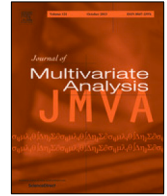




Contents lists available at ScienceDirect

## Journal of Multivariate Analysis

journal homepage: [www.elsevier.com/locate/jmva](http://www.elsevier.com/locate/jmva)

## Sub-dimensional Mardia measures of multivariate skewness and kurtosis

Joydeep Chowdhury<sup>a,\*</sup>, Subhajit Dutta<sup>b</sup>, Reinaldo B. Arellano-Valle<sup>c</sup>, Marc G. Genton<sup>a</sup><sup>a</sup> Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia<sup>b</sup> Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur 208016, India<sup>c</sup> Department of Statistics, Pontificia Universidad Católica de Chile, Santiago 22, Chile

## ARTICLE INFO

## Article history:

Received 29 November 2021

Received in revised form 21 July 2022

Accepted 21 July 2022

Available online 29 July 2022

## AMS 2020 subject classifications:

primary 62H15

secondary 62H12

## Keywords:

Asymptotic distribution

Measures of multivariate skewness and kurtosis

Multivariate normality test

Skew-normal distribution

Skew-t distribution

Symmetric distribution

## ABSTRACT

The Mardia measures of multivariate skewness and kurtosis summarize the respective characteristics of a multivariate distribution with two numbers. However, these measures do not reflect the sub-dimensional features of the distribution. Consequently, testing procedures based on these measures may fail to detect skewness or kurtosis present in a sub-dimension of the multivariate distribution. We introduce sub-dimensional Mardia measures of multivariate skewness and kurtosis, and investigate the information they convey about all sub-dimensional distributions of some symmetric and skewed families of multivariate distributions. The maxima of the sub-dimensional Mardia measures of multivariate skewness and kurtosis are considered, as these reflect the maximum skewness and kurtosis present in the distribution, and also allow us to identify the sub-dimension bearing the highest skewness and kurtosis. Asymptotic distributions of the vectors of sub-dimensional Mardia measures of multivariate skewness and kurtosis are derived, based on which testing procedures for the presence of skewness and of deviation from Gaussian kurtosis are developed. The performances of these tests are compared with some existing tests in the literature on simulated and real datasets.

© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

Consider a  $p$ -variate random vector  $\mathbf{X} = (X_1, \dots, X_p)^T$  from a multivariate distribution with mean vector  $\boldsymbol{\mu} \in \mathbb{R}^p$  and  $p \times p$  positive definite covariance matrix  $\boldsymbol{\Sigma}$ . Mardia [27] defined the measures of multivariate skewness and kurtosis:

$$\beta_{1,p} = E[\{(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})\}^3] \in \mathbb{R}_+, \quad \beta_{2,p} = E[\{(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})\}^2] \in \mathbb{R}_+, \quad (1)$$

respectively, where  $\mathbf{X}$  and  $\mathbf{Y}$  are independent and identically distributed. These measures are invariant under affine transformations and each provide a single number to summarize the skewness and the kurtosis of that  $p$ -dimensional distribution. Their sample counterparts have known asymptotic distribution and can be used to test for normality [27,28]. For the multivariate normal distribution, it is well known that  $\beta_{1,p} = 0$  and  $\beta_{2,p} = p(p+2)$ .

One drawback of the Mardia measures of multivariate skewness and kurtosis is that they summarize the information about skewness and kurtosis too much. Sub-dimensional distributions may exhibit evidence of skewness or kurtosis, which may not be reflected in the overall Mardia measures of multivariate skewness or kurtosis. For example, consider

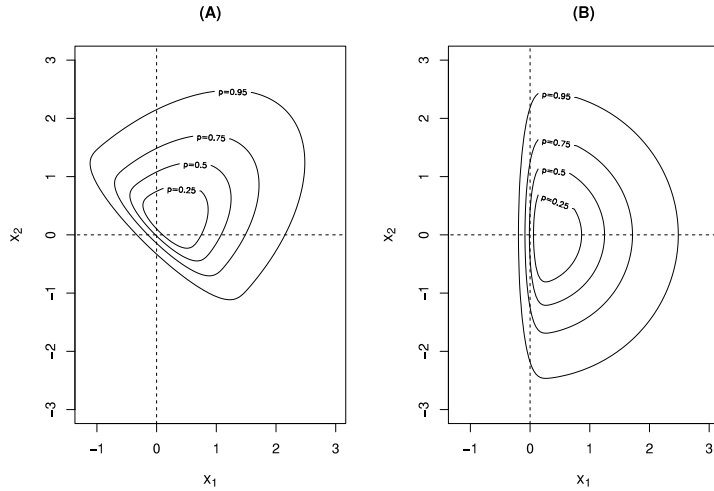
\* Corresponding author.

E-mail address: [joydeep.chowdhury@kaust.edu.sa](mailto:joydeep.chowdhury@kaust.edu.sa) (J. Chowdhury).

**Table 1**

P-values of the Mardia skewness test in some sub-dimensions of Fisher's iris setosa data. The variables denoted as 1, 2, 3 and 4 are sepal length, sepal width, petal length and petal width, respectively, of the iris setosa flower.

Sub-dimensions	(4)	(1, 4)	(2, 4)	(3, 4)	(1, 2, 3, 4)
p-values	0.001	0.012	0.019	0.018	0.236



**Fig. 1.** (A) Contour plot of  $\mathcal{SN}_2(\mathbf{0}, \Omega, \alpha)$ , which has marginal skewness  $\beta_{1,1,1} = \beta_{1,1,2} = 0.130$ . (B) Contour plot of  $\mathcal{SN}_2(\mathbf{0}, \mathbf{I}_2, \alpha^*)$ , which has marginal skewness  $\beta_{1,1,1} = 0.889$  and  $\beta_{1,1,2} = 0$ . Both the skew-normal distributions have the same Mardia measure of multivariate skewness  $\beta_{1,2} = 0.889$ .

the celebrated Fisher's iris data for the species 'iris setosa' [12]. In Table 1, the p-values of the Mardia test of skewness are presented for four sub-dimensions along with the whole dataset, that is, the sub-dimension (1, 2, 3, 4). Here, the variables 1, 2, 3 and 4 correspond to sepal length, sepal width, petal length and petal width, respectively. One can see that in the complete dataset, there is no significant evidence of skewness, while evidence of skewness in the reported sub-dimensions of dimension one (consisting of the fourth variable, namely, petal width) and of dimension two is quite strong at the 5% level as reflected by the corresponding p-values. This presence of skewness in the distribution of petal width of the species 'iris setosa' was also observed in [38]. Therefore, this motivates the investigation of the Mardia measures of multivariate skewness and kurtosis on sub-dimensional marginals.

Let  $\mathbf{X}_{qi}$  denote a subvector of dimension  $q$  for  $1 \leq q \leq p$  from the random vector  $\mathbf{X}$ , and let  $\mu_{qi}$  and  $\Sigma_{qi}$  be the corresponding entries of  $\mu$  and  $\Sigma$  for  $i \in \{1, \dots, Q_q\}$  with  $Q_q = \binom{p}{q}$ . We define the following sub-dimensional Mardia measures of multivariate skewness and kurtosis

$$\beta_{1,q,i} = E [ \{ (\mathbf{X}_{qi} - \mu_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{Y}_{qi} - \mu_{qi}) \}^3 ] \in \mathbb{R}_+, \quad \beta_{2,q,i} = E [ \{ (\mathbf{X}_{qi} - \mu_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \mu_{qi}) \}^2 ] \in \mathbb{R}_+, \tag{2}$$

for  $i \in \{1, \dots, Q_q\}$ , where  $\mathbf{X}_{qi}$  and  $\mathbf{Y}_{qi}$  are independent and identically distributed. When  $q = p$ , we get back  $\beta_{1,p,1} \equiv \beta_{1,p}$  and  $\beta_{2,p,1} \equiv \beta_{2,p}$ . For  $q \in \{1, \dots, p\}$ , we collect these measures in the following vectors

$$\mathbf{M}_{1,q} = (\beta_{1,q,1}, \dots, \beta_{1,q,Q_q})^\top \in \mathbb{R}_+^{Q_q}, \quad \mathbf{M}_{2,q} = (\beta_{2,q,1}, \dots, \beta_{2,q,Q_q})^\top \in \mathbb{R}_+^{Q_q}, \tag{3}$$

each of dimension  $Q_q = \binom{p}{q}$ . We call (3) the  $q$ -th vectors of sub-dimensional Mardia measures of multivariate skewness and kurtosis. For the multivariate normal distribution,  $\mathcal{N}_p(\mu, \Sigma)$ , we have

$$\mathbf{M}_{1,q}^{\mathcal{N}} = \mathbf{0}_{Q_q}, \quad \mathbf{M}_{2,q}^{\mathcal{N}} = q(q+2)\mathbf{1}_{Q_q}, \tag{4}$$

for all  $1 \leq q \leq p$ , where  $\mathbf{0}_{Q_q}$  and  $\mathbf{1}_{Q_q}$  are  $Q_q$ -dimensional vectors of zeros and of ones, respectively.

We further define

$$\mathbf{M}_1^* = (\mathbf{M}_{1,1}^\top, \mathbf{M}_{1,2}^\top, \dots, \mathbf{M}_{1,p}^\top)^\top = (\beta_{1,1,1}, \dots, \beta_{1,p,1})^\top \in \mathbb{R}_+^{2^p-1}, \tag{5}$$

$$\mathbf{M}_2^* = (\mathbf{M}_{2,1}^\top, \mathbf{M}_{2,2}^\top, \dots, \mathbf{M}_{2,p}^\top)^\top = (\beta_{2,1,1}, \dots, \beta_{2,p,1})^\top \in \mathbb{R}_+^{2^p-1}. \tag{6}$$

Here,  $\mathbf{M}_1^*$  and  $\mathbf{M}_2^*$  collect all the sub-dimensional Mardia measures of multivariate skewness and kurtosis.

Due to affine invariance, the Mardia measures of multivariate skewness and kurtosis cannot detect any difference between distributions which are affine transformations of one another, but whose marginal skewness and/or kurtosis values are very different. In Fig. 1, we depict such an example for the skew-normal distribution [3] with  $p = 2$ . Let  $\Omega = (\omega_{ij})$ , where  $\omega_{ij} = 0.5 + 0.5\mathbb{I}(i = j)$  and  $\alpha = (5, 5)^\top$ . Consider the skew-normal distribution  $\mathcal{SN}_2(\mathbf{0}, \Omega, \alpha)$  and

its ‘canonical form’  $\mathcal{SN}_2(\mathbf{0}, \mathbf{I}_2, \boldsymbol{\alpha}^*)$ ; see [1, Proposition 4] and [2, subsection 5.1.8]. The contour plots of the densities of  $\mathcal{SN}_2(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  and  $\mathcal{SN}_2(\mathbf{0}, \mathbf{I}_2, \boldsymbol{\alpha}^*)$  are presented in Fig. 1. From Propositions 3 and 4 in [1], it follows that the skew-normal distributions  $\mathcal{SN}_2(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  and  $\mathcal{SN}_2(\mathbf{0}, \mathbf{I}_2, \boldsymbol{\alpha}^*)$  have the same values of Mardia measure of multivariate skewness,  $\beta_{1,2} = 0.889$ . However, the distribution  $\mathcal{SN}_2(\mathbf{0}, \mathbf{I}_2, \boldsymbol{\alpha}^*)$  has all its skewness in its first component  $X_1$  with the marginal distribution of  $X_2$  being symmetric, while both the components of the distribution  $\mathcal{SN}_2(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  have the same marginal skewness. This further supports the study of Mardia measures of multivariate skewness and kurtosis on sub-dimensional marginals.

There are certain non-Gaussian distributions for which all the lower-dimensional marginals are Gaussian. An example of such a distribution, which is a trivariate generalized skew-normal distribution [16] with all the univariate and bivariate marginal distributions being standard normal, is given in [25]. Here, any procedure based on  $\mathbf{M}_{1,q}$  or  $\mathbf{M}_{2,q}$ , defined in (3), for  $q < 3$  would not be able to detect the presence of non-Gaussianity in the distribution. However, a procedure based on  $\mathbf{M}_1^*$  defined in (5) would be able to detect non-Gaussianity, since  $\mathbf{M}_1^*$  and  $\mathbf{M}_2^*$  incorporate the Mardia measures on all the sub-dimensions including the whole dimension. This further motivates developing procedures based on  $\mathbf{M}_1^*$  and  $\mathbf{M}_2^*$ , which are defined earlier in (5) and (6).

In this paper, we start by studying the forms of  $\mathbf{M}_{1,q}$  and  $\mathbf{M}_{2,q}$  defined in (3) for some parametric classes of non-Gaussian distributions, including, for example, the multivariate Student’s  $t$  distribution and the multivariate skew-normal and skew- $t$  distributions. Then, we propose tests of normality for each dimension  $q$  based on the maximum entry of  $\mathbf{M}_{1,q}$  and of  $\mathbf{M}_{2,q}$ , as well as global tests of multivariate normality based on the maximum entry of  $\mathbf{M}_1^*$  and of  $\mathbf{M}_2^*$  in (5) and (6). One important advantage of our approach is the ability of the test, when it rejects multivariate normality, to identify the dimension  $q$  and the associated sub-dimensions for which the rejection occurs.

The definitions of the Mardia measures require that the population covariance matrix  $\boldsymbol{\Sigma}$  is non-singular. On the other hand, the estimation of the Mardia measures requires the sample covariance matrix to be non-singular. Estimation based on  $g$ -inverses is considered in [29], but the null distributions of the Mardia tests get altered when the sample covariance matrix is singular. For these reasons, the Mardia measures and associated tests are not applicable in a high-dimensional setup, particularly when  $n \leq p$ . While the quantities  $\mathbf{M}_1^*$  and of  $\mathbf{M}_2^*$  cannot be used because they include the Mardia measures, the quantities  $\mathbf{M}_{1,q}$  and  $\mathbf{M}_{2,q}$  can be considered for small values of  $q$  even when  $n \leq p$ . If it is known that the sub-dimension supporting skewness or kurtosis has a small dimension  $q$  considerably less than the sample size  $n$ , then the quantities  $\mathbf{M}_{1,q}$  and  $\mathbf{M}_{2,q}$  can be employed to investigate the presence of skewness or kurtosis. Some possible ways in which the methodology can be extended to the general high-dimensional setup are discussed in Section 9.

This paper is organized as follows. The vectors of sub-dimensional Mardia measures of multivariate skewness and kurtosis in the case of some symmetric distributions are investigated in Section 2, whereas in the case of some skewed distributions are considered in Section 3. The invariance of these sub-dimensional measures under location-scale transformations is studied in Section 4. The new hypothesis tests are introduced in Section 5 and their asymptotic distributions are established in Section 6. The results of a Monte Carlo simulation study of the empirical sizes and powers of the new tests, as well as of the sub-dimensional detection, are reported in Section 7. Sub-dimensional data analyses of Fisher’s iris data and of wind speed data near a wind farm in Saudi Arabia are presented in Section 8. The paper ends with a discussion in Section 9. Some additional numerical results are provided in the supplementary material.

