ORIGINAL PAPER

DISCRIMINANT AND CLUSTER ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES DATA BY A CLASS OF DISPARITIES

Yan Liu[1], Hideaki Nagahata[1], Hirotaka Uchiyama, and Masanobu Taniguchi[2]

[1] Department of Applied Mathematics
Waseda University
Okubo, Shinjuku-ku, 169-8555

[2]Research Institute for Science and Engineering
Waseda University
Okubo, Shinjuku-ku, 169-8555

Email: hideaki.nagahata@suou.waseda.jp

Key Words: discriminant analysis; cluster analysis; high-dimensional data; time series data; disparity measure; jackknife type adjustment.

ABSTRACT

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.

1 INTRODUCTION

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 paper sheds light on the issue of discriminant analyses and cluster analyses of high dimensional dependent data. There have been a lot of fundamental results for discriminant and cluster analysis of independent and identically distributed 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 discriminant statistic and the quadratic discriminant statistic for classification of categories which have different covariance matrix. Recently many results were reported in the case of independent and identically distributed 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 "HDLSS" (high-dimensional, low-sample-size), $p/n \to \infty$ ($p \to \infty,\ n$ is fixed). Dempster (1958) defined the classical multivariate two sample significance test based on Hotelling's T^2 in high-dimension analysis. Saranadasa (1993) considered the same scale categories, common covariance matrix of a population and substituted the identity matrix for an inverse matrix of a sample covariance matrix. Bai and Saranadasa (1996) derived the asymptotic power of the classical Hotelling's T^2 test and Dempster's nonexact test for a two-sample problem and developed an asymptotically normally distributed test statistic. Hall et al. (2005) suggested a geometric representation of HDLSS data by using a non-standard type of asymptotics. A scale adjusted-type distance-based classifier for high-dimensional data was proposed by Chan and Hall (2009). Yata and Aoshima (2009, 2012) developed a pioneer method for principal component analysis for high-dimensional small sample setting, especially non-sparse settings that mean vectors and covariance matrices $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| \to \infty$ or $\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\| \to \infty$ and some eigenvalues $\lambda_i \to \infty$ as $p \to \infty$. A distance based discriminant classifier for high-dimensional low sample size was proposed in Aoshima and Yata (2014).

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 univariate time series from the point of view of frequency domain approach. Kakizawa et al. (1998) extended the Whittle likelihood to minimum discrimination information for the classification of multivariate 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 spectrum of linear processes. In this paper, we consider the following problem:

for 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 $\boldsymbol{\mu}$ is the expectation of and $\boldsymbol{f}(\lambda)$ is the spectral density matrix of the process $\{\boldsymbol{X}(t)\}$. We consider a distance-based classifier proposed by Saranadasa (1993) and a scale adjusted-type distance-based classifier developed by Chan and Hall (2009). Without the information of spectral densities $\boldsymbol{f}^{(1)}(\lambda)$ and $\boldsymbol{f}^{(2)}(\lambda)$ for π_1 and π_2, we propose a discriminant analysis by a new disparity with jackknife type adjustment and show that jackknife type adjustment reduces bias of a discriminant statistic. For dependent data, the jackknife type adjustment was studied by Carlstein (1986) and Künsch (1989). We elucidate the new asymptotics for high-dimensional time series where the dimension p is allowed to diverge. We also propose a cluster analysis method by the new disparity with jackknife type adjustment for real financial data. It indicates that we can classify companies on the Tokyo Stock Exchange by applying our method to the high-dimensional real data. The notations and symbols are listed in the following: E and Var denote the expectation and the variance with respect to a triplet of $(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$, where the data $\boldsymbol{X} = \{\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)\}$ is classified into π_1 or π_2 and independent of the training sample $\boldsymbol{X}^{(i)} = \{\boldsymbol{X}^{(i)}(1), \ldots, \boldsymbol{X}^{(i)}(n)\}$ from π_i for $i = 1, 2$; $\delta(i, j)$ denotes the Kronecker's delta; $c_{a_1,\ldots,a_k}(t_1, \ldots, t_k)$ denotes the joint cumulant $\mathrm{cum}\{X_{a_1}(t_1), \ldots, X_{a_k}(t_k)\}$; I_d denotes the d-dimensional identity matrix; $\to_p$ denotes the convergence in probability.

2 SETTING

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

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

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

We are interested in the case of $\|\boldsymbol{\mu}\|^2 \equiv \boldsymbol{\mu}'\boldsymbol{\mu} = O(p)$ and $\boldsymbol{f}^{(1)}(\lambda) \neq \boldsymbol{f}^{(2)}(\lambda)$. Also we have independent training samples $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ from π_1 and π_2 with size n_1 and n_2, respectively. For $\boldsymbol{X}$ and $\boldsymbol{X}^{(i)}$ $(i = 1, 2)$, 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 high-dimensional time series data, we use the following discriminant statistic:

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

For simplicity, denote $\boldsymbol{\Gamma} \equiv \boldsymbol{\Gamma}(\boldsymbol{X}, \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$. The classification rule is to classify $\boldsymbol{X}$ into π_1 if $\boldsymbol{\Gamma} < 0$ and into π_2 otherwise. To discuss the asymptotic property of $\boldsymbol{\Gamma}$, we impose the following assumptions.

Assumption 2.1. (i) If $\boldsymbol{X}$ is a Gaussian stationary process, suppose two categories π_1 and π_2 satisfy

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

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

(ii) If $\boldsymbol{X}$ is not Gaussian, then we further assume that all of the third and fourth order cumulants of the process $\{\boldsymbol{X}^{(l)}(t)\}$ satisfy

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

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

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

$$c_{a_1,a_2,a_3}(t_1, t_2) = \mathrm{cum}\{X_{a_1}(t), X_{a_2}(t + t_1), X_{a_3}(t + t_2)\},$$

$$c_{a_1,a_2,a_3,a_4}(t_1, t_2, t_3) = \mathrm{cum}\{X_{a_1}(t), X_{a_2}(t + t_1), X_{a_3}(t + t_2), X_{a_4}(t + t_3)\}.$$

Assumption 2.2. (i) Whenever both n_1 and n_2 diverge, for $\Delta \equiv \|\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}\|^2$, suppose

$$c_1 p < \Delta < c_2 p, \quad \text{for some constants } c_1,\, c_2 > 0.$$

(ii) Both n_1 and n_2 are large enough such that $p = o(n_1)$ and $p = o(n_2)$.

Assumption 2.2 (i) is satisfied if the absolute values of elements of $\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ do not increase as the dimension p increases, for example, are of order $O(1)$. To investigate the asymptotic behavior of misclassification probability of $\boldsymbol{\Gamma}$, first, we evaluate the expectation and variance. Proofs of lemmas and theorems are placed in Appendix. Assumption 2.2 (ii) is explained after Theorem 2.1.

