Tests of multinormality based on location vectors and scatter matrices

Annaliisa Kankainen[1]**, Sara Taskinen**[1]**, Hannu Oja**[2]

[1] Department of Mathematics and Statistics, FIN-40014 University of Jyväskylä, Finland, e-mail: kankaine@maths.jyu.fi, slahola@maths.jyu.fi
[2] Tampere School of Public Health, FIN-33014 University of Tampere, Finland, e-mail: Hannu.Oja@uta.fi

Abstract   Classical univariate measures of asymmetry such as Pearson's (mean-median)$/\sigma$ or (mean-mode)$/\sigma$ often measure the standardized distance between two separate location parameters and have been widely used in assessing univariate normality. Similarly, measures of univariate kurtosis are often just ratios of two scale measures. The classical standardized fourth moment and the ratio of the mean deviation to the standard deviation serve as examples. In this paper we consider tests of multinormality which are based on the Mahalanobis distance between two multivariate location vector estimates or on the (matrix) distance between two scatter matrix estimates, respectively. Asymptotic theory is developed to provide approximate null distributions as well as to consider asymptotic effiencies. Limiting Pitman efficiencies for contiguous sequences of contaminated normal distributions are calculated and the efficiencies are compared to those of the classical tests by Mardia. Simulations are used to compare finite sample efficiencies. The theory is also illustrated by an example.

Key words   Affine invariance–Kurtosis–Pitman efficiency–Skewness

1 Introduction

Classical multivariate analysis is mostly based on the assumption that the data come from a multivariate normal distribution. The tests of multinormality have therefore received much attention and several tests for assessing multinormality have been proposed in the literature. See e.g. Gnanadesikan (1977), Cox and Small (1978), Mardia (1980) and Romeu and Ozturk (1993), for reviews. If the hypothesis of multinormality is then rejected, alternative robust and nonparametric multivariate procedures can be used.

In the univariate case, standardized third and fourth moments are often used to indicate the skewness and kurtosis. The standardization is then made using the sample mean and sample variance. In this spirit, if $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is a random sample from a k-variate normal distribution $N(\boldsymbol{\mu}, \Sigma)$, the first step in assessing multivariate normality is to standardize the data vectors using the sample mean vector and sample covariance matrix

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum \boldsymbol{x}_i \;\; \text{and} \;\; S = \frac{1}{n-1}\sum(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T,$$

the regular estimates of $\boldsymbol{\mu}$ and Σ, respectively. Let then $\boldsymbol{z}_i = S^{-1/2}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$, be the *scaled residual*, and let

$$r_i = ||\boldsymbol{z}_i|| \;\; \text{and} \;\; u_i = ||\boldsymbol{z}_i||^{-1}\boldsymbol{z}_i$$

be its radius and the direction vector, $i = 1, \ldots, n$. Note that r_i^2 is the squared *Mahalanobis distance* between $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}$. Moreover, $r_{ij} = \boldsymbol{z}_i^T \boldsymbol{z}_j$ is known as the cosine of the *Mahalanobis angle* between the vectors $\boldsymbol{x}_i - \bar{\boldsymbol{x}}$ and $\boldsymbol{x}_j - \bar{\boldsymbol{x}}$. Mardia (1970) then defined multivariate measures of skewness and kurtosis by

$$b_1 = \text{ave}_{i,j}\{r_{ij}^3\} \;\; \text{and} \;\; b_2 = \text{ave}_i\{r_i^4\}.$$

Under the hypothesis of multinormality, the limiting distribution of $nb_1/6$ is a chi-square distribution with $k(k+1)(k+2)/6$ degrees of freedom and the limiting distribution of $(b_2 - k(k+2))/(8k(k+2)/n)^{1/2}$ is $N(0,1)$. For different definitions of multivariate skewness and kurtosis, see Malkovich and Afifi (1973) and Koziol (1986,1987).

Bera and John (1983) used certain third and forth moments of the scaled residuals,

$$T_r = \text{ave}_i\{z_{ri}^3\} \;\; \text{and} \;\; T_{rs} = \text{ave}_i\{z_{ri}^2 z_{si}^2\}, \;\; r,s = 1, \ldots, k,$$

to construct four different test statistics for multinormality and compared them with Mardia's skewness and kurtosis tests under multivariate Pearson family of distributions. Also Koziol (1982,1983,1986,1987,1993) used third and fourth moments to construct tests for multinormality. He noted that the smooth tests may be decomposed into individual components, which are distributed as independent χ^2 random variables and compared those components under different kinds of alternatives.

Classical univariate measures of asymmetry, like Pearson's β_1, are of the form (mean-median)/σ or (mean-mode)/σ and measure the distance between two location parameters (Pearson, 1895). These measures have been widely used in assessing univariate normality, see MacGillivray (1986) for survey of work on skewness. On the other hand, measures of univariate kurtosis are often just ratios of two scale measures. Examples include classical kurtosis measure, standardized fourth moment and the ratio of the

mean deviation to the standard deviation (Geary, 1935), among others. For measures of univariate skewness and kurtosis, see also Oja (1981).

In the paper we construct new test statistics for multinormality which are based on multivariate location vectors and scatter matrices. Our plan is as follows. In Section 2, location vectors and scatter matrices are introduced. We also give the general form of their influence functions and asymptotic variances and present some affine equivariant location and scatter matrix estimators. In Section 3, new statistics for testing multinormality are presented and their large sample properties are studied. Asymptotic efficiencies are given in Section 4 and in Section 5, simulations are used to compare the finite sample powers of tests. In Section 6, the theory is illustrated by an example and the paper is concluded with some comments in Section 7.

2 Location vectors and scatter matrices

Let $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a random sample from a k-variate distribution with cdf F. As usually, write F_n for the empirical cdf. Our tests of multinormality are based on two separate location vector or scatter matrix estimates. A k-vector valued statistic $T = T(X)$ is a *location vector* if it is affine equivariant, that is, if

$$T(A\boldsymbol{x}_1 + \boldsymbol{b}, \ldots, A\boldsymbol{x}_n + \boldsymbol{b}) \;=\; A\,T(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \;+\; \boldsymbol{b}.$$

We assume that it is possible to define (in a 'large' family of distributions) a corresponding statistical functional $T(F)$ so that $T(F_n) = T(X)$. A $k \times k$-matrix valued statistic $C = C(X)$ is a *scatter matrix* if it is affine equivariant, which now means, that

$$C(A\boldsymbol{x}_1 + \boldsymbol{b}, \ldots, A\boldsymbol{x}_n + \boldsymbol{b}) \;=\; A\,C(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\,A^T.$$

Assume again that a corresponding statistical functional $C(F)$ can be defined so that $C(F_n) = C(X)$.

It is easy to see that, if F is elliptic with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ (and they exist), then

$$T(F) = \boldsymbol{\mu} \;\; \text{and} \;\; C(F) = c_F \Sigma,$$

where c_F depends on distribution F and scatter functional C. If multinormality is assumed, then c_F is known for each C and correction factor can be used to guarantee the consistency to the regular covariance matrix. To conclude, under the multinormality, all location vectors estimate the same quantity $\boldsymbol{\mu}$ and all scatter matrices equipped with the correction factor estimate the covariance matrix Σ.

