

# Testing for normality in any dimension based on a partial differential equation involving the moment generating function

Norbert Henze · Jaco Visagie

**Abstract** We use a system of first-order partial differential equations that characterize the moment generating function of the  $d$ -variate standard normal distribution to construct a class of affine invariant tests for normality in any dimension. We derive the limit null distribution of the resulting test statistics, and we prove consistency of the tests against general alternatives. In the case  $d > 1$ , a certain limit of these tests is connected with two measures of multivariate skewness. The new tests show strong power performance when compared to well-known competitors, especially against heavy-tailed distributions, and they are illustrated by means of a real data set.

**Keywords** Moment generating function · test for multivariate normality · direct sum of Hilbert spaces · multivariate skewness · weighted  $L^2$ -statistic

## 1 Introduction

It is often of interest to check whether an observed  $d$ -dimensional dataset is compatible with the assumption of coming from a multivariate normal distribution. Such a model check is of practical interest to researchers, as many multivariate techniques rely on the assumption of multivariate normality (for

---

The second author's work is based on research supported by the National Research Foundation, South Africa (Research chair: Non-parametric, Robust Statistical Inference and Statistical Process Control, Grant number 71199). Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

N. Henze (Corresponding author)  
Karlsruhe Institute of Technology (KIT), Institute of Stochastics, Englerstr. 2, D-76313  
Karlsruhe, Germany  
E-mail: henze@kit.edu

I.J.H. Visagie  
University of Pretoria (UP), Department of Statistics, Private bag X20, Hatfield 0028, South  
Africa

short: multinormality). As a consequence, it is not surprising that there is ongoing interest in testing for multinormality, as can be witnessed by various recent papers on the subject, see for example Mecklin and Mundfrom (2005), Székely and Rizzo (2005), Farrell et al. (2007) and Joenssen and Vogel (2014) as well as the references therein. Research into the practical implementation of these tests is also undertaken regularly, see, e.g., Korkmaz et al. (2014) and Joenssen and Vogel (2014) regarding the implementation in the statistical software package R.

The purpose of this paper is not to review the multitude of tests that hitherto has been proposed (for an account of the state of the art regarding affine invariant procedures before 2002, see Henze (2002)), but to introduce and study a new class of tests that is based on a certain partial differential equation. To be specific, let  $X_1, X_2, \dots, X_n, \dots$  be independent and identically distributed (i.i.d.) copies of a  $d$ -dimensional random (column) vector  $X$ , the distribution of which is assumed to be absolutely continuous with respect to  $d$ -dimensional Lebesgue measure. All random vectors are defined on a common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Writing  $N_d(\mu, \Sigma)$  for the  $d$ -dimensional normal distribution with mean vector  $\mu$  and non-degenerate covariance matrix  $\Sigma$  and  $\mathcal{N}_d$  for the class of all non-degenerate  $d$ -variate normal distributions, a classical problem is to test the null hypothesis

$$H_0 : \text{The law of } X \text{ belongs to } \mathcal{N}_d, \quad (1)$$

against general alternatives. Since the class  $\mathcal{N}_d$  is closed with respect to full rank affine transformations, any genuine test statistic  $T_n = T_n(X_1, \dots, X_n)$  based on  $X_1, \dots, X_n$  should also be invariant with respect to such transformations, i.e., we should have  $T_n(AX_1 + b, \dots, AX_n + b) = T_n(X_1, \dots, X_n)$  for each nonsingular ( $d \times d$ )-matrix  $A$  and each  $b \in \mathbb{R}^d$ , see Henze (2002) for an account of the importance of affine invariance in connection with testing for multivariate normality. Writing  $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$  for the sample mean and  $S_n = n^{-1} \sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)^\top$  for the sample covariance matrix of  $X_1, \dots, X_n$ , where the superscript  $\top$  denotes the transpose of vectors and matrices, a necessary and sufficient condition for a test statistic  $T_n$  to be affine invariant is that it is based on the scalar products

$$Y_{n,i}^\top Y_{n,j} = (X_i - \bar{X}_n)^\top S_n^{-1} (X_j - \bar{X}_n), \quad i, j \in \{1, \dots, n\},$$

where  $Y_{n,j} = S_n^{-1/2} (X_j - \bar{X}_n)$ ,  $j = 1, \dots, n$ , are the so-called *scaled residuals* of  $X_1, \dots, X_n$ , see Henze (2002). Here,  $S_n^{-1/2}$  denotes the unique symmetric square root of  $S_n^{-1}$  which, due to the absolute continuity of the distribution of  $X$ , exists with probability one if  $n \geq d + 1$ , see Eaton and Perlman (1973). The latter condition is tacitly assumed to hold in what follows.

