Measures of multivariate skewness and kurtosis in high-dimensional framework

Takuma SUMIKAWA[*] Kazuyuki KOIZUMI[†] Takashi SEO[‡]

Abstract

In this paper, we propose new definitions for multivariate skewness and kurtosis when the covariance matrix has a block diagonal structure. These measures are based on the ones of Mardia (1970). We give the expectations and the variances for new multivariate sample measures of skewness and kurtosis. Further, we derive asymptotic distributions of statistics by new measures under multivariate normality. To evaluate accuracy of these statistics, numerical results are given by the Monte Carlo simulation. We consider the problem estimating for covariance structure. Pavlenko, Björkström and Tillander (2012) consider a method which approximate the inverse covariance matrix to block diagonal structure via gLasso. In this paper, we propose an improvement of Pavlenko, Björkström and Tillander (2012)'s method by using AIC. Finally, numerical results are shown in order to investigate probability that the new method select true model for covariance structure.

Key Words and Phrases: Multivariate Skewness; Multivariate Kurtosis; Normality test; Covariance structure approximation; Block diagonal matrix; AIC.

1 Introduction

In multivariate statical analysis, normality for sample is assumed in many cases. Hence, assessing for multivariate normality is an important problem. The graphical method and the statistical hypothesis testing are considered to this problem by many authors (see, e.g. Henze (2002); Thode (2002)). From aspects of calculation cost and simplicity, we focus on the testing theory based on multivariate skewness and kurtosis. There are various definitions of multivariate skewness and kurtosis (Mardia (1970), Malkovich and Afifi (1973), Srivastava (1984) and so on.). Mardia (1970, 1974) defined multivariate skewness and kurtosis as a natural extension of univariate case. To assess multivariate normality, sample measures of multivariate skewness and kurtosis have been defined and their asymptotic distributions under the multivariate normality have been given in Mardia (1970). Srivastava (1984) also has considered another definition for the sample measures by using principal component scores and derived their asymptotic null distributions. Recently, the sample measure of multivariate kurtosis of the form containing

[*]Dept. of Mathematical Information Science, Graduate School of Science, Tokyo University of Science, Kagurazaka 1-3, Shinjuku-ku, Tokyo 162-8601, Japan

[†]Dept. of International College of Arts and Sciences, Yokohama City University, Seto 22-2, Kanazawa-ku, Yokohama city, Kanagawa 236-0027, Japan

[‡]Dept. of Mathematical Information Science, Faculty of Science, Tokyo University of Science, Kagurazaka 1-3, Shinjuku-ku, Tokyo 162-8601, Japan

Mardia (1970) and Srivastava (1984) has been proposed by Miyagawa et al. (2012). Recently, Koizumi et al. (2009) proposed an omnibus test statistic MJB by using Mardia's and Srivastava's skewness and kurtosis for assessing multivariate normality. Improvements of MJB test statistic have been discussed by many authors (see, e.g. Enomoto et al. (2012); Koizumi et al. (2013)).

These theory have been considered when the sample size N is larger than the dimension p. Since sample covariance matrix S is singular, it cannot be used when the dimension p is larger than sample size N. In this paper, we propose new measures of multivariate skewness and kurtosis when the covariance structure is a block diagonal matrix and derive their asymptotic distributions under the multivariate normality.

Further, we consider to estimate for inverse covariance matrix to a block diagonal structure. Pavlenko et al. (2012) propose a gLasso estimator of inverse covariance matrix Ξ. Similar estimates of Ξ were also considered in Rothman et al. (2008) and Rütimann et al. (2009). By the Pavlenko's method, Ξ may be estimated as incomplete block diagonal matrix. In this paper, we give an improvement of Pavlenko's method by using Akaike's information criterion (AIC). AIC is proposed by Akaike (1973, 1974) as an estimator of the risk function based on Kullback-Leibler information (Kullback and Leibler (1951)). It is used for selecting the optimal model among the candidate models.

The organization of this paper is as following. In Section 2, we examine the definition of Mardia's (1970) multivariate skewness and kurtosis and their asymptotic distributions under the multivariate normality. In Section 3, we propose new multivariate skewness and kurtosis and derive their asymptotic distributions under the multivariate normality. In Section 4, we consider the Pavlenko's method for the block diagonal approximation and propose an AIC-method which is an improvement of Pavlenko's method by using AIC. In Section 5, numerical results are given by Monte Carlo simulation to evaluate the accuracy of the upper percentage points for new statistics proposed in Section 3. Further, we investigate correct selection rate (CSR) of AIC-method proposed in Section 4.

2 Mardia's multivariate skewness and multivariate kurtosis

First, we discuss measures of multivariate skewness and multivariate kurtosis defined by Mardia (1970). Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be random p-vectors with the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ. Then Mardia (1970) has defined population measures of multivariate skewness and kurtosis as

$$\beta_1 = \mathrm{E}[\{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\}^3], \tag{2.1}$$

$$\beta_2 = \mathrm{E}[\{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}^2], \tag{2.2}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent and identical random vectors. We note that $\beta_1 = 0$ and $\beta_2 = p(p+2)$ hold under multivariate normality. These measures are invariant under a nonsingular transformation

$$\boldsymbol{x} = A\boldsymbol{u} + \boldsymbol{b}, \tag{2.3}$$

where A is a nonsingular $p \times p$ matrix and $\boldsymbol{b}$ is a p-vector. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be sample observation vectors of size N from a multivariate population with the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ. And let $\overline{\boldsymbol{x}}$ and S be the sample mean vector and the sample covariance matrix based on sample size N as follows:

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{x}_j,$$

$$S = \frac{1}{N}\sum_{j=1}^{N} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})',$$

respectively. Then the sample measures of multivariate skewness and multivariate kurtosis in Mardia (1970) are defined as

$$b_1 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'S^{-1}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})\}^3, \tag{2.4}$$

$$b_2 = \frac{1}{N}\sum_{i=1}^{N}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'S^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^2. \tag{2.5}$$

These measures are invariant under a nonsingular transformation in (2.3). Then Mardia (1970) obtained the following lemma.

