Theory of Multivariate Time Series Analysis and Its Application

多変量時系列に対する理論とその応用

February, 2019

Hideaki NAGAHATA

長幡　英明

Supervisor: *Professor* Masanobu Taniguchi
Examiner: *Professor* Yoichi Nishiyama
and
Professor Yasutaka Shimizu

Theory of Multivariate Time Series Analysis and Its Application

多変量時系列に対する理論とその応用

February, 2019

Waseda University
Graduate School of Fundamental Science and Engineering
Department of Pure and Applied Mathematics,
Research on Mathematical Statistics, Time Series and Finance

Hideaki NAGAHATA

長幡　英明

Contents

Chapter 1

Introduction

Many methods for multivariate data have been developed. For example, Anderson (2003) and Rao (2009) proposed and developed many multivariate methods based on large i.i.d. sample approximations. Recently, Fujikoshi et al. (2011) expanded their results to high-dimensional i.i.d. framework. Also, in time series analysis, Taniguchi and Kakizawa (2000) and Brillinger (2001) discussed some multivariate methods. These statistical inferences are very important, but they can not be used for high-dimensional time series problems. This dissertation discusses some multivariate statistics for time series models with possibly high-dimensional setting. Specifically, we focus on statistics of discriminant analysis, cluster analysis and the analysis of variance (ANOVA).

This doctoral dissertation is organized as follows. Chapter 2 is based on Liu, Nagahata, Uchiyama and Taniguchi (2017). We discuss discriminant and cluster analysis for possibly high-dimensional time series. Discriminant and cluster analysis of high-dimensional time series data have been an urgent need in more and more academic fields. To settle the always-existing problem of bias in discriminant statistics for high-dimensional model, we introduce a new class of disparities with jackknife type adjustment for discriminant and cluster analysis. In numerical experiments, our proposed discriminant statistics result in smaller misclassification error rates than other existing classifiers. The performance is also verified by real data of companies on the Tokyo Stock Exchange. We conclude that our method is suitable for the discriminant and cluster analysis of high-dimensional dependent data.

In Chapter 3, we introduce an application of Liu et al. (2017) for real data. Recently, discriminant and cluster analysis of high-dimensional time series data have been developed for academic fields. This effective methods have been applied to genetic analysis and are expected to apply to many other fields. However applications of high-dimensional financial time series data are very poor. In this chapter, we study clustering of companies by using some financial indicators which are large dimensional and small observations. Specifically, we propose consistent classifiers for the problem of clustering. By computing them, we successfully draw some dendrogram of financial indicators. We conclude that the proposed method has a potential in applications for rating companies.

Chapter 4 discusses ANOVA for multivariate time series. This study establishes a new approach for ANOVA of time series. ANOVA has been sufficiently tailored for cases with independent observations, but there has recently been substantial demand across many fields for ANOVA in cases with dependent observations. For example, ANOVA for dependent observations is important to analyze differences among industry averages within financial data. Despite this demand, the study of ANOVA for dependent observations is more nascent than that of ANOVA for independent observations, and, thus, in this analysis, we study ANOVA for dependent observations. Specifically, we show the asymptotics of classical tests proposed for independent observations and give a sufficient condition for the observations to be asymptotically χ^2 distributed. If this sufficient condition is not satisfied, we suggest a likelihood ratio test based on the Whittle likelihood and derive an asymptotic χ^2 distribution of our test. Finally, we provide some numerical examples using simulated and real financial data as applications of these results.

In Chapter 5, we consider ANOVA for high-dimensional time series. Recently, there has been considerable demand for ANOVA of high-dimensional and dependent observations in many fields. For example, it is important to analyze differences among industry averages of financial data. However, ANOVA for these types of observations has been inadequately developed. In this chapter, we thus present a study of ANOVA for high-dimensional and dependent observations. Specifically, we present the asymptotics of classical test statistics proposed for independent observations and provide a sufficient condition for them to be asymptotically normal. Numerical examples for simulated and radioactive data are presented as applications of these results.

Chapter 6 introduces higher-order approximation of classical ANOVA models under high-dimensional time series setting. Now it is important to analyze differences among big data's averages of any areas of all over the world, for example, the financial industry, the manufacturing one, and so on. However, the numerical accuracy of ANOVA for these types of observations has been inadequately developed. In this chapter, we thus present a study on Edgeworth expansion of ANOVA for high-dimensional and dependent observations. Specifically, we present the second-order approximation of classical test statistics proposed for independent observations. We also give numerical examples for simulated high-dimensional time series data.

Throughout this dissertation, we define some of the notation. The set of all integers, positive integers and real numbers are denoted as $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{R}$, respectively. We denote the imaginary unit by $\mathbf{i}$. δ_{ij} and $\mathbb{1}$ denote Kronecker's delta and the indicator function. Let $\boldsymbol{A}'$, $\boldsymbol{B}^*$, $\boldsymbol{I}_p$, $\boldsymbol{O}_i$, and $\boldsymbol{O}_{i\times j}$ be the transpose of matrix $\boldsymbol{A}$, the conjugate transpose of matrix $\boldsymbol{B}$, the $p \times p$ identity matrix, the i-dimensional zero vector, and the $i \times j$ zero matrix, respectively. For the sequence of random vectors, $\xrightarrow{p}$ denotes the convergence in probability, and $\xrightarrow{d}$ denotes the convergence in distribution. Let $\boldsymbol{O}_{\mathrm{P}}(a_n)$ be an order of the probability that is, for a sequence of random variables $\{X_n\}$ and a sequence of real numbers$\{a_n\}$, $0 < a_n \in \mathbb{R}$, $\{a_n^{-1} X_n\}$ is bounded in probability, and let $\boldsymbol{O}_{\mathrm{P}}^U(\cdot)$ be a $p \times p$ matrix whose elements are probability order $\boldsymbol{O}_{\mathrm{P}}(\cdot)$ with respect to all elements uniformly. In addition, let $|\cdot|$ be the determinant of $\cdot$, $\|\cdot\|$ be the Euclidean norm of $\cdot$, respectively.

Chapter 2

Discriminant and cluster analysis of possibly high-dimensional time series data by a class of disparities

To make a statistical decision for high-dimensional time series data is a matter of great concern nowadays. High-dimensional dependent data are observed in many scientific fields, such as economics, finance, bioinformatics and so on. Previous research for statistical analysis of high-dimensional dependent data, however, was not sufficient. This chapter sheds light on the issue of discriminant and cluster analysis of high dimensional dependent data. There have been a lot of fundamental results for discriminant and cluster analysis of independent and identically distributed (i.i.d.) data or time series data in finite dimension case. Anderson (2003) developed the linear discriminant statistic based on the Mahalanobis distance and likelihood ratio. Rao (2009) proposed the linear and quadratic discriminant statistics for several normal populations $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$. Recently many results were reported in i.i.d. and high-dimensional data with dimension $p \to \infty$. The issue originates from the fact that the inverse matrix of the sample covariance matrix does not exist in high-dimensional, low sample size (HDLSS) situation, where $p/n \to \infty$. Dempster (1958) discussed the multivariate two sample significance test based on Hotelling's T^2 in high-dimension analysis. Saranadasa (1993) considered two normal populations $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ of sample size n_j for $j = 1, 2$, and, in the case $p/(n_1 + n_2 - 2) \to y \in (0, 1)$, evaluated the misclassification rate for the A-criterion, which is essentially the Euclidean classifier. Bai and Saranadasa (1996) also assumed $p/(n_1 + n_2 - 2) \to y \in (0, 1)$ and $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$, and derived the asymptotic powers of the classical Hotelling's T^2 test and Dempster's nonexact test for a two-sample problem. Hall et al. (2005) suggested a geometric representation of the i.i.d. and HDLSS data by using a non-standard type of asymptotics. A scale adjusted-type distance-based classifier for i.i.d. and high-dimensional data was proposed by Chan and Hall (2009) (see also Aoshima and Yata (2014)). Yata and Aoshima (2009, 2012) developed a method for principal component analysis in the i.i.d. and HDLSS data.

Discriminant analysis for finite dimensional stationary time series has a long history. For time domain approach, Gersch et al. (1979) developed a classifier based on the Kullback-Leibler information measure. Applying approximation based on the Whittle likelihood, Shumway (1982) proposed the method of discrimination for stationary time series using frequency domain approach. Kakizawa et al. (1998) extended the Whittle likelihood to minimum discrimination information for the classification of multivariate stationary time series. More details of discriminant analysis for time series can be found in Taniguchi and Kakizawa (2000). They mentioned both time domain and frequency domain approaches for the discriminant analysis and discussed the problem of discriminating linear processes.

In this chapter, we are concerned with high-dimensional stationary process $\{\boldsymbol{X}(t)\}$, which is supposed to belong to one of the two categories;

$$\pi_1 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(1)}(\lambda),$$

$$\pi_2 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(2)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(2)}(\lambda),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{f}(\lambda)$ are the mean vector and spectral density matrix of the process $\{\boldsymbol{X}(t)\}$, respectively. We then consider a distance-based classifier (see Chan and Hall (2009)). Using the multivariate analogue of the methodology of the jackknife in the stationary observations developed by Künsch (1989) (see also Carlstein (1986)), we propose a new classifier with bias adjustment in time series data. We primarily discuss the consistency of the distance-based classifier for multivariate stationary time series data where the dimension p is allowed to diverge, under suitable conditions on size of samples and training samples (n, n_1, n_2) and dimension p. We also conduct the cluster analysis for real financial data.

The remainder of the chapter is organized as follows. In Section 2.1, we show the misspecification rates of our discriminant statistics converge to 0 under suitable conditions on (p, n, n_1, n_2). In Section 2.2, we compare our proposed classifier with the existing classifiers through simulation studies. In Section 2.3, we conduct the cluster analysis for real financial data of companies on the Tokyo Stock Exchange. Proofs of Lemmas and Theorems are placed in Appendix.

2.1 Discriminant statistics for possibly high dimensional time series data

Let $\{\boldsymbol{X}(t) = (X_1(t), \dots, X_p(t))' \,;\, t \in \mathbb{Z}\}$ be a p-dimensional stationary process with mean vector $\boldsymbol{\mu}$ and autocovariance matrix function $\boldsymbol{R}(t) = \{R_{ij}(t)\,;\, i, j = 1, \dots, p\}$. Here the dimension p is allowed to diverge. Suppose we observe $\boldsymbol{X} = \{\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)\}$ from the stationary process $\{\boldsymbol{X}(t)\}$, which belongs to one of the following two categories

$$\pi_1 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}, \quad \boldsymbol{R}(t) = \boldsymbol{R}^{(1)}(t),$$

$$\pi_2 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(2)}, \quad \boldsymbol{R}(t) = \boldsymbol{R}^{(2)}(t).$$

Also we have independent training samples $\boldsymbol{X}^{(1)} = \{\boldsymbol{X}^{(1)}(1), \dots, \boldsymbol{X}^{(1)}(n_1)\}$ and $\boldsymbol{X}^{(2)} = \{\boldsymbol{X}^{(2)}(1), \dots, \boldsymbol{X}^{(2)}(n_2)\}$ from π_1 and π_2 with size n_1 and n_2, respectively. Write $\Delta \equiv \|\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}\|^2$. Throughout this chapter, the notation E and Var often denote the expectation and variance with respect to a triplet of $(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$. The following sample versions for fundamental quantities are introduced:

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}(t), \quad \bar{\boldsymbol{X}}^{(i)} = \frac{1}{n_i}\sum_{t=1}^{n_i} \boldsymbol{X}^{(i)}(t),$$

$$\boldsymbol{S}^{(i)} = \frac{1}{n_i - 1}\sum_{t=1}^{n_i} (\boldsymbol{X}^{(i)}(t) - \bar{\boldsymbol{X}}^{(i)})(\boldsymbol{X}^{(i)}(t) - \bar{\boldsymbol{X}}^{(i)})', \quad i = 1, 2.$$

To classify the time series data, we use the following discriminant statistic:

$$\Gamma(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big). \tag{2.1}$$

This statistic and its scale-adjusted version have been discussed by, e.g., Chan and Hall (2009) and Aoshima and Yata (2014) for the i.i.d. and HDLSS data. For simplicity, denote $\boldsymbol{\Gamma}(\boldsymbol{X}) \equiv \Gamma(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$. First, we evaluate the expectation and variance of $\boldsymbol{\Gamma}(\boldsymbol{X})$.

Lemma 2.1.1 *When $\boldsymbol{X}$ belongs to π_i, $E(\boldsymbol{\Gamma}(\boldsymbol{X})) = (-1)^i \Delta/2 + B$, where*

$$B = \frac{1}{2n_1}\sum_{u=1-n_1}^{n_1-1}\Big(1 - \frac{|u|}{n_1}\Big)\mathrm{tr}\boldsymbol{R}^{(1)}(u) - \frac{1}{2n_2}\sum_{u=1-n_2}^{n_2-1}\Big(1 - \frac{|u|}{n_2}\Big)\mathrm{tr}\boldsymbol{R}^{(2)}(u).$$

Let $c^{(l)}_{a_1a_2a_3}(t_1,t_2)$ and $c^{(l)}_{a_1a_2a_3a_4}(t_1,t_2,t_3)$ denote the third and the fourth order cumulants of the process $\{\boldsymbol{X}^{(l)}(t)\}$ in the category π_l for $l=1,2$ as follows:

$$c^{(l)}_{a_1a_2a_3}(t_1,t_2)=\mathrm{cum}\{X^{(l)}_{a_1}(t),X^{(l)}_{a_2}(t+t_1),X^{(l)}_{a_3}(t+t_2)\},$$

$$c^{(l)}_{a_1a_2a_3a_4}(t_1,t_2,t_3)=\mathrm{cum}\{X^{(l)}_{a_1}(t),X^{(l)}_{a_2}(t+t_1),X^{(l)}_{a_3}(t+t_2),X^{(l)}_{a_4}(t+t_3)\}.$$

Lemma 2.1.2 *When* $\boldsymbol{X}$ *belongs to* π_i,

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X}))&=\sum_{l=1}^{2}\frac{1}{nn_l}\mathrm{tr}\Big(\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\\
&+\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)+\frac{1}{n_{\bar{i}}}\sum_{u=1-n_{\bar{i}}}^{n_{\bar{i}}-1}\Big(1-\frac{|u|}{n_{\bar{i}}}\Big)\boldsymbol{R}^{(\bar{i})}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)\\
&+\frac{1}{2}\sum_{l=1}^{2}\Big[\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)^2\Big\}\\
&+(-1)^i\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}(\mu^{(1)}_j-\mu^{(2)}_j)\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c^{(l)}_{jkk}(t_1-s,t_2-s)\\
&+\frac{1}{4n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c^{(l)}_{jjkk}(t_1-s_1,s_2-s_1,t_2-s_1)\Big],\end{aligned}$$

where $(i,\bar{i})=(1,2),\ (2,1)$.

Remark 2.1.1 *Lemma 2.1.2 includes i.i.d. cases considered in Aoshima and Yata (2014). Note that the processes considered in their chapter is non-Gaussian with all vanishing third and fourth cumulants. In fact, if we set*

$$\boldsymbol{R}^{(l)}(u)=\begin{cases}\boldsymbol{\Sigma}^{(l)}, & \text{if } u=0,\\ 0, & \text{if } u\neq 0\end{cases}$$

for $l=1,2$, *we see that when* $n=1$ *and* $\boldsymbol{X}$ *belongs to* π_i,

$$\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X}))=\sum_{k=1}^{2}\frac{1}{n_k}\mathrm{tr}\Big\{\boldsymbol{\Sigma}^{(i)}\boldsymbol{\Sigma}^{(k)}\Big\}+\big(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}\big)'(\boldsymbol{\Sigma}^{(i)}+\frac{\boldsymbol{\Sigma}^{(j)}}{n_j})\big(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}\big)+\frac{1}{2}\sum_{l=1}^{2}\mathrm{tr}\Big\{\frac{1}{n_l^2}\Big(\boldsymbol{\Sigma}^{(l)}\Big)^2\Big\},$$

where $(i,j)=(1,2),(2,1)$.

Now, the classification rule is to classify $\boldsymbol{X}$ into π_1 if $\boldsymbol{\Gamma}(\boldsymbol{X}) < 0$ and into π_2 otherwise. To discuss the asymptotic property of $\boldsymbol{\Gamma}(\boldsymbol{X})$, we impose the following assumptions.

Assumption 2.1.1 *(i) $n_2 = c_0 n_1$ for some constant $c_0 > 0$.*

(ii) There exists $\eta \geq 0$ such that $c_1 p^\eta < \Delta < c_2 p^\eta$ for some constants $c_1 > 0$ and $c_2 > 0$.

Assumption 2.1.2 *(i) The autocovariance matrix function $\boldsymbol{R}^{(l)}(u)$ of the stationary process $\{\boldsymbol{X}^{(l)}(t)\}$ in the category π_l for $l = 1, 2$ satisfies*

$$\sum_{l=1}^{2} \sum_{t=-\infty}^{\infty} |R_{ij}^{(l)}(t)| < \infty$$

uniformly for $i, j = 1, \ldots, p$.

(ii) The third and fourth order of cumulants of the stationary process $\{\boldsymbol{X}^{(l)}(t)\}$ satisfy

$$\sum_{l=1}^{2} \sum_{t_1, t_2=-\infty}^{\infty} |c_{a_1 a_2 a_3}^{(l)}(t_1, t_2)| < \infty,$$

$$\sum_{l=1}^{2} \sum_{t_1, t_2, t_3=-\infty}^{\infty} |c_{a_1 a_2 a_3 a_4}^{(l)}(t_1, t_2, t_3)| < \infty$$

uniformly for $a_1, a_2, a_3, a_4 = 1, \ldots, p$.

Assumption 2.1.3 ($\boldsymbol{\Gamma}(\boldsymbol{X})$) *Suppose either of the following conditions is satisfied:*

(I) p is finite and $n_1, n \to \infty$ ($\eta \geq 0$);

(II) $p \to \infty$, and

- *(i) if $\eta > 1$, then, both n_1 and n are finite or infinite,*
- *(ii) if $\eta = 1$, then, $n_1 \to \infty$ and n is finite or infinite,*
- *(iii) if $1/2 < \eta < 1$, then, $n_1 \to \infty$ and n is finite or infinite, such that $p = o(n_1^{1/(1-\eta)})$,*
- *(iv) if $\eta = 1/2$, then, $n_1, n \to \infty$, such that $p = o(n_1^2)$,*
- *(v) if $0 \leq \eta < 1/2$, then, $n_1, n \to \infty$, such that $p = o\big((n_1^{1/(1-\eta)} n^{2/(1-2\eta)})/(n_1^{1/(1-\eta)} + n^{2/(1-2\eta)})\big)$.*

Let $P(i|j)$ be misclassification rate by the classification statistic (2.1) such that $\boldsymbol{X}$ belonging to π_j is erroneously assigned to π_i ($i \neq j$). We say that the statistic $\boldsymbol{\Gamma}(\boldsymbol{X})$ is a consistent classifier if $P(i|j) \to 0$ for $(i,j) = (1,2), (2,1)$. We obtain the following theorem.

Theorem 2.1.1 *Under Assumptions 2.1.1 – 2.1.3,* $\boldsymbol{\Gamma}(\boldsymbol{X})$ *is a consistent classifier.*

Condition under which $\boldsymbol{\Gamma}(\boldsymbol{X})$ becomes a consistent discriminant statistic is rather restrictive, due to the requirement $B/\Delta = O(p/(n_1\Delta)) = o(1)$. One may consider $\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}} = \boldsymbol{\Gamma}(\boldsymbol{X}) - B$, if there were the information of $\boldsymbol{R}^{(1)}(u)$ and $\boldsymbol{R}^{(2)}(u)$ beforehand. Although this situation seems unrealistic, a spectral density of Gaussian process with a recursive structure of autocovariances is usually considered in demonstrative research.

For the consistency of $\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}}$, we impose the following assumptions.

Assumption 2.1.4 ($\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}}$) *Suppose either of the following conditions is satisfied:*

(I) p *is finite and* $n_1, n \to \infty$ ($\eta \geq 0$)*;*

(II-1) $\{\boldsymbol{X}(t)\}$ *has all vanishing third and fourth cumulants,* $p \to \infty$, *and*

- *(i) if* $\eta > 1/2$, *then, both* n_1 *and* n *are finite or infinite,*
- *(ii) if* $\eta = 1/2$, *then,* $n_1, n \to \infty$,
- *(iii) if* $0 \leq \eta < 1/2$, *then,* $n_1, n \to \infty$, *such that* $p = o\big((n_1 n)^{2/(1-2\eta)}/(n_1^{2/(1-2\eta)} + n^{2/(1-2\eta)})\big)$;

(II-2) $\{\boldsymbol{X}(t)\}$ *has the non-vanishing third or fourth cumulants,* $p \to \infty$, *and*

- *(i) if* $\eta > 1$, *then, both* n_1 *and* n *are finite or infinite,*
- *(ii) if* $\eta = 1$, *then,* $n_1 \to \infty$ *and* n *is finite or infinite,*
- *(iii) if* $1/2 < \eta < 1$, *then,* $n_1 \to \infty$ *and* n *is finite or infinite, such that* $p = o(n_1^{4/(3-3\eta)})$,
- *(iv) if* $\eta = 1/2$, *then,* $n_1, n \to \infty$, *such that* $p = o(n_1^{8/3})$,
- *(v) if* $0 \leq \eta < 1/2$, *then,* $n_1, n \to \infty$, *such that* $p = o\big((n_1^{4/(3-3\eta)} n^{2/(1-2\eta)})/(n_1^{4/(3-3\eta)} + n^{2/(1-2\eta)})\big)$.

We provide the following result for $\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}}$.

Theorem 2.1.2 *Under Assumptions 2.1.1, 2.1.2 and 2.1.4,* $\mathbf{\Gamma}(\boldsymbol{X})_{\text{mod}}$ *is a consistent classifier.*

We turn to the jackknife type bias adjusted classifier. The jackknife estimates $\hat{\mathbf{\Sigma}}^{(k)}_{\text{Jack}}$ of $\text{Var}(\bar{\boldsymbol{X}}) = n_k^{-1}\sum_{u=1-n_k}^{n_k-1}(1-|u|/n_k)\boldsymbol{R}^{(k)}(u)$ for $k=1,2$ can be defined as the multivariate analogue of Theorem 3.1 of Künsch (1989), as follows:

$$\hat{\mathbf{\Sigma}}^{(k)}_{\text{Jack}}(u) = {n_k}^{-1}\sum_{u=1-l_k}^{l_k-1}\nu_{n_k}(u)/\nu_{n_k}(0)\hat{\boldsymbol{R}}^{(k)}(u),$$

$$\hat{\boldsymbol{R}}^{(k)}(u) = \sum_{t=1}^{n_k-|u|}\beta_{n_k}(t,u)(\boldsymbol{X}^{(k)}(t)-\hat{\boldsymbol{\mu}}^{(k)})(\boldsymbol{X}^{(k)}(t+|u|)-\hat{\boldsymbol{\mu}}^{(k)})',$$

$$\hat{\boldsymbol{\mu}}^{(k)} = \sum_{t=1}^{n_k}\alpha_{n_k}(t)\boldsymbol{X}^{(k)}(t),$$

where l_k is the length of the downweighted block, such that $l_k = l_k(n_k) \to \infty$ as $n_k \to \infty$. Here the weight functions are defined by

$$\alpha_{n_k}(t) = \frac{1}{\sum_{i=1}^{l_k}\omega_{n_k}(i)}(n_k-l_k+1)^{-1}\sum_{j=0}^{n_k-l_k}\omega_{n_k}(t-j),$$

$$\beta_{n_k}(t,u) = \nu_{n_k}(u)^{-1}(n_k-l_k+1)^{-1}\sum_{j=0}^{n_k-l_k}\omega_{n_k}(t-j)\omega_{n_k}(t+|u|-j), \quad |u|<l_{n_k},$$

$$\nu_{n_k}(u) = \sum_{j=1}^{l_k-|u|}\omega_{n_k}(j)\omega_{n_k}(j+|u|),$$

for $k=1,2$, where

$$\omega_{n_k}(i) = h\Big(\Big(i-\frac{1}{2}\Big)\Big/l_k\Big), \quad 1\le i\le l_k,$$

with kernel function $h(x)$ satisfying (i) $h:(0,1)\to(0,1)$, (ii) symmetric about $x=\frac{1}{2}$, (iii) increasing in a wide sense. We consider the jackknife type bias adjusted statistic

$$\mathbf{\Gamma}(\boldsymbol{X})_{\text{Jack}} = \mathbf{\Gamma}(\boldsymbol{X}) - \frac{1}{2}\text{tr}\hat{\mathbf{\Sigma}}^{(1)}_{\text{Jack}} + \frac{1}{2}\text{tr}\hat{\mathbf{\Sigma}}^{(2)}_{\text{Jack}},$$

under some combinations of the following four assumptions:

(B1) $h(x) = \mathbb{1}_{(0,1)}(x)$, and $\sum_{k=1}^{2}\sum_{u=-\infty}^{\infty}|u||R^{(k)}_{a_1a_2}(u)| < \infty$ for $a_1,a_2 = 1,\dots,p$.

(B2) $h{*}h$ is twice continuously differentiable around zero, and $\sum_{k=1}^{2}\sum_{u=-\infty}^{\infty} u^2|R_{a_1a_2}^{(k)}(u)| < \infty$ for $a_1, a_2 = 1, \dots, p$.
(B3) $E[|X_j^{(k)}(t)|^{6+\delta}] < \infty$, $k = 1, 2$, $j = 1, \dots, p$ for some $\delta > 0$.
(B4)$_\alpha$ $l_k = c_k' n_k^\alpha$, $k = 1, 2$ for some constants $c_k' > 0$.

Using Lemmas 2.1.1 and 2.1.2, the asymptotic properties of Γ_{Jack} are readily available as the multivariate extension of Theorem 3.2 of Künsch (1989), noting that

$$\mathrm{Var}(\Gamma_{\text{Jack}}) \le 3\Big[\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})) + \frac{1}{4}\mathrm{Var}(\mathrm{tr}\hat{\boldsymbol{\Sigma}}_{\text{Jack}}^{(1)}) + \frac{1}{4}\mathrm{tr}\hat{\boldsymbol{\Sigma}}_{\text{Jack}}^{(2)}\Big],$$

where, for $l = 1, 2$,

$$\mathrm{Var}(\mathrm{tr}\hat{\boldsymbol{\Sigma}}_{\text{Jack}}^{(l)}) \le \sum_{j=1}^{p}\mathrm{Var}(\hat{\Sigma}_{\text{Jack},jj}^{(l)}) + 2\sum_{j>k}\big[\mathrm{Var}(\hat{\Sigma}_{\text{Jack},jj}^{(l)})\mathrm{Var}(\hat{\Sigma}_{\text{Jack},kk}^{(l)})\big]^{1/2}.$$

Lemma 2.1.3 *Assumption 2.1.1 (i) holds. Assume that $n_1 \to \infty$ and $l_k = l_k(n_k) \to \infty$ for $k = 1, 2$. When $\boldsymbol{X}$ belongs to π_i, it holds that*

(i) if (B1) holds and $l_k = o(n_k^{1/2})$ for $k = 1, 2$, then

$$E(\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{Jack}}) = \frac{(-1)^i}{2}\Delta + O(l_1^{-1} n_1^{-1} p).$$

(ii) if (B2) holds and $l_k = o(n_k^{1/3})$ for $k = 1, 2$, then

$$E(\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{Jack}}) = \frac{(-1)^i}{2}\Delta + O(l_1^{-2} n_1^{-1} p).$$

(iii) if (B3) holds and $l_k = o(n_k)$ for $k = 1, 2$, then

$$\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{Jack}}) = \mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})) + O(l_1 n_1^{-3} p^2).$$

For the consistency of Γ_{Jack}, we impose the following assumptions.

Assumption 2.1.5 (Γ_{Jack}) *$n_1 \to \infty$, and either of the following conditions is satisfied:*

(I) p is finite and $n \to \infty$ ($\eta \ge 0$);

(II-B1) (B1) and (B4)$_\alpha$ hold with $0 < \alpha < 1/2$, $p \to \infty$, and

(i) if $\eta \ge 1$, then, n is finite or infinite,

(ii) *if* $1/2 < \eta < 1$*, then,* $p = o(n_1^{\iota_1(\alpha)/(1-\eta)})$ *and* n *is finite or infinite,*

(iii) *if* $\eta = 1/2$*, then,* $p = o(n_1^{2\iota_1(\alpha)})$ *and* $n \to \infty$*,*

(iv) *if* $0 \leq \eta < 1/2$*, then,* $n_1, n \to \infty$*, such that* $p = o\big((n_1^{\iota_1(\alpha)/(1-\eta)} n^{2/(1-2\eta)})/(n_1^{\iota_1(\alpha)/(1-\eta)} + n^{2/(1-2\eta)})\big)$*,*

where $\iota_1(\alpha) = \min\{(3-\alpha)/2, 1+\alpha\}$*;*

(II-B2-1) $\{\boldsymbol{X}^{(k)}(t)\}$ *has all vanishing third and fourth cumulants for* $k = 1, 2$*, (B2) and* $(B4)_\alpha$ *hold with* $0 < \alpha < 1/3$*,* $p \to \infty$*, and*

(i) *if* $\eta \geq 1$*, then,* n *is finite or infinite,*

(ii) *if* $1/2 < \eta < 1$*, then,* $p = o(n_1^{\iota_2(\alpha)/(1-\eta)})$ *and* n *is finite or infinite,*

(iii) *if* $\eta = 1/2$*, then,* $p = o(n_1^{2\iota_2(\alpha)})$ *and* $n \to \infty$*,*

(iv) *if* $0 \leq \eta < 1/2$*, then,* $n_1, n \to \infty$*, such that* $p = o\big((n_1^{\iota_2(\alpha)/(1-\eta)} n^{2/(1-2\eta)})/(n_1^{\iota_2(\alpha)/(1-\eta)} + n^{2/(1-2\eta)})\big)$*,*

where $\iota_2(\alpha) = \min\{(3-\alpha)/2, 1+2\alpha\}$*;*

(II-B2-2) $\{\boldsymbol{X}^{(k)}(t)\}$ *has the non-vanishing third or fourth cumulants for* $k = 1, 2$*, (B2) and* $(B4)_\alpha$ *hold with* $0 < \alpha < 1/3$*,* $p \to \infty$*, and*

(i) *if* $\eta \geq 1$*, then,* n *is finite or infinite,*

(ii) *if* $1/2 < \eta < 1$*, then,* $p = o(n_1^{\iota_3(\alpha)/(1-\eta)})$ *and* n *is finite or infinite,*

(iii) *if* $\eta = 1/2$*, then,* $p = o(n_1^{\iota_3(\alpha)/3})$ *and* $n \to \infty$*,*

(iv) *if* $0 \leq \eta < 1/2$*, then,* $n_1, n \to \infty$*, such that* $p = o\big((n_1^{\iota_3(\alpha)/(1-\eta)} n^{2/(1-2\eta)})/(n_1^{\iota_3(\alpha)/(1-\eta)} + n^{2/(1-2\eta)})\big)$*,*

where $\iota_3(\alpha) = \min\{4/3, 1+2\alpha\}$*.*

Theorem 2.1.3 *Under Assumptions 2.1.1, 2.1.2 and 2.1.5,* Γ_{Jack} *is a consistent classifier.*

Remark 2.1.2 *The jackknife type bias adjusted classifier having consistency for higher dimensional data is preferable. We can determine the optimal order of* l_k *by considering preferable jackknife type bias adjusted classifier. In case (II-B1),* $\alpha = 1/3$ *is preferable; in case (II-B2-1),* $\alpha = 1/5$ *is preferable; in case (II-B2-2),* $1/6 \leq \alpha < 1/3$ *is preferable.*

Remark 2.1.3 *The assumption for consistency is improved for higher dimension when (B2) holds than when (B1) holds, if we compare Assumption 2.1.5 (II-B2) with (II-B1). The difference between non-Gaussian process and Gaussian process can be seen if we compare (II-B2-2) with (II-B2-1).*

Remark 2.1.4 *Assumptions 2.1.3 - 2.1.5 may be very technical. Let us consider two special cases (a) n is finite, (b) $n = c'n_1$ for some constant $c' > 0$.*

In case (a), a necessary condition for the consistency of $\boldsymbol{\Gamma}(\boldsymbol{X})$, $\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}}$ or Γ_{Jack} is $\eta > 1/2$, where η is the order of squared distance between mean vectors $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ (see Assumption 2.1.1 (ii)). Although in the ideal case (II-1), there is not any restriction for the dimension p, when $\eta < 1$, p must satisfy $p = o(n_1^{1/(1-\eta)})$, $p = o(n_1^{4/(3-3\eta)})$, $p = o(n_1^{\iota_1(\alpha)/(1-\eta)})$, $p = o(n_1^{\iota_2(\alpha)/(1-\eta)})$, $p = o(n_1^{\iota_3(\alpha)/(1-\eta)})$ in case (II), (II-2), (II-B1), (II-B2-1), (II-B2-2), respectively. Note that $1 < \iota_1(\alpha) < \iota_3(\alpha) \leq \iota_2(\alpha) \leq 4/3$ when $0 < \alpha \leq 1/6$ and $1 < \iota_1(\alpha) < \iota_3(\alpha) = 4/3 < \iota_2(\alpha)$ when $1/6 < \alpha < 1/3$. In consequence, if $\{\boldsymbol{X}^{(k)}(t)\}$ has the non-vanishing third or fourth cumulants for $k = 1, 2$, Assumption 2.1.5 improves Assumption 2.1.3, and attains Assumption 2.1.4 when (B2) holds and $1/6 \leq \alpha < 1/3$, as $n_1 \to \infty$. Note that Assumption 2.1.5 attains Assumption 2.1.4 when (B1) holds only if $\alpha = 1/3$. On the other hand, if $\{\boldsymbol{X}^{(k)}(t)\}$ has all vanishing third and fourth cumulants for $k = 1, 2$, Assumption 2.1.5 does not attain Assumption 2.1.4 but always improves Assumption 2.1.3, as $n_1 \to \infty$.

