldsymbol{\theta}$ is equal to

$$\widehat{\boldsymbol{\theta}}^{OLS} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}. \tag{4.35}$$

The Hessian matrix of $S_1(\boldsymbol{\beta})$ is required to be positive definite to ensure that $\widehat{\boldsymbol{\beta}}$ in Eq (4.32) minimizes the function $S_1(\boldsymbol{\beta})$. The Hessian matrix of $S_1(\boldsymbol{\beta})$ is the second derivative of the function $S_1(\boldsymbol{\beta})$, and is given by (see Eq (4.13))

$$\frac{\partial^2 S_1(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = 2(\mathbf{X}\mathbf{X}^T \otimes \mathbf{I}_K). \tag{4.36}$$

This is obviously a positive definite matrix, and confirms that $\widehat{\boldsymbol{\beta}}$ is the minimizer of the sum of squares of the errors function $S_1(\boldsymbol{\beta})$ defined in Eq (4.29).

There is another way to derive the OLS estimator $\widehat{\boldsymbol{\theta}}^{OLS}$ obtained in Eq (4.35). First, consider the matrix autoregressive process as

$$\mathbf{Y}_t = \boldsymbol{\theta}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{4.37}$$

where $\boldsymbol{\theta}$ and $\mathbf{X}_t$ are defined in Eq (4.17). Then, by postmultiplying $\mathbf{X}_{t-1}^T$ on both sides of Eq (4.37), we have

$$\mathbf{Y}_t\mathbf{X}_{t-1}^T = \boldsymbol{\theta}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T + \boldsymbol{\varepsilon}_t\mathbf{X}_{t-1}^T. \tag{4.38}$$

Taking the expectation of both sides of this equation and using the fact that $E[\boldsymbol{\varepsilon}_t] = 0$ yields

$$E[\mathbf{Y}_t\mathbf{X}_{t-1}^T] = \boldsymbol{\theta}E[\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T]. \tag{4.39}$$

Now, $E[\mathbf{Y}_t\mathbf{X}_{t-1}^T]$ and $E[\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T]$ can be estimated, respectively, by

$$E[\widehat{\mathbf{Y}_t\mathbf{X}_{t-1}^T}] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_t\mathbf{X}_{t-1}^T = \frac{1}{N}\mathbf{Y}\mathbf{X}^T, \tag{4.40}$$

$$E[\widehat{\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T}] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T = \frac{1}{N}\mathbf{X}\mathbf{X}^T. \tag{4.41}$$

By substituting these estimated values from Eqs (4.40) and (4.41) into Eq (4.39), the following normal equation will be obtained

$$\frac{1}{N}\mathbf{Y}\mathbf{X}^T = \widehat{\boldsymbol{\theta}}\frac{1}{N}\mathbf{X}\mathbf{X}^T. \tag{4.42}$$

Hence, $\widehat{\boldsymbol{\theta}}^{OLS} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$, as before (in Eq (4.35)).

4.3.3 Generalized Least Squares (GLS) Estimation

The generalized least squares (GLS) method is another procedure to estimate unknown parameters of linear models. The purpose of using the GLS method is to provide an efficient unbiased estimator for the parameters by taking into account the heterogeneous variance of the errors. The OLS estimators in such cases are unbiased and consistent, but not efficient.

Therefore, let $\boldsymbol{\Sigma}_\varepsilon$ be the covariance matrix of the white noise matrix $\boldsymbol{\varepsilon}_t$. Then, the covariance matrix of the $\boldsymbol{\varepsilon}$ defined in Eq (4.17) is equal to

$$Var(\boldsymbol{\varepsilon}) = (\mathbf{I}_N \otimes \boldsymbol{\Sigma}_\varepsilon). \tag{4.43}$$

Furthermore, we assume that the $KS \times KS$ covariance matrix $\boldsymbol{\Sigma}_\varepsilon$ can be broken down and written as the Kronecker product of two covariance matrices $\boldsymbol{\Sigma}_{K\times K}$ and $\boldsymbol{\Omega}_{S\times S}$ with dimension $K \times K$ and $S \times S$, respectively. That is, we can write

$$\boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Omega}_{S\times S} \otimes \boldsymbol{\Sigma}_{K\times K}. \tag{4.44}$$

For the sake of brevity and convenience in notation, in the sequel the two covariance matrices $\boldsymbol{\Omega}_{S\times S}$ and $\boldsymbol{\Sigma}_{K\times K}$ will be shown without the subscripts. Hence, the variance-covariance matrix of $\boldsymbol{\varepsilon}$ given in Eq (4.43) can be rewritten as

$$Var(\boldsymbol{\varepsilon}) = (\mathbf{I}_N \otimes \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}). \tag{4.45}$$

Now, to find the GLS estimator, let $S_2(\boldsymbol{\beta})$ be the squared Mahalanobis distance of the residual vector $\boldsymbol{\omega}$ defined in Eq (4.27). That is,

$$\begin{aligned} S_2(\boldsymbol{\beta}) &= \boldsymbol{\omega}^T(\mathbf{I}_N \otimes \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})^{-1}\boldsymbol{\omega} \\ &= (\mathbf{y} - (\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta})^T(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{y} - (\mathbf{X}^T \otimes \mathbf{I}_K)\boldsymbol{\beta}). \end{aligned} \tag{4.46}$$

Then, applying the rule of Eq (4.2) and using the relationship between the defined variables in Eq (4.17), we have $\boldsymbol{\omega} = Vec(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X})$. Hence, $S_2(\boldsymbol{\beta})$ can be rewritten as

$$S_2(\boldsymbol{\beta}) = (Vec(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X}))^T(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Sigma}^{-1})(Vec(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X})). \tag{4.47}$$

Now, applying the equality given in Eq (4.3) to Eq (4.47), we have

$$\begin{aligned} S_2(\boldsymbol{\theta}) &= (Vec(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X}))^T(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Sigma}^{-1})(Vec(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X})) \\ &= tr[(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\theta}\mathbf{X})(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})^T]. \end{aligned} \tag{4.48}$$

By expanding the quadratic term inside the trace in Eq (4.48), and using the fact that the trace of the sum of matrices is the sum of the trace of each matrix (see Eq (4.12)), $S_2(\boldsymbol{\theta})$ in Eq (4.48) can be simplified as

$$\begin{aligned} S_2(\boldsymbol{\theta}) = tr[\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] - tr[\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] \\ - tr[\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] + tr[\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})]. \end{aligned} \tag{4.49}$$

Note that, in the last term in Eq (4.49) we used $(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})^T = (\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})$; this follows from Eq (4.11) and using the fact that the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are symmetric.

To derive the GLS estimator for the matrix parameter $\boldsymbol{\theta}$, we need to minimize the function $S_2(\boldsymbol{\theta})$ in Eq (4.49). Thus, it is required to take the derivative of the function $S_2(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, and set it to zero. To this end, first for the sake of compatibility of the notations of trace terms in Eq (4.49) with the matrix derivative rules of traces (given in Eqs (4.5) - (4.7)), let us rewrite the function $S_2(\boldsymbol{\theta})$.

The first term of $S_2(\boldsymbol{\theta})$ in Eq (4.49) is not a function of $\boldsymbol{\theta}$, so we can keep it as it is. In the second term of Eq (4.49), let $\mathbf{A} = \mathbf{Y}^T\boldsymbol{\Sigma}^{-1}$, $\mathbf{B} = \boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})$; then based on the rule for the trace of the product of matrices given in Eq (4.4), we have

$$tr(\mathbf{AB}) = tr[\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] = tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}] = tr(\mathbf{BA}). \tag{4.50}$$

Likewise, for the third term of $S_2(\boldsymbol{\theta})$ in Eq (4.49), let $\mathbf{A} = \mathbf{X}^T$, $\mathbf{B} = \boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})$; then by using Eq (4.4), we have

$$tr(\mathbf{AB}) = tr[\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] = tr[\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T] = tr(\mathbf{BA}). \tag{4.51}$$

Similarly, by using the same rule for the fourth term of Eq (4.49), and setting $\mathbf{A} = \mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}$ and $\mathbf{B} = \boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})$, we obtain

$$tr(\mathbf{AB}) = tr[\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] = tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}] = tr(\mathbf{BA}). \quad (4.52)$$

Then, adding Eqs (4.50) - (4.52), the function $S_2(\boldsymbol{\theta})$ can be rewritten as

$$\begin{aligned} S_2(\boldsymbol{\theta}) = {} & tr[\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})] - tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}] \\ & - tr[\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T] + tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}]. \end{aligned} \quad (4.53)$$

Now, by using the differentiation rules of traces given in Eqs (4.5)-(4.7), the derivative of each term in Eq (4.53) with respect to $\boldsymbol{\theta}$ is derived. Based on the rule in Eq (4.5), we have, for the second term,

$$\frac{\partial tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}]}{\partial\boldsymbol{\theta}} = \boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T. \quad (4.54)$$

Similarly, the derivative of the third term in Eq (4.53) with respect to $\boldsymbol{\theta}^T$ is given by

$$\frac{\partial tr[\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T]}{\partial\boldsymbol{\theta}^T} = \mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}; \quad (4.55)$$

then, by using of the derivative rule in Eq (4.9), the term becomes

$$\frac{\partial tr[\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T]}{\partial\boldsymbol{\theta}} = \boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T. \quad (4.56)$$

Finally, taking the derivative of the fourth term, using the rule given in Eq (4.6) by setting $\mathbf{A} = \boldsymbol{\theta}$, $\mathbf{B} = \mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T$, and $\mathbf{C} = \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1^T}$ leads to

$$\frac{\partial tr[\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}]}{\partial\boldsymbol{\theta}} = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T. \quad (4.57)$$

Thence, by combining Eqs (4.54)-(4.57), the first derivative of $S_2(\boldsymbol{\theta})$ in Eq (4.53) with respect to $\boldsymbol{\theta}$ is obtained as

$$\frac{\partial S_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\boldsymbol{\Sigma}^{-1}\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T + 2\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T. \tag{4.58}$$

By setting $\left.\frac{\partial S_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$, we have

$$-2\boldsymbol{\Sigma}^{-1}\Big(\mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T - \widehat{\boldsymbol{\theta}}\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T\Big) = 0. \tag{4.59}$$

Therefore, the GLS estimator $\widehat{\boldsymbol{\theta}}^{GLS}$ of the matrix parameter $\boldsymbol{\theta}$ is equal to

$$\widehat{\boldsymbol{\theta}}^{GLS} = \mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T(\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T)^{-1}. \tag{4.60}$$

Note that if the column (within) covariance $\boldsymbol{\Omega} = \mathbf{I}$, then the $\widehat{\boldsymbol{\theta}}^{GLS}$ of Eq (4.60) reduces to the $\widehat{\boldsymbol{\theta}}^{OLS}$ of Eq (4.35); i.e., the OLS and the GLS estimators are the same, $\widehat{\boldsymbol{\theta}}^{GLS} = \widehat{\boldsymbol{\theta}}^{OLS}$.

4.3.4 Mean Adjusted Least Squares Estimation

In this section, we derive the OLS and GLS estimators for the matrix autoregressive process when the mean of the model is adjusted. In section 4.3.4.1, the mean adjusted model will be introduced. We will obtain the OLS and GLS estimators of the model parameters in section 4.3.4.2 for the known adjusted mean. An estimator for the mean of the adjusted model will be proposed in section 4.3.4.3 when the mean of the model is unknown.

4.3.4.1 The Model

Assume that the matrix autoregressive process in Eq (4.16) has mean $\boldsymbol{\nu}$. Note that, the MAR(1) model in Eq (4.16) has intercept $\boldsymbol{\mu}$; however, the mean of the matrix $\mathbf{Y}_t$ is equal to $\boldsymbol{\nu}$, $E[\mathbf{Y}] = \boldsymbol{\nu}$. Then, the mean-adjusted form of the MAR model is given by

$$\mathbf{Y}_t - \boldsymbol{\nu} = \mathbf{A}(f(\mathbf{Y}_{t-1}) - f(\boldsymbol{\nu})) + \boldsymbol{\varepsilon}_t. \tag{4.61}$$

In order to consider this modification in the least squares approach, some notations defined in Eqs (4.17) - (4.24) are redefined as follows

$\mathbf{Y}^0 = (\mathbf{Y}_1 - \boldsymbol{\nu}, \mathbf{Y}_2 - \boldsymbol{\nu}, \cdots, \mathbf{Y}_N - \boldsymbol{\nu})$	with dimension	$(K \times SN)$	(4.62)
$\mathbf{A} = \mathbf{A}$	$''$	$(K \times (KS^2))$	(4.63)
$\mathbf{X}_t^0 = \left[\ f(\mathbf{Y}_t) - f(\boldsymbol{\nu})\ \right]$	$''$	$((KS^2) \times S)$	(4.64)
$\mathbf{X}^0 = (\mathbf{X}_0^0, \mathbf{X}_1^0, \cdots, \mathbf{X}_{N-1}^0)$	$''$	$((KS^2) \times SN$	(4.65)
$\mathbf{y}^0 = Vec(\mathbf{Y}^0)$	$''$	$(KSN \times 1)$	(4.66)
$\boldsymbol{\alpha} = Vec(\mathbf{A})$	$''$	$((K^2S^2) \times 1)$	(4.67)

Then, the mean-adjusted MAR(1) model defined in Eq (4.61) for $t = 1, 2, \ldots, N$, has the following compact form

$$\mathbf{Y}^0 = \mathbf{A}\mathbf{X}^0 + \boldsymbol{\varepsilon} \tag{4.68}$$

where $\boldsymbol{\varepsilon}$ is the same as in Eq (4.21).

4.3.4.2 Least Squares Estimation with Known Mean

Suppose that the mean of the MAR(1) model is as given in Eq (4.61), and suppose $\boldsymbol{\nu}$ is known. Then, by applying the same process used in section 4.3.1 for the unadjusted model, it can be shown that the OLS estimator of coefficient vector $\boldsymbol{\alpha}$ is analogous to Eq (4.32) and equals

$$\widehat{\boldsymbol{\alpha}}^{OLS} = ((\mathbf{X}^0\mathbf{X}^{0^T} \otimes \mathbf{I}_K)^{-1}\mathbf{X}^0 \otimes \mathbf{I}_K)\mathbf{y}^0. \tag{4.69}$$

Therefore, analogously with Eq (4.35), we have the OLS estimator of the parameter matrix $\mathbf{A}$, $\widehat{\mathbf{A}}^{OLS}$, as

$$\widehat{\mathbf{A}}^{OLS} = \mathbf{Y}^0\mathbf{X}^{0^T}(\mathbf{X}^0\mathbf{X}^{0^T})^{-1}. \tag{4.70}$$

Also, by following the same procedure as in section 4.3.3, it can be shown that the GLS estimator of the coefficient matrix $\mathbf{A}$, $\widehat{\mathbf{A}}^{GLS}$, in this case is equal to

$$\widehat{\mathbf{A}}^{GLS} = \mathbf{Y}^0(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^{0^T}(\mathbf{X}^0(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^{0^T})^{-1}. \tag{4.71}$$

This is analogous to Eq (4.60).