Lemma 2.1. Assume Assumptions 2.1 (i) and 2.2 (i). When $\boldsymbol{X}$ belongs to π_i,

$$E(\boldsymbol{\Gamma}) = \frac{(-1)^i}{2}\Delta + O(\min(n_1, n_2)^{-1}p).$$

Lemma 2.2. Assume Assumptions 2.1 and 2.2 (i). When $\boldsymbol{X}$ belongs to π_i,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\Gamma}) &= \sum_{k=1}^{2} \frac{1}{nn_k}\mathrm{tr}\Big(\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)R^{(i)}(u)\sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\Big)R^{(k)}(u)\Big) \\
&\quad + \frac{1}{2}\sum_{l=1}^{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)^2\Big\} + \Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)R^{(i)}(u) \\
&\qquad + \frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)R^{(j)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big) + O(n_1^{-2}p^2) + O(n_2^{-2}p^2) \quad (2)
\end{aligned}$$

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

Remark 2.1. Lemma 2.1 includes i.i.d. cases considered in Aoshima and Yata (2014). In fact, if we set

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

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

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

Suppose $P(i|j)$ is misclassification rate by the classification statistic (1) such that $\boldsymbol{X}$ belonging to π_j is erroneously assigned to π_i $(i \neq j)$. We obtained the following theorem.

Theorem 2.1. Under Assumptions 2.1 - 2.2, it holds that

(i) if p is finite or $p \to \infty$, and $n \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$, then

$$P(i|j) \to 0, \qquad (i, j) = (1, 2) \text{ or } (2, 1).$$

(ii) if $p \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$ with n finite, then

$$P(i|j) \to 0, \qquad (i,j) = (1,2) \text{ or } (2,1),$$

under assumption that

$$\left(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\right)' \boldsymbol{f}^{(i)}(0) \left(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\right) / \Delta^2 \to 0.$$

Remark 2.2. (i) Since the proof of Theorem 2.1 heavily depends on the positivity of $E(\boldsymbol{\Gamma})$, it is necessary to assume Assumption 2.2 to show the consistency of the statistic $\boldsymbol{\Gamma}$.

(ii) The case (i) of finite dimension p in Theorem 2.1 is not for high-dimension, but we give a comprehensive result for all cases of dimension p.

Conditions which $\boldsymbol{\Gamma}$ becomes a consistent discriminant statistic under, i.e., $P(i|j) \to 0$, are very restrictive because of Assumption 2.2 (ii). Let us write the Lemma 2.1 with the exact expression, we get

$$E(\boldsymbol{\Gamma}) = \frac{(-1)^i}{2}\Delta + \frac{1}{2n_1}\sum_{u=1-n_1}^{n_1-1}\left(1 - \frac{|u|}{n_1}\right)\mathrm{tr}R^{(1)}(u) - \frac{1}{2n_2}\sum_{u=1-n_2}^{n_2-1}\left(1 - \frac{|u|}{n_2}\right)\mathrm{tr}R^{(2)}(u). \tag{3}$$

Since for both $j = 1, 2$,

$$\frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\left(1 - \frac{|u|}{n_j}\right)\mathrm{tr}R^{(j)}(u)$$

is positive although the classification rule is to classify $\boldsymbol{X}$ into π_1 if $\boldsymbol{\Gamma} < 0$ and into π_2 otherwise, the classification rule is highly biased by the second and third terms in (3).

One possibility to reduce the bias in $\boldsymbol{\Gamma}$ is to modify (1) by

$$\begin{aligned}\boldsymbol{\Gamma}_{\mathrm{mod}} = &\left(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\right)'\left(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\right) \\ &- \frac{1}{2n_1}\sum_{u=1-n_1}^{n_1-1}\left(1 - \frac{|u|}{n_1}\right)\mathrm{tr}R^{(1)}(u) + \frac{1}{2n_2}\sum_{u=1-n_2}^{n_2-1}\left(1 - \frac{|u|}{n_2}\right)\mathrm{tr}R^{(2)}(u),\end{aligned} \tag{4}$$

if we have the information of $R^{(1)}(u)$ and $R^{(2)}(u)$ beforehand. Suppose $P^*(i|j)$ is misclassification rate by the classification statistic (4) such that $\boldsymbol{X}$ belonging to π_j is erroneously assigned to π_i $(i \neq j)$. We obtained the following theorem.

Theorem 2.2. Under Assumptions 2.1 and 2.2 (i), it holds that

(i) if p is finite or $p \to \infty$, and $n \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$, then

$$P^*(i|j) \to 0, \qquad (i,j) = (1,2) \text{ or } (2,1).$$

(ii) if $p \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$ with n finite,, then

$$P^*(i|j) \to 0, \qquad (i,j) = (1,2) \text{ or } (2,1),$$

under assumption that

$$\left(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\right)' \boldsymbol{f}^{(i)}(0)\left(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\right)/\Delta^2 \to 0.$$

(iii) if $p \to \infty$ and $n \to \infty$ with n_1 and n_2 finite, then

$$P^*(i|j) \to 0, \qquad (i,j) = (1,2) \text{ or } (2,1).$$

under assumptions that the absolute sum of all the third and fourth cumulants of $\boldsymbol{X}$ is $O(1)$ and for $l = 1, 2$,

$$\operatorname{tr}\left\{\left(\boldsymbol{f}^{(l)}(0)\right)^2\right\}/\Delta^2 \to 0,$$

and

$$\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c^{(l)}_{jjk}(u,v)/\Delta^2 \to 0, \qquad \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c^{(l)}_{jkk}(u,v)/\Delta^2 \to 0.$$

(iv) if $p \to \infty$ with n, n_1 and n_2 finite, then

$$P^*(i|j) \to 0, \qquad (i,j) = (1,2) \text{ or } (2,1).$$

under assumptions that for $l = 1, 2$,

$$\operatorname{tr}\left\{\left(\boldsymbol{f}^{(l)}(0)\right)^2\right\}/\Delta^2 \to 0,$$

and

$$\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c^{(l)}_{jjk}(u,v)/\Delta^2 \to 0, \qquad \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c^{(l)}_{jkk}(u,v)/\Delta^2 \to 0.$$

Remark 2.3. From Theorem 2.2, we can see that Assumption 2.2 (ii) and the information of $R^{(1)}(u)$ and $R^{(2)}(u)$ are in the relationship of trade-off. Knowing $R^{(1)}(u)$ and $R^{(2)}(u)$, the classification statistic under the classification rule can be applied to much richer situations.

Next, we provide the situation to reduce the bias coming from the terms $R^{(1)}(u)$ and $R^{(2)}(u)$. Let us consider the jackknife type adjustment to the auto-covariance of the training samples $\boldsymbol{X}^{(k)}$, i.e.,

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

The jackknife estimators $\hat{\Sigma}^{(k)}_{\text{Jack}}$ are given by

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

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

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

where l_k is the length of the downweighted block (see Künsch (1989)). Here the weight function are defined by

$$\alpha_{n_k}(t) = (\|\omega_{n_k}\|_1)^{-1}(n_k - l_k + 1)^{-1}\sum_{j=0}^{n_k-l_k} \omega_{n_k}(t-j),$$

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

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

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.