Lemma 1. (Mardia (1970)) *The expectation of* b_1 *in* (2.4) *and the expectation and the variance of* b_2 *in* (2.5) *under the multivariate normal population* $N_p(\boldsymbol{\mu}, \Sigma)$ *are given by*

$$E[b_1] = \frac{1}{N}p(p+1)(p+2) + o(N^{-1}), \tag{2.6}$$

$$E[b_2] = \frac{N-1}{N+1}p(p+2), \tag{2.7}$$

$$Var[b_2] = \frac{8}{N}p(p+2) + o(N^{-1}). \tag{2.8}$$

Proof. Mardia (1970) rewrite (2.4) as

$$b_1 = \sum_{r=1}^{p}\sum_{r'=1}^{p}\sum_{s=1}^{p}\sum_{s'=1}^{p}\sum_{t=1}^{p}\sum_{t'=1}^{p} S^{rr'}S^{ss'}S^{tt'}, \tag{2.9}$$

where

$$S^{-1} = \{S^{ij}\} \text{ and } M_{111}^{(rst)} = \frac{1}{N}\sum_{i=1}^{N}(x_{ri} - \overline{x}_r)(x_{si} - \overline{x}_s)(x_{ti} - \overline{x}_t).$$

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$. Since b_1 is invariant under linear transformation, we assume $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I_p$. S converges to Σ in probability. Therefore, from (2.9), we get

$$b_1 \xrightarrow{P} \sum_{r,s,t}\{M_{111}^{(rst)}\}^2, \tag{2.10}$$

in probability. We have

$$b_1 \xrightarrow{P} \{M_3^{(1)}\}^2 + \cdots + 3\{M_{21}^{(12)}\}^2 + \cdots + 6\{M_{111}^{(123)}\}^2 + \cdots \tag{2.11}$$

in probability by writing $M_{111}^{(rrr)} = M_3^{(r)}$ and $M_{111}^{(rss)} = M_{12}^{(rs)}$ $(r \neq s)$ in (2.10). Using the normality of the vector $\{M_3^{(1)}, \ldots, M_{21}^{(12)}, \ldots, M_{111}^{(123)}, \ldots\}$, Mardia (1970) derive

$$\mathrm{E}[b_1] = \frac{1}{N}p(p+1)(p+2) + o(N^{-1}).$$

Let $\boldsymbol{x}_r^* = (x_{r1}, x_{r2}, \ldots, x_{rN})'$ $(r = 1, 2, \ldots, p)$. We transform $\boldsymbol{x}_r^*$ to $\boldsymbol{\xi}_r^* = (\xi_{r1}, \xi_{r2}, \ldots, \xi_{rN})'$ by an orthogonal transformation $\boldsymbol{\xi}_r^* = C\boldsymbol{x}_r^*$ $(r = 1, 2, \ldots, p)$, where C is an orthogonal matrix with the first row as $(1/\sqrt{N}, 1/\sqrt{N}, \ldots, 1/\sqrt{N})$ and the second row as $(-a, \ldots, -a, 1/\sqrt{Na})$, a being $1/\sqrt{N(N-1)}$. Then we find

$$\mathrm{E}[b_2] = (N-1)^2\mathrm{E}[y],$$

where

$$y = \boldsymbol{\xi}_2'(\sum_{t=2}^{N} \boldsymbol{\xi}_t \boldsymbol{\xi}_t')^{-1}\boldsymbol{\xi}_2 \text{ with } \boldsymbol{\xi}_t' = (\xi_{1t}, \xi_{2t}, \ldots, \xi_{pt})' \ (t = 2, \ldots, N).$$

From Rao (1965, p. 459), we find the probability density function of y. Hence, Mardia (1970) give

$$\mathrm{E}[b_2] = \frac{N-1}{N+1}p(p+2).$$

Let $S = I + S^*$ so that to order $N^{-1}, \mathrm{E}(S^*) = 0$. On using

$$S^{-1} = (I + S^*)^{-1} = I - S^* + S^{*2} - \cdots$$

in (2.5), we obtain that

$$b_2 = \frac{1}{N}\sum_{i=1}^{N}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^2 - \frac{2}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'S^*(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) + \cdots. \tag{2.12}$$

Further, we rewrite (2.12) as

$$b_2 = \sum_{i=1}^{p} M_4^{(i)} + \sum_{i \neq j} M_{22}^{(ij)} - 2\sum_{i=1}^{p} M_2^{(i)*}M_4^{(i)} - 2\sum_{i \neq j} M_2^{(j)*}M_{22}^{(ij)} - 2\sum_{i=1}^{p}\sum_{j \neq k} M_{11}^{(jk)*}M_{211}^{(ijk)} \ldots, \tag{2.13}$$

where

$$M_{i_1,\ldots,i_s}^{(j_1,\ldots,j_s)} = \frac{1}{N}\sum_{i=1}^{N}\{\prod_{r=1}^{s}(x_{j_r i} - \overline{x}_{j_r})^{i_r}\}, \ M_2^{(i)*} = S_{ii}^*, \ M_{11}^{(ij)*} = S_{ij}^*$$

with $S^* = \{S_{ij}^*\}$. Using the normality of the vector $M_{i_1,\ldots,i_s}^{(j_1,\ldots,j_s)}$, Mardia (1970) derive

$$\mathrm{Var}[b_2] = \frac{8}{N}p(p+2) + o(N^{-1}).$$

□

In Lemma 1, we note that E$[b_1]$, Var$[b_2]$ are asymptotic results. By using Lemma 1, Mardia (1970) derived limiting distributions of b_1 and b_2 as following:

Theorem 1. (Mardia (1970)) *Let b_1 and b_2 in* (2.4) *and* (2.5) *are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$. Then, for large N,*

$$z_1 = \frac{N}{6} b_1^2 \tag{2.14}$$

has a χ^2-distribution with $p(p+1)(p+2)/6$ degrees of freedom and

$$z_2 = \frac{b_2 - \dfrac{N-1}{N+1} p(p+2)}{\sqrt{\dfrac{8}{N} p(p+2)}} \tag{2.15}$$

is distributed as $N(0, 1)$.

Proof. From (2.6), we find with the help of the well known results on the limiting distributions of quadratic forms that

$$N[\{M_3^{(1)}\}^2 + \cdots + 3\{M_{21}^{(12)}\}^2 + \cdots + 6\{M_{111}^{(123)}\}^2 + \cdots]/6$$

has a χ^2-distribution with $p(p+1)(p+2)/6$ degrees of freedom. On using results given by (2.7) and (2.8) and the central limit theorem Mardia (1970) derived the asymptotic normality of z_2. □

Let b_1 and b_2 in (2.4) and (2.5) are expressed as follows:

$$b_1 = N \sum_{i=1}^{N} \sum_{j=1}^{N} R_{ij}^3,$$
$$b_2 = N \sum_{i=1}^{N} R_{ii}^2,$$

where

$$R_{ij} = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})'(NS)^{-1}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}). \tag{2.16}$$

Then, Mardia (1974) obtained the following lemma about exact expectation and variance by deriving moments of R_{ij}.