4.3.4.3 Least Squares Estimation with Unknown Mean

In applications, usually the mean $\boldsymbol{\nu}$ of the process is unknown, and it needs to be estimated. The mean matrix $\boldsymbol{\nu}$ of the model in Eq (4.61) can be estimated by the matrix of sample means. That is,

$$\widehat{\boldsymbol{\nu}} = \overline{\mathbf{Y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_t. \tag{4.72}$$

Then, by substituting the $\widehat{\boldsymbol{\nu}}$ into the sample matrix $\mathbf{Y}^0$ in Eq (4.62), the OLS and the GLS estimators of coefficients given in Eqs (4.69)-(4.71) will be modified as follows. First, let $\widehat{\mathbf{Y}}^0$ and $\widehat{\mathbf{X}}_t^0$ be the estimated mean adjusted matrix observation of $\mathbf{Y}^0$ and mean adjusted matrix of past observations $\mathbf{X}_t^0$, given in Eqs 4.62 and 4.64, respectively; i.e.,

$$\widehat{\mathbf{Y}}^0 = (\mathbf{Y}_1 - \widehat{\boldsymbol{\nu}}, \mathbf{Y}_2 - \widehat{\boldsymbol{\nu}}, \cdots, \mathbf{Y}_N - \widehat{\boldsymbol{\nu}}) \qquad \text{with dimension} \quad (K \times SN) \tag{4.73}$$

$$\widehat{\mathbf{X}_t}^0 = \left[\ f(\mathbf{Y}_t) - f(\widehat{\boldsymbol{\nu}})\ \right] \qquad '' \quad ((KS^2) \times S) \tag{4.74}$$

$$\widehat{\mathbf{X}}^0 = (\widehat{\mathbf{X}}_0^0, \widehat{\mathbf{X}}_1^0, \cdots, \widehat{\mathbf{X}}_{N-1}^0) \qquad '' \quad ((KS^2) \times SN \tag{4.75}$$

Then, the OLS and GLS estimators of the coefficient matrix $\mathbf{A}$, $\widehat{\mathbf{A}}^{OLS}$ and $\widehat{\mathbf{A}}^{GLS}$, are analogous to Eqs (4.70) and (4.71), respectively, as

$$\widehat{\mathbf{A}}^{OLS} = \widehat{\mathbf{Y}}^0 \widehat{\mathbf{X}}^{0T} (\widehat{\mathbf{X}}^0 \widehat{\mathbf{X}}^{0T})^{-1} \tag{4.76}$$

$$\widehat{\mathbf{A}}^{GLS} = \widehat{\mathbf{Y}}^0 (\mathbf{I}_N \otimes \mathbf{\Omega}^{-1}) \widehat{\mathbf{X}}^{0T} (\widehat{\mathbf{X}}^0 (\mathbf{I}_N \otimes \mathbf{\Omega}^{-1}) \widehat{\mathbf{X}}^{0T})^{-1}. \tag{4.77}$$

4.3.5 The Least Squares Estimation for MAR(p)

The OLS and GLS estimators for the matrix parameters of the $\boldsymbol{\theta}$ and $\mathbf{A}$ for the MAR(1) model given in sections 4.3.3 and 4.3.4, respectively, can each be extended to the MAR(p) model by making the following modifications. In particular, the matrix parameter $\boldsymbol{\theta}$ and the variable $\mathbf{X}_t$ given in Eqs (4.18) and (4.19), respectively, are modified to

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_p) \qquad \text{with dimension} \qquad (K \times (KS^2p + S)) \tag{4.78}$$

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{I} \\ f(\mathbf{Y}_t) \\ f(\mathbf{Y}_{t-1}) \\ \vdots \\ f(\mathbf{Y}_{t-p+1}) \end{bmatrix} \qquad '' \qquad ((KS^2p + S) \times S) \tag{4.79}$$

$$\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_{N-1}) \qquad '' \qquad ((KS^2p + S) \times SN \tag{4.80}$$

$$\boldsymbol{\beta} = Vec(\boldsymbol{\theta}) \qquad '' \qquad ((K^2S^2p + KS) \times 1) \tag{4.81}$$

Then, the OLS and GLS estimators of the matrix parameter $\boldsymbol{\theta}$ are the same as those given in Eqs (4.35) and (4.60), respectively, except the variable $\mathbf{X}$ will be replaced by the new $\mathbf{X}$ defined in Eq (4.80); i.e.,

$$\widehat{\boldsymbol{\theta}}^{OLS} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} \tag{4.82}$$

$$\widehat{\boldsymbol{\theta}}^{GLS} = \mathbf{Y}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T(\mathbf{X}(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^T)^{-1}. \tag{4.83}$$

Likewise, for finding the least squares estimators for the mean adjusted MAR(p) model, the matrix parameter $\mathbf{A}$ and the matrix variable $\mathbf{X}_t^0$ in Eqs (4.62) and (4.63), respectively, are modified to

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_p) \qquad \text{with dimension} \qquad (K \times (KS^2p)) \tag{4.84}$$

$$\mathbf{X}_t^0 = \begin{bmatrix} f(\mathbf{Y}_t) - f(\boldsymbol{\nu}) \\ f(\mathbf{Y}_{t-1}) - f(\boldsymbol{\nu}) \\ \vdots \\ f(\mathbf{Y}_{t-p+1}) - f(\boldsymbol{\nu}) \end{bmatrix} \qquad '' \qquad ((KS^2p) \times S) \tag{4.85}$$

$$\mathbf{X}^0 = (\mathbf{X}_0^0, \mathbf{X}_1^0, \cdots, \mathbf{X}_{N-1}^0) \qquad '' \qquad ((KS^2p) \times SN \tag{4.86}$$

$$\boldsymbol{\alpha} = Vec(\mathbf{A}) \qquad '' \qquad ((K^2S^2p) \times 1) \tag{4.87}$$

Then, by considering these modifications, the OLS and GLS estimators of $\mathbf{A}$ in Eq (4.84) are the same as those given in Eqs (4.70) and (4.71), respectively; i.e.,

$$\widehat{\mathbf{A}}^{OLS} = \mathbf{Y}^0\mathbf{X}^{0^T}(\mathbf{X}^0\mathbf{X}^{0^T})^{-1} \tag{4.88}$$

$$\widehat{\mathbf{A}}^{GLS} = \mathbf{Y}^0(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^{0^T}(\mathbf{X}^0(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})\mathbf{X}^{0^T})^{-1}. \tag{4.89}$$

4.4 Maximum Likelihood Estimation (MLE)

In contrast to least squares estimation where we minimize the sum of squares of the residuals without any assumptions on the distributions of the error terms, in the maximum likelihood estimation (MLE) method, we need to assume a probability distribution for the residuals. The Gaussian (normal) probability distribution is often considered for the distribution of the residuals to perform MLE. The univariate and multivariate (vector variate) normal distributions are used for univariate and multivariate models, respectively. However, in our study we are dealing with a matrix variate model, which is a matrix time series process. Therefore, we need to consider an appropriate matrix variate probability distribution for the matrix white noise $\boldsymbol{\varepsilon}_t$ of the matrix time series $\mathbf{Y}_t$.

The probability distribution that can be applied to implement the MLE of the matrix time series is the matrix normal distribution. Then, with regards to this assumption on the matrix white noise terms of matrix time series, we build a likelihood function for the MAR models. First, let us define the matrix variate normal distribution briefly. We will write the matrix normal distribution of $\mathbf{X}_{m\times n}$ by $\mathbf{X} \sim N_{m\times n}(\mathbf{M}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})$, and a vector normal distribution of $\mathbf{X}_m$ by $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Definition 4.4.1 *(Gupta and Nagar, 2000) The random matrix $\mathbf{X}_{m\times n}$ has a matrix normal distribution with matrix mean $\mathbf{M}_{m\times n}$ and covariance matrix $\boldsymbol{\Omega}\otimes\boldsymbol{\Sigma}$ if $Vec(\mathbf{X}) \sim N_{mn}\big(Vec(\mathbf{M}), \boldsymbol{\Omega}\otimes\boldsymbol{\Sigma}\big)$, where $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{n\times n}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{m\times m}$ are positive definite matrices, and N_{mn} represents a mn-variate (vector variate) normal distribution.*

Theorem 4.4.1 *(Gupta and Nagar, 2000) The probability density function (p.d.f.) of the matrix normal distribution of $\mathbf{X}_{m\times n}$, $\mathbf{X} \sim N_{m\times n}(\mathbf{M}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})$, where $\mathbf{M}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Sigma}$ have dimension $m \times n$, $n \times n$, and $m \times m$, respectively, is given by*

$$p(\mathbf{X}|\mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}) = (2\pi)^{-nm/2}|\boldsymbol{\Omega}|^{-m/2}|\boldsymbol{\Sigma}|^{-n/2} \exp\{-\tfrac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{M})\boldsymbol{\Omega}^{-1}(\mathbf{X} - \mathbf{M})^T]\},$$
$$\mathbf{X} \in \mathbb{R}^{m\times n}, \mathbf{M} \in \mathbb{R}^{m\times n}, \boldsymbol{\Omega} > 0, \boldsymbol{\Sigma} > 0. \qquad (4.90)$$

Now, assume that the first order matrix autoregressive (MAR(1)) time series given in Eq (4.16) has a matrix normal distribution. In particular, assume that the $K \times S$ matrix white noise process $\boldsymbol{\varepsilon}_t$ of the matrix time series $\mathbf{Y}_t$ in Eq (4.16) has a matrix normal distribution with mean zero and covariance matrix $\mathbf{\Omega} \otimes \mathbf{\Sigma}$, where $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ have dimensions $S \times S$ and $K \times K$, respectively. That is,

$$\boldsymbol{\varepsilon}_t \sim N_{K\times S}(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{\Sigma}). \tag{4.91}$$

Then, by using the fact that $\boldsymbol{\varepsilon}_t$'s are independent, the matrix error $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_N)$ defined in Eq (4.21) has the matrix normal distribution with mean zero and covariance matrix $\mathbf{I}_N \otimes \mathbf{\Omega} \otimes \mathbf{\Sigma}$. It can be written as

$$\boldsymbol{\varepsilon} \sim N_{K\times SN}(0, (\mathbf{I}_N \otimes \mathbf{\Omega}) \otimes \mathbf{\Sigma}). \tag{4.92}$$

Hence, according to Theorem 4.4.1, the p.d.f. of $\boldsymbol{\varepsilon}$ is given by

$$p(\boldsymbol{\varepsilon}|\mathbf{\Omega}, \mathbf{\Sigma}) = (2\pi)^{-KSN/2}|\mathbf{\Omega}|^{-KN/2}|\mathbf{\Sigma}|^{-SN/2} \exp\{-\tfrac{1}{2}tr[\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}(\mathbf{I}_N \otimes \mathbf{\Omega}^{-1})\boldsymbol{\varepsilon}^T]\}. \tag{4.93}$$

In order to determine the MLE of the intercept, $\boldsymbol{\mu}$, of the matrix time series $\mathbf{Y}_t$ separately, the MAR(1) model defined in Eq (4.16) is rewritten as

$$\mathbf{Y}_t - \boldsymbol{\mu} = \mathbf{A}f(\mathbf{Y}_{t-1}) + \boldsymbol{\varepsilon}_t. \tag{4.94}$$

Moreover, in contrast to the least squares estimation in section 4.3, in this section we consider the MAR(p) model rather than MAR(1) model. However, the results of section 4.3 can easily be extended to MAR(p) models as was described in the end of section 4.3.

To this end, consider the MAR(p) model with intercept $\boldsymbol{\mu}$ as

$$\mathbf{Y}_t - \boldsymbol{\mu} = \mathbf{A}_1 f(\mathbf{Y}_{t-1}) + \mathbf{A}_2 f(\mathbf{Y}_{t-2}) + \ldots + \mathbf{A}_p f(\mathbf{Y}_{t-p}) + \boldsymbol{\varepsilon}_t, \tag{4.95}$$

and assume that N matrix sample observations $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N$ are available. Then, in a similar manner used for the least squares method in section 4.3, the notations in Eqs (4.17)-

(4.21) can be redefined for the MAR(p) model given in Eq (4.95) such that we can write all N observed matrix series in one compact linear model. That is,

$$\mathbf{Y} = (\mathbf{Y}_1 - \boldsymbol{\mu}, \mathbf{Y}_2 - \boldsymbol{\mu}, \cdots, \mathbf{Y}_N - \boldsymbol{\mu}) \qquad \text{with dimension} \qquad (K \times SN) \tag{4.96}$$

$$\boldsymbol{\theta} = (\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_p) \qquad '' \qquad (K \times (KS^2p)) \tag{4.97}$$

$$\mathbf{X}_t = \begin{bmatrix} f(\mathbf{Y}_t) \\ f(\mathbf{Y}_{t-1}) \\ \vdots \\ f(\mathbf{Y}_{t-p+1}) \end{bmatrix} \qquad '' \qquad ((KS^2p) \times S) \tag{4.98}$$

$$\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_{N-1}) \qquad '' \qquad ((KS^2p) \times SN \tag{4.99}$$

$$\boldsymbol{\varepsilon} := (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_N) \qquad '' \qquad (K \times SN) \tag{4.100}$$

Therefore, the compact form of the MAR(p) model in Eq (4.95) with N matrix observations can be written as

$$\mathbf{Y} = \boldsymbol{\theta}\mathbf{X} + \boldsymbol{\varepsilon} \tag{4.101}$$

where it is assumed that $\boldsymbol{\varepsilon}$ has the matrix normal distribution given in Eq (4.92). Hence, $\mathbf{Y}$ in Eq (4.101) has matrix normal distribution with mean $\boldsymbol{\theta}\mathbf{X}$ and covariance matrix $(\mathbf{I}_N \otimes \boldsymbol{\Omega}) \otimes \boldsymbol{\Sigma}$. That is,

$$\mathbf{Y} \sim N_{K \times SN}(\boldsymbol{\theta}\mathbf{X}, (\mathbf{I}_N \otimes \boldsymbol{\Omega}) \otimes \boldsymbol{\Sigma}). \tag{4.102}$$

Now, by applying the definition of the matrix normal distribution, we can write the likelihood function of the compact model given in Eq (4.101), or in particular the likelihood function of the MAR(p) model based on the N sample observations as follows,

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}|\mathbf{Y}, \mathbf{X}) &= (2\pi)^{-KSN/2}|\boldsymbol{\Omega}|^{-KN/2}|\boldsymbol{\Sigma}|^{-SN/2}\times \\
&\qquad \exp\{-\tfrac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\theta}\mathbf{X})(\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1})(\mathbf{Y}-\boldsymbol{\theta}\mathbf{X})^T]\} \\
&= (2\pi)^{-KSN/2}|\boldsymbol{\Omega}|^{-KN/2}|\boldsymbol{\Sigma}|^{-SN/2}\times \\
&\qquad \exp\{-\tfrac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))\boldsymbol{\Omega}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))^T]\}.
\end{aligned} \tag{4.103}
$$

Then, by taking the log function on both sides of Eq (4.103), the log-likelihood function of the MAR(p) model is obtained as

$$
\begin{aligned}
\ln L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}|\mathbf{Y}, \mathbf{X}) &= -\frac{1}{2}(KSN)\ \ln(2\pi) - \frac{1}{2}(KN)\ \ln|\boldsymbol{\Omega}| - \frac{1}{2}SN\ \ln|\boldsymbol{\Sigma}| \\
&\quad - \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))\boldsymbol{\Omega}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))^T].
\end{aligned} \tag{4.104}
$$

In order to obtain the MLEs of the parameters, the log-likelihood function should be maximized with respect to the parameters $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Sigma}$. For simplicity and ease of implementation, the quadratic quantity in the last term of the log-likelihood function in Eq (4.104) is expanded, and the log function is rewritten as

$$
\begin{aligned}
\ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X}) = &-\frac{1}{2}(KSN)\ \ln(2\pi) - \frac{1}{2}(KN)\ \ln|\boldsymbol{\Omega}| - \frac{1}{2}SN\ \ln|\boldsymbol{\Sigma}| \\
&- \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] + \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] \\
&+ \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] + \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T] \\
&- \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T] - \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T] \\
&+ \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T] - \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T] \\
&- \frac{1}{2}\sum_{t=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T]. \qquad (4.105)
\end{aligned}
$$

To derive the MLE, we need to set the first partial derivatives of the log-likelihood function to zero and solve them with respect to the parameters. For the sake of compatibility with our notations, and for ease of use of the trace derivative rules given in Eqs (4.5)-(4.7), let us reorder the terms after each trace operator in the log-likelihood function of Eq (4.105) by using the trace operator rule given in Eq (4.4). For instance, to obtain the derivative with respect to $\boldsymbol{\mu}$, by using the rule in Eq (4.4), we can reorder matrices after the trace operator in those terms that have $\boldsymbol{\mu}$. Hence, in the term $tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T]$, let $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ and $\mathbf{B} = \boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T$; then,

$$
tr[\mathbf{AB}] = tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] = tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T\boldsymbol{\Sigma}^{-1}] = tr[\mathbf{BA}]. \qquad (4.106)
$$