Here $\|\omega_{n_k}\|_1 = \sum_{i=1}^{l_k} \omega_{n_k}(i)$.

To consider the asymptotic properties of the jackknife type adjusted discriminant statistics, we follow Theorem 3.2 of Künsch (1989). We can relax Assumption 2.2 (ii) to the following assumption.

Assumption 2.2. (ii)' Both n_1 and n_2 are large enough such that $p = O(n_1)$ and $p = O(n_2)$.

Corollary 2.3. Under Assumptions 2.1 and 2.2 (i) and (ii)', we have the following results. Let $l_k = l_k(n_k) \to \infty$ when $n_k \to \infty$ for $k = 1, 2$. When $\boldsymbol{X}$ belongs to π_i, it holds that for $k = 1, 2$,

(i) if $h(x) = \mathbf{1}_{(0,1)}(x)$, $l_k = o(n_k^{1/2})$, then

$$E(\mathbf{\Gamma}_{\text{Jack}}) = \frac{(-1)^i}{2}\Delta + O(l_k^{-1}\, n_k^{-1}\, p).$$

(ii) if $h{*}h$ is twice continuously differentiable around zero, $l_k = o(n_k^{1/3})$, and if further $\sum u^2 |R^{(k)}(u)| < \infty$, then

$$E(\mathbf{\Gamma}_{\text{Jack}}) = \frac{(-1)^i}{2}\Delta + O(l_k^{-2}\, n_k^{-1}\, p).$$

(iii) if $E|X(t)|^{6+\delta} < \infty$ and $l_k = o(n_k)$ for $k = 1, 2$, then

$$\text{Var}(\mathbf{\Gamma}_{\text{Jack}}) \sim \text{Var}(\mathbf{\Gamma}).$$

3 SIMULATION STUDIES

To investigate performances of the statistics considered in Section 2, we compared the misclassification probability of the following five discriminant statistics $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, $\mathbf{\Gamma}_3$, $\mathbf{\Gamma}_4$, and $\mathbf{\Gamma}_5$ below. For independent and identically distributed (i.i.d.) case, Chan and Hall (2009) proposed the statistics $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$. $\mathbf{\Gamma}_1$ is a discriminant statistic that has no bias-correction. $\mathbf{\Gamma}_2$ is a statistic that has bias-correction for i.i.d. random variables. On the other hand, we considered the statistics $\mathbf{\Gamma}_3$, $\mathbf{\Gamma}_4$, and $\mathbf{\Gamma}_5$ for dependent data. $\mathbf{\Gamma}_3$ is a statistic that remove the exact bias, of $\mathbf{\Gamma}_1$ constructed from the stochastic process, from $E(\mathbf{\Gamma})$. Estimating sequences of $\{R^{(1)}(u)\}$ and $\{R^{(2)}(u)\}$, however, is a big problem. As a robust method against the property of independence and dependence, we proposed statistics $\mathbf{\Gamma}_4$ and $\mathbf{\Gamma}_5$. Statistics $\mathbf{\Gamma}_4$ and $\mathbf{\Gamma}_5$ are based on $\mathbf{\Gamma}_1$ with jackknife type adjustment for the bias. The kernel functions in $\mathbf{\Gamma}_4$ and $\mathbf{\Gamma}_5$ are $h_1(x) = 1_{(0,1)}(x)$ and $h_2(x) = 1/2\{1 - \cos(\pi x)\}$, respectively. The lengths of the downweighted blocks are $l_k = 5$. $h_1(x)$ is designed for the kernel satisfying Corollary 2.3 (i) while $h_2(x)$ for that satisfying Corollary 2.3 (ii).

$$\mathbf{\Gamma}_1 = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big); \tag{5}$$

$$\mathbf{\Gamma}_2 = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big) - \frac{1}{2n_1}\operatorname{tr}\boldsymbol{S}^{(1)} + \frac{1}{2n_2}\operatorname{tr}\boldsymbol{S}^{(2)}; \tag{6}$$

$$\begin{aligned}\mathbf{\Gamma}_3 = {} & \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big) \\ & - \frac{1}{2n_1}\sum_{u=1-n_1}^{n_1-1}\Big(1 - \frac{|u|}{n_1}\Big)\operatorname{tr}R^{(1)}(u) + \frac{1}{2n_2}\sum_{u=1-n_2}^{n_2-1}\Big(1 - \frac{|u|}{n_2}\Big)\operatorname{tr}R^{(2)}(u);\end{aligned} \tag{7}$$

$$\mathbf{\Gamma}_4 = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big) - \operatorname{tr}\hat{\Sigma}^{(1)}_{\text{Jack1}} + \operatorname{tr}\hat{\Sigma}^{(2)}_{\text{Jack1}}, \tag{8}$$

$$h_1(x) = 1_{(0,1)}(x);$$

$$\mathbf{\Gamma}_5 = \Big(\bar{\boldsymbol{X}} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big) - \operatorname{tr}\hat{\Sigma}^{(1)}_{\text{Jack2}} + \operatorname{tr}\hat{\Sigma}^{(2)}_{\text{Jack2}}, \tag{9}$$

$$h_2(x) = \frac{1}{2}\{1 - \cos(\pi x)\}.$$

We used moving average (MA) model, autoregressive (AR) model and autoregressive-moving average (ARMA) model for our simulation study. Table 1 provides the further detail settings of the models.