In case (b), all three classifiers $\boldsymbol{\Gamma}(\boldsymbol{X})$, $\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{mod}}$ and Γ_{Jack} work for finite p. As $p \to \infty$ and $0 \leq \eta \leq 1/2$, p must satisfy $p = o(n_1^{1/(1-\eta)})$, $p = o(n_1^{2/(1-2\eta)})$, $p = o(n_1^{4/(3-3\eta)})$, $p = o(n_1^{\iota_1(\alpha)/(1-\eta)})$, $p = o(n_1^{\iota_2(\alpha)/(1-\eta)})$, $p = o(n_1^{\iota_3(\alpha)/(1-\eta)})$ in case (II), (II-1), (II-2), (II-B1), (II-B2-1), (II-B2-2), respectively. Consequently, if $\{\boldsymbol{X}^{(k)}(t)\}$ has the non-vanishing third or fourth cumulants for $k = 1, 2$, Assumption 2.1.5 improves Assumption 2.1.3 and attains Assumption 2.1.4 when (B2) holds and $1/6 \leq \alpha < 1/3$, as $n_1 \to \infty$. Note that $\iota_2(\alpha)/(1-\eta) < 2/(1-2\eta)$ for $0 < \alpha < 1/3$. Assumption 2.1.5 does not attain Assumption 2.1.4 but always improves Assumption 2.1.3 if $\{\boldsymbol{X}^{(k)}(t)\}$ has all vanishing third and fourth cumulants for $k = 1, 2$, as $n_1 \to \infty$.

2.2 Simulation Studies

To investigate the performance of the statistics considered in Section 2.1, we compared the misclassification rates of the following five discriminant statistics: $\boldsymbol{\Gamma}(\boldsymbol{X})_1 =$

$\mathbf{\Gamma}(\boldsymbol{X})$;

$$\begin{aligned}
\mathbf{\Gamma}(\boldsymbol{X})_2 &= \mathbf{\Gamma}(\boldsymbol{X})_1 - \frac{1}{2n_1}\mathrm{tr}\,\boldsymbol{S}^{(1)} + \frac{1}{2n_2}\mathrm{tr}\,\boldsymbol{S}^{(2)};\\
\mathbf{\Gamma}(\boldsymbol{X})_3 &= \mathbf{\Gamma}(\boldsymbol{X})_1 - B;\\
\mathbf{\Gamma}(\boldsymbol{X})_4 &= \Gamma_1 - \frac{1}{2}\mathrm{tr}\hat{\mathbf{\Sigma}}^{(1)}_{\mathrm{Jack1}} + \frac{1}{2}\mathrm{tr}\hat{\mathbf{\Sigma}}^{(2)}_{\mathrm{Jack1}}, \quad h_1(x) = 1_{(0,1)}(x);\\
\mathbf{\Gamma}(\boldsymbol{X})_5 &= \Gamma_1 - \frac{1}{2}\mathrm{tr}\hat{\mathbf{\Sigma}}^{(1)}_{\mathrm{Jack1}} + \frac{1}{2}\mathrm{tr}\hat{\mathbf{\Sigma}}^{(2)}_{\mathrm{Jack2}}, \quad h_2(x) = \frac{1}{2}\{1-\cos(\pi x)\}.
\end{aligned}$$

Chan and Hall (2009) discussed Γ_1 and Γ_2, where $\Gamma_1 = \Gamma$ is a statistic without bias correction, and $\mathbf{\Gamma}(\boldsymbol{X})_2$ is a statistic after the bias correction in the i.i.d. case. We consider an infeasible statistic $\mathbf{\Gamma}(\boldsymbol{X})_3 = \Gamma_{\mathrm{mod}}$ for the stationary case, that removes the term B. As a feasible version, we thus introduce two statistics $\mathbf{\Gamma}(\boldsymbol{X})_4$ and $\mathbf{\Gamma}(\boldsymbol{X})_5$, on the basis of Künsch (1989)'s jackknife bias adjustment.

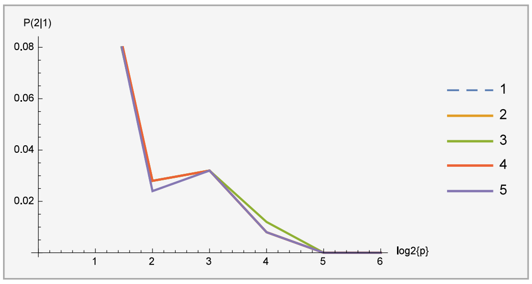
We used moving average (MA) model, autoregressive (AR) model and autoregressive moving average (ARMA) model for our simulation study. Gaussian MA(1), AR(1), and ARMA(1,1) processes are given by $\boldsymbol{X}^{(i)}(t) = \boldsymbol{\mu}^{(i)} + \boldsymbol{\epsilon}_t + \mathbf{\Theta}_1\boldsymbol{\epsilon}_{t-1}$, $\boldsymbol{X}^{(i)}(t) - \Xi_1\boldsymbol{X}^{(i)}(t-1) = (\boldsymbol{I}_p - \Xi_1)\,\boldsymbol{\mu}^{(i)} + \boldsymbol{\epsilon}_t$, and $\boldsymbol{X}^{(i)}(t) - \Xi_1\boldsymbol{X}^{(i)}(t-1) = (\boldsymbol{I}_p - \Xi_1)\,\boldsymbol{\mu}^{(i)} + \boldsymbol{\epsilon}_t + \mathbf{\Phi}_1\boldsymbol{\epsilon}_{t-1}$, respectively, where $\boldsymbol{\mu}^{(1)} = \mathbf{0}_p$, while $\boldsymbol{\mu}^{(2)}$ is p-dimensional vector such that the first $\lfloor p^{2/3} \rfloor$ elements are 1 and 0 else, $\mathbf{\Theta}_1 = ((p-1)/p)\,\boldsymbol{I}_p$, and $\mathbf{\Xi}_1 = \mathbf{\Phi}_1 = (1/p)\,\boldsymbol{I}_p$, where $\boldsymbol{I}_p$ denotes the $p \times p$ identity matrix. The covariance matrix of the innovation process $\{\boldsymbol{\epsilon}_t\}$ in the category π_i is $\mathbf{\Sigma}^{(i)}_{jk} = 0.1^{(|j-k|^{1/3})}$ for $i = 1, 2$.

case	categories		bias	corrected bias		
	π_1	π_2		Γ_2	Γ_4	Γ_5
(a)	AR(1)	MA(1)	-0.8406	0.2861	0.2906	0.1263
(b)	AR(1)	ARMA(1,1)	-0.0103	-0.0162	-0.0100	-0.0115

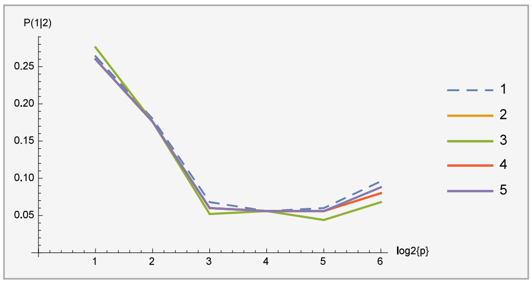
Table 2.1: Simulation settings for each case. The bias is B. Corrected bias is $\mathbf{\Gamma}(\boldsymbol{X})_i - \mathbf{\Gamma}(\boldsymbol{X})_1$ for $i = 2, 4, 5$.

We carried out two numerical simulations (a) and (b) in finite training sample cases. The simulations were repeated for 250 times, where $(n, n_1, n_2) = (100, 10, 10)$. The lengths of the down-weighted blocks were chosen as $l_1 = l_2 = 5$. The bias and corrected biases of $\mathbf{\Gamma}(\boldsymbol{X})_2$, $\mathbf{\Gamma}(\boldsymbol{X})_4$ and $\mathbf{\Gamma}(\boldsymbol{X})_5$ when $p = 64$ are shown in Table 2.1. The misclassification rates when $p = 2^j$, $j = 1, \ldots, 6$, in simulations (a) and (b), are shown in Figure 2.1. Table 2.1 supports Lemma 2.1.3 (i) and (ii). Figure

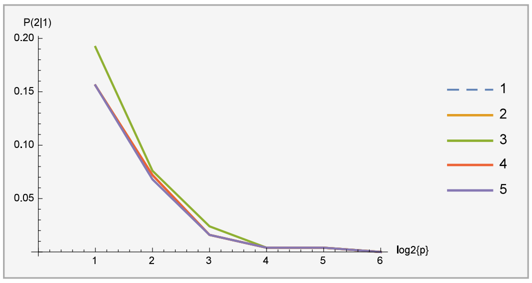
2.1 shows the consistency of discriminant statistics as Theorems 2.1.1 – 2.1.3. The jackknife type adjusted discriminant statistics $\mathbf{\Gamma}(\boldsymbol{X})_4$ and $\mathbf{\Gamma}(\boldsymbol{X})_5$ performed better than the other statistics in these simulations. Other simulation results are given in Supplementary Material.



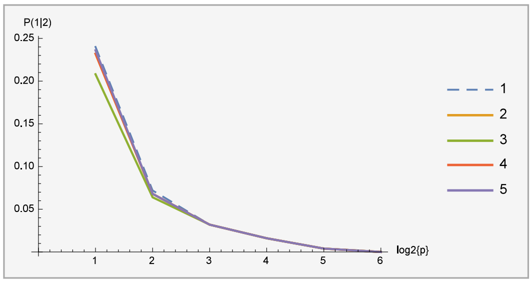
(a) $P(1|2)$ for the statistics 1, 2, 3, 4, and 5



(a) $P(2|1)$ for the statistics 1, 2, 3, 4, and 5



(b) $P(1|2)$ for the statistics 1, 2, 3, 4, and 5



(b) $P(2|1)$ for the statistics 1, 2, 3, 4, and 5

Figure 2.1: The misclassification rates in simulations (a) and (b).

2.3 Real data analysis

As cluster analysis, we used the financial data obtained by "NEEDS-FAME"* database. The data set consists of 15 cell lines and 42 dimension, which are 42 pieces of accounting information (balance sheet, profit and loss statement, cash flow statement, etc.) of companies listed with first and second sections of the Tokyo Stock Exchange in these 15 years. Here, we summarized some of these companies in Ta-

*NEEDS-FAME is a database of financial data. Waseda University has contracted with the company Nikkei Media Marketing, Inc. (http://www.nikkeimm.co.jp/solution/needs-fame/) to use the data.

ble 2.2. Our analysis is to make a dendrogram of these companies since dendrogram gives a visual representation of the hierarchical cluster. To classify this data

The first section S_1
denso, toyota, panasonic, sharp, hitachi, sony, canon, nissan, mazda, kyocera, ntt, nttdocomo, nikon, etc.

The second section S_2
mitani, chuogyorui, nihonseiki, maxvalutokai, daitogyorui, kitamura, sbshd, sbshokuhin, vitec, kansaisupermarket, etc.

Table 2.2: Companies of the first and second sections.

$\boldsymbol{X}^{[j]} = \{\boldsymbol{X}^{(j)}(1), \ldots, \boldsymbol{X}^{(j)}(15)\}$, $j = 1, \ldots, 42$, we computed the following disparity:

$$C(\boldsymbol{X}^{[j_1]}, \boldsymbol{X}^{[j_2]}) = (\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)})'(\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)}) + \left|\mathrm{tr}\hat{\boldsymbol{\Sigma}}^{(j_1)}_{\mathrm{Jack}} - \mathrm{tr}\hat{\boldsymbol{\Sigma}}^{(j_2)}_{\mathrm{Jack}}\right|, \qquad (2.2)$$

where $\boldsymbol{X}_i = \bar{\boldsymbol{X}}^{(i)}$, and $\hat{\boldsymbol{\Sigma}}^{(k)}_{\mathrm{Jack}}$ is defined by $\hat{\boldsymbol{\Sigma}}^{(k)}_{\mathrm{Jack}}(u) = n_k{}^{-1} \sum_{u=1-l_k}^{l_k-1} \nu_{n_k}(u)/\nu_{n_k}(0)\hat{\boldsymbol{R}}^{(k)}(u)$ with kernel function $h_1(x) = \mathbb{1}_{(0,1)}(x)$. The lengths of the down-weighted blocks were set as $l_k = 2$. Clearly, the measure (2.2) satisfies the conditions (i) $C(\boldsymbol{X}, \boldsymbol{X}) = 0$ and (ii) $C(\boldsymbol{X}, \boldsymbol{Y}) = C(\boldsymbol{Y}, \boldsymbol{X})$. The motivation of the measure (2.2) aims at the classification of $\{\boldsymbol{\mu}^{(j)}, \mathrm{tr}\boldsymbol{f}^{(j)}(0)\}$, where $\boldsymbol{f}^{(j)}$ is the spectral density matrix. We obtained dendrograms shown in Figures 2.2 – 2.4. As can be seen from Figure 2.2, the companies in the second section companies formed an exact crowd (red characters in Figure 2.2). That means that our cluster method by the disparity (2.2) can classify the first section and the second section very well. In Figure 2.3, Toyota and NTT are far apart from other companies. This result might be natural since we think Toyota and NTT are the most representative two companies of all Japanese big companies. We also found Mitani corporation ("mitani" in Figure 2.4) is apart from the other companies. Mitani corporation is the most big company in the second section.

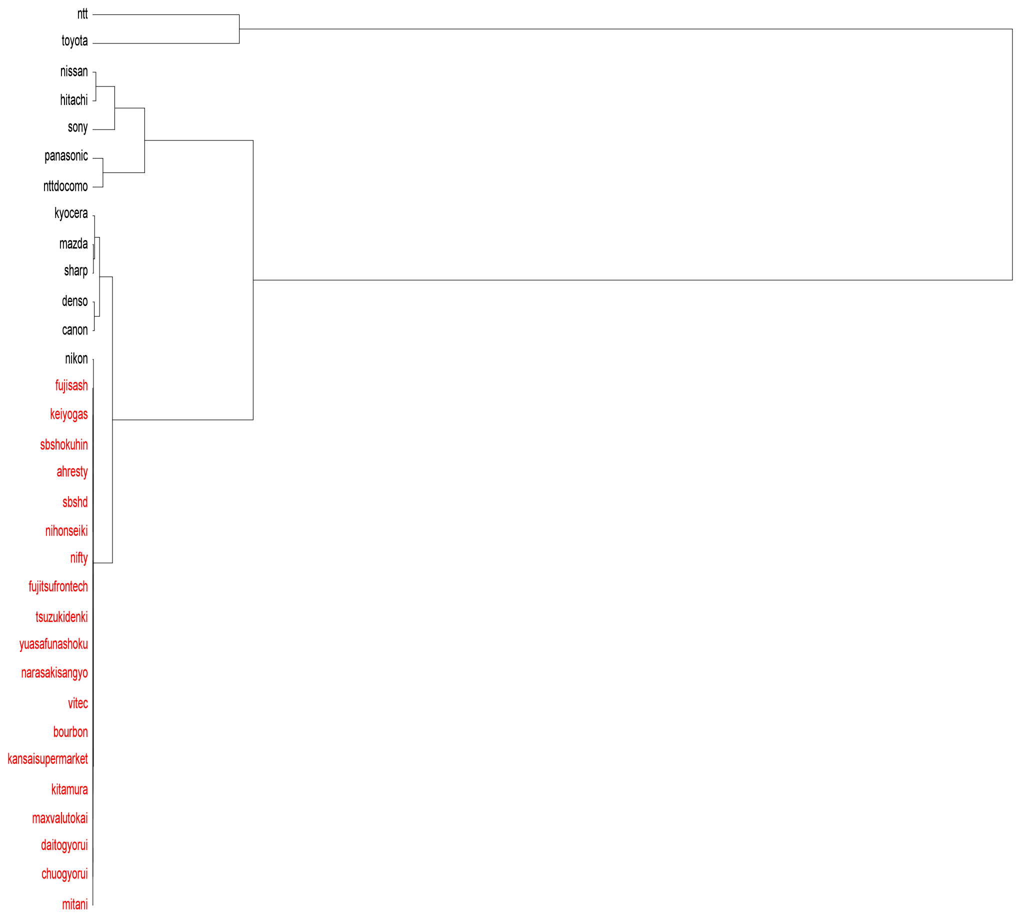


Figure 2.2: The cluster analysis of the first and second sections. The first section companies are in the black. The second section companies are in the red.

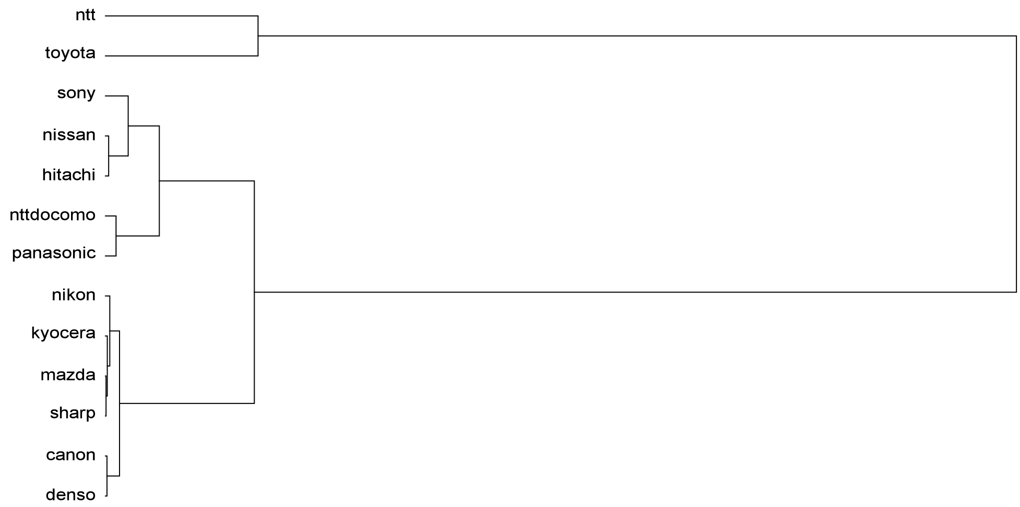


Figure 2.3: The cluster analysis of companies in the first section.

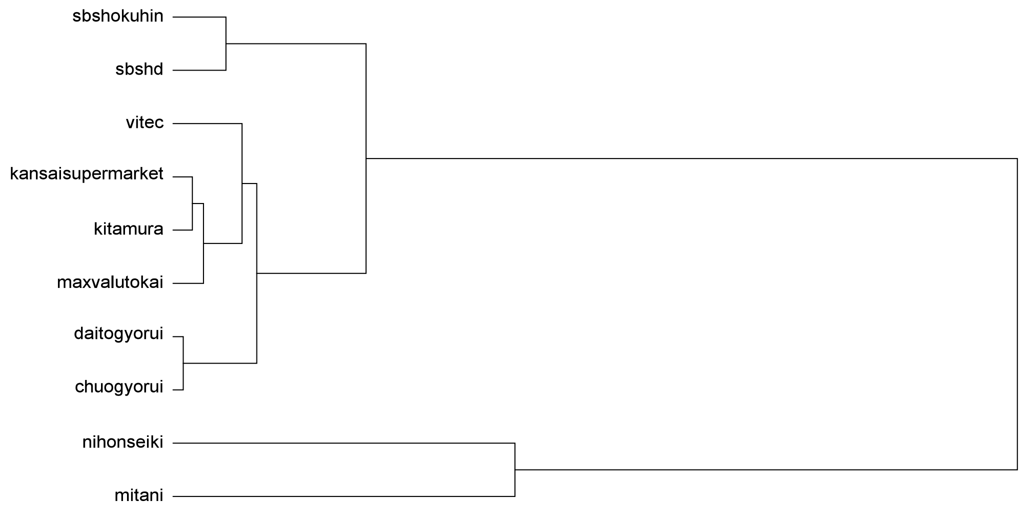


Figure 2.4: The cluster analysis of companies in the second section.

2.4 Appendix

In Appendix, we provide the proofs of the results in Section 2.1. The details are available in Supplementary Material.

Proof (Proof of Lemma 2.1.1) *When* $\boldsymbol{X}$ *belongs to* π_i*, we see that*

$$\begin{aligned}
E\,\boldsymbol{\Gamma}(\boldsymbol{X}) &= EE\,(\boldsymbol{\Gamma}(\boldsymbol{X})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) \\
&= E\Big(\boldsymbol{\mu}^{(i)} - \frac{\bar{\boldsymbol{X}}^{(1)}+\bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)}\Big) \\
&= \frac{(-1)^i}{2}\Delta + B.
\end{aligned}$$

Proof (Proof of Lemma 2.1.2) *Let* $(i,j)=(1,2),(2,1)$*. Suppose that* $\boldsymbol{X}$ *belongs to* π_i*. Then, from the law of total variance, we have*

$$\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})) = E\,\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) + \mathrm{Var}(E\,(\boldsymbol{\Gamma}(\boldsymbol{X})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})). \tag{2.3}$$

It is easy to see that the first term in (2.3) is given by

$$\begin{aligned}
&E\,\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) = E\,\mathrm{Var}\{\bar{\boldsymbol{X}}'(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}\} \\
=\;& E\,(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})'\mathrm{Var}(\bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)}) \\
=\;& \mathrm{tr}\Big\{\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)\Big)\sum_{k=1}^{2}\frac{1}{n_k}\sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\Big)\boldsymbol{R}^{(k)}(u)\Big\} \\
&+\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big).
\end{aligned}$$

The second term in (2.3) *can be evaluated by*

$$\begin{aligned}
&\mathrm{Var}(E\,(\boldsymbol{\Gamma}(\boldsymbol{X})|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})) \\
=\;& \mathrm{Var}\Big\{\Big(\boldsymbol{\mu}^{(i)}-\frac{\bar{\boldsymbol{X}}^{(1)}+\bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)}\Big)\Big\} \\
=\;& \sum_{l=1}^{2}\Big[\mathrm{Var}\Big\{(\boldsymbol{\mu}^{(i)})'\bar{\boldsymbol{X}}^{(l)}\Big\} - \mathrm{Cov}\Big\{(\boldsymbol{\mu}^{(i)})'\bar{\boldsymbol{X}}^{(l)},(\bar{\boldsymbol{X}}^{(l)})'\bar{\boldsymbol{X}}^{(l)}\Big\} + \frac{1}{4}\mathrm{Var}\Big\{(\bar{\boldsymbol{X}}^{(l)})'\bar{\boldsymbol{X}}^{(l)}\Big\}\Big] \\
=\;& \sum_{l=1}^{2}\Big[\boldsymbol{\mu}^{(i)'}\Big\{\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big\}\boldsymbol{\mu}^{(i)}
\end{aligned}$$

$$+\frac{1}{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)^2\Big\}+\boldsymbol{\mu}^{(l)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}$$

$$-2\boldsymbol{\mu}^{(i)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}-\Big\{\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jkk}^{(l)}(t_1-s,t_2-s)\mu_j^{(i)}\Big\}$$

$$+\frac{1}{4n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}\Big\{c_{jjkk}^{(l)}(t_1-s_1,s_2-s_1,t_2-s_1)$$

$$+2\mu_k^{(l)}\Big(c_{jjk}^{(l)}(t_1-s_1,s_2-s_1)\Big)+2\mu_j^{(l)}\Big(c_{jkk}^{(l)}(s_2-s_1,t_2-s_1)\Big)\Big\}\Big]$$

$$=\ (\boldsymbol{\mu}^{(i)}-\boldsymbol{\mu}^{(j)})'\Big\{\frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)\boldsymbol{R}^{(j)}(u)\Big\}(\boldsymbol{\mu}^{(i)}-\boldsymbol{\mu}^{(j)})$$

$$+\sum_{l=1}^{2}\Big[\frac{1}{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)^2\Big\}$$

$$+(-1)^i\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}(\mu_j^{(1)}-\mu_j^{(2)})\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jkk}^{(l)}(t_1-s,t_2-s)$$

$$+\frac{1}{4n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jjkk}^{(l)}(t_1-s_1,s_2-s_1,t_2-s_1)\Big].$$

Proof (Proofs of Theorems 2.1.1 - 2.1.3) *Suppose $\boldsymbol{X}$ belongs to π_i. Using Lemma 2.1.1 - 2.1.3, it suffices to show that $E[\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger/\Delta]=(-1)^i/2+o(1)$ and $\mathrm{Var}[\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger/\Delta]=o(1)$ (hence $\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger/\Delta=(-1)^i/2+o_p(1)$) holds if Assumption 2.1.3 holds for $\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger=\boldsymbol{\Gamma}(\boldsymbol{X})$ (or if Assumption 2.1.4 holds for $\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger=\boldsymbol{\Gamma}(\boldsymbol{X})_{\mathrm{mod}}$ or if Assumption 2.1.5 holds for $\boldsymbol{\Gamma}(\boldsymbol{X})^\dagger=\Gamma_{\mathrm{Jack}}$), provided that Assumptions 2.1.1 and 2.1.2 hold.*

2.5 Supplementary Material

In Supplementary Material, we present the basic theoretical results in Section 2.1 and the other simulation results in Sections 2.2 and 2.3.

2.5.1 Basic theoretical results

First, we give the basic theoretical results, which are listed in Lemma A.

Lemma A 1 *When $\boldsymbol{X}$ belongs to π_i, $(i=1,2)$, for any $t=1,\dots n$,*

(i) $E\boldsymbol{X}(t) = \boldsymbol{\mu}^{(i)}$.

(ii) $E\bar{\boldsymbol{X}} = \boldsymbol{\mu}^{(i)}$.

(iii) $\mathrm{Var}\,\boldsymbol{X}(t) = \boldsymbol{R}^{(i)}(0)$.

(iv) $\mathrm{Var}\,\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)$.
For $i, j = 1, 2$,

(v) $E\bar{\boldsymbol{X}}^{(i)} = \boldsymbol{\mu}^{(i)}$.

(vi) $\mathrm{Cov}(\bar{\boldsymbol{X}}^{(i)}, \bar{\boldsymbol{X}}^{(j)}) = \frac{1}{n_i}\sum_{u=1-n_i}^{n_i-1}\Big(1-\frac{|u|}{n_i}\Big)\boldsymbol{R}^{(i)}(u)\delta(i,j)$.

(vii) *Under* π_j $(j=1,2)$, $\mathrm{Cov}(\boldsymbol{X}(t), \bar{\boldsymbol{X}}^{(i)}) = \frac{1}{n_i}\sum_{s=1}^{n_i}\boldsymbol{R}^{(i)}(s-t)\delta(i,j)$.

(viii) $E(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}) = \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}$.

(ix) $\mathrm{Var}(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}) = \sum_{i=1}^{2}\frac{1}{n_i}\sum_{u=1-n_i}^{n_i-1}\Big(1-\frac{|u|}{n_i}\Big)\boldsymbol{R}^{(i)}(u)$.