Lemma 2. (Mardia (1974)) *The exact expectation of b_1 and the exact expectation and the variance of b_2 when the population is $N_p(\boldsymbol{\mu}, \Sigma)$ are given by*

$$E[b_1] = \frac{p(p+2)}{(N+1)(N+3)} \{(N+1)(p+1) - 6\}, \tag{2.17}$$

$$E[b_2] = \frac{N-1}{N+1} p(p+2), \tag{2.18}$$

$$Var[b_2] = \frac{8p(p+2)}{(N+1)^2(N+3)(N+5)} (N-p-1)(N-p+1). \tag{2.19}$$

Proof. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$. Since b_1 and b_2 are invariant under linear transformation, we assume without any loss of generality that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I_p$. Further, let us write $\boldsymbol{x}_r^* = (x_{r1}, x_{r2}, \ldots, x_{rN})'$ $(r = 1, 2, \ldots, p)$. We transform $\boldsymbol{x}_r^*$ to $\boldsymbol{\xi}_r^* = (\xi_{r1}, \xi_{r2}, \ldots, \xi_{rN})'$ by the Helmert orthogonal transformation so that

$$
\begin{aligned}
\xi_{r,i-1} &= \sqrt{\frac{i-1}{i}} \left\{ -x_{r,i} + \frac{1}{i-1} \sum_{k=1}^{i-1} x_{r,k} \right\} \ (i = 2, 3, \ldots, N), \\
\xi_{r,N} &= \sqrt{N}\, \overline{x}_r.
\end{aligned}
$$

Then, we can rewrite (2.16) as

$$
R_{ij} = \left(-a_i \boldsymbol{z}_{i-1} + \sum_{k=i}^{N-1} b_k \boldsymbol{z}_k \right)' \left(-a_j \boldsymbol{z}_{j-1} + \sum_{k=j}^{N-1} b_k \boldsymbol{z}_k \right), \tag{2.20}
$$

where

$$
a_i = \sqrt{\frac{i-1}{i}},\ b_i = \frac{1}{\sqrt{i(i+1)}},\ \boldsymbol{z}_i = T^{-1} \boldsymbol{\xi}_i \ (i = 1, 2, \ldots, N) \text{ and } NS = TT'. \tag{2.21}
$$

And, the moments of $\boldsymbol{z}_k$ can be obtained (see, e.g. Khatri (1959); Khatri and Pillai (1967)). Using the moments of R_{ij}, Mardia (1974) derive (2.17), (2.18) and (2.19). □

By using Lemma 2 and Theorem 1, Mardia (1974) gave the following theorem:

Theorem 2. (Mardia (1974)) *Let b_1 and b_2 are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$ Then, for large N*

$$
z_1^* = \frac{NK}{6} b_1 \tag{2.22}
$$

where $K = \frac{(p+1)(N+1)(N+3)}{N\{(N+1)(p+1)-6\}}$ has a χ^2-distribution with $p(p+1)(p+2)/6$ degrees of freedom and

$$
z_2^* = \frac{\{(N+1)b_2 - p(p+2)(N-1)\}\sqrt{(N+3)(N+5)}}{\sqrt{8p(p+2)(N-3)(N-p-1)(N-p+1)}} \tag{2.23}
$$

is distributed as $N(0, 1)$.

Proof. From (2.14) and $\mathrm{E}[z_1^*] = p(p+1)(p+2)/6$, z_1^* is asymptotically distributed as χ^2-distribution. On using the results given by (2.18) and (2.19) and the central limit theorem z_2^*'s asymptotic distribution is standard normal. □

3 Multivariate skewness and multivariate kurtosis when covariance matrix Σ is block diagonal structure

In this section, we propose new measures of multivariate skewness and kurtosis. These measures can be used even when the dimension p is larger than sample size N. This is called high-dimensional framework. Let $\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(k)})'$ and $\boldsymbol{y} = (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \dots, \boldsymbol{y}^{(k)})'$ are random p-vectors with the mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \dots, \boldsymbol{\mu}^{(k)})'$ and the covariance matrix $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_k)$. $\boldsymbol{x}^{(l)}$, $\boldsymbol{y}^{(l)}$ and $\boldsymbol{\mu}^{(l)}$ are p_l-vectors ($p_l < N,\ l = 1, 2, \dots, k$), $\sum_{l=1}^{k} p_l = p$, and Σ_l is a $p_l \times p_l$ matrix. This covariance structure is called block diagonal structure. Then we propose new population measures of multivariate skewness and kurtosis which are natural extensions of Mardia's (1970) measures as follows:

$$\beta_{h,1} = \mathrm{E}\left[\sum_{l=1}^{k}\{(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})'\Sigma_l{}^{-1}(\boldsymbol{y}^{(l)} - \boldsymbol{\mu}^{(l)})\}^3\right],$$

$$\beta_{h,2} = \mathrm{E}\left[\sum_{l=1}^{k}\{(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})'\Sigma_l{}^{-1}(\boldsymbol{x}^{(l)} - \boldsymbol{\mu}^{(l)})\}^2\right],$$

where $\boldsymbol{x}^{(l)}$ and $\boldsymbol{y}^{(l)}$ are independent and identical random vectors. We note that $\beta_{h,1} = 0$ and $\beta_{h,2} = \sum_{l=1}^{k} p_l(p_l + 2)$ hold under the multivariate normality. These measures are invariant under a nonsingular transformation

$$\boldsymbol{x} = A^*\boldsymbol{u} + \boldsymbol{b}, \tag{3.1}$$

where $A^* = \mathrm{diag}(A_1, A_2, \dots, A_k)$, A_l is a nonsingular $p_l \times p_l$ matrix ($l = 1, 2, \dots, k$) and $\boldsymbol{b}$ is a p-vector. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N$ be sample observation vectors of size N from a multivariate population with the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ, where $\boldsymbol{x}_j = (\boldsymbol{x}_j^{(1)}, \boldsymbol{x}_j^{(2)}, \dots, \boldsymbol{x}_j^{(k)})'$ ($j = 1, 2, \dots, N$). Let $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}^{(1)}, \overline{\boldsymbol{x}}^{(2)}, \dots, \overline{\boldsymbol{x}}^{(k)})'$, $S = \mathrm{diag}(S_1, S_2, \dots, S_k)$ be the sample mean vector and the sample covariance matrix based on sample size N as follows:

$$\overline{\boldsymbol{x}}^{(l)} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{x}_j^{(l)},$$

$$S_l = \frac{1}{N}\sum_{j=1}^{N}(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})'.$$

Then we propose sample measures of multivariate skewness and kurtosis as

$$b_{h,1} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{k}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})'S_l^{-1}(\boldsymbol{x}_j^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^3, \tag{3.2}$$

$$b_{h,2} = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{k}\{(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})'S_l^{-1}(\boldsymbol{x}_i^{(l)} - \overline{\boldsymbol{x}}^{(l)})\}^2. \tag{3.3}$$

By using results in Lemma 1 , we obtained the following lemma.