Similarly, by using the same rule, in the term $tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T]$, let $\mathbf{A} = \boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}$ and $\mathbf{B} = \boldsymbol{\mu}^T$; in the quantity $tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T]$, set $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ and $\mathbf{B} = \boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T$; in $tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T]$, let $\mathbf{A} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}$ and $\mathbf{B} = \boldsymbol{\mu}^T$; and finally in the term $tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T]$, let $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ and $\mathbf{B} = \boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T]$. Thence, the Eq (4.105) can be rewritten as

$$
\begin{aligned}
\ln L(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}|\mathbf{Y}, \mathbf{X}) = &-\frac{1}{2}(KSN)\ \ln(2\pi) - \frac{1}{2}(KN)\ \ln|\boldsymbol{\Omega}| - \frac{1}{2}SN\ \ln|\boldsymbol{\Sigma}| \\
&- \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] + \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T\boldsymbol{\Sigma}^{-1}] \\
&+ \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T] + \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}] \\
&- \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}] - \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}] \\
&+ \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T] - \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}] \\
&- \frac{1}{2}\sum_{i=1}^{N} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T]. \qquad (4.107)
\end{aligned}
$$

Now, the first partial derivative of each trace function of the log-likelihood function in Eq (4.107) with respect to $\boldsymbol{\mu}$ is derived as follows. By using the derivatives rule given in Eq (4.5), we have

$$\frac{\partial\ tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T\boldsymbol{\Sigma}^{-1}]}{\partial\boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}, \qquad (4.108)$$

$$\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}^T} = \boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T\boldsymbol{\Sigma}^{-1}, \qquad (4.109)$$

$$\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}^T} = \boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}, \qquad (4.110)$$

$$\frac{\partial\ tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}]}{\partial\boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}. \qquad (4.111)$$

Moreover, by using the matrix derivative rule in Eq (4.9) and results obtained in Eqs (4.109) and (4.110), the derivatives of $tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}]$ and $tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}]$ with respect to $\boldsymbol{\mu}$, respectively, are given by

$$\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}} = \Big(\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}^T}\Big)^T = \boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}, \tag{4.112}$$

$$\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}} = \Big(\frac{\partial\ tr[\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}]}{\partial\boldsymbol{\mu}^T}\Big)^T = \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}.\boldsymbol{\Omega}^{-1}. \tag{4.113}$$

Finally, the derivative of $tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}]$ with respect to $\boldsymbol{\mu}$ can be obtained by using the matrix derivative rule in Eq (4.6), as

$$\frac{\partial\ tr[\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}]}{\partial\boldsymbol{\mu}} = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}. \tag{4.114}$$

Thence, the derivative of the log-likelihood function given in Eq (4.107) with respect to $\boldsymbol{\mu}$ is obtained as

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X})}{\partial\boldsymbol{\mu}} &= \sum_{i=1}^{N}(\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1})\\ &= \boldsymbol{\Sigma}^{-1}(\sum_{i=1}^{N}(\mathbf{Y}_t - \boldsymbol{\theta}\mathbf{X}_{t-1}) - N\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}.\end{aligned} \tag{4.115}$$

By setting $\frac{\partial \ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X})}{\partial\boldsymbol{\mu}}\Big|_{\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Omega}},\widehat{\boldsymbol{\Sigma}}} = 0$, we have

$$\widehat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Y}_t - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1}). \tag{4.116}$$

Let $\bar{\mathbf{Y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_t$ and $\bar{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_{t-1}$, then the estimator of the $K \times S$ intercept matrix $\boldsymbol{\mu}$ given in Eq (4.116) is equivalent to

$$\widehat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} - \widehat{\boldsymbol{\theta}}\bar{\mathbf{X}}. \tag{4.117}$$

Accordingly, the derivatives of the log-likelihood function given in Eq (4.107) with respect to $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Omega}$, respectively, are given by

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X})}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{N}(\boldsymbol{\Sigma}^{-1}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T) \\ &= \boldsymbol{\Sigma}^{-1}[\sum_{i=1}^{N}(\mathbf{Y}_t-\boldsymbol{\mu})\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T - \boldsymbol{\theta}\sum_{i=1}^{N}(\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T)],\end{aligned} \tag{4.118}$$

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X})}{\partial \boldsymbol{\Sigma}} &= -\frac{1}{2}SN\ \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}[\sum_{i=1}^{N}(\frac{1}{2}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T - \frac{1}{2}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T \\ &\quad - \frac{1}{2}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T - \frac{1}{2}\mathbf{Y}_t\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T - \frac{1}{2}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{Y}_t^T \\ &\quad + \frac{1}{2}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T + \frac{1}{2}\boldsymbol{\mu}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T + \frac{1}{2}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}^T \\ &\quad + \frac{1}{2}\boldsymbol{\theta}\mathbf{X}_{t-1}\boldsymbol{\Omega}^{-1}\mathbf{X}_{t-1}^T\boldsymbol{\theta}^T)]\boldsymbol{\Sigma}^{-1} \\ &= -\frac{1}{2}NS\ \boldsymbol{\Sigma}^{-1} \\ &\quad + \frac{1}{2}\boldsymbol{\Sigma}^{-1}[\sum_{i=1}^{N}(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\theta}\mathbf{X}_{t-1})\boldsymbol{\Omega}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\theta}\mathbf{X}_{t-1})^T]\boldsymbol{\Sigma}^{-1} \\ &= -\frac{1}{2}[NS\ \boldsymbol{\Sigma} - \sum_{i=1}^{N}(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\theta}\mathbf{X}_{t-1})\boldsymbol{\Omega}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\theta}\mathbf{X}_{t-1})^T],\end{aligned} \tag{4.119}$$

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Omega},\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X})}{\partial \boldsymbol{\Omega}} &= -\frac{1}{2}NK\ \boldsymbol{\Omega}^{-1} \\ &\quad + \frac{1}{2}\boldsymbol{\Omega}^{-1}[\sum_{i=1}^{N}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))]\boldsymbol{\Omega}^{-1} \\ &= -\frac{1}{2}[NK\ \boldsymbol{\Omega} - \sum_{i=1}^{N}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_t-\boldsymbol{\mu}-(\boldsymbol{\theta}\mathbf{X}_{t-1}))].\end{aligned} \tag{4.120}$$

Therefore, by setting to zero the partial derivatives given in Eqs (4.118)-(4.120), the MLEs are obtained as

$$\widehat{\boldsymbol{\theta}} = \Big[\sum_{i=1}^{N}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}})\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{t-1}^T\Big]\Big[\sum_{i=1}^{N}(\mathbf{X}_{t-1}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{t-1}^T)\Big]^{-1}, \tag{4.121}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{NS}\sum_{i=1}^{N}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1})\widehat{\boldsymbol{\Omega}}^{-1}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1})^T, \tag{4.122}$$

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{NK}\sum_{i=1}^{N}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1})^T\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1}). \tag{4.123}$$

By substituting $\widehat{\boldsymbol{\mu}}$ given in Eq (4.117) into Eq (4.121), the estimated coefficient matrix parameters $\boldsymbol{\theta}$ can be simplified as

$$\widehat{\boldsymbol{\theta}} = \Big[\sum_{i=1}^{N}(\mathbf{Y}_t\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{t-1}^T) - \bar{\mathbf{Y}}\widehat{\boldsymbol{\Omega}}^{-1}\bar{\mathbf{X}}^T\Big]\Big[\sum_{i=1}^{N}(\mathbf{X}_{t-1}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{t-1}^T) - \bar{\mathbf{X}}\widehat{\boldsymbol{\Omega}}^{-1}\bar{\mathbf{X}}^T\Big]^{-1}. \tag{4.124}$$

Now, assume that the column or within covariance matrix $\boldsymbol{\Omega}$ of the $K \times S$ matrix white noise process $\boldsymbol{\varepsilon}_t$ in Eq (4.91) is an identity matrix. That is,

$$\boldsymbol{\varepsilon}_t \sim N_{K\times S}(\mathbf{0}, \mathbf{I}_S \otimes \boldsymbol{\Sigma}). \tag{4.125}$$

Then, the maximum likelihood estimators of the parameters $\boldsymbol{\mu}$, $\boldsymbol{\theta}$, and $\boldsymbol{\Sigma}$, respectively, are given by

$$\widehat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} - \widehat{\boldsymbol{\theta}}\bar{\mathbf{X}}. \tag{4.126}$$

$$\widehat{\boldsymbol{\theta}} = \Big[\sum_{i=1}^{N}(\mathbf{Y}_t\mathbf{X}_{t-1}^T) - \bar{\mathbf{Y}}\bar{\mathbf{X}}^T\Big]\Big[\sum_{i=1}^{N}(\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T) - \bar{\mathbf{X}}\bar{\mathbf{X}}^T\Big]^{-1}, \tag{4.127}$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{NS}\sum_{i=1}^{N}(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1})(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\theta}}\mathbf{X}_{t-1})^T. \tag{4.128}$$

Remember that $\boldsymbol{\mu}$ is the intercept of the matrix time series $\mathbf{Y}_t$. Hence, $\widehat{\boldsymbol{\mu}}$ is the MLE estimator of the intercept. For determining the MLE estimator for the mean of the matrix time series $\mathbf{Y}_t$, Eq (3.168) in chapter 3.8 can be used. It is given by

$$\widehat{\boldsymbol{\upsilon}}_X = \widehat{\boldsymbol{\mu}}_X + \frac{\widehat{\mathbf{A}}}{(\mathbf{I}_{KS^2p} - \widehat{\mathbf{B}})}\, C(\widehat{\boldsymbol{\mu}}). \tag{4.129}$$

Chapter 5

Numerical Study

In this chapter, numerical and simulation studies are conducted to compare different matrix autoregressive of order one (MAR(1)) models when they have different coefficient matrices. In fact, like univariate and vector time series, the structure of the autocorrelation functions of MAR models is dependent on the configuration of the coefficient matrices.

Recall from chapter 3 (see Eq (3.23)) that a $K \times S$ matrix time series $\mathbf{Y}_t$ is said to have a matrix autoregressive of order one (MAR(1)) model if it is given by

$$\mathbf{Y}_t = \boldsymbol{\mu} + \sum_{j=1}^{S} \sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1} \mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{5.1}$$

where $\boldsymbol{\mu}$ is the $K \times S$ intercept matrix, $\mathbf{A}_r^j$, $r, j = 1, 2, \ldots, S$, are $K \times S$ coefficient matrices defined in Eq (3.21), $\mathbf{E}_{rj}$, $r, j = 1, 2, \ldots, S$, are the $S \times S$ matrices defined in Eq (3.22), and $\boldsymbol{\varepsilon}_t$ is $K \times S$ matrix error terms, where K is the number of variables (rows of the matrix series $\mathbf{Y}_t$), and S is the number of multiple series (columns of the matrix series $\mathbf{Y}_t$).

In section 5.1, we will show how to simulate an MAR(1) model with length N by considering a matrix normal distribution for the matrix error terms $\boldsymbol{\varepsilon}_t$, $t = 1, 2, \ldots, N$. The process of calculating the autocorrelation function (ACF) of MAR(1) models is briefly described in section 5.2. Finally, the chapter will finish by illustrating a numerical study. The computer codes for this study are in Appendix B.

5.1 Simulation of Matrix Autoregressive Models

Let $\mathbf{Y}_t$ be an MAR(1) model with dimension $K \times S$ as given in Eq (5.1) by

$$\mathbf{Y}_t \;=\; \boldsymbol{\mu} + \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_{t-1}\mathbf{E}_{rj} + \boldsymbol{\varepsilon}_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{5.2}$$

where $\boldsymbol{\varepsilon}_t$ is the white noise matrix that has a matrix normal distribution with mean $\mathbf{0}$, row (within) covariance matrix $\boldsymbol{\Sigma}$, and column (between) covariance matrix $\boldsymbol{\Omega}$ with dimension $K \times K$ and $S \times S$, respectively. That is,

$$\boldsymbol{\varepsilon}_t \sim N_{K\times S}(\mathbf{0}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}). \tag{5.3}$$

Assume that all coefficient matrices $\mathbf{A}_r^j$, the intercept matrix $\boldsymbol{\mu}$, and the covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are known. To simulate MAR models with given parameters, as was discussed in section 3.4.1, first we can generate a matrix of independent standard normal random numbers $\mathbf{Z}_1$; then, by premultiplying and postmultiplying this matrix of independent random numbers by the squared roots of the row and column covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, respectively, we obtain the first matrix observation. That is,

$$\boldsymbol{\varepsilon}_1 = \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{Z}_1 \boldsymbol{\Omega}^{\frac{1}{2}}; \tag{5.4}$$

then, by assuming that $\mathbf{Y}_0 \equiv \mathbf{0}$, we can obtain the first MAR observation as

$$\mathbf{Y}_1 = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_1. \tag{5.5}$$

In other words, the first MAR observation is the sum of the matrix of the intercept and a matrix normal observation. The second MAR observation can be obtained by using the first MAR observation $\mathbf{Y}_1$, and applying Eq (5.2) by adding a new random normal matrix observation. That is,

$$\mathbf{Y}_2 \;=\; \boldsymbol{\mu} + \sum_{j=1}^{S}\sum_{r=1}^{S} \mathbf{A}_r^j \mathbf{Y}_1 \mathbf{E}_{rj} + \boldsymbol{\varepsilon}_2 \tag{5.6}$$

where, now, $\boldsymbol{\varepsilon}_2$ is a new matrix normal observation given by $\boldsymbol{\varepsilon}_2 = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}_2\boldsymbol{\Omega}^{\frac{1}{2}}$. This simulation process can be continued to obtain an MAR(1) time series with length N.

5.2 Autocorrelation function of MAR(1)

In this section, we briefly explain how to find the autocorrelation functions of a MAR(1) model by using the Yule-Walker equations. These autocorrelation functions can be used for further investigations such as in structural diagnostics studies.

Again, assume that all parameters of the MAR(1) model, i.e., the coefficient matrices $\mathbf{A}_r^j$, and the covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are known. First, by using Eq (3.177), we can obtain $\boldsymbol{\Psi}(0)$ as follows

$$\boldsymbol{\Psi}(0) = \sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S} (\mathbf{A}_r^j \otimes \mathbf{E}_{j'r'})\boldsymbol{\Psi}(0)(\mathbf{E}_{rj} \otimes \mathbf{A}_{r'}^{j'T}) + \boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}, \tag{5.7}$$

or equivalently (see Eq (3.179))

$$Vec(\boldsymbol{\Psi}(0)) = \Big(\mathbf{I}_{K^2S^2} - \sum_{j=1}^{S}\sum_{r=1}^{S}\sum_{j'=1}^{S}\sum_{r'=1}^{S} (\mathbf{E}_{jr} \otimes \mathbf{A}_{r'}^{j'}) \otimes (\mathbf{A}_r^j \otimes \mathbf{E}_{j'r'})\Big)^{-1} Vec(\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}}), \tag{5.8}$$

where, from Eq (3.5), we have $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} = \mathbf{T}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. Note that, here, we have $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$; therefore, $\boldsymbol{\Psi}_{\boldsymbol{\varepsilon}} = \mathbf{T}^{-1}(\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})$. Then, by applying Eq (3.174), the lag function $\boldsymbol{\Psi}(h)$ can be obtained for all lags $h = 1, 2, \ldots,$ as

$$\boldsymbol{\Psi}(h) = \sum_{j=1}^{S}\sum_{r=1}^{S} (\mathbf{A}_r^j \otimes \mathbf{I})\boldsymbol{\Psi}(h-1)(\mathbf{E}_{rj} \otimes \mathbf{I}_K). \tag{5.9}$$

Eventually, the autocovariance function of the MAR(1) model at lag h, $\mathbf{\Gamma}(h)$, can be obtained by premultiplying the transformation matrix $\mathbf{T}$ as follows

$$\mathbf{\Gamma}(h) = \mathbf{T}\mathbf{\Psi}(h). \tag{5.10}$$

Alternatively, the autocovariance function $\mathbf{\Gamma}(h)$ can be estimated by using the empirical autocovariance function. Assume that the matrix samples $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N$ and presample matrix observations $\mathbf{Y}_{-p+1}, \mathbf{Y}_{-p+2}, \ldots, \mathbf{Y}_0$ are available. First, the mean of the series, $\boldsymbol{\upsilon}$, can be estimated by

$$\widehat{\boldsymbol{\upsilon}} = \bar{\mathbf{Y}} = \frac{1}{N+p} \sum_{t=-p+1}^{N} \mathbf{Y}_t. \tag{5.11}$$

Then, we can estimate the lag function $\mathbf{\Psi}(h)$ as

$$\widehat{\mathbf{\Psi}}(h) = \frac{1}{N+p-h} \sum_{t=-p+1}^{N-h} (\mathbf{Y}_{t+h} - \bar{\mathbf{Y}}) \otimes (\mathbf{Y}_t - \bar{\mathbf{Y}})^T. \tag{5.12}$$

Eventually, similarly to Eq (5.10), the estimated autocovariance function $\widehat{\mathbf{\Gamma}}(h)$ can be obtained by using the transformation matrix $\mathbf{T}$ as

$$\widehat{\mathbf{\Gamma}}(h) = \mathbf{T}\widehat{\mathbf{\Psi}}(h). \tag{5.13}$$

5.3 Illustration

In this section, we will consider different combinations of eigenvalues for coefficient matrices $\mathbf{A}_r^j$, $r, j = 1, 2, \ldots, S$, in Eq (5.1), and review the structure of the autocorrelation function

(ACF) and partial autocorrelation function (PACF) of matrix autoregressive processes of order one (MAR(1)) for true models. For convenience, as in section 3.4.1, let us consider MAR(1) processes with dimension $K \times S = 3 \times 2$. Then, all coefficient matrices $\mathbf{A}_r^j$'s are $K \times K = 3 \times 3$ matrices, one for each series $j = 1, 2$.