The novel idea for constructing a test of  $H_0$  is the following: Suppose  $X$  is a  $d$ -dimensional random vector, and assume that the moment generating function (MGF)  $m(t) = \mathbb{E}[\exp(t^\top X)]$  exists for each  $t \in \mathbb{R}^d$  and satisfies the

system of partial differential equations

$$\frac{\partial m(t)}{\partial t_j} = t_j m(t), \quad t = (t_1, \dots, t_d)^\top \in \mathbb{R}^d, \quad j = 1, \dots, d. \quad (2)$$

Writing  $y'(t)$  for the gradient of a function  $y : \mathbb{R}^d \rightarrow \mathbb{R}$  at  $t$ , (2) may be succinctly written as  $m'(t) = tm(t)$ ,  $t \in \mathbb{R}^d$ . Taking into account that  $m(0) = 1$ , the unique solution of this equation in the case  $d = 1$  is  $m(t) = \exp(t^2/2)$ , which is the MGF of the standard normal distribution. If  $d > 1$ , and we fix  $t_2, \dots, t_d$ , the solution of (2) for  $j = 1$  is

$$m(t) = c_2(t_2, \dots, t_d) \cdot e^{t_1^2/2} \quad (3)$$

for some function  $c_2 : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ . We thus have  $c_2(t_2, \dots, t_d) = e^{-t_1^2/2} \cdot m(t)$ ,  $t \in \mathbb{R}^d$ , which shows that  $c_2$  is differentiable. Moreover,

$$\frac{\partial}{\partial t_2} c_2(t_2, \dots, t_d) = e^{-t_1^2/2} \cdot \frac{\partial}{\partial t_2} m(t).$$

Using (2) with  $j = 2$  then gives

$$t_2 = \frac{\frac{\partial}{\partial t_2} c_2(t_2, \dots, t_d)}{c_2(t_2, \dots, t_d)}, \quad t \in \mathbb{R}^d,$$

the solution of which is  $c_2(t_2, \dots, t_d) = c_3(t_3, \dots, t_d) \cdot \exp(t_2^2/2)$  for some function  $c_3 : \mathbb{R}^{d-2} \rightarrow \mathbb{R}$ . Inserting this expression into (3) and continuing this way, we finally obtain

$$m(t) = \prod_{j=1}^d e^{t_j^2/2} = e^{\|t\|^2/2}, \quad t \in \mathbb{R}^d,$$

where  $\|\cdot\|$  denotes the Euclidean norm in  $\mathbb{R}^d$ . Notice that this unique solution of (2) is the MGF of the standard normal distribution  $N_d(0, I_d)$ , where  $I_d$  is the unit matrix of order  $d$ .

If  $X$  has some non-degenerate normal distribution, the scaled residuals  $Y_{n,1}, \dots, Y_{n,n}$  should be approximately independent, with a distribution close to  $N_d(0, I_d)$ , at least for large  $n$ . Hence, a natural approach for testing  $H_0$  is to consider the empirical MGF

$$M_n(t) = \frac{1}{n} \sum_{j=1}^n e^{t^\top Y_{n,j}} \quad (4)$$

of  $Y_{n,1}, \dots, Y_{n,n}$ , and to employ the weighted  $L^2$ -statistic

$$T_{n,\gamma} := n \int_{\mathbb{R}^d} \|M'_n(t) - tM_n(t)\|^2 w_\gamma(t) dt, \quad (5)$$

where

$$w_\gamma(t) = \exp(-\gamma\|t\|^2), \quad t \in \mathbb{R}^d. \quad (6)$$

Rejection of  $H_0$  is for large values of  $T_{n,\gamma}$ .

We stress that other choices than  $w_\gamma$  are conceivable in (5). The extremely appealing feature of using the weight function  $w_\gamma$ , however, is that with this choice the test statistic  $T_{n,\gamma}$  allows for an explicit form that does not involve any integration and is given below. The role of  $\gamma > 0$  will be discussed later.