Under multinormality, the *influence functions* of the location and scatter functionals have simple expressions: If $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$, $r = ||\boldsymbol{z}||$ and $\boldsymbol{u} = ||\boldsymbol{z}||^{-1}\boldsymbol{z}$, then

$$IF(\boldsymbol{x}; T, F) = \gamma_T(r)\Sigma^{1/2}\boldsymbol{u} \tag{1}$$

and

$$IF(\boldsymbol{x}; C, F) = \alpha_C(r) \Sigma^{1/2} \boldsymbol{u}\boldsymbol{u}^T \Sigma^{1/2} - \beta_C(r) \Sigma. \qquad (2)$$

For robust functionals T and C, functions $\gamma_T(r)$ and $\alpha_C(r)$ and $\beta_C(r)$ are continuous and bounded. In the paper we assume that, under multinormality, the estimates $T(X)$ and $C(X)$ are consistent to $\boldsymbol{\mu}$ and Σ, respectively, and that the limiting distributions of

$$\sqrt{n}(T(X) - \boldsymbol{\mu}) \quad \text{and} \quad \sqrt{n}vec(C(X) - \Sigma)$$

are multinormal. Then the asymptotic covariance matrices of $T(X)$ and $vec(C(X))$ are given by

$$\tau_1 \Sigma \quad \text{and} \quad \tau_2(I_{k^2} + I_{k.k})(\Sigma \otimes \Sigma) + \tau_3 vec(\Sigma)vec(\Sigma)^T$$

where

$$\tau_1 = \frac{E[\gamma_T^2(r)]}{k}, \quad \tau_2 = \frac{E[\alpha_C^2(r)]}{k(k+2)}$$

and

$$\tau_3 = \frac{E[\alpha_C^2(r)]}{k(k+2)} - \frac{2}{k}E[\alpha_C(r)\beta_C(r)] + E[\beta_C^2(r)].$$

See e.g. Tyler (1982).

   Besides sample mean and sample covariance matrix, several affine equivariant location vector and scatter matrix estimators are introduced in the literature. Location and scatter M-functionals (Maronna, 1976; Huber, 1981) are defined as functionals $T(F)$ and $C(F)$ which satisfy implicit equations

$$T(F) = [E[v_1(r)]]^{-1} E[v_1(r)\boldsymbol{x}]$$

and

$$C(F) = E[v_2(r)(\boldsymbol{x} - T(F))(\boldsymbol{x} - T(F))^T],$$

where

$$r^2 = ||\boldsymbol{x} - T(F)||_{C(F)}^2 = (\boldsymbol{x} - T(F))^T [C(F)]^{-1} (\boldsymbol{x} - T(F))$$

and v_1 and v_2 are real-valued functions on $[0, \infty)$. As an example of M-functionals, consider Huber's M-functionals $M(q)$ that use

$$v_1(r) = \begin{cases} 1, & r \le c \\ c/r, & r > c \end{cases} \quad \text{and} \quad \sigma v_2(r) = \begin{cases} 1, & r^2 \le c^2 \\ c^2/r^2, & r^2 > c^2, \end{cases}$$

where c is a tuning constant defined so that $q = Pr(\chi_k^2 \le c^2)$. The scaling factor σ is chosen so that $E[v_2(r)r^2] = k$, where r^2 is a random variable from a χ_k^2 distribution. Then $C(F) = \Sigma$ at multivariate normal population. The influence functions of location and scatter M-functionals are derived in Huber (1981; Section 8.7). Note also that choosing $v_1(r) = 1/r$ and $v_2(r) = k/r^2$ in M-estimation equations yield the transformation-retransformation

spatial median (Hettmansperger and Randles, 2002). The influence function of this median is given in Ollila et al. (2003).

If now $T_1(F)$ and $C_1(F)$ are any location and scatter functionals then one-step M-functionals of location and scatter, starting from $T_1(F)$ and $C_1(F)$, are given by

$$T_2(F) = [E[v_1(r)]]^{-1} E[v_1(r)\boldsymbol{x}]$$

and

$$C_2(F) = E[v_2(r)(\boldsymbol{x} - T_1(F))(\boldsymbol{x} - T_1(F))^T],$$

where $r^2 = ||\boldsymbol{x} - T_1(F)||^2_{C_1(F)}$. Again, to quarantee the consistency of $C_2(F)$ to Σ at multivariate normal distribution, the weight function v_2 should be scaled so that $E[v_2(r)r^2] = k$, where r^2 is a random variable from a χ^2_k distribution. The influence functions of one-step M-functionals may be written using the influence functions of initial estimators. The influence functions are given by (1) and (2) with

$$\gamma_{T_2}(r) = E[v_1(r)]^{-1} \left[v_1(r)r - k^{-1}E[v_1'(r)r]\gamma_{T_1}(r) \right],$$

$$\alpha_{C_2}(r) = v_2(r)r^2 - \frac{E[v_2'(r)r^3]}{k(k+2)}\alpha_{C_1}(r)$$

and

$$\beta_{C_2}(r) = \frac{E[v_2'(r)r^3]}{2k(k+2)} \left[\alpha_{C_1}(r) - (k+2)\beta_{C_1}(r) \right] + 1.$$

Location and scatter S-functionals (Rousseeuw and Leroy, 1987; Davies, 1987) are defined as the solutions $T(F)$ and $C(F)$ to the problem of minimizing $det(C(F))$ subject to

$$E[\rho(r)] = b,$$

where $r^2 = ||\boldsymbol{x} - T(F)||^2_{C(F)}$ and $\rho(r)$ is bounded, increasing and non-negative function. The constant b is generally chosen to be $b = E[\rho(r)]$ where r^2 is a random variable from a χ^2_k. Then $C(F) = \Sigma$ at multinormal population. An example of function ρ is Tukey's biweight ρ-function

$$\rho(r) = \begin{cases} \frac{r^2}{2} - \frac{r^4}{2c^2} + \frac{r^6}{6c^4}, & |r| \leq c \\ \frac{c^2}{6}, & |r| > c. \end{cases}$$

To attain $\epsilon\%$ breakdown point, the tuning constant c is chosen so that $b = \epsilon\rho(c)$. For the influence functions of S-functionals, see e.g Lopuhaä (1989).

3 New test statistics

As before let $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a random sample from a k-variate distribution with cdf F. We wish to test the null hypothesis of multinormality, that is, F is the cdf of $N(\boldsymbol{\mu}, \Sigma)$ for some (unknown) $\boldsymbol{\mu}$ and Σ.

As in the univariate case, the Mahalanobis difference between two location vectors may be used to indicate skewness. Also, two scatter matrices may be used to construct a multivariate measure of kurtosis. Therefore we define our test statistics as follows.

Definition 1 *(i)   Let T_1 and T_2 be two separate location vectors and let C be a scatter matrix. Then a test statistic for multinormality (to detect skewness) is given by*

$$U = (T_1 - T_2)^T C^{-1} (T_1 - T_2).$$

(ii)   Let C_1 and C_2 be two separate scatter matrices equipped with the correction factor. Then a test statistic for multinormality (to detect kurtosis) is given by

$$W = ||C_1^{-1}C_2 - I_k||^2$$
$$= \left[Tr((C_1^{-1}C_2)^2) - \frac{1}{k}Tr^2(C_1^{-1}C_2) \right] + \frac{1}{k}\left[Tr(C_1^{-1}C_2) - k \right]^2,$$

where $|| \cdot ||^2 = Tr(\cdot^T \cdot)$.

Note that as U is based on the distance between two location parameters, it measures the skewness in the direction of $C^{-1/2}(T_1 - T_2)$. The first part of W, $Tr((C_1^{-1}C_2)^2) - k^{-1}Tr^2(C_1^{-1}C_2)$, is the variance of the eigenvalues of $C_1^{-1}C_2$ and therefore measures the difference on the shapes of scatter matrix estimates. The second part clearly measures the difference in their scales. Note also that using the eigenvalues and eigenvectors of $C_1^{-1}C_2$ or equivalently $C_1^{-1/2}C_2C_1^{-1/2}$, it is also possible to find the direction that maximizes the kurtosis. Since

$$\frac{\boldsymbol{u}^T C_2 \boldsymbol{u}}{\boldsymbol{u}^T C_1 \boldsymbol{u}}$$

is a descriptive statistic for the kurtosis of linear combination $\boldsymbol{u}^T \boldsymbol{x}$, the vector $\boldsymbol{u}$ that maximizes this quantity gives the direction of maximal kurtosis. Now writing $\boldsymbol{u} = C_1^{-1/2}\boldsymbol{v}$ the problem reduces to maximizing

$$\frac{\boldsymbol{v}^T C_1^{-1/2} C_2 C_1^{-1/2} \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}},$$

that is, the so called Rayleigh quotient of $C_1^{-1/2}C_2C_1^{-1/2}$. It is well known that the direction $\boldsymbol{v}$ that maximizes this quotient is the eigenvector corresponding to the largest eigenvalue of $C_1^{-1/2}C_2C_1^{-1/2}$.