## 2. Sub-dimensional Mardia measures for some symmetric distributions

First, let  $\mathbf{X} \stackrel{d}{=} \mathbf{S} \circ \mathbf{T} \in \mathbb{R}^p$  be a standardized symmetric random vector, in which  $\circ$  represents the componentwise (Hadamard) product,  $\mathbf{S} = \text{sign}(\mathbf{X}) = (\text{sign}(X_1), \dots, \text{sign}(X_p))^T \sim \mathcal{U}\{-1, +1\}^p$  (discrete uniform), and  $\mathbf{S}$  is independent of  $\mathbf{T} = |\mathbf{X}| = (|X_1|, \dots, |X_p|)^T$ . Let  $\mathbf{X}_\pi = \mathbf{S}_\pi \circ \mathbf{X} = \mathbf{S}_\pi \circ \mathbf{S} \circ \mathbf{T}$ , where  $\mathbf{S}_\pi$  is formed by a permutation  $\pi = (\pi_1, \dots, \pi_p)$  of the components of  $\mathbf{S}$ . Note that  $\mathbf{s} \circ \mathbf{t} = \mathbf{D}(\mathbf{s})\mathbf{t} = \mathbf{D}(\mathbf{t})\mathbf{s}$ , where  $\mathbf{D}(\mathbf{a}) = \text{diag}(a_1, \dots, a_p)$ . Suppose that  $\mathbf{X}$  has finite fourth moment with  $E(\mathbf{X}) = \mathbf{0}_p$  and  $\text{Var}(\mathbf{X}) = \mathbf{I}_p$ . This implies  $E(\|\mathbf{X}\|^2) = p$ , where  $\|\cdot\|$  denotes the Euclidean norm. Also, let  $\mathbf{X}_{\pi_*} = \mathbf{S}_{\pi_*} \circ \mathbf{X}$ , where  $\pi_*$  represents a permutation where no component remains in its original position. For example, for  $p = 3$ ,  $\pi \in \{(1, 2, 3), (2, 1, 3), (2, 3, 1), (3, 2, 1), (3, 1, 2), (1, 3, 2)\}$  and  $\pi_* \in \{(2, 3, 1), (3, 1, 2)\}$ . Since  $\mathbf{S}_{\pi_*} \circ \mathbf{S}$  has mean vector  $\mathbf{0}_p$  and covariance matrix  $\mathbf{I}_p$ , we then have  $E(\mathbf{X}_{\pi_*}) = E(\mathbf{S}_{\pi_*} \circ \mathbf{S}) \circ E(|\mathbf{X}|) = \mathbf{0}_p$  and  $\text{Var}(\mathbf{X}_{\pi_*}) = \mathbf{I}_p$ . Also, obviously  $\|\mathbf{X}_{\pi_*}\| \stackrel{d}{=} \|\mathbf{X}_\pi\| \stackrel{d}{=} \|\mathbf{X}\|$ , hence  $E(\|\mathbf{X}_{\pi_*}\|^k) = E(\|\mathbf{X}_\pi\|^k) = E(\|\mathbf{X}\|^k)$ ,  $k \geq 1$ . So, for the random vector  $\mathbf{X}_{\pi_*}$ , we have  $\beta_{1,p} = 0$  and  $\beta_{2,p} = E(\|\mathbf{X}\|^4)$ . In particular,

1. If  $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ , with probability density function (pdf) given by  $f(\mathbf{x}) = (2\pi)^{-p/2} \exp(-\|\mathbf{x}\|^2)$ ,  $\mathbf{x} \in \mathbb{R}^p$ , then  $\|\mathbf{X}_{\pi_*}\|^2 \stackrel{d}{=} \|\mathbf{X}\|^2 \sim \chi_p^2$  and  $\beta_{2,p} = p(p + 2)$ . Therefore, for any  $q$ -dimensional subvector  $\mathbf{X}_q$  of  $\mathbf{X}$ , we have  $\mathbf{X}_q \sim \mathcal{N}_q(\mathbf{0}_q, \mathbf{I}_q)$  and so  $\beta_{1,q} = \beta_{1,p} = 0$  and  $\beta_{2,q} = q(q + 2)$ . Therefore, the  $q$ th vectors of sub-dimensional Mardia measures of multivariate skewness and kurtosis are as in (4).
2. If  $\mathbf{X}$  is spherically distributed with pdf  $f(\mathbf{x}) = h(\|\mathbf{x}\|^2)$ ,  $\mathbf{x} \in \mathbb{R}^p$ , for some density generator function  $h$ , i.e.,  $h(u) > 0$  for  $u > 0$  and  $\int_0^\infty u^{p/2-1} h(u) du = \Gamma(p/2)/\pi^{p/2}$ , and  $\text{Var}(\mathbf{X}) = \sigma^2 \mathbf{I}_p$ , then  $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(p)}$ , where  $R \stackrel{d}{=} \|\mathbf{X}\|$ , with  $E(R^2) = p\sigma^2$ , and  $R$  is independent of  $\mathbf{U}^{(p)} \stackrel{d}{=} \mathbf{X}/\|\mathbf{X}\|$ , the uniform vector on the  $p$ -dimensional unit sphere. Therefore,  $\beta_{1,p} = 0$  by symmetry and

$$\beta_{2,p} = \frac{p^2 E(R^4)}{\{E(R^2)\}^2} = p(p + 2)(\kappa + 1),$$

where  $\kappa = \gamma_2 = \{E(X_1^4) - 3\}/3$  is the excess of kurtosis in the corresponding spherical univariate distribution, which can be computed from the relation

$$\kappa + 1 = \frac{\beta_{2,p}}{p(p+2)} = \frac{p}{p+2} \frac{E(R^4)}{\{E(R^2)\}^2},$$

with the assumption  $E(R^4) < \infty$ . Moreover, any  $q$ -dimensional subvector  $\mathbf{X}_q$  of  $\mathbf{X}$ , is also spherically distributed with stochastic representation  $\mathbf{X}_q = R_q \mathbf{U}^{(q)}$ , where  $R_q = \sqrt{B_q} R$  with Beta  $B_q \sim \mathcal{B}(q/2, (p-q)/2)$  and  $R, B_q$  and  $\mathbf{U}^{(q)}$  are mutually independent. Using the fact that  $E(B_q^s) = \{\Gamma(p/2)\Gamma(q/2+s)\}/\{\Gamma(q/2)\Gamma(p/2+s)\}$ , we find again that  $\beta_{1,q} = 0$  and

$$\beta_{2,q} = \frac{q^2 E(B_q^2) E(R^4)}{\{E(B_q) E(R^2)\}^2} = \frac{q(q+2)}{p(p+2)} \beta_{2,p} = q(q+2)(\kappa + 1).$$

Hence, for spherical distributions

$$\mathbf{M}_{1,q}^{SPH} = \mathbf{0}_{0q}, \quad \mathbf{M}_{2,q}^{SPH} = q(q+2)(\kappa + 1)\mathbf{1}_{0q},$$

for all  $1 \leq q \leq p$ . For example:

(a) If  $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ , then  $R^2 \sim \chi_p^2$ , with  $E(R^2) = p, E(R^4) = p(p+2)$  and so

$$\kappa = \frac{p}{p+2} \frac{E(R^4)}{\{E(R^2)\}^2} - 1 = 0.$$

(b) If  $\mathbf{X} \sim t_p(\mathbf{0}_p, \mathbf{I}_p, \nu)$ , where  $t_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$  denotes the multivariate Student's  $t$  distribution with location vector  $\boldsymbol{\xi}$ , dispersion matrix  $\boldsymbol{\Omega}$ ,  $\nu$  degrees of freedom with  $\nu > 4$  and pdf  $c_p(\nu)|\boldsymbol{\Omega}|^{-p/2} \{1 + (\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\xi})/\nu\}^{-(\nu+p)/2}$ ,  $\mathbf{x} \in \mathbb{R}^p$ , where  $c_p(\nu) = \Gamma\{(\nu+p)/2\}/\{\Gamma(\nu/2)(\nu\pi)^{p/2}\}$ , then  $R^2 \sim pF_{p,\nu}$  with

$$E(R^{2k}) = p^k \left(\frac{\nu}{p}\right)^k \frac{\Gamma(p/2+k)\Gamma(\nu/2-k)}{\Gamma(p/2)\Gamma(\nu/2)}, \quad \nu \geq 2k,$$

thus  $\kappa = 2/(\nu - 4)$ .

(c) If  $\mathbf{X} \sim \mathcal{EP}_p(\mathbf{0}_p, \mathbf{I}_p, \nu)$ , where  $\mathcal{EP}_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$  denotes the multivariate exponential power distribution, with location vector  $\boldsymbol{\xi}$ , dispersion matrix  $\boldsymbol{\Omega}$  and kurtosis parameter  $\nu > 0$ , and pdf  $c_p(\nu)|\boldsymbol{\Omega}|^{-p/2} \exp[-\{(\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\xi})\}^\nu/2]$ ,  $\mathbf{x} \in \mathbb{R}^p$ , where  $c_p(\nu) = \{p\Gamma(p/2)\}/\{\Gamma(p/(2\nu) + 1)2^{p/(2\nu)+1}\pi^{p/2}\}$ , then  $R^2 \stackrel{d}{=} V^{1/\nu}$  with Gamma  $V \sim \mathcal{G}(p/(2\nu), 1/2)$ . Thus, we find that

$$E(R^{2k}) = \frac{2^{k/\nu} \Gamma\{(p+2k)/(2\nu)\}}{\Gamma\{p/(2\nu)\}}, \quad k \geq 1,$$

so

$$\kappa = \frac{p}{p+2} \frac{\Gamma\{(p+4)/(2\nu)\}\Gamma\{p/(2\nu)\}}{[\Gamma\{(p+2)/(2\nu)\}]^2} - 1.$$

In this case, note that  $\kappa = \kappa^{(p)}$  depends on the dimension  $p$ . This is due to the fact that each marginal distribution of a  $p$ -variate exponential power distribution depends on  $p$ . In particular, the univariate marginal distributions are not equivalent to the univariate one obtained by putting  $p = 1$ . Another relevant characteristic of this distribution is that it allows both lighter ( $\nu > 1$ ) and heavier ( $0 < \nu \leq 1$ ) tails than the normal distribution.

- The class of random vectors  $\mathbf{X}_{\pi-\pi_*}$  is also very interesting because  $E(\mathbf{X}_{\pi-\pi_*}) \neq \mathbf{0}_p$ , but some of its marginal distributions have zero mean vector. For instance, for  $p = 3$ , a random vector in this class is given by  $(S_{11}|X_1|, S_{23}|X_2|, S_{23}|X_3|)^\top$ , where  $S_{ij} = S_i S_j$ , in which the first component has mean  $E(|X_1|) > 0$ , while the remaining components have zero mean.

### 3. Sub-dimensional Mardia measures for some skewed distributions

Consider a generalized skew-normal [16] random vector  $\mathbf{X} \sim \mathcal{GSN}_p(\mathbf{0}_p, \boldsymbol{\Omega}, \boldsymbol{\lambda})$  with probability density function  $f_{\mathbf{X}}(\mathbf{x}) = 2\phi_p(\mathbf{x}; \boldsymbol{\Omega})G(\boldsymbol{\lambda}^\top \mathbf{x}), \mathbf{x} \in \mathbb{R}^p$  and some absolutely continuous symmetric cumulative distribution function  $G$  defined on  $\mathbb{R}$ . Since  $\beta_{1,p}$  and  $\beta_{2,p}$  are invariant with respect to non-singular linear transformations, they can be computed by using the canonical representations  $\mathbf{Z} = \boldsymbol{\Gamma}\mathbf{X}$  and  $\mathbf{Z}' = \boldsymbol{\Gamma}\mathbf{Y}$  of  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively, in which  $\boldsymbol{\Gamma}$  is a  $p \times p$  orthonormal matrix such that  $\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top = \mathbf{I}_p$  and  $\boldsymbol{\Gamma}\boldsymbol{\lambda} = \lambda_* \mathbf{e}_{1,p}$ , where  $\lambda_* = \sqrt{\boldsymbol{\lambda}^\top \boldsymbol{\Omega} \boldsymbol{\lambda}}$  and  $\mathbf{e}_1$  is the first  $p$ -dimensional unit vector. Then,  $Z_1 \sim \mathcal{GSN}_1(0, 1, \lambda_*)$  which has mean  $\mu_*$  and variance  $1 - \mu_*^2$ , and  $Z_1$  is independent of  $\mathbf{Z}_2 = (Z_2, \dots, Z_p)^\top \sim \mathcal{N}_{p-1}(\mathbf{0}_{p-1}, \mathbf{I}_{p-1})$ , where

$$\mu_* = 2\lambda_* E\{G(\lambda_* Y)\} = 2 \frac{dE\{G(\lambda_* Y)\}}{d\lambda_*},$$

with  $Y \sim \mathcal{HN}_1(0, 1)$  (half-normal). The same holds for  $Z'_1$  and  $\mathbf{Z}'_2 = (Z'_2, \dots, Z'_p)^\top$ . Note that  $\boldsymbol{\mu}_Z = \mu_* \mathbf{e}_1$  and  $\boldsymbol{\Sigma}_Z = \mathbf{I}_p - \mu_*^2 \mathbf{e}_1 \mathbf{e}_1^\top$ . Since  $(\mathbf{X} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) = (\mathbf{Z} - \boldsymbol{\mu}_Z)^\top \boldsymbol{\Sigma}_Z^{-1} (\mathbf{Z} - \boldsymbol{\mu}_Z) = Z_{01} Z'_{01} + \mathbf{Z}_2^\top \mathbf{Z}'_2$ , where  $Z_{01} = (Z_1 - \mu_*) / \sqrt{1 - \mu_*^2}$  and  $Z'_{01} = (Z'_1 - \mu_*) / \sqrt{1 - \mu_*^2}$  are standardized and independent random variables

$$\begin{aligned} \beta_{1,p} &= E\{(Z_{01} Z'_{01} + \mathbf{Z}_2^\top \mathbf{Z}'_2)^3\} = E\{(Z_{01} Z'_{01})^3\} + 3E\{(Z_{01} Z'_{01})^2 (\mathbf{Z}_2^\top \mathbf{Z}'_2)\} + 3E\{(Z_{01} Z'_{01}) (\mathbf{Z}_2^\top \mathbf{Z}'_2)^2\} + E\{(\mathbf{Z}_2^\top \mathbf{Z}'_2)^3\} \\ &= E\{(Z_{01} Z'_{01})^3\} = \gamma_{1,1}^3, \end{aligned}$$

where  $\gamma_{1,1} \equiv \gamma_1 = \sqrt{\beta_{1,1}} = E(Z_{01}^3)$ . Similarly, we have  $(\mathbf{X} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) = Z_{01}^2 + \mathbf{Z}_2^\top \mathbf{Z}_2$  and thus

$$\beta_{2,p} = E\{(Z_{01}^2 + \mathbf{Z}_2^\top \mathbf{Z}_2)^2\} = E(Z_{01}^4) + 2E(Z_{01}^2) E\{(\mathbf{Z}_2^\top \mathbf{Z}_2)\} + E\{(\mathbf{Z}_2^\top \mathbf{Z}_2)^2\} = E(Z_{01}^4) = \gamma_{2,1} + p(p + 2),$$

with  $\gamma_{2,1} \equiv \gamma_2 = E(Z_{01}^4) - 3 = \beta_{2,1} - 3$  being the excess of kurtosis. In this case, we can also consider the Mardia multivariate excess of kurtosis index  $\gamma_{2,p} = \beta_{2,p} - p(p + 2)$  which equals  $\gamma_{2,1}$ .

In the skew-normal case with  $G(x) = \Phi(x)$ , we have  $\gamma_1 = b(2b^2 - 1)\gamma_*^{3/2}$  and  $\gamma_2 = 2b^2(2 - 3b^2)\gamma_*^2$ , where  $\gamma_* = \lambda_*^2 / \{1 + (1 - b^2)\lambda_*^2\}$ ,  $b = \sqrt{2/\pi}$  and  $\lambda_* = \sqrt{\boldsymbol{\lambda}^\top \boldsymbol{\Omega} \boldsymbol{\lambda}}$  as defined above. However, the indices  $\gamma_1$  and  $\gamma_2$  corresponding to the subvectors of  $\mathbf{X}$  are different through subvectors with different dimensions. As was shown in [1], if  $\mathbf{X} \sim \mathcal{SN}_p(\mathbf{0}_p, \boldsymbol{\Omega}, \boldsymbol{\lambda})$  and  $\mathbf{A}$  is a  $q \times p$  fixed matrix, then  $\mathbf{A}\mathbf{X} \sim \mathcal{SN}_q(\mathbf{0}_q, \boldsymbol{\Omega}_A, \boldsymbol{\lambda}_A)$ , where  $\boldsymbol{\Omega}_A = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top$  and  $\boldsymbol{\lambda}_A = \{(\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top)^{-1} \mathbf{A}\boldsymbol{\Omega}\boldsymbol{\lambda}\} \{1 + \boldsymbol{\lambda}^\top (\boldsymbol{\Omega} - \boldsymbol{\Omega}_A^\top \boldsymbol{\Omega}_A^{-1} \boldsymbol{\Omega}) \boldsymbol{\lambda}\}^{-1/2}$ . In particular, if we consider the partition  $\mathbf{X} = (\mathbf{X}_q^\top, \mathbf{X}_{p-q}^\top)^\top$  with the corresponding partition for the scale matrix  $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_{rs})_{r,s=q,p-q}$  and the skewness vector  $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_q^\top, \boldsymbol{\lambda}_{p-q}^\top)^\top$ , we then have

$$\mathbf{X}_q = \mathbf{A}\mathbf{X} \sim \mathcal{SN}_q(\mathbf{0}_q, \boldsymbol{\Omega}_{qq}, \boldsymbol{\lambda}^{(q)}), \quad \boldsymbol{\lambda}^{(q)} = \frac{\boldsymbol{\lambda}_q + \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q} \boldsymbol{\lambda}_{p-q}}{\sqrt{1 + \boldsymbol{\lambda}_{p-q}^\top \boldsymbol{\Omega}_{p-q,p-q} \boldsymbol{\lambda}_{p-q}}},$$

where  $\mathbf{A}\mathbf{A}^\top = \mathbf{I}_q$  and  $\boldsymbol{\Omega}_{p-q,p-q;q} = \boldsymbol{\Omega}_{p-q,p-q} - \boldsymbol{\Omega}_{p-q,q} \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q}$ . The associated canonical transformation to  $\mathbf{X}_q$  has summary skewness parameter

$$\lambda_*^{(q)} = \sqrt{\boldsymbol{\lambda}^{(q)\top} \boldsymbol{\Omega}_{qq} \boldsymbol{\lambda}^{(q)}} = \sqrt{\frac{(\boldsymbol{\lambda}_q + \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q} \boldsymbol{\lambda}_{p-q})^\top \boldsymbol{\Omega}_{qq} (\boldsymbol{\lambda}_q + \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q} \boldsymbol{\lambda}_{p-q})}{1 + \boldsymbol{\lambda}_{p-q}^\top \boldsymbol{\Omega}_{p-q,p-q} \boldsymbol{\lambda}_{p-q}}}.$$

Therefore, we have  $\beta_{1,q,i} = \gamma_1^{2(q)}$  and  $\beta_{2,q,i} = \gamma_2^{(q)} + q(q + 2)$ , where for  $k = 1, 2$ ,  $\gamma_k^{(q)} = \gamma_k$  for  $q = p$ , and for  $q < p$  it must be computed as  $\gamma_k$  but with  $\lambda_*$  replaced by  $\lambda_*^{(q)}$ . In particular, if  $\boldsymbol{\Omega} = \mathbf{I}_p$ , then  $\lambda_*^{(q)} = \sqrt{(\boldsymbol{\lambda}_q^\top \boldsymbol{\lambda}_q) / (1 + \boldsymbol{\lambda}_{p-q}^\top \boldsymbol{\lambda}_{p-q})}$  with  $\lambda_*^{(q)} = \lambda_*$  if  $q = p$ , and  $\lambda_*^{(q)} = \lambda_i / \sqrt{1 + \sum_{j \neq i} \lambda_j^2}$  for the  $i$ th marginal component if  $q = 1$ .