Using the relations

$$\int_{\mathbb{R}^d} e^{t^\top \alpha} w_\gamma(t) dt = \left(\frac{\pi}{\gamma}\right)^{d/2} \exp\left(\frac{\|\alpha\|^2}{4\gamma}\right), \quad (7)$$

$$\int_{\mathbb{R}^d} e^{t^\top \alpha} t^\top \alpha w_\gamma(t) dt = \left(\frac{\pi}{\gamma}\right)^{d/2} \frac{\|\alpha\|^2}{2\gamma} \exp\left(\frac{\|\alpha\|^2}{4\gamma}\right),$$

$$\int_{\mathbb{R}^d} e^{t^\top \alpha} \|t\|^2 w_\gamma(t) dt = \left(\frac{\pi}{\gamma}\right)^{d/2} \exp\left(\frac{\|\alpha\|^2}{4\gamma}\right) \left(\frac{d}{2\gamma} + \frac{\|\alpha\|^2}{4\gamma^2}\right), \quad (8)$$

and putting  $Y_{n,j,k}^+ = Y_{n,j} + Y_{n,k}$ , the test statistic defined in (5) takes the form

$$T_{n,\gamma} = \frac{1}{n} \left(\frac{\pi}{\gamma}\right)^{d/2} \sum_{j,k=1}^n \exp\left(\frac{\|Y_{n,j,k}^+\|^2}{4\gamma}\right) \left(Y_{n,j}^\top Y_{n,k} - \frac{\|Y_{n,j,k}^+\|^2}{2\gamma} + \frac{d}{2\gamma} + \frac{\|Y_{n,j,k}^+\|^2}{4\gamma^2}\right), \quad (9)$$

which is amenable to computational purposes. Notice that  $T_{n,\gamma}$  is affine invariant.

The remainder of this paper is structured as follows: Section 2 deals with the convergence in distribution of  $T_{n,\gamma}$  under  $H_0$ , and Section 3 is devoted to the problem of consistency of the new test (which seems to be new even in the univariate case). In Section 4 we let  $\gamma$  tend to infinity while keeping  $n$  fixed. Under this setting,  $T_{n,\gamma}$  converges to a certain linear combination of two well-known measures of multivariate skewness. In Section 5 we compare the finite-sample power behavior of the new test to that of several classical and recent tests for both univariate and multivariate normality. Section 6 illustrates the procedures by means of a real data set. Section 7 presents some conclusions, while Section 8 contains several technical proofs which do not form part of the main text.

## 2 The limit null distribution of $T_{n,\gamma}$

In this section, we derive the limit distribution of the test statistic  $T_{n,\gamma}$  defined in (5) under the null hypothesis (1). In view of affine invariance, we will assume  $\mathbb{E}(X) = 0$  and  $\mathbb{E}(XX^\top) = I_d$ . Put

$$W_n(t) := \sqrt{n}(M_n'(t) - tM_n(t)) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top Y_{n,j}} (Y_{n,j} - t), \quad t \in \mathbb{R}^d, \quad (10)$$

and let  $L^2 := L^2(\mathbb{R}^d, \mathcal{B}^d, w_\gamma(t)dt)$  denote the separable Hilbert space of (equivalence classes of) measurable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  that are square integrable

with respect to the finite measure on the  $\sigma$ -field  $\mathcal{B}^d$  of Borel sets of  $\mathbb{R}^d$ , given by the weight function  $w_\gamma$ . The inner product and the resulting norm on  $L^2$  will be denoted by

$$\langle f, g \rangle = \int_{\mathbb{R}^d} f(t) g(t) w_\gamma(t) dt, \quad \|f\|_{L^2} = \sqrt{\langle f, f \rangle},$$

respectively. Since  $W_n(t)$  in (10) is  $\mathbb{R}^d$ -valued, we consider the Hilbert space  $\mathcal{H}$ , which is the  $d$ -fold (orthogonal) direct sum  $\mathcal{H} := L^2 \oplus \cdots \oplus L^2$ , consisting of all ordered  $d$ -tuples  $f = (f_1, \dots, f_d) \in L^2 \times \cdots \times L^2$ , equipped with the inner product

$$\langle f, g \rangle_{\oplus} := \langle f_1, g_1 \rangle_{L^2} + \cdots + \langle f_d, g_d \rangle_{L^2},$$

where  $f = (f_1, \dots, f_d)$ ,  $g = (g_1, \dots, g_d) \in \mathcal{H}$ . Notice that the norm  $\|\cdot\|_{\oplus}$  on  $\mathcal{H}$  satisfies

$$\|f\|_{\oplus}^2 = \sum_{i=1}^d \|f_i\|_{L^2}^2 = \int_{\mathbb{R}^d} \|f(t)\|^2 w_\gamma(t) dt, \quad f = (f_1, \dots, f_d) \in \mathcal{H}.$$