Since our measures of skewness and kurtosis are based on affine equivariant location vectors and scatter matrices, the invariance of test statistics easily follows.

Lemma 1 *The statistics U and W are affine invariant.*

As the test statistics are affine invariant, their critical values for different dimensions and different sample sizes may in principle be tabulated. The limiting null distributions may also be used to find approximate p-values, see the following theorem.

Theorem 1 *(i) Under the null hypothesis, the limiting distribution of nU is that of $\eta_1 U_1$, where $U_1 \sim \chi_k^2$ and $\eta_1 = k^{-1}E_F[(\gamma_{T_1}(r) - \gamma_{T_2}(r))^2]$ with $r^2 \sim \chi_k^2$.*

(ii) Under the null hypothesis, the limiting distribution of nW is that of

$$\eta_2\,W_1\;+\;\eta_3\,W_2,$$

where $W_1 \sim \chi_{k(k+1)/2-1}^2$ and $W_2 \sim \chi_1^2$ are independent,

$$\eta_2 = 2\,[k(k+2)]^{-1}E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))^2]$$

and

$$\eta_3 = k^{-1}E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))^2] - 2E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))(\beta_{C_2}(r) - \beta_{C_1}(r))]$$
$$+ kE_F[(\beta_{C_2}(r) - \beta_{C_1}(r))^2].$$

The expected values are again calculated for $r^2 \sim \chi_k^2$.

Finally note that, using special choices of location and scatter estimators, it is possible to obtain generalizations of classical Mardia's measures of skewness and kurtosis. If the multivariate measure of skewness is constructed so that T_1 and C are the sample mean vector and sample covariance matrix and T_2 is the one-step M-estimator that uses T_1 and C as initial estimators and weight function $v_1(r) = r^2$, then the resulting skewness measure is easily seen to be

$$b_{1,new} = ave_i\{r_i^2\}^{-2}ave_{i,j}\{r_i^2 r_j^2 r_{ij}\}.$$

Note that this measure is equivalent to that introduced in Mori et al. (1993). The limiting distribution of $nb_{1,new}$ may be computed using Theorem 1 and it simplifies to that of $\eta_1 U_1$, where $U_1 \sim \chi_k^2$ and $\eta_1 = 2(k+2)/k^2$. See also Henze (1997).

If the new kurtosis measure is such that it uses the sample covariance matrix as C_1 and one-step M-estimator based on the sample mean vector and sample covariance matrix with weight function $v_2(r) = (k+2)^{-1}r^2$ as C_2, then we obtain the generalization of the classical Mardia's measure of kurtosis, that is,

$$b_{2,new} = (k+2)^{-2}ave_{i,j}\{r_i^2 r_j^2 r_{ij}^2\} - 2(k+2)^{-1}ave_i\{r_i^4\} + k.$$

Using Theorem 1, the limiting distribution of $nb_{2,new}$ is that of $\eta_2 W_1 + \eta_3 W_2$, where $W_1 \sim \chi^2_{k(k+1)/2-1}$ and $W_2 \sim \chi^2_1$ and

$$\eta_2 = \frac{4(k+4)}{(k+2)^2} \quad \text{and} \quad \eta_3 = \frac{8}{k+2}.$$

4 Limiting efficiencies

In this section, the contaminated normal model $CN(\boldsymbol{\mu}, \Sigma)$ is used for efficiency comparisons. Due to the affine invariance property, it is not a restriction to assume that the observation comes from standard multinormal distribution $N(\mathbf{0}, I_k)$ with probability $(1 - \epsilon)$ and from $N(\boldsymbol{\mu}, \Sigma)$ with probability ϵ. In the following we write G for the cdf of the contamination $N(\boldsymbol{\mu}, \Sigma)$ and let $\epsilon = \delta/\sqrt{n}$ depend on n. The null hypotheses case is then given by $H_0 : \delta = 0$, and we consider a contiguous sequence of alternatives $H_n : \epsilon = \delta/\sqrt{n}$ for a fixed $\delta > 0$. For the contiguity of this sequence, see the proof of Theorem 2.

Theorem 2 (i) *Under the sequence of alternatives H_n, the limiting distribution of nU is that of $\eta_1 U_1^*$ where U_1^* has a noncentral χ^2_k distribution with noncentrality parameter*

$$C_U(\delta) = \frac{\delta^2}{\eta_1} \big|\big| E_G\big[(\gamma_{T_1}(r) - \gamma_{T_2}(r))\boldsymbol{u}\big]\big|\big|^2,$$

where $r = ||\boldsymbol{x}||$, $\boldsymbol{u} = ||\boldsymbol{x}||^{-1}\boldsymbol{x}$ with $\boldsymbol{x}$ having a $N(\boldsymbol{\mu}, \Sigma)$ distribution.

(ii) Under the sequence of alternatives H_n, the limiting distribution of nW is that of

$$\eta_2 W_1^* + \eta_3 W_2^*,$$

where W_1^* and W_2^* are independent noncentral chi-squared variables with $k(k+1)/2 - 1$ and 1 degrees of freedom and noncentrality parameters

$$C_{W_1}(\delta) = \frac{\delta^2}{\eta_2}\big(Tr(E^2) - \frac{1}{k}Tr^2(E)\big) \quad \text{and} \quad C_{W_2}(\delta) = \frac{\delta^2}{\eta_3 k}Tr^2(E),$$

where

$$\begin{aligned} E &= E_G[IF(\boldsymbol{x}; C_1^{-1}C_2, F)] \\ &= E_G\big[(\alpha_{C_2}(r) - \alpha_{C_1}(r))\boldsymbol{u}\boldsymbol{u}^T - (\beta_{C_2}(r) - \beta_{C_1}(r))I_k\big] \end{aligned}$$

and $r = ||\boldsymbol{x}||$, $\boldsymbol{u} = ||\boldsymbol{x}||^{-1}\boldsymbol{x}$ with $\boldsymbol{x}$ having a $N(\boldsymbol{\mu}, \Sigma)$ distribution.

The contaminated normal model is a natural alternative to normal model. It is flexible containing alternatives of different types (skewness and kurtosis) depending on the choices of $\boldsymbol{\mu}$ and Σ. Contiguous alternative sequences for skewness and kurtosis are obtained by choices

$$H_{1n} : \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)^T \quad \text{and} \quad \Sigma = I_k$$

and
$$H_{2n}: \quad \boldsymbol{\mu} = \mathbf{0} \quad \text{and} \quad \Sigma = diag(\sigma_1^2, \ldots, \sigma_k^2).$$

In the next we compare our new skewness measures to Mardia's skewness b_1 using the alternative sequences H_{1n}. Further, new kurtosis measures are compared to Mardia's b_2 using the alternative sequences H_{2n}, therefore we need the following results, see e.g Mardia (1974).

Theorem 3 *(i) Under the sequence of alternatives H_{1n}, the limiting distribution of $nb_1/6$ is a noncentral chi-squared distribution with $k(k+1)(k+2)/6$ degrees of freedom and noncentrality parameter*

$$C_{b_1}(\delta) = \frac{\delta^2}{6} ||\boldsymbol{\mu}||^3.$$

(ii) Under the sequence of alternatives H_{2n}, the limiting distribution of $n(b_2 - k(k+2))^2/(8k(k+2))$ is a noncentral chi-squared distribution with 1 degrees of freedom and noncentrality parameter

$$C_{b_2}(\delta) = \frac{\delta^2 \left[2\, Tr((I_k - \Sigma)^2) - (Tr(I_k - \Sigma))^2 \right]^2}{8k(k+2)}.$$

Nyblom and Mäkeläinen (1983) extented the definition of Pitman efficiency in cases where the limiting distributions of test statistics are of different types. In our case, the asymptotic relative efficiency of U with respect to b_1 is then easily seen to be

$$ARE(U, b_1) = \left(\frac{\delta_1}{\delta_2} \right)^2, \tag{3}$$

where δ_1 and δ_2 are such that for chosen limiting size α and power β

$$P(\chi^2_{k(k+1)(k+2)/6}(C_{b_1}(\delta_1)) > \chi^2_{k(k+1)(k+2)/6, 1-\alpha})$$
$$= P(\chi^2_k(C_U(\delta_2)) > \chi^2_{k, 1-\alpha}) = \beta.$$

Note that the limiting efficiency may now depend on the chosen size and power.