Next, we consider the multivariate skew- $t$  distribution as described in [2, section 6.2]. Let  $\mathbf{X} \sim \mathcal{ST}_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$ , where  $\mathcal{ST}_p$  denotes a  $p$ -dimensional skew- $t$  distribution with degrees of freedom  $\nu$ , location vector  $\boldsymbol{\xi}$ , scale matrix  $\boldsymbol{\Omega}$  and skewness vector  $\boldsymbol{\alpha}$ . Define

$$\delta_* = \left( \frac{\boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}}{1 + \boldsymbol{\alpha}^\top \boldsymbol{\Omega} \boldsymbol{\alpha}} \right)^{1/2}, \quad b_\nu = \frac{\sqrt{\nu} \Gamma\{(\nu - 1)/2\}}{\sqrt{\pi} \Gamma(\nu/2)}.$$

Based on  $\delta_*$  and  $b_\nu$ , we set  $\mu_* = b_\nu \delta_*$  and  $\sigma_*^2 = \{\nu/(\nu - 2)\} - \mu_*^2$ . The Mardia measures of multivariate skewness and kurtosis [2] for  $\mathbf{X}$  are

$$\beta_{1,p} = \beta_1^* + 3(p - 1) \frac{\mu_*^2}{(\nu - 3)\sigma_*^2} \text{ if } \nu > 3, \tag{7}$$

$$\beta_{2,p} = \beta_2^* + (p^2 - 1) \frac{\nu - 2}{\nu - 4} + \frac{2(p - 1)}{\sigma_*^2} \left\{ \frac{\nu}{\nu - 4} - \frac{(\nu - 1)\mu_*^2}{\nu - 3} \right\} - p(p + 2) \text{ if } \nu > 4, \tag{8}$$

where

$$\beta_1^* = \frac{\mu_*^2}{\sigma_*^3} \left\{ \frac{\nu(3 - \delta_*^2)}{\nu - 3} - \frac{3\nu}{\nu - 2} + 2\mu_*^2 \right\}^2, \quad \beta_2^* = \frac{1}{\sigma_*^4} \left\{ \frac{3\nu^2}{(\nu - 2)(\nu - 4)} - \frac{4\mu_*^2 \nu(3 - \delta_*^2)}{\nu - 3} + \frac{6\mu_*^2 \nu}{\nu - 2} - 3\mu_*^4 \right\}.$$

Next, denote the subvector corresponding to  $\beta_{1,q,i}$  and  $\beta_{2,q,i}$  as  $\mathbf{X}_q$ , and consider the partition  $\mathbf{X} = (\mathbf{X}_q^\top, \mathbf{X}_{p-q}^\top)^\top$  with the corresponding partitions for the location vector  $\boldsymbol{\xi} = (\boldsymbol{\xi}_q^\top, \boldsymbol{\xi}_{p-q}^\top)^\top$ , scale matrix  $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_{rs})_{r,s=q,p-q}$  and the skewness vector  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_q^\top, \boldsymbol{\alpha}_{p-q}^\top)^\top$  as in the earlier part of this section. Then

$$\mathbf{X}_q \sim \mathcal{ST}_q(\boldsymbol{\xi}_q, \boldsymbol{\Omega}_{qq}, \boldsymbol{\alpha}^{(q)}, \nu), \quad \boldsymbol{\alpha}^{(q)} = \frac{\boldsymbol{\alpha}_q + \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q} \boldsymbol{\alpha}_{p-q}}{\sqrt{1 + \boldsymbol{\alpha}_{p-q}^\top \boldsymbol{\Omega}_{p-q,p-q} \boldsymbol{\alpha}_{p-q}}},$$

where  $\boldsymbol{\Omega}_{p-q,p-q;q} = \boldsymbol{\Omega}_{p-q,p-q} - \boldsymbol{\Omega}_{p-q,q} \boldsymbol{\Omega}_{qq}^{-1} \boldsymbol{\Omega}_{q,p-q}$ . Now,  $\beta_{1,q,i}$  and  $\beta_{2,q,i}$  are obtained by replacing  $\boldsymbol{\alpha}$  and  $\boldsymbol{\Omega}$  with  $\boldsymbol{\alpha}^{(q)}$  and  $\boldsymbol{\Omega}_{qq}$ , respectively, in the formulae given in (7) and (8).

#### 4. Invariance under location-scale transformations

It is well known that the Mardia measures of multivariate skewness and kurtosis (1) are invariant under affine transformations. We show next that the sub-dimensional Mardia measures of multivariate skewness and kurtosis (2) are only invariant under location and scale transformation, unless the multivariate distribution is spherically invariant.

Let  $MD^2(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$  be the squared Mahalanobis distance. Consider the affine transformation  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ ,  $\mathbf{A} \in \mathbb{R}^{p \times p}$  with  $|\mathbf{A}| \neq 0$  and  $\mathbf{b} \in \mathbb{R}^p$ . Then, it is immediate that  $MD^2(\mathbf{Y}) = MD^2(\mathbf{A}\mathbf{X} + \mathbf{b}) = MD^2(\mathbf{X}) = MD^2(\mathbf{Z})$ , where  $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ .

Next, partition  $\mathbf{X} = (\mathbf{X}_q^\top, \mathbf{X}_{p-q}^\top)^\top$  and similarly for  $\mathbf{Y}$  and  $\mathbf{Z}$ . Then,  $\mathbf{X}_q = \mathbf{B}_q\mathbf{X}$ , where  $\mathbf{B}_q = (\mathbf{I}_q, \mathbf{O}) \in \mathbb{R}^{q \times p}$  with  $\text{rank}(\mathbf{B}_q) = q$  and  $\mathbf{O}$  being the null matrix. Similarly,  $\mathbf{Y}_q = \mathbf{B}_q\mathbf{Y}$  and  $\mathbf{Z}_q = \mathbf{B}_q\mathbf{Z}$ . Therefore,

$$MD^2(\mathbf{X}_q) = (\mathbf{B}_q\mathbf{X} - \mathbf{B}_q\boldsymbol{\mu})^\top (\mathbf{B}_q\boldsymbol{\Sigma}\mathbf{B}_q^\top)^{-1}(\mathbf{B}_q\mathbf{X} - \mathbf{B}_q\boldsymbol{\mu}) = (\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{B}_q^\top (\mathbf{B}_q\boldsymbol{\Sigma}\mathbf{B}_q^\top)^{-1}\mathbf{B}_q(\mathbf{X} - \boldsymbol{\mu}) \\ = \mathbf{Z}^\top \boldsymbol{\Sigma}^{1/2} \mathbf{B}_q^\top (\mathbf{B}_q\boldsymbol{\Sigma}\mathbf{B}_q^\top)^{-1} \mathbf{B}_q \boldsymbol{\Sigma}^{1/2} \mathbf{Z} = \mathbf{Z}^\top \mathbf{P}\mathbf{Z},$$

where  $\mathbf{P} \geq 0$  is a  $p \times p$  orthogonal projection matrix, that is, symmetric with  $\mathbf{P}^2 = \mathbf{P}$  and  $\text{rank}(\mathbf{P}) = q$ . Similarly,  $MD^2(\mathbf{Y}_q) = \mathbf{Z}^\top \mathbf{P}_A \mathbf{Z}$  with  $\mathbf{P}_A = \boldsymbol{\Sigma}^{1/2} \mathbf{A}^\top \mathbf{B}_q^\top (\mathbf{B}_q \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^\top \mathbf{B}_q^\top)^{-1} \mathbf{B}_q \mathbf{A} \boldsymbol{\Sigma}^{1/2}$ , where  $\mathbf{P}_A \geq 0$  is a  $p \times p$  orthogonal projection matrix, that is, symmetric with  $\mathbf{P}_A^2 = \mathbf{P}_A$  and  $\text{rank}(\mathbf{P}_A) = q$ . Therefore, if  $\mathbf{P}_A = \mathbf{P}$  then  $MD^2(\mathbf{X}_q) = MD^2(\mathbf{Y}_q) = MD^2(\mathbf{Z}_q) = \mathbf{Z}_q^\top \mathbf{Z}_q$ . In particular, this holds when the matrix  $\mathbf{A}$  is diagonal, which proves invariance under location and scale transformations.

More generally, simple calculations show that  $\mathbf{P}_A = \mathbf{P}$  if and only if

$$\mathbf{A}_{12} \boldsymbol{\Sigma}_{21} \mathbf{A}_{11}^\top + \mathbf{A}_{11} \boldsymbol{\Sigma}_{12} \mathbf{A}_{12}^\top + \mathbf{A}_{12} \boldsymbol{\Sigma}_{22} \mathbf{A}_{12}^\top = \mathbf{O}. \tag{9}$$

In addition to the trivial case  $\mathbf{A}_{12} = \mathbf{O}$ , Eq. (9) holds, for instance, when  $\mathbf{A}_{12} \perp (\boldsymbol{\Sigma}_{21}, \boldsymbol{\Sigma}_{22})$ , where  $\perp$  means orthogonal.

Moreover, if  $\mathbf{Z}$  is spherically distributed, that is,  $\Gamma \mathbf{X} \stackrel{d}{=} \mathbf{X}$  for any orthogonal matrix  $\Gamma$ , then the invariance holds for any affine transformation with  $|\mathbf{A}| \neq 0$ . Indeed, in this case  $\mathbf{P}_A = \Gamma_A^\top \mathbf{D}_A \Gamma_A$  with  $\mathbf{D}_A = \mathbf{B}_q^\top \mathbf{B}_q$  for some orthogonal matrix  $\Gamma_A$  and therefore

$$\mathbf{Z}^\top \mathbf{P}_A \mathbf{Z} = \mathbf{Z}^\top \Gamma_A^\top \mathbf{B}_q^\top \mathbf{B}_q \Gamma_A \mathbf{Z} = (\Gamma_A \mathbf{Z})^\top \mathbf{B}_q^\top \mathbf{B}_q (\Gamma_A \mathbf{Z}) \stackrel{d}{=} \mathbf{Z}^\top \mathbf{B}_q^\top \mathbf{B}_q \mathbf{Z} = \mathbf{Z}_q^\top \mathbf{Z}_q,$$

and similarly when  $\mathbf{A} = \mathbf{I}_p$ .

In summary, although the Mardia measures of multivariate skewness and kurtosis are invariant under affine transformations, this is generally not the case for the sub-dimensional measures as shown in this section and illustrated in the simple case of a bivariate skew-normal distribution in Fig. 1. Therefore, the sub-dimensional Mardia measures of multivariate skewness and kurtosis are informative for testing normality in the sub-dimensions as proposed in the next section.

#### 5. Sub-dimensional estimation and testing of hypotheses

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be a random sample from a probability distribution  $F$  in  $\mathbb{R}^p$ . Let  $\mathbf{X}_{qij}$  denote a sub-vector of dimension  $q$  obtained from  $\mathbf{X}_j$  for  $j \in \{1, \dots, n\}$ . The dimension of  $\mathbf{X}_{qij}$  corresponds to that of  $\mathbf{X}_{qi}$  defined in Section 1, where the random vector  $\mathbf{X}$  has distribution  $F$ . More precisely, suppose  $\mathbf{X}_{qi} = \mathbf{P}_{qi}\mathbf{X}$ , where  $\mathbf{P}_{qi}$  is a diagonal matrix whose  $i$ th diagonal element is either 1 or 0 (depending on whether the  $i$ th coordinate of  $\mathbf{X}$  is included in  $\mathbf{X}_{qi}$  or not). Then,  $\mathbf{X}_{qij} = \mathbf{P}_{qi}\mathbf{X}_j$  for all  $j$ . We assume that the covariance matrix  $\boldsymbol{\Sigma}$  of the distribution  $F$  is positive definite. This implies that the covariance matrices of all sub-vectors  $\mathbf{X}_{qi}$  are positive definite, and hence all of the quantities  $\beta_{1,q,i}$  and  $\beta_{2,q,i}$  defined in (2) are well defined for the underlying distribution  $F$ . Further, we assume that  $n > p$  so that the sample covariances corresponding to all of the sub-vectors  $\mathbf{X}_{qi}$  are invertible almost surely. This is necessary for defining the sample estimators of  $\beta_{1,q,i}$  and  $\beta_{2,q,i}$ . The dimension  $p$  is fixed in this setup. Let  $\bar{\mathbf{X}}_{qi}$  and  $\mathbf{S}_{qi}$  be the sample mean and the sample covariance matrix of the observations  $\mathbf{X}_{qi1}, \dots, \mathbf{X}_{qin}$ , respectively. The sample estimators of  $\beta_{1,q,i}$  and  $\beta_{2,q,i}$  defined in (2) are

$$b_{1,q,i} = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qik} - \bar{\mathbf{X}}_{qi})\}^3, \quad b_{2,q,i} = \frac{1}{n} \sum_{j=1}^n \{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})\}^2,$$

respectively. Consequently, the estimates of  $\mathbf{M}_{1,q}$  and  $\mathbf{M}_{2,q}$  defined in (3) are  $\mathbf{m}_{1,q} = (b_{1,q,1}, \dots, b_{1,q,Q_q})^\top$  and  $\mathbf{m}_{2,q} = (b_{2,q,1}, \dots, b_{2,q,Q_q})^\top$ , respectively. Similarly, the estimates of  $\mathbf{M}_1^*$  and  $\mathbf{M}_2^*$  defined in (5) and (6) are  $\mathbf{m}_1^* = (b_{1,1,1}, \dots, b_{1,p,1})^\top$  and  $\mathbf{m}_2^* = (b_{2,1,1}, \dots, b_{2,p,1})^\top$ , respectively.

The quantities  $\mathbf{m}_{1,q}$  and  $\mathbf{m}_{2,q}$  provide information about the skewness and kurtosis present in all the  $q$ -dimensional sub-vectors constructed from the sample. On the other hand,  $\mathbf{m}_1^*$  and  $\mathbf{m}_2^*$  reflect the skewness and kurtosis present in all possible sub-dimensions of the sample. Based on these quantities, tests of skewness and kurtosis can be constructed.

##### 5.1. Testing skewness

From (2.4) in [27], it follows that  $\beta_{1,q,i} \geq 0$  for all  $q, i$ . However, for symmetric distributions,  $\beta_{1,q,i} = 0$  for all  $q, i$  as discussed in Section 2. So, if any of the sub-dimensions bears skewness, we shall have  $\beta_{1,q,i} > 0$  for the  $(q, i)$ -pair

corresponding to that sub-dimension, which implies that  $\max\{\beta_{1,q,i} | q \in \{1, \dots, p\}, i \in \{1, \dots, Q_q\}\} > 0$ . Therefore, a hypothesis for testing skewness in all sub-dimensions of the distribution can be formulated as follows:

$$H_0^{(s)} : \max_{q,i} \beta_{1,q,i} = 0, \quad H_A^{(s)} : \max_{q,i} \beta_{1,q,i} > 0. \tag{10}$$

The above hypothesis is equivalent to

$$\tilde{H}_0^{(s)} : \beta_{1,q,i} = 0 \text{ for all } q, i, \quad \tilde{H}_A^{(s)} : \beta_{1,q,i} > 0 \text{ for some } q, i.$$

The usual Mardia skewness test [27] only tests for

$$\bar{H}_0^{(s)} : \beta_{1,p} = 0, \quad \bar{H}_A^{(s)} : \beta_{1,p} > 0,$$

and thus it may be less efficient in providing information about skewness supported on a smaller sub-dimension. In this aspect, testing for (10) can be expected to be more efficient than the usual Mardia skewness test.

