1

Fundamental concepts and issues in multivariate time series analysis

With the development of computers and the internet, we have had a data explosion. For example, a study of monthly cancer rates in the United States during the past 10 years can involve 50 or many hundreds or thousands of time series depending on whether we investigate the cancer rates for states, cities, or counties. Multivariate time series analysis methods are needed to properly analyze these data in a study, and these are different from standard statistical theory and methods based on random samples that assume independence. Dependence is the fundamental nature of the time series. The use of highly correlated high-dimensional time series data introduces many complications and challenges. The methods and theory to solve these issues will make up the content of this book.

1.1 Introduction

In studying a phenomenon, we often encounter many variables, $Z_{i,t}$, where $i = 1, 2, ..., m$, and the observations are taken according to the order of time, $t$. For convenience we use a vector, $Z_t = [Z_{1,t}, Z_{2,t}, ..., Z_{m,t}]'$, to denote the set of these variables, where $Z_{i,t}$ is the $i$th component variable at time $t$ and it is a random variable for each $i$ and $t$. The time $t$ in $Z_t$ can be continuous and any value in an interval, such as the time series of electric signals and voltages, or discrete and be a specific time point, such as the daily closing price of various stocks or the total monthly sales of various products at the end of each month. In practice, even for a continuous time series, we take observations only at digitized discrete time points for analysis. Hence, we will consider only discrete time series in this book, and with no loss of generalizability, we will consider $Z_{i,t}$, for $i = 1, 2, ..., m, t = 0, \pm 1, \pm 2, ..., $ and hence $Z_t = [Z_{1,t}, Z_{2,t}, ..., Z_{m,t}]'$, $t = 0, \pm 1, \pm 2, ...$.

We call $Z_t = [Z_{1,t}, Z_{2,t}, ..., Z_{m,t}]'$ a multivariate time series or a vector time series, where the first subscript refers to a component and the second subscript refers to the time. The fundamental characteristic of a multivariate time series is that its observations depend not only on component $i$ but also time $t$. The observations between $Z_{i,s}$ and $Z_{j,t}$ can be correlated when $i \neq j$, ...
regardless of whether the times \( s \) and \( t \) are the same or not. They are vector-valued random variables. Most standard statistical theory and methods based on random samples are not applicable, and different methods are clearly needed. The body of statistical theory and methods for analyzing these multivariate or vector time series is referred to as multivariate time series analysis.

Many issues are involved in multivariate time series analysis. They are different from standard statistical theory and methods based on a random sample that assumes independence and constant variance. In multivariate time series, \( \mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]' \), a fundamental phenomenon is that dependence exists not only in terms of \( i \) but also in terms of \( t \). In addition, we have the following important issues to consider:

1. **Fundamental concepts and representations related to dependence.**
   We will introduce the variance–covariance and correlation matrix functions, vector white noise processes, vector autoregressive and vector moving average representations, vector autoregressive models, vector moving average models, and vector autoregressive moving average models.

2. **Relationship between several multivariate time series.**
   A multiple regression is known to be a useful statistical model that describes a relationship between a response variable and several predictor variables. The error term in the model is normally assumed to be uncorrelated noise with zero mean and constant variance. We will extend the results to a multivariate time series regression model where both response variables, and predictor variables are vectors. More importantly, not only are all components in the multivariate regression equation time series variables, but also the error term follows a correlated time series process.

3. **Dimension reduction and model simplification.**
   Without losing information, we will introduce useful methods of dimension reduction and representation including principal components and factor analysis in time series.

4. **Representations of time variant variance–covariance structure.**
   Unlike most classical linear methods, where the variance of error term has been assumed to be constant, in time series analysis a non-constant variance often occurs and generalized autoregressive conditional heteroscedasticity (GARCH) models are been introduced. The literature and theory of GARCH models for univariate time series was introduced in chapter 15 of Wei (2006). In this book, we will extend the results to the multivariate GARCH models.

5. **Repeated measurement phenomenon.**
   Many fields of study, including medicine, biology, social science, and education, involve time series measurements of treatments for different subjects. They are multivariate time series but often relatively short, and the applications of standard time series methods are difficult, if not impossible. We will introduce some methods and various models that are useful for analyzing repeated measures data. Empirical examples will be used as illustrations.