In the next the efficiencies of four new skewness measures are compared to Mardia's skewness b_1 using the alternative sequence $H_{1n} : \boldsymbol{\mu} = (\mu_1, 0, \ldots, 0)^T$ and $\Sigma = I_k$. In all three cases, location vector T_1 and scatter matrix C were chosen to be the sample mean vector and sample covariance matrix. As T_2 we use Huber's M-estimator $M(0.9)$ in $\beta_{1,M}$, 25 % breakdown S-estimator with biweight ρ-function in $\beta_{1,S}$ and transformation-retransformation spatial median in $\beta_{1,T}$. Finally, as pointed out in previous section, the one-step M-estimator with sample mean vector and sample covariance matrix as initial estimators and weight function $v_1(r) = r^2$ is used as T_2 in $b_{1,new}$. The asymptotic relative efficiencies were computed using (3) with $\alpha = 0.05$ and $\beta = 0.5$. Simulations with 10000 replications were used to compute the noncentrality parameters given in Theorem 2. Since the
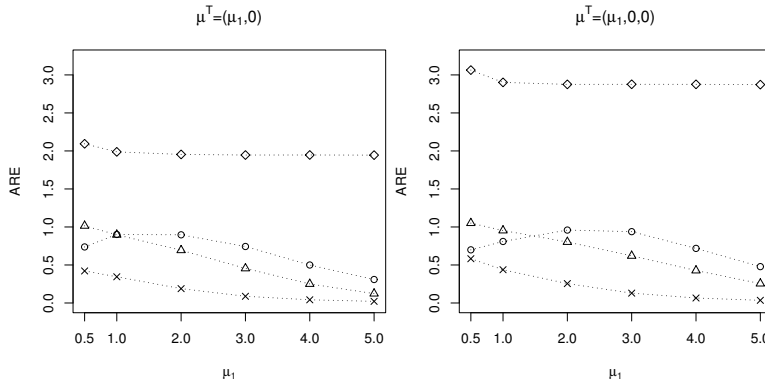
Fig. 1 The asymptotic relative efficiencies of $b_{1,new}$ ($\diamond$), $\beta_{1,M}$ ($\circ$), $\beta_{1,S}$ ($\triangle$) and $\beta_{1,T}$ ($\times$) relative to Mardia's skewness b_1 for different values of μ_1 in the bivariate (left panel) and 3-variate (right panel) cases.

resulting simulated noncentrality parameters appeared to be very unstable for very small μ_1, we only considered the values $\mu_1 = 0.5, 1, 2, 3, 4$ and 5.

In Figure 1, the asymptotic relative efficiencies are illustrated for different values of μ_1 and for dimensions $k = 2$ and $k = 3$. In both cases, the $b_{1,new}$ test, that is, the generalization of the Mardia's skewness measure is the most efficient one. For small μ_1 the efficiencies of $\beta_{1,M}$ and $\beta_{1,S}$ are tolerable, but as μ_1 increases the efficiencies decrease. The efficiency of $\beta_{1,T}$ is very low for all μ_1. All efficiencies seem to increase with dimension k.

Since the test based one-step M-estimator with $v_1(r) = r^2$ appeared to be the most efficient one, the effect of using one-step M-estimator with weigth function $v_1(r) = r^4$ on efficiencies was studied. The resulting test appeared to be the most efficient one and the efficiencies increased drastically with μ_1. Note, however, that the one-step M-estimator with weigth function $v_1(r) = r^4$ is highly sensitive to outliers, therefore this choice yields to very non-robust test of skewness.

Let us now compare our new kurtosis measures to Mardia's kurtosis b_2 using the alternative seguences $H_{2n} : \boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = diag(\sigma_1^2, \ldots, \sigma_k^2)$. The asymptotic relative efficiency of W with respect to b_2 is given by

$$ARE(W, b_2) = \left(\frac{\delta_1}{\delta_2}\right)^2, \tag{4}$$

where δ_1 and δ_2 are chosen so that

$$P(\eta_2 \, \chi^2_{k(k+1)/2-1}(C_{W_1}(\delta_2)) + \eta_3 \, \chi^2_1(C_{W_2}(\delta_2)) > c_\alpha)$$
$$= P(\chi^2_1(C_{b_2}(\delta_1)) > \chi^2_{1,1-\alpha}) = \beta$$

and the critical value c_α is given by

$$P(\eta_2 \, \chi^2_{k(k+1)/2-1}(0) + \eta_3 \, \chi^2_1(0) > c_\alpha) = \alpha. \tag{5}$$

In our efficiency comparisons, we use three different W statistics. For every statistic, C_1 was chosen to be sample covariance matrix and as C_2 we use Huber's M-estimator $M(0.9)$ in W_M, 25 % breakdown S-estimator with biweight ρ-function in W_S and one-step M-estimator based on the sample mean vector and sample covariance matrix with weight function $v_2(r) = (k+2)^{-1}r^2$ in $b_{2,new}$. The asymptotic relative efficiencies were computed using (4) with $\alpha = 0.05$ and $\beta = 0.5$. The tail probabilities of linear combinations of chi-squared random variables were computed using the algorithm developed in Field (1993) and the noncentrality parameters given in Theorem 2 were obtained through simulations.
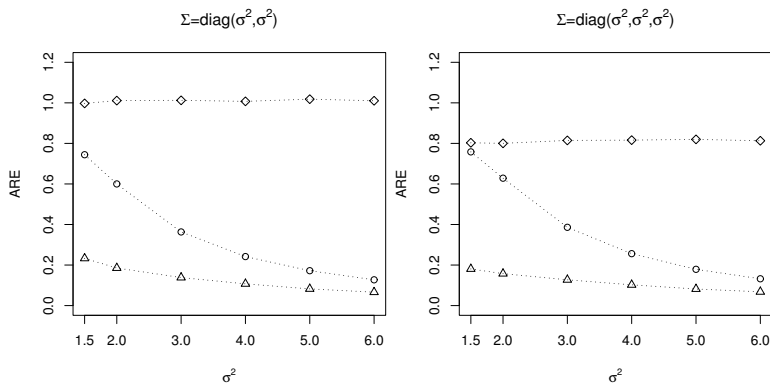

Fig. 2 The asymptotic relative efficiencies of $b_{2,new}$ ($\diamond$), W_M ($\circ$) and W_S ($\triangle$) relative to Mardia's kurtosis b_2 for different values of σ^2 in the bivariate (left panel) and 3-variate (right panel) cases.
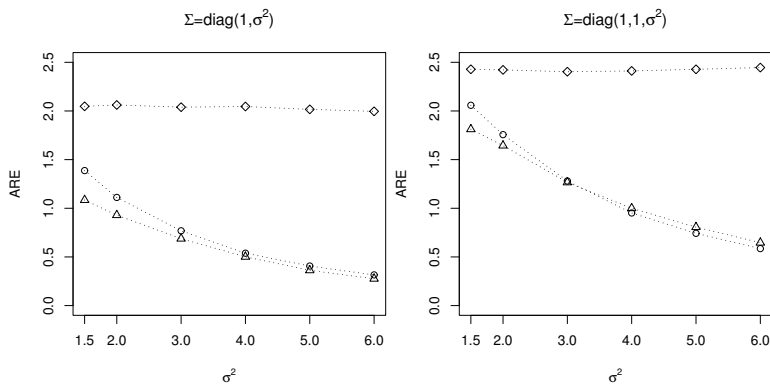

Fig. 3 The asymptotic relative efficiencies of $b_{2,new}$ ($\diamond$), W_M ($\circ$), W_S ($\triangle$) relative to Mardia's kurtosis b_2 for different values of σ^2 in the bivariate (left panel) and 3-variate (right panel) cases.