Since  $W_n$  is a random element of  $\mathcal{H}$  and  $T_n = \|W_n\|_{\oplus}^2$ , the aim is to prove  $W_n \xrightarrow{\mathcal{D}} W$  for some centred Gaussian random element of  $\mathcal{H}$ . By the Continuous Mapping Theorem (CMT), we would then have  $T_n \xrightarrow{\mathcal{D}} \|W\|_{\oplus}^2$ . Here and in the sequel,  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution of random elements (especially: of random variables). Moreover,  $o_{\mathbb{P}}(1)$  refers to convergence in probability to zero of random elements  $\mathcal{H}$ . The main result of this section is as follows.

**Theorem 1** (Convergence of  $W_n$ )

Suppose that  $X$  has some non-degenerate  $d$ -variate normal distribution. If the weight function  $w_\gamma$  in (6) satisfies  $\gamma > 2$ , there is some centred Gaussian random element  $W$  of  $\mathcal{H}$  having covariance (matrix) kernel  $K(s, t) = \mathbb{E}[W(s)W(t)^\top]$ , where

$$K(s, t) = e^{(\|s\|^2 + \|t\|^2)/2} \left( e^{s^\top t} (ts^\top + \mathbf{I}_d) - ts^\top - (1 + s^\top t) \mathbf{I}_d \right), \quad (11)$$

$s, t \in \mathbb{R}^d$ , so that  $W_n \xrightarrow{\mathcal{D}} W$  as  $n \rightarrow \infty$ .

**Corollary 1** The limit distribution of  $T_{n,\gamma}$  as  $n \rightarrow \infty$  under  $H_0$  is that of

$$\|W\|_{\oplus}^2 = \int_{\mathbb{R}^d} \|W(t)\|^2 w_\gamma(t) dt.$$

PROOF of Theorem 1. To highlight the main idea of the proof, let

$$Z_j(t) := e^{t^\top Y_{n,j}} (Y_{n,j} - t), \quad Z_j^\circ(t) := e^{t^\top X_j} (X_j - t), \quad (12)$$

and put

$$W_n^\circ(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j^\circ(t), \quad t \in \mathbb{R}^d. \quad (13)$$

Notice that  $Z_1^\circ, Z_2^\circ, \dots$  are i.i.d. (centred) random elements of  $\mathcal{H}$ . Writing  $X := X_1$ , and putting  $\rho := 2(\pi/\gamma)^{d/2}$ , the definition of  $Z_1^\circ$  and (7), (8) give

$$\begin{aligned} \mathbb{E}\|Z_1^\circ\|_\oplus^2 &= \mathbb{E}\left[\int_{\mathbb{R}^d} e^{2t^\top X} \|X - t\|^2 w_\gamma(t) dt\right] \\ &\leq 2\mathbb{E}\left[\|X\|^2 \int_{\mathbb{R}^d} e^{2t^\top X} w_\gamma(t) dt\right] + 2\mathbb{E}\left[\int_{\mathbb{R}^d} e^{2t^\top X} \|t\|^2 w_\gamma(t) dt\right] \\ &= \rho\mathbb{E}\left[\|X\|^2 \exp\left(\frac{\|X\|^2}{\gamma}\right)\right] + \rho\mathbb{E}\left[\|X\|^2 \exp\left(\frac{\|X\|^2}{\gamma}\right) \left(\frac{d}{2\gamma} + \frac{\|X\|^2}{\gamma^2}\right)\right]. \end{aligned} \quad (14)$$

Since  $X$  has a standard normal distribution in  $\mathbb{R}^d$ , the first expectation on the right-hand side is

$$\frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \|x\|^2 \exp\left(-\left(\frac{1}{2} - \frac{1}{\gamma}\right) \|x\|^2\right) dt.$$

This is finite if, and only if,  $\gamma > 2$ . Since the same conclusion can be drawn for the second expectation, the condition  $\gamma > 2$  implies  $\mathbb{E}\|Z_1^\circ\|_\oplus^2 < \infty$ .