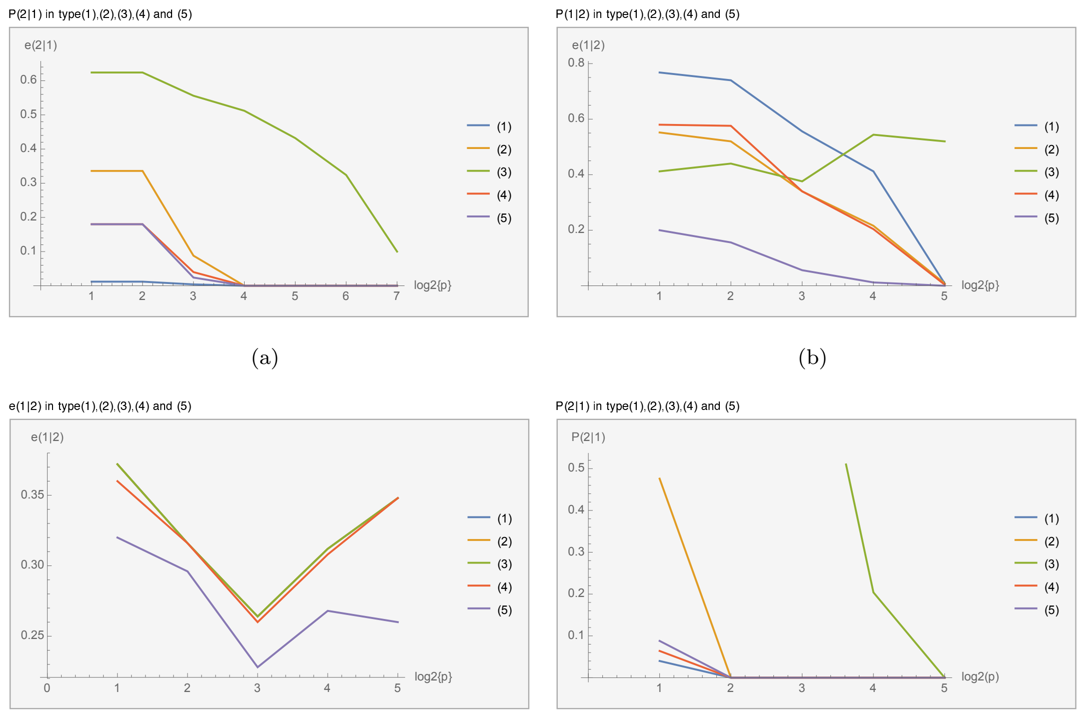
model under each category π_i		mean vector
MA(2)	$\boldsymbol{X}^{(i)}(t) = \boldsymbol{\mu}^{(i)} + \sum_{j=0}^{2}\Theta_j\boldsymbol{\epsilon}_{t-j}$	$\boldsymbol{\mu}^{(1)} = (0, 0, ..., 0)$,
AR(1)	$\boldsymbol{X}^{(i)}(t) - \Xi_1\boldsymbol{X}^{(i)}(t-1) = \boldsymbol{\mu}^{(i)} + \boldsymbol{\epsilon}(t)$	$\boldsymbol{\mu}^{(2)} = (\underbrace{1, 1, ...1}_{\lfloor p^{2/3}\rfloor}, \underbrace{0, 0, ..., 0}_{p-\lfloor p^{2/3}\rfloor})$
ARMA(1, 1)	$\boldsymbol{X}^{(i)}(t) - \Xi_1\boldsymbol{X}^{(i)}(t-1) = \boldsymbol{\mu}^{(i)} + \sum_{j=0}^{1}\Theta_j\boldsymbol{\epsilon}(t-j)$	

Table 1: Detailed setting for each category. The mean vector of category π_1 is vector $\mathbf{0}$, while the mean vector of category π_2 is that whose first $\lfloor p^{2/3}\rfloor$ elements are 1 and 0 else. For simplicity, we used Gaussian MA(2), AR(1) and ARMA(1,1) process in our simulations.

case number	categories		scale c_i	
	π_1	π_2	π_1	π_2
(a)	AR(1)	AR(1)	1	100
(b)	AR(1)	MA(2)	1	30
(c)	MA(2)	MA(2)	1	10
(d)	AR(1)	ARMA(1, 1)	0.8	1.2

Table 2: Simulation settings for each case. Covariance matrix of innovation process $\{\boldsymbol{\epsilon}_t\}$ in category (i) is as follows: $(\Sigma^{(i)})_{jk} = c_i\{0.1^{(|j-k|^{(1/3)})}\}$. Suppose Φ is a diagonal matrix $((p-1)/p)\, I_p$. $\Theta_0 = I_p$, $\Theta_1 = \Xi_1 = \Phi$ and $\Theta_2 = \Phi^2$.

We carried out 4 numerical experiments (a), (b), (c) and (d) for comparison, whose details are given in Table 2. The numerical results of (a), (b), (c) and (d) are shown in Figure 1 below. These simulation results support our theory strongly. The jackknife type adjusted discriminant statistics $\mathbf{\Gamma}_4$ and $\mathbf{\Gamma}_5$ are clearly admissible and performed better than the other statistics in these simulations. Other simulation results are given in Supplementary Material after the main text.



(a) (b)

(c) (d)

Figure 1: The misclassification rates in four simulations (a), (b), (c) and (d).

4 REAL DATA ANALYSIS

In our cluster analysis, we applied the jackknife type adjusted disparity $\mathbf{\Gamma}_4$ to the real 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 and 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 Table 3. The purpose of our simulation is to make a dendrogram of these companies. A dendrogram gives a visual representation of the hierarchical cluster. To classify the real data, we applied the following disparity to the real data:

$$C(\boldsymbol{X}_1, \boldsymbol{X}_2) = (\boldsymbol{X}_1 - \boldsymbol{X}_2)'(\boldsymbol{X}_1 - \boldsymbol{X}_2) + \left|\mathrm{tr}\hat{\Sigma}^{(1)}_{\mathrm{Jack}} - \mathrm{tr}\hat{\Sigma}^{(2)}_{\mathrm{Jack}}\right|, \tag{10}$$

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 3: Companies of the first and second sections.

where $\boldsymbol{X}_i = \bar{\boldsymbol{X}}^{(i)}$, and $\hat{\Sigma}^{(k)}_{\text{Jack}}$ is defined by $\hat{\Sigma}^{(k)}_{\text{Jack}}(u) = {n_k}^{-1} \sum_{u=1-l_k}^{l_k-1} \nu_{n_k}(u)/\nu_{n_k}(0)\hat{R}^{(k)}(u)$ with kernel function $h_1(x) = 1_{(0,1)}(x)$. The lengths of the downweighted blocks are set as $l_k = 2$. The cluster function (10) satisfies following conditions.

(i) $C(\boldsymbol{X}, \boldsymbol{X}) = 0$,

(ii) $C(\boldsymbol{X}, \boldsymbol{Y}) = C(\boldsymbol{Y}, \boldsymbol{X})$.

A generalized distance function $C(\boldsymbol{X}_1, \boldsymbol{X}_2)$ has been considered in De Leon and Carriere (2005). We obtained the results shown in Figure 8, Figure 9, and Figure 10. As can be seen in Figure 8, the companies in the second section companies formed an exact crowd (red characters in Figure 8). That means that our cluster method by the disparity (10) can classify the first section and the second section very well. In Figure 9, 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 10) is apart from the other companies. Mitani corporation is the most big company in the second section.

