Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1 – 29
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

A family of estimators for multivariate kurtosis in a nonnormal linear regression model

Hirokazu Yanagihara*

Department of Social Systems and Management, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

Received 14 January 2005
Available online 1 August 2005

Abstract

In this paper, we propose a new estimator for a kurtosis in a multivariate nonnormal linear regression model. Usually, an estimator is constructed from an arithmetic mean of the second power of the squared sample Mahalanobis distances between observations and their estimated values. The estimator gives an underestimation and has a large bias, even if the sample size is not small. We replace this squared distance with a transformed squared norm of the Studentized residual using a monotonic increasing function. Our proposed estimator is defined by an arithmetic mean of the second power of these transformed squared norms with a correction term and a tuning parameter. The correction term adjusts our estimator to an unbiased estimator under normality, and the tuning parameter controls the sizes of the squared norms of the residuals. The family of our estimators includes estimators based on ordinary least squares and predicted residuals. We verify that the bias of our new estimator is smaller than usual by constructing numerical experiments.

AMS 1991 subject classification: 62H12; 62J05

Keywords: Bias correction; Hotelling's T^2 distribution; Mahalanobis distance; Monotonic increasing function; Multivariate linear model; Nonnormality; Predicted residuals; Studentized residuals; Tuning parameter

* Corresponding author. Fax: +81 29 853 6451.
E-mail address: yanagi@sk.tsukuba.ac.jp.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.05.015

1. Introduction

We consider a nonnormal multivariate linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Theta} + \mathscr{E}\boldsymbol{\Sigma}^{1/2}, \tag{1}$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$ is an $n \times p$ observation matrix of p response variables, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ is an $n \times k$ design matrix of k explanatory variables with full rank k, $\boldsymbol{\Theta}$ is a $k \times p$ unknown parameter matrix and $\mathscr{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)'$ is an $n \times p$ error matrix. It is assumed that each vector $\boldsymbol{\varepsilon}_i$ is $i.i.d.$ with $\mathrm{E}[\boldsymbol{\varepsilon}_i] = \boldsymbol{0}$ and $\mathrm{Cov}[\boldsymbol{\varepsilon}_i] = \boldsymbol{I}_p$.

In model (1), the multivariate kurtosis (see e.g., [15,21]) is defined by

$$\begin{aligned} \kappa_4^{(1)} &= \mathrm{E}\left[\left(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right)^2\right] - p(p+2) \\ &= \mathrm{E}\left[\left\{(\boldsymbol{y}_i - \boldsymbol{\Theta}'\boldsymbol{x}_i)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\Theta}'\boldsymbol{x}_i)\right\}^2\right] - p(p+2). \end{aligned} \tag{2}$$

Note that it is defined by the expectation of the second power of the squared Mahalanobis distance between an observation and its mean. The kurtosis is one of the important tools for measuring the nonnormality of the distribution, that is, the tail weight of the distribution. It is known that the kurtosis is 0 under a normal assumption, i.e., $\boldsymbol{\varepsilon}_i \sim i.i.d.\ \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Because most ordinary statistical procedures are proposed under the normal assumption, we must correct these procedures by using an estimator of the kurtosis when the influence of nonnormality is large. For example, when the true population distribution is nonnormal, we can improve the approximation of a null distribution of the test statistics, which are obtained under the normal assumption ([5–7,28–31,32], etc.). Especially, the asymptotic null distribution of the test statistic for a covariance structure depends on the kurtosis (see e.g., [1,16,23,25,27,33]). Therefore, the test statistic for a covariance structure must be corrected for nonnormality in actual use ([2,17,25,35] and so on). As indicated by these previous studies, we can see that it is important to estimate the kurtosis.

For a multivariate distribution, Mardia [21] proposed an estimator of kurtosis constructed from the arithmetic mean of the second power of the squared sample Mahalanobis distances between the observations and their estimated values. However, Mardia's estimator is not useful because it is well known that it underestimates. Moreover, it has a large bias even if the sample size n is not small. For a univariate linear model, Pukelsheim [24] proposed an unbiased estimator of kurtosis when the variance is known. For a multivariate case, Isogai [14,15] suggested an estimator which is constructed from the sample Mahalanobis distance with Kaplan's correction term [18]. Unfortunately, both estimators are imperfect because the expectation of an inverse sample covariance matrix is not evaluated properly. Koziol [20] proposed another measure of kurtosis, and Henze [10] and Klar [19] studied its limiting distribution. But we cannot use this measure as an estimator of kurtosis, because it does not converge to the true value of a kurtosis. Therefore, it is necessary to find a more useful estimator of kurtosis.

Because we believe that the squared sample Mahalanobis distance gives an underestimation for Mardia's estimator, we propose in this paper a new estimator based on another squared distance; that is, we replace the squared sample Mahalanobis distance with the transformed squared norm of multivariate Studentized residuals [3, p. 18] using a monotonic increasing function. From this transformation, we consider that the distance between an observation and its estimated value is larger than the sample Mahalanobis distance. Moreover, our estimator incorporates a correction term and a tuning parameter. The correction term adjusts our estimator to an unbiased estimator under normality. The tuning parameter controls the sizes of the squared norms of the residuals in order to make the bias smaller than that of an ordinary estimator. Furthermore, the family of our estimators includes usual estimators based on ordinary least-squares and predicted residuals.

This paper is organized in the following way. In Section 2, we explain some notations and the condition of numerical studies in our paper. In Section 3, we describe some properties of the mean of Mardia's estimator. Our main result is given in Section 4. In Section 5, we discuss our conclusions. Some numerical studies are also given to show how well our modification works. Technical details are provided in the appendix.

2. Preliminaries

2.1. Higher-order cumulants and assumptions

In this subsection, we describe some moments in a multivariate distribution and assumptions for deriving valid asymptotic expansions of biases of estimators of kurtosis.

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$ be a $p \times 1$ random vector from $\boldsymbol{\varepsilon}_i$ $(i = 1, \ldots, n)$. Then, the lth multivariate moment of $\boldsymbol{\varepsilon}$, $\mu_{a_1\cdots a_l}$, is defined by

$$\mu_{a_1\cdots a_l} = \mathrm{E}[\varepsilon_{a_1} \cdots \varepsilon_{a_l}].$$

Similarly, the corresponding lth multivariate cumulant of $\boldsymbol{\varepsilon}$ is given by $\kappa_{a_1\cdots a_l}$. The relations between moments and cumulants become the following equations.

$$\mu_{abc} = \kappa_{abc}, \quad \mu_{abcd} = \kappa_{abcd} + \sum_{[3]} \delta_{ab}\delta_{cd},$$
$$\mu_{abcdef} = \kappa_{abcdef} + \sum_{[10]} \kappa_{abc}\kappa_{def} + \sum_{[15]} \delta_{ab}\kappa_{cdef} + \sum_{[15]} \delta_{ab}\delta_{cd}\delta_{ef},$$

where δ_{ab} is the Kronecker delta, i.e., $\delta_{aa} = 1$ and $\delta_{ab} = 0$ for $a \neq b$ and $\sum_{[j]}$ is the sum of all possible j combinations, i.e., $\sum_{[3]} \delta_{ab}\delta_{cd} = \delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}$.

Also, we consider sums of multivariate cumulants as

$$\kappa_4^{(1)} = \sum_{ab}^{p} \kappa_{aabb}, \quad \kappa_{3,3}^{(1)} = \sum_{abc}^{p} \kappa_{abc}^2, \quad \kappa_{3,3}^{(2)} = \sum_{abc}^{p} \kappa_{aab}\kappa_{bcc},$$
$$\kappa_6^{(1)} = \sum_{abc}^{p} \kappa_{aabbcc}, \quad \kappa_{4,4}^{(1)} = \sum_{abcd}^{p} \kappa_{abcd}^2, \quad \kappa_{4,4}^{(2)} = \sum_{abcd}^{p} \kappa_{aabc}\kappa_{bcdd}, \tag{3}$$

and sums of multivariate moments as

$$\mu_4^{(1)} = \sum_{ab}^{p} \mu_{aabb}, \quad \mu_{3,3}^{(1)} = \sum_{abc}^{p} \mu_{abc}^2, \quad \mu_{3,3}^{(2)} = \sum_{abc}^{p} \mu_{aab}\mu_{bcc},$$

$$\mu_6^{(1)} = \sum_{abc}^{p} \mu_{aabbcc}, \quad \mu_{4,4}^{(1)} = \sum_{abcd}^{p} \mu_{abcd}^2, \quad \mu_{4,4}^{(2)} = \sum_{abcd}^{p} \mu_{aabc}\mu_{bcdd}, \tag{4}$$

where the notation $\sum_{a_1 a_2 \cdots}^{p}$ means $\sum_{a_1=1}^{p} \sum_{a_2=1}^{p} \cdots$. From these equations, we obtain the relations between the sums of the multivariate moments and cumulants as

$$\begin{aligned}
\mu_4^{(1)} &= \kappa_4^{(1)} + p(p+2),\\
\mu_{3,3}^{(1)} &= \kappa_{3,3}^{(1)},\\
\mu_{3,3}^{(2)} &= \kappa_{3,3}^{(2)},\\
\mu_6^{(1)} &= \kappa_6^{(1)} + 2\left\{2\kappa_{3,3}^{(1)} + 3\kappa_{3,3}^{(2)}\right\} + 3(p+4)\kappa_4^{(1)} + p(p+2)(p+4),\\
\mu_{4,4}^{(1)} &= \kappa_{4,4}^{(1)} + 6\kappa_4^{(1)} + 3p(p+2),\\
\mu_{4,4}^{(2)} &= \kappa_{4,4}^{(2)} + 2(p+2)\kappa_4^{(1)} + p(p+2)^2.
\end{aligned} \tag{5}$$

Next, in order to guarantee the validity of the asymptotic expansions of expectations, we assume that the following assumptions, which are the same as those in Wakaki et al. [28], are held until the end of this paper.

Assumption. Let ϕ_n be the smallest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$ and $M_n = \max\{\|\boldsymbol{x}_i\| : i = 1, \ldots, n\}$, where $\|\boldsymbol{x}\|$ denotes the Euclidean norm of vector $\boldsymbol{x}$. Then, the assumptions 1, 2, 3 and 4 are:

1. For some integer $s \geqslant 3$, $\mathrm{E}[\|\boldsymbol{\varepsilon}\|^s] < \infty$.
2. For some integer $s \geqslant 3$, $\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{x}_i\|^s < \infty$.
3. $\liminf_{n\to\infty} \frac{\phi_n}{n} > 0$,
4. For some constant $0 < \delta \leqslant 1/2$, $M_n = \mathrm{O}(n^{1/2-\delta})$.

2.2. *Three residuals*

In this subsection, we describe used three residuals in our paper.

(i) *Ordinary least-squares residual*: Let $\hat{\boldsymbol{\varepsilon}}_i$ $(i = 1, \ldots, n)$ be the $p \times 1$ estimated residuals standardized by a sample covariance matrix, i.e.,

$$\hat{\boldsymbol{\varepsilon}}_i = \hat{\boldsymbol{\Sigma}}^{-1/2}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right), \tag{6}$$

where

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{Y}.$$

Here, $\boldsymbol{P}_A = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ denotes the projection matrix to the space spanned by the columns of $\boldsymbol{A}$.

(ii) *Internally multivariate Studentized residual*: We consider the following estimated residual $\tilde{\boldsymbol{\varepsilon}}_i$ $(i = 1, \ldots, n)$.

$$\tilde{\boldsymbol{\varepsilon}}_i = \frac{1}{\sqrt{1 - (\boldsymbol{P}_X)_{ii}}}\boldsymbol{S}^{-1/2}(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i), \tag{7}$$

where $(\boldsymbol{A})_{ij}$ denotes the (i, j)th element of a matrix $\boldsymbol{A}$, and $\boldsymbol{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$, i.e.,

$$\boldsymbol{S} = \frac{1}{n-k} \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_X)\boldsymbol{Y}.$$

Note that the covariance matrix of $\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i$ is $\{1 - (\boldsymbol{P}_X)_{ii}\}\boldsymbol{\Sigma}$. Thus, the covariance matrix of $\{1 - (\boldsymbol{P}_X)_{ii}\}^{-1/2}(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i)$ is corrected to $\boldsymbol{\Sigma}$. In the univariate case, Cook and Weisberg [3, p. 18] called such a residual the internally Studentized residual. Our residual is a multivariate version of this. Therefore, we call $\tilde{\boldsymbol{\varepsilon}}_i$ the ith internally multivariate Studentized residual.