We will study the patterns of the ACF and PACF of stationary MAR(1) models. Recall that an MAR(1) model is stationary if all eigenvalues of the coefficient matrix $\mathbf{B}$ defined in Eq (3.37) have modulus less than one. Also recall from Eq (3.37) that the relationship between the coefficient matrix $\mathbf{B}$ and the coefficient matrices $\mathbf{A}_r^j$, $r, j = 1, 2 \ldots, S$, is given by

$$\mathbf{B} = \sum_{j=1}^{S} \sum_{r=1}^{S} (\mathbf{E}_{jr} \otimes \mathbf{A}_r^j) = \begin{pmatrix} \mathbf{A}_1^1 & \mathbf{A}_2^1 & \ldots & \mathbf{A}_S^1 \\ \mathbf{A}_1^2 & \mathbf{A}_2^2 & \ldots & \mathbf{A}_S^2 \\ \vdots & & & \vdots \\ \mathbf{A}_1^S & \mathbf{A}_2^S & \ldots & \mathbf{A}_S^S \end{pmatrix}. \tag{5.14}$$

Note that, having all eigenvalues of coefficient matrices $\mathbf{A}_r^j$ in modulus less than one does not guarantee that eigenvalues of coefficient matrix $\mathbf{B}$ are in modulus less than one. Therefore, to generate a stationary MAR(1), we choose the eigenvalues of coefficient matrices $\mathbf{A}_r^j$, $r, j = 1, 2 \ldots, S$, such that the eigenvalues of the coefficient matrix $\mathbf{B}$ be less than one in modulus.

To generate a 3×2 stationary MAR(1) model, first assume that all 3×3 coefficient matrices $\mathbf{A}_r^j$'s are known and given. Then, as explained in section 5.1, we can simulate a stationary MAR(1) by using the given coefficient matrices, and the known intercept matrix $\boldsymbol{\mu}$, the given 3×3 row (within) and the given 2×2 column (between) matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, respectively. For the sake of consistency in our comparisons, let us to fix these matrix parameters as follows

$$\boldsymbol{\mu} = \begin{pmatrix} 12.16 & 34.24 \\ -42.37 & -27.69 \\ 4.26 & 36.12 \end{pmatrix}, \ \boldsymbol{\Omega} = \begin{pmatrix} 3.57 & -0.68 \\ -0.68 & 1.45 \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} 1.82 & -0.42 & -0.67 \\ -0.42 & 1.93 & -0.39 \\ -0.67 & -0.39 & 2.57 \end{pmatrix}. \tag{5.15}$$

5.3.1 Negative Eigenvalues

Assume all eigenvalues of all 3×3 coefficient matrices $\mathbf{A}_r^j$'s are values between negative one and zero. In other words, let $\lambda_{rj,1}, \lambda_{rj,2}, \ldots, \lambda_{rj,K}$ be the eigenvalues of $\mathbf{A}_r^j$, $r, j = 1, 2 \ldots, S$; then, assume that for all $i = 1, 2, \ldots, K$, $-1 < \lambda_{rj,i} < 0$. Also assume that the intercept matrix $\boldsymbol{\mu}$, the row and column covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are known and have the same values as in Eq (5.15). We choose randomly numbers between negative one and zero to set eigenvalues $\lambda_{rj,i}$, $r, j = 1, 2$, and $i = 1, 2, 3$. These values are set as follows

$$\begin{aligned} &\lambda_{11,1} = -0.29, \lambda_{11,2} = -0.17, \lambda_{11,3} = -0.01; \quad &\lambda_{12,1} = -0.57, \lambda_{12,2} = -0.56, \lambda_{12,3} = -0.001 \\ &\lambda_{21,1} = -0.18, \lambda_{21,2} = -0.80, \lambda_{21,3} = -0.42; \quad &\lambda_{22,1} = -0.54, \lambda_{22,2} = -0.08, \lambda_{22,3} = -0.55. \end{aligned} \tag{5.16}$$

Now, we need to generate the coefficient matrices $\mathbf{A}_r^j$ with these given sets of eigenvalues. To this end, for each set of given eigenvalues, we first generate the matrix $\mathbf{D}$ with entries of uniform random values between zero and one, then we put the given eigenvalues on the diagonal of $\mathbf{D}$, and zero on the lower off diagonal elements. Therefore, $\mathbf{D}$ is a upper triangular matrix that has given eigenvalues in its diagonal entries. Furthermore, we generate an invertible matrix $\mathbf{S}$ with the same dimension of a given set of eigenvalues by using the "*genPositiveDefMat*" function in the software package R. Eventually, we generate the coefficient matrix $\mathbf{A}_r^j$ such that it has eigenvalues $\lambda_{rj,i}, i = 1, 2, \ldots, K$, by $\mathbf{A}_r^j = \mathbf{SDS}^{-1}$.

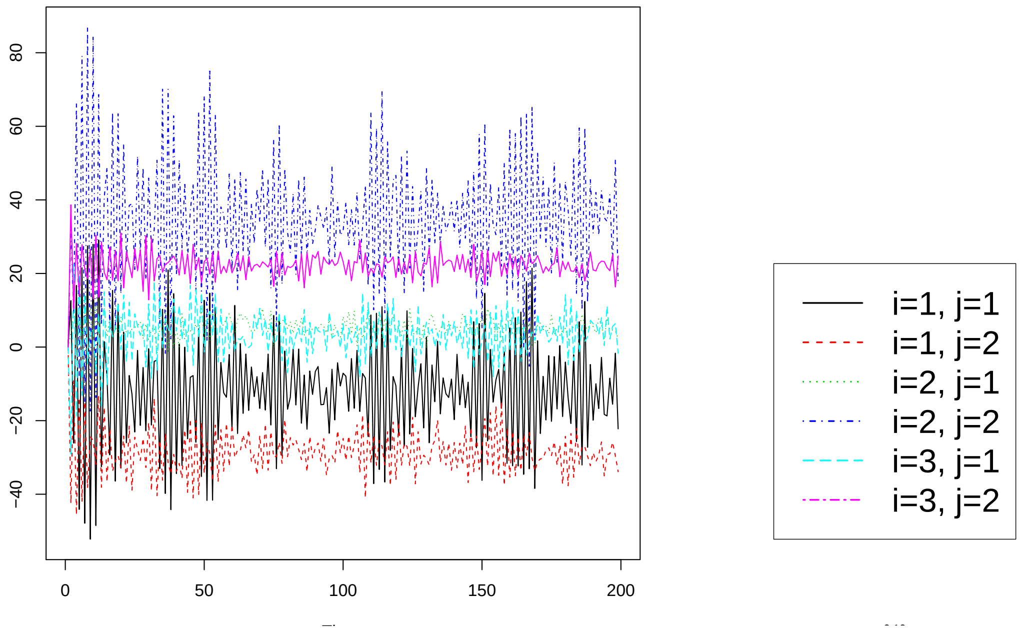

Figure 5.1: A stationary MAR(1) model with negative eigenvalues of coefficient matrices, $K = 3, S = 2$

Their corresponding coefficient matrices $\mathbf{A}_r^j$ are given by

$$\mathbf{A}_1^1 = \begin{pmatrix} -0.20 & 0.60 & 0.70 \\ -0.01 & -0.24 & 0.67 \\ 0.00 & 0.02 & -0.02 \end{pmatrix}, \qquad \mathbf{A}_1^2 = \begin{pmatrix} -0.54 & -0.01 & 0.27 \\ 0.02 & -0.58 & 0.20 \\ 0.06 & -0.05 & -0.01 \end{pmatrix}$$
$$\mathbf{A}_2^1 = \begin{pmatrix} -0.18 & 0.05 & 0.81 \\ 0.02 & -0.81 & 0.13 \\ 0.00 & -0.01 & -0.42 \end{pmatrix}, \qquad \mathbf{A}_2^2 = \begin{pmatrix} -0.64 & 1.07 & 0.61 \\ -0.05 & 0.00 & 0.43 \\ 0.00 & 0.01 & -0.53 \end{pmatrix}. \tag{5.17}$$

Figure 5.1 shows a simulated stationary MAR(1) series with those known parameters in Eq (5.15) and the coefficient matrices $\mathbf{A}_r^j$s given in Eq (5.17). The small box on the right hand of Figure 5.1 identifies the corresponding series of the $K \times S = 3 \times 2$ matrix time series $\mathbf{Y}_t$.

Then, we calculate the cross-ACF for the stationary MAR(1) series plotted in Figure 5.1 by using the Yule-Walker equations described in section 5.2 of this chapter. First, note that, from Eq (5.4), we have $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{\Omega} \otimes \mathbf{\Sigma}$; then, by using the relationship between $\mathbf{\Psi}_{\boldsymbol{\varepsilon}}$ and $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ in Eq (3.5), we can calculate $\mathbf{\Psi}_{\boldsymbol{\varepsilon}}$ as follows

$$\mathbf{\Psi}_{\boldsymbol{\varepsilon}} = \mathbf{T}^{-1}\mathbf{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{T}^{-1}(\mathbf{\Psi} \otimes \mathbf{\Psi}) \tag{5.18}$$

where $\mathbf{T}$ is the transformation matrix defined in Eq (3.6). Therefore, because now all of the parameters of Yule-Walker equations are known, we can calculate the lag functions $\mathbf{\Psi}(h)$, $h = 1, 2, \ldots$. Once we have these lag functions, the autocovariance lag function $\mathbf{\Gamma}(h)$ can be found from Eq (5.13).

This cross-ACF is given as a level plot in Figure 5.2. Note, because $K = 3$ and $S = 2$ we have $K \times S = 6$ series. Therefore, $6 \times 6 = 36$ individual cross-ACFs exist. In other words, for each lag, we have a 6×6 cross-autocorrelation matrix (see Figure 5.2). All 36 individual ACFs are extracted from the level plot in Figure 5.2, and are shown in Figure 5.3 for lags $h = 1, 2, \ldots, 40$.

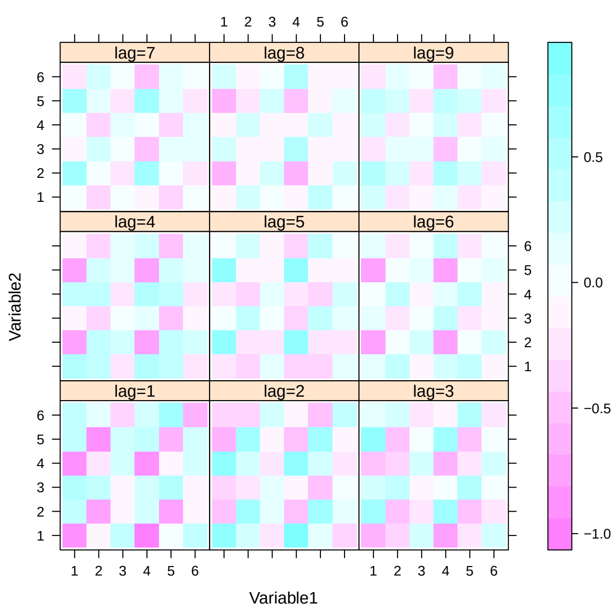

Figure 5.2: The ACF of the MAR(1) series of Figure 5.1 (negative eigenvalues)

Now, for each lag h, if the level plot of the autocorrelation function matrix in Figure 5.2 is split into four equal $K \times K = 3 \times 3$ parts (submatrices), then the two diagonal matrices will represent the autocorrelations of each vector time series (within the variables of each vector series, i.e., for $j = 1, 2 = S$) at lag h, and the offdiagonal matrices will represent the cross-autocorrelations between the vector series $S = 1$ and $S = 2$ at lag h.

Figures 5.3 (a) and (d) show the extracted individual ACFs of the diagonal submatrices after splitting the level plot in Figure 5.2 into four parts. In fact, Figures 5.3 (a) shows the individual ACFs between variables of the vector series $S = 1$, and Figure 5.3 (d) shows the individual ACFs between variables of the vector series $S = 2$.

On the top of each plot in Figures 5.3, there is a pair of numbers that specifies the exact individual ACF of series. These pairs are obtained by $\big(i + (j-1)K, i' + (j'-1)K\big)$, where $i, i' = 1, 2, \ldots, K$, and $j, j' = 1, 2, \ldots, S$. In our example, we have $K = 3$, and $S = 2$. Therefore, in Figure 5.3 (a) where $S = 1$, the pairs on top of each plot are obtained by $(i, i'), i, i' = 1, 2, 3$; and in Figure 5.3 (d), where $S = 2$, the pairs on top of the plots are obtained by $(i + 3, i' + 3), i, i' = 1, 2, 3$.

Similarly, these pairs show the cross-ACFs between series of the vector time series $S = 1$ and the vector time series $S = 2$. In particular, in Figure 5.3 (b) we have $j = 1$, and $j' = 2$; hence these pairs are obtained by $(i, i' + 3), i, i' = 1, 2, 3$; and in Figure 5.3 ($c$) where $j = 2$, and $j' = 1$, they are obtained by $(i + 3, i'), i, i' = 1, 2, 3$.

Note that the plots in Figures 5.3 (a) and (d), when $i = i'$ (diagonal plots) are autocorrelation functions of the individual variables in each of the vector time series $S = 1$ and $S = 2$, respectively. However, plots when $i \neq i'$ (off-diagonal plots) are the cross-autocorrelation functions within the variables in each vector time series. On the other hand, the plots in Figures 5.3 (b) and (c), are cross-autocorrelation functions between the series of the two vector time series $S = 1$ and $S = 2$.