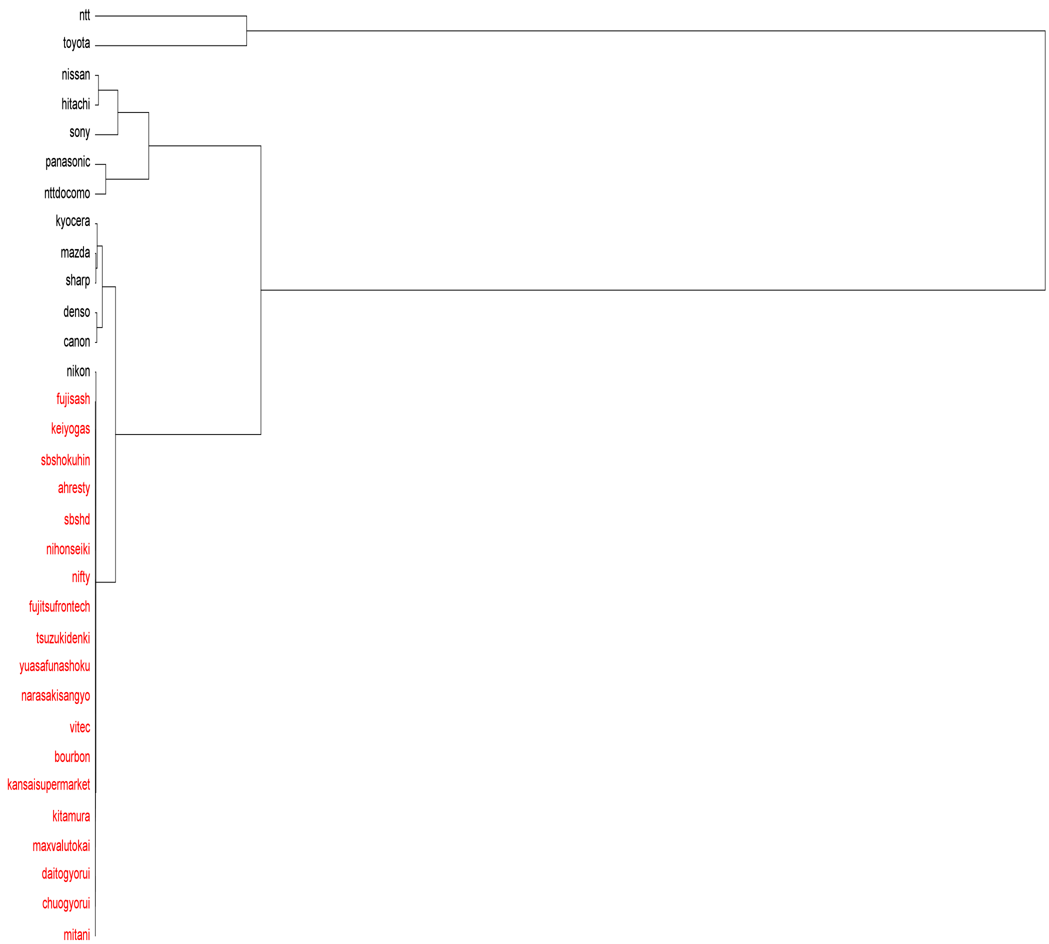


Figure 2: The cluster analysis of the first and second sections. The first section companies are in the black. The second section companies are in the red.

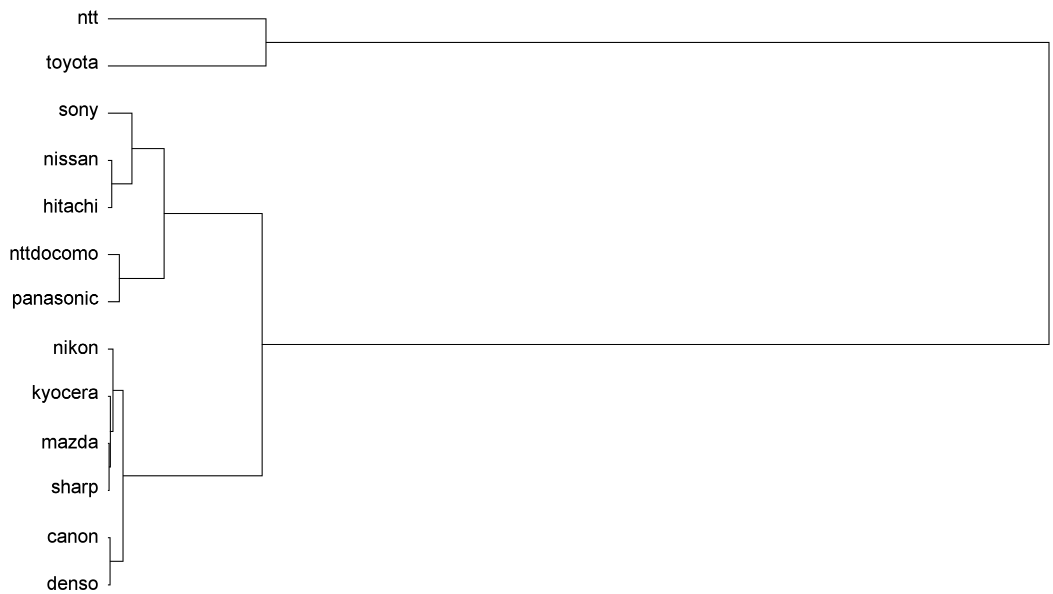


Figure 3: The cluster analysis of companies in the first section.

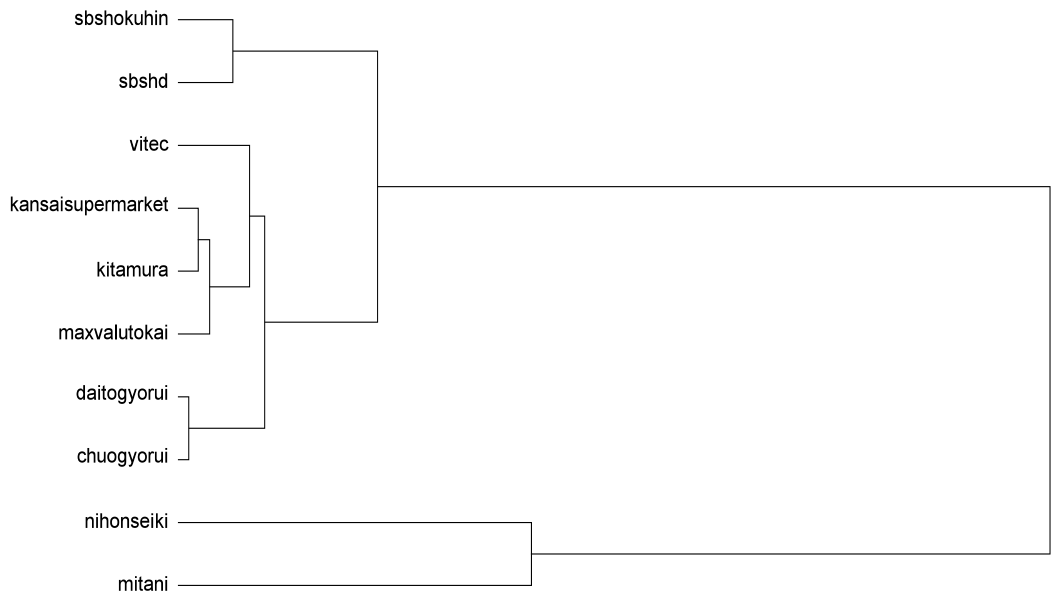


Figure 4: The cluster analysis of companies in the second section.