The null hypothesis  $H_0^{(s)}$  in (10) should be rejected when the maximum of  $b_{1,q,i}$  is large. However, directly comparing the  $b_{1,q,i}$ s is not proper, because they have different means and standard deviations under the null hypothesis. When the underlying distribution  $F$  is Gaussian, it follows from Equation (2.26) in [27] that the asymptotic expectation and asymptotic standard deviation of  $nb_{1,q,i}$  are  $q(q + 1)(q + 2)$  and  $\sqrt{12q(q + 1)(q + 2)}$ , respectively. So, while comparing the quantities  $b_{1,q,i}$ , it is appropriate to center and scale them first. Because our aim is to detect non-Gaussian features in the sample, we center and scale  $b_{1,q,i}$  using its asymptotic expectation and standard deviation under Gaussianity, and consider

$$\tilde{b}_{1,q,i} = \frac{nb_{1,q,i} - q(q + 1)(q + 2)}{\sqrt{12q(q + 1)(q + 2)}}, \quad q \in \{1, \dots, p\}, \quad i \in \{1, \dots, Q_q\}.$$

We reject  $H_0^{(s)}$  in (10) when  $\max_{q,i} \tilde{b}_{1,q,i}$  is large, and we shall denote this test as the MaxS test. Let  $\tilde{\mathbf{m}}_{1,q}$  and  $\tilde{\mathbf{m}}_1^*$  be the centered and scaled analogues of  $\mathbf{m}_{1,q}$  and  $\mathbf{m}_1^*$ , formed by replacing  $b_{1,q,i}$  by  $\tilde{b}_{1,q,i}$ . Note that  $\max_{q,i} \tilde{b}_{1,q,i}$  is the maximum of the coordinates of  $\tilde{\mathbf{m}}_1^*$ . To find the  $p$ -value of the MaxS test or to construct the cutoff for a given level, we need the asymptotic null distribution of  $\max_{q,i} \tilde{b}_{1,q,i}$ , which is derived in Section 6.

Sometimes, it may be the case that there is information about the possible presence of skewness in the sub-vectors of a fixed dimension, say,  $q_0$ , where  $q_0 < p$ , but it is not known exactly which subvector has a skewed distribution. Then, it would be judicious to construct the test based on  $\max_i \tilde{b}_{1,q_0,i}$  only. In such a situation, the hypothesis is

$$H_0^{q_0,(s)} : \max_i \beta_{1,q_0,i} = 0, \quad H_A^{q_0,(s)} : \max_i \beta_{1,q_0,i} > 0. \tag{11}$$

Here, the null hypothesis  $H_0^{q_0,(s)}$  in (11) is rejected for large values of  $\max_i \tilde{b}_{1,q_0,i}$ . We denote this test as the MaxS $_{q_0}$  test. To find the  $p$ -value (or, the cutoff) for rejection at a given level, we derive the asymptotic distribution of  $\max_i \tilde{b}_{1,q_0,i}$  in Section 6.

### 5.2. Testing kurtosis

Similar to tests of skewness, tests of kurtosis can be also constructed based on the  $b_{2,q,i}$  quantities. For a  $p$ -dimensional Gaussian distribution, the value of the Mardia measure of multivariate kurtosis  $\beta_{2,p}$  is  $p(p + 2)$  (see [27]). The quantity  $\beta_{2,p}$  measures how heavy-tailed the distribution is. For distributions with tails heavier than the Gaussian distribution, e.g., the multivariate Student's  $t$  distributions, we have  $\beta_{2,p} > p(p + 2)$ . Similarly, for distributions with lighter tails than the Gaussian distribution, we have  $\beta_{2,p} < p(p + 2)$ . However, the Mardia test of kurtosis [27] again checks the overall kurtosis of all the dimensions, and would not be efficient in checking whether the distribution in a particular sub-dimension deviates from Gaussianity in terms of kurtosis.

Suppose one wants to test whether the kurtosis of the distribution of any sub-dimension deviates from the Gaussian distribution. If this is the case, then we have  $|\beta_{2,q,i} - q(q + 2)| > 0$ , where the pair  $(q, i)$  corresponds to that sub-dimension. For this scenario, the suitable hypothesis is

$$H_0^{(k)} : \max_{q,i} |\beta_{2,q,i} - q(q + 2)| = 0, \quad H_A^{(k)} : \max_{q,i} |\beta_{2,q,i} - q(q + 2)| > 0. \tag{12}$$

Like in the case of  $b_{1,q,i}$ ,  $b_{2,q,i}$  has different expectation and standard deviation for different  $q$ . When the underlying distribution  $F$  is Gaussian, from the derivations in subsection 3.2 in [27], the asymptotic expectation and the asymptotic standard deviation of  $b_{2,q,i}$  are  $q(q + 2)$  and  $\sqrt{\{8q(q + 2)\}/n}$ . So, we center and scale  $b_{2,q,i}$ , and consider

$$\tilde{b}_{2,q,i} = \frac{b_{2,q,i} - q(q + 2)}{\sqrt{\{8q(q + 2)\}/n}}, \quad q \in \{1, \dots, p\}, \quad i \in \{1, \dots, Q_q\}.$$

The null hypothesis  $H_0^{(k)}$  in (12) is rejected if  $\max_{q,i} |\tilde{b}_{2,q,i}|$  is large. We denote this test as the MaxK test. To find the  $p$ -value (or the cutoff) of the test at a given level, we use the asymptotic null distribution of  $\max_{q,i} |\tilde{b}_{2,q,i}|$  derived in Section 6.

Now, suppose one knows that there is possible deviation from the Gaussian distribution in terms of kurtosis in some sub-dimension with dimension  $q_0 < p$ , but the exact sub-dimension is unknown. In such a case, the appropriate hypothesis is

$$H_0^{q_0, (k)} : \max_i |\beta_{2, q_0, i} - q_0(q_0 + 2)| = 0, \quad H_A^{q_0, (k)} : \max_i |\beta_{2, q_0, i} - q_0(q_0 + 2)| > 0. \tag{13}$$

Here, the null hypothesis  $H_0^{q_0, (k)}$  in (13) is rejected if  $\max_i |\tilde{b}_{2, q_0, i}|$  is large. This test is denoted as the  $\text{MaxK}_{q_0}$  test. We employ the asymptotic distribution of  $\max_i |\tilde{b}_{2, q_0, i}|$  derived in Section 6 to find the  $p$ -value (or the cutoff) of the test at a given level.

### 6. Asymptotic distributions and implementation of tests

In this section, the asymptotic distributions of the quantities  $\max_{q,i} \tilde{b}_{1, q, i}$ ,  $\max_i \tilde{b}_{1, q_0, i}$ ,  $\max_{q,i} |\tilde{b}_{2, q, i}|$  and  $\max_i |\tilde{b}_{2, q_0, i}|$  introduced in Section 5 are derived. Based on the respective asymptotic distributions, the implementations of the  $\text{MaxS}$ ,  $\text{MaxS}_{q_0}$ ,  $\text{MaxK}$  and  $\text{MaxK}_{q_0}$  tests corresponding to (10), (11), (12) and (13), respectively, are described. To derive the aforementioned asymptotic distributions, we shall use certain linearizations corresponding to the terms  $b_{1, q, i}$  and  $b_{2, q, i}$ .

#### 6.1. Skewness

Given any  $q, i$ , define

$$h_{qi}(\mathbf{x}, \mathbf{y}) = \{(\mathbf{x} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{qi})\}^3 - 3(\mathbf{x} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{qi})(\mathbf{x} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{qi}) - 3(\mathbf{y} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{qi})(\mathbf{x} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{qi}) + 3(q + 2)(\mathbf{x} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{qi}).$$

Define the integral operator  $\mathbf{h}_{qi}$  by

$$\{\mathbf{h}_{qi}(g)\}(\mathbf{x}) = E \{h_{qi}(\mathbf{x}, \mathbf{X}_{qi})g(\mathbf{X}_{qi})\}. \tag{14}$$

If  $E(\|\mathbf{X}\|^3) < \infty$ , then the integral operator  $\mathbf{h}_{qi}$  is well defined on  $L_2[\mathbf{X}_{qi}]$ , the space of measurable functions  $g$  which are square integrable with respect to the distribution of  $\mathbf{X}_{qi}$ . Under appropriate assumptions, it can be established that  $\mathbf{h}_{qi}$  has only finitely many nonzero eigenvalues, where the number  $K(q)$  of nonzero eigenvalues depends on  $q$  (see [4]). Let  $\lambda_{qik}$  be the eigenvalues and  $f_{qik}(\cdot)$  be the corresponding eigenfunctions of  $\mathbf{h}_{qi}$  for  $k \in \{1, \dots, K(q)\}$ . Then, we have for all  $\mathbf{x}, \mathbf{y}$ ,

$$h_{qi}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{K(q)} \lambda_{qik} f_{qik}(\mathbf{x}) f_{qik}(\mathbf{y}). \tag{15}$$

The function  $h_{qi}(\mathbf{x}, \mathbf{y})$  is closely related to the quantity  $b_{1, q, i}$ . The eigenvalues and eigenfunctions of  $h_{qi}(\mathbf{x}, \mathbf{y})$  are used to establish a relationship between  $b_{1, q, i}$  and the average of independent random vectors in the following theorem. This linearization will be used to derive the asymptotic distributions of the quantities  $\max_{q,i} \tilde{b}_{1, q, i}$  and  $\max_i \tilde{b}_{1, q_0, i}$ . Define

$$m_{4, qi} = E \left[ \{(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})\}^2 \right], \quad m_{6, qi} = E \left[ \{(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})\}^3 \right].$$

We have the following theorem on  $b_{1, q, i}$ . Note that when  $q = 1$ ,  $\mathbf{X}_{qi} = X_i$ , the  $i$ th component of  $\mathbf{X}$  for  $i \in \{1, \dots, p\}$ . Let  $\mu_i$  be the  $i$ th component of  $\boldsymbol{\mu}$  and  $X_{ij}$  be the  $i$ th component of the observation vector  $\mathbf{X}_j$  for  $i \in \{1, \dots, p\}$  and  $j \in \{1, \dots, n\}$ .

**Theorem 1.** Suppose that  $E(\|\mathbf{X}\|^6) < \infty$  and the distribution of  $\mathbf{X}$  is elliptical. Let  $\sigma_i^2 = E\{(X_i - \mu_i)^2\}$  for  $i \in \{1, \dots, p\}$ . For  $j \in \{1, \dots, n\}$ , define  $\mathbf{u}_{qij}$  as

$$\mathbf{u}_{qij} = \begin{cases} \sigma_i^{-3}(X_{ij} - \mu_i)\{(X_{ij} - \mu_i)^2 - 3\sigma_i^2\} & \text{if } q = 1, \\ (\sqrt{\lambda_{qi}} f_{qi1}(\mathbf{X}_{qij}), \dots, \sqrt{\lambda_{qiK(q)}} f_{qiK(q)}(\mathbf{X}_{qij}))^\top & \text{otherwise.} \end{cases}$$

Let  $\bar{\mathbf{u}}_{qi} = n^{-1} \sum_{j=1}^n \mathbf{u}_{qij}$ . Then

$$nb_{1, q, i} = \|\sqrt{n}\bar{\mathbf{u}}_{qi}\|_2^2 + o_p(1)$$

as  $n \rightarrow \infty$ . Further, for  $q > 1$ ,  $K(q) = q + q(q - 1)(q + 4)/6$ ,

$$\lambda_{qik} = \begin{cases} (3/q) \{m_{6, qi}/(q + 2) - 2m_{4, qi} + (q + 2)q\}, & k \in \{1, \dots, q\}, \\ 6m_{6, qi}/\{q(q + 2)(q + 4)\}, & k \in \{(q + 1), \dots, K(q)\}, \end{cases}$$

and  $E\{f_{qik}(\mathbf{X}_{qij})\} = 0$  for all  $q, i, j$  and  $k$ .



**Proof.** When  $q = 1$ , from the arguments in the proof of Theorem 1 in [15], it follows that, as  $n \rightarrow \infty$ ,

$$nb_{1,q,i} = \left[ n^{-1/2} \sum_{j=1}^n \sigma_i^{-3} (X_{ij} - \mu_i) \{ (X_{ij} - \mu_i)^2 - 3\sigma_i^2 \} \right]^2 + o_p(1) = (\sqrt{n}\bar{\mathbf{u}}_{qi})^2 + o_p(1).$$

For  $q > 1$ , let  $\mathbf{Y}_{qij} = \Sigma_{qi}^{-1/2} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})$  for all  $i, j$ . Then  $\mathbf{Y}_{qij}$  has an elliptical distribution, which is identical over  $j$  for fixed  $i$ , with  $E(\mathbf{Y}_{qij}) = \mathbf{0}$  and  $E(\mathbf{Y}_{qij}\mathbf{Y}_{qij}^\top) = \mathbf{I}_q$ . Also,  $E(\|\mathbf{Y}_{qij}\|^6) < \infty$  for all  $i, j$ , and hence  $\mathbf{Y}_{qij}$ s satisfy the conditions of Lemma 2.1 in [4]. From an application of this lemma, we have

$$nb_{1,q,i} = n^{-1} \sum_{j=1}^n \sum_{k=1}^n h_{qi}(\mathbf{X}_{qij}, \mathbf{X}_{qik}) + o_p(1) \tag{16}$$

as  $n \rightarrow \infty$ . Next, from the arguments in the proof of Theorem 2.2 in [4], it follows that the integral operator  $\mathbf{h}_{qi}$  defined in (14) has only two non-zero distinct eigenvalues, which are

$$\gamma_{qi1} = (3/q) \{ m_{6,qi}/(q+2) - 2m_{4,qi} + (q+2)q \}, \quad \gamma_{qi2} = 6m_{6,qi} / \{ q(q+2)(q+4) \},$$

with associated multiplicities  $\nu_{qi1} = q$  and  $\nu_{qi2} = q(q-1)(q+4)/6$ , respectively. So, we can take  $\lambda_{qik} = \gamma_{qi1}$  for  $k \in \{1, \dots, \nu_{qi1}\}$  and  $\lambda_{qik} = \gamma_{qi2}$  for  $k \in \{\nu_{qi1} + 1, \dots, (\nu_{qi1} + \nu_{qi2})\}$ . Consequently,  $K(q) = q + q(q-1)(q+4)/6$ . From (15) and (16), we get

$$\begin{aligned} nb_{1,q,i} &= n^{-1} \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^{K(q)} \lambda_{qik} f_{qik}(\mathbf{X}_{qij}) f_{qik}(\mathbf{X}_{qil}) + o_p(1) = \sum_{k=1}^{K(q)} \left\{ n^{-1/2} \sum_{j=1}^n \sqrt{\lambda_{qik}} f_{qik}(\mathbf{X}_{qij}) \right\}^2 + o_p(1) \\ &= \|\sqrt{n}\bar{\mathbf{u}}_{qi}\|_2^2 + o_p(1) \end{aligned}$$

as  $n \rightarrow \infty$ . Finally, from the arguments in the proof of Theorem 2.1 in [18], it follows that  $E\{f_{qik}(\mathbf{X}_{qij})\} = 0$  for all  $q, i, j$  and  $k$ .  $\square$

Define  $\mathbf{U}_j = (\mathbf{u}_{11j}^\top, \dots, \mathbf{u}_{qj}^\top, \dots, \mathbf{u}_{pj}^\top)^\top$  and  $\mathbf{U}_{qj} = (\mathbf{u}_{q1j}^\top, \dots, \mathbf{u}_{qQj}^\top)^\top$ . Let  $G(\cdot)$  and  $G_q(\cdot)$  be such that

$$G(\mathbf{U}_j) = \max \left\{ \frac{\|\mathbf{u}_{11j}\|_2^2 - 6}{\sqrt{72}}, \dots, \frac{\|\mathbf{u}_{qj}\|_2^2 - q(q+1)(q+2)}{\sqrt{12q(q+1)(q+2)}}, \dots, \frac{\|\mathbf{u}_{pj}\|_2^2 - p(p+1)(p+2)}{\sqrt{12p(p+1)(p+2)}} \right\},$$

and

$$G_q(\mathbf{U}_{qj}) = \max \left\{ \frac{\|\mathbf{u}_{q1j}\|_2^2 - q(q+1)(q+2)}{\sqrt{12q(q+1)(q+2)}}, \dots, \frac{\|\mathbf{u}_{qQj}\|_2^2 - q(q+1)(q+2)}{\sqrt{12q(q+1)(q+2)}} \right\}.$$

Clearly,  $G(\cdot)$  and  $G_q(\cdot)$  are continuous functions. From these observations, we derive the asymptotic null distributions of the test statistics  $\max_{q,i} \tilde{b}_{1,q,i}$  for the MaxS test and  $\max_i \tilde{b}_{1,q_0,i}$  for the MaxS<sub>q0</sub> test in the next theorem.