Lemma 3. *When p_l and k are fixed, the expectation of $b_{h,1}$ in (3.2) and the expectation and the* *variance of $b_{h,2}$ in (3.3) when the population is $N_p(\boldsymbol{\mu}, \Sigma)$ are given by*

$$E[b_{h,1}] = \frac{1}{N}\sum_{l=1}^{k} p_l(p_l+1)(p_l+2) + o(N^{-1}), \tag{3.4}$$

$$E[b_{h,2}] = \frac{N-1}{N+1}\sum_{l=1}^{k} p_l(p_l+2), \tag{3.5}$$

$$Var[b_{h,2}] = \frac{8}{N}\sum_{l=1}^{k} p_l(p_l+2) + o(N^{-1}). \tag{3.6}$$

Proof. Let $\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(k)})'$ be a random vectors from $N_p(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \dots, \boldsymbol{\mu}^{(k)})'$ and $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_k)$. $\boldsymbol{x}^{(l)}$, $\boldsymbol{y}^{(l)}$ and $\boldsymbol{\mu}^{(l)}$ are p_l-vectors ($p_l < N$, $l = 1, 2, \dots, k$), $\sum_{l=1}^{k} p_l = p$, and Σ_l is a $p_l \times p_l$ matrix. The probability density function of $\boldsymbol{x}$ is defined as

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]. \tag{3.7}$$

Since $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{l=1}^{k}(\boldsymbol{x}^{(l)}-\boldsymbol{\mu}^{(l)})'\Sigma_l^{-1}(\boldsymbol{x}^{(l)}-\boldsymbol{\mu}^{(l)})$ and $|\Sigma| = \prod_{l=1}^{k}|\Sigma_l|$, we rewrite (3.7) as

$$f(\boldsymbol{x}) = \prod_{l=1}^{k} \frac{1}{(2\pi)^{\frac{p_l}{2}}|\Sigma_l|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}^{(l)}-\boldsymbol{\mu}^{(l)})'\Sigma_l^{-1}(\boldsymbol{x}^{(l)}-\boldsymbol{\mu}^{(l)})\right].$$

Hence, we find the independence of $\boldsymbol{x}^{(l)}$ and $\boldsymbol{x}^{(l')}$ ($l \neq l'$, $l, l' = 1, 2, \dots, k$). By using results in Lemma 1 and independence of $\boldsymbol{x}^{(l)}$ and $\boldsymbol{x}^{(l')}$ ($l \neq l'$, $l, l' = 1, 2, \dots, k$), we obtain (3.4), (3.5) and (3.6). □

By using Lemma 3, we derived the following theorem:

Theorem 3. *Let $b_{h,1}$ and $b_{h,2}$ in (3.2) and (3.3) are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$ Then, for large N,*

$$z_{h,1} = \frac{N}{6} b_{h,1} \tag{3.8}$$

has a χ^2-distribution with $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ degrees of freedom and

$$z_{h,2} = \frac{b_{h,2} - \dfrac{N-1}{N+1}\sum_{l=1}^{k} p_l(p_l+2)}{\sqrt{\dfrac{8}{N}\sum_{l=1}^{k} p_l(p_l+2)}} \tag{3.9}$$

is distributed as $N(0, 1)$.

Proof. From Theorem 1, we derive (3.8). On using the results given by (3.5) and (3.6) and the central limit theorem we find that (3.9). □

Further, by similar way of Lemma 2 and 3, we obtained the following lemma.

Lemma 4. *When p_l and k are fixed, the exact expectation of $b_{h,1}$ in* (3.2) *and the exact expectation and the exact variance of $b_{h,2}$ in* (3.3) *when the population is $N_p(\boldsymbol{\mu}, \Sigma)$ are given by*

$$E[b_{h,1}] = \sum_{l=1}^{k} \frac{p_l(p_l+2)}{(N+1)(N+3)}\{(N+1)(p_l+1)-6\}, \tag{3.10}$$

$$E[b_{h,2}] = \sum_{l=1}^{k} \frac{N-1}{N+1} p_l(p_l+2), \tag{3.11}$$

$$Var[b_{h,2}] = \sum_{l=1}^{k} \frac{8p_l(p_l+2)}{(N+1)^2(N+3)(N+5)}(N-p_l-1)(N-p_l+1). \tag{3.12}$$

By using Lemma 4, we derived the following theorem:

Theorem 4. *Let $b_{h,1}$ and $b_{h,2}$ in* (3.2) *and* (3.3) *are sample measures of multivariate skewness and multivariate kurtosis on the basis of random samples of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$ Then, for large N,*

$$z^*_{h,1} = \frac{N}{6}\sum_{l=1}^{k} \frac{(p_l+1)(N+1)(N+3)}{N\{(N+1)(p_l+1)-6\}} b_{h,1} \tag{3.13}$$

has a χ^2-distribution with $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ degrees of freedom and

$$z^*_{h,2} = \frac{\{(N+1)b_{h,2} - \sum_{l=1}^{k} p_l(p_l+2)(N-1)\}\sqrt{(N+3)(N+5)}}{\sqrt{8\sum_{l=1}^{k} p_l(p_l+2)(N-3)(N-p_l-1)(N-p_l+1)}} \tag{3.14}$$

is distributed as $N(0,1)$.

Proof. From (3.8) and $\mathrm{E}[z^*_{h,1}] = \sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$, we derive (3.13). On using the results given by (3.11) and (3.12) and the central limit theorem we obtain (3.14). □

We consider a statistic based on Wilson-Hilferty transformation (Wilson and Hilferty (1931)), an effective and simple transform of $z^*_{h,1}$ to normality, where $z^*_{h,1}$ is defined in Theorem (3.13). It is well known as normalizing transformation to fast convergence in distribution. Then, we derive the following theorem:

Theorem 5. *Let $b_{h,1}$ in* (3.2) *be a sample measure of multivariate skewness on the basis of random samples of size N drawn from $N_p(\boldsymbol{\mu}, \Sigma)$ Then*

$$z_{wh} = \left\{ \left(\frac{z^*_{h,1}}{f}\right)^{\frac{1}{3}} - 1 + \frac{2}{9f} \right\} \Big/ \sqrt{\frac{2}{9f}}, \ f = \frac{1}{6}\sum_{l=1}^{k} p_l(p_l+1)(p_l+2) \tag{3.15}$$

is distributed as $N(0,1)$ when $f \to \infty$ after $N \to \infty$.