6. **Space and time series modeling.**
   In many multivariate time series applications, the components \( i \) in \( Z_{i,t} \) refer to regions or locations. For example, in a crime study, the observations can be the crime rates of
different counties in a state, and in a market study, one could look at the price of a certain product in different regions. Thus, the analysis will involve both regions and time, and we will construct space and time series models.

7. Multivariate spectral analysis.
Similar to univariate time series analysis where one can study a time series through its autocovariance/autocorrelation functions and lag relationships or through its spectrum properties, we can study a multivariate time series through a time domain or a frequency domain approach. In the time domain approach we use covariance/correlation matrices, and in the frequency domain approach we will use spectral matrices. We will introduce spectral analysis for both multivariate stationary and nonstationary time series.

Because of high-speed internet and the power and speed of the new generation of computers, a researcher now faces some very challenging phenomena. First, he/she must deal with an ever-increasing amount of data. To find useful information and hidden patterns underlying the data, a researcher may use various data-mining methods and techniques. Adding a time dimension to these large databases certainly introduces new aspects and challenges. In multivariate time series analysis, a very natural issue is the high dimension problem where the number of parameters may exceed the length of the time series. For example, a simple second order vector autoregressive VAR(2) model for the 50 states in the USA will involve more than 5000 parameters, and the length of the time series may be much shorter. For example, the length of the monthly observations for 20 years is only 240. Traditional time series methods are not designed to deal with these kinds of high-dimensional variables. Even with today’s computer power and speed, there are many difficult problems that remain unsolved. As most statistical methods are developed for a random sample, the use of highly correlated time series data certainly introduces a new set of complications and challenges, especially for a high-dimensional data set.

The methods and theory to solve these issues will be the focus of this book. Examples and applications will be carefully chosen and presented.

1.2 Fundamental concepts
The \( m \)-dimensional vector time series process, \( \mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]' \), is a stationary process if each of its component series is a univariate stationary process and its first two moments are time-invariant. Just as a univariate stationary process or model is characterized by its moments such as mean, autocorrelation function, and partial autocorrelation function, a stationary vector time series process or model is characterized by its mean vector, correlation matrix function, and partial correlation matrix function.

1.2.1 Correlation and partial correlation matrix functions
Let \( \mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{m,t}]' \), \( t = 0, \pm 1, \pm 2, \ldots \) be a \( m \)-dimensional stationary real-valued vector process so that \( E(Z_{i,t}) = \mu_i \) is constant for each \( i = 1, 2, \ldots, m \) and the cross-covariance between \( Z_{i,t} \) and \( Z_{j,s} \), for all \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, m \), are functions only of the time difference \( (s - t) \). Hence, we have the mean vector
and the lag $k$ covariance matrix

$$\Gamma(k) = \text{Cov}(Z_t, Z_{t+k}) = E\left[(Z_t - \mu)(Z_{t+k} - \mu)\right]$$

$$= E \begin{bmatrix} Z_{1,t} - \mu_1 \\ Z_{2,t} - \mu_2 \\ \vdots \\ Z_{m,t} - \mu_m \end{bmatrix} \begin{bmatrix} Z_{1,t+k} - \mu_1, Z_{2,t+k} - \mu_2, \ldots, Z_{m,t+k} - \mu_m \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{1,1}(k) & \gamma_{1,2}(k) & \cdots & \gamma_{1,m}(k) \\ \gamma_{2,1}(k) & \gamma_{2,2}(k) & \cdots & \gamma_{2,m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,1}(k) & \gamma_{m,2}(k) & \cdots & \gamma_{m,m}(k) \end{bmatrix},$$

where

$$\gamma_{i,j}(k) = E(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j)$$

for $k = 0, \pm 1, \pm 2, \ldots, i = 1, 2, \ldots, m,$ and $j = 1, 2, \ldots, m$. As a function of $k$, $\Gamma(k)$ is called the covariance matrix function for the vector process $Z_t$. Also, $i = j$, $\gamma_{i,i}(k)$ is the autocovariance function for the $i$th component process, $Z_{i,t}$; and $i \neq j$, $\gamma_{i,j}(k)$ is the cross-covariance function between component series $Z_{i,t}$ and $Z_{j,t}$. The matrix $\Gamma(0)$ can be easily seen to be the contemporaneous variance–covariance matrix of the process.