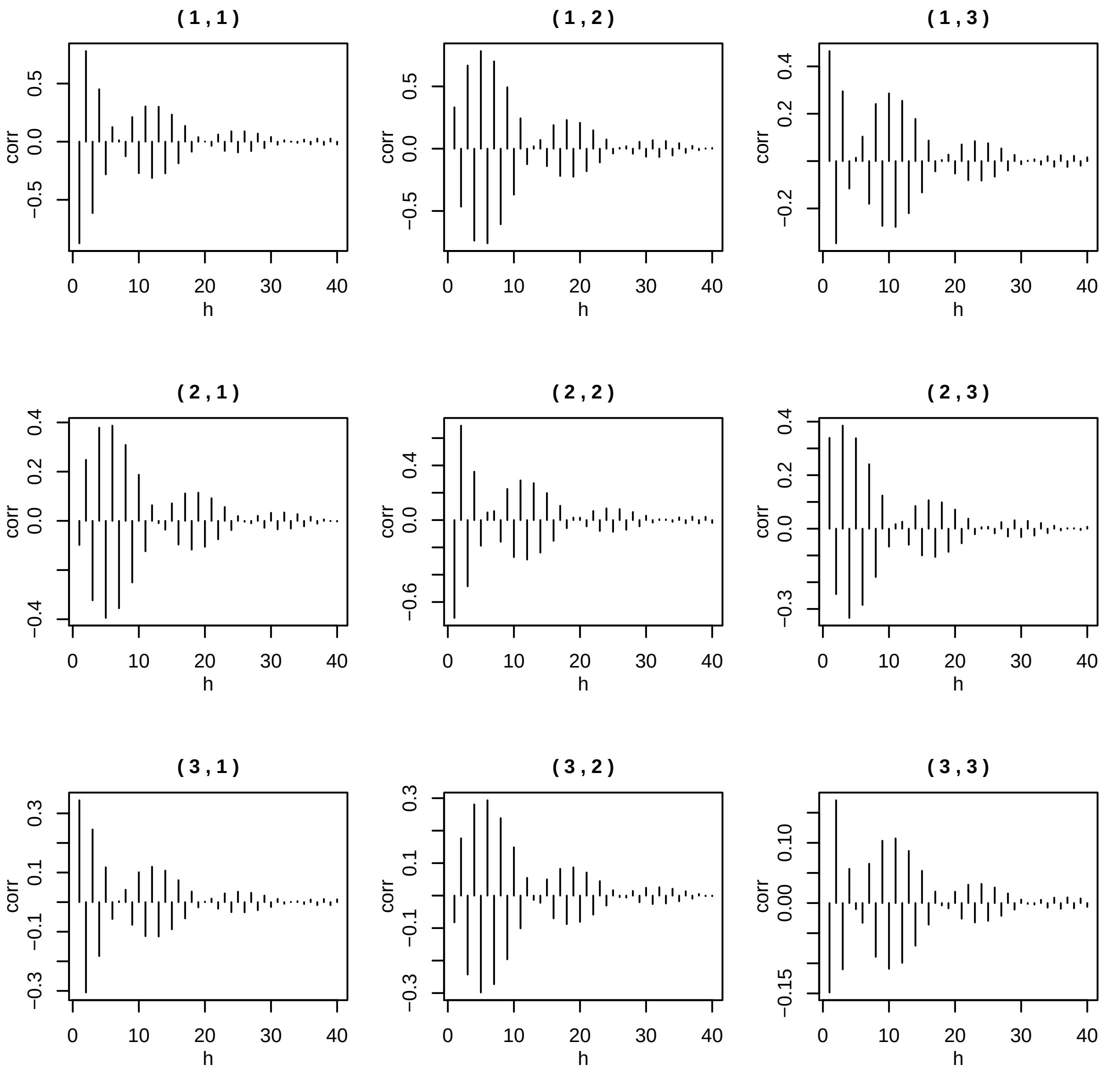

Figure 5.3: (a) Autocorrelations functions for lag $h = 1, 2, \ldots, 40$ for series in Figure 5.1 when $S = 1$ (negative eigenvalues)

5.3.2 Positive Eigenvalues

Now, assume that all eigenvalues of all 3×3 coefficient matrices $\mathbf{A}_r^j$'s have values between zero and one, i.e., $0 < \lambda_{rj,i} < 1$, $i = 1, 2, \ldots, K$. Then, similarly to the case of negative eigenvalues in section 5.3.1, we used the same known intercept matrix $\boldsymbol{\mu}$, and the same known row and column covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, respectively, as given in Eq (5.15). Then, by

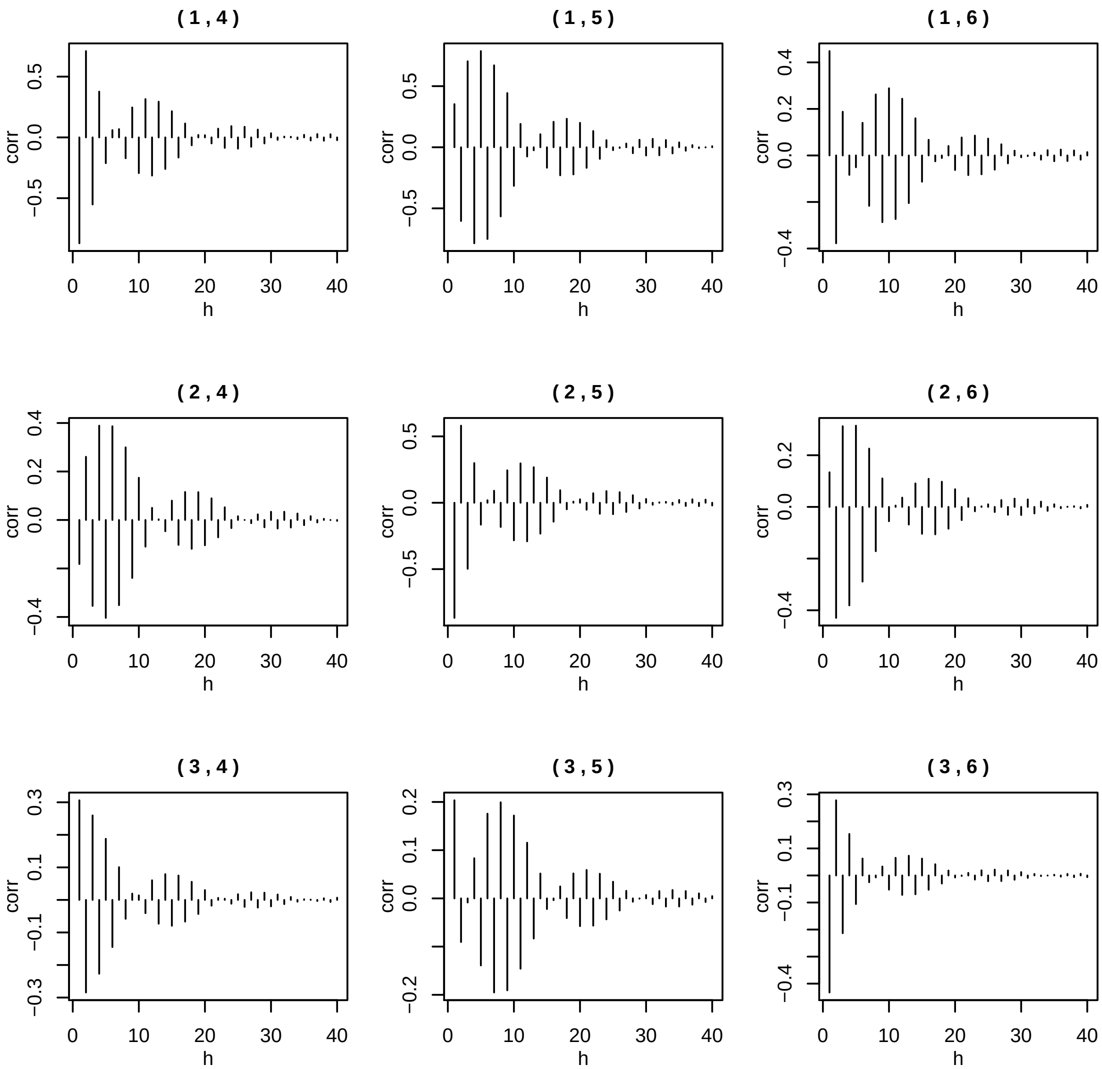

Figure 5.3: (b) Cross-autocorrelations functions between series of $S = 1$ and series of $S = 2$ for lag $h = 1, 2, \ldots, 40$ (negative eigenvalues)

using these parameters, we simulate a stationary MAR(1) series, when the eigenvalues of all coefficient matrices are positive. Figure 5.4 shows a simulated MAR(1) model with

$$\lambda_{11,1} = 0.22, \lambda_{11,2} = 0.26, \lambda_{11,3} = 0.31; \quad \lambda_{12,1} = 0.69, \lambda_{12,2} = 0.30, \lambda_{12,3} = 0.005$$

$$\lambda_{21,1} = 0.41, \lambda_{21,2} = 0.23, \lambda_{21,3} = 0.46; \quad \lambda_{22,1} = 0.09, \lambda_{22,2} = 0.53, \lambda_{22,3} = 0.02. \qquad (5.19)$$

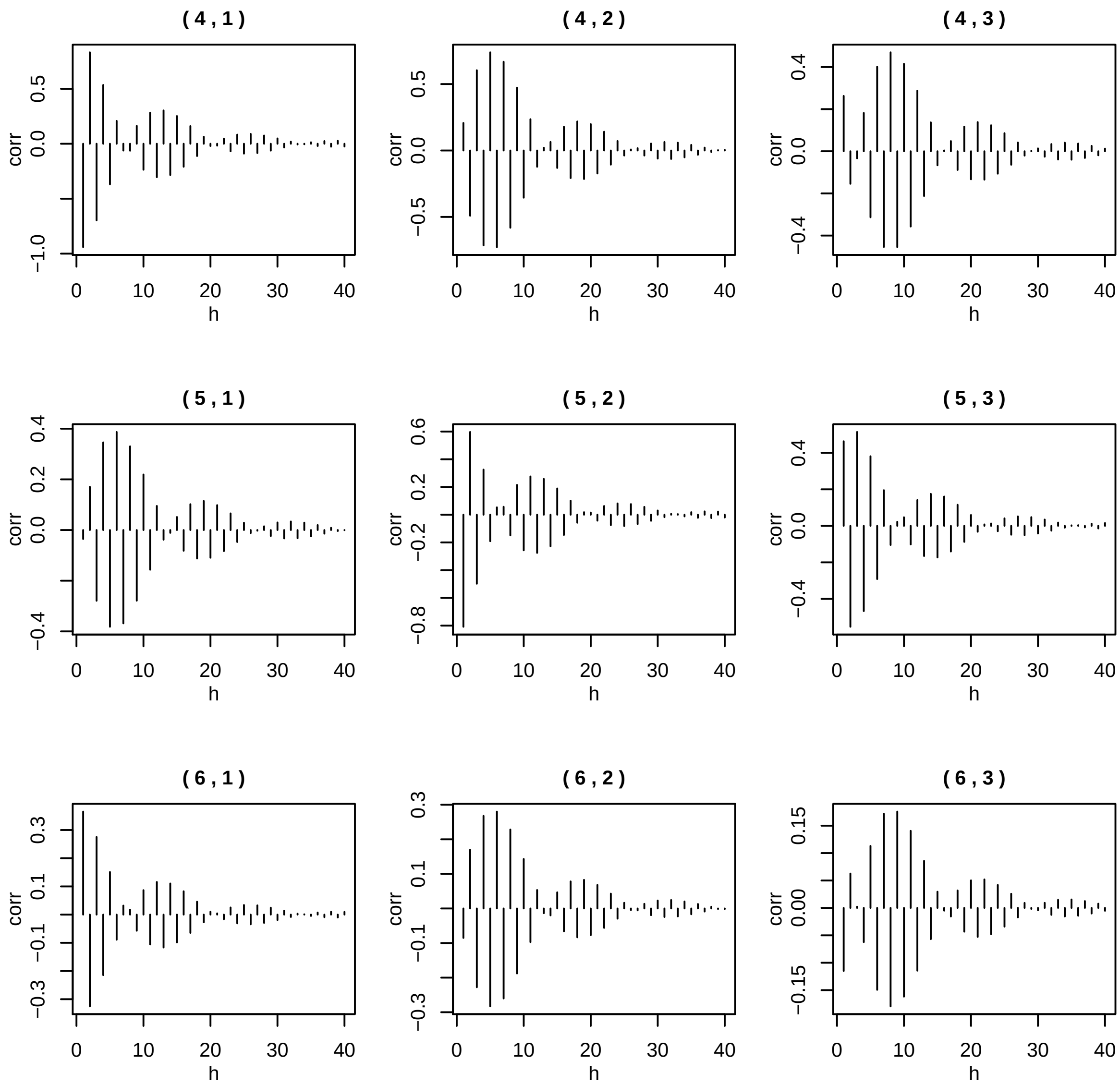

Figure 5.3: (c) Cross-autocorrelations functions between series of $S = 2$ and series of $S = 1$ for lag $h = 1, 2, \dots, 40$ (negative eigenvalues)

The coefficient matrices $\mathbf{A}_r^j$ with these given eigenvalues are generated similarly as for those in the negative eigenvalue case by $\mathbf{A}_r^j = \mathbf{SDS}^{-1}$, where $\mathbf{D}$ is an arbitrary upper triangular matrix with given eigenvalues set on diagonal entries, and $\mathbf{S}$ is an invertible matrix generated by using the "*genPositiveDefMat*" function in the software package R. We choose uniform random values between zero and one for upper off diagonal entries of $\mathbf{D}$.

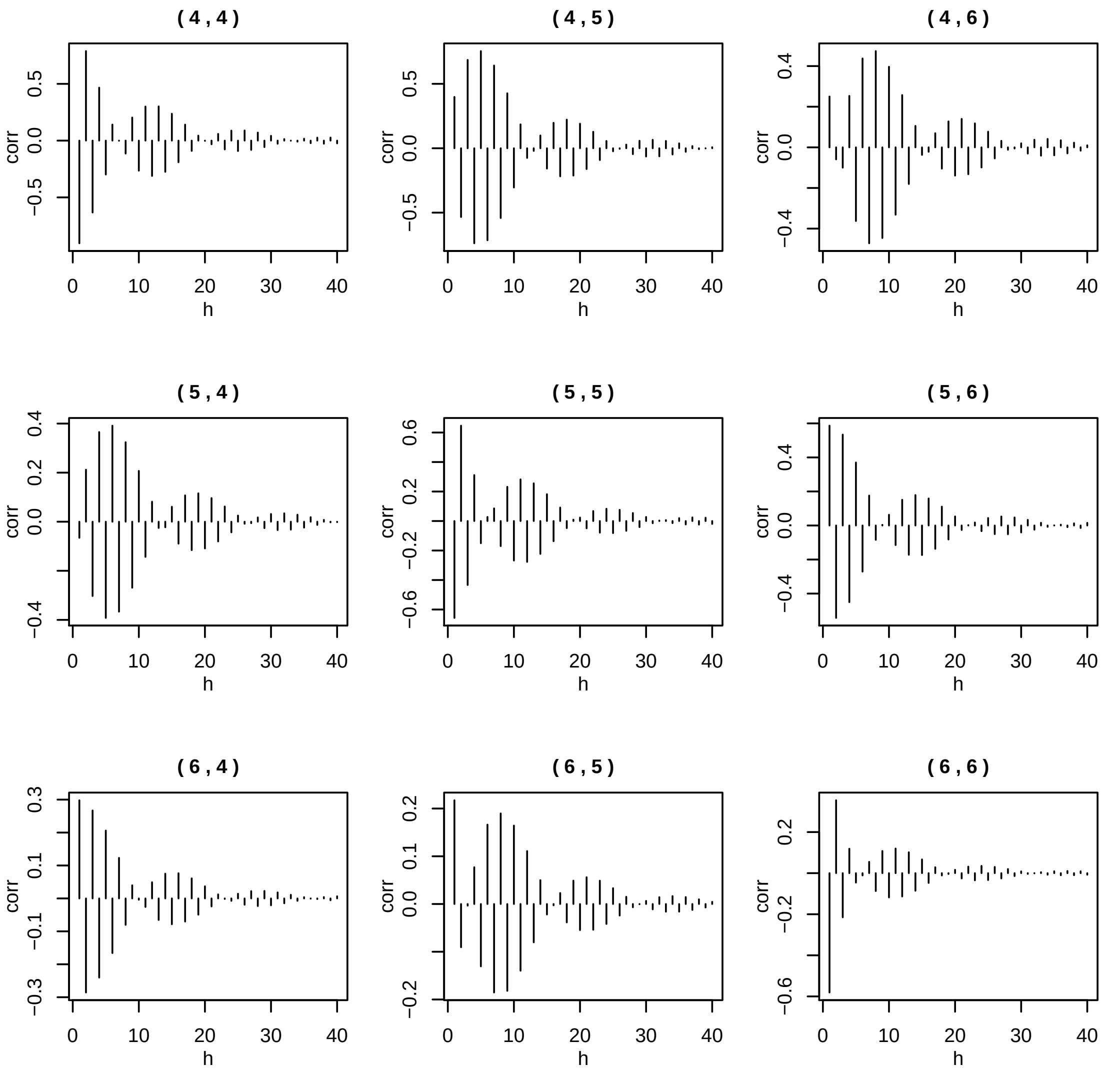

Figure 5.3: (d) Autocorrelations functions for lag $h = 1, 2, \ldots, 40$ for series when $S = 2$ (negative eigenvalues)