(x) *Also,*

$$E(\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)})'(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}) = \Big(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}\Big)'\Big(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\Big)$$
$$+\frac{1}{n_2}\sum_{u=1-n_2}^{n_2-1}\Big(1-\frac{|u|}{n_2}\Big)\mathrm{tr}\boldsymbol{R}^{(2)}(u) - \frac{1}{n_1}\sum_{u=1-n_1}^{n_1-1}\Big(1-\frac{|u|}{n_1}\Big)\mathrm{tr}\boldsymbol{R}^{(1)}(u).$$

(xi) *Further,*

$$\begin{aligned}
&\mathrm{Var}(\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)})'(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}) \\
= \;& \sum_{l=1}^{2}\frac{1}{n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}\Big\{c_{jjkk}^{(l)}(t_1-s_1, s_2-s_1, t_2-s_1) \\
&+\mu_k^{(l)}\Big(c_{jjk}^{(l)}(t_1-s_1, s_2-s_1) + c_{jjk}^{(l)}(t_1-s_1, t_2-t_1)\Big) \\
&+\mu_j^{(l)}\Big(c_{jkk}^{(l)}(s_2-s_1, t_2-s_1) + c_{jkk}^{(l)}(s_2-t_1, t_2-t_1)\Big) \\
&+R_{jk}^{(l)}(s_2-s_1)R_{jk}^{(l)}(t_2-t_1) + R_{jk}^{(l)}(t_2-s_1)R_{jk}^{(l)}(s_2-t_1) \\
&+R_{jk}^{(l)}(s_2-s_1)\mu_j^{(l)}\mu_k^{(l)} + R_{jk}^{(l)}(t_2-t_1)\mu_j^{(l)}\mu_k^{(l)} + R_{jk}^{(l)}(t_2-s_1)\mu_j^{(l)}\mu_k^{(l)} + R_{jk}^{(l)}(s_2-t_1)\mu_j^{(l)}\mu_k^{(l)}\Big\}.
\end{aligned}$$

In Gaussian case, we can simplify the equation by

$$\begin{aligned}&\mathrm{Var}(\bar{\boldsymbol{X}}^{(1)}+\bar{\boldsymbol{X}}^{(2)})'(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})\\ =\ & 2\sum_{l=1}^{2}\Big[\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)^2\Big\}+2\boldsymbol{\mu}^{(l)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}\Big\}\Big].\end{aligned}$$

(xii) *In addition,*

$$\begin{aligned}&\mathrm{Cov}\Big(\boldsymbol{\mu}^{(i)'}(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)}),(\bar{\boldsymbol{X}}^{(1)}+\bar{\boldsymbol{X}}^{(2)})'(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})\Big)\\ =\ & \sum_{l=1}^{2}\Big\{\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jkk}^{(l)}(t_1-s,t_2-s)\mu_j^{(i)}+\frac{2}{n_l}\boldsymbol{\mu}^{(i)\prime}\Big(\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}\Big\}.\end{aligned}$$

In Gaussian case, the equation is equivalent to

$$\begin{aligned}&\mathrm{Cov}\Big(\boldsymbol{\mu}^{(i)'}(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)}),(\bar{\boldsymbol{X}}^{(1)}+\bar{\boldsymbol{X}}^{(2)})'(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})\Big)\\ =\ & 2\sum_{l=1}^{2}\boldsymbol{\mu}^{(i)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}.\end{aligned}$$

2.5.2 Evaluation of each term involved in $\mathrm{Var}(\boldsymbol{\Gamma}(\boldsymbol{X}))$

Let $|\boldsymbol{R}^{(l)}(t)|=\big(|R_{ij}^{(l)}(t)|\big)$ for $l=1,2$. Note that

$$\sum_{k=1}^{2}\frac{1}{nn_k}\mathrm{tr}\Big(\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)\sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\Big)\boldsymbol{R}^{(k)}(u)\Big)=O(n^{-1}n_1^{-1}p),$$

$$\begin{aligned}\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)\boldsymbol{R}^{(i)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big) &\le n^{-1}\Delta\Big[\mathrm{tr}\Big\{\Big(\sum_{t=-\infty}^{\infty}|\boldsymbol{R}^{(i)}(t)|\Big)^2\Big\}\Big]^{1/2}\\ &= O(n^{-1}p^{1/2}\,\Delta),\end{aligned}$$

and similarly,

$$\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)\boldsymbol{R}^{(j)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)=O(n_1^{-1}p^{1/2}\,\Delta),$$

$$\frac{1}{2}\sum_{l=1}^{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)\boldsymbol{R}^{(l)}(u)\Big)^2\Big\} \le \big(\min(n_1,n_2)\big)^{-2}\sum_{l=1}^{2}\mathrm{tr}\Big\{\Big(\sum_{t=-\infty}^{\infty}|\boldsymbol{R}^{(l)}(t)|\Big)^2\Big\}$$
$$= O(n_1^{-2}p).$$

Also, we have

$$\begin{aligned}
&\Big|\sum_{l=1}^{2}(-1)^i\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}(\mu_j^{(1)}-\mu_j^{(2)})\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jkk}^{(l)}(t_1-s,t_2-s)\Big| \\
\le\ & \big(\min(n_1,n_2)\big)^{-3}\sum_{j=1}^{p}\sum_{k=1}^{p}|\mu_j^{(1)}-\mu_j^{(2)}|\sum_{l=1}^{2}\Big(n_l\max_{j,k}\sum_{t_1,t_2=-\infty}^{\infty}|c_{jkk}^{(l)}(t_1,t_2)|\Big) \\
\le\ & \big(\min(n_1,n_2)\big)^{-3}\sum_{k=1}^{p}(p\,\Delta)^{1/2}\sum_{l=1}^{2}\Big(n_l\max_{j,k}\sum_{t_1,t_2=-\infty}^{\infty}|c_{jkk}^{(l)}(t_1,t_2)|\Big) \\
=\ & O(n_1^{-2}p^{3/2}\,\Delta^{1/2}),
\end{aligned}$$

and

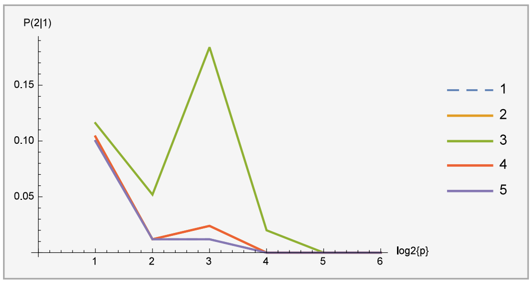
$$\begin{aligned}
&\Big|\sum_{l=1}^{2}\frac{1}{4n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l}c_{jjkk}^{(l)}(t_1-s_1,s_2-s_1,t_2-s_1)\Big| \\
\le\ & \sum_{l=1}^{2}n_l^{-4}p^2\Big(n_l\max_{j,k}\sum_{t_1,t_2,t_3=-\infty}^{\infty}|c_{jjkk}^{(l)}(t_1,t_2,t_3)|\Big) = O(n_1^{-3}p^2).
\end{aligned}$$

2.5.3 Other simulation results by the discriminants statistics $\boldsymbol{\Gamma}(\boldsymbol{X})_1 - \boldsymbol{\Gamma}(\boldsymbol{X})_5$

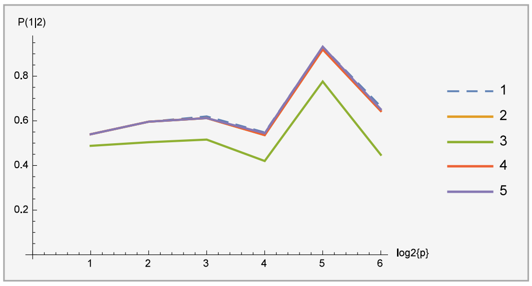
In this section, we give other simulation results by the discriminants statistics $\boldsymbol{\Gamma}(\boldsymbol{X})_1$ – $\boldsymbol{\Gamma}(\boldsymbol{X})_5$. The setting of each case (c), (d) and (e) is given in Table 2.3. AR(1) and ARMA(1,1) are the same as what are defined in Section 2.2. MA(q) model is defined by $\boldsymbol{X}^{(i)}(t) = \boldsymbol{\mu}^{(i)} + \boldsymbol{\epsilon}_t + \sum_{j=1}^{q}\boldsymbol{\Theta}_j\boldsymbol{\epsilon}_{t-j}$, where $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \cdots = \boldsymbol{\Theta}_q = ((p-1)/p)\boldsymbol{I}_p$ for $q = 5, 10$.

case	categories		bias	corrected bias		
	π_1	π_2		Γ_2	Γ_4	Γ_5
(c)	MA(5)	MA(10)	-25.2546	1.53793	1.55901	0.691619
(d)	AR(1)	MA(5)	-10.6733	1.37466	1.38702	0.614256
(e)	AR(1)	ARMA(1,1)	-0.824897	0.242366	0.261391	0.10194

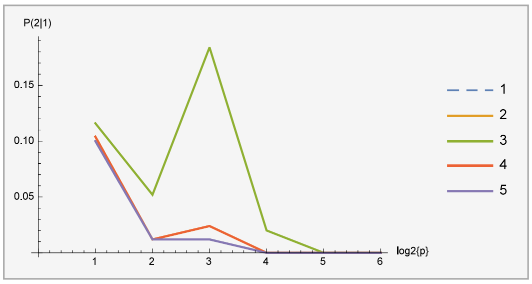
Table 2.3: Simulation settings for each case (c), (d) and (e). $\boldsymbol{\Theta}_1$, $\boldsymbol{\Xi}_1$ are diagonal matrices $(1/p)\,\boldsymbol{I}_p$ and $\boldsymbol{\Phi}_1 = ((1-\sqrt{p})/\sqrt{p})\,\boldsymbol{I}_p$.



(c) $P(1|2)$ for the statistics 1, 2, 3, 4, and 5

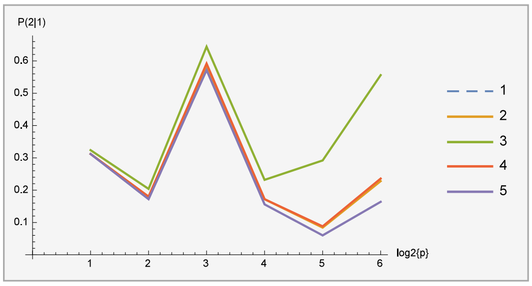


(c) $P(2|1)$ for the statistics 1, 2, 3, 4, and 5

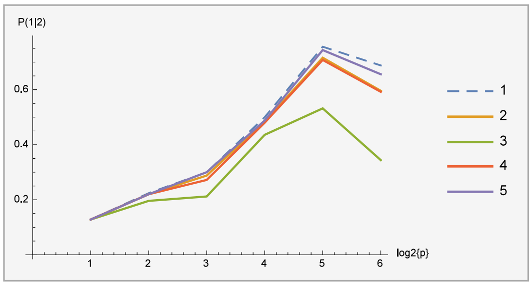


(d) $P(1|2)$ for the statistics 1, 2, 3, 4, and 5

(d) $P(2|1)$ for the statistics 1, 2, 3, 4, and 5



(e) $P(1|2)$ for the statistics 1, 2, 3, 4, and 5



(e) $P(2|1)$ for the statistics 1, 2, 3, 4, and 5

Figure 2.5: The misclassification rates in three simulations (c), (d), and (e).

Chapter 3

Classification for high-dimensional financial time series by a class of disparities

Altman (1968) and Altman and Brenner (1981) used discriminant analysis for a bankruptcy prediction by financial data of companies. Specifically, they constructed a linear discriminant model whose coefficients are estimated by scores calculated from some financial indicators like liquidity assets, equity capital, net income, and so on. The mathematical method as stated above is strongly required for the investigation of credit risk in many banks. Because, if there is such a method, even amateur can decide which bank one can loan.

In mathematical statistics, discriminant analysis is of deciding on the membership of an observed individual to one of a given set of populations. Mainly, for independent samples, discriminant analysis has been developed by e.g., Anderson (2003) and Rao (2009). In addition, discriminant analysis for finite dimensional stationary time series has a history. For example, Taniguchi and Kakizawa (2000) developed both time domain and frequency domain approaches for the discriminant analysis and discussed the problem of discriminating linear processes. Recently in i.i.d. high-dimensional data with dimension $p \to \infty$, Chan and Hall (2009) and Aoshima and Yata (2014) discussed the problem of classification, and proposed scale adjusted-type distance-based classifier. For dependent observations Liu, Nagahata, Uchiyama and Taniguchi (2017) are concerned with high-dimensional stationary process $\{\boldsymbol{X}(t)\}$, which is supposed to belong to one of the two categories;

$$\pi_1 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(1)}(\lambda),$$

$$\pi_2 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(2)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(2)}(\lambda),$$

where $\boldsymbol{\mu}$ and $\boldsymbol{f}(\lambda)$ are the mean vector and spectral density matrix of the process $\{\boldsymbol{X}(t)\}$, respectively. Using the multivariate analogue of the methodology of the jackknife in the stationary observations developed by Künsch (1989), Liu et al. (2017) proposed a new classifier with bias adjustment in time series data. Primarily they discuss the consistency of a distance-based classifier for multivariate stationary time series data where the dimension p is allowed to diverge, under suitable conditions on size of samples and training samples (n, n_1, n_2) and dimension p.

In recent years, stored big data encourages rapid development of discriminant analysis for high-dimensional independent and dependent observations. However applications of the method proposed by Liu et al. (2017) to high-dimensional financial time series are unexplored.

In this chapter, we show an effectiveness of theory in Liu et al. (2017) for high-dimensional financial data.

This chapter is organized as follows. In Section 3.1, we review the misspecification rates of our discriminant statistics. They converge to 0 under some appropriate conditions when $p \to \infty$. In Section 3.2, we examine the cluster analysis for real financial data of companies in the Tokyo Stock Exchange.

3.1 Discriminant theory for high-dimensional time series

Let $\{\boldsymbol{X}(t) = (X_1(t), \ldots, X_p(t))', \, t \in \mathbb{Z}\}$ be a p-dimensional stationary process with mean vector $\boldsymbol{\mu}$, spectral density matrix $\boldsymbol{f}(\lambda)$ and autocovariance matrix function $\boldsymbol{R}(t) = \{R_{ij}(t)\, , \, i, j = 1, \ldots, p\}$. Here the dimension p is allowed to be $p \to \infty$. Suppose that we observe $\boldsymbol{X} = \{\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)\}$ from the stationary process $\{\boldsymbol{X}(t)\}$, which belongs to one of the two categories

$$\pi_1 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(1)}(\lambda), \quad \boldsymbol{R}(t) = \boldsymbol{R}^{(1)}(t),$$

$$\pi_2 : \boldsymbol{\mu} = \boldsymbol{\mu}^{(2)}, \quad \boldsymbol{f}(\lambda) = \boldsymbol{f}^{(2)}(\lambda), \quad \boldsymbol{R}(t) = \boldsymbol{R}^{(2)}(t).$$

Also we have independent training samples $\boldsymbol{X}^{(1)} = \{\boldsymbol{X}^{(1)}(1), \ldots, \boldsymbol{X}^{(1)}(n_1)\}$ and $\boldsymbol{X}^{(2)} = \{\boldsymbol{X}^{(2)}(1), \ldots, \boldsymbol{X}^{(2)}(n_2)\}$ from π_1 and π_2 with size n_1 and n_2, respectively. Write $\Delta \equiv \|\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}\|^2$. Throughout this chapter, the notations E and Var are the expectation and variance with respect to a triplet of $(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$, respectively.

The following sample versions for fundamental quantities are introduced:

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{X}(t), \quad \bar{\boldsymbol{X}}^{(i)} = \frac{1}{n_i}\sum_{t=1}^{n_i}\boldsymbol{X}^{(i)}(t),$$

$$\boldsymbol{S}^{(i)} = \frac{1}{n_i - 1}\sum_{t=1}^{n_i}(\boldsymbol{X}^{(i)}(t) - \bar{\boldsymbol{X}}^{(i)})(\boldsymbol{X}^{(i)}(t) - \bar{\boldsymbol{X}}^{(i)})', \quad i = 1, 2.$$

To classify the time series data $\boldsymbol{X}$, we use the discriminant statistic:

$$\Gamma(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big). \tag{3.1}$$

Then, the classification rule is to classify $\boldsymbol{X}$ into π_1 if $\boldsymbol{\Gamma}(\boldsymbol{X}) < 0$ and into π_2 otherwise. To discuss the asymptotic property of $\boldsymbol{\Gamma}(\boldsymbol{X})$, we impose the following assumptions.

Assumption 3.1.1 *(i) $n_2 = c_0 n_1$ for some constant $c_0 > 0$.*

(ii) There exists $\eta \geq 0$ such that $c_1 p^\eta < \Delta < c_2 p^\eta$ for some constants $c_1 > 0$ and $c_2 > 0$.

Assumption 3.1.2 *(i) The autocovariance matrix function $\boldsymbol{R}^{(l)}(u)$ of the stationary process $\{\boldsymbol{X}^{(l)}(t)\}$ in the category π_l for $l = 1, 2$ satisfies*

$$\sum_{l=1}^{2}\sum_{t=-\infty}^{\infty}|R_{ij}^{(l)}(t)| < \infty$$

uniformly for $i, j = 1, \ldots, p$.

(ii) The third and fourth order of cumulants of the stationary process $\{\boldsymbol{X}^{(l)}(t)\}$ satisfy

$$\sum_{l=1}^{2}\sum_{t_1,t_2=-\infty}^{\infty}|c_{a_1a_2a_3}^{(l)}(t_1, t_2)| < \infty,$$

$$\sum_{l=1}^{2}\sum_{t_1,t_2,t_3=-\infty}^{\infty}|c_{a_1a_2a_3a_4}^{(l)}(t_1, t_2, t_3)| < \infty$$

uniformly for $a_1, a_2, a_3, a_4 = 1, \ldots, p$.

Assumption 3.1.3 *$p \to \infty$, and if $\eta > 1$, then, both n_1 and n are finite or infinite.*

Let $P(i|j)$ be the misclassification rate by the classification statistic (3.1) such that $\boldsymbol{X}$ belonging to π_j is erroneously assigned to π_i $(i \neq j)$. We say that the statistic $\boldsymbol{\Gamma}(\boldsymbol{X})$ is a consistent classifier if $P(i|j) \to 0$ for $(i,j) = (1,2), (2,1)$. So we introduce the following result, which is due to Liu et al. (2017).

Theorem 3.1.1 *Under Assumptions 3.1.1 – 3.1.3,* $\boldsymbol{\Gamma}(\boldsymbol{X})$ *is a consistent classifier.*

Next we introduce a bias adjusted classifier of Jackknife type. The jackknife estimator $\hat{\boldsymbol{\Sigma}}^{(k)}_{\text{Jack}}$ of $\text{Var}(\bar{\boldsymbol{X}}) = n_k^{-1}\sum_{u=1-n_k}^{n_k-1}(1-|u|/n_k)\boldsymbol{R}^{(k)}(u)$ for $k = 1, 2$ can be defined as a multivariate analogue of Theorem 3.1 of Künsch (1989), i.e.,

$$\hat{\boldsymbol{\Sigma}}^{(k)}_{\text{Jack}}(u) = {n_k}^{-1} \sum_{u=1-l_k}^{l_k-1} \nu_{n_k}(u)/\nu_{n_k}(0)\hat{\boldsymbol{R}}^{(k)}(u),$$

$$\hat{\boldsymbol{R}}^{(k)}(u) = \sum_{t=1}^{n_k-|u|} \beta_{n_k}(t,u)(\boldsymbol{X}^{(k)}(t) - \hat{\boldsymbol{\mu}}^{(k)})(\boldsymbol{X}^{(k)}(t+|u|) - \hat{\boldsymbol{\mu}}^{(k)})',$$

$$\hat{\boldsymbol{\mu}}^{(k)} = \sum_{t=1}^{n_k} \alpha_{n_k}(t)\boldsymbol{X}^{(k)}(t),$$

where l_k is the length of downweighted block, such that $l_k = l(n_k) \to \infty$ as $n_k \to \infty$. Here the weight functions are defined by

$$\alpha_{n_k}(t) = \frac{1}{\sum_{i=1}^{l_k}\omega_{n_k}(i)}(n_k - l_k + 1)^{-1}\sum_{j=0}^{n_k-l_k}\omega_{n_k}(t-j),$$

$$\beta_{n_k}(t,u) = \nu_{n_k}(u)^{-1}(n_k - l_k + 1)^{-1}\sum_{j=0}^{n_k-l_k}\omega_{n_k}(t-j)\omega_{n_k}(t+|u|-j), \quad |u| < l_{n_k},$$

$$\nu_{n_k}(u) = \sum_{j=1}^{l_k-|u|}\omega_{n_k}(j)\omega_{n_k}(j+|u|),$$

for $k = 1, 2$, where

$$\omega_{n_k}(i) = h\Big(\Big(i - \frac{1}{2}\Big)\Big)\Big/ l_k\Big), \quad 1 \leq i \leq l_k,$$

with kernel function $h(x)$ satisfying (i) $h : (0,1) \to (0,1)$, (ii) symmetric about $x = \frac{1}{2}$, (iii) increasing in a wide sense. Then the bias adjusted statistic of Jackknife type is

$$\boldsymbol{\Gamma}(\boldsymbol{X})_{\text{Jack}} = \boldsymbol{\Gamma}(\boldsymbol{X}) - \frac{1}{2}\text{tr}\hat{\boldsymbol{\Sigma}}^{(1)}_{\text{Jack}} + \frac{1}{2}\text{tr}\hat{\boldsymbol{\Sigma}}^{(2)}_{\text{Jack}}. \tag{3.2}$$

Further we assume $n_1 \to \infty$, and

Assumption 3.1.4 *For $k = 1, 2$, $E|X_j^{(k)}(t)|^{6+\delta} < \infty$, $j = 1, \ldots, p$ for some $\delta > 0$, and $l_k = c_k' n_k^\alpha$ for some constants $c_k' > 0$ and $0 < \alpha \leq 1$. Further, suppose $p \to \infty$, $\{\boldsymbol{X}^{(k)}(t)\}$ is non-Gaussian with $\sum_{u=-\infty}^{\infty} |u||R_{a_1 a_2}^{(k)}(u)| < \infty$ for $k = 1, 2$, $a_1, a_2 = 1, \ldots, p$, $h(x) = \mathbb{1}_{(0,1)}(x)$ and $1/3 \leq \alpha < 1/2$, and if $\eta \geq 1$, then, n is allowed to be finite or infinite.*

The following result is due to Liu et al. (2017).

Theorem 3.1.2 *Under Assumptions 3.1.1, 3.1.2 and 3.1.4, Γ_{Jack} is a consistent classifier.*

3.2 Application to financial time series

As cluster analysis, we use financial data obtained by "NEEDS-FinancialQUEST"*. Here, we analyze financial data of companies of the first and second sections in Table 3.1.

The data set consists of 21 cell lines and 42 dimension, i.e., 42 financial indicators (balance sheet, profit and loss statement, cash flow statement set at Table 3.2 of Appendix) of companies listed in the Tokyo Stock Exchange in these 21 years. Since there exist missing values in the data set, we use the last observation carried forward (LOCF) which carries the last existing data to the next missing value.

Our analysis is to draw a dendrogram of these companies since dendrogram provides a visual understanding of the hierarchical cluster. Hierarchical clustering is widely used in data analysis.

Specifically, by Theorems 3.1.1 and 3.1.2, (3.1) and (3.2) are the consistent classifiers, so we apply them to cluster analysis. To classify the data $\boldsymbol{X}^{[j]} = \{\boldsymbol{X}^{(j)}(1), \ldots, \boldsymbol{X}^{(j)}(15)\}$ with $\dim\{\boldsymbol{X}^{(j)}(t)\} = p$, $j = 1, \ldots, 14$, we computed the following disparity:

$$C(\boldsymbol{X}^{[j_1]}, \boldsymbol{X}^{[j_2]}) = (\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)})'(\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)}), \tag{3.3}$$

$$C(\boldsymbol{X}^{[j_1]}, \boldsymbol{X}^{[j_2]}) = (\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)})'(\bar{\boldsymbol{X}}^{(j_1)} - \bar{\boldsymbol{X}}^{(j_2)}) + \left|\mathrm{tr}\hat{\boldsymbol{\Sigma}}_{\mathrm{Jack}}^{(j_1)} - \mathrm{tr}\hat{\boldsymbol{\Sigma}}_{\mathrm{Jack}}^{(j_2)}\right|, \tag{3.4}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{Jack}}^{(k)}$ is defined by $\hat{\boldsymbol{\Sigma}}_{\mathrm{Jack}}^{(k)}(u) = {n_k}^{-1} \sum_{u=1-l_k}^{l_k-1} \nu_{n_k}(u)/\nu_{n_k}(0)\hat{\boldsymbol{R}}^{(k)}(u)$ with kernel function $h_1(x) = \mathbb{1}_{(0,1)}(x)$. The lengths of the down-weighted blocks are set as $l_k = 2$. First, we assign two elements of minimum distance into one cluster. In the

*NEEDS-FinancialQUEST is a database of financial data. Waseda University has contracted with the company Nikkei Media Marketing, Inc. (`http://finquest.nikkeidb.or.jp/ver2/ip_waseda/`) to use the data. database.

Companies of the first section
1. SEKISUI HOUSE
2. MORINAGA
3. TORAY INDUSTRIES
4. TAKEDA PHARMACEUTICAL
5. SHOWA SHELL SEKIYU
6. BRIDGESTONE
7. ASAHI GLASS
8. NIPPON STEEL & SUMITOMO METAL
9. CITIZEN WATCH
10. KEISEI ELECTRIC RAILWAY
11. NEC
12. TOSHIBA
Companies of the second section
13. SHARP
14. FDK
15. PIXELA
16. YONEX
17. FUJI FURUKAWA ENGINEERING & CONSTRUCTION
18. BOURBON
19. C.UYEMURA
20. TOA OIL
21. MORISHITA JINTAN
22. NISHIKAWA RUBBER
23. GEOSTR
24. MAXVALU NISHINIHON

Table 3.1.

case of Figure 3.1, the first cluster is {C.UYEMURA, GEOSTR}, and this distance is 2.13641×10^{20}. Next we can iterate this procedure and it is seen that the second cluster is {MORISHITA JINTAN, PIXELA} with distance 4.00659×10^{20}. Furthermore, we define the distance between clusters as the minimum of the distances between each element of the clusters. For instance, the distance between {C.UYEMURA, GEOSTR} and {MORISHITA JINTAN, PIXELA} is given by the minimum of the {C.UYEMURA, MORISHITA JINTAN}, {C.UYEMURA, PIXELA}, {GEOSTR, MORISHITA JINTAN} and {GEOSTR, PIXELA}, and which is 4.62408×10^{20}. Then we iteratively define the distances between all the clusters.

By comparison of Figures 3.1 and 3.2 it is seen that the classification by (3.4) is very good for huge companies. Figure 3.1, by the new disparity (3.4), provides clear classification between NIPPON STEEL & SUMITOMO METAL (written by red character in the figures) and the next cluster (TOSHIBA, NEC). (Actually, NIPPON STEEL & SUMITOMO METAL, TOSHIBA, and NEC are huge companies, although recently TOSHIBA and NEC have become unstable. TOSHIBA's rating was downgraded to CCC^{+} by Standard & Poor's, and NEC went into a economic slump and announced a lot of deficit businesses and sold their own building.) We can understand that this result is supported by a successful reduction in bias of (3.4).

Furthermore, it is interesting to observe that Figure 3.1 shows that (3.4) could classify almost of the first and second section companies. (The second section companies are written by blue character in the figure.) This result is not true for all the first and second section companies but shows an evidence of the good classifier for high dimensional financial data.

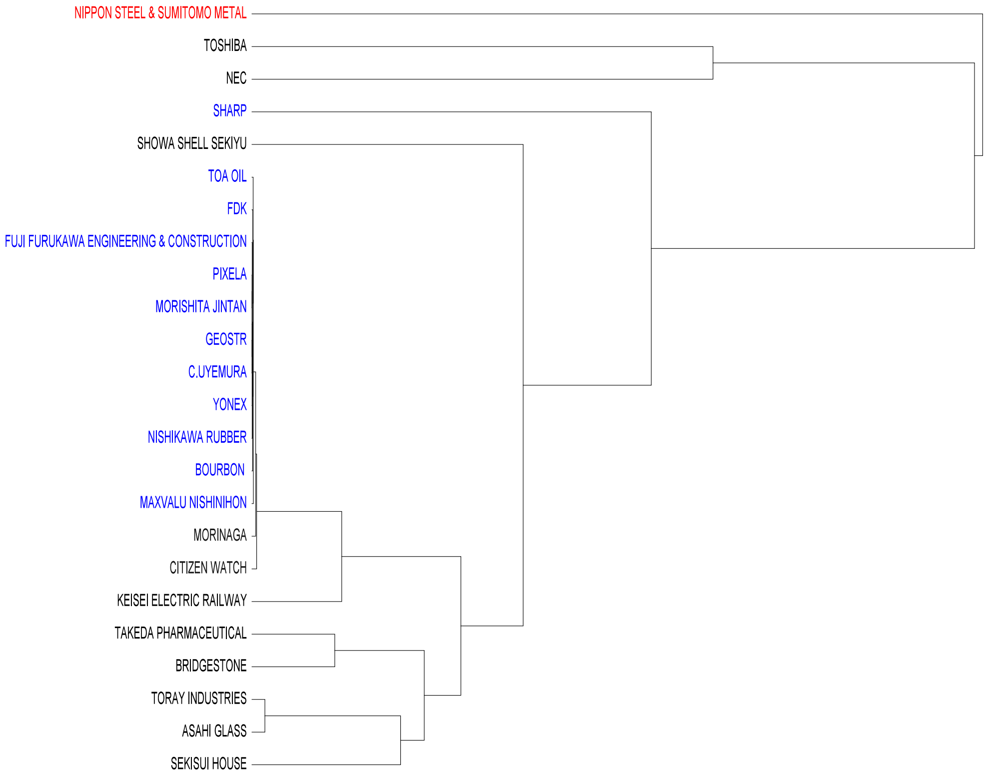


Figure 3.1: The cluster analysis by the new disparity (3.4)

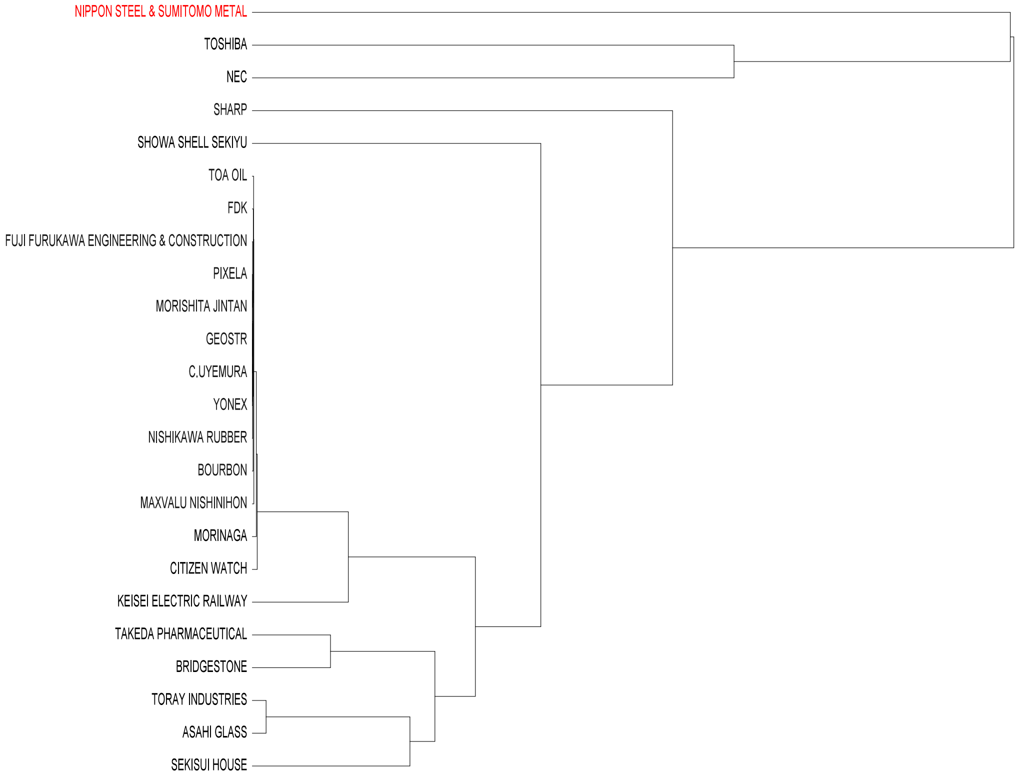


Figure 3.2: The cluster analysis by the disparity (3.3)

Appendix

Financial indicator (The following * means an indicator of Cumulative Total.)
1. Current Assets
2. Cash and Deposit/Cash and Cash Equivalents
3. Notes and Accounts Receivable Trade/Accounts Receivable and Other ShortTerm Claims
4. Accounts Receivable Trade
5. Inventories
6. Raw Materials and Supplies
7. Raw Materials
8. Noncurrent Assets
9. Property, Plant and Equipment
10.Depreciable Property
11.Buildings and Structures
12.Machinery, Equipment and Vehicles
13.Vessels / Vehicles / Delivery Equipment
14.Investments and Other Assets
15.Other Investments and Other Assets/Other Noncurrent Assets
16.Total Assets
17.Current Liabilities
18.Accounts Payable Other and Accrued Expenses
19.Accounts Payable
20.Accrued Expenses
21.Income Taxes Payable
22.Noncurrent Liabilities
23.Total Liabilities
24.Capital Stock
25.Retained Earnings Brought Forward
26.Liabilities and Net Assets (Japanese Standard)/Liabilities, Minority Interests, Shareholders' Equity,Total/Shareholders' Equity and Liabilities,Total
27.Equity Capital (Japanese Standard)/Shareholders' Equity, Total (US GAAP)/Capital of Attributable to Owners of the Parent (IFRS)
28.Net Sales / Operating Revenue*
29.Selling, General and Administrative Expenses*
30.Operating Income*
31.NonOperating Income*
32.Interest and Dividends Income*
33.Interest Income / Discount Revenue / Interest on Securities*
34.Other NonOperating Income*
35.NonOperating Expenses*
36.Interest Expenses / Discount on Notes*
37.Ordinary Income/Income before Income Taxes and Others (IFRS)*
38.Net Income before Income Taxes and Others*
39.Income before Income Taxes and Others*
40.Income Taxes [Cumulative Total]
41.Income TaxesCurrent*
42.Net Income attributable to Parent Company's Shareholders (Consolidated) / Net Income (Unconsolidated)*

Table 3.2: Financial indicators

Chapter 4

Analysis of variance for multivariate time series

Analysis of variance (ANOVA) is a fundamental theory in statistics with a long history. This method is used to test the null hypothesis that the means of three or more populations, or the within-group means, are all equal. In other words, ANOVA indicates whether or not the within-group means are equivalent.