(iii) *Externally multivariate Studentized residual*: Let $\boldsymbol{Y}_{(-i)}$ and $\boldsymbol{X}_{(-i)}$ be obtained from $\boldsymbol{Y}$ and $\boldsymbol{X}$ by deleting $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$, respectively, and let $\tilde{\boldsymbol{\varepsilon}}_{i[-i]}$ be the Studentized predicted residuals defined by

$$\tilde{\boldsymbol{\varepsilon}}_{i[-i]} = \frac{1}{\sqrt{1 + \boldsymbol{x}_i'(\boldsymbol{X}_{(-i)}'\boldsymbol{X}_{(-i)})^{-1}\boldsymbol{x}_i}} \boldsymbol{S}_{[-i]}^{-1/2} \left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}_{[-i]}'\boldsymbol{x}_i\right), \tag{8}$$

where $\hat{\boldsymbol{\Theta}}_{[-i]}$ and $\boldsymbol{S}_{[-i]}$ are unbiased estimators of $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ based on $\boldsymbol{Y}_{(-i)}$ and $\boldsymbol{X}_{(-i)}$, i.e.,

$$\hat{\boldsymbol{\Theta}}_{[-i]} = \left(\boldsymbol{X}_{(-i)}'\boldsymbol{X}_{(-i)}\right)^{-1} \boldsymbol{X}_{(-i)}'\boldsymbol{Y}_{(-i)},$$
$$\boldsymbol{S}_{[-i]} = \frac{1}{n-k-1}\boldsymbol{Y}_{(-i)}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{X_{(-i)}})\boldsymbol{Y}_{(-i)}.$$

From the adjusting term $\sqrt{1 + \boldsymbol{x}_i'(\boldsymbol{X}_{(-i)}'\boldsymbol{X}_{(-i)})^{-1}\boldsymbol{x}_i}$ in (8), the covariance of $\tilde{\boldsymbol{\varepsilon}}_{i[-i]}$ is adjusted to $\boldsymbol{\Sigma}$. Note that,

$$1 + \boldsymbol{x}_i'(\boldsymbol{X}_{(-i)}'\boldsymbol{X}_{(-i)})^{-1}\boldsymbol{x}_i = \frac{1}{1 - (\boldsymbol{P}_X)_{ii}}.$$

From Fujikoshi et al. [8], we obtain

$$\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}_{[-i]}'\boldsymbol{x}_i = \frac{1}{1 - (\boldsymbol{P}_X)_{ii}} \left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right).$$

Therefore, we can see that the Studentized predicted residual $\tilde{\boldsymbol{\varepsilon}}_{i[-i]}$ is equivalent to

$$\tilde{\boldsymbol{\varepsilon}}_{i[-i]} = \frac{1}{\sqrt{1 - (\boldsymbol{P}_X)_{ii}}} \boldsymbol{S}_{[-i]}^{-1/2} \left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right). \tag{9}$$

This residual is a multivariate version of the ith externally Studentized residual [3, p. 20]. Therefore, we call the residual (9) the ith externally multivariate Studentized residual.

2.3. Setting of numerical studies

In this subsection, we describe the condition of numerical studies in our paper. Since our model is a multivariate nonnormal one, we prepare the data model, which was proposed by Yuan and Bentler [34], for generating multivariate nonnormal data.

Data model: Let $w_1, \ldots, w_q$ $(q \geqslant p)$ be independent random variables with $\mathrm{E}[w_j] = 0$, $\mathrm{E}[w_j^2] = 1$ and the mth cumulants ψ_m $(m = 3, 4, \ldots)$, and $\boldsymbol{w} = (w_1, \ldots, w_q)'$. Let

r be a random variable which is independent of $\boldsymbol{w}$, $\mathrm{E}[r^2] = 1$ and the mth moments β_m $(m = 3, 4, \ldots)$. Then, we generate an error vector by

$$\boldsymbol{\varepsilon} = r\boldsymbol{A}'\boldsymbol{w},$$

where $\boldsymbol{A}$ is a $q \times p$ matrix defined by $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)'$ with a full rank p and $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_p$. In addition, the cumulants of this model become

$$\begin{aligned}
\kappa_4^{(1)} &= \beta_4\psi_4 a_4^{(1)} + (\beta_4 - 1)p(p+2),\\
\kappa_{3,3}^{(1)} &= \beta_3^2\psi_3^2 a_{3,3}^{(1)},\\
\kappa_{3,3}^{(2)} &= \beta_3^2\psi_3^2 a_{3,3}^{(2)},\\
\kappa_6^{(1)} &= \beta_6\psi_6 a_6^{(1)} + 2(\beta_6 - \beta_3^2)\psi_3^2(2a_{3,3}^{(1)} + 3a_{3,3}^{(2)})\\
&\quad +3(p+4)(\beta_6 - \beta_4)\psi_4 a_4^{(1)} + p(p+2)(p+4)(\beta_6 - 3\beta_4 + 2),\\
\kappa_{4,4}^{(1)} &= \beta_4^2\psi_4^2 a_{4,4}^{(1)} + 6(\beta_4^2 - \beta_4)\psi_4 a_4^{(1)} + 3p(p+2)(\beta_4^2 - 2\beta_4 + 1),\\
\kappa_{4,4}^{(2)} &= \beta_4^2\psi_4^2 a_{4,4}^{(2)} + 2(p+2)(\beta_4^2 - \beta_4)\psi_4 a_4^{(1)} + p(p+2)^2(\beta_4^2 - 2\beta_4 + 1),
\end{aligned}$$

where

$$a_4^{(1)} = \sum_{i=1}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_i)^2,\ a_{3,3}^{(1)} = \sum_{ij}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_j)^3,\ a_{3,3}^{(2)} = \sum_{ij}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_i)(\boldsymbol{a}_i'\boldsymbol{a}_j)(\boldsymbol{a}_j'\boldsymbol{a}_j),$$
$$a_6^{(1)} = \sum_{i=1}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_i)^3,\ a_{4,4}^{(1)} = \sum_{ij}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_j)^4,\ a_{4,4}^{(2)} = \sum_{ij}^{q}(\boldsymbol{a}_i'\boldsymbol{a}_i)(\boldsymbol{a}_i'\boldsymbol{a}_j)^2(\boldsymbol{a}_j'\boldsymbol{a}_j).$$

Let χ_f be a random variable from the χ^2 distribution with f degrees of freedom and let $\boldsymbol{A}_0$ be a $(p+1) \times p$ matrix defined by

$$\boldsymbol{A}_0 = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{1}_p' \end{pmatrix} (\boldsymbol{I}_p + \boldsymbol{1}_p\boldsymbol{1}_p')^{-1/2},$$

where $\boldsymbol{1}_p$ is a $p \times 1$ vector, and all of the vector's elements are 1. Then, the error vector $\boldsymbol{\varepsilon} = r\boldsymbol{A}_0'\boldsymbol{w}$ is written as

$$\boldsymbol{\varepsilon} = r(\boldsymbol{I}_p - \rho\boldsymbol{1}_p\boldsymbol{1}_p')\begin{pmatrix} w_1 + w_{p+1} \\ \vdots \\ w_p + w_{p+1} \end{pmatrix},$$

where

$$\rho = \frac{1}{p}\left(1 - \frac{1}{\sqrt{p+1}}\right).$$

Also, the above equations lead to

$$a_4^{(1)} = \frac{p^2}{p+1},\quad a_{3,3}^{(1)} = \frac{p(p^3 + p^2 - p + 3)}{(p+1)^3},\quad a_{3,3}^{(2)} = \frac{4p^3}{(p+1)^3},$$
$$a_6^{(1)} = \frac{p^3}{(p+1)^2},\quad a_{4,4}^{(1)} = \frac{p(p^3+1)}{(p+1)^3},\quad a_{4,4}^{(2)} = \frac{p^3}{(p+1)^2}.$$

Then we generate error vectors with the following five models,

(1) *Model 1* (*Normal distribution*): $w_j \sim \mathrm{N}(0, 1)$, $r = 1$ and $\boldsymbol{A} = \boldsymbol{I}_p$ ($\kappa_4^{(1)} = 0$).

(2) *Model 2* (*t-distribution*): $w_j \sim \mathrm{N}(0, 1)$, $r = \sqrt{6/\chi_8^2}$ and $\boldsymbol{A} = \boldsymbol{I}_p$ ($\kappa_4^{(1)} = p(p+2)/2$).

(3) *Model 3* (*Uniform distribution*): w_j is generated from a uniform $(-5, 5)$ distribution divided by the standard deviation $5/\sqrt{3}$, $r = 1$ and $\boldsymbol{A} = \boldsymbol{A}_0$ ($\kappa_4^{(1)} = -1.2 \times p^2(p+1)^{-1}$).

(4) *Model 4* (*χ^2 distribution*): w_j is generated from a χ^2 distribution with 4 degrees of freedom standardized by mean 4 and standard deviation $2\sqrt{2}$, $r = \sqrt{6/\chi_8^2}$ and $\boldsymbol{A} = \boldsymbol{A}_0$ ($\kappa_4^{(1)} = 4.5 \times p^2(p+1)^{-1} + p(p+2)/2$).

(5) *Model 5* (*Log-normal distribution*): w_j is generated from a lognormal distribution such that $\log w_i \sim \mathrm{N}(0, 1/4)$ standardized by mean $e^{1/4}$ and standard deviation $e^{1/2}\sqrt{e^{1/4} - 1}$, $r = \sqrt{6/\chi_8^2}$ and $\boldsymbol{A} = \boldsymbol{A}_0$ ($\kappa_4^{(1)} = 1.5 \times p^2(p+1)^{-1}(e + 2e^{3/4} + 3e^{1/2} - 6) + p(p+2)/2$).

3. On Mardia's estimator

In this section, we show some characteristics of the mean of an usual estimator of kurtosis, as proposed by Mardia [21]. For an asymptotic distribution of this estimator, see Henze [9].

Mardia [21] proposed an estimator of a multivariate kurtosis (2), using an arithmetic mean of the second power of the squared sample Mahalanobis distances between $\boldsymbol{y}_i$ and $\hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i$, i.e.,

$$\begin{aligned}\hat{\kappa}_4^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i\right)^2 - p(p+2)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right)'\hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right)\right\}^2 - p(p+2), \end{aligned} \tag{10}$$

where $\hat{\boldsymbol{\varepsilon}}_i$ is the ordinary least-squares residual given by (6).

In accordance with Mardia's research, we write $n^{-1}\sum_{i=1}^{n}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^2$ with $b_{2,p}$.

Sometimes the estimator (10) reveals a latent problem when it is used to estimate a large kurtosis. Let $\hat{\mathscr{E}}$ be the estimated error matrix as $\hat{\mathscr{E}} = (\hat{\boldsymbol{\varepsilon}}_1, \ldots, \hat{\boldsymbol{\varepsilon}}_n)'$. Then $n^{-1}\hat{\mathscr{E}}\hat{\mathscr{E}}'$ is an idempotent matrix. From the property of the idempotent matrix, we can see that $0 < \hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i < n$. This makes the inequality $p^2 \leqslant b_{2,p} < np$. Therefore, we obtain the following inequality.

$$-2p \leqslant \hat{\kappa}_4^{(1)} < p(n-p-2). \tag{11}$$

Consequently, we can see that it is impossible to estimate a kurtosis which is larger than $p(n-p-2)$. For example, when $p = 1$, we cannot estimate the kurtosis of the log-normal distribution with location parameter 0 and dispersion parameter 1 without at least 114 samples, because the kurtosis is $\kappa_4^{(1)} \approx 110.94$.

For more detailed properties, we first study an asymptotic expansion of the mean of $\hat{\kappa}_4^{(1)}$. We obtain the asymptotic expansion of $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ in the following theorem.