Figure 2 illustrates the efficiencies in cases $\Sigma = \sigma^2 I_2$ and $\Sigma = \sigma^2 I_3$ for different values of σ^2. Now as $k = 2$, the $b_{2,new}$ test is equally powerful with Mardia's kurtosis, but in case $k = 3$ Mardia's kurtosis performs better than the other tests. These results were, however, expected since only the second part of test statistic W detects this kind of unnormality. For small σ^2, W_M is much more efficient than W_S, but as σ^2 increases, these tests become equally efficient.

In Figure 3 the efficiencies are given in cases $\Sigma = diag(1, \sigma^2)$ and $\Sigma = diag(1, 1, \sigma^2)$ for different values of σ^2. In both cases the $b_{2,new}$ test performs better than the other tests. As σ^2 is small, also W_M and W_S tests are more efficient than b_2 but again as σ^2 increases, their efficiencies decrease.

5 Simulation study

In this section, we compare the empirical powers of skewness and kurtosis measures defined in the last section with the powers of b_1 and b_2 through a simple simulation study. The skewness measures were compared using 3-variate samples of sizes $n = 50$ and 200 from a contaminated normal distributions with $\epsilon = 0.1$ and $G = N((\mu_1, 0, 0)^T, I_3)$ with selected values of μ_1. For the kurtosis measures we used $G = N(\mathbf{0}, diag(1, 1, \sigma^2))$ with selected values of σ^2. The test statistics were then computed. Corresponding p-values of skewness measures were obtained using the chi-square approximations to the null distributions. Since the convergence of all kurtosis measures to the null distributions appeared to be very slow even in the case $n = 200$, the empirical critical values were simulated using 10000 multinormal samples. Resulting critical values (as well as asymptotic ones obtained using (5)) are given in Table 2 in Appendix B. The process was replicated 1500 times and empirical powers of the level 0.10 tests are illustrated in Figures 4 and 5.

Consider first the simulation results in Figure 4. As $n = 50$ and μ_1 is small, no big differences may be seen between different tests. As μ_1 increases, the b_1 and $\beta_{1,S}$ tests become more powerful than the others, but as n increases, the empirical power of $b_{1,new}$ is highest as suggested by the limiting efficiencies. The $\beta_{1,S}$ test is equally powerful with Mardia's test and $\beta_{1,M}$ and $\beta_{1,T}$ perform also very well. The sizes of $\beta_{1,M}$, $\beta_{1,S}$ and $\beta_{1,T}$ are close to the designated size 0.10 in both cases. In the case $n = 50$, the size of b_1 was below 0.10, therefore a correction factor given in Mardia (1974) was used to speed up the convergence to the null distribution.

The results in Figure 5 show that as $n = 50$, W_M and W_S tests are the most powerful ones and perform much better than b_2 and $b_{2,new}$. This may be due to the fact that b_2 and $b_{2,new}$ converge very slowly to the normal distribution. As $n = 200$, the powers of Mardia's type tests increase, but the tests based on robust scatter estimators are still slightly more powerful. In both cases, the sizes of tests are close to the designated size 0.10.
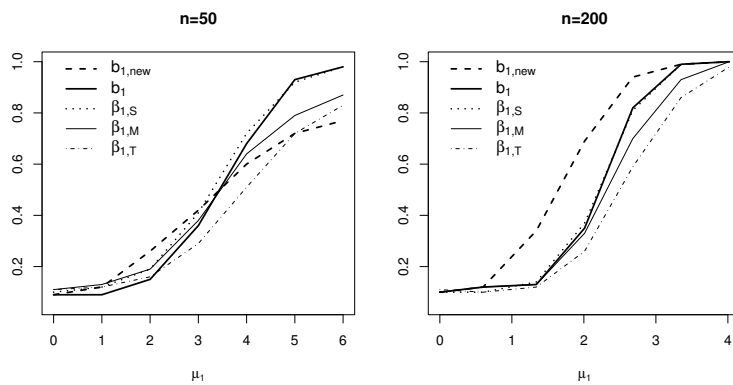
Fig. 4 Empirical powers of the level 0.10 tests to detect skewness for the contaminated normal distribution with $\epsilon = 0.1$ and contamination $N((\mu_1, 0, 0)^T, I_3)$.
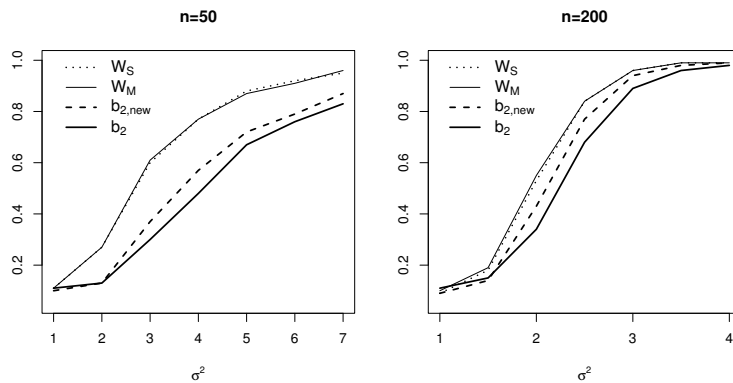

Fig. 5 Empirical powers of the level 0.10 tests to detect kurtosis using contaminated normal distribution with $\epsilon = 0.1$ and contamination $N(\mathbf{0}, diag(1, 1, \sigma^2))$.

6 A real data example

In our example we consider the same data as in Kankainen et al. (2004), that is, two bivariate data sets of $n = 50$ observations and the combined data set of $n = 100$ observations. The data sets are explained and illustrated in Figure 6. We use Mardia's skewness b_1 and kurtosis b_2 to test the null hypothesis of bivariate normality. The results are compared to those given by new skewness and kurtosis measures $\beta_{1,S}$ and W_S. These measures were included in the comparisons, since in spite of their low limiting efficiencies, the finite-sample powers appeared to be very competitive with those of Mardia's measures.

The values of the test statistics were calculated for the three data sets ('Boys', 'Girls' and 'Combined'). The p-values for b_1 and $\beta_{1,S}$ were obtained using the chi-square approximations to the null distributions. For b_2 and
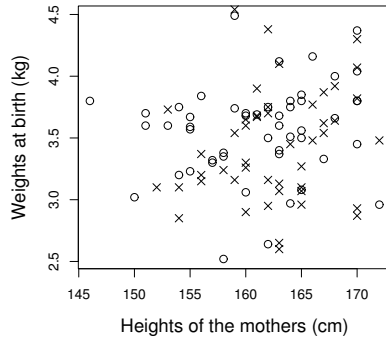
Fig. 6 The height of the mother and the birth weight of the child measured on 50 boys who have siblings (o) and 50 girls (x) who do not have siblings.

Table 1 Test statistics for the three data sets.

	Boys	Girls	Combined
	Test statistic	Test statistic	Test statistic
b_1	3.034	3.376	9.242^*
$\beta_{1,S}$	0.976	3.065	$5.473^\dagger$
b_2	0.016	0.281	2.275
W_S	1.530	1.889	$3.769^\dagger$

$* \ p \leq 0.10, \ \dagger \ p \leq 0.05$

W_S, empirical critical values were used to approximate p-values. See Table 1 for the results.