**Theorem 2.** Let  $\Omega$  and  $\Omega_q$  be the dispersion matrices of  $\mathbf{U}_1$  and  $\mathbf{U}_{q1}$ , respectively. Let  $\mathbf{W}$  and  $\mathbf{W}_{q_0}$  be zero-mean Gaussian random vectors with dispersion matrices  $\Omega$  and  $\Omega_{q_0}$ , respectively. Assume  $E(\|\mathbf{X}\|^6) < \infty$  and the distribution of  $\mathbf{X}$  is elliptical. Then,  $\max_{q,i} n\tilde{b}_{1,q,i} \xrightarrow{d} G(\mathbf{W})$  and  $\max_i n\tilde{b}_{1,q_0,i} \xrightarrow{d} G_{q_0}(\mathbf{W}_{q_0})$  as  $n \rightarrow \infty$ .

**Proof.** Since we have  $E\{f_{qik}(\mathbf{X}_{qij})\} = 0$  for all  $q, i, j$  and  $k$  from Theorem 1, it follows that  $E(\mathbf{U}_{qj}) = \mathbf{0}$  and  $E(\mathbf{U}_j) = \mathbf{0}$  for all  $j$ . Further, the distributions of  $\mathbf{U}_{qj}$  are independent and identical for all  $j$ , and the same is true for the distributions of  $\mathbf{U}_j$ . Define  $\bar{\mathbf{U}} = n^{-1} \sum_{j=1}^n \mathbf{U}_j$  and  $\bar{\mathbf{U}}_q = n^{-1} \sum_{j=1}^n \mathbf{U}_{qj}$ . It follows from the multivariate central limit theorem that  $\sqrt{n}\bar{\mathbf{U}} \xrightarrow{d} \mathbf{W}$  and  $\sqrt{n}\bar{\mathbf{U}}_{q_0} \xrightarrow{d} \mathbf{W}_{q_0}$  as  $n \rightarrow \infty$ . Now, note that  $\max_{q,i} n\tilde{b}_{1,q,i} = G(\sqrt{n}\bar{\mathbf{U}})$  and  $\max_i n\tilde{b}_{1,q_0,i} = G_{q_0}(\sqrt{n}\bar{\mathbf{U}}_{q_0})$ . Since  $G(\cdot)$  and  $G_q(\cdot)$  are continuous, the proof of the theorem is completed from an application of the continuous mapping theorem.  $\square$

We implement the tests of skewness under Gaussianity of the null hypotheses. To derive the p-values of the tests of the hypotheses described in (10) and (11), we need to first estimate  $\Omega$  and  $\Omega_{q_0}$ , the dispersion matrices of  $\mathbf{U}_j$  and  $\mathbf{U}_{q_0j}$ , respectively. However, the random vectors  $\mathbf{U}_j$  and  $\mathbf{U}_{q_0j}$  are constituted of  $\mathbf{u}_{qij}$ , whose definition involve unknown population quantities  $\sigma_i, \mu_i, \lambda_{qik}$  and  $f_{qik}(\mathbf{X}_{qij})$ . So, we substitute  $\mathbf{u}_{qij}$  in  $\mathbf{U}_j$  and  $\mathbf{U}_{q_0j}$  by  $\hat{\mathbf{u}}_{qij}$  for all  $q, i$  and  $j$  to form  $\hat{\mathbf{U}}_j$  and  $\hat{\mathbf{U}}_{q_0j}$ , respectively, where  $\hat{\mathbf{u}}_{qij}$  is constructed based on the sample observations. The construction procedure of  $\hat{\mathbf{u}}_{qij}$  is described below.

Let  $\bar{X}_i$  and  $s_i^2$  denote the sample mean and sample variance, respectively, of the  $i$ th components  $X_{ij}$  of the observation vectors  $\mathbf{X}_j$ , where  $j \in \{1, \dots, n\}$  and  $i \in \{1, \dots, p\}$ . When  $q = 1$ ,  $\mathbf{u}_{qij}$  is a univariate random variable, and it can be verified that under Gaussianity,  $\text{Var}(\mathbf{u}_{qij}) = 6$  for all  $i$  and  $j$ . For  $q = 1$ , we construct  $\hat{\mathbf{u}}_{qij}$  ensuring that its sample variance is also

equal to 6. We first compute  $\tilde{u}_{ij} = s_i^{-3}(X_{ij} - \bar{X}_i)\{(X_{ij} - \bar{X}_i)^2 - 3s_i^2\}$  for all  $j$ , and then set  $\hat{\mathbf{u}}_{qij} = \{\sqrt{6}/\tilde{s}_i\} \tilde{u}_{ij}$ , where  $\tilde{s}_i^2$  denotes the sample variance of  $\tilde{u}_{ij}$ . In this way, it is ensured that the sample variance of  $\hat{\mathbf{u}}_{qij}$  is equal to  $\text{Var}(\mathbf{u}_{qij})$  under Gaussianity for all  $i$  and  $j$  when  $q = 1$ . Next, for  $q > 1$ ,  $\lambda_{qik}$  is derived under Gaussianity, and it can be verified that  $\lambda_{qik} = 6$  for all  $q, i, k$  in this case. Recall that the function  $f_{qik}(\cdot)$  is the eigenfunction of the integral operator  $\mathbf{h}_{qi}$  corresponding to the eigenvalue  $\lambda_{qik}$ . From the arguments in Theorem 2.2 in [4], it follows that  $f_{qik}(\cdot)$  are spherical harmonic functions (see [11]). Explicit expressions of these spherical harmonic functions can be derived (see [5,21]). However, their numerical approximation may be unstable due to the involvement of hypergeometric functions (see [21, p. 1554]). For this reason and ease of computation, the random variables  $f_{qik}(\mathbf{X}_{qij})$  in the definition of  $\mathbf{u}_{qij}$  are substituted in the following way. First, we form the  $n \times n$  matrix  $\hat{\mathbf{H}}_{qi} = (\hat{h}_{qi}(\mathbf{X}_{qij}, \mathbf{X}_{qil}))$ , where

$$\begin{aligned} \hat{h}_{qi}(\mathbf{X}_{qij}, \mathbf{X}_{qil}) &= \{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi})\}^3 + 3(q+2)\{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi})\} \\ &\quad - 3(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi}) \\ &\quad - 3(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi})(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1}(\mathbf{X}_{qil} - \bar{\mathbf{X}}_{qi}). \end{aligned}$$

We then compute the  $K(q)$  eigenvectors of the matrix  $\hat{\mathbf{H}}_{qi}$  corresponding to the  $K(q)$  eigenvalues with the largest magnitudes, and arrange them by the descending order of the magnitudes of their corresponding eigenvalues. Each computed eigenvector is multiplied by  $\sqrt{n}$  to maintain its correspondence with the eigenfunctions of the integral operator  $\mathbf{h}_{qi}$ . Each such vector obtained, denoted as  $\hat{f}_{qik}$ , substitutes the vector  $(f_{qik}(\mathbf{X}_{qi1}), \dots, f_{qik}(\mathbf{X}_{qin}))^\top$ . From the components of the vectors  $\hat{f}_{qik}$ , we compute  $\hat{\mathbf{u}}_{qij}$  analogous to how  $\mathbf{u}_{qij}$  is defined using the components of the vectors  $(f_{qik}(\mathbf{X}_{qi1}), \dots, f_{qik}(\mathbf{X}_{qin}))^\top$  and take  $\lambda_{qik} = 6$ . Then, we construct the vectors  $\hat{\mathbf{U}}_j$  and  $\hat{\mathbf{U}}_{q0j}$  using  $\hat{\mathbf{u}}_{qij}$  in place of  $\mathbf{u}_{qij}$  in  $\mathbf{U}_j$  and  $\mathbf{U}_{q0j}$ , respectively. Finally, the dispersion matrices  $\Omega$  and  $\Omega_{q0}$  are estimated by

$$\begin{aligned} \hat{\Omega} &= (n-1)^{-1} \sum_{j=1}^n \left( \hat{\mathbf{U}}_j - n^{-1} \sum_{l=1}^n \hat{\mathbf{U}}_l \right) \left( \hat{\mathbf{U}}_j - n^{-1} \sum_{l=1}^n \hat{\mathbf{U}}_l \right)^\top, \\ \hat{\Omega}_{q0} &= (n-1)^{-1} \sum_{j=1}^n \left( \hat{\mathbf{U}}_{q0j} - n^{-1} \sum_{l=1}^n \hat{\mathbf{U}}_{q0l} \right) \left( \hat{\mathbf{U}}_{q0j} - n^{-1} \sum_{l=1}^n \hat{\mathbf{U}}_{q0l} \right)^\top. \end{aligned}$$

To compute the  $p$ -values of the MaxS test for (10), we generate 1000 independent zero-mean Gaussian random vectors  $\tilde{\mathbf{W}}_1, \dots, \tilde{\mathbf{W}}_{1000}$  with the dispersion matrix  $\hat{\Omega}$ . The proportion of the values  $G(\tilde{\mathbf{W}}_1), \dots, G(\tilde{\mathbf{W}}_{1000})$  larger than  $\max_{q,i} n \tilde{b}_{1,q,i}$  is taken as the  $p$ -value of the null hypothesis in (10). The  $p$ -value of the test  $\text{MaxS}_{q0}$  in (11) is derived similarly.

### 6.2. Kurtosis

Next, we derive the asymptotic null distributions of  $\max_{q,i} |\tilde{b}_{2,q,i}|$  and of  $\max_i |\tilde{b}_{2,q0,i}|$ . Here also, we first derive a linearization of  $b_{2,q,i}$ .

**Theorem 3.** Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent and identically distributed random vectors in  $\mathbb{R}^p$  with  $E(\|\mathbf{X}\|^8) < \infty$ . Let  $\mathbf{A}_{qi} = \Sigma_{qi}^{-1} E\{(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1}(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})(\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top\}$ . Define

$$\mathbf{Z}_{qij} = \begin{bmatrix} \left\{ (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi}) \right\}^2 - E \left[ \left\{ (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi}) \right\}^2 \right] \\ (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi}) - E \left\{ (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi}) \right\} \end{bmatrix},$$

and

$$\mathbf{a}_{qi} = \left( 1, -2, -4E \left[ \left\{ (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi}) \right\} \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi}) \right] \right)^\top.$$

Then,  $n^{1/2}(b_{2,q,i} - \beta_{2,q,i}) = n^{-1/2} \sum_{j=1}^n \mathbf{a}_{qi}^\top \mathbf{Z}_{qij} + o_p(1)$  as  $n \rightarrow \infty$ .

**Proof.** Recall that  $\beta_{2,q,i}$  and  $b_{2,q,i}$  are invariant under location transformations. So, without loss of generality, we can assume  $\boldsymbol{\mu} = E(\mathbf{X}) = \mathbf{0}$ , which means  $\boldsymbol{\mu}_{qi} = \mathbf{0}$  and  $\text{Var}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^\top) = \Sigma$ .

The arguments are similar to those in the proof of Theorem 2.1 in [19]. In this proof,  $\mathbf{Y}_n = O_p(b_n)$  means that the sequence  $\mathbf{Y}_n/b_n$  is bounded in Euclidean/matrix norm, while  $\mathbf{Y}_n = o_p(b_n)$  means that  $b_n^{-1} \mathbf{Y}_n \xrightarrow{p} \mathbf{0}$ . Let

$$\mathbf{B}_{qi} = n^{1/2} \left\{ n^{-1} \sum_{j=1}^n (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})(\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top - \Sigma_{qi} \right\}.$$

From the weak law of large numbers, we get  $\mathbf{B}_{qi} = O_p(1)$  for all  $q, i$ . Also,

$$n^{1/2} (\mathbf{S}_{qi} - \Sigma_{qi}) = \mathbf{B}_{qi} - n^{1/2} (\bar{\mathbf{X}}_{qi} - \boldsymbol{\mu}_{qi}) (\bar{\mathbf{X}}_{qi} - \boldsymbol{\mu}_{qi})^\top. \tag{17}$$

From the multivariate central limit theorem, we have that  $n^{1/2} (\bar{\mathbf{X}}_{qi} - \boldsymbol{\mu}_{qi}) (\bar{\mathbf{X}}_{qi} - \boldsymbol{\mu}_{qi})^\top = O_p(n^{-1/2})$ , which, applied to (17), yields

$$\mathbf{S}_{qi} = \Sigma_{qi} + n^{-1/2} \mathbf{B}_{qi} + O_p(n^{-1}), \text{ or, } \Sigma_{qi}^{-1} \mathbf{S}_{qi} = \mathbf{I}_q + n^{-1/2} \Sigma_{qi}^{-1} \mathbf{B}_{qi} + O_p(n^{-1}).$$

Note that  $n^{1/2} (\mathbf{S}_{qi} - \Sigma_{qi}) = \mathbf{B}_{qi} - n^{1/2} \bar{\mathbf{X}}_{qi} \bar{\mathbf{X}}_{qi}^\top$  and  $n^{1/2} \bar{\mathbf{X}}_{qi} \bar{\mathbf{X}}_{qi}^\top = O_p(n^{-1/2})$  by the multivariate central limit theorem. It follows that

$$\mathbf{S}_{qi} = \Sigma_{qi} + n^{-1/2} \mathbf{B}_{qi} + O_p(n^{-1}),$$

and thus

$$\Sigma_{qi}^{-1} \mathbf{S}_{qi} = \mathbf{I}_q + n^{-1/2} \Sigma_{qi}^{-1} \mathbf{B}_{qi} + O_p(n^{-1}).$$

This means  $\Sigma_{qi}^{-1} \mathbf{S}_{qi}$  is invertible for all sufficiently large  $n$  with probability approaching 1, and we have

$$(\Sigma_{qi}^{-1} \mathbf{S}_{qi})^{-1} = \mathbf{I}_q - n^{-1/2} \Sigma_{qi}^{-1} \mathbf{B}_{qi} + O_p(n^{-1}),$$

which implies that

$$\mathbf{S}_{qi}^{-1} = \Sigma_{qi}^{-1} - n^{-1/2} \Sigma_{qi}^{-1} \mathbf{B}_{qi} \Sigma_{qi}^{-1} + O_p(n^{-1}). \tag{18}$$

Now,

$$\begin{aligned} \{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1} (\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})\}^2 &= (\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij})^2 + 4(\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij})^2 + (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi})^2 \\ &\quad - 4(\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij} \bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij}) + 2(\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij} \bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi}) \\ &\quad - 4(\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij} \bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi}). \end{aligned} \tag{19}$$

Using this and the expression for the inverse in (18) above, we get for the first term in (19) to be

$$\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij} = \mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij} - n^{-1/2} \mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{B}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij} + O_p(n^{-1}).$$

Squaring both sides, we get

$$(\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij})^2 = (\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij})^2 - 2n^{-1/2} \mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{B}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{B}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij} + O_p(n^{-1}).$$

Using the fact that the trace of a matrix is invariant under cyclic permutations, we have

$$\text{tr}(\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{B}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij}) = \text{tr}\{\Sigma_{qi}^{-1} \mathbf{B}_{qi} (\Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top) (\Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top)\}.$$

Since  $n^{-1} \sum_{j=1}^n (\Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top) (\Sigma_{qi}^{-1} \mathbf{X}_{qij} \mathbf{X}_{qij}^\top) = \mathbf{A}_{qi} + o_p(1)$  with  $\mathbf{A}_{qi} = E\{(\Sigma_{qi}^{-1} \mathbf{X}_{qi} \mathbf{X}_{qi}^\top) (\Sigma_{qi}^{-1} \mathbf{X}_{qi} \mathbf{X}_{qi}^\top)\}$ , we now get

$$\frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij})^2 = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij})^2 - 2n^{-1/2} \text{tr}(\Sigma_{qi}^{-1} \mathbf{B}_{qi} \mathbf{A}_{qi}) + o_p(n^{-1}).$$

For the second term in (19), again using Eq. (18) and the fact that  $n^{-1} \sum_{j=1}^n \mathbf{X}_{qij} \mathbf{X}_{qij}^\top = O_p(1)$  (using the weak law of large numbers) we obtain

$$\frac{1}{n} \sum_{j=1}^n (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij})^2 = O_p(n^{-1/2}).$$