Gauss laid the foundations of this theory in the late 1800s, and Markoff continued this work in the early 1900s. Since these early studies, many test statistics for ANOVA and multivariate ANOVA (MANOVA) have been proposed, but these statistics have mainly focused on the case of independent data. For example, Hooke (1926) and Wishart (1938) applied these test statistics to practical cases, and Bishop (1939) and Box (1949) obtained general theoretical results by deriving asymptotic expansions of the null and non-null distributions of the likelihood ratio test statistic.

Recently, ANOVA has been further developed in many ways. Bai et al. (1990) provided a new ANOVA method of adjusting for the unknown parameter. Liu and Rao (1995) derived the asymptotic distribution of a statistic for the analysis of quadratic entropy (ANOQE) as a generalization of ANOVA, and Rao (2010) investigated some postulates and conditions for ANOQE. Fujikoshi et al. (2011) developed general asymptotic expansions of the null and non-null distributions of the likelihood ratio test, the Lawley-Hotelling test, and the Bartlett-Nanda-Pillai test in some high-dimensional settings. In a time series analysis, Shumway (1971) discussed the asymptotic relationship between the likelihood ratio test and the Lawley-Hotelling test. Brillinger (1973) developed a univariate and balanced ANOVA for time series.

Multivariate time series data is commonly analyzed to solve practical problems in such fields as economics, finance, bioinformatics, and so on, and now, in the era of big data, making statistical decisions with multivariate time series data has become a matter of great concern. We are interested in identifying the within-group means of real financial data that depend on the industry in question. However, as previously stated, the study of ANOVA for dependent data remains nascent even though the analysis for the second-order moments; estimation of autocovariance and spectral density have been well developed.

In this study, we consider the one-way MANOVA model (4.1) whose disturbance processes $\{\boldsymbol{\epsilon}_i\}$ are generated by a stationary process. The remainder of the chapter proceeds as follows. Section 4.1.1 describes our setting. In Section 4.1.2, we discuss the asymptotics of the previous test statistics for dependent observations and derive a sufficient condition for the observations to be asymptotically χ^2-distributed. In Section 4.1.3, we propose a new test statistic based on the Whittle likelihood and show that this statistic is asymptotically χ^2-distributed without the sufficient condition. Sections 4.2.1 and 4.2.2 simulate the classical tests and the proposed test and apply these tests to the daily log data of some stocks to confirm the theoretical results.

4.1 Asymptotic distributions of test statistics for dependent disturbances

4.1.1 Setting and general methods

Let $\boldsymbol{X}_{i1}, \dots \boldsymbol{X}_{in_i}$ be p-dimensional stretches observed from the following the one-way MANOVA model:

$$\boldsymbol{X}_{it} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it},\ t = 1, \dots, n_i,\ i = 1, \dots, q, \tag{4.1}$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ and $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{ip})'$. Here, we assume $\sum_{i=1}^{q} \boldsymbol{\alpha}_i = \mathbf{0}$. In what follows, we assume that the disturbance process $\boldsymbol{\epsilon}_i \equiv \{\boldsymbol{\epsilon}_{it} = (\epsilon_{it}^{(1)}, \cdots, \epsilon_{it}^{(p)})';$ $t = 1, \dots, n_i;\ i = 1, \dots, q\}$ is a stationary process with mean $\mathbf{0}$; common autocovariance matrix $\boldsymbol{\Gamma}(\cdot) = \{\Gamma_{j,k}(\cdot)\,;\, j, k = 1, \dots, p\}$, $(\Gamma_{j,k}(h) \equiv \mathrm{E}(\epsilon_{it}^{(j))} \epsilon_{i,t+h}^{(k)})$; and spectral density matrix $\boldsymbol{f}(\lambda)$. We also assume that the $\boldsymbol{\epsilon}_i,\ i = 1, \dots, q$ are mutually independent. This assumption is a type of homoscedasticity assumption on $\{\boldsymbol{\epsilon}_{it}\}$ that is often applied for a typical multivariate ANOVA (e.g., see Anderson Anderson (2003) Chapters 8 and 9).

Let $\{\boldsymbol{\epsilon}_{it}\}$ be generated from

$$\boldsymbol{\epsilon}_{it} = \sum_{j=0}^{\infty} \boldsymbol{A}(j) \eta_i(t-j),\ \sum_{j=0}^{\infty} \|\boldsymbol{A}(j)\|^2 < \infty,\ i = 1, \dots, q, \tag{4.2}$$

where the p-dimensional random vectors $\boldsymbol{\eta}_i(t) \overset{i.i.d.}{\sim} (\mathbf{0}, \boldsymbol{G})$ and $\boldsymbol{A}(j)$s are $p \times p$ constant matrices.

Then, $\{\boldsymbol{\epsilon}_{it}\}$ has the autocovariance matrix

$$\boldsymbol{\Gamma}(l) = \sum_{j=0}^{\infty} \boldsymbol{A}(j)\boldsymbol{G}\boldsymbol{A}(j+l)'$$

and the spectral density matrix

$$\boldsymbol{f}(\lambda) = \frac{1}{2\pi}\Big\{\sum_{j=0}^{\infty} \boldsymbol{A}(j)e^{\mathrm{i}j\lambda}\Big\}\boldsymbol{G}\Big\{\sum_{j=0}^{\infty} \boldsymbol{A}(j)e^{\mathrm{i}j\lambda}\Big\}^{*}.$$

Generally, this matrix is a complex-valued matrix, but $\boldsymbol{f}(0)$ is a real-valued matrix.

Now, we are interested in testing the hypothesis

$$H : \boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_q. \tag{4.3}$$

Under hypothesis (4.3), we will obtain an asymptotic distribution of the test statistics.

4.1.2 The three famous tests

In this section, we consider the likelihood ratio, Lawley-Hotelling, and Bartlett-Nanda-Pillai tests proposed for independent observations.

To write the test statistics, we introduce

$$\begin{aligned}
\bar{\boldsymbol{X}}_{i\cdot} &\equiv \frac{1}{n_i}\sum_{t=1}^{n_i} \boldsymbol{X}_{it},\ \ \bar{\boldsymbol{X}}_{\cdot\cdot} \equiv \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i} \boldsymbol{X}_{it},\\
\hat{\mathcal{S}}_H &\equiv \sum_{i=1}^{q} n_i(\bar{\boldsymbol{X}}_{i\cdot} - \bar{\boldsymbol{X}}_{\cdot\cdot})(\bar{\boldsymbol{X}}_{i\cdot} - \bar{\boldsymbol{X}}_{\cdot\cdot})',\\
\hat{\mathcal{S}}_E &\equiv \sum_{i=1}^{q}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_{i\cdot})',
\end{aligned}$$

where $n = n_1 + \cdots + n_q$. These statistics are known as the within-group mean (i-th treatment group), the grand mean, the sum of squares + products (SSP) for the hypothesis, and the SSP for the errors, respectively. In the case where the $\boldsymbol{\epsilon}_{it}$'s are mutually independent with respect to t, the following tests under normality have been proposed:

$$LR \equiv -n\log\{|\hat{\mathcal{S}}_E|/|\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H|\} \text{ (likelihood ratio test)}, \tag{4.4}$$
$$LH \equiv n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\} \text{ (Lawley-Hotelling test)}, \tag{4.5}$$
$$BNP \equiv n\mathrm{tr}\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H)^{-1} \text{ (Bartlett-Nanda-Pillai test)}. \tag{4.6}$$

Now, we suppose

Assumption 4.1.1 $\det\{\boldsymbol{f}(0)\} > 0$.

Let $\bar{\boldsymbol{\epsilon}}_{i\cdot} \equiv n_i^{-1}\sum_{t=1}^{n_i}\boldsymbol{\epsilon}_{it}$, and $\boldsymbol{Z} = (\sqrt{n_1}\bar{\boldsymbol{\epsilon}}_{1\cdot},\ldots,\sqrt{n_q}\bar{\boldsymbol{\epsilon}}_{q\cdot})$.

The following lemma comes from Hannan (1970) (p.208, p.221).

Lemma 4.1.1 *Under Assumption 4.1.1, as n_i, $i = 1,\ldots,q$, tend to ∞ so that $n_i/n \to \rho_i > 0$ as $n \to \infty$, if $\boldsymbol{\epsilon}_{it}$ is generated by generalized linear process (4.2), then*

(i) $\bar{\boldsymbol{\epsilon}}_{i\cdot} \xrightarrow{p} \mathbf{0}$ *for* $i = 1,\ldots,q$,

(ii) $$vec\{\boldsymbol{Z}\} \xrightarrow{\mathcal{L}} N\left[\boldsymbol{O}_{pq},\ 2\pi\begin{pmatrix} \boldsymbol{f}(0) & \boldsymbol{O}_{p\times p} & \cdots & \boldsymbol{O}_{p\times p} \\ \boldsymbol{O}_{p\times p} & \boldsymbol{f}(0) & \cdots & \boldsymbol{O}_{p\times p} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{O}_{p\times p} & \boldsymbol{O}_{p\times p} & \cdots & \boldsymbol{f}(0) \end{pmatrix}\right].$$

Theorem 4.1.1 *Assume that the processes $\{\boldsymbol{\epsilon}_{it}\}$ in (4.2) have the fourth-order cumulant and that*

$$\boldsymbol{\Gamma}(j) = \mathbf{0} \ for\ all\ j \neq 0. \tag{4.7}$$

Then, under H and Assumption 4.1.1, it holds that the tests LR, LH, and BNP are all asymptotically $\chi^2_{p(q-1)}$.

The proof may be found in Section 4.4.

Condition (4.7) means that the $\{\boldsymbol{\epsilon}_{it}\}$'s follow an uncorrelated process. Theorem 4.1.1 shows that the three test statistics proposed for independent observations can be applied to dependent observations satisfying (4.7). Here, it may be noted that condition (4.7) is not so severe because the following example describes a very practical, nonlinear time series model that satisfies (4.7).

Example 1 *Engle (2002) introduced the dynamic conditional correlation generalized autoregressive conditional heteroscedasticity (DCC-GARCH) (p, q) model defined as*

$$\boldsymbol{\epsilon}_{it} = \boldsymbol{H}_{it}{}^{1/2}\boldsymbol{\eta}_{it}, \quad \boldsymbol{\eta}_{it} \overset{i.i.d.}{\sim} (\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{H}_{it} = \boldsymbol{D}_{it}\boldsymbol{R}_{it}\boldsymbol{D}_{it}, \quad \boldsymbol{D}_{it} = diag\left[\sqrt{\sigma_{it}^{(1)}}, \cdots, \sqrt{\sigma_{it}^{(p)}}\right],$$

$$\boldsymbol{\epsilon}_{it} = \begin{pmatrix} \epsilon_{it}^{(1)} \\ \vdots \\ \epsilon_{it}^{(p)} \end{pmatrix}, \quad \sigma_{it}^{(j)} = c_j + a_j \sum_{l=1}^{q} \left\{\epsilon_{i,t-l}^{(j)}\right\}^2 + b_j \sum_{l=1}^{p} \sigma_{i,t-l}^{(j)},$$

$$\boldsymbol{R}_{it} = (diag\,[\boldsymbol{Q}_{it}])^{-1/2}\,\boldsymbol{Q}_{it}\,(diag\,[\boldsymbol{Q}_{it}])^{-1/2},$$

$$\tilde{\boldsymbol{\epsilon}}_{it} = \begin{pmatrix} \tilde{\epsilon}_{it}^{(1)} \\ \vdots \\ \tilde{\epsilon}_{it}^{(p)} \end{pmatrix}, \quad \tilde{\epsilon}_{it}^{(j)} = \frac{\epsilon_{it}^{(j)}}{\sqrt{\sigma_{it}^{(j)}}}, \quad \boldsymbol{Q}_{it} = (1-\alpha-\beta)\tilde{\boldsymbol{Q}} + \alpha\tilde{\boldsymbol{\epsilon}}_{i,t-1}\tilde{\boldsymbol{\epsilon}}'_{i,t-1} + \beta\boldsymbol{Q}_{i,t-1},$$

where $a_j, b_j, c_j, \alpha, \beta$, and $j = 1, \cdots, p$ are constants, and $\tilde{\boldsymbol{Q}}$, the unconditional correlation matrix, is a constant positive semidefinite matrix. Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $\{\boldsymbol{\epsilon}_{i,t-1}, \boldsymbol{\epsilon}_{i,t-2}, \cdots\}$. We assume that $\boldsymbol{H}_{it}$ is measurable with respect to $\mathcal{F}_{t-1}$ and that $\boldsymbol{\eta}_{it} \perp \mathcal{F}_{t-1}$, where $\perp$ signifies mutual independence. This model is easier to compute than the usual vector-GARCH model because it has fewer unknown parameters than does the usual vector-GARCH model.

4.1.3 Likelihood ratio test based on the Whittle likelihood

In Section 4.1.2, we saw that the classical tests LR, LH, and BNP are asymptotically χ^2-distributed when (4.7) holds. However if we want to test H for general disturbances, the LR, LH, and BNP tests are not available. For this case, we propose a new test based on the Whittle likelihood. In what follows, we use the same notations as in Section 4.1.2.

It is known that Whittle's approximation to the Gaussian likelihood function is given by

$$l(\boldsymbol{\mu}, \alpha_i) \equiv -\frac{1}{2}\sum_{i=1}^{q}\sum_{s=0}^{n_i-1} \mathrm{tr}\left\{\boldsymbol{I}_i(\lambda_s)\boldsymbol{f}(\lambda_s)^{-1}\right\},$$

where $\lambda_s = 2\pi s/n_i$ and

$$\boldsymbol{I}_i(\lambda) \equiv \frac{1}{2\pi n_i}\left\{\sum_{t=1}^{n_i}(\boldsymbol{X}_{it} - \boldsymbol{\mu} - \boldsymbol{\alpha}_i)\mathrm{e}^{\mathrm{i}\lambda t}\right\}\left\{\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu} - \boldsymbol{\mu} - \boldsymbol{\alpha}_i)\mathrm{e}^{\mathrm{i}\lambda u}\right\}^*.$$

The derivation of the integral version for each i can be found in, for example, p.52-53 of Taniguchi and Kakizawa (2000).

Under H, from $\frac{\partial l(\boldsymbol{\mu},\mathbf{0})}{\partial \boldsymbol{\mu}} = \mathbf{0}$, $\frac{\partial l(\boldsymbol{\mu},\boldsymbol{\alpha}_i)}{\partial \boldsymbol{\mu}} = \mathbf{0}, and \ \frac{\partial l(\boldsymbol{\mu},\boldsymbol{\alpha}_i)}{\partial \boldsymbol{\alpha}_i} = \mathbf{0}$, we can see that the solutions are

$$\begin{aligned} \boldsymbol{\mu} &= \hat{\boldsymbol{\mu}} \equiv \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}\boldsymbol{X}_{it}, \\ \boldsymbol{\alpha}_i &= \hat{\boldsymbol{\alpha}}_i \equiv \frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}}). \end{aligned}$$

To solve the problem of testing H, we introduce the test statistic

$$WLR \equiv 2\left\{l(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\alpha}}_i) - l(\hat{\boldsymbol{\mu}},\mathbf{0})\right\},$$

which can be used when condition (4.7) does not hold.

Showing that

$$WLR = \sum_{i=1}^{q}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\boldsymbol{f}(0)\right\}^{-1}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i, \tag{4.8}$$

we obtain the following theorem.

Theorem 4.1.2 *Under H and Assumption 4.1.1, if $\boldsymbol{\epsilon}_{it}$ is generated by the generalized linear process (4.2), then the test WLR is asymptotically $\chi^2_{p(q-1)}$-distributed.*

The proof may be found in Section 4.4.

Thus, this new test based on the Whittle likelihood is asymptotically χ^2-distributed even if condition (4.7) does not hold.

Next, to propose a practical version of WLR, we consider

$$\hat{\boldsymbol{f}}_i(\lambda) = \frac{1}{2\pi}\sum_{t=-(n_i-1)}^{n_i-1} w_t e^{-\mathrm{i}t\lambda}\left(1-\frac{|t|}{n_i}\right)\boldsymbol{C}_i(t), \tag{4.9}$$

$$w_t = w\left(\frac{t}{M(n_i)}\right), \tag{4.10}$$

$$\boldsymbol{C}_i(t) = \frac{1}{n_i}\sum_{s=1}^{n_i-t}(\boldsymbol{X}_{is}-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{i\ s+t}-\hat{\boldsymbol{X}}_{i\cdot})',$$

where $w(x)$ is a continuous, even function, $M(n_i)$ is a sequence of integers, and $\hat{\boldsymbol{X}}_{i\cdot}$ in $\boldsymbol{C}_i(t)$ is defined as in Section 4.1.2. As in Hannan (1970), we impose the following assumption.

Assumption 4.1.2 *(i)*

$$\sum_{t_1=-\infty}^{\infty}\sum_{t_2=-\infty}^{\infty}\sum_{t_3=-\infty}^{\infty}|c_{j,k,l,m}^{(i)}(0,t_1,t_2,t_3)|<\infty, \quad \sum_{t=-\infty}^{\infty}|t|^{\nu}\|\mathbf{\Gamma}(t)\|<\infty,\ \nu\geq 0$$

for $i=1,\cdots,q$ *and* $j,k,l,m=1,\cdots,p$,

(ii) $w(x)$ *in (4.10) is continuous and uniformly bounded,* $w(0)=1$, $|w(x)|<1$, $\int_{-\infty}^{\infty}w^2(x)dx<\infty$, *and for some integer* $\nu\geq 1$,

$$\lim_{x\to 0}\frac{1-w(x)}{|x|^{\nu}}<\infty,$$

(iii) $M(n_i)\to\infty$, $\{M(n_i)\}^{\nu}/n_i\to 0$, *for the same* ν.

Here, $c_{j,k,l,m}^{(i)}(t_1,t_2,t_3,t_4)=\mathrm{cum}\{\epsilon_{it_1}^{(j)},\epsilon_{it_2}^{(k)},\epsilon_{it_3}^{(l)},\epsilon_{it_4}^{(m)}\}$ *for* $i=1,\cdots,q$ *and* $j,k,l,m=1,\cdots,p$.

The following lemma is due to Hannan (1970) (p.280, p.283, and p.331).

Lemma 4.1.2 *Under Assumption 4.1.2, as* n_i, $i=1,\ldots,q$, *tend to* ∞, $\hat{\boldsymbol{f}}_i(\lambda)\xrightarrow{p}\boldsymbol{f}(\lambda)$ *for* $i=1,\cdots,q$.

Using Lemma 4.1.2, we can replace $\boldsymbol{f}(0)$ in (4.8) by $\hat{\boldsymbol{f}}_i(0)$:

$$WLR^*=\sum_{i=1}^{q}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\hat{\boldsymbol{f}}_i(0)\right\}^{-1}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i. \tag{4.11}$$

The following result immediately follows from Slutsky's theorem.

Theorem 4.1.3 *Under* H *and Assumptions 4.1.1 and 4.1.2, if* $\boldsymbol{\epsilon}_{it}$ *is generated by generalized linear process (4.2), then the test* WLR^* *is asymptotically* $\chi^2_{p(q-1)}$*-distributed.*

WLR^* in Theorem 4.1.3 is useful to practical applications because it can be computed directly from observations.

4.2 Numerical studies

4.2.1 Simulation to test the theoretical results

We conduct numerical studies of the tests LR, LH, BNP, and WLR^*, which are given by equations (4.4), (4.5), (4.6), and (4.11), respectively. In this section, our purpose is to confirm whether or not the four tests are well-approximated by the $\chi^2_{p(q-1)}$ distribution under the null hypothesis H and to determine their powers under the alternative hypothesis (i.e., $\boldsymbol{\alpha}_i,\ i = 1, \cdots, q$ are not the same) in the cases of both uncorrelated and dependent observations. DCC-$GARCH(1,1)$ is an example of an uncorrelated process (see Engle (2002)), and $VAR(1)$ is an example process with a high level of dependence.

First, we introduce the following six simulation process steps.

1. Generate a two-dimensional time series and three data groups with lengths 1000, 1500, and 2000, given by $\{\boldsymbol{X}_{1,1}, \dots, \boldsymbol{X}_{1,1000}\}, \{\boldsymbol{X}_{2,1}, \dots, \boldsymbol{X}_{2,1500}\}$, and $\{\boldsymbol{X}_{3,1}, \dots, \boldsymbol{X}_{3,2000}\}$, respectively, from one-way MANOVA model (1) with DCC-$GARCH(1,1)$ and $VAR(1)$ disturbance processes.

2. Set $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3 = \mathbf{0}$ when the null hypothesis H is valid, and set $\boldsymbol{\alpha}_1 = (-0.1, -0.1)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, and $\boldsymbol{\alpha}_3 = (0.1, 0.1)$ or $\boldsymbol{\alpha}_1 = (-1.0, -1.0)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, and $\boldsymbol{\alpha}_3 = (1.0, 1.0)'$ when the alternative hypothesis is valid.

3. Calculate $\hat{\mathcal{S}}_E$ and $\hat{\mathcal{S}}_H$ for (4.4), (4.5), and (4.6) as well as the smoothed periodogram to estimate $\boldsymbol{f}(0)$ for (4.8).

4. Calculate the test statistics $T = LR$, LH, BNP, and WLR^*.

5. Repeat steps 1-3 1000 times independently, and obtain $\{T^{(1)}, \dots, T^{(1000)}\}$.

6. Make a Q-Q plot using $\{T^{(1)}, \dots, T^{(1000)}\}$.

The data used in step 1 can be thought of as the virtual financial data of two companies for three industries. We generate the disturbance process $\{\boldsymbol{\epsilon}_{it}\}$ of observations $\{\boldsymbol{X}_{it}\}$ using DCC-$GARCH(1,1)$ and $VAR(1)$. The innovation terms of these processes are set as Gaussian. Specifically, we generate the two-dimensional simulation from one-way MANOVA model (1) with $\boldsymbol{\mu} = (1,1)'$. $VAR(1)$ is defined by $\boldsymbol{\epsilon}_{it}$:

$$\boldsymbol{\epsilon}_{it} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \boldsymbol{\epsilon}_{i,t-1} + \boldsymbol{u}_{it}, \quad \boldsymbol{u}_{it} \overset{i.i.d.}{\sim} N\left(\mathbf{0}, \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}\right).$$

We introduce $DCC\text{-}GARCH(1,1)$ in $\boldsymbol{\epsilon}_{it}$:

$$\boldsymbol{\epsilon}_{it} = \boldsymbol{H}_{it}^{1/2}\boldsymbol{\eta}_{it}, \quad \boldsymbol{\eta}_{it} \overset{i.i.d.}{\sim} N\left(\boldsymbol{0}, \boldsymbol{I}_p\right),$$
$$\boldsymbol{H}_{it} = \boldsymbol{D}_{it}\boldsymbol{R}_{it}\boldsymbol{D}_{it}, \quad \boldsymbol{D}_{it} = diag\left[\sqrt{\sigma_{it}^{(1)}}, \sqrt{\sigma_{it}^{(2)}}\right],$$
$$\boldsymbol{\epsilon}_{it} = \begin{pmatrix} \epsilon_{it}^{(1)} \\ \epsilon_{it}^{(2)} \end{pmatrix}, \quad \sigma_{it}^{(1)} = 0.003 + 0.2\left\{\epsilon_{i,t-1}^{(1)}\right\}^2 + 0.75\sigma_{i,t-1}^{(1)},$$
$$\sigma_{it}^{(2)} = 0.005 + 0.3\left\{\epsilon_{i,t-1}^{(2)}\right\}^2 + 0.6\sigma_{i,t-1}^{(2)},$$
$$\boldsymbol{R}_{it} = \left(diag\left[\boldsymbol{Q}_{it}\right]\right)^{-1/2}\boldsymbol{Q}_{it}\left(diag\left[\boldsymbol{Q}_{it}\right]\right)^{-1/2},$$
$$\tilde{\boldsymbol{\epsilon}}_{it} = \begin{pmatrix} \tilde{\epsilon}_{it}^{(1)} \\ \tilde{\epsilon}_{it}^{(2)} \end{pmatrix}, \quad \tilde{\epsilon}_{it}^{(j)} = \frac{\epsilon_{it}^{(j)}}{\sqrt{\sigma_{it}^{(j)}}}, \quad \boldsymbol{Q}_{it} = 0.1\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} + 0.1\tilde{\boldsymbol{\epsilon}}_{i,t-1}\tilde{\boldsymbol{\epsilon}}_{i,t-1}' + 0.8\boldsymbol{Q}_{i,t-1},$$

where $j = 1, 2$ (see Engle (2002)). We used Mathematica's code "ARprocess" and the R package "ccgarch" for these algorithms. Under the null hypothesis, the Q-Q plot for $\chi^2_{p(q-1)}$ is given by Fig. 4.1. We show the powers of the four tests under the alternative hypothesis in Tables 4.1-4.2. The cutoff points of the $10\%, 5\%$, and 1% significance levels are calculated using the $\chi^2_{p(q-1)}$ distribution, as in Theorems 4.1.1 and 4.1.3.

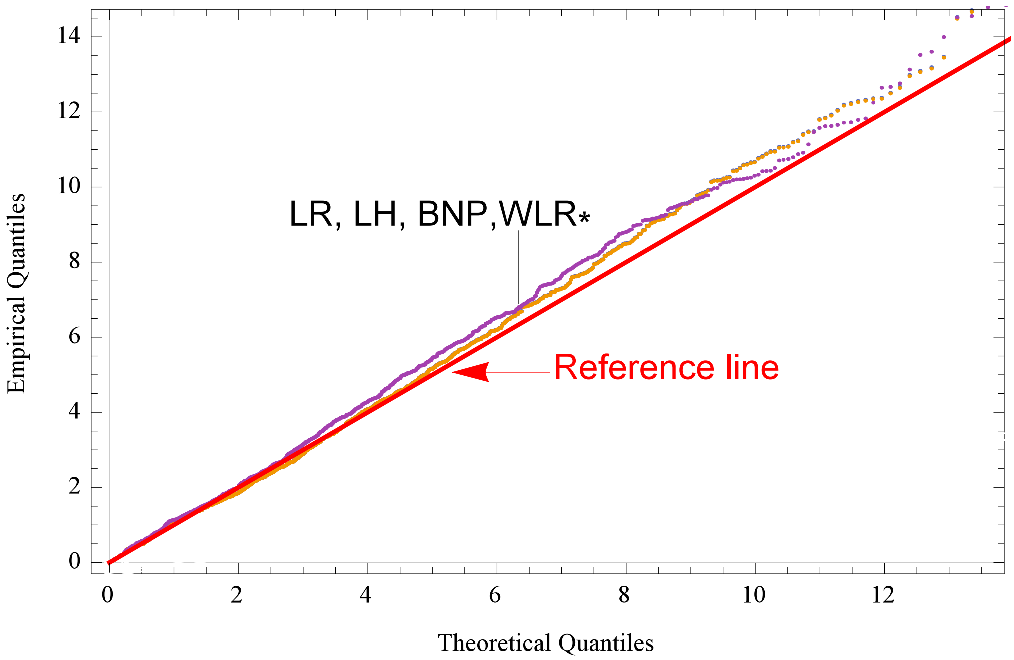
Under the null hypothesis, all of the tests are $\chi^2_{p(q-1)}$-distributed for uncorrelated observations. For dependent observations, only WLR^* is $\chi^2_{p(q-1)}$-distributed, and, evidently, this test was more effective than the classical tests, as Fig. 4.1 shows. On the other hand, under the alternative hypothesis, if the data are uncorrelated, the powers of all of the tests are sufficient in all settings, as indicated by the strong numerical evidence shown in Tables 4.1-4.2. In addition, for dependent observations, the power of WLR^* is sufficient, as shown in Tables 4.1-4.2.

Disturbance process	Test statistic	Significance level		
		10%	5%	1%
DCC-GARCH(1,1)	LR	1.000	1.000	1.000
	LH	1.000	1.000	1.000
	BNP	1.000	1.000	1.000
	WLR^*	1.000	1.000	1.000
VAR(1)	LR	$0.522^\dagger$	$0.410^\dagger$	$0.162^\dagger$
	LH	$0.522^\dagger$	$0.410^\dagger$	$0.162^\dagger$
	BNP	$0.522^\dagger$	$0.410^\dagger$	$0.162^\dagger$
	WLR^*	0.549	0.426	0.211

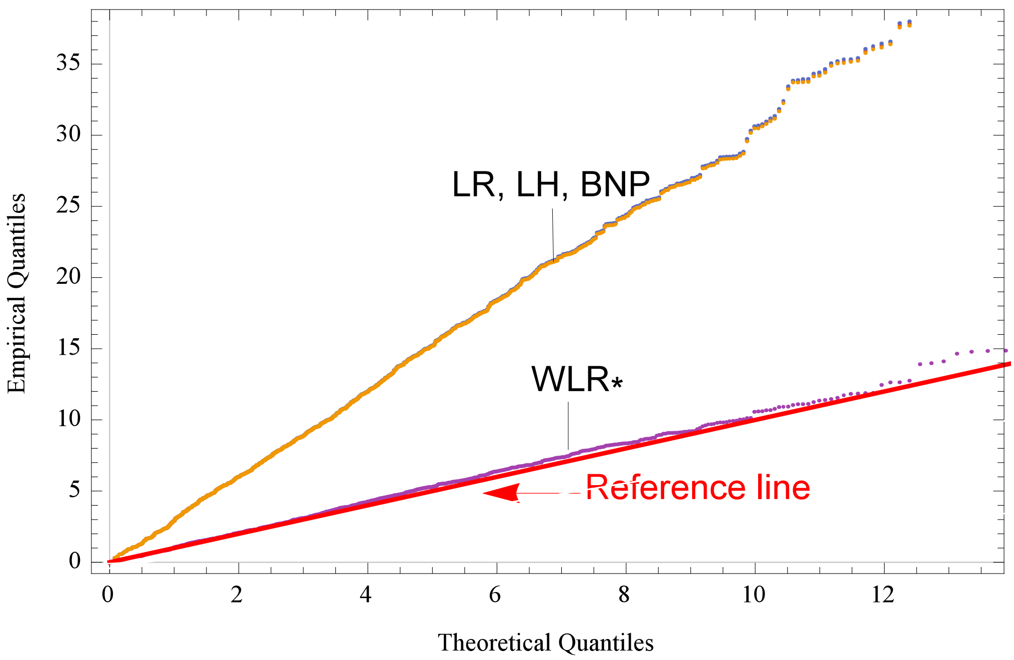
Table 4.1: The powers of the test statistics under the alternative hypothesis $\boldsymbol{\alpha}_1 = (-0.05, -0.05)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.05, 0.05)'$ (†: we computed the cutoff point from the empirical distributions of the classical tests because if the disturbance process is a dependent observation, the asymptotic null distribution is quite different from a $\chi^2_{p(q-1)}$ distribution, as in Theorem 4.1.1 and Fig. 4.1)

Disturbance process	Test statistic	Significance level		
		10%	5%	1%
DCC-GARCH(1,1)	LR	1.000	1.000	1.000
	LH	1.000	1.000	1.000
	BNP	1.000	1.000	1.000
	WLR^*	1.000	1.000	1.000
VAR(1)	LR	$0.975^\dagger$	$0.948^\dagger$	$0.848^\dagger$
	LH	$0.975^\dagger$	$0.948^\dagger$	$0.848^\dagger$
	BNP	$0.975^\dagger$	$0.948^\dagger$	$0.848^\dagger$
	WLR^*	0.980	0.957	0.881

Table 4.2: The powers of the test statistics under the alternative hypothesis $\boldsymbol{\alpha}_1 = (-0.1, -0.1)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.1, 0.1)'$ (†: we computed the cutoff point from the empirical distributions of the classical tests because if the disturbance process is a dependent observation, the asymptotic null distribution is quite different from the $\chi^2_{p(q-1)}$ distribution, as in Theorem 4.1.1 and Fig. 4.1)



(a) Uncorrelated observations from $DCC\text{-}GARCH(1,1)$



(b) Dependent observations from $VAR(1)$

Figure 4.1: The graphs show Q-Q plots whose theoretical quantiles are given by a $\chi^2_{p(q-1)}$ distribution, and these empirical quantiles are calculated using LR, LH, BNP, and WLR^*. The reference line is a straight line of vector (1,1).