Proof. The statistic (3.13) converge in distribution to χ^2-distribution with $\sum_{l=1}^{k} p_l(p_l+1)(p_l+2)/6$ degrees of freedom under large N. By evaluating the leading terms of characteristic function of (3.13) with large dimension p and under large N, we obtain (3.15). □

4 Covariance structure approximation

In this section, we propose a new method of estimation for block diagonal structure. Let $\boldsymbol{x}$ be a random p-vectors from $N_p(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be sample observation vectors of size N from $N_p(\boldsymbol{\mu}, \Sigma)$. Assume $\Xi = \mathrm{diag}(\Xi_1, \Xi_2, \ldots, \Xi_k) = \Sigma^{-1}$ where Ξ_l is a $p_l \times p_l$ matrix, $p_l < N,\ l = 1, 2, \ldots, k$ and $\sum_{l=1}^{k} p_l = p$. Our purpose is to get a estimator of Ξ.

To ensure that the estimator of Ξ exists and be sparsity we make the following assumptions about the covariance matrix Σ.

Existence. There exist such a constant $\varepsilon > 0$ that

$$0 < \varepsilon \leq \phi_{\min}(\Sigma) \leq \phi_{\max}(\Sigma) < \frac{1}{\varepsilon},$$

where $\phi_{\min}(\Sigma)$ and $\phi_{\max}(\Sigma)$ are the smallest and the largest eigenvalues of Σ, respectively. This condition ensures that Ξ exists.

Sparsity. Let $A = \{(i,j) : \xi_{ij} \neq 0, i > j\}$ denote the set of non-zero off-diagonal entries of Ξ. For the number of A-elements, we assume that

$$\#A < \frac{p(p-1)}{2},$$

where $\#A$ means the number of set A. This assumption is to ensure sparsity of Ξ.

Then, Pavlenko et al. (2012) proposed a gLasso estimator of Ξ as the minimizer of the penalized negative log-likelihood

$$\widehat{\Xi}_\lambda = \arg\min_{\Xi > 0} \{\mathrm{tr}(\Xi\widehat{\Sigma}) - \log|\Xi| + \lambda||\Xi^-||_1\},$$

where $\widehat{\Sigma}$ is the maximum likelihood estimator of Σ, $\Xi^- = \Xi - \mathrm{diag}(\Xi)$, $||\Xi^-||_1 = \sum_{i<j} |\xi_{i,j}|$ is ℓ_1-norm of Ξ^-, λ is a non-negative tuning parameter, and λ is the order $\sqrt{\log p/N}$ (see Rothman et al. (2008)). This estimator is similar to the original gLasso introduced in Friedman et al. (2007) (they used $||\Xi||_1$ instead of $||\Xi^-||_1$).

Further, following the modification to fast convergence be considered by Pavlenko et al. (2012). Let $\mathcal{K}$ denote the inverse of correlation matrix and Γ denote the diagonal matrix of the standard deviations. Then, a gLasso estimator of $\mathcal{K}$ be defined as

$$\widetilde{\mathcal{K}}_\lambda = \arg\min_{\mathcal{K} > 0} \{\mathrm{tr}(\mathcal{K}\widehat{\mathcal{K}}^{-1}) - \log|\mathcal{K}| + \lambda||\mathcal{K}^-||_1\}, \tag{4.1}$$

where $\widehat{\mathcal{K}}^{-1}$ is the estimated correlation matrix. Since $\mathcal{K} = (\kappa_{i,j}) = \Gamma\Xi\Gamma$, the estimator of Ξ be given by

$$\widetilde{\Xi}_\lambda = \widehat{\Gamma}^{-1}\widetilde{\mathcal{K}}_\lambda\widehat{\Gamma}^{-1}, \tag{4.2}$$

where $\widehat{\Gamma}$ is a sample estimator of Γ. We call this procedure gLasso-method.

However, these estimators cannot necessarily estimate Ξ to the block diagonal structure. Then, we propose an AIC-method of making Ξ the block diagonal matrix by using Akaike's information criterion (AIC). AIC is defined as

$$\begin{aligned} \mathrm{AIC} = &- 2 \times (\text{The maximum log-likelihood}) \\ &+ 2 \times (\text{The number of parameters}). \end{aligned} \tag{4.3}$$

The model which makes AIC the minimum is considered to be the optimal model. Our method of estimation for block diagonal structure is following:

(A.1) We calculate $\widetilde{\Xi}_\lambda$by gLasso estimator in (4.1) and (4.2).

(A.2) Candidate models are determined from obtained $\widetilde{\Xi}_\lambda$.

(A.3) AICs for all candidate models are calculated by (4.3).

(A.4) We select the optimal model by values of AICs.

Hence, a block diagonal estimation of Ξ be attained.

An example of the proposed AIC-method is given. Parameters are the following: $p = 6, N = 10, \lambda = 0.29$ and population is $N_p(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = \mathbf{0}, \Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3)$, $\Sigma_l = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ $(l = 1, 2, 3)$ and $\rho = 0.85$. Then

$$\widetilde{\Xi}_\lambda = \begin{pmatrix} 0.73 & -0.35 & 0 & 0 & 0 & 0.03 \\ -0.35 & 0.68 & 0 & 0 & 0 & 0.04 \\ 0 & 0 & 0.83 & -0.26 & 0 & 0 \\ 0 & 0 & -0.26 & 0.77 & 0.11 & 0 \\ 0 & 0 & 0 & 0.11 & 0.72 & -0.40 \\ 0.03 & 0.04 & 0 & 0 & -0.40 & 0.60 \end{pmatrix}.$$

is calculated by glasso pakage in R. Next, we consider how to decide candidate models. When we decide candidate models, we need the following rule:

(R.1) (The number of 0 in each block matrix) ≤ 2.

(R.2) If the number of 0 is not contained in block matrix which has not overlapped under (R.1), the size of this matrix do not make small.

(R.3) If block matrix which satisfy (R.1) has overlapped, we fix one block matrix and make others small.