Theorem 1. *Suppose that assumptions* 1, 2, 3 *and* 4 *are held. Then* $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ *is expanded up to the order* n^{-1} *as*

$$\mathrm{E}[\hat{\kappa}_4^{(1)}] = \kappa_4^{(1)} + \frac{1}{n}[\kappa_{4,4}^{(1)} + 2\kappa_{4,4}^{(2)} - 2\kappa_6^{(1)} + 4\{(a_1-2)\kappa_{3,3}^{(1)} + (a_1-3)\kappa_{3,3}^{(2)}\} \\ -(2p+2k+11)\kappa_4^{(1)} - 2p(p+2)] + \mathrm{O}(n^{-2}), \tag{12}$$

where all κ*'s are given by* (3) *and coefficient* a_1 *is defined by*

$$a_1 = \frac{1}{n}\mathbf{1}_n'\boldsymbol{P}_{\boldsymbol{X}}\mathbf{1}_n, \quad (0 < a_1 \leqslant 1). \tag{13}$$

Proof. An asymptotic expansion of $\mathrm{E}[b_{2,p}]$ is obtained in (A.1) in Appendix A.1. Note that $\hat{\kappa}_4^{(1)} = b_{2,p} - p(p+2)$. Therefore, we obtain the result in Theorem 1. □

Next, we consider the expectation of $\hat{\kappa}_4^{(1)}$ under the special case that $\boldsymbol{\varepsilon}_i$ is *i.i.d.* $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ (the proof is described in Appendix A.3).

Theorem 2. *If* $\boldsymbol{\varepsilon}_i$ *is i.i.d.* $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$, *then the exact mean of* $\hat{\kappa}_4^{(1)}$ *becomes the following.*

$$\mathrm{E}[\hat{\kappa}_4^{(1)}] = \frac{-\{k^2 + 2(n-k) - a_2\}p(p+2)}{(n-k)(n-k+2)},$$

where coefficient a_2 *is defined by*

$$a_2 = n\sum_{i=1}^{n}\{(\boldsymbol{P}_{\boldsymbol{X}})_{ii}\}^2.$$

It is easy to see that $k^2 \leqslant a_2 < nk$. However, under the assumptions 2, 3 and 4, $k^2 + 2(n-k) - a_2$ tends to be positive, because $a_2 = \mathrm{O}(1)$. Therefore, under normality, the mean of $\hat{\kappa}_4^{(1)}$ tends to be negative in most cases. Moreover, the bias becomes large when dimension p becomes large. Especially, when $\boldsymbol{X} = \mathbf{1}_n$, we obtain the following special case of Theorem 2. This result coincides with Mardia's result [21].

Corollary 1. *If* $\boldsymbol{\varepsilon}_i$ *is i.i.d.* $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ *and* $\boldsymbol{X} = \mathbf{1}_n$, *then the exact mean of* $\hat{\kappa}_4^{(1)}$ *becomes the following.*

$$\mathrm{E}[\hat{\kappa}_4^{(1)}] = -\frac{2p(p+2)}{n+1}.$$

Proof. Note that $k = 1$ and $(\boldsymbol{P}_{\boldsymbol{X}})_{ii} = n^{-1}$ when $\boldsymbol{X} = \mathbf{1}_n$. Therefore, we obtain the result in Corollary 1. □

Before concluding this section, we show some of our simulation results. Data models 1–5, which are used in our simulations, are described in Section 2.3. First, we studied the convergence of expectation of Mardia's estimator (10). Fig. 1 shows $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ in the cases $p = 2$

and 8. Simulated values of $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ for sample size $n = 30, 40, 50, 60, 70, 80, 90, 100,$ $120, 150, 200, 300$ and 500 were obtained from $30,000, 30,000, 30,000, 10,000, 10,000,$ $10,000, 10,000, 5,000, 5,000, 5,000, 3,000, 3,000$ and $3,000$ times iterations, respectively. We used $\mathbf{1}_n$ as the design matrix $\boldsymbol{X}$ in all the cases. In this figure, the solid and broken lines denote $\kappa_4^{(1)}$ and $\mathrm{E}[\hat{\kappa}_4^{(1)}]$, respectively, and the dotted line denotes the theoretical mean which is obtained from the asymptotic expansion of $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ until the n^{-1} term in (12). The theoretical mean is an inversely proportional function with respect to the sample size n, although it disappears from some places in Fig. 1. From this figure, we notice that Mardia's estimator has a large bias, although the true kurtosis is of moderate size. The bias still exists, even if the sample size becomes huge. Moreover, it underestimates in most cases, except in model 3; that is, the true kurtosis is negative. The underestimation becomes severe when $p = 8$. On the other hand, it seems that most theoretical means are smaller than the true kurtosis, except for model 3. And, when $\kappa_4^{(1)}$ is large, these biases are still large, even if the sample size is huge. Accordingly, we consider the reason why Mardia's estimator underestimates and $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ does not easily converge to $\kappa_4^{(1)}$. In addition, if $\kappa_4^{(1)}$ is larger, a difference between $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ and the theoretical mean is larger. It means that $\mathrm{E}[\hat{\kappa}_4^{(1)}]$ is severely dependent not only n^{-1} term but n^{-2} or n^{-3} terms in the expansion. Next, we studied the mean, standard deviation (SD), and square root of the mean square error (RMSE) of $\hat{\kappa}_4^{(1)}$ from 30,000 times iteration. These are shown in Table 1. We used $n \times 3$ and $n \times 5$ design matrices, whose first columns were $\mathbf{1}_n$ and next columns were generated by $\mathrm{U}(-1, 1)$. From this table, we can see that, if the number of explanatory variables k is larger, the bias is larger. Moreover, if the true kurtosis is larger, the standard deviation and RMSE are larger. However, the increment of the standard deviation is much smaller than that of the RMSE. Finally, we notice that the standard deviation of $\hat{\kappa}_4^{(1)}$ is bigger when n is larger. Because $\hat{\kappa}_4^{(1)}$ has upper and lower bounds (11), the variance of $\hat{\kappa}_4$ is small when n is small, although an asymptotic variance depends on the eighth cumulant of $\boldsymbol{\varepsilon}$ (for an asymptotic variance of $\hat{\kappa}_4^{(1)}$, see [9]). The standard deviation approaches the asymptotic one becoming large, because the upper bound of $\hat{\kappa}_4^{(1)}$ is wider if n is larger.

4. Main result

In the previous section, we explained that Mardia's estimator gives an underestimation and has a large bias, even if the sample size is not small. We think that the following is the reason for this fact: the squared sample Mahalanobis distance between an observation and its estimate tends to be small when n is not large, because this has restrictions, i.e., $\sum_{i=1}^{n} \hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i = np$ and $0 < \hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i < n$. Therefore, the probability of an event, that is, a large outcome of $\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i$, becomes very small. This result shows $\hat{\kappa}_4^{(1)}$ give an underestimation. In this section, we extend $\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i$ by a monotonic increasing function with a turning parameter, which controls its size. By using this, we propose a new estimator of kurtosis. Moreover, we adjust the new estimator of kurtosis to an unbiased estimator under normality. It is known that kurtosis is a measure of the discrepancy from normality. Hence, it is desirable that the mean of the estimator is exactly 0 under normality.

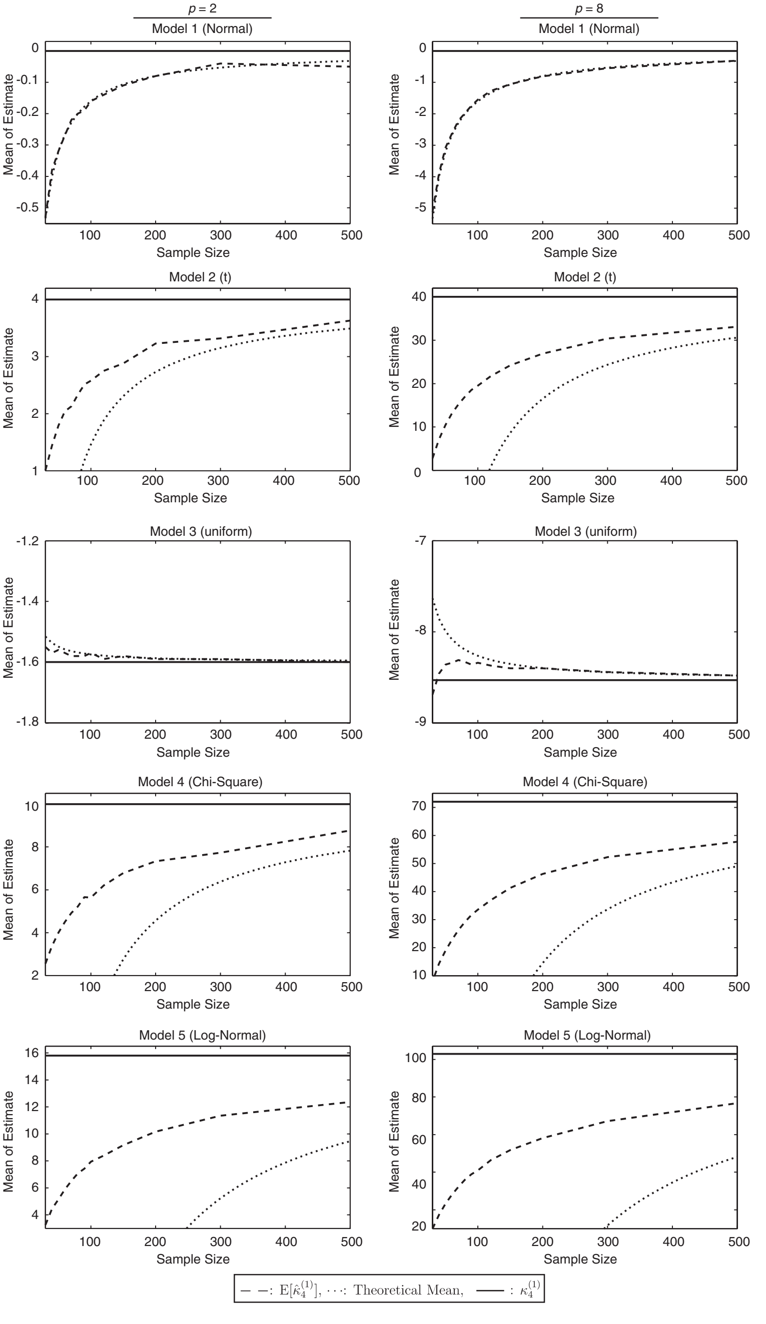

Fig. 1. The convergence of $E[\hat{\kappa}_4^{(1)}]$.

Table 1
Mean, standard deviation, and RMSE for Mardia's estimator

p	k	n		Models 1	2	3	4	5
2			$\kappa_4^{(1)}$	0.00	4.00	−1.60	10.00	15.80
	3	30	Mean	−0.53	0.74	−1.42	2.07	2.71
			(SD)	(1.09)	(1.99)	(0.70)	(2.87)	(3.33)
			(RMSE)	(1.22)	(3.82)	(0.73)	(8.43)	(13.51)
		50	Mean	−0.33	1.57	−1.46	3.62	4.72
			(SD)	(0.95)	(2.32)	(0.52)	(3.74)	(4.57)
			(RMSE)	(1.01)	(3.35)	(0.54)	(7.40)	(11.98)
	5	30	Mean	−0.56	0.51	−1.29	1.59	2.11
			(SD)	(1.08)	(1.78)	(0.76)	(2.53)	(2.89)
			(RMSE)	(1.21)	(3.92)	(0.82)	(8.78)	(13.99)
		50	Mean	−0.33	1.42	−1.36	3.25	4.26
			(SD)	(0.95)	(2.22)	(0.56)	(3.49)	(4.26)
			(RMSE)	(1.00)	(3.41)	(0.60)	(7.60)	(12.30)
8			$\kappa_4^{(1)}$	0.00	40.00	−8.53	72.00	102.92
	3	30	Mean	−5.37	1.00	−8.20	5.17	6.74
			(SD)	(2.66)	(4.43)	(2.04)	(5.40)	(5.90)
			(RMSE)	(5.99)	(39.25)	(2.07)	(67.05)	(96.36)
		50	Mean	−3.21	9.01	−7.92	16.85	20.57
			(SD)	(2.62)	(6.33)	(1.66)	(8.33)	(9.51)
			(RMSE)	(4.15)	(31.63)	(1.77)	(55.78)	(82.90)
	5	30	Mean	−5.46	−0.40	−7.69	2.86	4.20
			(SD)	(2.65)	(4.04)	(2.14)	(4.82)	(5.20)
			(RMSE)	(6.06)	(40.61)	(2.30)	(69.31)	(98.86)
		50	Mean	−3.27	7.71	−7.44	14.67	18.07
			(SD)	(2.60)	(5.93)	(1.74)	(7.67)	(8.77)
			(RMSE)	(4.17)	(32.83)	(2.06)	(57.85)	(85.30)

By using the ith internally multivariate Studentized residual, we define the new estimator of multivariate kurtosis as follows.