4.2.2 Application to real financial data

We apply LR, LH, BNP, and WLR^* to the daily log data of some stocks in the NEEDS-FinancialQUEST database, which can be found at (http://finquest.nikkeidb.or.jp/ver2/online/index_en.htm). This data set consists of three groups with two dimensions and about 2500 - 5000 cell lines. We choose three groups, the electric appliance, film, and financial industries, which have 2450 - 4911 data points from May 22, 1997; May 22, 1997; and May 22, 2007 to May 22, 2017, respectively. Each industry includes two companies, as shown in Table 4.3. (The time ranges are the same for the two time series within a group, but the time ranges differ between groups.) We assume that time series data belonging to different groups are independent, as is the setting in Section 4.1.1. In fact, the correlations between the data from the three groups are weak, as shown in Table 4.4.

Industry	Electric appliance
Companies	NEC
	TOSHIBA
Industry	Film
Companies	TOEI
	TOHO
Industry	Finance
Companies	SUMITOMO MITSUI FINANCIAL GROUP (SMFG)
	MITSUBISHI UFJ FINANCIAL GROUP (MUFG)

Table 4.3: Names of the six selected stocks from the three industries

	NEC	TOSHIBA	TOEI	TOHO	SMFG	MUFG
NEC	1	0.698972	0.0507566	-0.0876778	-0.147324	-0.185421
TOSHIBA	0.698972	1	0.147726	0.0436181	-0.157357	-0.103833
TOEI	0.0507566	0.147726	1	0.874348	0.163815	0.275132
TOHO	-0.0876778	0.0436181	0.874348	1	-0.0160858	0.105472
SMFG	-0.147324	-0.157357	0.163815	-0.0160858	1	0.975848
MUFG	-0.185421	-0.103833	0.275132	0.105472	0.975848	1

Table 4.4: Correlations of the six selected stocks from the three industries

Industry	Within-group mean estimate	Overall average
NEC	6.27636	
TOSHIBA	6.20365	
TOEI	6.26408	6.93906
TOHO	7.48471	6.68281
SMFG	8.27675	
MUFG	6.36006	

Table 4.5: The within-group mean estimates and the overall averages of the three industries

To test hypothesis H, we examine whether the tests exceed the 5% significance level using Theorems 4.1.1 and 4.1.3.

We find that all of the tests reject hypothesis H, and their P-values are all around 1.0.

However, in view of Theorem 4.1.1, the classical tests LR, LH, BNP do not hold because these financial data do not satisfy condition (4.7). In all of these cases, we also examine the sample autocorrelation function of these daily log data. This function is very high, as shown in Fig. 4.2, so we cannot use the tests LR, LH, BNP, as implied by Theorem 4.1.1. Hence, only WLR^* is recommended in this case, and it can be applied to a variety of data, as previously described in Sections 4.1.3 and 4.2.1.

4.3 Conclusion

Considering a MANOVA model where the disturbances are generated by generalized linear processes, we provided a sufficient condition, condition (4.7), for each of the classical tests (i.e., the likelihood ratio, Lawely-Hotelling, and Bartlett-Nanda-Pillai tests) to have a χ^2 asymptotic null distribution. We also introduced a new likelihood ratio test statistic based on the Whittle likelihood, WLR^*, which is asymptotically χ^2-distributed even if condition (4.7) does not hold. We conducted some interesting numerical studies to show that our test statistic WLR^* is numerically stable under the null and alternative hypotheses. Finally, we applied the four test statistics to real financial data and found numerical evidence that our new test statistic WLR^* works well.

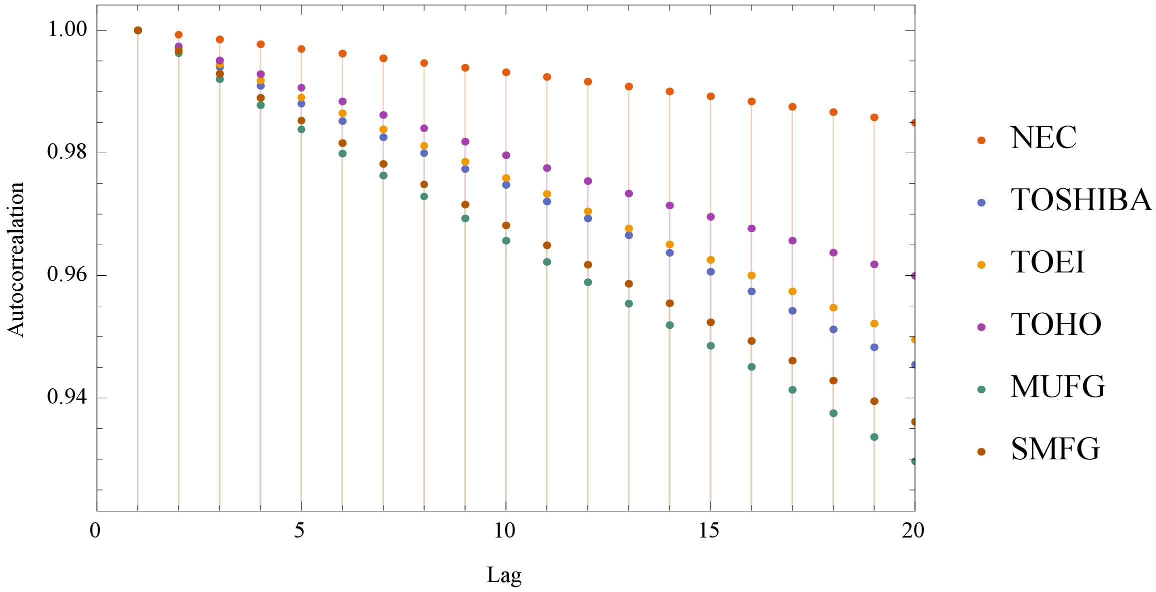


Figure 4.2: The dots in the graph show the sample autocorrelations of the daily log data of the given six stocks in the legends.

4.4 Proofs of Theorems

Proof (of Theorem 1) *By the transformation* $\boldsymbol{X}_{it} \to \boldsymbol{\Gamma}(0)^{-1/2}\boldsymbol{X}_{it}$, *we observe that the three tests, LR, LH and BNP, are invariant under linear transformation. Hence, without loss of generality, we may assume* $\boldsymbol{\Gamma}(0) = \boldsymbol{I}_p$. *Then,*

$$\begin{aligned}\frac{1}{n}\hat{\mathcal{S}}_E &= \sum_{i=1}^{q}\left(\frac{n_i}{n}\right)\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})', \\ &= \sum_{i=1}^{q}\rho_i\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})'.\end{aligned}$$

For each i, *rewriting*

$$\begin{aligned}&\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\bar{\boldsymbol{X}}_{i\cdot})' \\ &= \frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i+\boldsymbol{\alpha}_i-\bar{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i+\boldsymbol{\alpha}_i-\bar{\boldsymbol{X}}_{i\cdot})' \qquad (4.12)\end{aligned}$$

and applying Corollary 1 (p.205 of Hannan (1970)) and Theorem 6 (p.210 of Hannan (1970)) to (4.12), we observe that

$$\frac{1}{n}\hat{\mathcal{S}}_E = \boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}\left(\frac{1}{\sqrt{n_i}}\right). \tag{4.13}$$

Next, note that

$$\begin{aligned} LR &= n\log\left\{|\boldsymbol{I}_p - \hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}|\right\}, & (4.14)\\ LH &= \mathrm{tr}\left\{\hat{\mathcal{S}}_H\left(\frac{1}{n}\hat{\mathcal{S}}_E\right)^{-1}\right\}, & (4.15)\\ BNP &= \mathrm{tr}\left\{\hat{\mathcal{S}}_H\left(\frac{1}{n}\hat{\mathcal{S}}_E+\frac{1}{n}\hat{\mathcal{S}}_H\right)^{-1}\right\}. & (4.16)\end{aligned}$$

Substituting (4.13) into (4.14), (4.15), and (4.16), and noting that $d\log|\boldsymbol{F}| = \mathrm{tr}\boldsymbol{F}^{-1}d\boldsymbol{F}$ *(Magnus and Neudecker (1999)), we can show that the stochastic expansion of the three tests* $LR, LH, BNP(=T)$ *is given by*

$$T = \mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}\boldsymbol{Z}' + \boldsymbol{O}_{\mathrm{P}}\left(\frac{1}{\sqrt{n_i}}\right),$$

where $\boldsymbol{\Omega} = \boldsymbol{I}_q - \boldsymbol{\rho}\boldsymbol{\rho}'$*, with* $\boldsymbol{\rho} = (\sqrt{n_1/n}, \dots, \sqrt{n_q/n})'$ *(for the i.i.d. case, e.g., Fujikoshi et al. (2011) (p.164-165)). Since* $\boldsymbol{\Gamma}(j) = \boldsymbol{0}$, $(j \neq 0)$*, it follows from Lemma 1 that*

$$\mathrm{vec}\{\boldsymbol{Z}\} \xrightarrow{\mathcal{L}} N\left[\begin{pmatrix}\boldsymbol{0}\\ \vdots\\ \boldsymbol{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{f}(0) & \boldsymbol{O}_{p\times p} & \cdots & \boldsymbol{O}_{p\times p}\\ \boldsymbol{O}_{p\times p} & \boldsymbol{f}(0) & \cdots & \boldsymbol{O}_{p\times p}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{O}_{p\times p} & \boldsymbol{O}_{p\times p} & \cdots & \boldsymbol{f}(0)\end{pmatrix}\right],$$

which leads to the conclusion that $\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}\boldsymbol{Z}'$ *is asymptotically* $\chi^2_{p(q-1)}$ *by Rao (2009). Hence, the results follow.*

Proof (of Theorem 2) *Under* H*, we obtain*

$$\begin{aligned}\frac{\partial l(\boldsymbol{\mu},\boldsymbol{0})}{\partial\boldsymbol{\mu}} &= -\frac{1}{2}\sum_{i=1}^{q}\left[\sum_{s=0}^{n_i-1}\boldsymbol{f}(\lambda_s)^{-1}\left\{\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(-\mathrm{e}^{\mathbf{i}\lambda_s t})\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\boldsymbol{\mu})\mathrm{e}^{-\mathbf{i}\lambda_s u}\right.\right.\\ &+ \left.\left.\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\mu})\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}(-\mathrm{e}^{-\mathbf{i}\lambda_s u})\right\}\right].\end{aligned}$$

Noting that

$$\frac{1}{n_i}\sum_{t=1}^{n_i} \mathrm{e}^{\boldsymbol{i}\lambda_s t} = \begin{cases} 1 & (s=0) \\ 0 & (s \neq 0), \end{cases} \tag{4.17}$$

we can see that $\frac{\partial l(\boldsymbol{\mu},\mathbf{0})}{\partial \boldsymbol{\mu}} = \mathbf{0}$ *leads to the solution*

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i} \boldsymbol{X}_{it} \ .$$

Next,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\mu},\boldsymbol{\alpha}_i)}{\partial \boldsymbol{\alpha}_i} &= -\frac{1}{2}\sum_{s=0}^{n_i-1} \boldsymbol{f}(\lambda_s)^{-1}\left\{\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(-\mathrm{e}^{\mathbf{i}\lambda_s t})\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\boldsymbol{\mu}-\boldsymbol{\alpha}_i)\mathrm{e}^{-\mathbf{i}\lambda_s u}\right. \\ &+ \left.\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\mu}-\boldsymbol{\alpha}_i)\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}(-\mathrm{e}^{-\mathbf{i}\lambda_s u})\right\} = \mathbf{0} \end{aligned}$$

leads to

$$\boldsymbol{\alpha}_i = \frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\mu}).$$

Similarly, from $\frac{\partial l(\boldsymbol{\mu},\boldsymbol{\alpha}_i)}{\partial \boldsymbol{\mu}} = \mathbf{0}$*, we obtain*

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i).$$

As a solution, we may take

$$\begin{aligned} \boldsymbol{\alpha}_i &= \hat{\boldsymbol{\alpha}}_i \equiv \frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}}), \\ \boldsymbol{\mu} &= \hat{\boldsymbol{\mu}} \equiv \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}\boldsymbol{X}_{it} \ . \end{aligned}$$

From the above, it follows that

$$
\begin{aligned}
WLR &= 2\left\{l(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\alpha}}_i)-l(\hat{\boldsymbol{\mu}},\mathbf{0})\right\}\\
&= \sum_{i=1}^{q}\sum_{s=0}^{n_i-1}\mathrm{tr}\left\{\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}})\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\hat{\boldsymbol{\mu}})'\mathrm{e}^{-\mathbf{i}\lambda_s u}\boldsymbol{f}(\lambda_s)^{-1}\right.\\
&\quad\left.-\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}}-\hat{\boldsymbol{\alpha}}_i)\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\hat{\boldsymbol{\mu}}-\hat{\boldsymbol{\alpha}}_i)'\mathrm{e}^{-\mathbf{i}\lambda_s u}\boldsymbol{f}(\lambda_s)^{-1}\right\}\\
&= \sum_{i=1}^{q}\sum_{s=0}^{n_i-1}\mathrm{tr}\left[-\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}\hat{\boldsymbol{\alpha}}_i\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}\hat{\boldsymbol{\alpha}}_i'\mathrm{e}^{\mathbf{i}\lambda_s u}\boldsymbol{f}(\lambda_s)^{-1}\right.\\
&\qquad +\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}\hat{\boldsymbol{\alpha}}_i\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\hat{\boldsymbol{\mu}})'\mathrm{e}^{-\mathbf{i}\lambda_s u}\boldsymbol{f}(\lambda_s)-1\\
&\qquad \left.+\frac{1}{2\pi n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}})\mathrm{e}^{\mathbf{i}\lambda_s t}\sum_{u=1}^{n_i}\hat{\boldsymbol{\alpha}}_i'\mathrm{e}^{-\mathbf{i}\lambda_s u}\boldsymbol{f}(\lambda_s)^{-1}\right].
\end{aligned}
$$

Recalling (4.17), we obtain

$$
\begin{aligned}
WLR &= \sum_{i=1}^{q}\mathrm{tr}\left[-\frac{1}{2\pi}\hat{\boldsymbol{\alpha}}_i\sum_{u=1}^{n_i}\hat{\boldsymbol{\alpha}}_i'\boldsymbol{f}(0)^{-1}\right.\\
&\qquad +\hat{\boldsymbol{\alpha}}_i\frac{1}{2\pi}\sum_{u=1}^{n_i}(\boldsymbol{X}_{iu}-\hat{\boldsymbol{\mu}})'\boldsymbol{f}(0)^{-1}\\
&\qquad \left.+\frac{1}{2\pi}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{\mu}})\hat{\boldsymbol{\alpha}}_i'\boldsymbol{f}(0)^{-1}\right]\\
&= \sum_{i=1}^{q}\mathrm{tr}\left[-\hat{\boldsymbol{\alpha}}_i n_i\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\boldsymbol{f}(0)\right\}^{-1}\right.\\
&\qquad \left.+2\hat{\boldsymbol{\alpha}}_i n_i\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\boldsymbol{f}(0)\right\}^{-1}\right]\\
&= \sum_{i=1}^{q}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i'\left\{2\pi\boldsymbol{f}(0)\right\}^{-1}\sqrt{n_i}\hat{\boldsymbol{\alpha}}_i. \qquad (4.18)
\end{aligned}
$$

Notice that

$$
\hat{\boldsymbol{\alpha}}_i=\bar{\boldsymbol{\alpha}}_i-\frac{1}{n}\sum_{i=1}^{q}n_i\bar{\boldsymbol{\alpha}}_i,\quad \text{where } \bar{\boldsymbol{\alpha}}_i=\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\mu}).
$$

Then, (4.18) is a mean-corrected quadratic form. Since $\sqrt{n_i}\bar{\boldsymbol{\alpha}}_i \xrightarrow{d} N(\mathbf{0}, 2\pi \boldsymbol{f}(0))$ *(see Hannan (1970)(p.208)), we observe that* $WLR \xrightarrow{d} \chi^2_{p(q-1)}$.

Chapter 5

Analysis of variance for high-dimensional time series

Analysis of variance (ANOVA) is a hypothesis testing method for the null hypothesis of "no treatment effect". ANOVA is employed to test the null hypothesis that the means of three or more populations—the within-group means—are all equal.

ANOVA has a long history in statistics: its foundations were established by Gauss in the late 1800s and later by Markoff in the early 1900s. Many test statistics for ANOVA and multivariate analysis of variance (MANOVA) have been proposed, primarily for independent observations. Early applications can be found in Hooke (1926) and Wishart (1938). In addition, Bishop (1939) and Box (1949) obtained general theoretical results. They derived asymptotic expansions of the null and non-null distributions of likelihood ratio test statistics. Moreover, Bai et al. (1990) proposed an ANOVA method of adjusting for the unknown parameter, whereas Liu and Rao (1995) derived an asymptotic distribution of statistics for analysis of quadratic entropy (ANOQE) as a generalization of ANOVA. In addition, Rao (2010) investigated some ANOQE postulates and conditions. Furthermore, Fujikoshi et al. (2011) developed general asymptotic expansions of the null and non-null distributions of the likelihood ratio test statistic, Lawley-Hotelling test statistic, and Bartlett-Nanda-Pillai test statistic under high-dimensional settings. On the contrary, in a time-series analysis, Shumway (1971) discussed the asymptotic relationship between the likelihood ratio test statistic and Lawley-Hotelling test statistic, whereas Brillinger (1973) developed a time-series univariate and balanced ANOVA.

The analysis of multivariate time series data is common in practical problems, such as those in economics, finance, and bioinformatics. In the current era of big data, the formulation of statistical decisions for high-dimensional time series data has become increasingly important. In the present study, our focus is on the equivalence of the within-group means of real financial data, which are industry-dependent.

In this chapter, we consider the one-way MANOVA model (5.1), whose disturbance processes $\{\boldsymbol{\epsilon}_i\}$ are generated by a high-dimensional stationary process. The remainder of this chapter is organized as follows. Sections 5.1 and 5.3 describe the setting of this study and the asymptotics of fundamental statistics, respectively. In Section 5.2, we discuss the asymptotics of existing test statistics for high-dimensional dependent observations and derive a sufficient condition for them to be asymptotically normal. Sections 5.4.1 and 5.4.2 present a simulation of classical test statistics and the application of these test statistics to radioactive observations. The simulation results confirm the theoretical results.

5.1 Problems and Preliminaries

Throughout, we consider the MANOVA model under which a q-tuple of p-dimensional time series $\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{in_i}$, $i = 1, \dots, q$ satisfies

$$\boldsymbol{X}_{it} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it}, \quad t = 1, \cdots, n_i, \quad i = 1, \cdots, q, \tag{5.1}$$

where the disturbances $\boldsymbol{\epsilon}_i \equiv \{\boldsymbol{\epsilon}_{i1}, \cdots \boldsymbol{\epsilon}_{in_i}\}$ are kth-order stationary with mean $\mathbf{0}$, lag u autocovariance matrix $\boldsymbol{\Gamma}(u) = (\Gamma_{jk}(u))_{1 \le j,k \le p}$, $u \in \mathbb{Z}$. Moreover, $\{\boldsymbol{\epsilon}_i\}$, $i = 1, \cdots, q$ are mutually independent. This is a standard assumption, which is called the homoscedasticity (e.g., Ch. 8.9 of Anderson (2003)). Here, $\boldsymbol{\mu}$ is the global mean of the model (5.1), and $\boldsymbol{\alpha}_i$ denotes the effect of the ith treatment, which measures the deviation from $\boldsymbol{\mu}$ satisfying $\sum_{i=1}^{q} \boldsymbol{\alpha}_i = \mathbf{0}$. Because the treatment effects sum up to zero, we now consider the problem of testing:

$$H: \ \boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_q = \mathbf{0} \text{ vs. } A: \ \boldsymbol{\alpha}_i \neq \mathbf{0} \text{ for some } i. \tag{5.2}$$

The null hypothesis H implies all the effects are zero.

For our high-dimensional dependent observations, we use the following Lawley-Hotelling test statistic (LH), likelihood ratio test statistic (LR), and Bartlett-Nanda-Pillai test statistic (BNP):

$$\begin{aligned} \mathrm{LH} &\equiv n \mathrm{tr} \hat{\mathcal{S}}_H \hat{\mathcal{S}}_E^{-1}, \\ \mathrm{LR} &\equiv -n \log |\hat{\mathcal{S}}_E| / |\hat{\mathcal{S}}_E + \hat{\mathcal{S}}_H|, \\ \mathrm{BNP} &\equiv n \mathrm{tr} \hat{\mathcal{S}}_H (\hat{\mathcal{S}}_E + \hat{\mathcal{S}}_H)^{-1}, \end{aligned}$$

where

$$\hat{\mathcal{S}}_H \equiv \sum_{i=1}^{q} n_i(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})', \text{ and } \hat{\mathcal{S}}_E \equiv \sum_{i=1}^{q}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})', \text{ with}$$

$$\hat{\boldsymbol{X}}_{i\cdot} = \frac{1}{n_i}\sum_{t=1}^{n_i}\boldsymbol{X}_{it} \text{ and } \hat{\boldsymbol{X}}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}\boldsymbol{X}_{it}.$$

Here, $\hat{\mathcal{S}}_H$ and $\hat{\mathcal{S}}_E$ are called the between-group sums of squares and products (SSP) and the within-group SSP, respectively.

We derive the null asymptotic distribution of the three test statistics under the following assumptions:

Assumption 5.1.1

$$\frac{p^{3/2}}{\sqrt{n}} \to 0 \ \text{as } n, p \to \infty, \tag{5.3}$$

$$\frac{n_i}{n} \to \rho_i > 0 \ \text{as } n \to \infty. \tag{5.4}$$

Here, condition (5.4) means the sample size of the ith group n_i, and the total sample size n of all the groups are asymptotically of the same order.

Assumption 5.1.2 *For the p-vectors $\boldsymbol{\epsilon}_{it} = (\epsilon_{it}^{(1)}, \cdots, \epsilon_{it}^{(p)})'$ given in (5.1), there exists an $\ell > 0$ with*

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} \{1 + |t_j|^{\ell}\}|c^i_{a_1,\cdots,a_k}(t_1, \cdots, t_{k-1})| < \infty,$$

for $j = 1, \cdots, k-1$ and any k-tuple $a_1, \cdots, a_k \in \{1, \cdots, p\}$ and $i = 1, \cdots, q$, when $k = 2, 3, \cdots$. Here, $c^i_{a_1,\cdots,a_k}(t_1, \cdots, t_{k-1}) = \mathrm{cum}\{\epsilon_{it_1}^{(a_1)}, \cdots, \epsilon_{it_k}^{(a_k)}\}$.

If $\epsilon_{it}^{(a_{m_1})}, \cdots, \epsilon_{it}^{(a_{m_h})}$ for any h-tuple $m_1, \cdots, m_h \in \{1, \cdots, k\}$ are independent of $\epsilon_{it}^{(a_{m_{h+1}})}, \cdots, \epsilon_{it}^{(a_{m_k})}$ for the remaining $(k-h)$-tuple $m_{h+1}, \cdots, m_k \in \{1, \cdots, k\}$, then $c^i_{a_{m_1},\cdots,a_{m_k}}(t_{m_1}, \cdots, t_{m_k-1}) = 0$ (Brillinger (2001), p. 19). Assumption 5.1.2 implies that, if the time points of a group of $\epsilon_{it_l}^{(a_*)}$'s are well separated from the remaining time points of $\epsilon_{it_s}^{(a_*)}$'s, the values of $c^i_{a_1,\cdots,a_k}(t_1, \cdots, t_{k-1})$ become small (and hence summable) (see Brillinger (2001, p.19)). This property is natural for stochastic processes with short memory. Nevertheless, some readers may believe that Assumption

2 is very restrictive, but it is not so. A sequence of polynomial processes was introduced by Nisio (1960):

$$X_L(t) = \sum_{J=0}^{L} \sum_{u_1,\cdots,u_J} a_J(t-u_1,\cdots,t-u_J)W(u_1)\cdots W(u_J), \tag{5.5}$$

where the a_J's are absolutely summable, and $\{W(u)\} \overset{i.i.d.}{\sim} N(0,1)$. Nisio (1960) showed that, if a process $\{Y(t)\}$ is strictly stationary and ergodic, we can find a sequence of polynomial processes that converges to $\{Y(t)\}$ in law. Evidently, the sequence of polynomial processes (5.5) satisfies Assumption 5.1.2. We also present a very practical nonlinear time series model in (5.7) below, which satisfies Assumption 5.1.2.

Assumption 5.1.3

$$\mathbf{\Gamma}(j) = \mathbf{0} \textit{ for all } j \neq 0. \tag{5.6}$$

This assumption implies that the disturbance process $\{\boldsymbol{\epsilon}_i\}$ is an uncorrelated process. Here, it may be noted that the condition (5.6) is of course restrictive, but includes some practical nonlinear time-series models, like the $DCC\text{-}GARCH(q,r)$:

$$\begin{gathered}
\boldsymbol{\epsilon}_{it} = \boldsymbol{H}_{it}{}^{1/2}\boldsymbol{\eta}_{it}, \quad \boldsymbol{\eta}_{it} \overset{i.i.d.}{\sim} (\mathbf{0}, \boldsymbol{I}_p), \\
\boldsymbol{H}_{it} = \boldsymbol{D}_{it}\boldsymbol{R}_{it}\boldsymbol{D}_{it}, \quad \boldsymbol{D}_{it} = diag\left[\sqrt{\sigma_{it}^{(1)}},\cdots,\sqrt{\sigma_{it}^{(p)}}\right], \\
\boldsymbol{\epsilon}_{it} = \begin{pmatrix} \epsilon_{it}^{(1)} \\ \vdots \\ \epsilon_{it}^{(p)} \end{pmatrix}, \quad \sigma_{it}^{(j)} = c_j + a_j \sum_{l=1}^{r} \left\{\epsilon_{i,t-l}^{(j)}\right\}^2 + b_j \sum_{l=1}^{q} \sigma_{i,t-l}^{(j)}, \\
\boldsymbol{R}_{it} = \left(diag\left[\boldsymbol{Q}_{it}\right]\right)^{-1/2} \boldsymbol{Q}_{it} \left(diag\left[\boldsymbol{Q}_{it}\right]\right)^{-1/2}, \\
\tilde{\boldsymbol{\epsilon}}_{it} = \begin{pmatrix} \tilde{\epsilon}_{it}^{(1)} \\ \vdots \\ \tilde{\epsilon}_{it}^{(p)} \end{pmatrix}, \quad \tilde{\epsilon}_{it}^{(j)} = \frac{\epsilon_{it}^{(j)}}{\sqrt{\sigma_{it}^{(j)}}}, \quad \boldsymbol{Q}_{it} = (1-\alpha-\beta)\tilde{\boldsymbol{Q}} + \alpha\tilde{\boldsymbol{\epsilon}}_{i,t-1}\tilde{\boldsymbol{\epsilon}}_{i,t-1}' + \beta\boldsymbol{Q}_{i,t-1},
\end{gathered} \tag{5.7}$$

where $\tilde{\boldsymbol{Q}}$, the unconditional correlation matrix, is a constant positive semidefinite matrix, and $\boldsymbol{H}_{it}$'s are measurable with respect to $\boldsymbol{\eta}_{i,t-1}, \boldsymbol{\eta}_{i,t-2}, \cdots$, (see Engle (2002)) satisfies (5.6). Owing to Giraitis et al. (2000, formula (2.3)), a typical component is expressed as

$$\sum_{l=0}^{\infty} \sum_{j_l<j_{l-1}<\cdots<j_1<t} b_{t-j_1}\cdots b_{j_{l-1}-j_l}\eta_{j_1}\cdots\eta_{j_l} \tag{5.8}$$

where η_j's are i.i.d. with $E\eta_j^2 < \infty$. It is not difficult to observe that *DCC-GARCH*(p,q) satisfies Assumption 5.1.2.

Remark 5.1.1 *As described above, we replaced independent disturbances in the classical ANOVA by dependent ones like white noise and GARCH type disturbances.*

5.2 Main Results

In what follows, our discussion for $\boldsymbol{X}_{it}$ remains valid for the case when we apply a linear transformation $\{\boldsymbol{\Gamma}(0)\}^{-1/2}$ to $\boldsymbol{X}_{it}$ because the three test statistics LH, LR, and BNP are invariant under linear transformation. Hence, without loss of generality, we may assume $\boldsymbol{\Gamma}(0) = \boldsymbol{I}_p$, and $\boldsymbol{\mu} = \boldsymbol{0}$. We can derive the stochastic expansion of the standardized versions T_1, T_2, and T_3 of the three test statistics LH, LR, and BNP, respectively:

$$T_1 \equiv \frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1} - \sqrt{p}(q-1)\right\}, \text{ (LH)} \tag{5.9}$$

$$T_2 \equiv -\frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\log|\hat{\mathcal{S}}_E|/|\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H| + \sqrt{p}(q-1)\right\}, \text{ (LR)} \tag{5.10}$$

$$T_3 \equiv \frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H)^{-1} - \sqrt{p}(q-1)\right\}. \text{ (BNP)} \tag{5.11}$$

This section provides the asymptotic theory for the three test statistics, T_1, T_2, and T_3. Lemmas and their proofs are presented in Section 5.3.

Theorem 5.2.1 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis H, we have the stochastic expansion:*

$$T_i = \frac{1}{\sqrt{2(q-1)}}\left\{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H - \sqrt{p}(q-1)\right\} + \boldsymbol{o}_{\mathrm{P}}(1), \quad (i = 1, 2, 3).$$

Proof *From Lemma 5.3.2, it is easily observed that*

$$T_1 = \frac{1}{\sqrt{2(q-1)}}\left\{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H - \sqrt{p}(q-1)\right\} + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^{3/2}}{\sqrt{n}}\right), \tag{5.12}$$

whose error term becomes $\boldsymbol{o}_{\mathrm{P}}(1)$ by Assumption 5.1.1. For T_2, first, note that

$$d\log|\boldsymbol{F}| = \mathrm{tr}(\boldsymbol{F}^{-1}d\boldsymbol{F}),$$

(e.g., Magnus and Neudecker (1999)). Then, a modification of Proposition 6.1.5 of Brockwell and Davis (1991), and Lemmas 5.3.1 and 5.3.2 show

$$T_2 = \frac{1}{\sqrt{2(q-1)}}\left\{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H - \sqrt{p}(q-1)\right\} + \boldsymbol{o}_{\mathrm{P}}\left(1\right). \tag{5.13}$$

For T_3, in view of Lemmas 5.3.1 and 5.3.2, we obtain

$$\begin{aligned} n\mathrm{tr}\{\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E + \hat{\mathcal{S}}_H)^{-1}\} &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left(\frac{1}{n}\hat{\mathcal{S}}_E + \frac{1}{n}\hat{\mathcal{S}}_H\right)^{-1}\right] \\ &= \mathrm{tr}\left\{\hat{\mathcal{S}}_H\left(\boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^U\left(\frac{1}{\sqrt{n}}\right)\right)^{-1}\right\} \\ &= \mathrm{tr}\left\{\hat{\mathcal{S}}_H\left(\boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^U\left(\frac{1}{\sqrt{n}}\right)\right)\right\}, \end{aligned} \tag{5.14}$$

which leads to

$$T_3 = \frac{1}{\sqrt{2(q-1)}}\left\{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H - \sqrt{p}(q-1)\right\} + \boldsymbol{o}_{\mathrm{P}}\left(1\right). \tag{5.15}$$

Using Lemmas 5.3.1 and 5.3.2 and Theorem 5.2.1, we derive the asymptotic distribution of LH, LR, and BNP.