The corresponding coefficient matrices $\mathbf{A}_r^j$ of the given eigenvalues in Eq 5.19 are obtained as follows

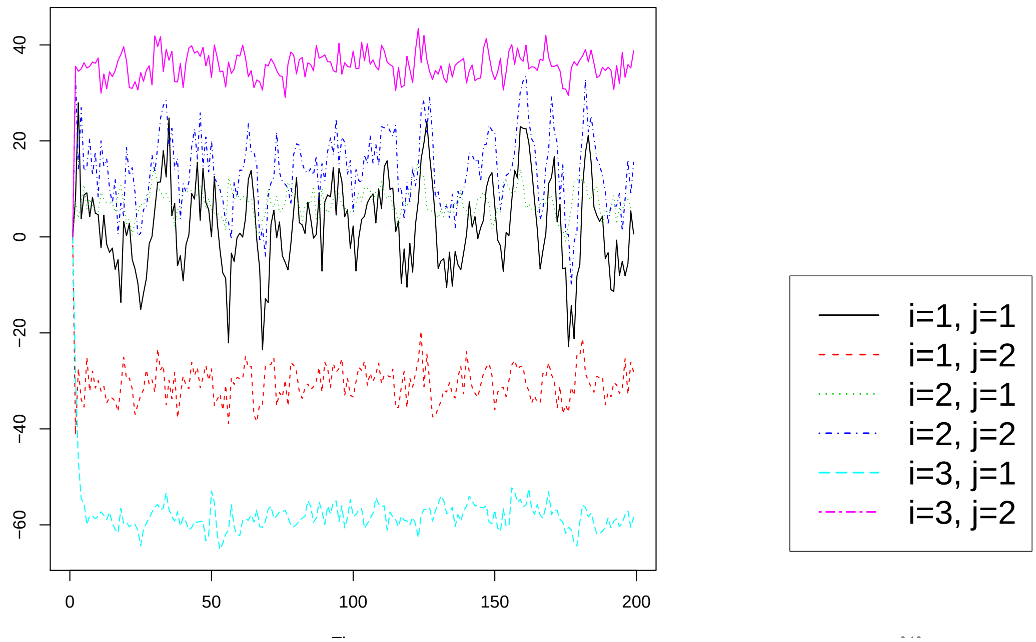

Figure 5.4: A stationary MAR(1) model with positive eigenvalues of coefficient matrices, $K = 3, S = 2$

$$\mathbf{A}_1^1 = \begin{pmatrix} 0.09 & 0.38 & 0.73 \\ -0.16 & 0.11 & 0.37 \\ -0.08 & 0.12 & 0.58 \end{pmatrix}, \qquad \mathbf{A}_1^2 = \begin{pmatrix} 0.68 & 0.61 & 0.63 \\ 0.05 & 0.21 & 0.69 \\ -0.02 & 0.01 & 0.11 \end{pmatrix}$$
$$\mathbf{A}_2^1 = \begin{pmatrix} 0.43 & 0.82 & 0.38 \\ -0.02 & 0.17 & 0.02 \\ 0.00 & 0.09 & 0.51 \end{pmatrix}, \qquad \mathbf{A}_2^2 = \begin{pmatrix} 0.07 & 0.15 & 0.24 \\ 0.01 & 0.51 & 0.11 \\ 0.01 & 0.07 & 0.07 \end{pmatrix}. \tag{5.20}$$

As in the negative eigenvalues case, we obtain the cross-ACF of the MAR(1) series in Figure 5.4 by using the Yule-Walker equations, and the cross-ACF is shown as a level plot in Figure 5.5.

Eventually, we extracted all 36 individual ACFs from the level plot in Figure 5.5, and they are shown in Figures 5.6. Figures 5.6 (a) and (d) show the individual AFCs for within

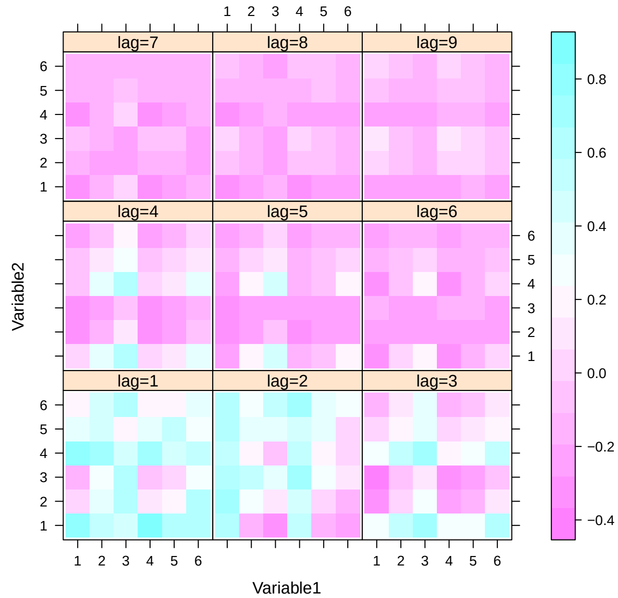

Figure 5.5: The ACF of the MAR(1) series of Figure 5.4 (positive eigenvalues)

variables of the vector series $S = 1$ and $S = 2$, respectively. On the other hand, they are the extracted individual ACFs of the diagonal autocorrelation matrices of the split level plot in Figure 5.5. Split level plot means, similarly to that the negative eigenvalues case, the autocorrelation matrix of the level plot at lag h is split into four parts (submatrices). Figures 5.6 (b) and (c) show the the offdiagonal cross-autocorrelation matrices of the split level plot in Figure 5.5 between the vector series $S = 1$ and $S = 2$.

Similarly to the negative eigenvalues case, the pairs on top of the plots are representing the individual series. They are obtained by using $\big(i + (j-1)K, i' + (j'-1)K\big)$ where $i, i' = 1, 2, \ldots, K$, and $j, j' = 1, 2, \ldots, S$. For example, in Figure 5.6 (a), $\big(i + (j-1)K, i' + (j'-1)K\big) = (1, 1)$ (top left plot) corresponds to $j = j' = 1$ and $i = i' = 1$, and $\big(i + (j-1)K, i' + (j'-1)K\big) = (1, 3)$ (top right plot) corresponds to $j = j' = 1$ and $i = 1, i' = 3$. Similarly, in Figure 5.6 (b), $\big(i + (j-1)K, i' + (j'-1)K\big) = (3, 4)$ (bottom left plot) corresponds to $j = 1$, $j' = 2$, and $i = 3, i' = 1$.

Note that, analogous to the negative eigenvalues case, the plots in Figures 5.6 (a) and (d), when $i = i'$ (diagonal plots) are autocorrelation functions of the individual variables in each of the vector time series $S = 1$ and $S = 2$, respectively. However, plots when $i \neq i'$ (off-diagonal plots) are the cross-autocorrelation functions within the variables in each vector time series. Furthermore, the plots in Figures 5.6 (b) and (c), are cross-autocorrelation functions between the series of the two vector time series $S = 1$ and $S = 2$.

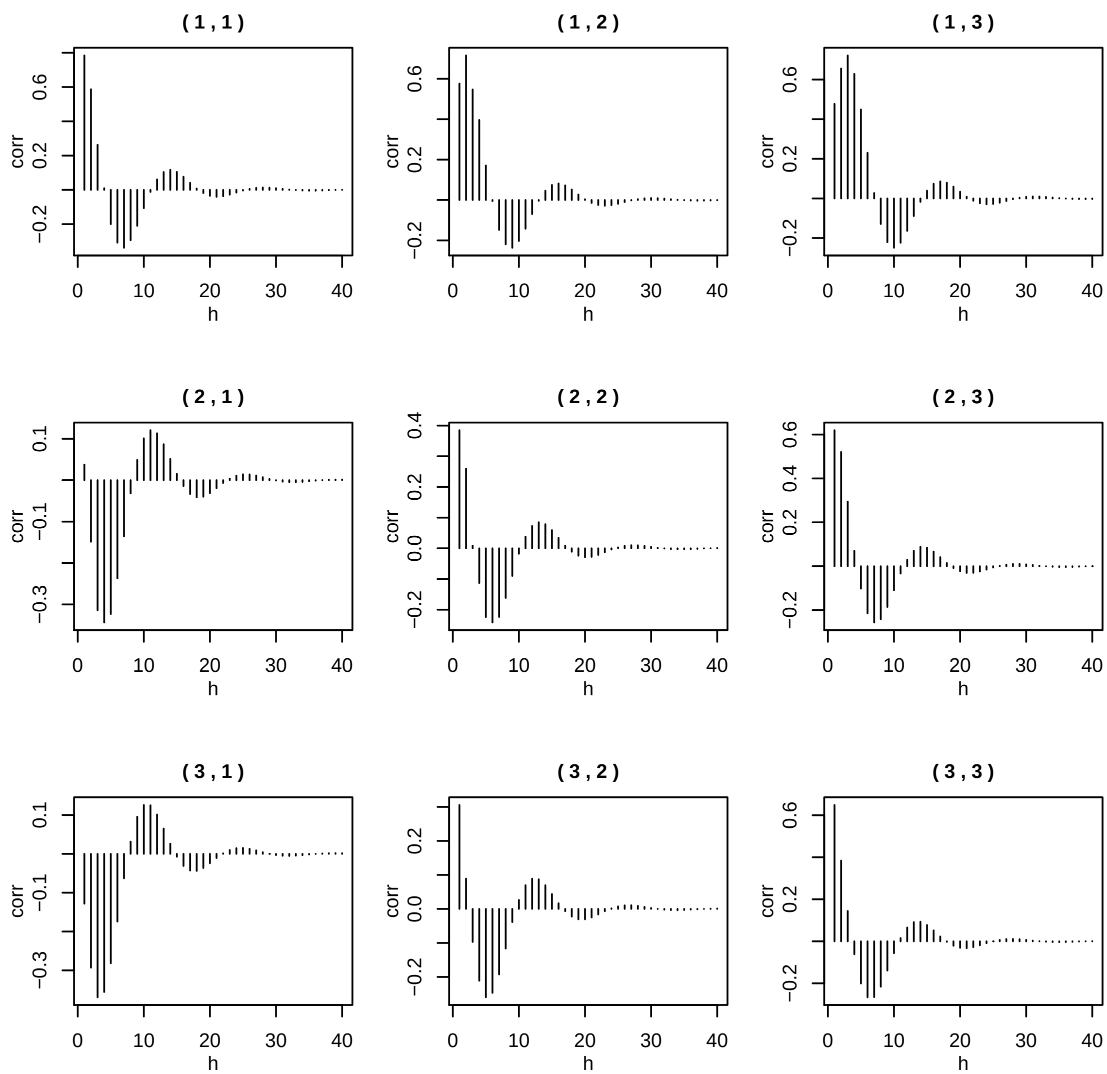

Figure 5.6: (a) Cross-autocorrelations functions for lag $h = 1, 2, \ldots, 40$ for series in Figure 5.4 when $S = 1$ (positive eigenvalues)

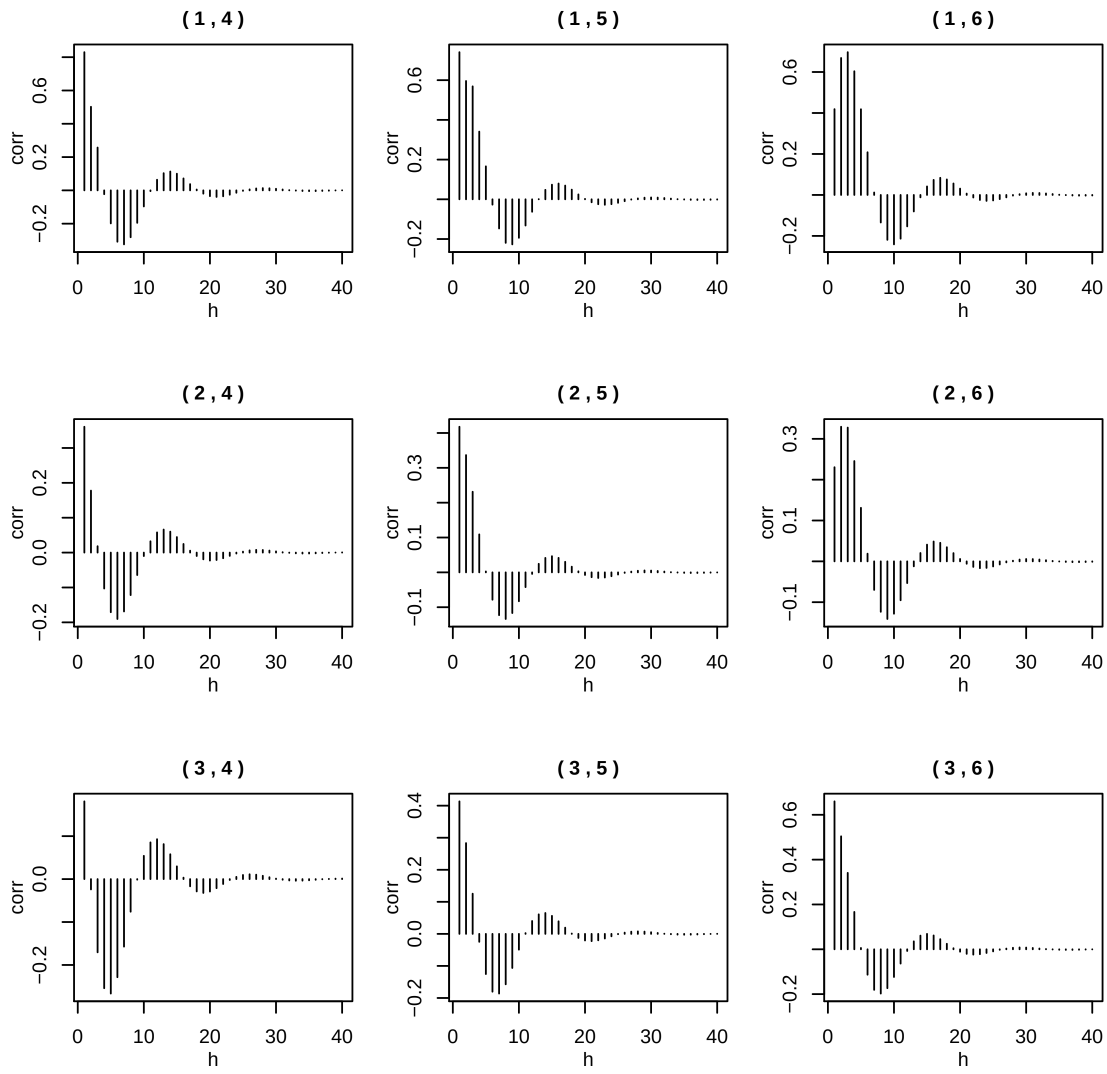

Figure 5.6: (b) Cross-autocorrelations functions between series of $S = 1$ and series of $S = 2$ for lag $h = 1, 2, \ldots, 40$ (positive eigenvalues)