Definition. Let $f(x;\lambda)$ $(0 \leqslant \lambda \leqslant 1)$ be the monotonic increasing function on $x \geqslant 0$, which is given by

$$f(x;\lambda) = \frac{x}{\{1 - x/(n-k)\}^{\lambda}}. \tag{14}$$

Then, the new estimator of the multivariate kurtosis is defined by

$$\tilde{\kappa}_4^{(1)}(\lambda) = \frac{c(\lambda)}{n} \sum_{i=1}^{n} \{f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i;\lambda)\}^2 - p(p+2), \tag{15}$$

where $\tilde{\boldsymbol{\varepsilon}}_i$ is the ith internally multivariate Studentized residual given by (7) and coefficient $c(\lambda)$ is given by

$$c(\lambda) = \frac{(n-k-4\lambda)(n-k-4\lambda+2)\Gamma\left(\frac{n-k-4\lambda}{2}\right)\Gamma\left(\frac{n-k-p}{2}\right)}{(n-k)^2\Gamma\left(\frac{n-k-p-4\lambda}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}. \tag{16}$$

Here, $\Gamma(x)$ is the Gamma function given by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t.$$

Coefficient $c(\lambda)$ in (16) makes our estimator unbiased under normality, as stated in the following theorem (the proof of this is given in Appendix A.4).

Theorem 3. *If* $\boldsymbol{\varepsilon}_i$ is *i.i.d.* $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$, *then the mean of* $\tilde{\kappa}_4^{(1)}(\lambda)$ *is exactly* 0, *i.e.*, $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)] = 0$. *Therefore, our estimator is always an unbiased estimator under normality.*

Note that when $\lambda = 0$,

$$f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i; 0) = \tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i = \frac{(n-k)\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i}{n\{1-(\boldsymbol{P_X})_{ii}\}}, \tag{17}$$

where $\hat{\boldsymbol{\varepsilon}}_i$ is the ith ordinary least-squares residual given by (6). We can see that $\tilde{\kappa}_4^{(1)}(0)$ is essentially an estimator based on ordinary least-squares residuals. On the other hand, from Theorem A.3 in Appendix A.2, we can see that

$$\tilde{\boldsymbol{\varepsilon}}_{i[-i]}'\tilde{\boldsymbol{\varepsilon}}_{i[-i]} = \frac{(n-k-1)\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i}{(n-k)\{1-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i/(n-k)\}},$$

where $\tilde{\boldsymbol{\varepsilon}}_{i[-i]}$ is the ith externally multivariate residual given by (8). Thus, when $\lambda = 1$, we obtain the following equation from (A.3).

$$f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i; 1) = \frac{\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i}{1-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i/(n-k)} = \frac{n-k}{n-k-1}\tilde{\boldsymbol{\varepsilon}}_{i[-i]}'\tilde{\boldsymbol{\varepsilon}}_{i[-i]}.$$

We can see that $\tilde{\kappa}_4^{(1)}(1)$ is essentially an estimator based on the predicted residuals. Therefore, the family of our estimators includes the ones based on ordinary least-squares and predicted residuals.

When $\lambda = 0$, $1/2$ and 1, our estimator is simplified, as in the following theorem.

Theorem 4. *If* $\lambda = 0, 1/2$ *and* 1, *then* $\tilde{\kappa}_4^{(1)}(\lambda)$ *becomes as simple as*

$$\tilde{\kappa}_4^{(1)}(0) = \frac{(n-k+2)}{n(n-k)}\sum_{i=1}^n \left(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)^2 - p(p+2),$$

$$\tilde{\kappa}_4^{(1)}(1/2) = \frac{(n-k-p-2)}{n(n-k)} \sum_{i=1}^{n} \frac{\left(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i\right)^2}{1-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i/(n-k)} - p(p+2),$$

$$\tilde{\kappa}_4^{(1)}(1) = \frac{(n-k-p-2)(n-k-p-4)}{n(n-k)^2} \sum_{i=1}^{n} \left\{ \frac{\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i}{1-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i/(n-k)} \right\}^2 - p(p+2).$$

Proof. Note that

$$c(0) = \frac{n-k+2}{n-k}, \quad c(1/2) = \frac{n-k-p-2}{n-k},$$
$$c(1) = \frac{(n-k-p-2)(n-k-p-4)}{(n-k)^2}.$$

Therefore, we obtain the results in Theorem 4. $\square$

Furthermore, we obtain a more special case when $\boldsymbol{X} = \mathbf{1}_n$.

Corollary 2. *When $\boldsymbol{X} = \mathbf{1}_n$, then $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$ become as simple as*

$$\tilde{\kappa}_4^{(1)}(0) = \frac{(n+1)}{n(n-1)} \sum_{i=1}^{n} \left\{ (\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \right\}^2 - p(p+2),$$

$$\tilde{\kappa}_4^{(1)}(1/2) = \frac{(n-p-3)}{n} \sum_{i=1}^{n} \frac{\left\{ (\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \right\}^2}{n-1-(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})} - p(p+2),$$

$$\begin{aligned}\tilde{\kappa}_4^{(1)}(1) = {} & \frac{(n-p-3)(n-p-5)}{n} \sum_{i=1}^{n} \left\{ \frac{(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})}{n-1-(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})} \right\}^2 \\ & -p(p+2),\end{aligned}$$

where

$$\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'.$$

Proof. Note, if $\boldsymbol{X} = \mathbf{1}_n$, then $1 - (\boldsymbol{P}_X)_{ii} = (n-1)/n$ and $k = 1$. By recalling the relation between $\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i$ and $\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i$ in (17), we can see the equation $\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i = \tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i$ when $\boldsymbol{X} = \mathbf{1}_n$. Therefore, the result in Corollary 2 is obtained. $\square$

From Corollary 2, we can see that $\tilde{\kappa}_4^{(1)}(0)$ with $\boldsymbol{X} = \mathbf{1}_n$ coincides with the adjusted Mardia's estimator, which is an unbiased estimator under normality.

We obtain an asymptotic expansion of $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$ in a general nonnormal case as well.

Theorem 5. *Suppose that assumptions* 1, 2, 3 *and* 4 *are held. Then* $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$ *is expanded up to the order* n^{-1} *as*

$$\begin{aligned}\mathrm{E}\left[\tilde{\kappa}_4^{(1)}(\lambda)\right] &= \kappa_4^{(1)} + \frac{1}{n}\left[\kappa_{4,4}^{(1)} + 2\kappa_{4,4}^{(2)} - 2(1-\lambda)\kappa_6^{(1)}\right.\\ &\quad +4\left\{(a_1 - 2(1-\lambda))\kappa_{3,3}^{(1)} + (a_1 - 3(1-\lambda))\kappa_{3,3}^{(2)}\right\}\\ &\quad \left. -(2p+2k+9-4\lambda(p+4))\kappa_4^{(1)}\right] + \mathrm{O}(n^{-2}), \end{aligned} \tag{18}$$

where all κ*'s are given by* (3) *and coefficient* a_1 *is defined by* (13).

Proof. From Stirling's formula, we obtain the asymptotic expansion of $c(\lambda)$ in (16) as

$$c(\lambda) = 1 - \frac{2\{(p+4)\lambda - 1\}}{n} + \mathrm{O}(n^{-2}).$$

On the other hand, we derive the following perturbation expansion from (17).

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\{f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i;\lambda)\right\}^2 &= \frac{1}{n}\left(1-\frac{k}{n}\right)^2\sum_{i=1}^{n}\frac{(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^2}{\{1-(\boldsymbol{P_X})_{ii}\}^2}\left[1-\frac{\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i}{n\{1-(\boldsymbol{P_X})_{ii}\}}\right]^{-2\lambda}\\ &= \frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^2\left\{1+2(\boldsymbol{P_X})_{ii} - \frac{2}{n}\left(k-\lambda\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i\right)\right\} + \mathrm{O}_p(n^{-2}).\end{aligned}$$

Therefore, $\tilde{\kappa}_4^{(1)}(\lambda)$ is expanded as

$$\begin{aligned}\tilde{\kappa}_4^{(1)}(\lambda) &= \left(1-\frac{2\{(p+4)\lambda-1\}}{n}\right)b_{2,p} - p(p+2)\\ &\quad +\frac{2}{n}\sum_{i=1}^{n}\left\{(\boldsymbol{P_X})_{ii}-\frac{k}{n}\right\}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^2 + \frac{2\lambda}{n^2}\sum_{i=1}^{n}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^3 + \mathrm{O}_p(n^{-2}).\end{aligned}$$

Let $\mu_4^{(1)}$ and $\mu_6^{(1)}$ be the sums of the multivariate moments defined by (4). By using the relations of the moments in (5), the following expectations are derived.

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{P_X})_{ii}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^2\right] &= k\mu_4^{(1)} + \mathrm{O}(n^{-1}) = k\kappa_4^{(1)} + kp(p+2) + \mathrm{O}(n^{-1}),\\ \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i)^3\right] &= \mu_6^{(1)} + \mathrm{O}(n^{-1})\\ &= \kappa_6^{(1)} + 2\left\{2\kappa_{3,3}^{(1)} + 3\kappa_{3,3}^{(2)}\right\} + 3(p+4)\kappa_4^{(1)}\\ &\quad + p(p+2)(p+4) + \mathrm{O}(n^{-1}).\end{aligned}$$

From the above equations and the asymptotic expansion of $\mathrm{E}[b_{2,p}]$ in (A.1) in Appendix A.1, we obtain the result in Theorem 5. □

Before concluding this section, we show some of our simulation results. As stated previously, data models 1–5, which are used in our simulations, are described in Section 2.3.

Table 2
The best tuning parameter λ in models 2–5

	($n = 30$) Models				($n = 100$) Models			
p	2	3	4	5	2	3	4	5
2	0.76	0.60	0.72	0.79	0.76	0.59	0.77	0.80
8	0.71	0.55	0.68	0.71	0.71	0.54	0.71	0.75

First, we studied the effects of the tuning parameter λ and the dimension p for the mean, standard deviation, and RMSE of our new estimator. Sample sizes $n = 30$ and 100 were chosen. Because the results at $n = 100$ for the mean, standard deviation and RMSE are somewhere between those corresponding to $n = 30$, we just report those for $n = 30$ and $p = 2$ and 8 to save space. From the same reason, we report the results for $p = 2$ and 8, although the simulations were done in several dimensions $p = 1, 2, 3, 4, 6, 8$ and 10. In Fig. 2, the effects in the mean are given, and the solid and broken lines denote $\kappa_4^{(1)}$ and $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$, respectively. Moreover, the dotted line denotes the theoretical mean which is obtained from the asymptotic expansion of $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$ until the n^{-1} term in (18). The theoretical mean is a linear function with respect to the tuning parameter λ, although it disappears from some places in Fig. 2. In Fig. 3, the effects in the standard deviation and RMSE are given, and the dotted and broken lines denote the standard deviation and RMSE of $\tilde{\kappa}_4^{(1)}(\lambda)$, respectively. The design matrix was $\boldsymbol{X} = \mathbf{1}_n$, and we obtained simulated values from $30,000$ times iterations at $\lambda_l = (l-1)/99$ ($l = 1, \ldots, 100$). From those figures, we can see that there is a tuning parameter which makes the bias of $\tilde{\kappa}_4^{(1)}(\lambda)$ close to 0. We call such a tuning parameter the best λ. Table 2 shows the best λ in Fig. 2. It seems that the best λ does not depend on the sample size n so much, but it becomes small when p is large. Moreover, it seems that if the best λ is larger, the true $\kappa_4^{(1)}$ is larger. We can see that the best λ is around 0.7 when $\kappa_4^{(1)} > 0$. On the other hand, if $\kappa_4^{(1)} < 0$, as in model 3, the best λ becomes smaller than the one in the case $\kappa_4^{(1)} > 0$. The tendencies appeared in the theoretical means. On the other hand, the standard deviation grows whenever λ is large. Especially, it is huge when the true $\kappa_4^{(1)}$ is large, i.e., in a model with a high dimension, such as model 5. Also, RMSE is large, as is the standard deviation when $\kappa_4^{(1)}$ is large. However, as p becomes larger, λ makes the smallest RMSE. It seems that such a λ exists around $\lambda = 1/2$.