Under these rules, we find four candidate models in this case. For example,

$$\begin{pmatrix} 0.73 & -0.35 & 0 & 0 & 0 & 0 \\ -0.35 & 0.68 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.83 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.77 & 0.11 & 0 \\ 0 & 0 & 0 & 0.11 & 0.72 & -0.40 \\ 0 & 0 & 0 & 0 & -0.40 & 0.60 \end{pmatrix}$$

is model$(2, 1, 3)$(=model(p_1, p_2, p_3)) and there are model$(2, 2, 2)$, model$(2, 3, 1)$ and model$(2, 1, 2, 1)$. We calculate AIC for each candidate model, respectively. In this case, AIC in (4.3) becomes

$$\mathrm{AIC} = N(p \log 2\pi - \log |S^{-1}| + p) + 2d,$$

where S is the maximum likelihood estimator of Σ, d is the number of free parameters of a model. S^{-1} and AIC of the model$(2,1,3)$ be calculated as

$$S^{-1}(2,1,3) = \begin{pmatrix} 5.64 & -5.02 & 0 & 0 & 0 & 0 \\ -5.02 & 5.09 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.08 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.01 & 3.02 & -2.10 \\ 0 & 0 & 0 & 3.02 & 9.79 & -7.70 \\ 0 & 0 & 0 & -2.10 & -7.70 & 6.56 \end{pmatrix},$$

$$\mathrm{AIC}(2,1,3) = 161.4.$$

In similar way, AICs of model$(2,2,2)$, model$(2,3,1)$ and model$(2,1,2,1)$ are calculated as

$$\mathrm{AIC}(2,2,2) = 151.2,\ \mathrm{AIC}(2,3,1) = 170.5,\ \mathrm{AIC}(2,1,2,1) = 184.5.$$

Since AIC$(2,2,2)$ is the smallest value in this example, model$(2,2,2)$ is the optimal model. In this case, true model is selected.

5 Simulation studies

5.1 Approximation accuracy of new multivariate skewness and kurtosis

In this subsection, we investigate accuracies of upper percentage points of proposed statistics $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ based on new multivariate skewness $b_{h,1}$ in (3.2) and kurtosis in (3.3) by the Monte Carlo simulation study. For each block, we assume that $\boldsymbol{\mu}^{(l)} = \mathbf{0}$ and $\Sigma_l = I_{p_l}$ $(l = 1, 2, \ldots, k)$ without loss of generality from (3.1). Simulation parameters are as follows:

$$p = 200, 300, 400, 500, 1000, 2000,\ p_l = 5, 10, 20,\ N = 50, 100, 200, 400, 800.$$

As a numerical experiment, we carry out 10,000 and 1,000 replications for the case of $N < 400$ and $N \geq 400$, respectively. But for the cases of $p = 1000, 2000$, we carry out 1,000 replications for all parameters.

It may be seen from Tables 1-7 that Type I Error probabilities of all statistics converge to the level of significance α when N is large. These results show that Theorems 3-5 hold. We note that $z^*_{h,1}$ and z_{wh} are improvements of $z_{h,1}$ for all parameters. We turn out that $z^*_{h,2}$ is an improvement of $z_{h,2}$ when $N \leq 400$. $z_{h,2}$ and $z^*_{h,2}$ are the almost same approximate accuracy when $N = 800$. Through this simulation, we recommend $z^*_{h,1}$ and $z_{w,h}$ for the skewness test. And when $N \leq 400$, we recommend $z^*_{h,2}$ for the kurtosis test.

5.2 Correct Selection Rate of AIC-method

In this subsection, we investigate correct selection rate (CSR) of AIC-method and gLasso-method by simulation studies, respectively. CSR of AIC-method calculated by using algorithm (A.1)-(A.4) in Section 4 is the probability of selecting the true model. CSR of gLasso-method calculated by (4.1) and (4.2) is the probability of selecting the true model. We decide candidate models under the condition (R.1)-(R.3) in Section 4. As a numerical experiment, we carry out 100 replications. Simulation parameters are the following: $p = 10,\ N = 10, 20,\ \lambda = \sqrt{\log p/N}$. We consider two cases for the covariance structure of population.

- (Case 1) $\boldsymbol{x} \sim N_p(\mathbf{0}, \Sigma)$,

  $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \Sigma_5),\ \Sigma_l = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\ (l = 1, 2, 3, 4, 5),\ \rho = 0.9.$

- (Case 2) $\boldsymbol{x} \sim N_p(\mathbf{0}, \Sigma)$,

  $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4),\ \Sigma_s = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}\ (s = 1, 3),\ \Sigma_t = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\ (t = 2, 4),\ \rho = 0.9.$

Table 8 give CSR for the Case 1 by AIC-method and gLasso-method. Table 9 give CSR for the Case 2 by AIC-method and gLasso-method. From Tables 8 and 9, we note that our method improve gLasso-method by using AIC. Even when N is small, CSR of AIC-method is quite higher than the one of gLasso-method.

6 Conclusion

In this paper, we considered tests for the multivariate normality when $p > N$. We proposed new definitions for multivariate skewness and kurtosis as natural extensions of Mardia's measures, and derived their asymptotic distributions under the multivariate normal population. Approximate accuracies of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ were evaluated by Monte Carlo simulation.

And we considered the problem to estimate for the covariance structure. There is gLasso-method in Pavlenko et al. (2012) for this problem. We proposed an AIC-method which is an improvement of gLasso-method by using an information criterion AIC. Finally, correct selection rates of AIC-method were given by simulation.

Table 1 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.1$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.002	0.128	0.133	0.058	0.104
		100	0.014	0.120	0.120	0.076	0.099
		200	0.040	0.106	0.109	0.087	0.100
		400	0.068	0.113	0.111	0.108	0.110
		800	0.072	0.090	0.094	0.097	0.099
	10	50	0.000	0.125	0.124	0.036	0.100
		100	0.004	0.120	0.129	0.065	0.099
		200	0.024	0.112	0.118	0.084	0.102
		400	0.051	0.101	0.130	0.091	0.101
		800	0.079	0.111	0.129	0.104	0.109
	20	50	0.000	0.099	0.095	0.010	0.106
		100	0.000	0.128	0.124	0.047	0.106
		200	0.007	0.123	0.124	0.073	0.101
		400	0.040	0.122	0.121	0.096	0.114
		800	0.066	0.119	0.117	0.098	0.106
300	5	50	0.000	0.131	0.124	0.060	0.106
		100	0.010	0.119	0.117	0.077	0.099
		200	0.033	0.111	0.109	0.086	0.100
		400	0.038	0.120	0.120	0.091	0.106
		800	0.033	0.110	0.120	0.088	0.100
	10	50	0.000	0.127	0.129	0.037	0.102
		100	0.001	0.128	0.124	0.070	0.102
		200	0.017	0.115	0.115	0.080	0.100
		400	0.043	0.111	0.108	0.090	0.100
		800	0.071	0.107	0.097	0.084	0.089
	20	50	0.000	0.100	0.094	0.010	0.107
		100	0.000	0.123	0.125	0.046	0.102
		200	0.003	0.118	0.127	0.069	0.099
		400	0.019	0.113	0.102	0.088	0.101
		800	0.054	0.128	0.106	0.095	0.100