Furthermore, for the third term in (19),  $(\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi})^2 = O_p(n^{-2})$ . Since,  $n^{-1} \sum_{j=1}^n \Sigma_{qi}^{-1} \mathbf{X}_{qij} (\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij}) = \tilde{\mathbf{a}}_{qi} + o_p(1)$  with  $\tilde{\mathbf{a}}_{qi} = E\{\Sigma_{qi}^{-1} \mathbf{X}_{qi} (\mathbf{X}_{qi}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qi})\}$ , it is easy to see that for the fourth term,

$$\frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij}) (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij}) = \tilde{\mathbf{a}}_{qi}^\top \bar{\mathbf{X}}_{qi} + O_p(n^{-1}).$$

Finally, for the last two terms,

$$\frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{qij}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij}) (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi}) = O_p(n^{-1}), \quad \frac{1}{n} \sum_{j=1}^n (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \mathbf{X}_{qij}) (\bar{\mathbf{X}}_{qi}^\top \mathbf{S}_{qi}^{-1} \bar{\mathbf{X}}_{qi}) = O_p(n^{-1}).$$

Summarizing, we obtain

$$b_{2,q,i} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij})^2 - 2n^{-1/2} \text{tr}(\Sigma_{qi}^{-1} \mathbf{B}_{qi} \mathbf{A}_{qi}) - 4\tilde{\mathbf{a}}_{qi}^\top \bar{\mathbf{X}}_{qi} + O_p(n^{-1}).$$

Note that

$$\text{tr}(\Sigma_{qi}^{-1} \mathbf{B}_{qi} \mathbf{A}_{qi}) = \text{tr}(\mathbf{B}_{qi} \mathbf{A}_{qi} \Sigma_{qi}^{-1}) = n^{1/2} \left\{ \frac{1}{n} \sum_{j=1}^n \mathbf{X}_{qij}^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij} - E(\mathbf{X}_{qi}^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qi}) \right\},$$

and using this we obtain

$$n^{1/2}(b_{2,q,i} - \beta_{2,q,i}) = n^{-1/2} \sum_{j=1}^n \mathbf{a}_{qi}^\top \mathbf{Z}_{qij} + o_p(1),$$

where

$$\mathbf{Z}_{qij} = \begin{bmatrix} (\mathbf{X}_{qij}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qij})^2 - E\{(\mathbf{X}_{qi}^\top \Sigma_{qi}^{-1} \mathbf{X}_{qi})^2\} \\ \mathbf{X}_{qij}^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qij} - E(\mathbf{X}_{qi}^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} \mathbf{X}_{qi}) \\ \mathbf{X}_{qij} \end{bmatrix}$$

is a  $(2 + q)$ -dimensional vector and  $\mathbf{a}_{qi} = (1, -2, -4\tilde{\mathbf{a}}_{qi}^\top)^\top$ .  $\square$

Tests of kurtosis are usually conducted for testing normality (see [19]). When the underlying distribution is Gaussian, we can simplify the quantities  $\mathbf{A}_{qi}$  and  $\mathbf{a}_{qi}$  described in Theorem 3, and consequently, the linearization becomes simpler.

**Corollary 1.** Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent and identically distributed Gaussian random vectors in  $\mathbb{R}^p$ . Define  $Y_{qij} = \{(\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})\}^2 - 2(q+2)(\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})$ . Then,  $n^{1/2}(b_{2,q,i} - \beta_{2,q,i}) = n^{-1/2} \sum_{j=1}^n \{Y_{qij} - E(Y_{qij})\} + o_p(1)$  as  $n \rightarrow \infty$ .

**Proof.** Define  $\tilde{\mathbf{X}}_{qij} = \Sigma_{qi}^{-1/2} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})$ . Under the assumption of the corollary, the  $\tilde{\mathbf{X}}_{qij}$ s are independent zero-mean Gaussian random vectors with the identity matrix as their dispersion matrix. It follows that

$$E \left[ \left\{ (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi})^\top \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi}) \right\} \Sigma_{qi}^{-1} (\mathbf{X}_{qi} - \boldsymbol{\mu}_{qi}) \right] = \Sigma_{qi}^{-1/2} E \left\{ \left( \tilde{\mathbf{X}}_{qi1}^\top \tilde{\mathbf{X}}_{qi1} \right) \tilde{\mathbf{X}}_{qi1} \right\} = \Sigma_{qi}^{-1/2} \times \mathbf{0} = \mathbf{0}. \tag{20}$$

Next, we have

$$(\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi})^\top \mathbf{A}_{qi} \Sigma_{qi}^{-1} (\mathbf{X}_{qij} - \boldsymbol{\mu}_{qi}) = \tilde{\mathbf{X}}_{qij}^\top E \left[ \tilde{\mathbf{X}}_{qij} \left( \tilde{\mathbf{X}}_{qij}^\top \tilde{\mathbf{X}}_{qij} \right) \tilde{\mathbf{X}}_{qij} \right] \tilde{\mathbf{X}}_{qij} = (q+2) \tilde{\mathbf{X}}_{qij}^\top \tilde{\mathbf{X}}_{qij}. \tag{21}$$

The proof follows from (20), (21) and the linearization in Theorem 3.  $\square$

The asymptotic null distributions of  $\max_{q,i} |\tilde{b}_{2,q,i}|$  and  $\max_i |\tilde{b}_{2,q_0,i}|$  are derived from the linearization in Corollary 1.

**Theorem 4.** Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent and identically distributed Gaussian random vectors in  $\mathbb{R}^p$ . Define

$$\tilde{\mathbf{Y}}_j = \left( \frac{Y_{11j}}{\sqrt{24}}, \dots, \frac{Y_{qij}}{\sqrt{8q(q+2)}}, \dots, \frac{Y_{p1j}}{\sqrt{8p(p+2)}} \right)^\top, \quad \tilde{\mathbf{Y}}_{qj} = \left( \frac{Y_{q1j}}{\sqrt{8q(q+2)}}, \dots, \frac{Y_{qQaj}}{\sqrt{8q(q+2)}} \right)^\top,$$

where  $Y_{qij}$  is as defined in Corollary 1. Let  $\Gamma$  and  $\Gamma_q$  be the dispersion matrices of  $\tilde{\mathbf{Y}}_1$  and  $\tilde{\mathbf{Y}}_{q1}$ , respectively. Let  $\mathbf{W}$  and  $\mathbf{W}_q$  be zero-mean Gaussian random vectors with dispersion matrices  $\Gamma$  and  $\Gamma_q$ , respectively. Then,  $\sqrt{n} \max_{q,i} |\tilde{b}_{2,q,i}| \xrightarrow{d} \|\mathbf{W}\|_\infty$  and  $\sqrt{n} \max_i |\tilde{b}_{2,q_0,i}| \xrightarrow{d} \|\mathbf{W}_q\|_\infty$  as  $n \rightarrow \infty$ , respectively, where  $\|\cdot\|_\infty$  is the  $l_\infty$  norm in the Euclidean space, i.e.,  $\|\mathbf{v}\|_\infty = \max_i |v_i|$  with  $v_i$  being the  $i$ th component of the vector  $\mathbf{v}$ .

**Proof.** Under the assumption of Gaussianity in the theorem,  $\beta_{2,q,i} = q(q+2)$  for all  $q$ . Since the  $l_\infty$  norm in the Euclidean space is continuous, the proof follows from an application of the multivariate central limit theorem in Corollary 1 and then applying the continuous mapping theorem.  $\square$

The tests of kurtosis are implemented under Gaussianity of the null hypotheses. To compute the p-values for the MaxK test in (12) and the MaxK<sub>q0</sub> test in (13), we need to estimate the dispersion matrices  $\Gamma$  and  $\Gamma_q$  of the random vectors  $\tilde{\mathbf{Y}}_j$  and  $\tilde{\mathbf{Y}}_{qj}$ . But,  $\tilde{\mathbf{Y}}_j$  and  $\tilde{\mathbf{Y}}_{qj}$  are constituted of  $Y_{qij}$ , whose definition involves unknown population quantities  $\boldsymbol{\mu}_{qi}$  and  $\Sigma_{qi}$ . So, we substitute  $Y_{qij}$  in  $\tilde{\mathbf{Y}}_j$  and  $\tilde{\mathbf{Y}}_{qj}$  by

$$\hat{Y}_{qij} = \{(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1} (\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})\}^2 - 2(q+2)(\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi})^\top \mathbf{S}_{qi}^{-1} (\mathbf{X}_{qij} - \bar{\mathbf{X}}_{qi}),$$

**Table 2**

Estimated sizes of the tests defined in Sections 5.1, 5.2 and 6.3 based on 1000 replicates at 5% nominal level for the underlying distribution being 5-variate Gaussian  $\mathcal{N}_5(\mathbf{0}, \Sigma)$ , where  $\Sigma = (\sigma_{ij})$  with  $\sigma_{ij} = 0.5 + 0.5\mathbb{I}(i = j)$ .

Test	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
MaxS	0.048	0.056	0.048	0.047	0.044
MaxK	0.043	0.049	0.058	0.060	0.049
MaxSK	0.044	0.053	0.057	0.055	0.045
MaxS <sub>1</sub>	0.041	0.053	0.043	0.051	0.046
MaxS <sub>2</sub>	0.050	0.062	0.051	0.055	0.040
MaxS <sub>3</sub>	0.034	0.059	0.055	0.055	0.035
MaxS <sub>4</sub>	0.030	0.050	0.043	0.055	0.038
MaxS <sub>5</sub>	0.015	0.041	0.043	0.045	0.048
MaxK <sub>1</sub>	0.047	0.062	0.053	0.058	0.048
MaxK <sub>2</sub>	0.021	0.034	0.045	0.057	0.050
MaxK <sub>3</sub>	0.014	0.018	0.030	0.049	0.051
MaxK <sub>4</sub>	0.031	0.038	0.036	0.059	0.045
MaxK <sub>5</sub>	0.151	0.088	0.060	0.069	0.050
MaxSK <sub>1</sub>	0.048	0.065	0.053	0.049	0.048
MaxSK <sub>2</sub>	0.033	0.049	0.049	0.049	0.050
MaxSK <sub>3</sub>	0.025	0.039	0.039	0.052	0.044
MaxSK <sub>4</sub>	0.022	0.035	0.035	0.055	0.039
MaxSK <sub>5</sub>	0.070	0.058	0.046	0.057	0.041

which is obtained by replacing  $\mu_{qi}$  and  $\Sigma_{qi}$  in the expression of  $Y_{qij}$  by their estimates  $\bar{X}_{qi}$  and  $S_{qi}$ , respectively. Now, it can be derived that under Gaussianity,  $\text{Var}(Y_{qi1}) = 8q(q + 2)$  for all  $q, i$ . So, the diagonal entries of the estimates of  $\Gamma$  and  $\Gamma_q$ , denoted as  $\hat{\Gamma}$  and  $\hat{\Gamma}_q$ , respectively, are fixed to be 1. The off-diagonal entries of  $\hat{\Gamma}$  and  $\hat{\Gamma}_q$  are obtained from the sample correlations of  $\hat{Y}_{q_1 i_1 j}$  and  $\hat{Y}_{q_2 i_2 j}$  for the corresponding indices  $q_1, i_1$  and  $q_2, i_2$ . Next, we generate 1000 independent zero mean Gaussian random vectors  $\tilde{\mathbf{W}}_1, \dots, \tilde{\mathbf{W}}_{1000}$  with dispersion matrix  $\hat{\Gamma}$ . The  $p$ -value of the MaxK test in (12) is the proportion of the values  $\|\tilde{\mathbf{W}}_1\|_\infty, \dots, \|\tilde{\mathbf{W}}_{1000}\|_\infty$  larger than  $\max_{q,i} |\tilde{b}_{2,q,i}|$ . The  $p$ -value of the MaxK<sub>q<sub>0</sub></sub> test in (13) is computed similarly.

6.3. Testing Gaussianity based on both skewness and kurtosis

Based on the tests of skewness and kurtosis, a test of Gaussianity can be constructed. Analogous to (10) and (12), the null hypothesis here is

$$H_0^{(g)} : \text{The underlying distribution is Gaussian.} \tag{22}$$

Similarly, analogous to (11) and (13), the null hypothesis is

$$H_0^{q_0 \cdot (g)} : \text{All } q_0\text{-dimensional subsets of the data follow some Gaussian distribution.} \tag{23}$$

Here, if at least one of the null hypotheses of the corresponding skewness test or kurtosis test is rejected, then we reject the null hypothesis of Gaussianity (after Bonferroni correction). For example, if any of the MaxS test and the MaxK test rejects their null hypotheses, then (22) is also rejected. The test for (22) is denoted as MaxSK test. Similarly, if any of the MaxS<sub>q<sub>0</sub></sub> test or the MaxK<sub>q<sub>0</sub></sub> test rejects their null hypothesis for a fixed  $q_0$ , then (23) is rejected, and we denote this test by MaxSK<sub>q<sub>0</sub></sub> test.

7. Simulation study

In this section, the performance of our proposed tests are investigated in terms of the estimated sizes and powers using some simulated models. The estimated powers of our tests are compared with the corresponding Mardia tests. We also compare the estimated powers of ours tests with several tests of Gaussianity.

7.1. Estimated sizes

For estimating the sizes of our tests, we consider  $\mathbf{X}_1, \dots, \mathbf{X}_n$  being a random sample from the  $p$ -variate Gaussian distribution  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ , where  $\Sigma = (\sigma_{ij})$  with  $\sigma_{ij} = 0.5 + 0.5\mathbb{I}(i = j)$ . We take  $p = 5$ . The number of replicates to estimate the sizes of our tests is taken as 1000. The sample size  $n$  is varied.

From Table 2, it can be seen that the tests MaxS, MaxK and MaxSK have estimated sizes close to the 5% nominal level for  $n = 50$  and up. The estimated sizes of MaxS<sub>5</sub> and MaxK<sub>2</sub>, MaxK<sub>3</sub>, MaxK<sub>5</sub> and MaxSK<sub>3</sub>, MaxSK<sub>4</sub> deviate slightly from the 5% nominal level for  $n = 50$  and 100, and require higher sample sizes to converge to the nominal level.

In section 1 of the supplementary material, estimated sizes for some other values of the dimension  $p$  are presented, namely  $p = 3, 4$  which correspond to the dimensions of the two datasets analyzed in Section 8. There also, it is found that

the estimated sizes of the majority of the tests, including the tests MaxS, MaxK and MaxSK, are close to the 5% nominal level for  $n = 50$  and up. A few tests, like  $\text{MaxK}_2$ ,  $\text{MaxK}_3$ ,  $\text{MaxSK}_2$ ,  $\text{MaxSK}_3$ , require higher sample sizes for their estimated sizes to reach close to the nominal level. In section 3 of the supplementary material, the estimated sizes of the tests are presented in an equicorrelation model, where the pairwise-correlation between the components of the vector  $\mathbf{X}$  is high. There, it is found that the estimated sizes of nearly all of the tests, including the tests MaxS, MaxK and MaxSK, are close to the nominal level for  $n = 100$  and up. When the sample size is lower, i.e.,  $n = 50$ , some deviations of the estimated sizes from the nominal level are observed.

### 7.2. Estimated powers

We now compare the powers of the tests with the usual Mardia skewness (MS) and kurtosis (MK) tests along with several tests of normality; see [8] for a recent review.

The following tests of Gaussianity are considered for the comparison of performances. The test by Henze and Zirkler [20] is denoted as the HZ test. The test of normality developed by Royston [34–37] is denoted as the R test. The testing procedure described by Doornik and Hansen [9] is denoted as the DH test. The skewness-based test of normality described by Kankainen, Taskinen and Oja [22] is denoted as the  $\text{KTO}_S$  test, and the kurtosis-based test of normality described in the same paper is denoted as the  $\text{KTO}_K$  test. The test developed by Bowman and Shenton [6] is denoted as the BS test. The testing procedure studied by Villasenor Alva and Estrada [39] is denoted as the VE test. The test of normality developed by Zhou and Shao [42] is denoted as the ZS test. The testing procedure based on the measure of skewness proposed in [30] is denoted as the MRS test. The test for multivariate skewness described in [26] is denoted as the MAS test, while the test for multivariate kurtosis described by the same authors is denoted as the MAK test. The test of kurtosis based on the measure of multivariate kurtosis proposed in [24] is denoted as the KK test.