Next, we looked at the special $\tilde{\kappa}_4^{(1)}(\lambda)$, i.e., $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$. Fig. 4 shows the convergence of several estimators, $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$, $\tilde{\kappa}_4^{(1)}(1)$ and $\hat{\kappa}_4^{(1)}$ in the cases $p = 2$ and $p = 8$ with $\boldsymbol{X} = \mathbf{1}_n$. Simulated values of expectations of estimators for sample size $n = 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 300$ and 500 were obtained from $30,000, 30,000, 30,000, 10,000, 10,000, 10,000, 10,000, 5,000, 5,000, 5,000, 3,000, 3,000$ and $3,000$ times iterations, respectively. In this figure, the solid line denotes $\kappa_4^{(1)}$, and $\circ$, ∇, $+$ and $\triangle$ denote $\mathrm{E}[\hat{\kappa}_4^{(1)}]$, $\mathrm{E}[\tilde{\kappa}_4^{(1)}(0)]$, $\mathrm{E}[\tilde{\kappa}_4^{(1)}(1/2)]$ and $\mathrm{E}[\tilde{\kappa}_4^{(1)}(1)]$, respectively. From this figure, when $\kappa_4^{(1)} > 0$, we can see that $\tilde{\kappa}_4^{(1)}(0)$ and $\tilde{\kappa}_4^{(1)}(1/2)$ give underestimations and $\tilde{\kappa}_4^{(1)}(1)$ gives an overestimation. This tendency is reversed when $\kappa_4^{(1)} < 0$. Furthermore,

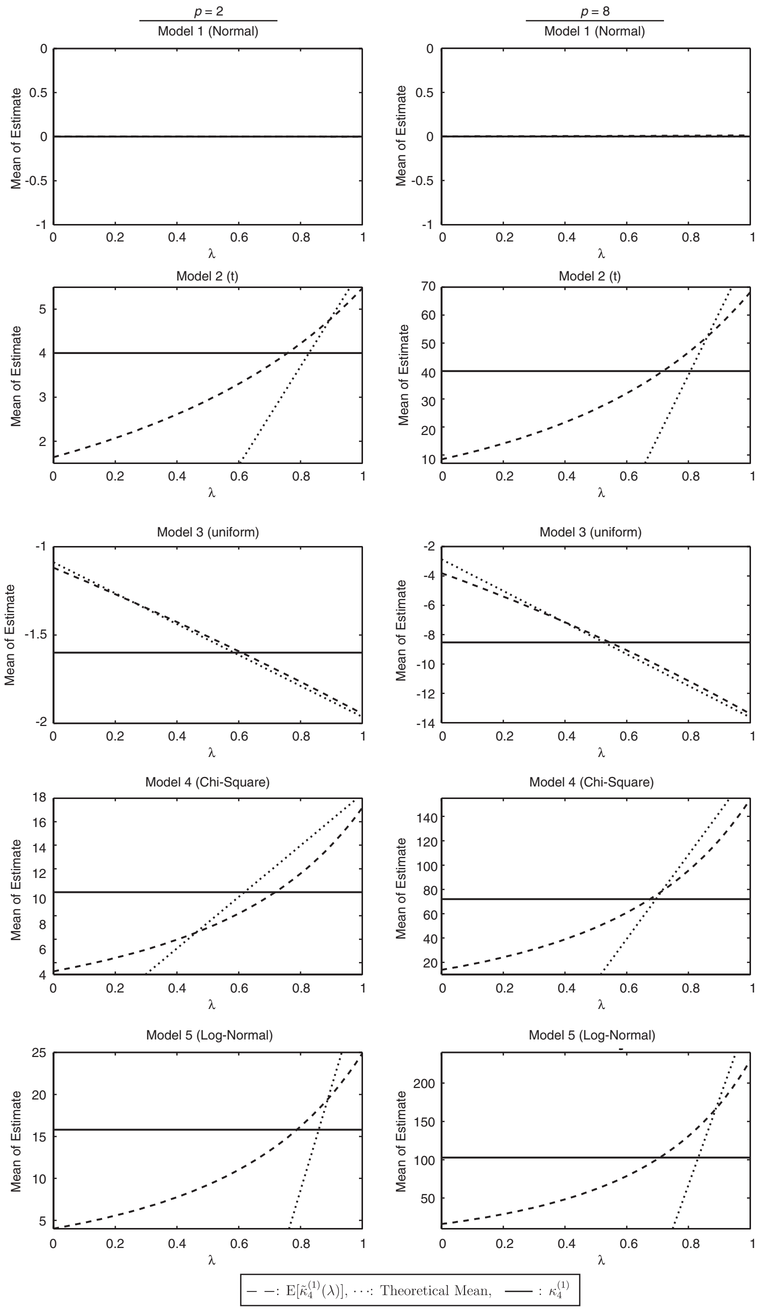

Fig. 2. Effects of λ and p on the mean of $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$ ($n = 30$).

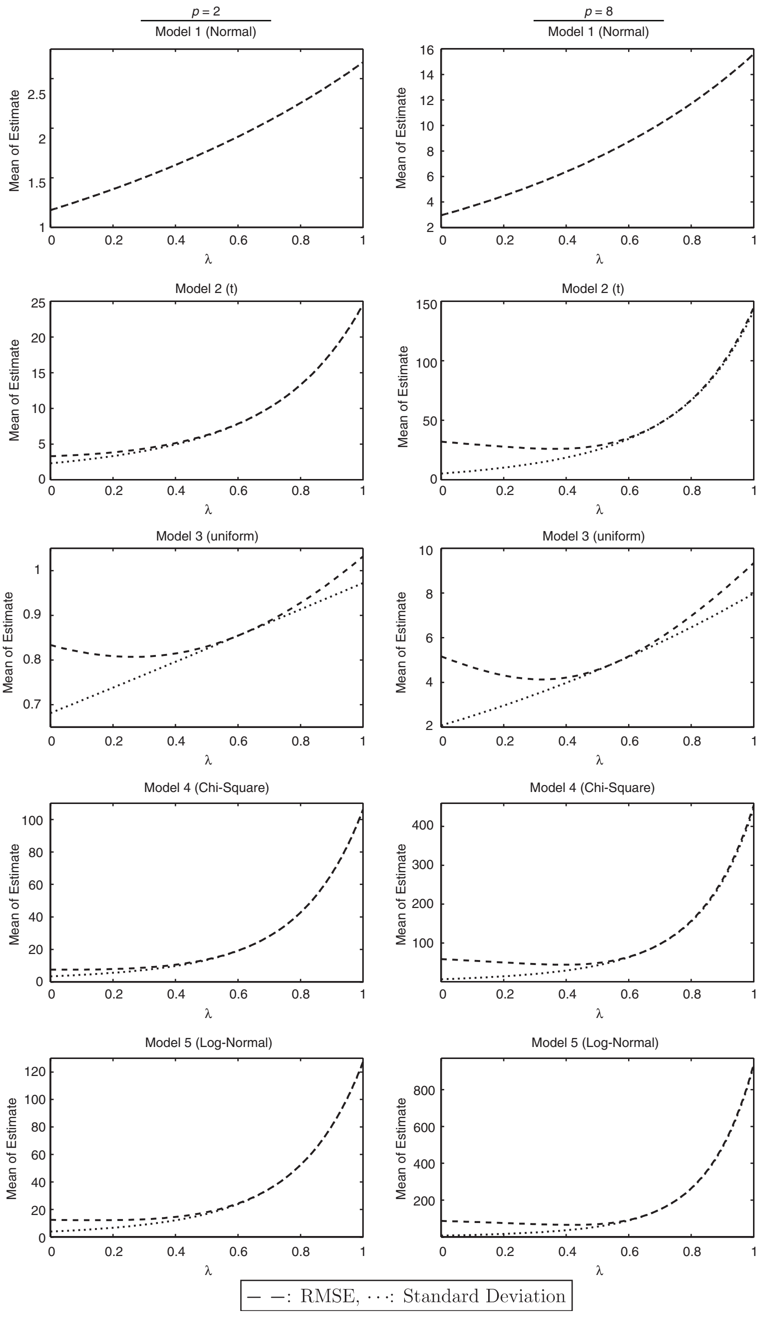

Fig. 3. Effects of λ and p on the standard deviations and RMSEs of $\mathrm{E}[\tilde{\kappa}_4^{(1)}(\lambda)]$ ($n = 30$).

$\tilde{\kappa}_4^{(1)}(1/2)$ has the smallest bias in most cases, although it gives an underestimation. Both estimators $\tilde{\kappa}_4^{(1)}(0)$ and $\hat{\kappa}_4^{(1)}$ give similar estimates, but the bias of $\tilde{\kappa}_4^{(1)}(0)$ is smaller than that of $\hat{\kappa}_4^{(1)}$ when $\kappa_4^{(1)} > 0$. Finally, we studied the means, standard deviations, and RMSEs of $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$ from 30,000 times iteration in the cases $p = 2$ and $p = 8$, which are shown in Tables 3a and b. We used $n \times 3$ and $n \times 5$ explanatory variable matrices, whose first columns were $\mathbf{1}_n$ and next columns were generated by $\mathrm{U}(-1, 1)$. From this table, we can see that, if the number of explanatory variables k is larger, the bias is larger. The tendencies of λ and p for the bias, standard deviation and RMSE are almost same as those in Figs. 2 and 3. By comparing Table 1 with Tables 3a and b, we can see that $\hat{\kappa}_4^{(1)}$ has the smallest standard deviation compared to the other estimators. However, the bias and RMSE of $\hat{\kappa}_4^{(1)}$ are larger than the ones of $\tilde{\kappa}_4^{(1)}(0)$ when $\kappa_4^{(1)} > 0$. In addition, we notice that the standard deviation of $\tilde{\kappa}_4^{(1)}(0)$ is larger if the sample size n is larger, although the ones of $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$ are smaller. This points out that the distribution of $\tilde{\kappa}_4^{(1)}(0)$ is almost same as the one of $\hat{\kappa}_4^{(1)}$.

5. Conclusion

We proposed a new estimator of kurtosis (15) by replacing the squared sample Mahalanobis distance with the transformed squared norm of multivariate internally Studentized residuals (7) using a monotonic increasing function (14). The correction term $c(\lambda)$ (16) adjusts our estimator to an unbiased estimator under normality. In a general case, we can make a nearly unbiased estimator by controlling the tuning parameter λ. Furthermore, we verified the reduction of bias by constructing simulation experiments. However, one serious problem, which is how to choose the tuning parameter λ, remains. The solution is very difficult. One of solutions for this problem is that we chose λ to make the theoretical bias (18) 0. But, if we use this method for the optimization of λ, it is necessary to estimate more higher-order cumulants. It is more difficult to estimate higher-order cumulants than the kurtosis. Moreover, although we can choose the best λ to make the bias of $\tilde{\kappa}_4^{(1)}(\lambda)$ close to 0, its variance becomes large.