Table 2 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.1$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.126	0.127	0.054	0.103
		100	0.006	0.124	0.121	0.078	0.103
		200	0.025	0.114	0.118	0.086	0.098
		400	0.058	0.124	0.102	0.088	0.097
		800	0.058	0.083	0.114	0.100	0.101
	10	50	0.000	0.130	0.127	0.037	0.102
		100	0.001	0.125	0.124	0.073	0.110
		200	0.013	0.117	0.114	0.081	0.099
		400	0.035	0.111	0.112	0.084	0.093
		800	0.063	0.105	0.094	0.084	0.090
	20	50	0.000	0.105	0.097	0.009	0.106
		100	0.000	0.127	0.128	0.047	0.102
		200	0.001	0.123	0.130	0.072	0.103
		400	0.016	0.112	0.128	0.086	0.103
		800	0.037	0.116	0.114	0.091	0.100
500	5	50	0.000	0.127	0.134	0.055	0.102
		100	0.004	0.124	0.117	0.076	0.104
		200	0.023	0.114	0.108	0.088	0.101
		400	0.044	0.113	0.113	0.100	0.105
		800	0.063	0.110	0.103	0.102	0.104
	10	50	0.000	0.128	0.119	0.039	0.109
		100	0.000	0.128	0.122	0.063	0.098
		200	0.010	0.117	0.117	0.086	0.103
		400	0.015	0.110	0.109	0.093	0.102
		800	0.053	0.098	0.098	0.108	0.113
	20	50	0.000	0.098	0.099	0.009	0.103
		100	0.000	0.128	0.126	0.044	0.099
		200	0.001	0.115	0.124	0.073	0.103
		400	0.009	0.115	0.115	0.088	0.098
		800	0.051	0.126	0.126	0.098	0.107

Table 3 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.05$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.000	0.072	0.072	0.021	0.055
		100	0.006	0.063	0.068	0.033	0.050
		200	0.018	0.056	0.056	0.043	0.052
		400	0.029	0.060	0.054	0.055	0.060
		800	0.038	0.047	0.044	0.049	0.051
	10	50	0.000	0.070	0.071	0.011	0.053
		100	0.001	0.063	0.072	0.026	0.049
		200	0.009	0.063	0.063	0.039	0.052
		400	0.029	0.058	0.080	0.041	0.045
		800	0.040	0.061	0.056	0.056	0.059
	20	50	0.000	0.050	0.048	0.002	0.055
		100	0.000	0.074	0.071	0.017	0.053
		200	0.002	0.068	0.073	0.031	0.052
		400	0.016	0.067	0.066	0.044	0.058
		800	0.031	0.066	0.058	0.048	0.056
300	5	50	0.000	0.076	0.072	0.023	0.057
		100	0.004	0.066	0.065	0.034	0.052
		200	0.015	0.060	0.060	0.043	0.052
		400	0.017	0.064	0.058	0.043	0.054
		800	0.014	0.057	0.058	0.042	0.051
	10	50	0.000	0.076	0.075	0.011	0.054
		100	0.000	0.073	0.068	0.031	0.054
		200	0.007	0.064	0.062	0.036	0.048
		400	0.022	0.057	0.055	0.040	0.048
		800	0.028	0.056	0.055	0.039	0.044
	20	50	0.000	0.053	0.049	0.001	0.054
		100	0.000	0.071	0.071	0.017	0.053
		200	0.001	0.064	0.071	0.029	0.049
		400	0.008	0.060	0.065	0.040	0.051
		800	0.027	0.068	0.049	0.050	0.055

Table 4 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.05$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.070	0.073	0.021	0.052
		100	0.003	0.068	0.068	0.035	0.054
		200	0.011	0.060	0.063	0.040	0.049
		400	0.027	0.059	0.051	0.046	0.052
		800	0.028	0.036	0.063	0.051	0.052
	10	50	0.000	0.075	0.068	0.012	0.052
		100	0.000	0.069	0.068	0.030	0.057
		200	0.006	0.067	0.063	0.037	0.050
		400	0.016	0.056	0.057	0.033	0.041
		800	0.025	0.052	0.057	0.042	0.045
	20	50	0.000	0.052	0.049	0.001	0.054
		100	0.000	0.074	0.073	0.016	0.053
		200	0.001	0.070	0.073	0.032	0.051
		400	0.004	0.055	0.072	0.042	0.049
		800	0.019	0.055	0.050	0.049	0.056
500	5	50	0.000	0.075	0.075	0.019	0.053
		100	0.001	0.069	0.063	0.034	0.051
		200	0.009	0.062	0.060	0.042	0.051
		400	0.019	0.048	0.048	0.049	0.056
		800	0.032	0.049	0.049	0.048	0.054
	10	50	0.000	0.072	0.068	0.013	0.054
		100	0.000	0.068	0.068	0.027	0.052
		200	0.004	0.065	0.065	0.040	0.055
		400	0.006	0.061	0.061	0.043	0.052
		800	0.028	0.051	0.051	0.043	0.049
	20	50	0.000	0.049	0.048	0.002	0.053
		100	0.000	0.073	0.070	0.014	0.050
		200	0.000	0.067	0.069	0.030	0.053
		400	0.002	0.058	0.058	0.044	0.050
		800	0.018	0.068	0.068	0.050	0.057

Table 5 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.01$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
200	5	50	0.000	0.020	0.018	0.002	0.011
		100	0.001	0.015	0.018	0.006	0.012
		200	0.003	0.013	0.013	0.007	0.011
		400	0.004	0.008	0.014	0.010	0.011
		800	0.008	0.014	0.011	0.009	0.010
	10	50	0.000	0.018	0.021	0.001	0.012
		100	0.000	0.017	0.020	0.003	0.010
		200	0.001	0.014	0.014	0.006	0.012
		400	0.008	0.020	0.015	0.007	0.011
		800	0.006	0.009	0.015	0.014	0.017
	20	50	0.000	0.010	0.010	0.000	0.012
		100	0.000	0.019	0.018	0.001	0.012
		200	0.000	0.016	0.018	0.004	0.010
		400	0.002	0.016	0.015	0.010	0.010
		800	0.005	0.014	0.012	0.010	0.011
300	5	50	0.000	0.021	0.021	0.002	0.013
		100	0.001	0.017	0.017	0.004	0.011
		200	0.002	0.014	0.016	0.008	0.012
		400	0.003	0.016	0.010	0.008	0.011
		800	0.003	0.013	0.016	0.009	0.013
	10	50	0.000	0.019	0.020	0.001	0.011
		100	0.000	0.019	0.020	0.003	0.013
		200	0.001	0.016	0.015	0.006	0.010
		400	0.004	0.015	0.014	0.008	0.010
		800	0.003	0.007	0.017	0.011	0.012
	20	50	0.000	0.010	0.009	0.000	0.013
		100	0.000	0.021	0.018	0.002	0.012
		200	0.001	0.014	0.019	0.004	0.009
		400	0.001	0.013	0.018	0.006	0.010
		800	0.007	0.017	0.012	0.009	0.010