The covariance matrix function has the following properties:

1. $\Gamma(k) = \Gamma'(-k)$. This follows because

$$\gamma_{i,j}(k) = E\left[(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j)\right] = E\left[(Z_{j,t+k} - \mu_j)(Z_{i,t} - \mu_i)\right] = \gamma_{j,i}(-k).$$

2. $|\gamma_{i,j}(k)| \leq [\gamma_{i,i}(0)\gamma_{j,j}(0)]^{1/2}$, for all $i, j = 1, \ldots, m$, because of the Cauchy–Schwarz inequality.

3. The covariance matrix function is positive semidefinite in the sense that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \Gamma(t_i - t_j) \alpha_j \geq 0,$$
for any set of time points $t_1, t_2, \ldots, t_n$ and any set of real vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$. The result follows immediately from $\text{Var}\left[\sum_{i=1}^{n} \alpha_i' Z_i \right] \geq 0$.

The correlation matrix function for the vector process is defined by

$$\rho(k) = D^{-1/2} \Gamma(k) D^{-1/2} = \begin{Bmatrix} \rho_{i,j}(k) \end{Bmatrix}$$  \hspace{1cm} (1.4)$$

for $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, m$, where $D$ is the diagonal matrix in which the $i$th diagonal element is the variance of the $i$th process; that is, $D = \text{diag} \left[ \gamma_1(0), \gamma_2(0), \ldots, \gamma_m(0) \right]$. Thus, the $i$th diagonal element of $\rho(k)$ is the autocorrelation function for the $i$th component series $Z_{i,t}$ whereas the $(i,j)$th off-diagonal element of $\rho(k)$ is the cross-correlation function between component series $Z_{i,t}$ and $Z_{j,t}$.

Similarly, the correlation matrix functions have the following properties:

1. $\rho(k) = \rho(-k)$.
2. The correlation matrix function is positive semidefinite so that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i' \rho(t_i-t_j) \alpha_j \geq 0,$$  \hspace{1cm} (1.5)$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any set of real vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Other than the correlation matrix function, another useful function for describing a vector time series process is the partial correlation matrix function. The concept of a partial correlation matrix function was introduced much later, and there are different versions.

Heyse and Wei (1985) extended the definition of univariate partial autocorrelation to vector time series and derived the correlation matrix between $Z_{t+1}$, $Z_{t+s-1}$. This partial correlation matrix is defined as the correlation between the residual vectors

$$U_{s-1,t+s} = Z_{t+s} - \alpha_{s-1,1} Z_{t+s-1} - \cdots - \alpha_{s-1,s-1} Z_{t+1}$$

$$= \begin{cases} Z_{t+s} - \sum_{j=1}^{s-1} \alpha_{s-1,j} Z_{t+s-j}, & s \geq 2, \\ Z_{t+1}, & s = 1, \end{cases}$$  \hspace{1cm} (1.6)$$

and

$$V_{s-1,t} = Z_{t} - \beta_{s-1,1} Z_{t+1} - \cdots - \beta_{s-1,s-1} Z_{t+s-1}$$

$$= \begin{cases} Z_{t} - \sum_{j=1}^{s-1} \beta_{s-1,j} Z_{t+j}, & s \geq 2, \\ Z_{t+1}, & s = 1. \end{cases}$$  \hspace{1cm} (1.7)$$