Theorem 5.2.2 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis H,*

$$T_i \xrightarrow{d} N(0,1), \quad (i = 1, 2, 3).$$

Proof *From Lemma 5.1.2, it is evident that*

$$\frac{1}{\sqrt{p}}E\{\mathrm{tr}\hat{\mathcal{S}}_H\} = \sqrt{p}(q-1) + \boldsymbol{O}\left(\frac{p^{3/2}}{\sqrt{n}}\right). \tag{5.16}$$

Moreover, from (5.27) and (5.31), it is not difficult to show

$$\mathrm{cum}\left\{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H, \frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H\right\} = 2(q-1) + \boldsymbol{o}\left(1\right). \tag{5.17}$$

Because $p^{-1/2}\mathrm{tr}\hat{\mathcal{S}}_H$ *is a finite linear combination of*

$$
\begin{aligned}
A_{ii} &\equiv \frac{1}{\sqrt{p}} n_i(\hat{\boldsymbol{X}}_{i\cdot} - \boldsymbol{X}_{\cdot\cdot})'(\hat{\boldsymbol{X}}_{i\cdot} - \boldsymbol{X}_{\cdot\cdot}) \\
&= \frac{1}{\sqrt{p}}\frac{1}{n_i}\sum_{l_1=1}^{p}\sum_{t_1=1}^{n_i}\epsilon_{it_1}^{(l_1)}\sum_{s_1=1}^{n_i}\epsilon_{is_1}^{(l_1)} - \frac{2}{\sqrt{p}}\frac{1}{n}\sum_{l_2=1}^{p}\sum_{t_2=1}^{n_i}\epsilon_{it_2}^{(l_2)}\sum_{j=1}^{q}\sum_{s_2=1}^{n_j}\epsilon_{js_2}^{(l_2)} \\
&\quad + \frac{1}{\sqrt{p}}\frac{n_i}{n^2}\sum_{l_3=1}^{p}\sum_{j_1=1}^{q}\sum_{t_3=1}^{n_{j_1}}\epsilon_{j_1t_3}^{(l_3)}\sum_{j_2=1}^{q}\sum_{s_3=1}^{n_{j_2}}\epsilon_{j_2s_3}^{(l_3)} \\
&\approx \frac{1}{\sqrt{p}}\frac{1}{n}\sum_{l_1=1}^{p}\sum_{t_1=1}^{n_i}\sum_{s_1=1}^{n_i}\epsilon_{it_1}^{(l_1)}\epsilon_{is_1}^{(l_1)} - \frac{2}{\sqrt{p}}\frac{1}{n}\sum_{l_2=1}^{p}\sum_{j=1}^{q}\sum_{t_2=1}^{n_i}\sum_{s_2=1}^{n_j}\epsilon_{it_2}^{(l_2)}\epsilon_{js_2}^{(l_2)} \\
&\quad + \frac{1}{\sqrt{p}}\frac{\rho_i}{n}\sum_{l_3=1}^{p}\sum_{j_1=1}^{q}\sum_{j_2=1}^{q}\sum_{t_3=1}^{n_{j_1}}\sum_{s_3=1}^{n_{j_2}}\epsilon_{j_1t_3}^{(l_3)}\epsilon_{j_2s_3}^{(l_3)} \qquad (5.18) \\
&\equiv B_{ii} + C_{ii} + D_{ii},
\end{aligned}
$$

we can evaluate $\mathrm{cum}^{(J)}\left\{p^{-1/2}\mathrm{tr}\hat{\mathcal{S}}_H, \cdots, p^{-1/2}\mathrm{tr}\hat{\mathcal{S}}_H\right\}$ *by using properties of the cumulant (see Brillinger (2001, p.19)). Next, we show that the J-th order* $(J \geq 3)$ *cumulants of* A_{ii} *tends to zero. It suffices thereby to focus on the term* B_{ii}*, because* C_{ii} *and* D_{ii} *can be treated similarly. Hence, we only show the J-th order* $(J \geq 3)$ *cumulants of* B_{ii} *converging to zero. In fact,*

$$
\mathrm{cum}^{(J)}\{B_{i_1i_1}, B_{i_2i_2}, \cdots, B_{i_Ji_J}\} = \Bigg[p^{-\frac{J}{2}}n^{-J}\sum_{l_1=1}^{p}\cdots\sum_{l_J=1}^{p}\sum_{t_1=1}^{n_{i_1}}\cdots\sum_{t_J=1}^{n_{i_J}}\sum_{s_1=1}^{n_{i_1}}\cdots\sum_{s_J=1}^{n_{i_J}} \mathrm{cum}^{(J)}\left\{\epsilon_{i_1t_1}^{(l_1)}\epsilon_{i_1s_1}^{(l_1)}, \cdots, \epsilon_{i_Jt_J}^{(l_J)}\epsilon_{i_Js_J}^{(l_J)}\right\}\Bigg]. \qquad (5.19)
$$

Brillinger (2001) shows that $\mathrm{cum}^{(J)}\{\ ,\cdots,\ \}$ *is expressed as the indecomposable sum of products of* $\mathrm{cum}\{\epsilon_{it}^{(l)} : (i,t) \in \nu\}$*. A typical main-order term is*

$$
\mathrm{cum}^{(J)}\left\{\epsilon_{i_1s_1}^{(l_1)}, \epsilon_{i_2t_2}^{(l_2)}\right\}\mathrm{cum}\left\{\epsilon_{i_2s_2}^{(l_2)}, \epsilon_{i_3t_3}^{(l_3)}\right\}\cdots\mathrm{cum}\left\{\epsilon_{i_Js_J}^{(l_J)}, \epsilon_{i_1t_1}^{(l_1)}\right\}. \qquad (5.20)
$$

Therefore, the main-order term of (5.19) for the typical cumulant (5.20) is

$$
p^{-\frac{J}{2}}n^{-J}\sum_{l_1=1}^{p}\cdots\sum_{l_J=1}^{p}\sum_{t_1=1}^{n_{i_1}}\cdots\sum_{t_J=1}^{n_{i_J}}\sum_{s_1=1}^{n_{j_1}}\cdots\sum_{s_J=1}^{n_{j_J}} c_{l_1l_2}(t_2 - s_1)c_{l_2l_3}(t_3 - s_2)\cdots c_{l_Jl_1}(t_1 - s_J)
$$

$$= p^{-\frac{J}{2}} \sum_{l=1}^{p} \sum_{r_1} \cdots \sum_{r_J} c_{ll}(r_1) c_{ll}(r_2) \cdots c_{ll}(r_J) \quad \text{(by Assumption 5.1.3 and } \boldsymbol{\Gamma}(0) = \boldsymbol{I}_p)$$
$$= \boldsymbol{O}\left(p^{-\frac{J}{2}+1}\right) \quad \text{(by Assumption 5.1.2).}$$

Finally, we use a method by Brillinger (2001) that under Assumption 5.1.1, all the cumulants of order greater than 2 are shown to converge to 0; therefore, the characteristic functions of the standardized test statistics converge to $\exp\{-t^2/2\}$ *(see Theorem 2.3.1 of Brillinger (2001, p.19); e.g., the proof of Theorem 5.10.1 of Brillinger (2001, p.417-18).). Hence, the asymptotic normality for* T_i *is shown.*

5.3 Asymptotics of Fundamental Statistics for Main Results

In this section, we provide lemmas and their proofs. In what follows, we use the same linear transformation as in Section 5.2. First, the stochastic expansion of $n^{-1}\hat{\mathcal{S}}_E$ and $\hat{\mathcal{S}}_H$ can be given.

Lemma 5.3.1 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis* H*,*

$$\frac{1}{n}\hat{\mathcal{S}}_E = \boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^U\left(\frac{1}{\sqrt{n}}\right), \tag{5.21}$$
$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^U\left(\frac{1}{\sqrt{n}}\right), \tag{5.22}$$
$$\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^U(1). \tag{5.23}$$

Proof *Write (5.21) as*

$$\frac{1}{n}\hat{\mathcal{S}}_E = \sum_{i=1}^{q}\left(\frac{n_i}{n}\right)\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})',$$
$$= \sum_{i=1}^{q}\rho_i\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it} - \hat{\boldsymbol{X}}_{i\cdot})'. \tag{5.24}$$

In what follows, for each i, rewrite:

$$
\begin{aligned}
&\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})' \\
&=\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i+\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i+\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})' \\
&=\frac{1}{n_i}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i)(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i)'+(\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})' \\
&\quad+\frac{1}{n_i}\sum_{t=1}^{n_i}\Big\{(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i)(\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})'+(\boldsymbol{\alpha}_i-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\boldsymbol{\alpha}_i)'\Big\} \\
&=\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{C}\ (say),
\end{aligned}
\tag{5.25}
$$

where $\boldsymbol{A}=\{A_{jk}\}$, $\boldsymbol{B}=\{B_{jk}\}$, *and* $\boldsymbol{C}=\{C_{jk}\}$. *We observe*

$$
\begin{aligned}
&E\{\boldsymbol{A}\}=\boldsymbol{I}_p \quad \text{and} \\
&Cov\{A_{jk},A_{lm}\} \\
&\quad=\frac{1}{n_i}\sum_{s=-n_i+1}^{n_i-1}\left(1-\frac{|s|}{n_i}\right)\{c_{jl}(s)c_{km}(s)+c_{jm}(s)c_{kl}(s)+c^i_{jklm}(0,s,s)\} \quad (5.26)\\
&\quad=\boldsymbol{O}\left(n_i^{-1}\right)=\boldsymbol{O}\left(n^{-1}\right) \quad \text{uniformly in } j,k,l,m \text{ by Assumption 5.1.2.}
\end{aligned}
$$

Hence, $\boldsymbol{A}=\boldsymbol{I}_p+\boldsymbol{O}^U_{\mathrm{P}}\left(1/\sqrt{n}\right)$. *Next, we observe*

$$
\begin{aligned}
&E(\hat{\boldsymbol{X}}_{i\cdot})=\boldsymbol{\alpha}_i \quad \text{and} \\
&Cov\{\hat{\boldsymbol{X}}_{i\cdot},\hat{\boldsymbol{X}}_{i\cdot}\} \\
&\quad=\left\{\frac{1}{n_i}\sum_{s=-n_i+1}^{n_i-1}\left(1-\frac{|s|}{n_i}\right)c_{jk}(s)\right\} \quad (5.27)\\
&\quad=\boldsymbol{O}^U\left(\frac{1}{n_i}\right).
\end{aligned}
$$

Thus,

$$
\boldsymbol{B}=\boldsymbol{O}^U_{\mathrm{P}}\left(\frac{1}{n}\right). \tag{5.28}
$$

Because $\boldsymbol{C}$ *is the matrix of cross-product terms between* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, application of Schwarz's inequality to each component of* $\boldsymbol{C}$ *yields* $\boldsymbol{C}=\boldsymbol{O}^U_{\mathrm{P}}\left(1/\sqrt{n}\right)$. *Therefore,*

$$
\frac{1}{n}\hat{\mathcal{S}}_E=\boldsymbol{I}_p+\boldsymbol{O}^U_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right),
$$

and

$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \left\{\boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^{U}\left(\frac{1}{\sqrt{n}}\right)\right\}^{-1} = \{\boldsymbol{I}_p - \boldsymbol{M}_n\}^{-1} \text{ (say)}.$$

It is known that

$$\{\boldsymbol{I}_p - \boldsymbol{M}_n\}^{-1} = \sum_{k=0}^{\infty} \boldsymbol{M}_n^k$$

(see p. 169 of Magnus and Neudecker (1999)). From Assumption 5.1.1, it follows that

$$\boldsymbol{M}_n^k = \boldsymbol{O}_{\mathrm{P}}\left(n^{-\frac{k}{2}}\right)\boldsymbol{H}$$

where $\boldsymbol{H}$ *is a* $p \times p$*-matrix and* $\boldsymbol{H} = \boldsymbol{O}_{\mathrm{P}}^{U}(1)$*. Then, we obtain*

$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^{U}\left(\frac{1}{\sqrt{n}}\right). \tag{5.29}$$

Next, we show $\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^{U}(1)$*. To this end, we recall*

$$\hat{\mathcal{S}}_H = \sum_{i=1}^{q} n_i(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})'. \tag{5.30}$$

From (5.27), we observe that $\hat{\boldsymbol{X}}_{i\cdot} = \boldsymbol{\alpha}_i + \boldsymbol{O}_{\mathrm{P}}^{U}\left(1/\sqrt{n_i}\right)$, $\sum_{i=1}^{q}\boldsymbol{\alpha}_i = \boldsymbol{0}$ *and, similarly,* $\hat{\boldsymbol{X}}_{\cdot\cdot} = \boldsymbol{O}_{\mathrm{P}}^{U}(1/\sqrt{n})$*. Thus, we have*

$$\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^{U}(1). \tag{5.31}$$

Lemma 5.3.2 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis* H*,*

$$\begin{aligned} n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\} &= \mathrm{tr}\hat{\mathcal{S}}_H + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^2}{\sqrt{n}}\right) \\ &= p(q-1) + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^2}{\sqrt{n}}\right). \end{aligned}$$

Proof *From Lemma 5.3.1, it follows that*

$$\begin{aligned} n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\} &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left(\frac{1}{n}\hat{\mathcal{S}}_E\right)^{-1}\right] \\ &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left\{\boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^{U}\left(\frac{1}{\sqrt{n}}\right)\right\}\right] \\ &= \mathrm{tr}\hat{\mathcal{S}}_H + \mathrm{tr}\left\{\hat{\mathcal{S}}_H\ \boldsymbol{O}_{\mathrm{P}}^{U}\left(\frac{1}{\sqrt{n}}\right)\right\}. \end{aligned}$$

From (5.31),

$$n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\} = \mathrm{tr}\hat{\mathcal{S}}_H + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^2}{\sqrt{n}}\right). \tag{5.32}$$

Under H, we may assume $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\alpha}_i = \mathbf{0}$, $i = 1, \cdots, q$. Consequently, recalling (5.27) and (5.30), we obtain

$$E[n\mathrm{tr}\{\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1}\}] = p(q-1) + \boldsymbol{O}\left(\frac{p^2}{\sqrt{n}}\right), \tag{5.33}$$

which completes the proof.

5.4 Numerical Studies

5.4.1 Simulation to verify the finite sample performance

We conduct numerical studies of the test statistics T_i, $i = 1, 2, 3$, which are given by equations (5.9), (5.10), and (5.11), respectively. In this section, our purpose is to confirm whether the null distributions of three test statistics are well approximated by $N(0, 1)$ and to evaluate their powers under the alternative hypothesis (i.e., $\boldsymbol{\alpha}_i$, $i = 1, \cdots, q$ are not the same) in cases of an uncorrelated disturbance. *DCC-GARCH*$(1, 1)$ is a typical example of an uncorrelated process (see Engle (2002)).

First, we introduce the following five simulation process steps.

1 Set $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3 = \mathbf{0}$ when the null hypothesis H is assumed. Additionally, set (i) $\boldsymbol{\alpha}_1 = (-0.1, \cdots, -0.1)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.1, \cdots, 0.1)$, (ii) $\boldsymbol{\alpha}_1 = (-0.01, \cdots, -0.01)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.01, \cdots, 0.01)$, or (iii) $\boldsymbol{\alpha}_1 = (-0.001, \cdots, -0.001)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, and $\boldsymbol{\alpha}_3 = (0.001, \cdots, 0.001)'$ when the alternative hypothesis is assumed.

2 Generate 50-dimensional $\{\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{1,k}\}, \{\boldsymbol{X}_{2,1}, \ldots, \boldsymbol{X}_{2,k}\}, \{\boldsymbol{X}_{3,1}, \ldots, \boldsymbol{X}_{3,k}\}$, k =50, 100, 500, 2500, or 7500 with *DCC-GARCH*$(1, 1)$ disturbance.

3 Calculate the test statistics T_i, $i = 1, 2, 3$.

4 Repeat steps 2 and 3 1,000 times independently and obtain $\{T_i^{(1)}, \ldots, T_i^{(1000)}; i = 1, 2, 3\}$.

5 Write the tables of the rejection rate and the power by using $\{T_i^{(1)}, \ldots, T_i^{(1000)}; i = 1, 2, 3\}$.

The data used in step 2 can be considered virtual radioactive observations of 50 points for three regions. We generate the disturbance process $\{\boldsymbol{\epsilon}_{it}\}$ of observations $\{\boldsymbol{X}_{it}\}$ using $DCC\text{-}GARCH(1,1)$. The innovation terms of these processes are assumed to be Gaussian. Specifically, we generate the 50-dimensional simulation from one-way MANOVA model (1) with a 50-dimensional vector $\boldsymbol{\mu}$ whose elements are all 1. The scenarios of $DCC\text{-}GARCH(q,r)$ (see (5.7)) in $\boldsymbol{\epsilon}_{it}$ are

$$
\begin{gathered}
p = 50,\ i = 1,2,3,\ t = 1,\cdots,2500 \text{ or } 7500,\\
j = 1,\cdots,50,\\
q = r = 1,\\
a_j = 0.2,\ b_j = 0.7,\ c_j = 0.002,\\
\alpha = 0.1,\ \beta = 0.8,\\
\tilde{Q}_{kl} = 0.7^{(|k-l|)},
\end{gathered}
$$

where $\tilde{\boldsymbol{Q}}_{kl}$ is the (k,l)-element of $\tilde{\boldsymbol{Q}}$. We use the Mathematica code and the "ccgarch" package of R for this algorithm. Under the null hypothesis, the rejection rate of the test statistics is given by Table 5.1. We show the powers of the three test statistics under the alternative hypothesis in Tables 5.2-5.4. The cutoff points of the $10\%, 5\%$, and 1% significance levels are calculated using the $N(0,1)$ distribution, as in Theorems 5.2.2.

Under the null hypothesis, it is evident that all the classical test statistics are effective for 2500 or more uncorrelated observations and especially BNP outperformed the other classical test statistics and works well for 100 or more observations, as shown in Table 5.1. On the contrary, under the alternative hypothesis, if the data are uncorrelated, the powers of all of the test statistics are sufficient for 500 or more observations, as indicated by the strong numerical evidence shown in Tables 5.2 and 5.3.

Sample size of each group	Test Statistic	Significance Level 10%	5%	1%
50	T_1 (LH)	0.906	0.879	0.781
	T_2 (LR)	0.615	0.493	0.313
	T_3 (BNP)	0.089	0.037	0.010
100	T_1 (LH)	0.564	0.460	0.286
	T_2 (LR)	0.323	0.212	0.098
	T_3 (BNP)	0.107	0.060	0.015
500	T_1 (LH)	0.161	0.100	0.032
	T_2 (LR)	0.134	0.082	0.022
	T_3 (BNP)	0.101	0.055	0.015
2500	T_1 (LH)	0.114	0.062	0.015
	T_2 (LR)	0.108	0.058	0.014
	T_3 (BNP)	0.100	0.057	0.014
7500	T_1 (LH)	0.109	0.058	0.012
	T_2 (LR)	0.107	0.055	0.012
	T_3 (BNP)	0.105	0.054	0.012

Table 5.1: Rejection rate of the test statistics under the null hypothesis $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3 = \mathbf{0}$.

Sample size of each group	Test Statistic	Significance Level 10%	5%	1%
50	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	0.993	0.941	0.665
100	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000
500	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000
2500	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000
7500	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000

Table 5.2: Powers of the test statistics under the alternative hypothesis (i) $\boldsymbol{\alpha}_1 = (-0.1, \cdots, -0.1)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.1, \cdots, 0.1)'$.

Sample size of each group	Test Statistic	Significance Level		
		10%	5%	1%
50	T_1 (LH)	0.960	0.939	0.890
	T_2 (LR)	0.768	0.684	0.520
	T_3 (BNP)	0.200	0.108	0.020
100	T_1 (LH)	0.844	0.770	0.606
	T_2 (LR)	0.636	0.518	0.346
	T_3 (BNP)	0.352	0.230	0.086
500	T_1 (LH)	0.978	0.959	0.907
	T_2 (LR)	0.972	0.950	0.864
	T_3 (BNP)	0.958	0.934	0.821
2500	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000
7500	T_1 (LH)	1.000	1.000	1.000
	T_2 (LR)	1.000	1.000	1.000
	T_3 (BNP)	1.000	1.000	1.000

Table 5.3: Powers of the test statistics under the alternative hypothesis (ii) $\boldsymbol{\alpha}_1 = (-0.01, \cdots, -0.01)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.01, \cdots, 0.01)'$.

Sample size of each group	Test Statistic	Significance Level		
		10%	5%	1%
50	T_1 (LH)	0.915	0.873	0.771
	T_2 (LR)	0.578	0.473	0.297
	T_3 (BNP)	0.087	0.035	0.004
100	T_1 (LH)	0.565	0.454	0.299
	T_2 (LR)	0.325	0.243	0.091
	T_3 (BNP)	0.098	0.052	0.015
500	T_1 (LH)	0.201	0.125	0.047
	T_2 (LR)	0.162	0.093	0.035
	T_3 (BNP)	0.126	0.077	0.019
2500	T_1 (LH)	0.161	0.107	0.033
	T_2 (LR)	0.158	0.103	0.028
	T_3 (BNP)	0.152	0.098	0.025
7500	T_1 (LH)	0.285	0.175	0.080
	T_2 (LR)	0.282	0.175	0.077
	T_3 (BNP)	0.280	0.174	0.076

Table 5.4: Powers of the test statistics under the alternative hypothesis (iii) $\boldsymbol{\alpha}_1 = (-0.001, \cdots, -0.001)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.001, \cdots, 0.001)'$

5.4.2 Application to radioactive observations

We applied $T_i, i = 1, 2, 3$ to radioactive observations in Fukushima Prefecture, Japan, which can be found at http://emdb.jaea.go.jp/emdb/en/portals/b138/. This series of data was created based on the Air Dose Rate Measurement in Fukushima Prefecture, which is publicly available on the Monitoring Information of Environmental Radioactivity Level website of the Nuclear Regulation Authority.

We divided Fukushima Prefecture into lower, middle, and upper areas, each of which had 50 observation points, as shown in Figure 5.1. In each area, the effective dose was observed at hourly intervals from 12:00 a.m., April 1, 2017 to 11:50 p.m., May 31, 2017. In other words, this data set consisted of three groups with 50 dimensions and 7,973 cell lines. We assumed that the time series data belonging to different groups were independent, as is the setting in Section 5.1.

In terms of the focus of our general methods, we were interested in the equality of the within-group means. We therefore employed ANOVA for a high-dimensional time series. To test hypothesis H, we examined whether the test statistics exceeded the 10% significance level using Theorem 5.2.2.

We determined that all the test statistics rejected hypothesis H, and their P-values were around 0. If our modeling was correct, this result implied that the effective dose of the lower and middle areas (red and blue points in Figure 5.1) were significantly higher than that of the upper area (green points in Figure 5.1).

In addition, we verified the condition (5.6). In view of the sample autocorrelation shown in Figure 5.2, this function is very low. Additionally, we verified the condition (5.6) by using $\tilde{Q}(\hat{r})$ $(m = 20)$ in Ljung and Box (1978) to jointly test for autocorrelations at multiple lags. The null hypothesis for this test was that the autocorrelations were jointly zero. The minimum of the P-values of this test was 0.828494; thus, it is far from any significance level $(0.1, 0.05, 0.01)$. This implies that these radioactive observations satisfied the condition (5.6). Accordingly, we could use the test statistics $T_i, i = 1, 2, 3$, as implied by Theorem 5.2.2. Furthermore, the power of the test statistics $T_i, i = 1, 2, 3$, was adequate. Basically, if the effect of the ith treatment $\boldsymbol{\alpha}_i$ became large, the test statistics could detect the difference between the within-group means, as shown in Tables 5.2-5.4. At this point, the difference between the within-group means in the radioactive observations was larger than $\boldsymbol{\alpha}_1 = (-0.01, \cdots, -0.01)'$, $\boldsymbol{\alpha}_2 = \mathbf{0}$, $\boldsymbol{\alpha}_3 = (0.01, \cdots, 0.01)'$, as shown in Table 5.3, where the power of $T_i, i = 1, 2, 3$, is quite high.

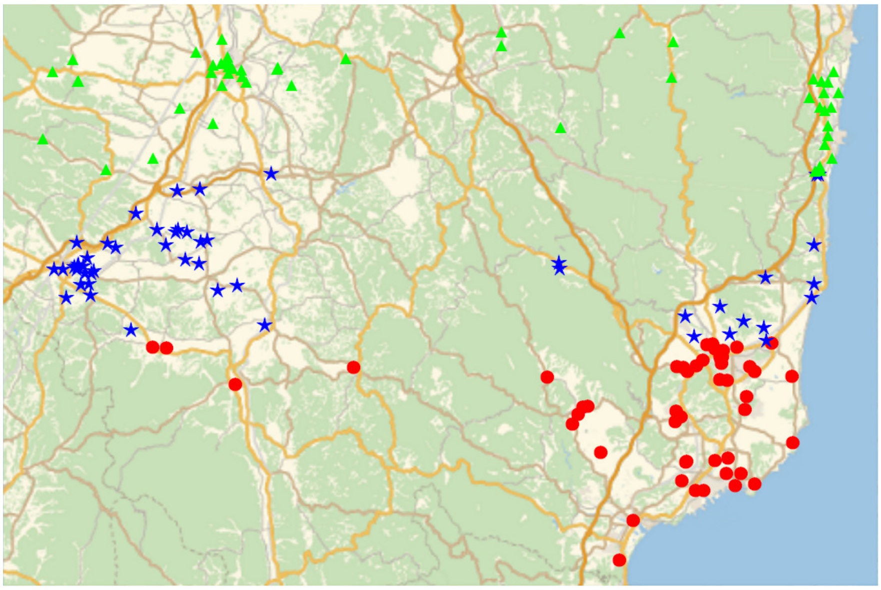


Figure 5.1: Map of Fukushima Prefecture, where red *circles*, blue *stars*, and green *triangles* are the observation points for the effective dose in lower, middle, and upper areas, respectively.

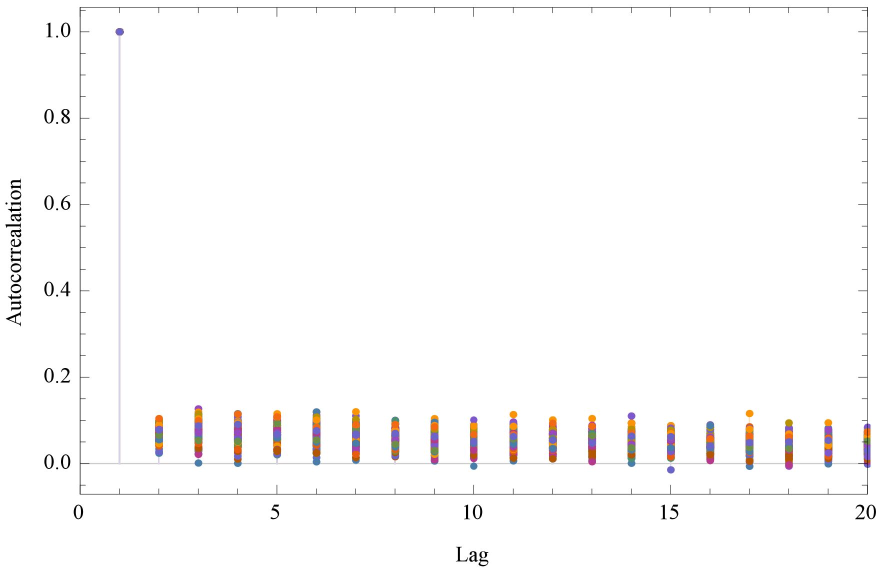


Figure 5.2: Dots in the graph show the sample autocorrelations of the radioactive observations.

Chapter 6

Higher-order approximation of the distribution of test statistics for high-dimensional time-series ANOVA models

Analysis of variance (ANOVA) is a type of hypothesis testing method for the null hypothesis of "no treatment effect". It is generally used to test the null hypothesis that the means of three or more populations of within-group means are all equal. Moreover, this method shows whether the within-group means are equal.

ANOVA has a long history in statistics. Gauss founded it in the late 1800s, and Markoff developed it in the early 1900s. Many test statistics for ANOVA and multivariate analysis of variance (MANOVA) have been proposed, primarily under independent disturbances of a MANOVA model. The early applications can be found in Hooke (1926) and Wishart (1938). In addition, Bishop (1939) and Box (1949) obtained general theoretical results. They derived asymptotic expansions of the null and non-null distributions of the likelihood ratio test-statistics. Bhattacharya and Rao (1986) discussed higher-order approximations (Edgeworth expansions) and their validity. Furthermore, Fujikoshi et al. (2011) developed higher-order asymptotic expansions of the null and non-null distributions of the likelihood ratio test statistic, Lawley-Hotelling test statistic, and Bartlett-Nanda-Pillai test statistic under high-dimensional and i.i.d. settings. Moreover, in a time-series analysis, Taniguchi and Kakizawa (2000) discussed the Edgeworth expansions for various statistics. Recently, under a high-dimensional time-series setting, Nagahata and Taniguchi (2018) discussed the first-order asymptotics of Lawley-Hotelling test statistic, likelihood ratio test statistic, and Bartlett-Nanda-Pillai test statistic.

In the current era of big data, an analysis of high-dimensional time-series data is required in practical problems, such as those in economics, finance, and bioinformatics. Especially, the accuracy of statistical decisions for high-dimensional time-series data has become increasingly important. Many data analysts need accurate methods for the equivalence of the within-group means of big data, because this analysis is very basic. MANOVA will be useful for these needs. However, from the viewpoint of the numerical accuracy of approximations, higher-order asymptotics of ANOVA test statistics for high-dimensional data are not adequately developed. In the present study, we focus on Edgeworth expansions of distributions of Lawley-Hotelling test statistic, likelihood ratio test statistic, and Bartlett-Nanda-Pillai test statistic.

In this chapter, we consider a one-way MANOVA model whose disturbance process is generated by a high-dimensional stationary process.

6.1 Problems and Preliminaries

Throughout this chapter, we consider the MANOVA model under which a q-tuple of p-dimensional time series $\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{in_i}$, $i = 1, \ldots, q$ satisfies

$$\boldsymbol{X}_{it} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{it}, \quad t = 1, \cdots, n_i, \quad i = 1, \cdots, q, \tag{6.1}$$

where $\mu \in \mathbb{R}^p$ is the global mean of the model (6.1), the disturbances $\boldsymbol{\epsilon}_i \equiv \{\boldsymbol{\epsilon}_{i1}, \cdots \boldsymbol{\epsilon}_{in_i}\}$ are kth-order stationary with mean $\mathbf{0}$, lag u autocovariance matrix $\boldsymbol{\Gamma}(u) = (\Gamma_{jk}(u))_{1 \le j,k \le p}$, $u \in \mathbb{Z}$, and n_i is the observation length of the ith group. Furthermore, the total observation length of all groups $n = \sum_{i=1}^{q} n_i$ and $\{\boldsymbol{\epsilon}_i\}$, $i = 1, \cdots, q$ are mutually independent. We impose a further standard assumption, which is called homoscedasticity (e.g., Ch. 8.9 of Anderson (2003)). Now $\boldsymbol{\alpha}_i$ denotes the effect of the ith treatment, which measures the deviation from $\boldsymbol{\mu}$ satisfying $\sum_{i=1}^{q} \boldsymbol{\alpha}_i = \mathbf{0}$. Because the treatment effects sum to zero, we discuss the problem of testing:

$$H: \ \boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_q = \mathbf{0} \text{ vs. } A: \ \boldsymbol{\alpha}_i \neq \mathbf{0} \text{ for some } i. \tag{6.2}$$

The null hypothesis H implies that all effects are zero.