ACKNOWLEDGEMENTS

This research was supported by Grant-in-Aid for JSPS Fellows Grant Number 26·7404, Grant-in-Aid for Scientific Research (A) (15H02061) & Challenging Exploratory Research (26540015). We are also grateful to Fumiya Akashi for his helpful comments on earlier versions of this paper.

APPENDIX

In Appendix, we provide proofs of Lemmas 2.1 and 2.2 and Theorems 2.1–**??** in the paper. The preparation Lemma A for all proofs is given in the end of Appendix.

Proof of Lemma 2.1.

Proof. From Lemma A (ii) and (x), when $\boldsymbol{X}$ belongs to π_i, we see that

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

$$
\begin{aligned}
= \quad & \frac{(-1)^i}{2}\Delta - \frac{1}{2n_2}\sum_{u=1-n_2}^{n_2-1}\Big(1-\frac{|u|}{n_2}\Big)\mathrm{tr}R^{(2)}(u) \\
& + \frac{1}{2n_1}\sum_{u=1-n_1}^{n_1-1}\Big(1-\frac{|u|}{n_1}\Big)\mathrm{tr}R^{(1)}(u).
\end{aligned}
$$

For $i = 1, 2$, it holds from Assumption 2.1 that

$$
\frac{1}{2n_i}\sum_{u=1-n_i}^{n_i-1}\Big(1-\frac{|u|}{n_i}\Big)\mathrm{tr}R^{(i)}(u) \le \frac{1}{2n_i}\sum_{u=1-n_i}^{n_i-1}\mathrm{tr}R^{(i)}(u) \le \frac{p}{2n_i}\max_{j=1,\ldots,p}\sum_{u=-\infty}^{\infty}\mathrm{tr}|R^{(i)}_{jj}(u)|.
$$

By Assumption 2.2 (i), we obtain the desired result. $\square$

Proof of Lemma 2.2.

Proof. From the law of total variance, when $\boldsymbol{X}$ belongs to π_i,

$$
\mathrm{Var}(\boldsymbol{\Gamma}) = E\,\mathrm{Var}(\boldsymbol{\Gamma}|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) + \mathrm{Var}(E\,(\boldsymbol{\Gamma}|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})). \tag{11}
$$

The first term at the right hand side in (11) can be evaluated by

$$
E\,\mathrm{Var}(\boldsymbol{\Gamma}|\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) = E\,\mathrm{Var}\{\bar{\boldsymbol{X}}'(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})\}
$$

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

Last equation follows Lemma A (iv), (viii) and (ix). From Assumption 2.1,

$$
\begin{aligned}
\frac{1}{n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l} c^{(l)}_{jjk}(t_1-s_1,s_2-s_1) \quad &= \quad O(n_l^{-2}p^2), \\
\frac{1}{n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l} c^{(l)}_{jjkk}(t_1-s_1,s_2-s_1,t_2-s_1) \quad &= \quad O(n_l^{-3}p^2),
\end{aligned}
$$

and

$$
\frac{1}{n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l} c^{(l)}_{jkk}(t_1-s,t_2-s) = O(n_l^{-2}p^2)
$$

for $l = 1, 2$. All the terms related to the third order cumulants have the similar evaluation. The second term in (11) can be evaluated by

$$
\begin{aligned}
& \operatorname{Var}(E(\boldsymbol{\Gamma}|\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})) \\
= \ & \operatorname{Var}\Big\{\Big(\boldsymbol{\mu}^{(i)} - \frac{\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)}}{2}\Big)'\Big(\bar{\boldsymbol{X}}^{(2)} - \bar{\boldsymbol{X}}^{(1)}\Big)\Big\} \\
= \ & \boldsymbol{\mu}^{(i)'}\Big\{\sum_{l=1}^{2}\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big\}\boldsymbol{\mu}^{(i)} \\
& + \sum_{l=1}^{2}\Big[\frac{1}{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)^2\Big\} + \boldsymbol{\mu}^{(l)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)}\Big\}\Big] \\
& - \sum_{l=1}^{2}2\boldsymbol{\mu}^{(i)'}\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)\boldsymbol{\mu}^{(l)} + O(n_1^{-2}p^2) + O(n_2^{-2}p^2) \\
= \ & (\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^{(j)})'\Big\{\frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)R^{(j)}(u)\Big\}(\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^{(j)}) \\
& + \sum_{l=1}^{2}\frac{1}{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)^2\Big\} + O(n_1^{-2}p^2) + O(n_2^{-2}p^2),
\end{aligned}
$$

where $j \neq i$. The second equation follows Lemma A (ix), (xi) and (xii). Combining two equations, we have

$$
\begin{aligned}
\operatorname{Var}_i(\boldsymbol{\Gamma}) = & \sum_{k=1}^{2}\frac{1}{nn_k}\mathrm{tr}\Big(\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)R^{(i)}(u)\sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\Big)R^{(k)}(u)\Big) \\
& + \frac{1}{2}\sum_{l=1}^{2}\mathrm{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)^2\Big\} + \Big(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)R^{(i)}(u) \\
& \qquad + \frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)R^{(j)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}\Big) + O(n_1^{-2}p^2) + O(n_2^{-2}p^2).
\end{aligned}
$$

□

Proof of Theorem 2.1.

Proof. Let Δ^* be

$$
\Delta^* = \Delta + (-1)^i\Big\{\frac{1}{n_1}\sum_{u=1-n_1}^{n_1-1}\Big(1-\frac{|u|}{n_1}\Big)\mathrm{tr}R^{(1)}(u) - \frac{1}{n_2}\sum_{u=1-n_2}^{n_2-1}\Big(1-\frac{|u|}{n_2}\Big)\mathrm{tr}R^{(2)}(u)\Big\}.
$$

To show the consistency of the discriminant statistic $\mathbf{\Gamma}$, first, we have to show the positivity of Δ^* under Assumptions 2.1 and 2.2. From Lemma 2.1 and Assumption 2.2 (ii), it can be seen that

$$\Delta^* = \Delta + o_p(1). \tag{12}$$

From Assumption 2.2 (i), Δ^* is positive for sufficiently large n_1 and n_2.

From Lemmas 2.1 and 2.2, and by Chebyshev's inequality, we have, when $\boldsymbol{X}$ belongs to π_i,

$$\frac{\mathbf{\Gamma}}{\Delta^*} = \frac{(-1)^i}{2} + o_p(1),$$

Combining with (12), it holds under both situations (i) and (ii) in theorem that

$$P(i|j) \to_p 0, \quad \text{if } i \neq j.$$

$\square$

Proof of Theorem 2.2.