By a Hilbert space central limit theorem (CLT), see e.g. Kundu et al. (2000), there is some centred Gaussian random element  $W^\circ$  of  $\mathcal{H}$  such that  $W_n^\circ \xrightarrow{\mathcal{D}} W^\circ$ . The idea now is to approximate  $W_n$  by a suitable random element  $\widetilde{W}_n$  of  $\mathcal{H}$  so that  $\|W_n - \widetilde{W}_n\|_\oplus = o_{\mathbb{P}}(1)$  as  $n \rightarrow \infty$ , and  $\widetilde{W}_n(t) = n^{-1/2} \sum_{j=1}^n \widetilde{Z}_j(t)$ , where  $\widetilde{Z}_1, \widetilde{Z}_2, \dots$  are i.i.d. centred random elements of  $\mathcal{H}$  satisfying  $\mathbb{E}\|\widetilde{Z}_1\|_\oplus^2 < \infty$ . The assertion would then follow from the CLT in Hilbert spaces and Slutsky's lemma. To this end, put

$$\Delta_{n,j} = Y_{n,j} - X_j = \left(S_n^{-1/2} - \mathbf{I}_d\right) X_j - S_n^{-1/2} \overline{X}_n, \quad j = 1, \dots, n. \quad (15)$$

A Taylor expansion gives

$$e^{t^\top \Delta_{n,j}} = 1 + t^\top \Delta_{n,j} + \frac{1}{2} (t^\top \Delta_{n,j})^2 \exp(\Theta_{n,j} t^\top \Delta_{n,j}), \quad (16)$$

where  $|\Theta_{n,j}| \leq 1$ . By some algebra, it follows that

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} (X_j - t) + A_n(t) + B_n(t) + C_n(t), \quad (17)$$

where

$$A_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} t^\top \Delta_{n,j} (X_j - t), \quad (18)$$

$$B_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} (1 + t^\top \Delta_{n,j}) \Delta_{n,j}, \quad (19)$$

$$C_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} \frac{1}{2} (t^\top \Delta_{n,j})^2 \exp(\Theta_{n,j} t^\top \Delta_{n,j}) (X_j - t + \Delta_{n,j}). \quad (20)$$

Notice that the first term on the right-hand side of (17) is  $W_n^\circ(t)$ , as given in (13). By Proposition 1, we have  $\|C_n\|_\oplus = o_{\mathbb{P}}(1)$ , and Proposition 2 yields

$$B_n(t) = -\frac{1}{\sqrt{n}} \exp\left(\frac{\|t\|^2}{2}\right) \sum_{j=1}^n \left(\frac{X_j X_j^\top t - t}{2} + X_j\right) + o_{\mathbb{P}}(1).$$

In view of Proposition 3, (17) implies  $W_n(t) = n^{-1/2} \sum_{j=1}^n \tilde{Z}_j(t) + o_{\mathbb{P}}(1)$ , where

$$\tilde{Z}_j(t) = e^{t^\top X_j} (X_j - t) - e^{\|t\|^2/2} (X_j X_j^\top t - t + X_j). \quad (21)$$

A straightforward calculation gives  $\mathbb{E}\tilde{Z}_j(t) = 0$ ,  $t \in \mathbb{R}^d$ . To show that  $\mathbb{E}\|\tilde{Z}_j\|_\oplus^2 < \infty$ , notice that (21) reads (for  $j = 1$  and putting  $X := X_1$ )  $\tilde{Z}_1(t) = Z_1^\circ(t) - U(t)$ , where  $Z_1^\circ(t)$  is given in (12), and  $U(t) = e^{\|t\|^2/2} (X X^\top t - t + X)$ . Then

$$\begin{aligned} \mathbb{E}\|\tilde{Z}_1\|_\oplus^2 &= \mathbb{E}\left[\int_{\mathbb{R}^d} \|Z_1^\circ(t) - U(t)\|^2 w_\gamma(t) dt\right] \\ &\leq 2\mathbb{E}\left[\int_{\mathbb{R}^d} \|Z_1^\circ(t)\|^2 w_\gamma(t) dt\right] + 2\mathbb{E}\left[\int_{\mathbb{R}^d} \|U(t)\|^2 w_\gamma(t) dt\right]. \end{aligned}$$