Since the boys in the study had siblings and girls did not, it was expected that observations concerning boys and girls separately came from different bivariate normal distributions. This is seen by using all skewness and kurtosis measures. Moreover, due to the difference between boys and girls, the combined data set was assumed to be a mixture of two different bivariate normal distributions. This is confirmed using our new skewness and kurtosis measures. Mardia's skewness rejects the normality hypothesis on significance level 0.10 and the p-value for $\beta_{1,S}$ test is even less than 0.05. When we use Mardia's kurtosis the normality hypothesis is not rejected, but the W_S test gives p-value which is less than 0.05.

7 Final remarks

The paper develops affine invariant tests for multivariate skewness and kurtosis. The tests for skewness use the Mahalanobis distance between two

location vectors and the tests for kurtosis are based on the comparison of two scatter matrices. As in the univariate case, these skewness and kurtosis measures are independent and may therefore be combined to obtain a test for multinormality. The resulting test is just a sum of skewness and kurtosis measures introduced here and follows a weighted chi-squared distribution.

   In the paper, new tests with several choices of affine equivariant location and scatter estimates are considered in more detail. It is shown that using the so called one-step M-estimators with special choices of weight functions and sample mean vector and covariance matrix as initial estimators in our test constructions yield the generalizations of classical Mardia's measures of skewness and kurtosis. The limiting efficiencies of these measures appear to be very good as compared to the classical Mardia's tests. As illustrated by a simulation study, the finite-sample behavior of tests based on robust scatter estimators is also satisfactory.

Appendix A: Proofs of the results

Proof (Lemma 1)

The affine invariance of U and W follows using $T(AX + \boldsymbol{b}) = AT(X) + \boldsymbol{b}$ and $C(AX + \boldsymbol{b}) = AC(X)A^T$. Then

$U(AX + \boldsymbol{b})$

$= (AT_1(X) + \boldsymbol{b} - (AT_2(X) + \boldsymbol{b}))^T \left(AC_1(X)A^T\right)^{-1} (AT_1(X) + \boldsymbol{b} - (AT_2(X) + \boldsymbol{b}))$

$= (T_1(X) - T_2(X))^T A^T (A^T)^{-1} C_1^{-1}(X)A^{-1}A (T_1(X) - T_2(X)) = U(X)$

and

$W(AX + \boldsymbol{b})$

$= ||C_1^{-1}(AX + \boldsymbol{b})C_2(AX + \boldsymbol{b}) - I_k||^2 Tr[(C_1^{-1}(AX + \boldsymbol{b})C_2(AX + \boldsymbol{b}))^2]$

$\quad - 2\,Tr^2[C_1^{-1}(AX + \boldsymbol{b})C_2(AX + \boldsymbol{b})] + k$

$= Tr[((AC_1(X)A^T)^{-1}AC_2(X)A^T)^2] - 2Tr^2[(AC_1(X)A^T)^{-1}AC_2(X)A^T] + k$

$= Tr[(A^T)^{-1}C_1^{-1}(X)A^{-1}AC_2(X)A^T(A^T)^{-1}C_1^{-1}(X)A^{-1}AC_2(X)A^T]$

$\quad - 2\,Tr^2[(A^T)^{-1}C_1^{-1}(X)A^{-1}AC_2(X)A^T] + k$

$= Tr[(C_1^{-1}(X)C_2(X))^2] - 2Tr^2[C_1^{-1}(X)C_2(X)] + k = W(X)$

Proof (Theorem 1)

(i) The test statistics are affine invariant, therefore we may assume that $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_k$. Write then $\boldsymbol{x} = r\boldsymbol{u}$, where $r = ||\boldsymbol{x}||$ and $\boldsymbol{u} = ||\boldsymbol{x}||^{-1}\boldsymbol{x}$ and r and $\boldsymbol{u}$ are independent.

Let us denote $V(F) = C^{-1/2}(F)(T_1(F) - T_2(F))$. Then the influence function of $V(F)$ is

$$IF(\boldsymbol{x}; V, F) = (\gamma_{T_1}(r) - \gamma_{T_2}(r))\boldsymbol{u}.$$

Write then $V = V(F_n)$ for the estimate corresponding to the functional $V(F)$. Now according to Huber (1981), as $V(F)$ is sufficiently regular,

$$\sqrt{n}(V - V(F)) = \frac{1}{\sqrt{n}} \sum_i IF(\boldsymbol{x}_i; V, F) + o_p(1). \tag{6}$$

Thus under H_0, $V(F) = 0$, and by the central limit theorem, the limiting distribution of $\sqrt{n}V$ is multinormal with mean $\mathbf{0}$ and covariance matrix

$$E_F[IF(\boldsymbol{x}; V, F)IF(\boldsymbol{x}; V, F)^T] = \frac{E_F[(\gamma_{T_1}(r) - \gamma_{T_2}(r))^2]}{k}I_k$$

If we now denote $\eta_1 = k^{-1}E_F[(\gamma_{T_1}(r) - \gamma_{T_2}(r))^2]$, then the limiting distribution of $nU = nV^TV$ is that of $\eta_1 U_1$, where $U_1 \sim \chi_k^2$.

(ii) Write
$$W = ||C_1^{-1}C_2 - I_k||^2 = ||H||^2.$$

Note first that the influence function of H is

$$IF(\boldsymbol{x}; H, F) = IF(\boldsymbol{x}; C_1^{-1}C_2, F) = (\alpha_{C_2}(r) - \alpha_{C_1}(r))\boldsymbol{u}\boldsymbol{u}^T - (\beta_{C_2}(r) - \beta_{C_1}(r))I_k.$$

As before, since $H(F)$ is sufficiently regular, we may write $H = H(F_n)$ as in (6). The elements of $\sqrt{n}H$ are thus normally distributed and the asymptotic variances and covariances are given by

$$ASV(h_{ij}; F) = \frac{E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))^2]}{k(k+2)},$$

$$ASC(h_{ii}, h_{jj}; F) = \frac{E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))^2]}{k(k+2)} + E_F[(\beta_{C_2}(r) - \beta_{C_1}(r))^2]$$
$$- \frac{2E_F[(\alpha_{C_2}(r) - \alpha_{C_1}(r))(\beta_{C_2}(r) - \beta_{C_1}(r))]}{k}$$

and
$$ASV(h_{ii}; F) = ASC(h_{ii}, h_{jj}; F) + 2ASV(h_{ij}; F).$$

Note that the diagonal and off-diagonal elements of H are independent.

Consider now the test statistic,

$$nW_1 = n\big[Tr((C_1^{-1}C_2)^2) - k^{-1}Tr^2(C_1^{-1}C_2)\big],$$

which may be rewritten as

$$nW_1 = nTr(H^2) - \frac{n}{k}Tr^2(H)$$
$$= 2n\sum\sum_{i<j}h_{ij}^2 + n\sum_i h_{ii}^2 - \frac{n}{k}(\sum_i h_{ii})^2.$$

Write then $b^2 = 2ASV(h_{ij}; F)$ and $d^2 = ASC(h_{ii}, h_{jj}; F)$. Since $\sqrt{n}h_{ij} \sim N(0, b^2/2)$ and the elements h_{ij}, $i \neq j$, are independently distributed, the limiting distribution of $2n\sum\sum_{i<j}h_{ij}^2$ under H_0 is that of $\eta_2 W_a$, where $W_a \sim \chi^2_{k(k-1)/2}$.