Let $C_{UV}(s)$ be the covariance between $V_{s-1,t}$ and $U_{s-1,t+s}$, i.e. $C_{UV}(s) = \text{Cov}(V_{s-1,t}, U_{s-1,t+s})$. Heyse and Wei (1985) showed that
\[
C_{UU}(0) = \text{Var}(U_{s-1,t+s}) = E(U_{s-1,t+s}U_{s-1,t+s}')
\]
\[
= \Gamma(0) - [\Gamma'(1), \Gamma'(2), \ldots, \Gamma'(s-1)]
\]
\[
C_{VV}(0) = \text{Var}(V_{s-1,t}) = E(V_{s-1,t}V_{s-1,t}')
\]
\[
= \Gamma(0) - [\Gamma(1), \Gamma(s-2), \ldots, \Gamma'(1)]
\]

where \( \Gamma(k) = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}') \). Thus, the partial autocorrelation matrix function is

\[
P(s) = [D_{V}(s)]^{-1}C_{UU}(s)[D_{U}(s)]^{-1},
\]

where \( D_{V}(s) \) is the diagonal matrix in which the \( i \)th diagonal element is the square root of the \( i \)th diagonal element of \( C_{VV}(0) \) and \( D_{U}(s) \) is similarly defined for \( C_{UU}(0) \).

Tiao and Box (1981) defined the partial autoregression matrix at lag \( s \), denoted by \( \Phi_{x,s} \), to be the last matrix coefficient when the data is fitted to a vector autoregressive process of order \( s \).

It can be shown that

\[
\Phi_{x,s} = C'_{UU}(s)[D_{U}(s)]^{-1}.
\]

Ansley and Newbold (1979) defined the multivariate partial autocorrelation matrix at lag \( s \) to be

\[
Q(s) = [W_{U}(s)]^{-1}C'_{UU}(s)[W_{V}(s)]^{-1},
\]
where $W_U(s)$ and $W_V(s)$ are the symmetric square roots of $C_{UU}(0)$ and $C_{VV}(0)$, respectively, defined such that $|W_U(s)|^2 = C_{UU}(0)$ and $|W_V(s)|^2 = C_{VV}(0)$. However, it should be noted that although $P(s)$, $\Phi(s)$, and $Q(s)$ all share the same cut-off property for vector VAR($s$) models, the elements of $P(s)$ are proper correlation coefficients but those of $\Phi(s)$ and $Q(s)$ are not correlation coefficients except when $m=1$; that is, except in the univariate case in which $P(s) = \Phi(s) = Q(s)$.

For more details, we refer readers to Wei (2006, chapter 16).

### 1.2.2 Vector white noise process

The $m$-dimensional vector process, $a_t$, is said to be a vector white noise process with mean vector $0$ and covariance matrix function $\Sigma$ if

$$E[a_t a_{t+k}'] = \begin{cases} 
\Sigma, & \text{if } k = 0, \\
0, & \text{if } k \neq 0, 
\end{cases}$$

where $\Sigma$ is a $m \times m$ symmetric positive definite matrix. Note that although the components of the white noise process are uncorrelated at different times, they may be contemporaneously correlated. It is a Gaussian white noise process if $a_t$ also follows a multivariate normal distribution. Unless mentioned otherwise, $a_t$ will be used to denote a Gaussian vector white noise process with mean vector $0$ and covariance matrix function $\Sigma$, VWN($0$, $\Sigma$), in this book.

### 1.2.3 Moving average and autoregressive representations of vector processes

A $m$-dimensional stationary vector time series process $Z_t$ that is purely nondeterministic can always be written as a linear combination of a sequence of vector white noises, that is

$$Z_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots = \mu + \sum_{\ell=0}^{\infty} \Psi_{\ell} a_{t-\ell} = \mu + (B) a_t,$$

where $\Psi(B) = \sum_{\ell=0}^{\infty} \Psi_{\ell} B^\ell$, $\Psi_0 = I$, and the sequence of $m \times m$ coefficient matrices $\Psi_{\ell}$ is square summable, $\sum_{\ell=0}^{\infty} ||\Psi_{\ell}||^2 < \infty$, in the sense that if we write $\Psi(B) = \sum_{\ell=0}^{\infty} \Psi_\ell B^\ell = [\psi_{i,j}(B)]$, and $\psi_{i,j}(B) = \sum_{\ell=0}^{\infty} \psi_{i,j,\ell} B^\ell$, we have $\sum_{\ell=0}^{\infty} ||\psi_{i,j,\ell}||^2 < \infty$, for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$. The $B$ is the backshift operator such that $B^j a_t = a_{t-j}$.