For our high-dimensional dependent observations, we use the Lawley-Hotelling test statistic $\tilde{T}_1$, likelihood ratio test statistic $\tilde{T}_2$, and Bartlett-Nanda-Pillai test statistic $\tilde{T}_3$:

$$\begin{aligned}\tilde{T}_1 &\equiv n\mathrm{tr}\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1},\\ \tilde{T}_2 &\equiv -n\log|\hat{\mathcal{S}}_E|/|\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H|,\\ \tilde{T}_3 &\equiv n\mathrm{tr}\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H)^{-1},\end{aligned}$$

where

$$\hat{\mathcal{S}}_H \equiv \sum_{i=1}^{q} n_i(\hat{\boldsymbol{X}}_{i\cdot}-\hat{\boldsymbol{X}}_{\cdot\cdot})(\hat{\boldsymbol{X}}_{i\cdot}-\hat{\boldsymbol{X}}_{\cdot\cdot})' \text{ and } \hat{\mathcal{S}}_E \equiv \sum_{i=1}^{q}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})' \text{ with}$$

$$\hat{\boldsymbol{X}}_{i\cdot} = \frac{1}{n_i}\sum_{t=1}^{n_i}\boldsymbol{X}_{it} \text{ and } \hat{\boldsymbol{X}}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{q}\sum_{t=1}^{n_i}\boldsymbol{X}_{it}.$$

Now, we call $\hat{\mathcal{S}}_H$ and $\hat{\mathcal{S}}_E$ the between-group sums of squares and products (SSP) and the within-group SSP, respectively. To derive the stochastic expansion of $n^{-1}\hat{\mathcal{S}}_E$ in Section 6.3, we introduce

$$\hat{\mathcal{S}}_i \equiv (n_i-1)^{-1}\sum_{t=1}^{n_i}(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})(\boldsymbol{X}_{it}-\hat{\boldsymbol{X}}_{i\cdot})', \tag{6.3}$$

$$\boldsymbol{V} = \sum_{i=1}^{q}\sqrt{\frac{n_i}{n}}\boldsymbol{V}_i,\ \ \boldsymbol{V}_i = \sqrt{n_i}(\hat{\mathcal{S}}_i-\boldsymbol{I}_p). \tag{6.4}$$

In addition, to derive the Edgeworth expansion of distributions of the three test statistics under H, we impose Assumptions 5.1.1, 5.1.2, and 5.1.3.

6.2 Main Results

In what follows, without loss of generality, we assume $\boldsymbol{\Gamma}(0)=\boldsymbol{I}_p$, and $\boldsymbol{\mu}=\boldsymbol{0}$ because the three test statistics $\tilde{T}_1$, $\tilde{T}_2$, and $\tilde{T}_3$ are invariant under linear transformation, our discussion for $\boldsymbol{X}_{it}$ remains valid for the case where we apply a linear transformation $\{\boldsymbol{\Gamma}(0)\}^{-1/2}$ to $\boldsymbol{X}_{it}$. We derive the stochastic expansion of the standardized versions T_1, T_2, and T_3 of the three test statistics $\tilde{T}_1$ (Lawley-Hotelling test statistic), $\tilde{T}_2$ (likelihood ratio test statistic), and $\tilde{T}_3$ (Bartlett-Nanda-Pillai test statistic), respectively:

$$T_1 \equiv \frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H\hat{\mathcal{S}}_E^{-1} - \sqrt{p}(q-1)\right\}, \tag{6.5}$$

$$T_2 \equiv -\frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\log|\hat{\mathcal{S}}_E|/|\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H| + \sqrt{p}(q-1)\right\}, \tag{6.6}$$

$$T_3 \equiv \frac{1}{\sqrt{2(q-1)}}\left\{\frac{n}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H(\hat{\mathcal{S}}_E+\hat{\mathcal{S}}_H)^{-1} - \sqrt{p}(q-1)\right\}. \tag{6.7}$$

This section provides their Edgeworth expansions. Lemmas and all proofs are provided in Section 6.3.

Theorem 6.2.1 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis H, we have the following Edgeworth expansions:*

$$\begin{aligned} P(T_i < z) &= \Phi(z) - \phi(z)\left\{p^{-1/2}\cdot\frac{c_3}{6}(z^2-1) + p^{-1}\cdot\frac{c_4}{24}(z^3-3z)\right\} \\ &\quad + \boldsymbol{o}\left(p^{-1}\right), \quad (i=1,2,3) \end{aligned} \tag{6.8}$$

where

$$\Phi(z) = \int_{-\infty}^{z}\phi(y)dy, \quad \phi(y) = (2\pi)^{-1/2}exp\left(-\frac{y^2}{2}\right),$$

and

$$c_3 = \left(\frac{2}{q-1}\right)^{3/2}\left\{q-3+3\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^2 - \sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^3\right\},$$

$$c_4 = \left(\frac{2}{q-1}\right)^{2}\left\{q-4+6\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^2 - 4\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^3 - \sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^4\right\}.$$

Remark 6.2.1 *This asymptotic result is an extended version of Fujikoshi et al. (2011) and Nagahata and Taniguchi (2018). Our setting in Section 6.1 shows we can apply this result to not only high-dimensional i.i.d. data (that was discussed in Fujikoshi et al. (2011)) but also high-dimensional time series data. Also, an approximation of the three test statistics T_i, $i = 1, 2, 3$ in Theorem 6.2.1 is more accurate than one of them in Nagahata and Taniguchi (2018) because we investigated the higher order asymptotic structure of T_i, $i = 1, 2, 3$ by using Edgeworth expansion method.*

6.3 Asymptotic theory for main results

In this section, we provide the lemmas and their proofs. In what follows, we use the same linear transformation as in Section 6.2. First, the stochastic expansion of $n^{-1}\hat{\mathcal{S}}_E$ and $\hat{\mathcal{S}}_H$ is given.

Lemma 6.3.1 *Suppose Assumptions 5.1.1-5.1.3. Then, under null hypothesis H, the following (6.9)-(6.11) hold true;*

$$\frac{1}{n}\hat{\mathcal{S}}_E = \boldsymbol{I}_p + \frac{1}{\sqrt{n}}\boldsymbol{V} - \frac{q}{n}\boldsymbol{I}_p + \boldsymbol{O}_{\mathrm{P}}^U\left(n^{-3/2}\right), \tag{6.9}$$

$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \boldsymbol{I}_p - \frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{1}{n}(\boldsymbol{V}^2 + q\boldsymbol{I}_p) + \boldsymbol{O}_{\mathrm{P}}^U\left(n^{-3/2}\right), \tag{6.10}$$

$$\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^U(1). \tag{6.11}$$

Proof (Lemma 6.3.1) *By (6.4), write $n^{-1}\hat{\mathcal{S}}_E$ as*

$$\begin{aligned}\frac{1}{n}\hat{\mathcal{S}}_E &= \frac{1}{n}\sum_{i=1}^{q}(n_i-1)\hat{\mathcal{S}}_i \\ &= \frac{1}{n}\sum_{i=1}^{q}(n_i-1)\left(\boldsymbol{I}_p + \frac{1}{\sqrt{n_i}}\boldsymbol{V}_i\right) \\ &= \boldsymbol{I}_p + \frac{1}{\sqrt{n}}\boldsymbol{V} - \frac{q}{n}\boldsymbol{I}_p - \frac{1}{n}\sum_{i=1}^{q}\frac{1}{\sqrt{n_i}}\boldsymbol{V}_i. \end{aligned} \tag{6.12}$$

In what follows, for each i, we will show $\boldsymbol{V}_i = \boldsymbol{O}_{\mathrm{P}}^U(1)$. By the null hypothesis H and $\boldsymbol{\mu} = \boldsymbol{0}$, we rewrite $\hat{\mathcal{S}}_i$ as follows:

$$\begin{aligned}\hat{\mathcal{S}}_i &= n_i(n_i-1)^{-1}\left(\frac{1}{n_i}\sum_{t=1}^{n_i}\boldsymbol{X}_{it}\boldsymbol{X}_{it}' - \hat{\boldsymbol{X}}_{i\cdot}\hat{\boldsymbol{X}}_{i\cdot}'\right) \\ &= n_i(n_i-1)^{-1}(\boldsymbol{A}-\boldsymbol{B}) \ \text{(say)}, \end{aligned} \tag{6.13}$$

where $\boldsymbol{A} = 1/n_i\sum_{t=1}^{n_i}\boldsymbol{X}_{it}\boldsymbol{X}_{it}'$ and $\boldsymbol{B} = \hat{\boldsymbol{X}}_{i\cdot}\hat{\boldsymbol{X}}_{i\cdot}'$. We observe

$$\begin{aligned}&E\{\boldsymbol{A}\} = \boldsymbol{I}_p \quad \text{and} \\ &Cov\{A_{jk}, A_{lm}\} \\ &= \frac{1}{n_i}\sum_{s=-n_i+1}^{n_i-1}\left(1-\frac{|s|}{n_i}\right)\{c_{jl}(s)c_{km}(s) + c_{jm}(s)c_{kl}(s) + c^i_{jklm}(0,s,s)\} \qquad (6.14) \\ &= \boldsymbol{O}\left(\frac{1}{n_i}\right) = \boldsymbol{O}\left(\frac{1}{n}\right) \quad \text{uniformly in } j,k,l,m \text{ by Assumption 5.1.2.}\end{aligned}$$

Hence, $\boldsymbol{A} = \boldsymbol{I}_p + \boldsymbol{O}^U_{\mathrm{P}}\left(1/\sqrt{n}\right)$. *Next, we observe*

$$\begin{aligned}
& E(\hat{\boldsymbol{X}}_{i\cdot}) = \boldsymbol{\alpha}_i \quad \text{and} \\
& Cov\{\hat{\boldsymbol{X}}_{i\cdot}, \hat{\boldsymbol{X}}_{i\cdot}\} \\
& \quad = \left\{ \frac{1}{n_i} \sum_{s=-n_i+1}^{n_i-1} \left(1 - \frac{|s|}{n_i}\right) c_{jk}(s) \right\} \qquad (6.15) \\
& \quad = \boldsymbol{O}^U\left(\frac{1}{n_i}\right).
\end{aligned}$$

Thus,

$$\boldsymbol{B} = \boldsymbol{O}^U_{\mathrm{P}}\left(\frac{1}{n}\right). \qquad (6.16)$$

Therefore,

$$\hat{\mathcal{S}}_i = \boldsymbol{I}_p + \boldsymbol{O}^U_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right),$$

and

$$\boldsymbol{V}_i = \boldsymbol{O}^U_{\mathrm{P}}(1). \qquad (6.17)$$

By using (6.12) and (6.17), we can get

$$\frac{1}{n}\hat{\mathcal{S}}_E = \boldsymbol{I}_p + \frac{1}{\sqrt{n}}\boldsymbol{V} - \frac{q}{n}\boldsymbol{I}_p + \boldsymbol{O}^U_{\mathrm{P}}\left(n^{-3/2}\right), \qquad (6.9)$$

and

$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \left\{\boldsymbol{I}_p + \frac{1}{\sqrt{n}}\boldsymbol{V} - \frac{q}{n}\boldsymbol{I}_p + \boldsymbol{O}^U_{\mathrm{P}}\left(n^{-3/2}\right)\right\}^{-1} = \{\boldsymbol{I}_p - \boldsymbol{M}_n\}^{-1} \text{ (say).}$$

It is known that

$$\{\boldsymbol{I}_p - \boldsymbol{M}_n\}^{-1} = \sum_{k=0}^{\infty} \boldsymbol{M}_n^k \qquad (6.18)$$

(see p. 169 of Magnus and Neudecker (1999)). From Assumption 5.1.1, it follows that

$$\begin{aligned}
\boldsymbol{M}_n^0 &= \boldsymbol{I}_p, \\
\boldsymbol{M}_n &= -\frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{q}{n}\boldsymbol{I}_p + \boldsymbol{O}^U_{\mathrm{P}}\left(n^{-3/2}\right), \\
\boldsymbol{M}_n^2 &= \frac{1}{n}\boldsymbol{V}^2 + \boldsymbol{O}^U_{\mathrm{P}}\left(n^{-3/2}\right), \\
\boldsymbol{M}_n^k &= \boldsymbol{O}_{\mathrm{P}}\left(n^{-\frac{k}{2}}\right)\boldsymbol{H}, \quad k \geq 3,
\end{aligned}$$

where $\boldsymbol{H}$ *is a* $p \times p$*-matrix and* $\boldsymbol{H} = \boldsymbol{O}_{\mathrm{P}}^{U}(1)$. *Then, we obtain*

$$\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1} = \boldsymbol{I}_p - \frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{1}{n}(\boldsymbol{V}^2 + q\boldsymbol{I}_p) + \boldsymbol{O}_{\mathrm{P}}^{U}\left(n^{-3/2}\right). \tag{6.10}$$

Next, we show $\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^{U}(1)$. *To this end, we recall*

$$\hat{\mathcal{S}}_H = \sum_{i=1}^{q} n_i(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})(\hat{\boldsymbol{X}}_{i\cdot} - \hat{\boldsymbol{X}}_{\cdot\cdot})'. \tag{6.19}$$

From (6.15), we observe that $\hat{\boldsymbol{X}}_{i\cdot} = \boldsymbol{\alpha}_i + \boldsymbol{O}_{\mathrm{P}}^{U}\left(1/\sqrt{n_i}\right)$, $\sum_{i=1}^{q}\boldsymbol{\alpha}_i = \boldsymbol{0}$, *and similarly,* $\hat{\boldsymbol{X}}_{\cdot\cdot} = \boldsymbol{O}_{\mathrm{P}}^{U}(1/\sqrt{n})$. *Thus, we have*

$$\hat{\mathcal{S}}_H = \boldsymbol{O}_{\mathrm{P}}^{U}(1). \tag{6.11}$$

Note that (6.9), (6.10), and (6.11) are derived for the multivariate i.i.d. case, e.g., (Fujikoshi et al., 2011, p.164).

Lemma 6.3.2 *Suppose Assumptions 5.1.1-5.1.3. Then, under null hypothesis* H, *it holds that*

$$\tilde{T}_i = U^{(0)} + \frac{1}{\sqrt{n}}U^{(1)} + \frac{1}{n}\left(U^{(2)} + \beta_i R^{(2)}\right) + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^{3/2}}{n}\right), \ i = 1, 2, 3, \tag{6.20}$$

where

$$\begin{aligned} U^{(0)} &= \mathrm{tr}\hat{\mathcal{S}}_H, \\ U^{(1)} &= -\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\}, \\ U^{(2)} &= \mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\}, \\ R^{(2)} &= \mathrm{tr}\{\hat{\mathcal{S}}_H^2\}, \ \textit{and} \\ (\beta_1, \beta_2, \beta_3) &= \left(0, -\frac{1}{2}, -1\right). \end{aligned}$$

Proof (Lemma 6.3.2) *From Lemma 6.3.1, it follows that*

$$\begin{aligned} \tilde{T}_1 &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left\{\frac{1}{n}\hat{\mathcal{S}}_E\right\}^{-1}\right] \\ &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left\{\boldsymbol{I}_p - \frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{1}{n}(\boldsymbol{V}^2 + q\boldsymbol{I}_p) + \boldsymbol{O}_{\mathrm{P}}^{U}\left(n^{-3/2}\right)\right\}\right] \\ &= \mathrm{tr}\hat{\mathcal{S}}_H - \frac{1}{\sqrt{n}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\} + \frac{1}{n}\mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\} + \mathrm{tr}\left\{\hat{\mathcal{S}}_H \cdot \boldsymbol{O}_{\mathrm{P}}^{U}\left(n^{-3/2}\right)\right\}. \end{aligned}$$

From (6.11),

$$\tilde{T}_1 = \mathrm{tr}\hat{\mathcal{S}}_H - \frac{1}{\sqrt{n}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\} + \frac{1}{n}\mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\} + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^{3/2}}{n}\right). \tag{6.20}$$

Next, to derive (6.20), first, note that for every matrix $\boldsymbol{F}$ and the matrix differential operator d

$$\begin{aligned} d\log|\boldsymbol{F}| &= \mathrm{tr}(\boldsymbol{F}^{-1}d\boldsymbol{F}), \\ d\boldsymbol{F}^{-1} &= -\boldsymbol{F}^{-1}(d\boldsymbol{F})\boldsymbol{F}^{-1}, \end{aligned}$$

and (6.18) (e.g., Magnus and Neudecker (1999)). Then, a modification of Proposition 6.1.5 of Brockwell and Davis (1991) and Lemma 6.3.1 shows that for

$$f := n\log\left|\boldsymbol{I}_p + \frac{1}{n}\hat{\mathcal{S}}_H\left\{\frac{1}{n}\hat{\mathcal{S}}_E^{-1}\right\}\right|,$$

we have that

$$f = \sum_{m=0}^{\infty}\frac{1}{m!}d^m f.$$

where d^m's are m-th differentials of f which are calculated by

$$\begin{aligned} d^0 f &= \mathrm{tr}\{\hat{\mathcal{S}}_H\} - \frac{1}{2n}\mathrm{tr}\{\hat{\mathcal{S}}_H^2\} + \boldsymbol{O}_{\mathrm{P}}\left(p\cdot n^{-2}\right), \\ d^1 f &= -\frac{1}{\sqrt{n}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\} + \frac{1}{n}\mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\} + \boldsymbol{O}_{\mathrm{P}}\left(p^2\cdot n^{-3/2}\right), \\ d^m f &= \boldsymbol{O}_{\mathrm{P}}\left(p\cdot n^{-2}\right), \quad m \geq 2. \end{aligned}$$

Thus, we obtain

$$\tilde{T}_2 = \mathrm{tr}\{\hat{\mathcal{S}}_H\} - \frac{1}{\sqrt{n}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\} + \frac{1}{n}\left[\mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\} - \frac{1}{2}\mathrm{tr}\{\hat{\mathcal{S}}_H^2\}\right] + \boldsymbol{O}_{\mathrm{P}}\left(p^2\cdot n^{-3/2}\right). \tag{6.20}$$

From Lemma 6.3.1 and (6.18), it follows that

$$\begin{aligned} \tilde{T}_3 &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left\{\frac{1}{n}\hat{\mathcal{S}}_E + \frac{1}{n}\hat{\mathcal{S}}_H\right\}^{-1}\right] \\ &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\left\{\boldsymbol{I}_p + \frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{1}{n}(\hat{\mathcal{S}}_H - q\boldsymbol{I}_p) + \boldsymbol{O}_{\mathrm{P}}^U\left(n^{-3/2}\right)\right\}^{-1}\right] \\ &= \mathrm{tr}\left[\hat{\mathcal{S}}_H\sum_{k=0}^{\infty}\left\{-\frac{1}{\sqrt{n}}\boldsymbol{V} - \frac{1}{n}(\hat{\mathcal{S}}_H - q\boldsymbol{I}_p) + \boldsymbol{O}_{\mathrm{P}}^U\left(n^{-3/2}\right)\right\}^k\right]. \end{aligned}$$

From (6.11),

$$\tilde{T}_3 = \mathrm{tr}\hat{\mathcal{S}}_H - \frac{1}{\sqrt{n}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\} + \frac{1}{n}[\mathrm{tr}\{\hat{\mathcal{S}}_H(\boldsymbol{V}^2 + q\boldsymbol{I}_p)\} - \mathrm{tr}\{\hat{\mathcal{S}}_H^2\}] + \boldsymbol{O}_{\mathrm{P}}\left(\frac{p^{3/2}}{n}\right) \quad (6.20)$$

(for the multivariate i.i.d. case, e.g., (Fujikoshi et al., 2011, p.164)).

Lemma 6.3.3 *Suppose Assumptions 5.1.1-5.1.3. Then, under the null hypothesis H, it holds that*

$$\begin{aligned}
&cum^{(J)}(\overbrace{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H, \cdots}^{K}, \overbrace{-\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\}, \cdots}^{L}, \overbrace{\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}^2\}, \cdots}^{M}, \overbrace{\frac{q}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H, \cdots}^{M_0}, \overbrace{\frac{\beta_i}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H^2, \cdots}^{N}) \\
&= \boldsymbol{O}\left(p^{1-J/2+N} \cdot n^{-2L-4M-M_0-N}\right) \quad (6.21)\\
&= \boldsymbol{o}\left(p^{1-J/2-6L-12M-3M_0-2N}\right), \quad (6.22)
\end{aligned}$$

where $K, L, M, M_0, N \geq 0$, $J = K + L + M + +M_0 + N \geq 1$ and

$$(\beta_1, \beta_2, \beta_3) = \left(0, -\frac{1}{2}, -1\right).$$

Proof (Lemma 6.3.3) *First, under $\boldsymbol{\mu} = \boldsymbol{0}$ and null hypothesis H, we prepare S_{jk} and V_{jk} as (j,k)th components of $\hat{\mathcal{S}}_H$ and $\boldsymbol{V}$, respectively:*

$$S_{jk} = \sum_{i_1=1}^{q} \frac{1}{n_{i_1}} \sum_{r=1}^{n_{i_1}} \sum_{s=1}^{n_{i_1}} \epsilon_{i_1 r}^{(j)} \epsilon_{i_1 s}^{(k)} - \frac{1}{n} \sum_{i_2=1}^{q} \sum_{i_3=1}^{q} \sum_{t=1}^{n_{i_2}} \sum_{u=1}^{n_{i_3}} \epsilon_{i_2 t}^{(j)} \epsilon_{i_3 u}^{(k)}, \quad (6.23)$$

$$V_{jk} = \frac{1}{\sqrt{n}} \sum_{i_4=1}^{q} \frac{n_{i_4}}{n_{i_4} - 1} \sum_{r=1}^{n_{i_4}} \epsilon_{i_4 r}^{(j)} \epsilon_{i_4 r}^{(k)} - \frac{1}{\sqrt{n}} \sum_{i_4=1}^{q} \frac{1}{n_{i_4} - 1} \sum_{s=1}^{n_{i_4}} \sum_{t=1}^{n_{i_4}} \epsilon_{i_4 s}^{(j)} \epsilon_{i_4 t}^{(k)} - \sqrt{n}\delta_{jk}. \quad (6.24)$$

Here, we can write

$$\begin{aligned}
&cum^{(J)}(\overbrace{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H, \cdots}^{K}, \overbrace{-\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\}, \cdots}^{L}, \overbrace{\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}^2\}, \cdots}^{M}, \overbrace{\frac{q}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H, \cdots}^{M_0}, \overbrace{\frac{\beta_i}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H^2, \cdots}^{N}) \\
&= (-1)^L q^{M_0} \beta_i^N \cdot p^{-J/2} n^{-L/2-M-M_0-N} \\
&\quad \times \ cum^{(J)}(\overbrace{\mathrm{tr}\hat{\mathcal{S}}_H, \cdots}^{K+M_0}, \overbrace{\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\}, \cdots}^{L}, \overbrace{\mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}^2\}, \cdots}^{M}, \overbrace{\mathrm{tr}\hat{\mathcal{S}}_H^2, \cdots}^{N}). \quad (6.25)
\end{aligned}$$

By (6.23) and (6.24), a typical term of the cumulant in (6.25) is

$$\sum_{j_{1,1}}^{p}\cdots\sum_{j_{1,K+M_0}}^{p}\sum_{j_{2,1}}^{p}\cdots\sum_{j_{2,L}}^{p}\sum_{j_{3,1}}^{p}\cdots\sum_{j_{3,M}}^{p}\sum_{j_{4,1}}^{p}\sum_{k_{4,1}}^{p}\cdots\sum_{j_{4,N}}^{p}\sum_{k_{4,N}}^{p}$$
$$\sum_{r_{1,1}}^{n_i}\sum_{s_{1,1}}^{n_i}\cdots\sum_{r_{1,K+M_0}}^{n_i}\sum_{s_{1,K+M_0}}^{n_i}\sum_{r_{2,1}}^{n_i}\sum_{s_{2,1}}^{n_i}\cdots\sum_{r_{2,L}}^{n_i}\sum_{s_{2,L}}^{n_i}\sum_{r_{3,1}}^{n_i}\sum_{s_{3,1}}^{n_i}\cdots\sum_{r_{3,M}}^{n_i}\sum_{s_{3,M}}^{n_i}$$
$$\sum_{r_{4,1}}^{n_i}\sum_{s_{4,1}}^{n_i}\sum_{t_{4,1}}^{n_i}\sum_{u_{4,1}}^{n_i}\cdots\sum_{r_{4,N}}^{n_i}\sum_{s_{4,N}}^{n_i}\sum_{t_{4,N}}^{n_i}\sum_{u_{4,N}}^{n_i}\boldsymbol{O}\left(n^{-K-5L/2-4M-M_0-2N}\right)$$
$$\times\ cum^{(J)}[\epsilon_{ir_{1,1}}^{(j_{1,1})}\epsilon_{is_{1,1}}^{(j_{1,1})},\cdots,\epsilon_{ir_{2,1}}^{(j_{2,1})}\epsilon_{is_{2,1}}^{(j_{2,1})},\cdots,\epsilon_{ir_{3,1}}^{(j_{3,1})}\epsilon_{is_{3,1}}^{(j_{3,1})},\cdots,$$
$$\epsilon_{ir_{4,1}}^{(j_{4,1})}\epsilon_{is_{4,1}}^{(k_{4,1})}\epsilon_{it_{4,1}}^{(k_{4,1})}\epsilon_{iu_{4,1}}^{(j_{4,1})},\cdots]. \tag{6.26}$$

By using the properties of the cumulant and Theorem 2.3.2 in (Brillinger, 2001, p.19-21), the cumulant appearing in (6.26) has a typical main-order term

$$\begin{aligned}
&\boldsymbol{O}\left(n^{-K-5L/2-4M-M_0-2N}\right)\ n_i^{K+L+M+M_0+2N}\\
&\times\sum_{j_{1,1}}^{p}\cdots\sum_{j_{1,K+M_0}}^{p}\sum_{j_{2,1}}^{p}\cdots\sum_{j_{2,L}}^{p}\sum_{j_{3,1}}^{p}\cdots\sum_{j_{3,M}}^{p}\sum_{j_{4,1}}^{p}\cdots\sum_{j_{4,N}}^{p}c_{j_{1,1}j_{1,2}}(0)\cdots c_{j_{1,K+M_0}j_{2,1}}(0)\\
&\quad\times\ c_{j_{2,1}j_{2,2}}(0)\cdots c_{j_{2,L}j_{3,1}}(0)\ c_{j_{3,1}j_{3,2}}(0)\cdots c_{j_{3,M}j_{4,1}}(0)\ c_{j_{4,1}j_{4,2}}(0)\cdots c_{j_{4,N}j_{1,1}}(0)\\
&\times\sum_{k_{4,1}}^{p}\cdots\sum_{k_{4,N}}^{p}c_{k_{4,1}k_{4,1}}(0)\cdots c_{k_{4,N}k_{4,N}}(0)\\
=\ &\boldsymbol{O}\left(n^{-K-5L/2-4M-M_0-2N}\right)\ n_i^{K+L+M+M_0+2N}\qquad(\text{By Assumption 5.1.3 and } \boldsymbol{\Gamma}(0)=\boldsymbol{I}_p)\\
&\times\sum_{j}^{p}c_{jj}(0)\cdots c_{jj}(0)\times\sum_{k_{4,1}}^{p}\cdots\sum_{k_{4,N}}^{p}c_{k_{4,1}k_{4,1}}(0)\cdots c_{k_{4,N}k_{4,N}}(0)\\
=\ &\boldsymbol{O}\left(p^{1+N}\cdot n^{-3L/2-3M}\right).
\end{aligned}\tag{6.27}$$

Thus, from (6.27), we rewrite a typical term of (6.25) as

$$\begin{aligned}
&cum^{(J)}(\overbrace{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H,\cdots}^{K},\overbrace{-\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\},\cdots}^{L},\overbrace{\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}^2\},\cdots}^{M},\overbrace{\frac{q}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H,\cdots}^{M_0},\overbrace{\frac{\beta_i}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H^2,\cdots}^{N}),\\
&=p^{-J/2}n^{-L/2-M-M_0-N}\boldsymbol{O}\left(p^{1+N}\cdot n^{-3L/2-3M}\right)\\
&=\boldsymbol{O}\left(p^{1-J/2+N}\cdot n^{-2L-4M-M_0-N}\right)\\
&=\boldsymbol{o}\left(p^{1-J/2-6L-12M-3M_0-2N}\right).\quad(\text{By Assumption 5.1.1})
\end{aligned}$$

Hence, we showed (6.21) and (6.22).

Lemma 6.3.4 *Suppose Assumptions 5.1.1-5.1.3. Define W_i for every $i = 1, 2, 3$ by*

$$\begin{aligned} W_i &= \frac{1}{\sqrt{2(q-1)}} \left\{ \frac{1}{\sqrt{p}} U^{(0)} + \frac{1}{\sqrt{p}n} U^{(1)} + \frac{1}{\sqrt{p}n} \left(U^{(2)} + \beta_i R^{(2)} \right) \right. \\ & \left. \quad - \sqrt{p}(q-1) \right\} \qquad (6.28) \\ &= \frac{1}{\sqrt{2(q-1)}} \left\{ \frac{1}{\sqrt{p}} \mathrm{tr} \hat{\mathcal{S}}_H - \frac{1}{\sqrt{p}n} \mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}\} + \frac{1}{\sqrt{p}n} \mathrm{tr}\{\hat{\mathcal{S}}_H \boldsymbol{V}^2\} + \frac{q}{\sqrt{p}n} \mathrm{tr} \hat{\mathcal{S}}_H \right. \\ & \left. \quad + \frac{\beta_i}{\sqrt{p}n} \mathrm{tr}\{\hat{\mathcal{S}}_H^2\} - \sqrt{p}(q-1) \right\}, \qquad (6.29) \\ (\beta_1, \beta_2, \beta_3) &= \left(0, -\frac{1}{2}, -1 \right). \end{aligned}$$

Then, under the null hypothesis H, the following (6.30)-(6.34) hold that

$$cum(W_i) = 0 + \boldsymbol{o}\left(p^{-1/2}\right), \qquad (6.30)$$

$$cum(W_i, W_i) = 1 + \boldsymbol{o}\left(p^{-1/2}\right), \qquad (6.31)$$

$$\begin{aligned} cum(W_i, W_i, W_i) = p^{-1/2} \left(\frac{2}{q-1} \right)^{3/2} \qquad (6.32) \\ \times \left\{ q - 3 + 3 \sum_{i=1}^{q} \left(\frac{n_i}{n} \right)^2 - \sum_{i=1}^{q} \left(\frac{n_i}{n} \right)^3 \right\} + \boldsymbol{o}\left(p^{-1/2}\right), \end{aligned}$$

$$\begin{aligned} cum^{(4)}(W_i, \cdots, W_i) = p^{-1} \left(\frac{2}{q-1} \right)^2 \qquad (6.33) \\ \times \left\{ q - 4 + 6 \sum_{i=1}^{q} \left(\frac{n_i}{n} \right)^2 - 4 \sum_{i=1}^{q} \left(\frac{n_i}{n} \right)^3 - \sum_{i=1}^{q} \left(\frac{n_i}{n} \right)^4 \right\} + \boldsymbol{o}\left(p^{-1}\right), \end{aligned}$$

$$cum^{(J)}(W_i, \cdots, W_i) = \boldsymbol{O}\left(p^{1-J/2}\right), \quad (J \geq 5) \qquad (6.34)$$

where (6.34) contains $K, L, M, M_0, N (\geq 0)$ of the first, second, third, fourth, and fifth terms of (6.29), respectively.