The diagonal elements h_{ii} are not independent, but we may write $\boldsymbol{h} = A\boldsymbol{z}$, where $\boldsymbol{h} = (h_{11}, \ldots, h_{kk})^T$, $\boldsymbol{z} = (z_1, \ldots, z_{k+1})^T$, $\sqrt{n}\boldsymbol{z} \sim N(\boldsymbol{0}, I_{k+1})$ and A is a $k \times (k+1)$ matrix

$$A = \begin{pmatrix} bI_k & d\boldsymbol{1}_k \end{pmatrix},$$

where $\boldsymbol{1}_k$ is a unit vector of length k. Now $ASC(h_{11}, \ldots, h_{kk}; F) = b^2 I_k + d^2 J_k$, where J_k is a $k \times k$ unit matrix, and using the above transformation we have that

$$n\sum_i h_{ii}^2 - \frac{n}{k}(\sum_i h_{ii})^2 = nb^2\big[\sum_i z_i^2 - \frac{1}{k}(\sum_i z_i)^2\big]$$
$$= nb^2\left(\frac{\boldsymbol{z}^T A_1^T A_1 \boldsymbol{z}}{b^2} - \frac{\boldsymbol{z}^T A_2^T A_2 \boldsymbol{z}}{b^2}\right), \tag{7}$$

where A_1 and A_2 are $k \times (k+1)$ matrices

$$A_1 = \begin{pmatrix} bI_k & \boldsymbol{0} \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} \frac{b}{k}J_k & \boldsymbol{0} \end{pmatrix}.$$

Since $b^{-2}A_1^T A_1$ is idempotent with $rank(b^{-2}A_1^T A_1) = k$ and $b^{-2}A_2^T A_2$ is idempotent with $rank(b^{-2}A_2^T A_2) = 1$, the limiting distribution of (7) under H_0 is that of $b^2 W_b$, where $W_b \sim \chi^2_{k-1}$ (Rao, 1965; Section 3b.4). Finally note that, since W_a and W_b are independent, the limiting distribution of nW_1 is that of $b^2(W_a + W_b)$.

Using the similar technique we may write

$$nW_2 = \frac{n}{k}\big[Tr(C_1^{-1}C_2) - k\big]^2 = \frac{n}{k}Tr^2(H) = \frac{n}{k}(\sum_i h_{ii})^2$$
$$= nk\big(\frac{b}{k}\sum_i z_i^2 + dz_{k+1}\big)^2 = n(b^2 + kd^2)\frac{\boldsymbol{z}^T A_3^T A_3 \boldsymbol{z}}{b^2 + kd^2},$$

where A_3 is a $k \times (k+1)$ matrix

$$A_3 = \begin{pmatrix} \frac{b}{k}J_k & d\boldsymbol{1}_k \end{pmatrix}.$$

Since $(b^2 + kd^2)^{-1}A_3^T A_3$ is idempotent with $rank((b^2 + kd^2)^{-1}A_3^T A_3) = 1$, the limiting distribution of nW_2 under H_0 is that of $(b^2 + kd^2)W_c$, where $W_c \sim \chi^2_1$. The result now follows, if we denote $\eta_2 = b^2$ and $\eta_3 = b^2 + kd^2$.

Proof (Theorem 2)

(i)   In the proof, we apply LeCam's third lemma. Write the density function of contaminated normal distribution as $f_\epsilon(\boldsymbol{x}_i) = (1 - \epsilon)f(\boldsymbol{x}_i) + \epsilon g(\boldsymbol{x}_i)$. Let us first show that $\epsilon = \delta/\sqrt{n}$ is contiguous to the null hypothesis: The optimal likelihood ratio test statistic for testing H_0 against H_n is

$$L = \sum_i \log \frac{f_\epsilon(\boldsymbol{x}_i)}{f_0(\boldsymbol{x}_i)} = \sum_i \log \frac{(1 - \epsilon)f(\boldsymbol{x}_i) + \epsilon g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)}$$

$$= \sum_i \log \left(1 + \left[\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right]\epsilon\right)$$

$$= \frac{\delta}{\sqrt{n}} \sum_i \left(\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right) - \frac{\delta^2}{2n} \sum_i \left(\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right)^2 + \dots$$

$$= \sqrt{n}\delta \frac{1}{n} \sum_i k_i - \frac{1}{2}\delta^2\sigma^2 + o_p(1),$$

where $E(k_i) = 0$ and $Var(k_i) = \sigma^2$ is bounded. Thus, by LeCam's first lemma, the sequence of alternatives is contiguous to the null hypothesis.

As in the proof of Theorem 1, write then

$$\sqrt{n}V = \frac{1}{\sqrt{n}} \sum_i IF(\boldsymbol{x}_i; V, F) + o_p(1).$$

Now under H_0,

$$\begin{pmatrix} \sqrt{n}V \\ L \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_1 I_k & E_F(\sqrt{n}VL) \\ E_F(\sqrt{n}VL)^T & E_F(L^2) \end{pmatrix}\right),$$

where

$$E_F(\sqrt{n}VL) = E_F\left[\frac{1}{\sqrt{n}} \sum_i IF(\boldsymbol{x}_i; V, F) \cdot \frac{\delta}{\sqrt{n}} \sum_i \left(\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right)\right]$$

$$= E_F\left[\frac{\delta}{n} \sum_i IF(\boldsymbol{x}_i; V, F) \cdot \left(\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right)\right]$$

$$+ E_F\left[\frac{\delta}{n} \sum\sum_{i \neq j} IF(\boldsymbol{x}_i; V, F) \cdot \left(\frac{g(\boldsymbol{x}_j)}{f(\boldsymbol{x}_j)} - 1\right)\right]$$

$$= \frac{\delta}{n} \sum_i E_F\left[IF(\boldsymbol{x}_i; V, F) \cdot \left(\frac{g(\boldsymbol{x}_i)}{f(\boldsymbol{x}_i)} - 1\right)\right]$$

$$= \delta E_G[IF(\boldsymbol{x}; V, F)].$$

Hence by LeCam's third lemma we have that under the alternative sequence, $\sqrt{n}V \xrightarrow{d} N(E_G[IF(\boldsymbol{x}_i; V, F)], \eta_1 I_k)$ and the limiting distribution of nU is

that of $\eta_1 U_1^*$ where U_1^* has a noncentral χ_k^2 distribution with noncentrality parameter

$$\frac{\delta^2}{\eta_1} E_G(IF(\boldsymbol{x}; V, F))^T E_G(IF(\boldsymbol{x}; V, F)) = \frac{\delta^2}{\eta_1} \|E_G(IF(\boldsymbol{x}; V, F))\|^2.$$

(ii)   As in the Proof of Theorem 1, decompose nW as

$$nb^2 \sum_{i<j} \sum \left(\frac{h_{ij}}{b/\sqrt{2}}\right)^2 + nb^2 \left(\frac{\boldsymbol{z}^T A_1^T A_1 \boldsymbol{z}}{b^2} - \frac{\boldsymbol{z}^T A_2^T A_2 \boldsymbol{z}}{b^2}\right)$$
$$+ n(b^2 + kd^2)\frac{\boldsymbol{z}^T A_3^T A_3 \boldsymbol{z}}{b^2 + kd^2} = b^2 W_a + b^2 W_b + (b^2 + kd^2)W_c.$$

If W_a^* is now computed under H_n, then the noncentality parameter of W_a^* is

$$\frac{2}{b^2} \sum_{i<j} \sum \left[E(\sqrt{n}h_{ij})\right]^2.$$

Notice next that using the transformation $\boldsymbol{h} = A\boldsymbol{z}$ defined in the previous proof we have that, $E\boldsymbol{z} = A^- E\boldsymbol{h}$, where $(k+1) \times k$ matrix

$$A^- = \left(\tfrac{1}{b}I_k \; \boldsymbol{0}\right)^T$$

is the generalized inverse of A. The noncentrality parameter of W_b^* is then

$$n\left(\frac{E(\boldsymbol{z}^T)A_1^T A_1 E(\boldsymbol{z})}{b^2} - \frac{E(\boldsymbol{z}^T)A_2^T A_2 E(\boldsymbol{z})}{b^2}\right)$$
$$= n\left(\frac{(A^- E(\boldsymbol{h}))^T A_1^T A_1 A^- E(\boldsymbol{h})}{b^2} - \frac{(A^- E(\boldsymbol{h}))^T A_2^T A_2 A^- E(\boldsymbol{h})}{b^2}\right)$$
$$= \frac{1}{b^2}\left[\sum_i \left(E(\sqrt{n}h_{ii})\right)^2 - \frac{1}{k}\left(\sum_i E(\sqrt{n}h_{ii})\right)^2\right],$$

and the noncentrality parameter of $W_1^* = W_a^* + W_b^*$ simplifies to

$$\frac{1}{\eta_2}\left[\sum_{i,j}\sum \left(E(\sqrt{n}h_{ij})\right)^2 - \frac{1}{k}\left(\sum_i E(\sqrt{n}h_{ii})\right)^2\right].$$