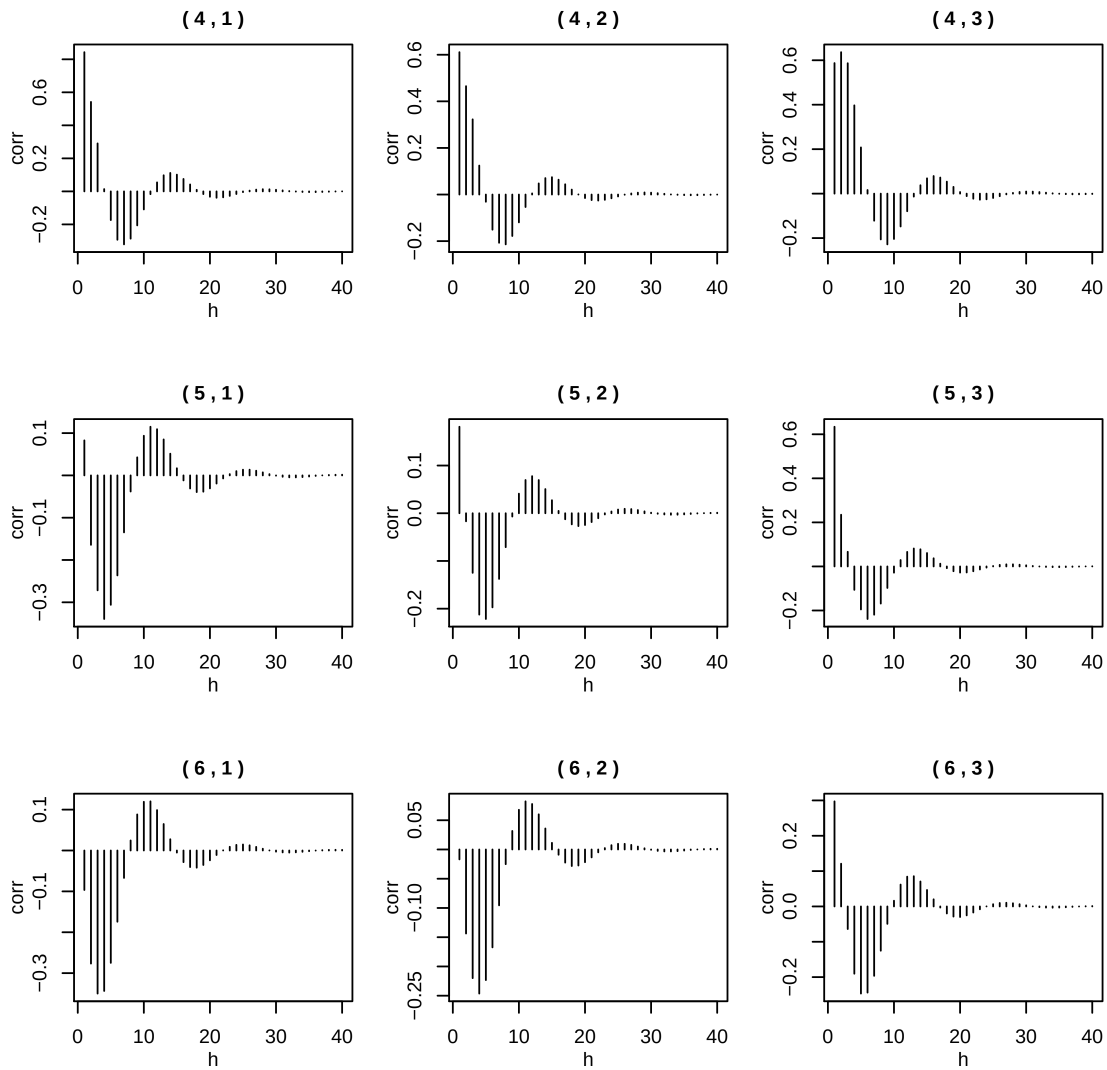

Figure 5.6: (c) Cross-autocorrelations functions between series of $S = 2$ and series of $S = 1$ for lag $h = 1, 2, \ldots, 40$ (positive eigenvalues)

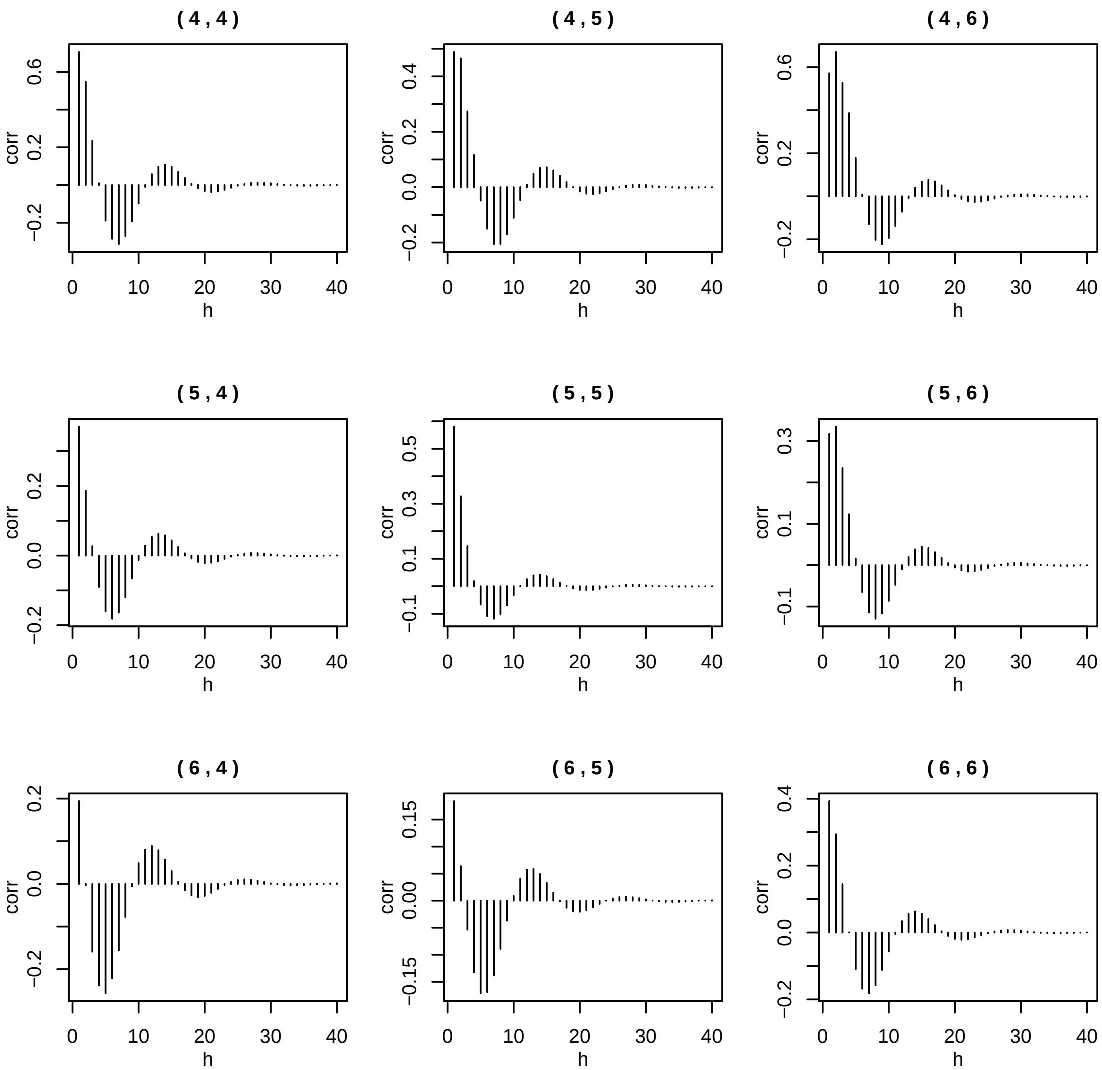

Figure 5.6: (d) Autocorrelations functions for lag $h = 1, 2, \ldots, 40$ for series when $S = 2$ (positive eigenvalues)

Chapter 6

Summary and Future Works

In this dissertation, we introduced a class of matrix time series models called matrix autoregressive (MAR) models for dealing with a new feature of time dependent data (matrix time series data). These types of data, essentially have two main components which can be considered as rows and columns of matrix observations (sometimes called two-way or transposable data). Moreover, they usually are collected over time; hence, they are time dependent, and therefore they constitute a matrix time series data set. If *"time"* be considered as a third component, then in the literature sometimes they are referred to as three-way data sets.

After introducing a matrix autoregressive model of order one and p for dealing with the situation where there are multiple sets of multivariate time series data, their infinite order moving average analogues are obtained, and this moving average representation is used to derive explicit expressions of cross-autocovariance and cross-autocorrelation functions of the MAR models. Stationarity conditions are also provided. We estimate the parameters of the proposed matrix time series models by ordinary and generalized least squares methods, and the maximum likelihood estimation method by considering the matrix normal distribution for the matrix error terms. By applying matrix time series models, the number of parameters to be estimated is dramatically decreased by increasing the dimension of the matrix time series data relative to thinking about the problem in a traditional way (columns-fold vector time series).

This is the beginning of the development of modeling and analysis of matrix time series data, and much exciting and interesting work remains to be done. Some of these important future work can be listed as follows:

- Test for determining the matrix autoregressive order and checking the model adequacy;
- Testing of normality of a matrix white noise process;
- Studying to find any possible relationship between the stationarity of some individual vector time series, $Y_{.jt}, j = 1, 2, \ldots, S$, and the stationarity of the matrix time series $\mathbf{Y}_t$;
- Studying to see if there is a connection between the cointegrated vector time series, $Y_{.jt}, j = 1, 2, \ldots, S$, and the stationary matrix time series $\mathbf{Y}_t = (Y_{.1t}, Y_{.2t}, \ldots, Y_{.St})$;
- Introducing finite matrix moving average (MMA), and matrix autoregressive moving average (MARMA) processes (mixed model);
- Estimation of MARMA models;
- Specification and checking the adequacy of MARMA models;
- Introducing matrix state space models;
- etc.

Bibliography

Allen, G. I. (2010): *Transposable Regularized Covariance Models with Applications to High-Dimensional Data*, Doctorial thesis, Stanford University.

Allen, G. I. and Tibshirani, R. (2012): Inference with transposable data: modelling the effects of row and column correlations, *Journal of the Royal Statistical Society*, B, 74, 4, 721-743.

Antognini, J. F., Buonocore, M. H., Disbrow, E. A. and Carstens, E. (1997): Isoflurane anesthesia blunts cerebral responses to noxious and innocuous stimuli: A fMRI study, *Life Sciences*, 61, 349-354.

Bartlett, M. S. (1946): On the theoretical specification of sampling properties of autocorrelated time series, *Journal of the Royal Statistical Society Supplement*, 8, 27-41.

Box, G. E. P. and Jenkins, G. M. (1970): *Time Series Analysis: Forecasting and Control*, San Francisco, Holden-Day.

Box, G. E. P. and Jenkins, G. M. (1976): *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994): *Time Series Analysis: Forecasting and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice Hall.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008): *Time Series Analysis, Forecasting and Control* (4th ed.), New York, Wiley.

Brillinger, D. R. (1975): *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart and Winston.

Brockwell, P. J. and Davis, R. A. (1991): *Time Series: Theory and Methods* (2nd ed.), New York: Springer.

Brockwell, P. J. and Davis, R. A. (2002): *Introduction to Time Series and Forecasting* (2nd ed.), New York: Springer.

Carvalho, C. M. and West, M. (2007a): Dynamic matrix-variate graphical models, *Bayesian Analysis*, 2, 69-98.

Carvalho, C. M. and West, M. (2007b): Dynamic matrix-variate graphical models - A synopsis. *In Bayesian Statistics VIII* (eds. J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith and M. West), Oxford University Press.

Champernowne, D. G. (1948): Sampling theory applied to autoregressive sequences, *Journal of the Royal Statistical Society*, B, 10, 204-231.

Cochrane, D. and G. H. Orcutt (1949): Applications of least squares regression to relationships containing autocorrelated errors, *Journal of the American Statistical Association*, 44, 32-61.

Dawid, A. P. (1981): Some matrix-variate distribution theory: Notational considerations and a Bayesian application, *Biometrika*, 68, 1, 265-274.

De Gooijer, J. G. and Hyndman R. J. (2006): 25 years of time series forecasting, *International Journal of Forecasting*, 22, 443-473.

Fox, E. B. and West, M. (2011): *Autoregressive Models for Variance Matrices: Stationary Inverse Wishart Processes*, arXiv:1107.5239.

Gupta, A. K. and Nagar, D. K. (2000): *Matrix Variate Distributions*, London, Chapman Hall/CRC.

Hannan, E. J. (1970): *Multiple Time Series*, New York, Wiley.

Harville, D. A. (1997): *Matrix Algebra from a Statisticians Perspective*, New York, Springer-Verlag.

Hoff, P. D. (2011): Separable covariance arrays via the Tucker product, with applications to multivariate relational data, *Bayesian Analysis*, 6, 2, 179-196.

Kadous, M. W. (1995): *Recognition of Australian Sign Language Using Instrumented Gloves*, Thesis University of New South Wales, 136 pages.

Kadous, M. W. (1999): Learning comprehensible descriptions and multivariate time series, In: *Proceedings of the Sixteenth International Conference on Machine Learning* (eds. I. Bratko and S. Dzeroski), Morgan Kaufmann Publishers, San Fransisco, 454-463.

Lütkepohl, H. (1991): *Introduction to Multiple Time Series Analysis*, Springer, Berlin.

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

Mann, H. B. and Wald, A. (1943): On the statistical treatment of linear stochastic difference equations, *Econometrica*, 11, 173-220.

Moran, P. A. P. (1947): Some theorems on time series, *Biometrika*, 34, 281-291.

Osborn, D. R. (1977): Exact and approximate maximum likelihood estimators for vector moving average processes, *Journal of the Royal Statistical Society*, B, 39, 114-118.

Parzen, E. (1969): Multiple time series modeling, In: *Multivariate Analysis-II* (ed. P. Krishnaiah), Academic Press, New York, 389-409.

Pourahmadi, M. (2001): *Foundations of Time Series Analysis and Prediction Theory*, Wiley, New York.

Priestley, M. B. (1981): *Spectral Analysis and Time Series*, Volumes 1 and 2, London, Academic Press.

Quenouille, M. H. (1957): *The Analysis of Multiple Time-Series* (2nd ed. 1968), London, Griffin.

Quintana, J. M. and West, M. (1987): Multivariate time series analysis: New techniques applied to international exchange rate data, *The Statistician*, 36, 275-281.

Roy, S. N. (1957): *Some Aspects of Multivariate Analysis*, New York, Wiley.

Seber, G. A. F. (2008): *A Matrix Handbook for Statisticians*, Hoboken, Wiley.

Shumway, R. H. (1988): *Applied Statistical Time Series Analysis*, Englewood Cliffs, Prentice Hall.

Shumway, R. H. and Stoffer, D. S. (2000): *Time Series Analysis and Its Applications*, Springer, New York.

Slutsky, E. (1937): The summation of random causes as the source of cyclic processes, *Econometrica*, 5, 105-146 (originally published in Russian in 1927).

Tiao, G. C. and Box, G. E. P. (1981): Modeling multiple time series with applications, *Journal of the American Statistical Association*, 76, 802-816.

Triantafyllopoulos, K. (2008): Missing observation analysis for matrix-variate time series data, *Statistics and Probability Letters*, 78, 2647-2653.

Tunniclille Wilson, G., (1973): The estimation of parameters in multivariate time series models, *Journal of the Royal Statistical Society* , B, 35, 76-85.

Viroli, C. (2012): On matrix-variate regression analysis, *Journal of Multivariate Analysis*, 111, 296-309.

Wall, D. De (1988): Matrix-variate distributions. In: *Encyclopedia of Statistical Sciences* (eds. S. Kotz, N. L. Johnson), 5, 326-333, New York, Wiley.

Wallis, K. F. (1977): Multiple time series analysis and the final form of econometric models, *Econometrica*, 45, 1481-1497.

Walker, G. T. (1931): On periodicity in series of related terms, *Proceedings of Royal Society, London Ser. A*, 131, 518-532.

Wang, H. and West, M. (2009): Bayesian analysis of matrix normal graphical models, *Biometrika*, 96(4), 821-834.

Wang, H. (2011): Bayesian analysis of gaussian graphical models for correlated sample, Unpublished manuscript.

Wei, W. W. S. (2006): *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.), Addison-Wesley, Boston.

West, M. and Harrison, J. (1999): *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York, Springer-Verlag.

Whittle, P. (1953): The analysis of multiple stationary time series, *Journal of the Royal Statistical Society*, B, 15 (1), 125-139.

Wigner, E. P. (1955): Characteristic vectors of bordered matrices with infinite dimensions, *Annals of Mathematics*, 62, 548-564.

Wigner, E. P. (1957): Statistical properties of real symmetric matrices with many dimensions, *Canadian Mathematical Congress Proceedings*, Toronto, University of Toronto Press, p 174.

Wigner, E. P. (1965): Distribution laws for the roots of a random Hemitian matrix, In: *Statistical Theory of Spectra: Fluctuations* (C. E. Porter, ed.), Academic Press, New York, 446-461.

Wigner, E. P. (1967): Random matrices in physics, *SIAM Review*, 9, 1-23.

Wishart, J. (1928): The generalized product moment distribution in samples from a normal multivarite population, *Biometrika*, A, 20, 32-52.

Wold, H. (1938): *A Study in the Analysis of Stationary Time Series*, Uppsala, Sweden, Almquist and Wicksell.