The HZ test, the R test and the DH test are implemented using the corresponding functions in the R (R version 4.1.3 (2022-03-10), [33]) package MVN [23]. The  $\text{KTO}_S$  and  $\text{KTO}_K$  tests are implemented using the functions in the R package ICS [31]. The BS test and the ZS test are implemented using their functions in the R package mvnormalTest [41]. The VE test is implemented using its function in the R package mvShapiroTest [17]. The MRS test is implemented using its function in the R package MultiSkew [13]. The MAS test, the MAK test and the KK test are implemented using their respective functions in the R package mnt [7]. The Mardia skewness and kurtosis tests are implemented based on the asymptotic distributions of the test statistics derived in [27] using the unbiased estimate of the population covariance matrix.

For the comparison of performances of the tests, we consider three simulation models. In each of the models, the non-Gaussian feature is supported on a small number of components of the random vector. Let  $\Sigma = (\sigma_{ij})$  with  $\sigma_{ij} = 0.5 + 0.5\mathbb{I}(i = j)$ . The dimension of the matrix  $\Sigma$  is to be determined based on the context. Let  $\mathbf{X} = (\mathbf{X}_q^\top, \mathbf{X}_{p-q}^\top)^\top$ , where  $\mathbf{X}_q$  and  $\mathbf{X}_{p-q}$  are independent and  $\mathbf{X}_{p-q}$  follows  $\mathcal{N}_{p-q}(\mathbf{0}, \Sigma)$ . Then,

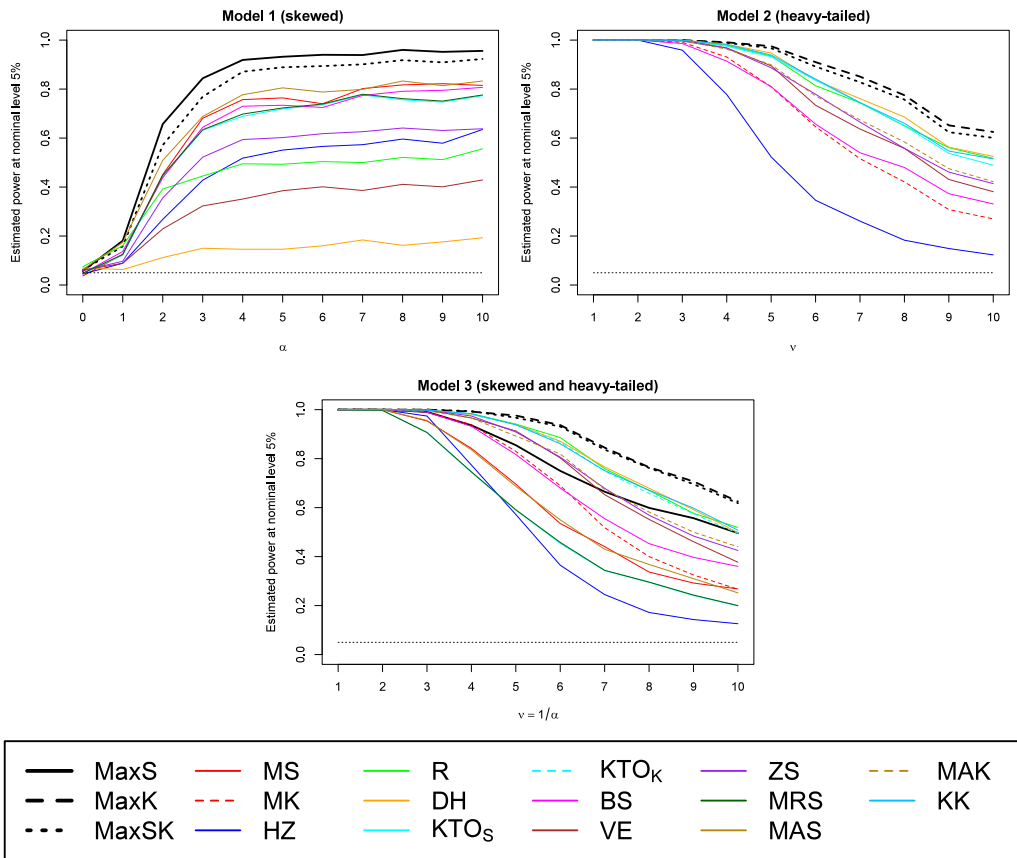
- Model 1 (skewed):  $\mathbf{X}_q \sim \mathcal{SN}_q(\mathbf{0}, \Sigma, \boldsymbol{\alpha})$ , where  $\boldsymbol{\alpha} = \alpha \mathbf{1}_q$ ;
- Model 2 (heavy-tailed):  $\mathbf{X}_q \sim t_q(\mathbf{0}, \Sigma, \nu)$ , where the degrees of freedom are  $\nu$ ;
- Model 3 (skewed and heavy-tailed):  $\mathbf{X}_q \sim \mathcal{ST}_q(\mathbf{0}, \Sigma, \boldsymbol{\alpha}, \nu)$ , where  $\boldsymbol{\alpha} = (1/\nu)\mathbf{1}_q$ .

It can be seen that the class of distributions in Model 1 is skewed-Gaussian distributions, while in Model 2, the class of distributions is symmetric heavy-tailed. In Model 3, the non-Gaussian distributions are both skewed and heavy-tailed.

We fix the sample size  $n = 200$ ,  $p = 5$  and  $q = 2$ . Then, we vary the values of  $\alpha$  or  $\nu$  to investigate the changes in power of the tests in the distributions. The plots of the estimated power curves of the tests are presented in Fig. 2. In the panel of Model 1, the kurtosis-based tests, namely MaxK test, the Mardia kurtosis (MK) test, the  $\text{KTO}_K$  test, the MAK test and the KK test are not included, as Model 1 is concerned with skewness only. Similarly, in the panel for Model 2, the skewness-based tests, namely the MaxS test, the Mardia skewness (MS) test, the  $\text{KTO}_S$  test, the MRS test and the MAS test are not included, as Model 2 is concerned with kurtosis only. However, in the plot for Model 3, all the tests are included.

From Fig. 2, it can be seen that in Model 1, the performances of MaxS and MaxSK tests are significantly better than all other tests. The power of the MaxSK test is slightly lower than the MaxS test. This is because MaxSK combines the MaxS and the MaxK tests using Bonferroni correction. If one of the tests does not exhibit a high power, then the power of MaxSK would be lower than the best performing test. Similar observations can be made in the panel for Model 2, where the estimated powers of the MaxK test and the MaxSK test are found to be better than all other tests, and the power curve of the MaxSK test is slightly below the power curve of the MaxK test. In Model 3, the estimated powers of the MaxSK test and the MaxK test are higher than other testing procedures, while the power exhibited by the MaxS test is lower in this model (this is because the skewness in Model 3 is rather weak). In all the cases, it can be clearly seen that our tests significantly outperform the other tests most of the time. The test for Gaussianity, i.e., the MaxSK test, always performs quite well and exhibits better performance than all the other tests.

Among the other testing procedures, the MAS test and the MAK test involve finding the projections which maximize the univariate skewness and kurtosis, respectively. Skewness-based projection pursuit has also been studied in [25]. The projection pursuit methods yield directions to project the data which would maximize the skewness or kurtosis. However, if the skewness or kurtosis is supported on only a sub-dimension of the data, the obtained direction vector may not directly help in identifying the particular sub-dimension. It is also notable that the two tests MAS and MAK are found to be less



**Fig. 2.** Estimated powers of MaxS, MaxK, MaxSK, MS, MK, HZ, R, DH, KTO<sub>S</sub>, KTO<sub>K</sub>, BS, VE, ZS, MRS, MAS, MAK and KK tests for 5% nominal level (horizontal dashed line) in Model 1 (top left), Model 2 (top right) and Model 3 (bottom) for  $n = 200$ ,  $p = 5$ ,  $q = 2$  based on 1000 replicates. The horizontal dotted lines near the bottom of the plots correspond to the nominal level of 5%.

powerful compared to our proposed tests in the above simulation study of estimated powers. The KTO<sub>K</sub> test proposed in [22] and implemented using the R package ICS [31] involves the ratio of the regular covariance matrix and the matrix of fourth moments, which is closely related to independent component analysis (see, e.g., section 2.4 in [32]). In independent component analysis, one tries to find a linear transform of the original multivariate data vectors so that the components of the transformed vectors are independent. However, the methodology of independent component analysis would not help in detecting the sub-dimension supporting the skewness or kurtosis present in the data due to the effect of taking linear transforms of the original data. The KTO<sub>K</sub> test is also found to be less powerful compared to our proposed tests in the above simulation study.

In [14], the authors conjectured that the normality test based on Mardia skewness is less powerful when skewness is present in a lower-dimensional space than the space of the underlying distribution, where the lower-dimensional space may be a few components of the random vector following the underlying distribution, or a projection to a lower-dimensional space. In the plots of the estimated powers, this phenomenon is clearly observed with the Mardia tests exhibiting significantly less power compared to our proposed tests and other tests in models where the non-Gaussian features are supported on only a small sub-dimension of the overall distribution. In [14], graphical methods to investigate skewness in such scenarios are described using skewness-based projection pursuit.

In section 2 of the supplementary material, estimated powers of the tests are presented for the smaller sample size  $n = 50$ . For this smaller sample size, it is found that in Model 1, which is the skewed model, the powers of all the tests decrease considerably, and a few tests exhibit higher powers compared to the MaxSK test. However, the power of the MaxS test is found to be higher than all these tests, except the R test. The R test sometimes exhibits higher estimated power compared to the MaxS test. In Model 2, which is the heavy-tailed model, again it is found that the estimated power of the R test is slightly higher than the MaxK and the MaxSK tests, although they are very close to each other. The other tests exhibit lower powers, but some of them are very close to the powers of the MaxK and MaxSK tests. In Model 3, which is the skewed and heavy-tailed model, it is found that the powers of the MaxS and MaxSK tests are considerably higher than all other tests. The power of the MaxK test is very close to that of the R test, which is significantly lower than the powers of the MaxS test and the MaxSK test.

**Table 3**

Assigned indices of all sub-dimensions of a 5-dimensional vector, where the numbers 1, 2, 3, 4 and 5 correspond to the five components of the vector.

Sub-dimension	Index	Sub-dimension	Index	Sub-dimension	Index
(1)	1	(2)	2	(3)	3
(4)	4	(5)	5	(1, 2)	6
(1, 3)	7	(1, 4)	8	(1, 5)	9
(2, 3)	10	(2, 4)	11	(2, 5)	12
(3, 4)	13	(3, 5)	14	(4, 5)	15
(1, 2, 3)	16	(1, 2, 4)	17	(1, 2, 5)	18
(1, 3, 4)	19	(1, 3, 5)	20	(1, 4, 5)	21
(2, 3, 4)	22	(2, 3, 5)	23	(2, 4, 5)	24
(3, 4, 5)	25	(1, 2, 3, 4)	26	(1, 2, 3, 5)	27
(1, 2, 4, 5)	28	(1, 3, 4, 5)	29	(2, 3, 4, 5)	30
(1, 2, 3, 4, 5)	31				

In section 3 of the supplementary material, the estimated powers of the tests are also presented in an equicorrelation model with high pairwise correlation between the components of the vector  $\mathbf{X}_i$ , for sample size  $n = 200$  and  $50$ . This experiment is carried out to evaluate the effect of the high pairwise correlation on the powers of the tests, and how this effect varies with sample size. It is found that for  $n = 200$ , the estimated powers of the MaxS, MaxSK and R tests increase sharply in Model 1, which is the skewed model, under high pairwise correlation. There is almost no difference among the powers of the MaxS, MaxSK and R tests in Model 1. However, in Model 2 and Model 3, the estimated powers of the R test is significantly lower than several other tests, and the estimated powers of the MaxK test and the MaxSK test are significantly higher than all other tests. On the other hand, when  $n = 50$ , it is found that in Model 1 under high pairwise correlation, the power of the R test is significantly higher than all of tests, and the powers of the MaxS and MaxSK tests are lower than that of the R test but significantly higher than the powers of all of the other tests. This strong dominance of the power of the R test is not maintained in case of Model 2 and Model 3 under high pairwise correlation, where several other tests exhibit higher powers compared to the R test. There, the MaxK and MaxSK tests exhibit the highest powers, but the powers of several other tests are close to them.

### 7.3. Detection of sub-dimensions supporting skewness and excess kurtosis

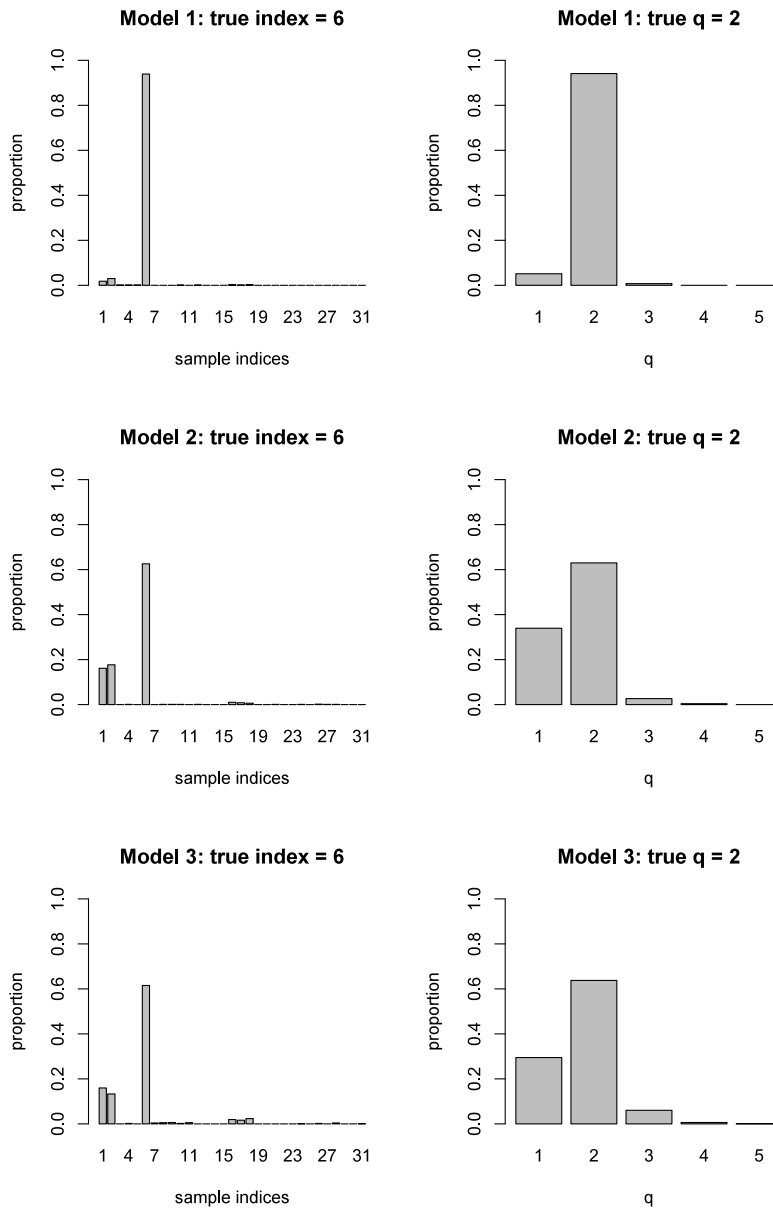
The testing procedures described earlier can be used to detect the sub-dimensions supporting non-Gaussian features in the data. Suppose the data are skewed, but skewness is supported only on a small sub-dimension of the data. Then, to detect the sub-dimension supporting skewness, we can first conduct the MaxS test. If the  $p$ -value of the MaxS test is small (say, lower than 5%), then there is statistical evidence of presence of skewness in the sample. Next, we find the sub-dimension corresponding to the maximum  $\tilde{b}_{1,q,i}$ , which is the detected sub-dimension supporting skewness in the data. Similarly, if a heavy-tailed component is present in a small sub-dimension of the data, and we wish to detect that sub-dimension, we can conduct the MaxK test and if it rejects Gaussianity, we find the sub-dimension which corresponds to the maximum  $|\tilde{b}_{2,q,i}|$ . If we want to detect the sub-dimension supporting a non-Gaussian distribution, we can use the MaxSK test in the following way. We first conduct the MaxSK test. If it rejects Gaussianity, then we find which  $p$ -value, whether for MaxS or MaxK, caused the rejection. If it is only one of the tests, say MaxS, then we detect the sub-dimension corresponding to the maximum  $\tilde{b}_{1,q,i}$ . Otherwise, if the  $p$ -values of both the MaxS test and the MaxK test are below 2.5% (due to Bonferroni correction on the 5% nominal level), then we find the sub-dimensions corresponding to the maximum  $\tilde{b}_{1,q,i}$  and the maximum  $|\tilde{b}_{2,q,i}|$ . The union of these two sub-dimensions is the detected sub-dimension supporting the non-Gaussian distribution.