Therefore, we recommend carefully choosing a λ that corresponds to each situation. $\tilde{\kappa}_4^{(1)}(1)$ may be used for testing normality, because it is very sensitive for nonnormality. The use of $\tilde{\kappa}_4^{(1)}(1)$ is supported by univariate research. In the univariate case, Imon [11–13] devised a testing method for normality by using a statistic similar to $\tilde{\kappa}_4^{(1)}(1)$, that is, a statistic based on the predicted residuals (8). Moreover, $\tilde{\kappa}_4^{(1)}(0)$ should be used if a large variance is not preferred, e.g., correction of the distribution of test statistics ([32] etc.). If it is considered only as an adjustment for the bias, the chosen value of λ should be around 0.5–0.7. When the RMSE is taken into account, we recommend using $\tilde{\kappa}_4^{(1)}(1/2)$.

In any case, the estimator of kurtosis must be unbiased at least under normality, because it is a measure for nonnormality. It is out of the question for the estimator to have a large bias, even though the data has normal distribution. On the other hand, there is another measure for kurtosis proposed by Koziol [20]. However, it does not become an estimator of kurtosis

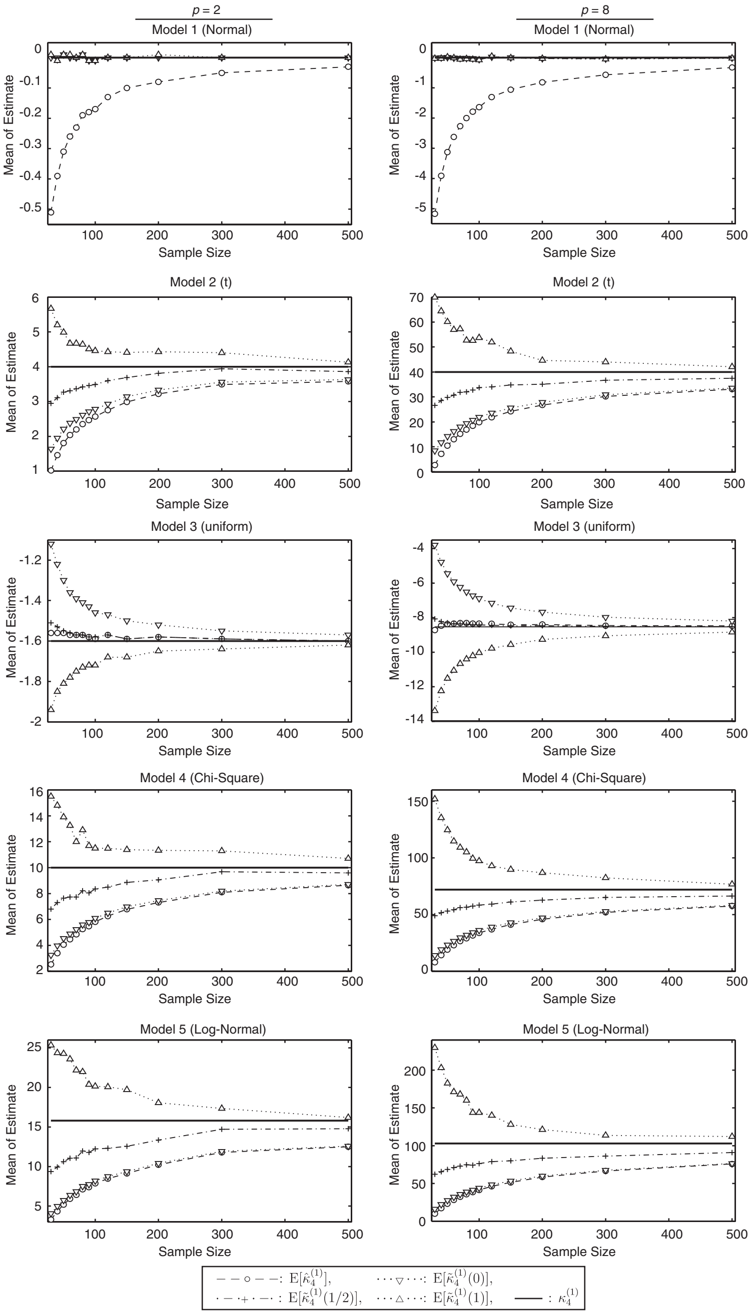

Fig. 4. Convergence of several estimators.

Table 3a
Means, standard deviations, and RMSEs for $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$, $p = 2$

				Models				
k	n			1	2	3	4	5
		$\kappa_4^{(1)}$		0.00	4.00	−1.60	10.00	15.80
3	30		Mean	0.01	1.36	−0.94	2.78	3.45
		$\tilde{\kappa}_4^{(1)}(0)$	(SD)	(1.17)	(2.09)	(0.77)	(3.02)	(3.51)
			(RMSE)	(1.17)	(3.37)	(1.02)	(7.83)	(12.83)
			Mean	0.01	2.48	−1.30	6.00	8.27
		$\tilde{\kappa}_4^{(1)}(1/2)$	(SD)	(1.80)	(5.19)	(0.98)	(11.04)	(16.25)
			(RMSE)	(1.80)	(5.40)	(1.03)	(11.74)	(17.91)
			Mean	0.01	4.55	−1.72	14.23	23.99
		$\tilde{\kappa}_4^{(1)}(1)$	(SD)	(2.73)	(16.10)	(1.23)	(69.24)	(166.44)
			(RMSE)	(2.73)	(16.11)	(1.23)	(69.37)	(166.64)
	50		Mean	−0.01	1.97	−1.18	4.09	5.25
		$\tilde{\kappa}_4^{(1)}(0)$	(SD)	(0.99)	(2.40)	(0.55)	(3.87)	(4.75)
			(RMSE)	(0.99)	(3.14)	(0.69)	(7.06)	(11.57)
			Mean	−0.01	2.92	−1.42	6.92	9.70
		$\tilde{\kappa}_4^{(1)}(1/2)$	(SD)	(1.28)	(5.36)	(0.61)	(10.15)	(15.89)
			(RMSE)	(1.28)	(5.46)	(0.64)	(10.61)	(17.02)
			Mean	−0.01	4.52	−1.68	12.38	20.48
		$\tilde{\kappa}_4^{(1)}(1)$	(SD)	(1.66)	(29.62)	(0.68)	(32.27)	(79.32)
			(RMSE)	(1.66)	(29.63)	(0.68)	(32.35)	(79.46)
5	30		Mean	0.00	1.13	−0.77	2.28	2.84
		$\tilde{\kappa}_4^{(1)}(0)$	(SD)	(1.15)	(1.86)	(0.83)	(2.62)	(3.00)
			(RMSE)	(1.15)	(3.42)	(1.17)	(8.15)	(13.30)
			Mean	0.01	2.12	−1.11	5.06	6.93
		$\tilde{\kappa}_4^{(1)}(1/2)$	(SD)	(1.86)	(4.58)	(1.13)	(9.63)	(13.66)
			(RMSE)	(1.86)	(4.95)	(1.24)	(10.82)	(16.28)
			Mean	0.02	3.93	−1.50	12.14	19.89
		$\tilde{\kappa}_4^{(1)}(1)$	(SD)	(3.00)	(14.05)	(1.52)	(62.78)	(122.31)
			(RMSE)	(3.00)	(14.05)	(1.52)	(62.81)	(122.38)
	50		Mean	−0.00	1.80	−1.07	3.71	4.75
		$\tilde{\kappa}_4^{(1)}(0)$	(SD)	(0.99)	(2.28)	(0.59)	(3.58)	(4.37)
			(RMSE)	(0.99)	(3.16)	(0.79)	(7.24)	(11.88)
			Mean	−0.00	2.71	−1.30	6.38	8.87
		$\tilde{\kappa}_4^{(1)}(1/2)$	(SD)	(1.28)	(4.69)	(0.67)	(10.38)	(14.71)
			(RMSE)	(1.28)	(4.86)	(0.74)	(11.00)	(16.26)
			Mean	−0.00	4.16	−1.55	11.92	18.87
		$\tilde{\kappa}_4^{(1)}(1)$	(SD)	(1.65)	(11.65)	(0.76)	(47.16)	(71.80)
			(RMSE)	(1.65)	(11.65)	(0.76)	(47.20)	(71.86)

Table 3b
Means, standard deviations, and RMSEs for $\tilde{\kappa}_4^{(1)}(0)$, $\tilde{\kappa}_4^{(1)}(1/2)$ and $\tilde{\kappa}_4^{(1)}(1)$, $p = 8$

				Models				
k	n			1	2	3	4	5
		$\kappa_4^{(1)}$		0.00	40.00	−8.53	72.00	102.92
3	30	$\tilde{\kappa}_4^{(1)}(0)$	Mean	−0.01	6.79	−3.03	11.25	12.94
			(SD)	(2.83)	(4.67)	(2.20)	(5.70)	(6.23)
			(RMSE)	(2.83)	(33.53)	(5.92)	(61.01)	(90.19)
		$\tilde{\kappa}_4^{(1)}(1/2)$	Mean	−0.03	23.02	−6.92	42.37	53.65
			(SD)	(7.62)	(23.20)	(5.30)	(37.84)	(51.46)
			(RMSE)	(7.62)	(28.75)	(5.54)	(48.06)	(71.25)
		$\tilde{\kappa}_4^{(1)}(1)$	Mean	−0.06	61.54	−11.88	136.44	209.10
			(SD)	(16.48)	(133.07)	(10.21)	(393.86)	(915.82)
			(RMSE)	(16.48)	(134.80)	(10.75)	(399.10)	(921.95)
	50	$\tilde{\kappa}_4^{(1)}(0)$	Mean	0.01	12.73	−4.88	20.88	24.76
			(SD)	(2.72)	(6.57)	(1.74)	(8.62)	(9.86)
			(RMSE)	(2.72)	(28.05)	(4.04)	(51.84)	(78.78)
		$\tilde{\kappa}_4^{(1)}(1/2)$	Mean	0.02	27.32	−7.61	48.83	63.23
			(SD)	(4.68)	(22.06)	(2.65)	(35.44)	(51.92)
			(RMSE)	(4.68)	(25.44)	(2.81)	(42.34)	(65.35)
		$\tilde{\kappa}_4^{(1)}(1)$	Mean	0.03	55.69	−10.72	113.05	175.19
			(SD)	(7.34)	(101.63)	(3.71)	(218.03)	(856.47)
			(RMSE)	(7.34)	(102.83)	(4.30)	(221.86)	(859.51)
5	30	$\tilde{\kappa}_4^{(1)}(0)$	Mean	−0.01	5.37	−2.39	8.86	10.30
			(SD)	(2.81)	(4.21)	(2.31)	(4.99)	(5.37)
			(RMSE)	(2.81)	(34.88)	(6.56)	(63.34)	(92.77)
		$\tilde{\kappa}_4^{(1)}(1/2)$	Mean	−0.01	19.68	−5.87	35.90	45.93
			(SD)	(8.13)	(21.89)	(6.09)	(33.07)	(43.66)
			(RMSE)	(8.13)	(29.87)	(6.65)	(48.96)	(71.79)
		$\tilde{\kappa}_4^{(1)}(1)$	Mean	0.01	54.92	−10.44	116.22	175.59
			(SD)	(19.13)	(159.23)	(12.68)	(297.36)	(624.50)
			(RMSE)	(19.13)	(159.92)	(12.83)	(300.63)	(628.71)
	50	$\tilde{\kappa}_4^{(1)}(0)$	Mean	−0.01	11.40	−4.34	18.60	22.13
			(SD)	(2.71)	(6.09)	(1.83)	(7.84)	(8.96)
			(RMSE)	(2.71)	(29.24)	(4.57)	(53.97)	(81.28)
		$\tilde{\kappa}_4^{(1)}(1/2)$	Mean	−0.01	25.06	−6.94	44.44	57.86
			(SD)	(4.75)	(20.62)	(2.89)	(32.42)	(46.79)
			(RMSE)	(4.75)	(25.46)	(3.30)	(42.55)	(64.96)
		$\tilde{\kappa}_4^{(1)}(1)$	Mean	−0.01	51.73	−9.93	103.52	159.71
			(SD)	(7.59)	(91.05)	(4.15)	(204.67)	(547.28)
			(RMSE)	(7.59)	(91.80)	(4.37)	(207.08)	(550.22)

in the multivariate case. The reason for this is described in Appendix A.5. Moreover, we can correct the bias of Mardia's estimator by using an asymptotic expansion of bias (12) as well as Maruyama and Seo [22]. However, this method is not realistic, because it is necessary to estimate more higher-order cumulants. Also, we can use the bootstrap method for the correction of bias (see e.g., [4, p. 138]). However, this method does not work well when we use $\hat{\kappa}_4^{(1)}$, because $\hat{\kappa}_4^{(1)}$ does not give a large value for any distribution. Therefore, the amount of reduced bias in the bootstrap correction becomes much smaller than what we expect. Considering these points, our estimator is better than ordinary estimators.