The above expectations may again be computed using LeCam's third lemma. The resulting noncentrality parameter for W_1^* is then

$$\frac{\delta^2}{\eta_2}\left[\sum_{i,j}\sum \left(E_G(IF(\boldsymbol{x}_i; [H]_{ij}, F))\right)^2 - \frac{1}{k}\left(\sum_i E_G(IF(\boldsymbol{x}_i; [H]_{ii}, F))\right)^2\right]$$
$$= \frac{\delta^2}{\eta_2}\left[Tr[E_G(IF(\boldsymbol{x}; H, F))^2] - \frac{1}{k}Tr^2[E_G(IF(\boldsymbol{x}; H, F))]\right].$$

Similarly, for W_c^* we have the noncentrality parameter

$$n\frac{E(\boldsymbol{z}^T)A_3^T A_3 E(\boldsymbol{z})}{b^2 + kd^2} = n\frac{(A^- E(\boldsymbol{h}))^T A_3^T A_3 A^- E(\boldsymbol{h})}{b^2 + kd^2}$$
$$= \frac{1}{b^2 + kd^2} \cdot \frac{1}{k}\Big(\sum_i E(\sqrt{n}h_{ii})\Big)^2 = \frac{1}{\eta_3 k}Tr^2[E_G(IF(\boldsymbol{x}; H, F))].$$

The result now follows, since $IF(\boldsymbol{x}; H, F) = IF(\boldsymbol{x}; C_1^{-1}C_2, F)$.

Appendix B: The empirical critical values for kurtosis measures

Table 2 Empirical critical values of Mardia's kurtosis b_2 and new kurtosis measures W_M, W_S and $b_{2,new}$ for $\alpha = 0.05$ and $\alpha = 0.10$ based on 10000 samples.

		$\alpha = 0.05$				$\alpha = 0.10$			
k	n	b_2	W_M	W_S	$b_{2,new}$	b_2	W_M	W_S	$b_{2,new}$
2	20	2.54	5.56	9.34	17.16	2.20	3.11	4.39	14.05
	50	2.98	3.58	4.32	13.92	2.31	2.16	2.66	11.05
	100	3.23	2.61	3.09	12.92	2.41	1.72	2.14	10.49
	200	3.40	2.33	2.75	12.89	2.44	1.55	1.91	9.97
	∞	3.84	1.89	2.23	13.00	2.71	1.47	1.74	10.38
3	20	3.60	6.87	9.55	19.90	3.18	3.67	4.46	17.31
	50	3.58	4.18	4.16	15.99	2.87	2.66	2.61	13.80
	100	3.65	3.06	3.00	15.44	2.78	2.13	2.15	13.08
	200	3.58	2.54	2.49	14.93	2.68	1.83	1.89	12.75
	∞	3.84	1.85	1.87	15.12	2.71	1.52	1.57	12.76
4	20	4.67	8.13	9.85	21.97	4.18	4.41	5.07	19.80
	50	4.33	4.66	4.09	18.05	3.47	2.99	2.69	16.02
	100	4.05	3.39	2.75	17.29	3.17	2.35	2.04	14.99
	200	3.97	2.72	2.30	16.99	2.92	2.04	1.82	14.92
	∞	3.84	1.88	1.71	17.12	2.78	1.81	1.69	14.41
5	20	5.90	9.08	5.56	23.86	5.33	4.90	2.94	21.93
	50	5.05	4.96	3.49	20.23	4.23	3.41	2.38	18.20
	100	4.61	3.64	2.69	19.68	3.63	2.60	2.02	17.29
	200	4.23	2.92	2.22	19.10	3.28	2.28	1.82	17.15
	∞	3.84	1.94	1.65	19.09	2.71	1.68	1.47	17.01
8	20	10.08	7.34	25.04	28.47	9.44	3.20	14.24	26.82
	50	7.79	6.06	3.04	26.18	6.72	4.47	2.36	24.10
	100	6.58	4.58	2.63	25.64	5.13	3.50	2.12	23.41
	200	5.44	3.53	2.19	25.31	4.08	2.77	1.88	23.23
	∞	3.84	2.18	1.69	24.97	2.71	1.96	1.56	23.10

References

Bera A, and John S (1983) Tests for multivariate normality with Pearson alternatives. Comm Statist Theory Methods 12:103–117

Cox DR, Small NJH (1978) Testing multivariate normality. Biometrika 65:263–272

Davies PL (1987) Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. Ann Statist 15, 1269-1292

Field C (1993) Tail areas of linear combinations of chi-squares and noncentral chi-squares. J Stat Comp Simul 45:243–248

Geary RC (1935) The Ratio of the Mean Deviation to the Standard Deviation as a Test of Normality. Biometrika 27(3/4):310–332

Gnanadesikan R (1977) Methods for statistical data analysis of multivariate observations. Wiley, New York

Henze N (1997) Limit laws for multivariate skewness in the sense of Mori, Rohatgi and Szekely. Stat Probab Lett 33:299–307

Hettmansperger TP, Randles RH (2002) A practical affine equivariant multivariate median. Biometrika 89:851–860

Huber PJ (1981) Robust statistics. Wiley, New York

Kankainen A, Taskinen S, Oja H (2004) On Mardia's tests of multinormality. in: Hubert M, Pison G, Stryuf A, Van Aelst S (eds) Statistics for industry and technology. Birkhauser, Basel, pp 153–164

Koziol JA (1982) A class of invariant procedures for assessing multivariate normality. Biometrika 69:423–427

Koziol JA (1983) On assessing multivariate normality. J R Statist Soc Ser B 45:358–361

Koziol JA (1986) Assessing multivariate normality: a compendium. Comm Statist Theory Methods 15:2763–2783

Koziol JA (1987) An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. Metrika 34:17–24

Koziol JA (1993) Probability plots for assessing multivariate normality. The Statistician 42:161–173

Lopuhaä HP (1989) On the relation between S-estimators and M-estimators of multivariate location and covariance. Ann Statist 17:1662–1683

MacGillvray HL (1986) Skewness and Asymmetry: Measures and Orderings. Ann Statist 14:994–1011

Malkovitch JF, Afifi AA (1973) On tests for multivariate normality. J Am Statist Assoc 68:176–179

Mardia KV (1970) Measures of multivariate skewness and kurtosis with applications. Biometrika 57:519–530

Mardia KV (1974) Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhyã Ser B 36(2):115–128

Mardia KV (1980) Tests of univariate and multivariate normality. Handbook of Statistics, vol 1 pp 279–320

Maronna RA (1976) Robust M-estimators of multivariate location and scatter. Ann Statist 4:51–67

Mori TF, Rohatgi VK, Szekely GJ (1993) On multivariate skewness and kurtosis. Theor Probab Appl 38:547–551

Nyblom J, Mäkeläinen T (1983) Comparisons of tests for the presence of random walk coefficients in a simple linear model. J Am Statist Assoc 78:856–864

Oja H (1981) On location, scale, skewness and kurtosis of univariate distributions. Scand J Statist 8:154-168

Ollila E, Hettmansperger TP, Oja H (2003) Affine equivariant multivariate sign methods. Under revision.

Pearson K (1895) Contributions to the mathematical theory of evolution II. Skew variation in homogeneous material. Phil. Trans. R. Soc. Lond. A 186:343–414

Rao CR (1965) Linear statistical inference and its applications. Wiley, New York

Romeu JL, Ozturk A (1993) A comparative study of goodness-of-fit tests for multivariate normality. J Multivariate Anal 46:309–334

Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York

Tyler DE (1982) Radial estimates and the test for sphericity. Biometrika 69:429–436