Table 6 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $\alpha = 0.01$

			Skewness			Kurtosis	
p	p_l	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
400	5	50	0.000	0.020	0.019	0.003	0.013
		100	0.000	0.018	0.019	0.005	0.011
		200	0.002	0.013	0.016	0.007	0.010
		400	0.005	0.012	0.010	0.008	0.009
		800	0.007	0.010	0.012	0.008	0.009
	10	50	0.000	0.020	0.019	0.001	0.013
		100	0.000	0.019	0.018	0.004	0.013
		200	0.001	0.019	0.014	0.007	0.010
		400	0.003	0.013	0.012	0.006	0.012
		800	0.002	0.011	0.010	0.004	0.005
	20	50	0.000	0.008	0.010	0.000	0.012
		100	0.000	0.020	0.020	0.001	0.011
		200	0.000	0.018	0.019	0.004	0.011
		400	0.001	0.016	0.020	0.008	0.012
		800	0.001	0.010	0.009	0.009	0.010
500	5	50	0.000	0.021	0.019	0.002	0.011
		100	0.000	0.019	0.016	0.005	0.010
		200	0.002	0.014	0.013	0.008	0.010
		400	0.004	0.007	0.007	0.008	0.009
		800	0.005	0.009	0.010	0.008	0.009
	10	50	0.000	0.019	0.018	0.001	0.013
		100	0.000	0.018	0.018	0.003	0.012
		200	0.000	0.018	0.018	0.007	0.011
		400	0.000	0.012	0.012	0.003	0.008
		800	0.006	0.010	0.010	0.004	0.005
	20	50	0.000	0.009	0.008	0.000	0.012
		100	0.000	0.019	0.019	0.001	0.010
		200	0.000	0.018	0.017	0.005	0.012
		400	0.000	0.011	0.011	0.006	0.013
		800	0.004	0.015	0.015	0.009	0.012

Table 7 The upper percentage points of $z_{h,1}$, $z^*_{h,1}$, z_{wh}, $z_{h,2}$ and $z^*_{h,2}$ for $p_l = 20$

			Skewness			Kurtosis	
p	α	N	$z_{h,1}$	$z^*_{h,1}$	z_{wh}	$z_{h,2}$	$z^*_{h,2}$
1000	0.1	50	0.000	0.110	0.110	0.030	0.130
		100	0.000	0.134	0.134	0.049	0.110
		200	0.000	0.125	0.125	0.068	0.097
		400	0.000	0.122	0.121	0.060	0.099
	0.05	50	0.000	0.060	0.060	0.000	0.080
		100	0.000	0.077	0.075	0.020	0.059
		200	0.000	0.065	0.065	0.030	0.048
		400	0.000	0.064	0.063	0.035	0.051
	0.01	50	0.000	0.020	0.020	0.000	0.030
		100	0.000	0.017	0.017	0.000	0.019
		200	0.000	0.010	0.010	0.008	0.016
		400	0.000	0.012	0.011	0.009	0.011
2000	0.1	50	0.000	0.111	0.111	0.008	0.109
		100	0.000	0.149	0.149	0.041	0.106
		200	0.000	0.119	0.118	0.076	0.110
		400	0.000	0.115	0.115	0.093	0.106
	0.05	50	0.000	0.058	0.056	0.001	0.056
		100	0.000	0.082	0.082	0.017	0.049
		200	0.000	0.070	0.070	0.031	0.055
		400	0.000	0.080	0.080	0.040	0.058
	0.01	50	0.000	0.012	0.012	0.001	0.012
		100	0.000	0.018	0.018	0.001	0.014
		200	0.000	0.024	0.023	0.007	0.011
		400	0.000	0.020	0.020	0.008	0.011

Table 8 Comparison of CSR (case 1)

N	gLasso-method	AIC-method
10	0.19	0.87
20	0.56	0.92

Table 9 Comparison of CSR (case 2)

N	gLasso-method	AIC-method
10	0.31	0.64
20	0.68	0.94

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, **1**, 267–281.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

[3] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

[4] Enomoto, R., Okamoto, N. and Seo, T. (2012). Multivariate normality test using Srivastava's skewness and kurtosis. *SUT Journal of Mathematics*, **48**, 103–115.

[5] Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.

[6] Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers*, **43**, 467–506.

[7] Khatri, C. G. (1959). On the mutual independence of certain statistics. *Annals of Mathematical Statistics*, **30**, 1258–1262.

[8] Khatri, C. G. and Pillai, K. C. S. (1967). On the moments of traces of two matrices in multivariate analysis. *Annals of the Institute of Statistical Mathematics*, **19**, 143–156.

[9] Koizumi, K., Hyodo, M. and Pavlenko, T. (2013). Modified Jarque-Bera type tests for multivariate normality in a high-dimensional framework. *(submitted).*

[10] Kullback, S. and Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.

[11] Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association*, **68**, 176–179.

[12] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

[13] Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B*, **36**, 115–128.

[14] Miyagawa, C., Koizumi, K. and Seo, T. (2011). A new multivariate kurtosis and its asymptotic distribution. *SUT Journal of Mathematics*, **47**, 55–71.

[15] Pavlenko, T., Björkström, A. and Tillander, A. (2012). Covariance structure approximation via gLasso in high-dimensional supervised classification. *Journal of Applied Statistics*, **39**, 1643–1666.

[16] Rao, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

[17] Rothman,.A., Bickel, P., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.

[18] Rütimann,.P. and Bühlmann,.P. (2009). High dimensional sparse covariance estimation via directed acyclic graphs. *Electronic Journal of Statistics*, **3**, 1133–1160.

[19] Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters*, **2**, 263–267.

[20] Thode, H. C. (2002). *Testing for normality.* Marcel Dekker.

[21] Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Mathematics*, **17**, 684–688.