Acknowledgements

The author thank two referees and the associate editor for their helpful comments.

Appendices

A.1. Asymptotic expansion of $\mathrm{E}[b_{2,p}]$

In this section, we obtain the asymptotic expansion of $\mathrm{E}[b_{2,p}]$ up to the order n^{-1}. Let

$$\boldsymbol{Z} = (\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{X}'\mathscr{E},\ \boldsymbol{V} = \frac{1}{\sqrt{n}}(\mathscr{E}'\mathscr{E} - n\boldsymbol{I}_p),\ \boldsymbol{q}_i = \sqrt{n}(\boldsymbol{X}'\boldsymbol{X})^{-1/2}\boldsymbol{x}_i,$$

then

$$\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right) = \boldsymbol{\varepsilon}_i - \frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{q}_i,$$
$$\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{I}_p - \frac{1}{\sqrt{n}}\boldsymbol{V} + \frac{1}{n}(\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{V}^2) + \mathrm{O}_p(n^{-3/2}).$$

Therefore, $b_{2,p}$ is expanded as

$$b_{2,p} = W_0 + \frac{1}{\sqrt{n}}W_1 + \frac{1}{n}W_2 + \mathrm{O}_p(n^{-3/2}),$$

where

$$\begin{aligned}
W_0 &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i)^2,\\
W_1 &= -\frac{2}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + 2\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right\},\\
W_2 &= \frac{1}{n}\sum_{i=1}^{n}\Big\{(\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i)^2 + 2\left(\boldsymbol{\varepsilon}_i'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i'\boldsymbol{V}^2\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{q}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right)\\
&\quad +4\left(\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{V}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i + \boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i + \boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{Z}'\boldsymbol{q}_i\right)\Big\}.
\end{aligned}$$

Note that

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right)^2\right] &= \mu_4^{(1)} = \kappa_4^{(1)} + p(p+2), \\ \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] &= \frac{1}{\sqrt{n}}\left(\mu_6^{(1)} - \mu_4^{(1)}\right) \\ &= \frac{1}{\sqrt{n}}\left\{\kappa_6^{(1)} + 2\left(2\kappa_{3,3}^{(1)} + 3\kappa_{3,3}^{(2)}\right)\right. \\ &\quad \left. +(3p+11)\kappa_4^{(1)} + p(p+2)(p+3)\right\}, \end{aligned}$$

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] = \frac{\mu_4^{(1)}}{n\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{q}_i = \frac{k}{\sqrt{n}}\left\{\kappa_4^{(1)} + p(p+2)\right\},$$

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\varepsilon}_i'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] &= \frac{\mu_4^{(1)}}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{q}_i + \mathrm{O}(n^{-1}) \\ &= k\left\{\kappa_4^{(1)} + p(p+2)\right\} + \mathrm{O}(n^{-1}), \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\varepsilon}_i'\boldsymbol{V}^2\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] &= \mu_{4,4}^{(2)} - \mu_4^{(1)} + \mathrm{O}(n^{-1}) \\ &= \kappa_{4,4}^{(2)} + (2p+3)\kappa_4^{(1)} + p(p+1)(p+2) + \mathrm{O}(n^{-1}), \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{V}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] &= \frac{\mu_{3,3}^{(2)}}{n^2}\sum_{ij}^{n}\boldsymbol{q}_i'\boldsymbol{q}_j + \mathrm{O}(n^{-1}) = a_1\kappa_{3,3}^{(2)} + \mathrm{O}(n^{-1}), \\ \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{q}_i\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right] &= \frac{p^2}{n^2}\sum_{ij}^{n}\left(\boldsymbol{q}_i'\boldsymbol{q}_j\right)^2 + \mathrm{O}(n^{-1}) = p^2k + \mathrm{O}(n^{-1}), \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i\right)^2\right] &= \mu_{4,4}^{(1)} - \mu_4^{(1)} + \mathrm{O}(n^{-1}) \\ &= \kappa_{4,4}^{(1)} + 5\kappa_4^{(1)} + 2p(p+2) + \mathrm{O}(n^{-1}), \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{V}\boldsymbol{\varepsilon}_i\right] &= \frac{\mu_{3,3}^{(1)}}{n^2}\sum_{ij}^{n}\boldsymbol{q}_i'\boldsymbol{q}_j + \mathrm{O}(n^{-1}) = a_1\kappa_{3,3}^{(1)} + \mathrm{O}(n^{-1}), \\ \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i'\boldsymbol{Z}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{Z}'\boldsymbol{q}_i\right] &= \frac{p}{n^2}\sum_{ij}^{n}\left(\boldsymbol{q}_i'\boldsymbol{q}_j\right)^2 + \mathrm{O}(n^{-1}) = kp + \mathrm{O}(n^{-1}). \end{aligned}$$

From the above expectations, we can obtain the following expectations.

$$\begin{aligned}
\mathrm{E}[W_0] &= \kappa_4^{(1)} + p(p+2),\\
\mathrm{E}[W_1] &= -\frac{2}{\sqrt{n}}\left\{\kappa_6^{(1)} + 2\left(2\kappa_{3,3}^{(1)} + 3\kappa_{3,3}^{(2)}\right) + (3p+2k+11)\kappa_4^{(1)}\right.\\
&\quad \left. + p(p+2)(p+2k+3)\right\},\\
\mathrm{E}[W_2] &= \kappa_{4,4}^{(1)} + 2\kappa_{4,4}^{(2)} + 4a_1\left(\kappa_{3,3}^{(1)} + \kappa_{3,3}^{(2)}\right) + (4p+2k+11)\kappa_4^{(1)}\\
&\quad + 2p(p+2)(p+2k+2) + \mathrm{O}(n^{-1}),
\end{aligned}$$

where a_1 is given by (13). Therefore, we obtain the asymptotic expansion of $\mathrm{E}[b_{2,p}]$ as the following theorem.

Theorem A.1. *Suppose that assumptions* 1, 2, 3 *and* 4 *are held. Then* $\mathrm{E}[b_{2,p}]$ *is expanded up to the order* n^{-1} *as*

$$\begin{aligned}
\mathrm{E}[b_{2,p}] &= \kappa_4^{(1)} + p(p+2) + \frac{1}{n}\left[\kappa_{4,4}^{(1)} + 2\kappa_{4,4}^{(2)} - 2\kappa_6^{(1)} + 4\left\{(a_1-2)\kappa_{3,3}^{(1)}\right.\right.\\
&\quad \left.\left. + (a_1-3)\kappa_{3,3}^{(2)}\right\} - (2p+2k+11)\kappa_4^{(1)} - 2p(p+2)\right] + \mathrm{O}(n^{-2}).
\end{aligned} \tag{A.1}$$

A.2. On predicted residuals

From the Lemma 3.1 in Fujikoshi et al. [8], we obtain the following lemma.

Lemma A.1. *Suppose that* $\boldsymbol{Y} \sim \mathrm{N}_{n\times p}(\boldsymbol{X\Theta}, \boldsymbol{\Sigma}\otimes\boldsymbol{I}_n)$. *Then*

(i) $\frac{1}{\sqrt{1-(\boldsymbol{P_X})_{ii}}}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right) \sim \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$,

(ii) $(n-k-1)\boldsymbol{S}_{[-i]} \sim \mathrm{W}_p(n-k-1, \boldsymbol{\Sigma})$,

(iii) $\frac{1}{\sqrt{1-(\boldsymbol{P_X})_{ii}}}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right)$ *and* $\boldsymbol{S}_{[-i]}$ *are independent of each other*.

From this Lemma A.1, we derive the following theorem.

Theorem A.2. *Suppose that* $\boldsymbol{Y} \sim \mathrm{N}_{n\times p}(\boldsymbol{X\Theta}, \boldsymbol{\Sigma}\otimes\boldsymbol{I}_n)$. *Let* U_i *denote the squared norm of the ith externally multivariate Studentized residual, i.e.*,

$$U_i = \tilde{\boldsymbol{\varepsilon}}_{i[-i]}'\tilde{\boldsymbol{\varepsilon}}_{i[-i]}. \tag{A.2}$$

Then, each U_i *is distributed according to Hotelling's* T^2 *distribution with* $N = n-k-1$ *degrees of freedom, whose probability density is*

$$g(u; N, p) = \frac{\Gamma\left(\frac{N+1}{2}\right)}{N\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p+1}{2}\right)}\left(\frac{u}{N}\right)^{p/2-1}\left(1+\frac{u}{N}\right)^{-(N+1)/2},$$

where $\Gamma(\cdot)$ *is the Gamma function.*

Proof. Let

$$z_i = \frac{1}{\sqrt{1-(\boldsymbol{P_X})_{ii}}}\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i\right), \quad \boldsymbol{W}_i = N\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}_{[-i]}\boldsymbol{\Sigma}^{-1/2}.$$

Then $U_i = Nz_i'\boldsymbol{W}_i^{-1}z_i$. By recalling Lemma A.1, we can see that z_i and $\boldsymbol{W}_i$ are mutually independent, and distributed according to $\mathrm{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $\mathrm{W}_p(N, \boldsymbol{I}_p)$, respectively. The general result in a multivariate distribution (see, e.g., [26, p. 190]) shows that $Nz_i'\boldsymbol{W}_i^{-1}z_i$ is distributed according to Hotelling's T^2 distribution with N degrees of freedom. Therefore, we obtain the result in Theorem A.2. □

On the other hand, there are the following relations between U_i and the squared norm of $\hat{\boldsymbol{\varepsilon}}_i$ in (6), and U_i and the squared norm of $\tilde{\boldsymbol{\varepsilon}}_i$ in (7), respectively.

Theorem A.3. *The squared norm of the ith externally Studentized residual* U_i (A.2) *can be rewritten by* $\hat{\boldsymbol{\varepsilon}}_i$ *or* $\tilde{\boldsymbol{\varepsilon}}_i$ *as*

$$U_i = \frac{N\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i/\{1-(\boldsymbol{P_X})_{ii}\}}{n\left[1-\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i/n\{1-(\boldsymbol{P_X})_{ii}\}\right]}, \quad U_i = \frac{N\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i}{(n-k)\left\{1-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i/(n-k)\right\}}. \tag{A.3}$$

Proof. From Fujikoshi et al. [8], we can see that

$$\begin{aligned}\boldsymbol{S}_{[-i]} &= \frac{n}{N}\hat{\boldsymbol{\Sigma}}^{1/2}\left[\boldsymbol{I}_p - \frac{1}{n\{1-(\boldsymbol{P_X})_{ii}\}}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\right]\hat{\boldsymbol{\Sigma}}^{1/2}\\ &= \frac{n-k}{N}\boldsymbol{S}^{1/2}\left(\boldsymbol{I}_p - \frac{1}{n-k}\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i'\right)\boldsymbol{S}^{1/2}.\end{aligned}$$

Hence

$$\begin{aligned}\boldsymbol{S}_{[-i]}^{-1} &= \frac{N}{n}\hat{\boldsymbol{\Sigma}}^{-1/2}\left[\boldsymbol{I}_p + \frac{\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'}{n\{1-(\boldsymbol{P_X})_{ii}\} - \hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i}\right]\hat{\boldsymbol{\Sigma}}^{-1/2}\\ &= \frac{N}{n-k}\boldsymbol{S}^{-1/2}\left(\boldsymbol{I}_p + \frac{\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i'}{n-k-\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i}\right)\boldsymbol{S}^{-1/2}.\end{aligned} \tag{A.4}$$