To investigate the performance of the detection procedure described above, we consider the three models described in Section 7.2. In Model 1, we fix  $\alpha = 5$  and conduct the detection procedure 1000 times on independent replicates. The proportion of times each of the sub-dimensions is detected as the one supporting the skewed distribution is computed. Also, the size of the sub-dimensions (denoted as  $q$ ) thus detected to support the skewed distribution is also recorded, and the proportions for the  $q$  values are computed. All the possible sub-dimensions from  $p = 5$  variables are assigned indices, which are presented in Table 3, and the proportions thus computed are plotted against the indices in the first row of Fig. 3. The estimated power of the MaxS test there is 0.922. It can be clearly seen that the highest proportions in the respective histograms are attained for the true sub-dimension and the true  $q$ .

Similarly, the procedure to detect the sub-dimension supporting the heavy-tailed distribution using the MaxK test is also carried out based on 1000 replicates in Model 2 with  $\nu = 5$ . The estimated power of the MaxK test there is 0.985. The histograms are presented in the second row of Fig. 3, and we again see that the highest proportions are attained for the true sub-dimension and true  $q$ .

Finally, the procedure to detect the sub-dimension supporting a non-Gaussian distribution using the MaxSK test is carried out in Model 3 taking  $\nu = 1/\alpha = 5$  and based on 1000 independent replicates. The estimated power of the MaxSK test is 0.974 there. The histograms are presented in the third row of Fig. 3, where we again see that the highest proportions are attained for the true sub-dimension and true value of  $q$ .





**Fig. 3.** Histograms of the detection rate of sub-dimensions in Model 1 using the centered and scaled skewness measure (first row), in Model 2 using the centered and scaled kurtosis measure (second row), in Model 3 using the procedure for detecting non-Gaussianity (third row). In all the cases,  $n = 200$ ,  $p = 5$  and the true sub-dimension supporting skewness or excess kurtosis is  $q = 2$ . Results based on 1000 replicates.

## 8. Sub-dimensional data analysis

### 8.1. Fisher's iris data

We revisit Fisher's iris dataset discussed in Section 1, where we considered a part of the data related to the species 'iris setosa' to demonstrate that the Mardia test of skewness fails to detect skewed features in sub-dimensions. In Table 4, the p-values of all the tests are presented. It can be seen that our test detects skewness in the data, while the Mardia test fails. Similar observations were made in [23]. The p-values of the  $MaxS_1$ ,  $MaxS_2$  and  $MaxSK_1$  tests, the R test, the DH test, the VG test and the ZS test are also smaller than the 5% nominal level. Next, we consider the whole Fisher's iris dataset and compute the p-values of all the tests. We find that our test of kurtosis can detect the deviation of kurtosis from Gaussian kurtosis, while the Mardia test of kurtosis fails. The p-values of the majority of the other tests are also smaller than the 5% nominal level.

**Table 4**

Estimated p-values of the tests defined in Sections 5.1, 5.2 and 6.3 and the tests in the literature mentioned in Section 7.2 in the data examples. Significant p-values at 5% level are in italics.

Test name	Iris setosa <i>p</i> = 4	Iris <i>p</i> = 4	Wind <i>p</i> = 3
MS	0.236	<i>0.000</i>	<i>0.004</i>
MK	0.448	0.611	0.338
MaxS	<i>0.001</i>	<i>0.000</i>	<i>0.032</i>
MaxK	0.360	<i>0.003</i>	<i>0.007</i>
MaxSK	<i>0.002</i>	<i>0.000</i>	<i>0.014</i>
MaxS <sub>1</sub>	<i>0.002</i>	0.317	<i>0.023</i>
MaxS <sub>2</sub>	<i>0.047</i>	<i>0.000</i>	<i>0.011</i>
MaxS <sub>3</sub>	0.132	<i>0.000</i>	<i>0.001</i>
MaxS <sub>4</sub>	0.235	<i>0.000</i>	–
MaxK <sub>1</sub>	0.244	<i>0.000</i>	<i>0.004</i>
MaxK <sub>2</sub>	0.236	0.068	0.124
MaxK <sub>3</sub>	0.570	0.258	0.367
MaxK <sub>4</sub>	0.436	0.568	–
MaxSK <sub>1</sub>	<i>0.004</i>	<i>0.000</i>	<i>0.008</i>
MaxSK <sub>2</sub>	0.094	<i>0.000</i>	<i>0.022</i>
MaxSK <sub>3</sub>	0.264	<i>0.000</i>	<i>0.002</i>
MaxSK <sub>4</sub>	0.470	<i>0.000</i>	–
HZ	0.050	<i>0.000</i>	0.097
R	<i>0.000</i>	<i>0.000</i>	<i>0.011</i>
DH	<i>0.000</i>	<i>0.000</i>	<i>0.022</i>
KT <sub>S</sub>	0.221	<i>0.040</i>	0.710
KT <sub>K</sub>	0.875	<i>0.046</i>	0.349
BS	0.060	<i>0.000</i>	<i>0.000</i>
VG	<i>0.012</i>	<i>0.000</i>	0.057
ZS	<i>0.020</i>	0.090	0.090
MRS	0.212	<i>0.038</i>	0.708
MAS	0.342	<i>0.034</i>	0.142
MAK	0.258	0.906	0.064
KK	0.060	0.312	0.080

Fisher’s iris dataset is generally modeled using a Gaussian distribution. However, our findings point to the non-Gaussianity of the data, and thus it may be judicious to use non-Gaussian and skewed distributions while analyzing this dataset.

### 8.2. Wind speed data in Saudi Arabia

We consider a trivariate windspeed dataset produced by Yip [40] with the Weather Research and Forecasting (WRF) model. The three components correspond to bi-weekly mid-day windspeed during the period 2009–2014 at three locations near Dumat Al Jandal, the first wind farm currently under construction in Saudi Arabia. It is important to study the distributional properties of this trivariate windspeed vector because they are crucial for understanding wind patterns that will influence the production of electricity by the nearby wind farm. In particular, it is of interest to assess whether a Gaussian distribution is suitable, or a non-Gaussian model needs to be developed.

The dataset consists of  $n = 156$  trivariate windspeed vectors. A Ljung–Box test reveals no indication of serial dependence, hence, the dataset is treated as a random sample from a three-dimensional distribution. The p-values of the various tests are listed in Table 4. At the 5% level, the Mardia tests support skewness but do not reject a Gaussian kurtosis. Our global tests, however, reject both symmetry and Gaussian kurtosis, suggesting that a non-Gaussian distribution would be more suitable to model these data. Looking at our sub-dimensional tests, we observe that skewness is rejected in all sub-dimensions, whereas Gaussian kurtosis is rejected only for the  $q = 1$  dimensional marginals. Among the other twelve tests of Gaussianity, only three reject Gaussianity whereas the remaining nine cannot.

In summary, our new tests suggest to use a non-Gaussian distribution to model these data. They provide additional information about the non-Gaussian behavior in sub-dimensional components of the trivariate distribution.

## 9. Discussion

We have developed some new tests of skewness and kurtosis which take into account the skewness and excess kurtosis present in the sub-dimensions of the data. It was demonstrated through analyses of simulated and real data that our tests

outperform the classical Mardia tests of skewness and kurtosis when the skewness and the excess kurtosis are present in a small sub-dimension of the variables under consideration. Moreover, our tests can also be used as tests of Gaussianity, and it was observed that as such, they outperform several popular tests of Gaussianity. We have further developed a methodology to detect the true sub-dimension when the skewness and the excess kurtosis are supported on a small sub-dimension of the data.

One limitation of our methodology is that it considers all the possible sub-dimensions, which is  $2^p - 1$ , to detect skewness or excess kurtosis. The number  $2^p - 1$  becomes large for even moderate values of  $p$ . So, the methodology is computationally intensive. Future research needs to develop suitable computational methods when the dimension  $p$  of the multivariate data is high to reduce the computation burden. In particular, methodology is required to be developed when the data are high-dimensional in nature, i.e.,  $p > n$ . We discuss some possible ways.

In the high-dimensional setup, in case it is known that the skewness or the non-Gaussian kurtosis can only possibly be supported on a sub-dimension  $q_0$  (with  $q_0$  relatively small compared to the sample size  $n$ ), then the testing procedures  $\text{MaxS}_{q_0}$ ,  $\text{MaxK}_{q_0}$  and  $\text{MaxSK}_{q_0}$  can be applied, which are described in Section 5.1, Section 5.2 and Section 6.3, respectively. However, in a setup where such information is not available, new procedures need to be developed. One way is to randomly select a fixed number of sub-dimensions from the collection of  $2^p - 1$  possible sub-dimensions, carry out the usual Mardia tests of skewness and kurtosis on those sub-dimensions, and then combine the results of those Mardia tests based on a multiple testing procedure to get the result of the overall test. A second possible way is to consider a fixed but suitably large number of random projections from the original dimension  $p$  to a smaller dimension  $p'$ , where our procedures can be applied. Then, our testing procedures  $\text{MaxS}$ ,  $\text{MaxK}$  and  $\text{MaxSK}$  can be applied on the  $p'$ -dimensional projected data for all the random projections, and the results for all the random projections can be combined using a multiple testing method. A third possible way can be to perform principal component analysis to reduce the dimension of the data from  $p$  to  $p'$ , and then apply our procedures on the dimension-reduced data.

The procedures proposed herein are developed based on the Mardia measures and tests, and these tests cannot detect non-Gaussianity if none of the sub-dimensional Mardia measures are able to detect it. One example of such a non-Gaussian distribution is given in [10]. There, all the sub-dimensional Mardia measures including the global Mardia measures coincide with the Gaussian distribution. In such a case, our methodology will not work, and different methods are required.

## CRedit authorship contribution statement

**Joydeep Chowdhury:** Conceptualization, Methodology, Writing. **Subhajit Dutta:** Conceptualization, Methodology, Writing. **Reinaldo B. Arellano-Valle:** Conceptualization, Methodology, Writing. **Marc G. Genton:** Conceptualization, Methodology, Writing.

## Acknowledgments

This research was supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia. We thank the associate editor and the anonymous reviewers for their helpful comments and suggestions.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2022.105089>. The supplementary material contains some additional simulations: estimated sizes for  $p = 3$  and 4, estimated powers for  $n = 50$ , and investigation of the performance of the tests under high pairwise correlation.

## References

- [1] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 61 (3) (1999) 579–602.
- [2] A. Azzalini, A. Capitanio, *The Skew-Normal and Related Families*, vol. 3, Cambridge University Press, 2014, p. 262.
- [3] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, *Biometrika* 83 (4) (1996) 715–726.
- [4] L. Baringhaus, N. Henze, Limit distributions for Mardia's measure of multivariate skewness, *Ann. Statist.* 20 (4) (1992) 1889–1902.
- [5] M.A. Blanco, M. Flórez, M. Bermejo, Evaluation of the rotation matrices in the basis of real spherical harmonics, *J. Mol. Struct. THEOCHEM* 419 (1–3) (1997) 19–27.
- [6] K.O. Bowman, L.R. Shenton, Omnibus test contours for departures from normality based on  $\sqrt{b_1}$  and  $b_2$ , *Biometrika* 62 (2) (1975) 243–250.
- [7] L. Butsch, B. Ebner, *mnt: Affine invariant tests of multivariate normality*, 2020, R package version 1.3.
- [8] W. Chen, M.G. Genton, Are you all normal? It depends!, *Internat. Statist. Rev.* (2022) in press.
- [9] J.A. Doornik, H. Hansen, An omnibus test for univariate and multivariate normality, *Oxf. Bull. Econ. Stat.* 70 (2008) 927–939.
- [10] S. Dutta, M.G. Genton, A non-Gaussian multivariate distribution with all lower-dimensional Gaussians and related families, *J. Multivariate Anal.* 132 (2014) 82–93.
- [11] A. Erdélyi, *Higher Transcendental Functions*, vol. 2, McGraw-Hill Book Company, 1953, p. 396.
- [12] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugen.* 7 (2) (1936) 179–188.
- [13] C. Franceschini, N. Loperfido, *MultiSkew: Measures, tests and removes multivariate skewness*, 2017, R package version 1.1.1.
- [14] C. Franceschini, N. Loperfido, *MaxSkew and MultiSkew: two R packages for detecting, measuring and removing multivariate skewness*, *Symmetry* 11 (8) (2019) 970.

- [15] J.L. Gastwirth, M.E.B. Owens, On classical tests of normality, *Biometrika* 64 (1) (1977) 135–139.
- [16] M.G. Genton, N.M.R. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, *Ann. Inst. Statist. Math.* 57 (2) (2005) 389–401.
- [17] E. Gonzalez-Estrada, J.A. Villasenor-Alva, *mvShapiroTest*: Generalized shapiro-wilk test for multivariate normality, 2013, R package version 1.0.
- [18] G.G. Gregory, Large sample theory for  $U$ -statistics and tests of fit, *Ann. Statist.* 5 (1) (1977) 110–123.
- [19] N. Henze, On Mardia's kurtosis test for multivariate normality, *Comm. Statist. Theory Methods* 23 (4) (1994) 1031–1045.
- [20] N. Henze, B. Zirkler, A class of invariant consistent tests for multivariate normality, *Comm. Statist. Theory Methods* 19 (10) (1990) 3595–3617.
- [21] A. Higuchi, Symmetric tensor spherical harmonics on the  $N$ -sphere and their application to the de Sitter group  $SO(N, 1)$ , *J. Math. Phys.* 28 (7) (1987) 1553–1566.
- [22] A. Kankainen, S. Taskinen, H. Oja, Tests of multinormality based on location vectors and scatter matrices, *Stat. Methods Appl.* 16 (3) (2007) 357–379.
- [23] S. Korkmaz, D. Goksuluk, G. Zararsiz, MVN: An R package for assessing multivariate normality, *R J.* 6 (2) (2014) 151–162, URL <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>.
- [24] J.A. Koziol, A note on measures of multivariate kurtosis, *Biom. J.* 31 (5) (1989) 619–624.
- [25] N. Loperfido, Skewness-based projection pursuit: A computational approach, *Comput. Statist. Data Anal.* 120 (2018) 42–57.
- [26] J.F. Malkovich, A.A. Afifi, On tests for multivariate normality, *J. Amer. Statist. Assoc.* 68 (341) (1973) 176–179.
- [27] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, *Biometrika* 57 (3) (1970) 519–530.
- [28] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, *Sankhyā* 36 (2) (1974) 115–128.
- [29] K. Mardia, P. Zemroch, Algorithm AS 84: Measures of multivariate skewness and kurtosis, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 24 (2) (1975) 262–265.
- [30] T.F. Móri, V.K. Rohatgi, G.J. Székely, On multivariate skewness and kurtosis, *Theory Probab. Appl.* 38 (3) (1993) 547–551.
- [31] K. Nordhausen, H. Oja, D.E. Tyler, Tools for exploring multivariate data: The package ICS, *J. Stat. Softw.* 28 (6) (2008) 1–31, URL <http://www.jstatsoft.org/v28/i06/>.
- [32] H. Oja, S. Sirkiä, J. Eriksson, Scatter matrices and independent component analysis, *Austrian J. Stat.* 35 (2&3) (2006) 175–189.
- [33] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [34] J.P. Royston, An extension of Shapiro and Wilk's  $W$  test for normality to large samples, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 31 (2) (1982) 115–124.
- [35] J.P. Royston, Some techniques for assessing multivariate normality based on the Shapiro-Wilk  $W$ , *J. R. Stat. Soc. Ser. C. Appl. Stat.* 32 (2) (1983) 121–133.
- [36] J.P. Royston, Approximating the Shapiro-Wilk  $W$ -test for non-normality, *Stat. Comput.* 2 (3) (1992) 117–119.
- [37] J.P. Royston, Remark AS R94: A remark on algorithm AS 181: The  $W$ -test for normality, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 44 (4) (1995) 547–551.
- [38] N.J.H. Small, Marginal skewness and kurtosis in testing multivariate normality, *J. R. Stat. Soc. Ser. C. Appl. Stat.* 29 (1) (1980) 85–87.
- [39] J.A. Villasenor Alva, E. González Estrada, A generalization of Shapiro-Wilk's test for multivariate normality, *Commun. Stat. Theory Methods* 38 (11) (2009) 1870–1883.
- [40] C.M.A. Yip, Statistical Characteristics and Mapping of Near-Surface and Elevated Wind Resources in The Middle East (Ph.D. thesis), KAUST, 2018, <http://dx.doi.org/10.25781/KAUST-02YSW>.
- [41] Y. Zhang, M. Zhou, Y. Shao, *mvnormalTest*: Powerful tests for multivariate normality, 2020, R package version 1.0.0.
- [42] M. Zhou, Y. Shao, A powerful test for multivariate normality, *J. Appl. Stat.* 41 (2) (2014) 351–363.