The infinite sum of random variables is defined as the limit in quadratic mean of the finite partial sums. Thus, $Z_t$ in Eq. (1.15) is defined such that

$$E \left[ \left( Z_t - \mu - \sum_{\ell=0}^{n} \Psi_{\ell} a_{t-\ell} \right)' \left( Z_t - \mu - \sum_{\ell=0}^{n} \Psi_{\ell} a_{t-\ell} \right) \right] \to 0 \text{ as } n \to \infty .$$

The Eq. (1.15) is known as the vector moving average (MA) representation.

For a given sequence of covariance matrices, $\Gamma(k)$, $k = 0, \pm 1, \pm 2, \ldots$, the covariance matrix generating function is defined as
\[ \Gamma(B) = \sum_{k=-\infty}^{\infty} \Gamma(k)B^k, \]  
(1.17)

where the covariance matrix of lag \( k \) is the coefficient of \( B^k \) and \( B^{-k} \). For a stationary vector process given in Eq. (1.15), it can be easily seen that

\[ \Gamma(B) = \sum_{k=-\infty}^{\infty} \Gamma(k)B^k = \Psi(B)\Sigma \Psi'(B^{-1}), \]  
(1.18)

where \( \Psi(B^{-1}) = \sum_{j=0}^{\infty} \Psi jB^{-j} \).

A vector time series process \( Z_t \) is said to be invertible if it can be written as a vector autoregressive (AR) representation

\[ Z_t = \theta_0 + \sum_{\ell=1}^{\infty} \Pi_{\ell} Z_{t-\ell} + \cdots + a_t, \]  
(1.19)

or equivalently,

\[ \Pi(B)Z_t = \theta_0 + a_t, \]  
(1.20)

where \( \Pi(B) = \sum_{\ell=0}^{\infty} \Pi_{\ell}B^\ell \), \( \Pi_0 = I \), so that the sequence of \( m \times m \) autoregressive coefficient matrices \( \Pi_{\ell} \) is absolutely summable in the sense that if we write \( \Pi(B) = \sum_{\ell=0}^{\infty} \Pi_{\ell}B^\ell = [\pi_{i,j}(B)] \), and \( \pi_{i,j}(B) = \sum_{\ell=0}^{\infty} \pi_{i,j,\ell}B^\ell \), we have \( \sum_{\ell=0}^{\infty} |\pi_{i,j,\ell}| < \infty \), for all \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, m \).

**Remarks**

1. Statistical software is needed to perform time series data analysis. We will provide the associated software code for the empirical examples used in the book at the end of chapters. Since R is a free software package supported by researchers through the R Foundation for Statistical Computing, we will use R (2018 version, Rx64 3.4.2) most of the time. Some procedures are not available or difficult to implement in R, so then SAS (2015) or MATLAB (matrix laboratory) will be used. Readers should know that even in the same software, available functions may vary from version to version. For various analyses of empirical examples in the book, readers can simply copy the software code and paste into the relevant software on your laptop to get the presented outputs. When you run into difficulty in using the code, it could be because required functions are missing in the software version you are using.

2. This book is designed for a research-oriented advanced time series analysis course. Problem sets normally found in an introductory course will not be provided. Instead, research-oriented projects will be suggested.
3. To illustrate the use of multivariate time series analysis methods introduced in the book, we have used very extensive empirical examples with data sets from “Bookdata” installed on the C drive of my laptop computer as shown in the software codes. The printed data sets are provided in the book appendix. To help readers, I have uploaded these data sets as “Data Set 2” on my website (http://astro.temple.edu/~wwei/). Sincerely thank to Wiley, you can also access the data sets through its provided website: www.wiley.com/go/wei/datasets.

Projects

1. Find a univariate time series book, for example, Wei (2006, ch. 1–13), to review your background.

2. Write a detailed study plan on multivariate time series analysis and applications so that you can evaluate your accomplishments at the end of the semester if you use this book in your course.

References