From (9), it notes that

$$U_i = \frac{1}{1-(\boldsymbol{P_X})_{ii}}(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i)'\boldsymbol{S}_{[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\Theta}}'\boldsymbol{x}_i).$$

Substituting Eqs. (A.4) into the above equation yields Theorem A.3. □

A.3. Exact mean of Mardia's estimator under normality

In this section, we give the proof of Theorem 2, i.e., the derivation of the exact mean of $\hat{\kappa}_4^{(1)}$ under normality. Let $T_i = \hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i/\{1-(\boldsymbol{P_X})_{ii}\}$. Note that $0 < T_i < n$. From (A.3), we can see that U_i (A.2) is a monotonic function of T_i in $0 < T_i < n$. Therefore,

$$T_i = \frac{n}{N} U_i \left(1 + \frac{U_i}{N}\right)^{-1}.$$

By using this equation, the expectation of T_i^2 is given by the following integral.

$$\mathrm{E}[T_i^2] = n^2 \int_0^\infty \frac{\Gamma\left(\frac{N+1}{2}\right)}{N\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p+1}{2}\right)} \left(\frac{u}{N}\right)^{(p+4)/2-1} \left(1 + \frac{u}{N}\right)^{-(N+5)/2} du.$$

Let $N_0 = N + 4$ and $p_0 = p + 4$. Note that

$$\Gamma\left(\frac{p_0}{2}\right) = \frac{p(p+2)}{4}\Gamma\left(\frac{p}{2}\right),$$
$$\Gamma\left(\frac{N_0+1}{2}\right) = \frac{(N+1)(N+3)}{4}\Gamma\left(\frac{N+1}{2}\right),$$
$$\Gamma\left(\frac{N_0-p_0+1}{2}\right) = \Gamma\left(\frac{N-p+1}{2}\right).$$

By using the above equations and replacing u/N with x/N_0, we derive

$$\mathrm{E}[T_i^2] = \frac{n^2 p(p+2)}{(N+1)(N+3)} \int_0^\infty g(x; N_0, p_0)dx = \frac{n^2 p(p+2)}{(N+1)(N+3)}.$$

Note that $b_{2,p} = n^{-1}\sum_{i=1}^n \{1 - (\boldsymbol{P_X})_{ii}\}^2 T_i^2$ and $n\sum_{i=1}^n \{1 - (\boldsymbol{P_X})_{ii}\}^2 = n^2 - 2kn + a_2$. Hence

$$\mathrm{E}[b_{2,p}] = \frac{(n^2 - 2kn + a_2)}{(n-k)(n-k+2)} p(p+2).$$

Recalling from equation $\hat{\kappa}_4^{(1)} = b_{2,p} - p(p+2)$, we obtain the result in Theorem 2.

A.4. Exact mean of $\hat{\kappa}_4^{(1)}(\lambda)$ under normality

In this section, we give the proof of Theorem 3, i.e., the derivation of an exact mean of $\tilde{\kappa}_4^{(1)}(\lambda)$ under normality. Let $\tilde{T}_i = \tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i$. Note that $0 < \tilde{T}_i < n - k$. From (A.3), we can see that U_i (A.2) is a monotonic function of $\tilde{T}_i$ in $0 < \tilde{T}_i < n - k$. Therefore,

$$\tilde{T}_i = \frac{n-k}{N} U_i \left(1 + \frac{U_i}{N}\right)^{-1}.$$

It makes

$$f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i; \lambda) = (n-k)\frac{U_i}{N}\left(1 + \frac{U_i}{N}\right)^{\lambda-1}.$$

Let $N_\lambda = N + 4(1-\lambda)$ and $p_0 = p + 4$. By using a method similar to that in Appendix A.3, the expectation of $\{f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i;\lambda)\}^2$ is given by

$$
\begin{aligned}
&\mathrm{E}\left[\left\{f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i;\lambda)\right\}^2\right]\\
&\quad = (n-k)^2\int_0^\infty \frac{\Gamma\left(\frac{N+1}{2}\right)}{N\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p+1}{2}\right)}\left(\frac{u}{N}\right)^{p_0/2-1}\left(1+\frac{u}{N}\right)^{-(N_\lambda+1)/2} du\\
&\quad = \frac{p(p+2)(n-k)^2\Gamma\left(\frac{N+1}{2}\right)\Gamma\left(\frac{N_\lambda-p_0+1}{2}\right)}{4\Gamma\left(\frac{N_\lambda+1}{2}\right)\Gamma\left(\frac{N-p+1}{2}\right)}.
\end{aligned}
$$

Note that $N_\lambda - p_0 + 1 = N - p - 4\lambda + 1$ and

$$
\Gamma\left(\frac{N_\lambda+1}{2}\right) = \frac{(N-4\lambda+1)(N-4\lambda+3)}{4}\Gamma\left(\frac{N-4\lambda+1}{2}\right).
$$

Recalling from $N = n-k-1$, we obtain

$$
\mathrm{E}\left[\left\{f(\tilde{\boldsymbol{\varepsilon}}_i'\tilde{\boldsymbol{\varepsilon}}_i;\lambda)\right\}^2\right] = \frac{p(p+2)}{c(\lambda)}.
$$

From the definition of $\tilde{\kappa}_4^{(1)}(\lambda)$ in (15), $\mathrm{E}\left[\tilde{\kappa}_4^{(1)}(\lambda)\right] = 0$ is derived.

A.5. Some comments on Koziol's measure

Koziol [20] proposed a variant measure of kurtosis as

$$
\tilde{b}_{2,p} = \frac{1}{n^2}\sum_{ij}^{n}(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_j)^4. \tag{A.5}
$$

The asymptotic distribution of this measure was studied by many authors, e.g., Henze [10] and Klar [19]. Suppose that assumptions 2, 3 and 4 are held. Then, the formula in (A.5) is expanded as

$$
\tilde{b}_{2,p} = \frac{1}{n^2}\sum_{ij}^{n}(\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j)^4 + \mathrm{O}_p(n^{-1/2}).
$$

From this expansion and the equation in (5), we obtain the asymptotic mean of $\tilde{b}_{2,p}$ as

$$
\mathrm{E}[\tilde{b}_{2,p}] = \mu_{4,4}^{(1)} + \mathrm{O}(n^{-1}) = \kappa_{4,4}^{(1)} + 6\kappa_4^{(1)} + 3p(p+2) + \mathrm{O}(n^{-1}).
$$

When $p = 1$, we rewrite $\kappa_4^{(1)}$ as κ_4. Then $\kappa_{4,4}^{(1)} = (\kappa_4+3)^2$. Note that $\tilde{b}_{2,p}$ is always positive. Therefore, we can see that $(\tilde{b}_{2,1})^{1/2} - 3$ becomes a consistent estimator of the kurtosis. However, $\{\kappa_4^{(1)}\}^2$ is not equivalent to $\kappa_{4,4}^{(1)}$ in the multivariate case. Therefore, $\tilde{b}_{2,p}$ does not become an estimator of $\kappa_4^{(1)}$ in the multivariate case.

References

[1] G.E.P. Box, Non-normality and tests on variances, Biometrika 40 (1953) 318–335.

[2] M.W. Browne, Asymptotically distribution-free methods for the analysis of covariance structures, British J. Math. Statist. Psych. 37 (1984) 62–83.

[3] R.D. Cook, S.W. Weisberg, Residuals and Influence in Regression, Chapman & Hall/CRC, New York, 1982.

[4] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, New York, 1993.

[5] Y. Fujikoshi, Transformations with improved chi-squared approximations, J. Multivariate Anal. 72 (2000) 249–263.

[6] Y. Fujikoshi, Asymptotic expansions for the distributions of multivariate basic statistics and one-way MANOVA tests under nonnormality, J. Statist. Plann. Inference 108 (2002) 263–282.

[7] Y. Fujikoshi, M. Ohmae, H. Yanagihara, Asymptotic approximations of the null distribution of the one-way ANOVA test statistic under nonnormality, J. Japan Statist. Soc. 29 (1999) 147–161.

[8] Y. Fujikoshi, T. Noguchi, M. Ohtaki, H. Yanagihara, Corrected versions of cross-validation criteria for selecting multivariate regression and growth curve models, Ann. Inst. Statist. Math. 55 (2003) 537–553.

[9] N. Henze, On Mardia's kurtosis test for multivariate normality, Comm. Statist. Theory Methods 23 (1994) 1031–1045.

[10] N. Henze, The asymptotic behavior of a variant of multivariate kurtosis, Comm. Statist. Theory Methods 23 (1994) 1047–1061.

[11] A.H.M.R. Imon, On PRESS residuals, J. Statist. Res. 33 (1999) 57–65.

[12] A.H.M.R. Imon, Modified residuals in tests for normality: a small sample study, J. Statist. Stud. 20 (2000) 21–25.

[13] A.H.M.R. Imon, Regression residuals, moments, and their use in tests for normality, Comm. Statist. Theory Methods 32 (2003) 1021–1034.

[14] T. Isogai, On a measure of multivariate skewness and a test for multivariate normality, Ann. Inst. Statist. Math. 34 (1982) 531–541.

[15] T. Isogai, On measures of multivariate skewness and kurtosis, Math. Japon. 28 (1983) 251–261.

[16] K. Ito, On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures, in: P.R. Krishnaiah (Ed.), Multivariate Analysis, vol. 2, Academic Press, New York, 1969, pp. 87–120.

[17] Y. Kano, Robust statistics for test-of-independence and related structural models, Statist. Probab. Lett. 15 (1992) 21–26.

[18] E.L. Kaplan, Tensor notation and the sampling cumulants of k-statistics, Biometrika 39 (1952) 319–323.

[19] B.A. Klar, Treatment of multivariate skewness, kurtosis, and related statistics, J. Multivariate Anal. 83 (2002) 141–165.

[20] J.A. Koziol, A note on measures of multivariate kurtosis, Biom. J. 31 (1989) 619–624.

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

[22] Y. Maruyama, T. Seo, Estimation of moment parameter in elliptical distributions, J. Japan Statist. Soc. 33 (2003) 215–229.

[23] R.J. Muirhead, C.M. Waternaux, Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, Biometrika 67 (1980) 31–43.

[24] F. Pukelsheim, Multilinear estimation of skewness and kurtosis in linear models, Metrika 27 (1980) 103–113.

[25] A. Satorra, P.M. Bentler, Scaling corrections for chi-square statistics in covariance structure analysis, in: Proceedings of the Business and Economic Statistics Section, American Statistical Association, Alexandria, 1988, pp. 308–313.

[26] M. Siotani, T. Hayakawa, Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, Columbus, OH, 1985.

[27] D.E. Tyler, Robustness and efficiency properties of scatter matrices, Biometrika 70 (1983) 411–420.

[28] H. Wakaki, H. Yanagihara, Y. Fujikoshi, Asymptotic expansion of the null distributions of test statistics for multivariate linear hypothesis under nonnormality, Hiroshima Math. J. 32 (2002) 17–50.

[29] H. Yanagihara, Asymptotic expansion of the null distribution of one-way ANOVA test statistic for heteroscedastic case under nonnormality, Comm. Statist. Theory Methods 29 (2000) 463–476.

[30] H. Yanagihara, Asymptotic expansions of the null distributions of three test statistics in a nonnormal GMANOVA model, Hiroshima Math. J. 31 (2001) 213–262.

[31] H. Yanagihara, Asymptotic expansion of the null distribution of test statistic for linear hypothesis in nonnormal linear model, J. Multivariate Anal. 84 (2003) 222–246.

[32] H. Yanagihara, K.-H. Yuan, Four improved statistics for contrasting means by correcting skewness and kurtosis, Brit. J. Math. Statist. Psych., in press.

[33] H. Yanagihara, T. Tonda, C. Matsumoto, The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption, J. Multivariate Anal., in press.

[34] K.-H. Yuan, P.M. Bentler, Generating multivariate distributions with specified marginal skewness and kurtosis, in: W. Bandilla, F. Faulbaum (Eds.), SoftStat' 97–Advances in Statistical Software 6–, Lucius and Lucius, Stuttgart, Germany, 1997, pp. 385–391.

[35] K.-H. Yuan, P.M. Bentler, On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions, Statist. Sinica 9 (1999) 831–853.