Proof. From Lemma 2.1, we see that

$$E(\mathbf{\Gamma}_{\text{mod}}) = \frac{(-1)^i}{2}\Delta.$$

From Lemma 2.2, it is easy to see that cases (i) and (ii) are the same as those in Theorem 2.1. For case (iii), it has to hold that

$$\Delta^{-2}\sum_{l=1}^{2}\operatorname{tr}\Big\{\Big(\frac{1}{n_l}\sum_{u=1-n_l}^{n_l-1}\Big(1-\frac{|u|}{n_l}\Big)R^{(l)}(u)\Big)^2\Big\} \to 0, \tag{13}$$

$$\Delta^{-2}\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n_j}\sum_{u=1-n_j}^{n_j-1}\Big(1-\frac{|u|}{n_j}\Big)R^{(j)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big) \to 0, \tag{14}$$

for $l = 1, 2$,

$$\frac{1}{\Delta^2 n_l^4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s_1=1}^{n_l}\sum_{s_2=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l} c_{jjk}^{(l)}(t_1-s_1, s_2-s_1) \to 0 \tag{15}$$

and

$$\frac{1}{\Delta^2 n_l^3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{s=1}^{n_l}\sum_{t_1=1}^{n_l}\sum_{t_2=1}^{n_l} c_{jkk}^{(l)}(t_1-s, t_2-s) \to 0. \tag{16}$$

It is not difficult to see that for $l = 1, 2$,

$$\operatorname{tr}\Big\{\Big(\boldsymbol{f}^{(l)}(0)\Big)^2\Big\}/\Delta^2 \to 0 \tag{17}$$

is a sufficient condition for (13). Furthermore, (17) implies that

$$\begin{aligned}(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{f}^{(j)}(0)(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}) &= \mathrm{tr}\Big\{(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)})(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{f}^{(j)}(0)\Big\} \\ &\leq \Delta\,\mathrm{tr}\Big(\boldsymbol{f}^{(j)}(0)^2\Big)^{1/2} = o(\Delta^2),\end{aligned}$$

which in turn implies (14). Also, assumptions

$$\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c_{jjk}^{(l)}(u,v)/\Delta^2 \to 0, \qquad \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} c_{jkk}^{(l)}(u,v)/\Delta^2 \to 0$$

implies that the third order terms in (15) and (16) go to 0 as $p \to \infty$. Thus, the conclusion holds.

For case (iv), we have to further show

$$\Delta^{-2}\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big)'\Big(\frac{1}{n}\sum_{u=1-n}^{n-1}\Big(1-\frac{|u|}{n}\Big)R^{(i)}(u)\Big)\Big(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)}\Big) \to 0. \tag{18}$$

However, (18) is also implied by (17), which concludes our theorem. □

Proof of Corollary 2.3.

Proof. From Theorem 3.2 in Künsch (1989), as $l_k \to \infty$, we can see that under (i),

$$E\Big(n_k \mathrm{tr}\hat{\Sigma}^{(k)}_{\mathrm{Jack}}\Big) - \sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\mathrm{tr}R^{(k)}(u)\Big) = O(l_k^{-1}p),$$

and under (ii),

$$E\Big(n_k \mathrm{tr}\hat{\Sigma}^{(k)}_{\mathrm{Jack}}\Big) - \sum_{u=1-n_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\mathrm{tr}R^{(k)}(u)\Big) = O(l_k^{-2}p),$$

for $k = 1, 2$. These expressions lead the conclusion. (iii) is a direct result of (3.12) in Künsch (1989).

□

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

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

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

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

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

For $i, j = 1, 2$,

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

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

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

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

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

(x) Also,

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

(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

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

(xii) In addition,

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

In Gaussian case, the equation is equivalent to

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

$$E_i\{\boldsymbol{\mu}^{(i)^{'}}(\bar{\boldsymbol{X}}^{(2)}-\bar{\boldsymbol{X}}^{(1)})\} = \sum_{k_1=1}^{p}\sum_{j=1}^{p}\{\mu_j^{(i)}\}^2\frac{1}{n_k}\sum_{u=1-u_k}^{n_k-1}\Big(1-\frac{|u|}{n_k}\Big)R_{jj}^{(i)}(u).$$

REFERENCES

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. New York: Wiley.

Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential analysis*, 30(4):356–399.

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

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.

De Leon, A. and Carriere, K. (2005). A generalized mahalanobis distance for mixed data. *Journal of Multivariate Analysis*, 92(1):174–185.

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

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.

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.

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.

Rao, C. R. (2009). *Linear statistical inference and its applications*, volume 22. John Wiley & Sons.

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). 1 discriminant analysis for time series. *Handbook of statistics*, 2:1–46.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series. Springer Series in Statistics.* Springer New York.

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. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of multivariate analysis*, 101(9):2060–2077.

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.

Yata, K. and Aoshima, M. (2013). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *Journal of multivariate analysis*, 117:313–331.

Supplementary Material

In Supplementary Material, we present the other simulation results in Sections 3 and 4.

4.1 Other simulation results by the discriminants statistics $\mathbf{\Gamma}_1 - \mathbf{\Gamma}_5$

In this section, we give other simulation results by the discriminants statistics $\mathbf{\Gamma}_1 - \mathbf{\Gamma}_5$. The setting of each case (e), (f) and (g) is given in Table 4.

case number	categories π_1	categories π_2	scale c_i π_1	scale c_i π_2
(e)	AR(1)	AR(1)	0.8	1.2
(f)	MA(2)	AR(1)	10	1
(g)	MA(2)	MA(2)	0.8	1.2

Table 4: Simulation settings for each case. Covariance matrix of innovation process $\{\boldsymbol{\epsilon}_t\}$ in category (i) is as follows: $(\Sigma^{(i)})_{jk} = c_i\{0.1^{(|j-k|^{(1/3)})}\}$. Suppose Φ is a diagonal matrix $((p-1)/p)\, I_p$. $\Theta_0 = I_p$, $\Theta_1 = \Xi_1 = \Phi$ and $\Theta_2 = \Phi^2$.

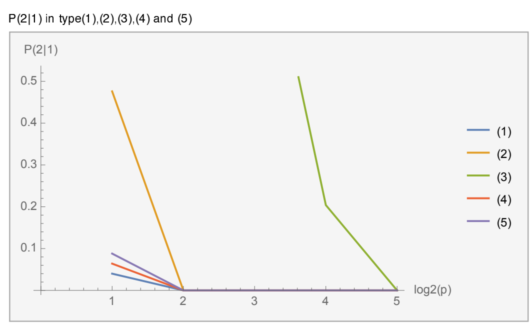


Figure 5: The misclassification rates in simulation (e).