Yule, G. U. (1921): On the time-correlation problem, with especial reference to the variate-difference correlation method, with discussion, *Journal of the Royal Statistical Society*, 84, 497-537.

Yule, G. U. (1926): Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society*, 89, 1-64.

Yule, G. U. (1927): On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, *Philosophical Transactions of the Royal Society*, A, 226, 267-226.

Appendix A

This appendix provides the R codes for the simulation in the stationarity study of matrix time series in chapter 3, section 3.4.1. In this program, we can generate a stationary or nonstationary matrix autoregressive of order one (MAR(1)) by controlling systematically the stationarity conditions of matrix time series data given in section 3.4.

```
# Simulating Matrix Time Series
################################################################################
#Assume that Y is matrix normal with mean matrix mu and row and column
#dispersion matrices Sigma and Gamma, respectively.

#Isn't Y = AZB + mu, where Z is a matrix of independent N(0, 1)'s, A is
#the square root matrix of Sigma (the dispersion matrix of the rows) and
#B is the square root matrix of Gamma. It should be easy to write this
#function in R.
################################################################################
K=3
S=2

# generate the two covairance matrix for Omega (Covaraince matrix of columens),
# and Sigma (the covaraince matrix if rows) Sigam.epislon will be the Kronecker
# product of Omega and Sigma (Look at the notations in the estimation part)

install.packages("clusterGeneration")
library(clusterGeneration)

Omega= genPositiveDefMat(dim=S, covMethod=c("eigen"), rangeVar=c(1,4), lambdaLow=1, ratioLambda=4)
Sigma=genPositiveDefMat(dim=K, covMethod=c("eigen"), rangeVar=c(1,5), lambdaLow=1, ratioLambda=6)

Sigma.epsilon=Omega$Sigma %x% Sigma$Sigma # Covaraince matrix of error matrix epsilon (K by S)

#finding the Squre root of Omega:
#Omega.eig <- eigen(Omega$Sigma)
#Omega.sqrt <- Omega.eig$vectors %*% diag(sqrt(Omega.eig$values)) %*% t(Omega.eig$vectors)

#finding the Squre root of Sigma:
#Sigma.eig <- eigen(Sigma$Sigma)
#Sigma.sqrt <- Sigma.eig$vectors %*% diag(sqrt(Sigma.eig$values)) %*% solve(Sigma.eig$vectors)
```

```
Mnorm=function(K, S){
        Z=matrix( rnorm(K*S,mean=0,sd=1), K, S)
        # Now we can generate the random normal matrix distribution with mean zero and
        # covariance matrix "Sigma.epsilon"
        epsilon=Sigma.sqrt%*% Z %*% Omega.sqrt
        return(list(epsilon))
        }

#A=Mnorm(K,S)
####################################################################################
# this function produce the  matrix E^{M\times N}_{ij}

matrix.E=function(M,N,i,j){
                E=matrix(nrow=M, ncol=N,0)
                E[i,j]=1
                return(E)
                        }
#matrix.E(4,3, 2,1)
######################################################################################
# generate coefficient matrix A_r^j
#Put the eigenvalues on the diagonal, 0 everywhere below the
#diagonal, anything you wish above the diagonal.
#If that's too simple, multiply on one side by S and on the other
#by S^(-1) where S is any invertible matrix.
#If you want the matrix to be symmetric, use 0 above the diagonal
#as well as below, and make S an orthogonal matrix.

library(MASS)
library(clusterGeneration)
#install.packages('eigeninv')
#library(eigeninv)

Arj=function(dim, ev= runif(dim, 0, 10)){
                D<- matrix(runif(dim^2), dim,dim)
                D[upper.tri(D)]=0
                diag(D)=ev
   S= genPositiveDefMat(dim=K, covMethod=c("eigen"), rangeVar=c(1,4), lambdaLow=1, ratioLambda=4)$Sigma
        A=S %*% D %*% solve(S)
        return(A)
        }
```

```
#Arj=function(dim, ev= runif(dim, 0, 10)){
#                A= eiginv(ev,dim)
#        return(A)
#        }

##############################################################################
#install.packages("matrixcalc")
#library("matrixcalc")
##############################################################################
K=3
S=2
N=200
#mu=matrix(runif(K*S, -44,54),K,S)

for(m in 1:100){
VectorSimu=matrix(nrow=N+1, ncol=K*S, 0)
Sample=array(dim=c(K,S,N+1),0)  # will hold the simulated matrix TS
A=array(dim=c(K,K,S^2))                 # will hold coefficient matrix A_r^j
Eigenvalues=vector(length=S^2*K) # will hold eigenvalues of all  A_r^j matrices
sumev=0                                  # will hold the sum of eigenvalues vectors of each A_r^j
for(i in 1:S^2){
        # the ranges of random eigenvalues impose for the coeficient matrices
                ev=runif(K,0,1)
                sumev=sumev+ev
                Eigenvalues[((i-1)*K+1):(i*K)]=ev
                A[,,i]=Arj(K,ev )
                }
Eigenvalues
sumev

# find the sum of all A_r^j to check what is the eigenvalues of the sum-matrix!
Atotal=matrix(nrow=K, ncol=K,0) # will hold sum of all coefficient matrices A_r^j
for(i in 1:S^2){
        Atotal=Atotal+A[,,i]
        }
EigenTotal=eigen(Atotal)
EigenTotal$values

# find the modulus of eigenvlus of total matrix
ModulusET=vector(length=K)
```

```
        for(i in 1:K){
 ModulusET[i]=sqrt(Re(EigenTotal$values[i])^2 + Im(EigenTotal$values[i])^2)
        }
ModulusET
#

AvectorTS=matrix(nrow=K*S, ncol=K*S)
        for(j in 1:S){
         for(r in 1:S){
          AvectorTS[((j-1)*K+1):(j*K) ,((r-1)*K+1):(r*K)]=A[,,(j*r)+(j-1)*(S-r)]
                }}

EivalVectorTS=eigen(AvectorTS)$values
EivalVectorTS

# find the modulus of eigenvlus of coeffeicnt matrix for Vector TS
ModulusEVTS=vector(length=K*S)
        for(i in 1:(K*S)){
 ModulusEVTS[i]=sqrt(Re(EivalVectorTS[i])^2 + Im(EivalVectorTS[i])^2)
        }
ModulusEVTS

########################################################################
# simulate 'N' Matrix Time Sereis put them in a array
for(i in 2:N+1){
                F=0
                for(j in 1:S){
                 for(r in 1:S){
                        F=F+A[,,(j*r)+(j-1)*(S-r)] %*% Sample[,,i-1] %*% matrix.E(S,S,r,j)
                                }}
                mr=Mnorm(K,S)
        Sample[,,i]=F+mr[[1]]+mu
        VectorSimu[i,]=AvectorTS %*% as.matrix(VectorSimu[i-1,])+vec(mr[[1]])+ vec(mu)
        #print(Sample[,,i])
        #print(VectorSimu[i,])
                }
########################################################################
# put arry in a matrix
FinalSample=matrix(nrow=N,ncol=K*S)
for(i in 1:K){
  for(j in 1:S){
```

```
        FinalSample[,(i*j)+(i-1)*(S-j)]=Sample[i,j,2:(N+1)]
                }}
# check the stationary condition
if(ModulusEVTS[1]<.95){
                        print(" ModulusEVTS:  ")
                        print(ModulusEVTS)
                         print(" Eigenvalues:  ")
                          print(Eigenvalues)
                         print(" ModulusET:  ")
                         print(ModulusET)

t=2:200
par(mfrow=c(1,2))

matplot(,FinalSample[2:200,], xlab="Time ",type="l")
matplot(,VectorSimu[2:200,], xlab="Time ", type="l")

 exit
write.table(VectorSimu, "StationaryMAR1-PositiveEigen.txt")
write.table(FinalSample, "FinalSample1-Statioanry-Pos.txt")
write.table(mu, "mu1-Statioanry.txt")
write.table(AvectorTS, "A1-Statioanry-PositiveEigen.txt")
write.table(Sigma.epsilon, "Sigma.epsilon-Statioanry.txt")
write.table(Omega$Sigma, "Omega-Statioanry.txt")
write.table(Sigma$Sigma, "Sigma-Statioanry.txt")
write.table(Eigenvalues, "Eigenvalues-PositiveEigen.txt")
                }#end of if
        }# end of for loop with counter "m"
}
```

Appendix B

The R codes that we used in chapter 5 to find autocorrelation functions of the matrix autoregressive model of order one (MAR(1)) are given here. Note that, for simulation we assumed that all parameters of the model are known. Then, we used Yule-Walker equations to find autocovariance and autocorrelation functions.

```
# find correlation matrix for MAR(1)
K=3
S=2

SigmaW=read.table("Sigma.epsilon-Statioanry-v2.txt")
AvectorTSW=read.table("A1-Statioanry-v2.txt")

SigmaW=read.table("Sigma.epsilon-Statioanry-Neg.txt")
AvectorTSW=read.table("A1-Statioanry-Neg.txt")


SigmaW=read.table("Sigma.epsilon-Statioanry-Pos.txt")
AvectorTSW=read.table("A1-Statioanry-NegativeEigen.txt")

SigmaW=Sigma.epsilon
#AvectorTSW=AvectorTS

AW=array(dim=c(K,K,S^2))         # will hold coefficient matrix A_r^j

for(j in 1:S){
        for(r in 1:S){
  AW[1:K,1:K,((j*r)+((j-1)*(S-r)))]= as.matrix(AvectorTSW[((j-1)*K+1):(j*K) ,((r-1)*K+1):(r*K)])
            }
          }
##################################################################################################
#generate matrix E_rj

matrix.E=function(M,N,i,j){
                E=matrix(nrow=M, ncol=N,0)
                E[i,j]=1
                return(E)
```

```
                }
###############################################################################################
install.packages("matrixcalc")
library(matrixcalc)
# Here we are using Yule-Walker equations
# Finally, my code is giving the correct answer of Yule-Walker equations for MAR(1) #Ok

T<- commutation.matrix( S, K )
#T will give the Transformation matrix T in the Matrix Time Series Models paper!

# Find the \psi(0) given in Eq  (9.9) and (9.10)
HH=matrix(nrow=K^2*S^2,ncol=K^2*S^2,0)
for(j in 1:S){
          for(r in 1:S){
                 for(j1 in 1:S){
                 for(r1 in 1:S){
 HH=HH+((matrix.E(S,S,j,r)%x%AW[,,(j1*r1)+(j1-1)*(S-r1)])%x%(AW[,,(j*r)+(j-1)*(S-r)]%x%matrix.E(S,S,j1,r1)))
        }
        }
        }}

ZZ=solve( diag(K^2*S^2)- HH )%*%vec(solve(T)%*% as.matrix(SigmaW))
Psi0=matrix(ZZ, K*S,K*S)
Gamma0=T%*%Psi0

h=40 # number of lags

D= diag(diag(Gamma0))
D=sqrt(D)                         # will be used to obtaion correlation matrix
R=array(dim=c(K*S,K*S,h))         # will hold Correlation matrix rho

    for(i in 1:h){
        P=get(paste("Psi",i-1,sep=""))
        Sum=matrix(nrow=K*S,ncol=K*S,0)
         for(j in 1:S){
           for(r in 1:S){
        Sum=Sum+ (AW[,,(j*r)+(j-1)*(S-r)]%x% diag(S))%*% P %*% (matrix.E(S,S,r,j)%x%diag(K) )
                 }}
          G1=T%*%Sum
          G2=T%*%Sum
# Here we are changing the Gamma to be same as the Gammaprime (Gamma when we consider the
```

```
# vec of K\times S matrix as vector time sereis) becasue the standard formula to obtain rho
# matrix does not properly work with Gamma that obtaiend from MAR (off diagonal blocks of
# Gamma of MAR are transpose of the off diagonal block elements Gammaprime of VAR )

        for(l in 1: S){
                for(m in 1:S){
                if(l!=m){
          G2[(((l-1)*K)+1):(l*K),(((m-1)*K)+1):(m*K)]=t(G2[(((l-1)*K)+1):(l*K),(((m-1)*K)+1):(m*K)])
                }
                }}


Ro=solve(D)%*%(G2)%*%solve(D)
# reorder the Ro matrix (off diagonal part) to get corrleation matrix for MAR models same as it was
# for Gamma in MAR cases

 for(l in 1: S){
                for(m in 1:S){
                if(l!=m){
          Ro[(((l-1)*K)+1):(l*K),(((m-1)*K)+1):(m*K)]=t(Ro[(((l-1)*K)+1):(l*K),(((m-1)*K)+1):(m*K)])
                }
                }}

        R[,,i]=Ro
        assign(paste('Psi', i, sep=''),  Sum)
        assign(paste('Gamma', i, sep=''),  G1)
        assign(paste('rho', i, sep=''),  Ro)
        }
#####################################################################################################
z <- cor(mtcars)
require(lattice)
levelplot(R,panel = panel.levelplot,region = TRUE)
names(R)=C(1:20)
idx=seq(1:6)
R1=setZ(R,idx)

#lag=c('lag=1','lag=2','lag=3','lag=4','lag=5','lag=6','lag=7','lag=8','lag=9','lag=10',
#'lag=11','lag=12','lag=13','lag=14','lag=15','lag=16','lag=17','lag=18','lag=19','lag=20')
lag=paste("lag=", 1:h, sep="")
dimnames(R)=list(1:6,1:6,lag)
```

```
levelplot(R[,,1:9], xlab="Variable1", ylab="Variable2",colorkey = TRUE,region = TRUE)

#plot1=levelplot(R[,,1], xlab="Variable1", ylab="Variable2",colorkey = TRUE,region = TRUE)
#plot2=levelplot(R[,,2], xlab="Variable1", ylab="Variable2",colorkey = TRUE,region = TRUE)
#plot3=levelplot(R[,,3], xlab="Variable1", ylab="Variable2",colorkey = TRUE,region = TRUE)
#plot4=levelplot(R[,,4], xlab="Variable1", ylab="Variable2",colorkey = TRUE,region = TRUE)

#grid.arrange(plot1,plot2,plot3,plot4,ncol=2 ,nrow=2)

# S=1, K=1,2,3
par(mfrow=c(3,3))

for(i in 4:6){
 for(j in 1:3){
 plot(R[i,j,],type="h",xlab='h',ylab='corr',mgp=c(2,1,0),main=paste('(',i,',',j,')'),cex.main=1)
}
}

plot(R[4,6,],type="h", xlab='h',ylab='corr')

par(mfrow=c(2,2))

# Here, I plot the ACF of some pairs

#par(mar=c(.1,2.1, .1, 2.1))
#par(oma=c(0,0,.25,.25))
par(mai=c(0.65,0.45,0.3,0.1)) # give enaough margin for plots
#par(mar=c(0.15,0.15, 0.15, 0.15))
#par(oma=c(.5,0.5,0.5,0.5))
#mar=c(5.1,4.1,4.1,2.1)
# oma=c(10,10,10,10)

e1=expression((i~',’~j)~'='(1,2)~~~~~~~ (i~plain("'")~',’~j~plain("'"))~'='(2,2))
plot(R[2,4,],type="h", xlab='lag', mgp=c(2,1,0), ylab="", main=e1 , cex.main=1)
abline(a=0,b=0)

#title(main= labelsX)
e1=expression((i~',’~j)~'='(3,1)~~~~~~~ (i~plain("'")~',’~j~plain("'"))~'='(3,1))
plot(R[5,5,],type="h", xlab='lag', mgp=c(2,1,0), ylab="", main=e1, cex.main=1)
abline(a=0,b=0)
```

```
#title(main="(5, 5)")
e1=expression((i~','~j)~'='(3,1)~~~~~~~ (i~plain("'")~','~j~plain("'"))~'='(3,2))
plot(R[5,6,],type="h", xlab='lag', mgp=c(2,1,0), ylab="",main=e1, cex.main=1)
abline(a=0,b=0)

#title(main="(5 , 6)")
e1=expression((i~','~j )~'='(3,2)~~~~~~~ (i~plain("'")~','~j~plain("'"))~'='(2,1))
plot(R[6,3,],type="h", xlab='lag', mgp=c(2,1,0), ylab="",main=e1, cex.main=1)
abline(a=0,b=0)


}
```