Proof (Lemma 6.3.4) *Now, from Lemma 6.3.3, we obtain from (6.28)*

$$cum(W_i) = \frac{1}{\sqrt{2(q-1)}} \left\{ \frac{1}{\sqrt{p}} \{E[U^{(0)}] - p(q-1)\} \right\} + \boldsymbol{o}\left(p^{-1/2}\right).$$

Here, under Assumptions 5.1.2 and 5.1.3, from (6.23), we get

$$\begin{aligned} E[U^{(0)}] &= \sum_{j=1}^{p} E[S_{jj}] \\ &= \sum_{j=1}^{p}\sum_{i_1=1}^{q}\sum_{s=-n_{i_1}+1}^{n_{i_1}-1}\left(1-\frac{|s|}{n_{i_1}}\right)c_{jj}(s) - \sum_{j=1}^{p}\sum_{i_2=1}^{q}\frac{n_{i_2}}{n}\sum_{r=-n_{i_2}+1}^{n_{i_2}-1}\left(1-\frac{|r|}{n_{i_2}}\right)c_{jj}(r) \\ &= p(q-1). \end{aligned} \tag{6.35}$$

Then, we can obtain

$$cum(W_i) = 0 + \boldsymbol{o}\left(p^{-1/2}\right). \quad (\text{By Assumption 5.1.1}) \tag{6.36}$$

Similarly, the main-order terms of $cum(W_i, W_i)$ *and* $cum(W_i, W_i, W_i)$ *can be computed as follows. From (6.11) and (6.15),*

$$\begin{aligned} cum(W_i, W_i) &= \frac{1}{2p(q-1)} cum(U^{(0)}, U^{(0)}) + \boldsymbol{o}\left(p^{-1/2}\right) \quad (\text{By Lemma 6.3.3}) \\ &= \frac{1}{2p(q-1)}\sum_{j=1}^{p}\sum_{k=1}^{p} cum(S_{jj}, S_{kk}) + \boldsymbol{o}\left(p^{-1/2}\right) \\ &= 1 + \boldsymbol{o}\left(p^{-1/2}\right). \end{aligned} \tag{6.37}$$

In addition, we can obtain

$$cum(W_i, W_i, W_i) = \{2p(q-1)\}^{-3/2} cum(U^{(0)}, U^{(0)}, U^{(0)}) + \boldsymbol{o}\left(p^{-1/2}\right), \quad (\text{By Lemma 6.3.3})$$

and

$$\begin{aligned} & cum(U^{(0)}, U^{(0)}, U^{(0)}) \\ = & \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} cum(S_{jj}, S_{kk}, S_{ll}) \\ = & \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}[\\ & \sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\sum_{i_3=1}^{q}\frac{1}{n_{i_1}}\frac{1}{n_{i_2}}\frac{1}{n_{i_3}}\sum_{r=1}^{n_{i_1}}\sum_{s=1}^{n_{i_1}}\sum_{t=1}^{n_{i_2}}\sum_{u=1}^{n_{i_2}}\sum_{v=1}^{n_{i_3}}\sum_{w=1}^{n_{i_3}} cum\{\epsilon_{i_1 r}^{(j)}\epsilon_{i_2 s}^{(j)}, \epsilon_{i_2 t}^{(k)}\epsilon_{i_2 u}^{(k)}, \epsilon_{i_3 v}^{(l)}\epsilon_{i_3 w}^{(l)}\} \\ & -3\frac{1}{n}\sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\sum_{i_3=1}^{q}\frac{1}{n_{i_1}}\frac{1}{n_{i_2}}\sum_{r=1}^{n_{i_1}}\sum_{s=1}^{n_{i_1}}\sum_{t=1}^{n_{i_2}}\sum_{u=1}^{n_{i_2}}\sum_{v=1}^{n_{i_3}}\sum_{w=1}^{n_{i_3}} cum\{\epsilon_{i_1 r}^{(j)}\epsilon_{i_2 s}^{(j)}, \epsilon_{i_2 t}^{(k)}\epsilon_{i_2 u}^{(k)}, \epsilon_{i_3 v}^{(l)}\epsilon_{i_3 w}^{(l)}\} \end{aligned}$$

$$+3\frac{1}{n^2}\sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\sum_{i_3=1}^{q}\frac{1}{n_{i_1}}\sum_{r=1}^{n_{i_1}}\sum_{s=1}^{n_{i_1}}\sum_{t=1}^{n_{i_2}}\sum_{u=1}^{n_{i_2}}\sum_{v=1}^{n_{i_3}}\sum_{w=1}^{n_{i_3}} cum\{\epsilon_{i_1r}^{(j)}\epsilon_{i_2s}^{(j)},\epsilon_{i_2t}^{(k)}\epsilon_{i_2u}^{(k)},\epsilon_{i_3v}^{(l)}\epsilon_{i_3w}^{(l)}\}$$

$$-\frac{1}{n^3}\sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\sum_{i_3=1}^{q}\sum_{r=1}^{n_{i_1}}\sum_{s=1}^{n_{i_1}}\sum_{t=1}^{n_{i_2}}\sum_{u=1}^{n_{i_2}}\sum_{v=1}^{n_{i_3}}\sum_{w=1}^{n_{i_3}} cum\{\epsilon_{i_1r}^{(j)}\epsilon_{i_2s}^{(j)},\epsilon_{i_2t}^{(k)}\epsilon_{i_2u}^{(k)},\epsilon_{i_3v}^{(l)}\epsilon_{i_3w}^{(l)}\}$$

$$]$$

$$\begin{aligned}
&= \sum_{j=1}^{p}\left\{\sum_{i_1=1}^{q}8c_{jj}(0)c_{jj}(0)c_{jj}(0) - 3\sum_{i_1=1}^{q}\frac{n_{i_1}}{n}\cdot 8c_{jj}(0)c_{jj}(0)c_{jj}(0)\right.\\
&\quad \left. +3\sum_{i_1=1}^{q}\left(\frac{n_{i_1}}{n}\right)^2 8c_{jj}(0)c_{jj}(0)c_{jj}(0) - \sum_{i_1=1}^{q}\left(\frac{n_{i_1}}{n}\right)^3\cdot 8c_{jj}(0)c_{jj}(0)c_{jj}(0)\right\}\\
&\quad +\boldsymbol{O}\left(p\cdot n^{-1}\right)\\
&= 8p\left\{q-3+3\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^2-\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^3\right\}+\boldsymbol{O}\left(p\cdot n^{-1}\right). \qquad (6.38)
\end{aligned}$$

Therefore,

$$cum(W_i,W_i,W_i) = p^{-1/2}\left(\frac{2}{q-1}\right)^{3/2}\left\{q-3+3\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^2-\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^3\right\}+\boldsymbol{o}\left(p^{-1/2}\right). \qquad (6.39)$$

Similarly, we can compute

$$\begin{aligned}
cum^{(4)}(W_i,\cdots,W_i) &= \{2p(q-1)\}^{-1}cum^{(4)}(U^{(0)},\cdots,U^{(0)})+\boldsymbol{o}\left(p^{-2}\right)\\
&= \{2p(q-1)\}^{-1}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{j_3=1}^{p}\sum_{j_4=1}^{p}cum(S_{j_1j_1},\cdots,S_{j_4j_4})+\boldsymbol{o}\left(p^{-1}\right)\\
&= p^{-1}\left(\frac{2}{q-1}\right)^2\left\{q-4+6\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^2-4\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^3-\sum_{i=1}^{q}\left(\frac{n_i}{n}\right)^4\right\}\\
&\quad +\boldsymbol{o}\left(p^{-1}\right). \qquad (6.40)
\end{aligned}$$

Hence, (6.30), (6.31), (6.32), and (6.33) were shown (from (6.36), (6.37), (6.39), and (6.40)). Furthermore, we discuss the J*th order for* $J \geq 5$ *cumulant* $cum^{(J)}(W_i,\cdots,W_i)$*. From Lemma 6.3.3, we obtain*

$$cum^{(J)}(W_i,\ldots,W_i) = \sum_{\substack{K,L,M,M_0,N;\\ K+L+M+M_0+N=J}}\{2(q-1)\}^{-J/2}$$

$$
\times cum^{(J)}(\overbrace{\frac{1}{\sqrt{p}}\mathrm{tr}\hat{\mathcal{S}}_H,\cdots}^{K},\overbrace{-\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}\},\cdots}^{L},\overbrace{\frac{1}{\sqrt{pn}}\mathrm{tr}\{\hat{\mathcal{S}}_H\boldsymbol{V}^2\},\cdots}^{M},\overbrace{\frac{q}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H,\cdots}^{M_0},\overbrace{\frac{\beta_i}{\sqrt{pn}}\mathrm{tr}\hat{\mathcal{S}}_H^2,\cdots}^{N})
$$

$$
\begin{aligned}
&= \sum_{\substack{K,L,M,M_0,N;\\ K+L+M+M_0+N=J}} \boldsymbol{O}\left(p^{1-J/2+N}\cdot n^{-2L-4M-M_0-N}\right) \\
&= \max_{K,L,M,M_0,N} \boldsymbol{O}\left(p^{1-J/2+N}\cdot n^{-2L-4M-M_0-N}\right) \\
&= \boldsymbol{O}\left(p^{1-J/2}\right). \quad (L=M=M_0=N=0)
\end{aligned}
$$

Then, (6.34) was shown.

Remark 6.3.1 *Nagahata and Taniguchi (2018) also evaluated the high-order cumulants of T_i, $i=1,2,3$ but there is a big difference between this chapter and Nagahata and Taniguchi (2018). The order of the stochastic expansion in Lemma 6.3.2 is higher than that in Nagahata and Taniguchi (2018), so we needed to derive asymptotics of W_i as in Lemmas 6.3.3 and 6.3.4.*

Proof (Theorem 6.2.1) *The Edgeworth expansion for a multivariate time series is derived by (Taniguchi and Kakizawa, 2000, p.168-170). We extend it to the case of high-dimensional time series. First, by the Taylor expansion and Lemma 6.3.4, we write the characteristic function of W_i $(i=1,2,3)$ in Lemma 6.3.4 as*

$$
\begin{aligned}
&E[exp\{\mathrm{it}W_i\}] \\
&= exp\Big\{cum(W_i)(\mathrm{it}) + \frac{1}{2}cum(W_i,W_i)(\mathrm{it})^2 + \frac{1}{6}cum(W_i,W_i,W_i)(\mathrm{it})^3 \\
&\qquad + \frac{1}{24}cum^{(4)}(W_i,\cdots,W_i)(\mathrm{it})^4 + \cdots\Big\} \\
&= exp\left(-\frac{t^2}{2}\right)\times\left\{1 + p^{-1/2}\cdot\frac{1}{6}cum(W_i,W_i,W_i)(\mathrm{it})^3 + p^{-1}\cdot\frac{1}{24}cum^{(4)}(W_i,\cdots,W_i)(\mathrm{it})^4\right\} \\
&+\boldsymbol{o}\left(p^{-1/2}\right). \\
&= exp\left(-\frac{t^2}{2}\right)\times\left\{1 + p^{-1/2}\cdot\frac{c_3}{6}(\mathrm{it})^3 + p^{-1}\cdot\frac{c_4}{24}(\mathrm{it})^4\right\} + \boldsymbol{o}\left(p^{-1/2}\right). \qquad (6.41)
\end{aligned}
$$

Inverting (6.41) by the Fourier inverse transform, we have

$$
P(W_i < z) \;=\; \Phi(z) - \phi(z)\left\{p^{-1/2}\cdot\frac{c_3}{6}(z^2-1) + p^{-1}\cdot\frac{c_4}{24}(z^3-3z)\right\} + \boldsymbol{o}\left(p^{-1/2}\right),
$$

where

$$\Phi(z) = \int_{-\infty}^{z} \phi(y)dy, \quad \phi(y) = (2\pi)^{-1/2} exp\left(-\frac{y^2}{2}\right).$$

Here, from Lemma 6.3.2, we observe that

$$E[exp\{\mathrm{it}T_i\}] = E[exp\{\mathrm{it}W_i\}] + \boldsymbol{o}\,(1)\,.$$

This implies (6.8), so we complete the proof.

6.4 Simulation to verify the finite sample performance

We simulate the Edgeworth expansions of distributions of $T_i,\ i = 1, 2, 3$, which are given by Theorem 6.2.1. In this section, our purpose is to show that their Edgeworth expansions $P(T_i < z),\ i = 1, 2, 3$ in (6.8) are more numerically accurate than the first-order approximation, that is, $\Phi(z)$ in (6.8). Specifically, in the case of an uncorrelated disturbance that is assumed by Assumption 5.1.3, DCC-$GARCH(1, 1)$ is a typical example of that process (see Engle (2002)). Therefore, we introduce the following five simulation process steps.

1 Set $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3 = \mathbf{0}$ for the null hypothesis H.

2 Generate 20-dimensional $\{\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{1,5000}\}, \{\boldsymbol{X}_{2,1}, \ldots, \boldsymbol{X}_{2,5000}\}, \{\boldsymbol{X}_{3,1}, \ldots, \boldsymbol{X}_{3,5000}\}$, with DCC-$GARCH(1, 1)$ disturbance.

3 Calculate the test statistics $T_i,\ i = 1, 2, 3$.

4 Repeat steps 2 and 3 1,000 times independently and obtain $\{T_i^{(1)}, \ldots, T_i^{(1000)};\ i = 1, 2, 3\}$.

5 Calculate $\hat{F}_{i,n}(z),\ i = 1, 2, 3$, which is the empirical distribution of $\{T_i^{(1)}, \ldots, T_i^{(1000)};\ i = 1, 2, 3\}$.

6 Write the plot of $|\hat{F}_{i,n}(z) - \Phi(z)|$ and $|\hat{F}_{i,n}(z) - P(T_i < z)|,\ i = 1, 2, 3$, which are plotted by dotted and thick lines, respectively, in Figures 6.1, 6.3, and 6.5.

7 Write the plot of $\{|\hat{F}_{i,n}(z) - \Phi(z)| - |\hat{F}_{i,n}(z) - P(T_i < z)|\},\ i = 1, 2, 3$, by a dotted line, in Figures 6.2, 6.4, and 6.6.

We set the 20-dimensional simulation from one-way MANOVA model (6.1) with a 20-dimensional vector $\boldsymbol{\mu}' = (1, \cdots, 1)'$ and generate the disturbance process $\{\boldsymbol{\epsilon}_{it}\}$ of observations $\{\boldsymbol{X}_{it}\}$ in (6.1) by using $DCC\text{-}GARCH(1,1)$, whose innovation term is assumed to be Gaussian. The scenarios of $DCC\text{-}GARCH(q,r)$ (see (**??**)) in $\boldsymbol{\epsilon}_{it}$ are

$$\begin{gathered} p = 20,\ i = 1,2,3,\ t = 1,\cdots,5000, \\ j = 1,\cdots,20, \\ q = r = 1, \\ a_j = 0.2,\ b_j = 0.7,\ c_j = 0.002, \\ \alpha = 0.1,\ \beta = 0.8, \\ \tilde{Q}_{kl} = 0.7^{(|k-l|)}, \end{gathered}$$

where $\tilde{\boldsymbol{Q}}_{kl}$ is the (k,l)-element of $\tilde{\boldsymbol{Q}}$. We set the observation length $n_i = 5000,\ i = 1,2,3$, because Table 1 of Section 5.1 in Nagahata and Taniguchi (2018) demonstrates that T_i are stable for $n_i = 2500$ or more uncorrelated observations $(i = 1,2,3)$. The Mathematical code and the "ccgarch" package of R are used for this algorithm. We compare the numerical accuracy of $P(T_i < z)$ with $\Phi(z)$ based on $\hat{F}_n(z)$ by using $|\hat{F}_{i,n}(z) - \Phi(z)|$, $|\hat{F}_{i,n}(z) - P(T_i < z)|$ (see Figures 6.1, 6.3, and 6.5), and $\{|\hat{F}_{i,n}(z) - \Phi(z)| - |\hat{F}_{i,n}(z) - P(T_i < z)|\},\ i = 1,2,3$ (see Figures 6.2, 6.4, and 6.6).

Figures 6.2, 6.4, and 6.6 indicate that the Edgeworth expansions $P(T_i < z)$ of T_i work better than the normal approximation $\Phi(z)$ from the perspective of numerical accuracy.

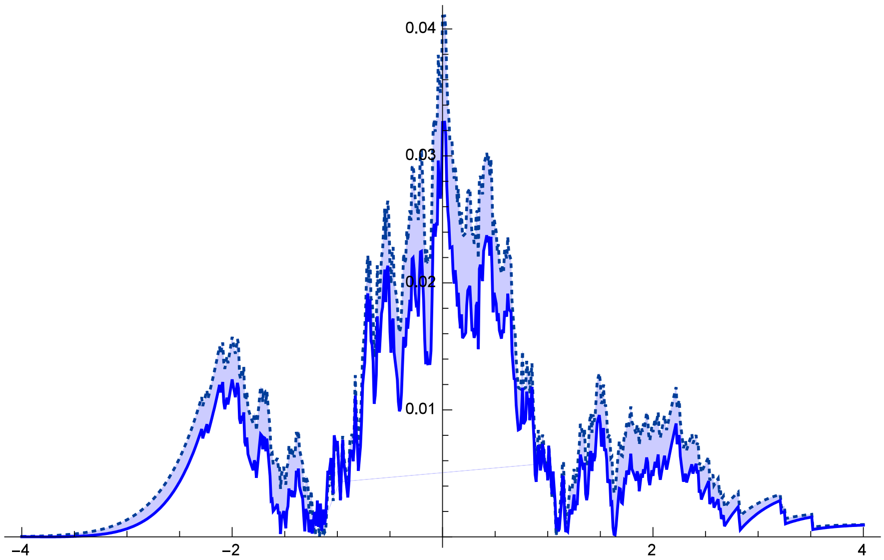


Figure 6.1: Plot of $|\hat{F}_{1,n}(z) - \Phi(z)|$ and $|\hat{F}_{1,n}(z) - P(T_1 < z)|$ by dotted and thick lines, respectively.

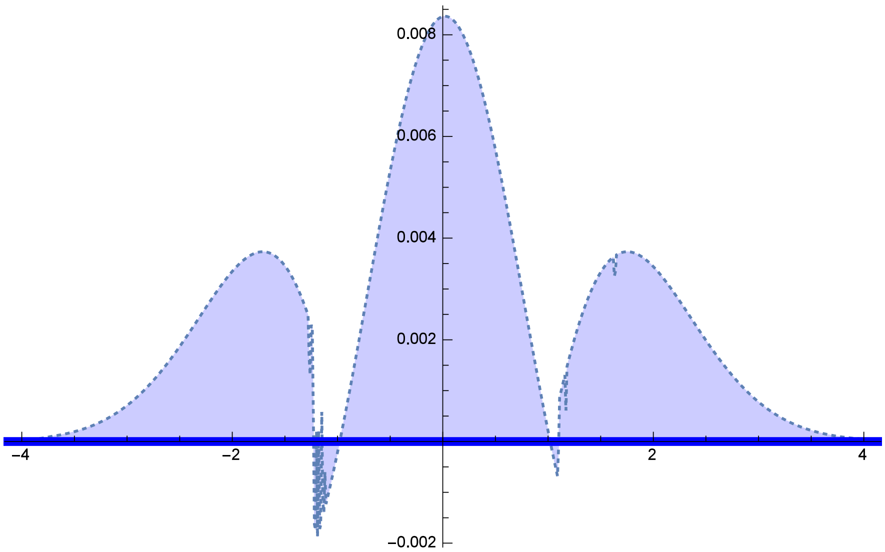


Figure 6.2: Plot of $\{|\hat{F}_{1,n}(z) - \Phi(z)| - |\hat{F}_{1,n}(z) - P(T_1 < z)|\}$ by a dotted line.

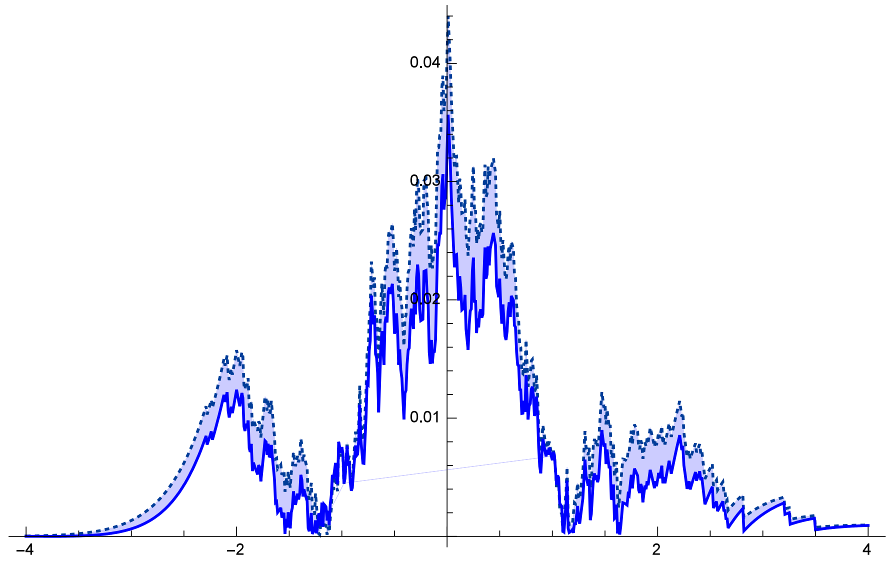


Figure 6.3: Plot of $|\hat{F}_{2,n}(z) - \Phi(z)|$ and $|\hat{F}_{2,n}(z) - P(T_2 < z)|$ by a dotted line and a thick one, respectively.

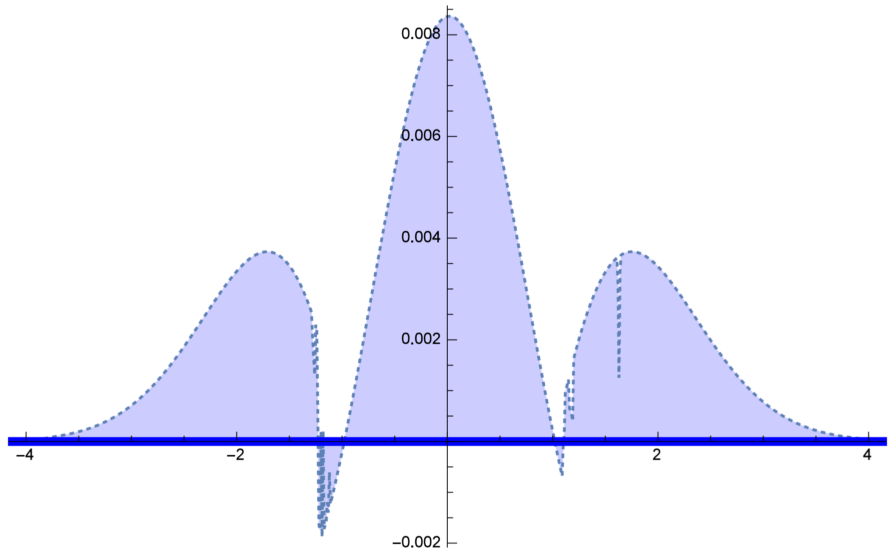


Figure 6.4: Plot of $\{|\hat{F}_{2,n}(z) - \Phi(z)| - |\hat{F}_{2,n}(z) - P(T_2 < z)|\}$ by a dotted line.

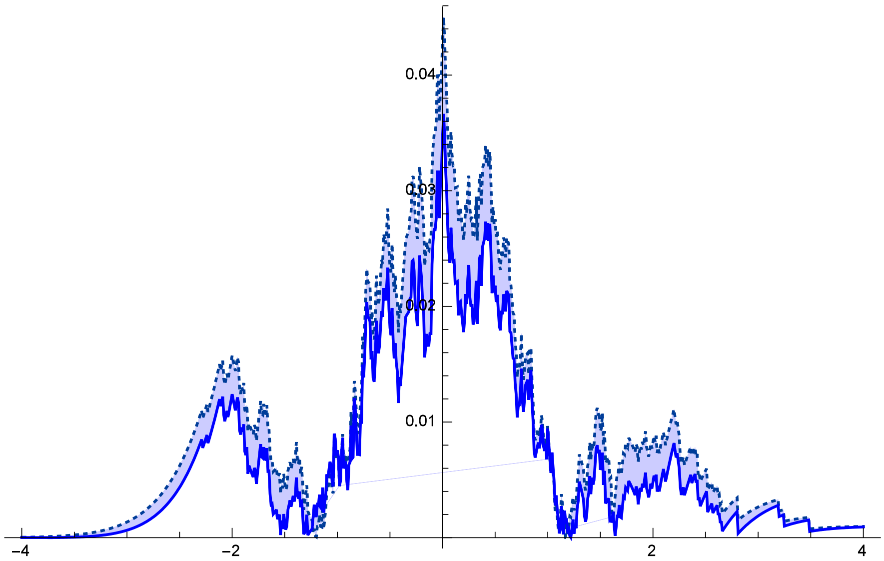


Figure 6.5: Plot of $|\hat{F}_{3,n}(z) - \Phi(z)|$ and $|\hat{F}_{3,n}(z) - P(T_3 < z)|$ by a dotted line and a thick one, respectively.

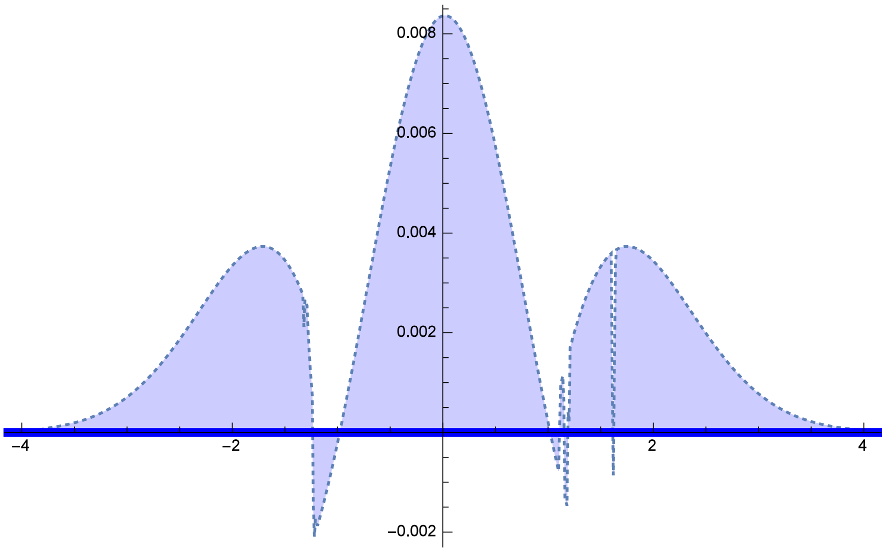


Figure 6.6: Plot of $\{|\hat{F}_{3,n}(z) - \Phi(z)| - |\hat{F}_{3,n}(z) - P(T_3 < z)|\}$ by a dotted line.

Acknowledgment

The author would like to express his deepest appreciation to his supervisor Professor Masanobu Taniguchi of Waseda university for successive guidance, advices and wholehearted encouragement. Professor Taniguchi supervised the author during his master and doctoral courses, and helped the author's life when the author was very uneasy about his career as a researcher.

The author acknowledges Professor Yoichi Nishiyama and Yasutaka Shimizu of Waseda University for their constructive feedback and critical reading of earlier drafts of this dissertation.

The author wishes to express his sincere appreciation to my seniors Research Associate Fumiya Akashi of Waseda University and Research Associate Yan Liu of Kyoto University for many advices.

A special thanks to my family. Words can not express how grateful I am to my father for all of supporting and encouraging. I would also like to thank to my wife. Thank you for supporting me for everything, and especially I can not thank you enough for encouraging me throughout this experience.

Bibliography

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4):589–609.

Altman, E. I. and Brenner, M. (1981). Information effects and stock market response to signs of firm deterioration. *Journal of Financial and Quantitative Analysis*, 16(01):35–51.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd Edition*. John Wiley & Sons, New York.

Aoshima, M. and Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Annals of the Institute of Statistical Mathematics*, 66(5):983–1010.

Bai, Z., Rao, C. R., and Yin, Y. (1990). Least absolute deviations analysis of variance. *Sankhyā: The Indian Journal of Statistics, Series A*, 52:166–177.

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal approximation and asymptotic expansions*, volume 64. SIAM.

Bishop, D. J. (1939). On a comprehensive test for the homogeneity of variances and covariances in multivariate problems. *Biometrika*, 31:31–55.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36:317–346.

Brillinger, D. R. (1973). The analysis of time series collected in an experimental design. *Multivariate Analysis*, III:241–256.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*, volume 36. SIAM, San Francisco.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media, New York.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179.

Chan, Y.-B. and Hall, P. (2009). Scale adjustments for classifiers in high-dimensional, low sample size settings. *Biometrika*, 96(2):469–478.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29(4):995–1010.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.*, 20(3):339–350.

Fujikoshi, Y., Ulyanov, V. V., and Shimizu, R. (2011). *Multivariate statistics: High-Dimensional and Large-Sample Approximations*, volume 760. John Wiley & Sons, Hoboken, New Jersey.

Gersch, W., Martinelli, F., Yonemoto, J., Low, M., and Mc Ewan, J. (1979). Automatic classification of electroencephalograms: Kullback-Leibler nearest neighbor rules. *Science*, 205(4402):193–195.

Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary arch models: dependence structure and central limit theorem. *Econometric Theory*, 16(1):3–22.

Hall, P., Marron, J., and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):427–444.

Hannan, E. J. (1970). *Multiple Time Series*, volume 38. John Wiley & Sons, New York.

Hooke, B. G. (1926). A third study of the english skull with special reference to the farringdon street crania. *Biometrika*, 18:1–55.

Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

Liu, Y., Nagahata, H., Uchiyama, H., and Taniguchi, M. (2017). Discriminant and cluster analysis of possibly high-dimensional time series data by a class of disparities. *Communications in Statistics-Simulation and Computation*, 46(10):455–468.

Liu, Z. and Rao, C. R. (1995). Asymptotic distribution of statistics based on quadratic entropy and bootstrapping. *J. Statist. Plann. Inference*, 43:1–18.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.

Nagahata, H. and Taniguchi, M. (2018). Analysis of variance for high-dimensional time series. *Statistical Inference for Stochastic Processes*, 21(2):455–468.

Nisio, M. (1960). On polynomial approximation for strictly stationary processes. *Journal of the mathematical Society of Japan*, 12(2):207–226.

Rao, C. R. (2009). *Linear Statistical Inference and Its Applications*, volume 22. John Wiley & Sons.

Rao, C. R. (2010). Quadratic entropy and analysis of diversity. *Sankhya A*, 72:70–80.

Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of d-and a-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices. *Journal of multivariate analysis*, 46(1):154–174.

Shumway, R. (1982). Discriminant analysis for time series. *Handbook of statistics*, 2:1–46.

Shumway, R. H. (1971). On detecting a signal in n stationarily correlated noise series. *Technometrics*, 13:499–519.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer.

Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika*, 30:16–28.

Yata, K. and Aoshima, M. (2009). PCA consistency for non-gaussian data in high dimension, low sample size context. *Communications in Statistics–Theory and Methods*, 38(16-17):2634–2652.

Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of multivariate analysis*, 105(1):193–215.

List of main papers

1. Nagahata, H. (-). Higher order approximation of the distribution of test statistics for high-dimensional time series ANOVA models. *To appear in Scientiae Mathematicae Japonicae.*

2. Nagahata, H. and Taniguchi, M. (2018). Analysis of variance for high-dimensional time series. *Statistical Inference for Stochastic Processes*, 21(2):455–468.

3. Nagahata, H. and Taniguchi, M. (2018). Analysis of variance for multivariate time series. *METRON*, 76(1):69–82.

4. Liu, Y., Nagahata, H., Uchiyama, H., and Taniguchi, M. (2017). Discriminant and cluster analysis of possibly high-dimensional time series data by a class of disparities. *Communications in Statistics-Simulation and Computation*, 46(10):8014-8027.

List of papers

1. Nagahata, H. (2017). Classification for high-dimensional finacncial time series by a class of disparities. *Advances in Science Technology and Environmentology Special Issue, 2017*, B14:47-55.

2. Nagahata, H., Suzuki, T., Usami, Y., Yokoyama, A., Ito, J., Hasegawa, F., Taniguchi, M. (2012). Faces for financial time series data. *Advances in Science Technology and Environmentology Special Issue, 2012*, B8:1-13.