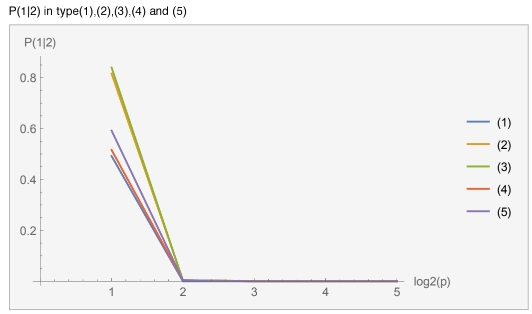


Figure 6: The misclassification rates in simulation (f).

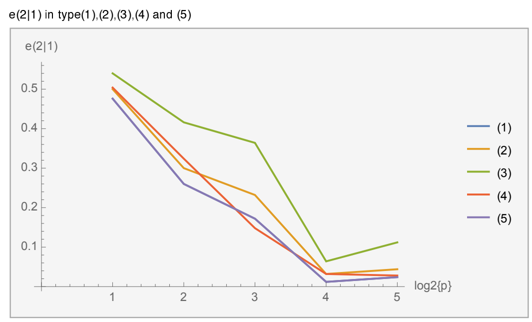


Figure 7: The misclassification rates in simulation (g).

4.2 Real data cluster analysis by Euclidean distance

For comparison, we also provide in Figures 8, 9 and 10 the results of real data cluster analysis by the Euclidean distance.

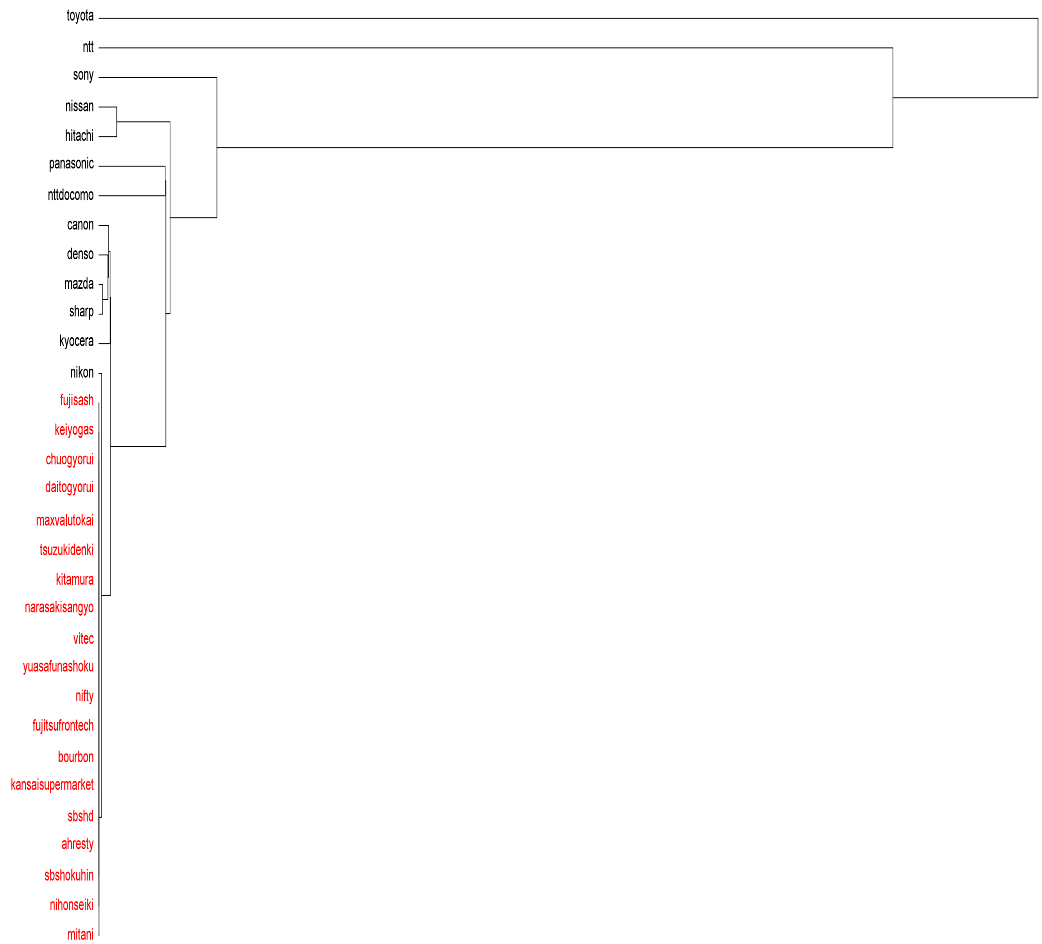


Figure 8: (Euclid) The cluster analysis of the first and second sections. The first section companies are in the black. The second section companies are in the red.

Figure 9: (Euclid) The cluster analysis of companies in the first section.

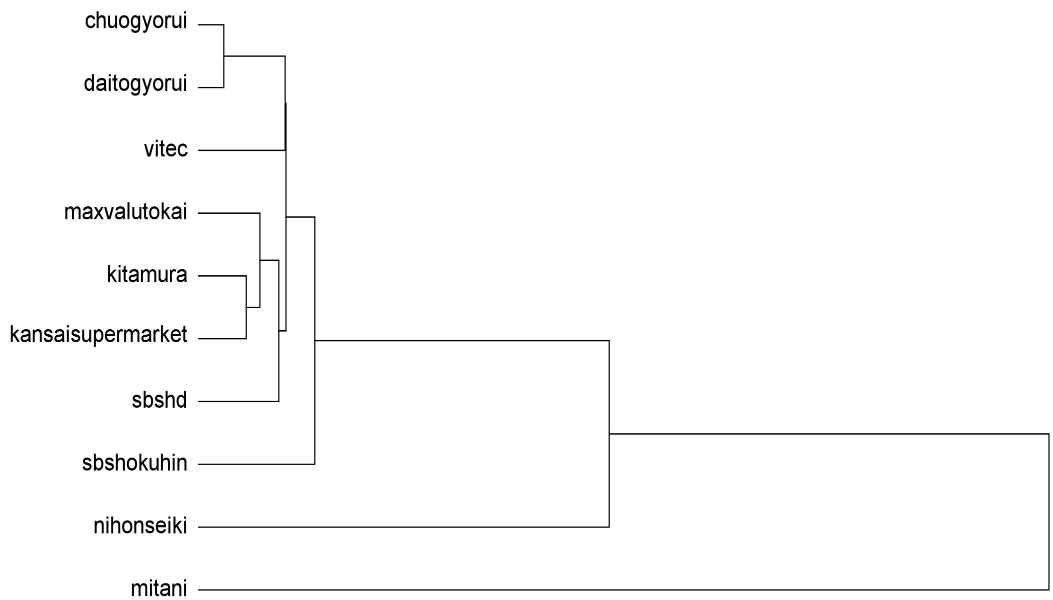


Figure 10: (Euclid) The cluster analysis of companies in the second section.