The first expectation on the right-hand side was tackled in (14) and was seen to be finite if  $\gamma > 2$ . As for the second expectation, notice that, by the Cauchy-Schwarz-inequality,  $\|U(t)\|^2 \leq 2e^{\|t\|^2} (\|X - t\|^2 + \|t\|^2 \|X\|^4)$ . In view of  $\|X - t\|^2 \leq 2\|X\|^2 + 2\|t\|^2$ , the second expectation on the right-hand side is easily seen to be finite if  $\gamma \geq 1$ . Thus, in view of the condition  $\gamma > 2$ , we have  $\mathbb{E}\|\tilde{Z}_j\|_\oplus^2 < \infty$ . Hence,  $\tilde{Z}_1, \tilde{Z}_2, \dots$  are i.i.d. centred random elements of  $\mathcal{H}$ , and the CLT in Hilbert spaces yields the assertion, since the kernel  $K$  figuring in (11) is given by  $K(s, t) = \mathbb{E}[\tilde{Z}_1(s)\tilde{Z}_1(t)^\top]$  (for details in computing  $K(s, t)$ , see Proposition 4).  $\square$

The following result provides some information on the limit null distribution of  $T_{n,\gamma}$ .

**Theorem 2** Let  $T_{\infty,\gamma}$  be a random variable with the distribution of  $\|W\|_\oplus^2$ , where  $W$  is given in Theorem 1. We then have

$$\mathbb{E}(T_{\infty,\gamma}) = \left(\frac{\pi}{\gamma-2}\right)^{d/2} \left(d + \frac{d}{2(\gamma-2)}\right) - \left(\frac{d(d+1)}{2(\gamma-1)} + d\right) \left(\frac{\pi}{\gamma-1}\right)^{d/2}.$$

$\gamma$	2.5	3	4	5	7
$\mathbb{E}(T_{\infty,\gamma})$	2.6013	0.7787	0.2022	0.0861	0.0277
$\mathbb{V}(T_{\infty,\gamma})$	4.7153	0.5430	0.0458	0.0094	0.0011

**Table 1:** Expectation and variance of the limit null distribution when  $d = 1$

PROOF. By Fubini's theorem, it follows that

$$\mathbb{E}[T_{\infty,\gamma}] = \int_{\mathbb{R}^d} \mathbb{E}\|W(t)\|^2 w_{\gamma}(t) dt.$$

Writing  $\text{tr}(D)$  for the trace of a square matrix  $D$ , we have

$$\begin{aligned} \mathbb{E}\|W(t)\|^2 &= \mathbb{E}[W(t)^{\top}W(t)] = \mathbb{E}[\text{tr}(W(t)^{\top}W(t))] \\ &= \mathbb{E}[\text{tr}(W(t)W(t)^{\top})] = \text{tr}(\mathbb{E}[W(t)W(t)^{\top}]) = \text{tr}(K(t,t)) \\ &= e^{\|t\|^2} \left( e^{\|t\|^2} (d + \|t\|^2) - (\|t\|^2 + d\|t\|^2) - d \right). \end{aligned}$$

Now, some straightforward algebra yields the assertion.  $\square$

In the univariate case  $d = 1$ , we also obtained an explicit expression for the variance of  $T_{\infty,\gamma}$ , by using the relation

$$\mathbb{V}(T_{\infty,\gamma}) = 2 \iint_{\mathbb{R}^2} K^2(s,t) w_{\gamma}(s) w_{\gamma}(t) ds dt.$$

By tedious calculations, it follows that

$$\mathbb{V}(T_{\infty,\gamma}) = 2\pi \left( \beta^{-1} + \beta^{-3} + \delta + \delta^3 + \frac{1}{4}(\beta^2 + 2)\delta^5 - 4\eta - 12\eta^3 - 16(2\beta^2 + 1)\eta^5 \right),$$

where  $\beta = \gamma - 1$ ,  $\delta = (\beta^2 - 1)^{-1/2}$  and  $\eta = (4\beta^2 - 1)^{-1/2}$ .

Table 1 contains expectation and variance of  $T_{\infty,\gamma}$  in the univariate case for various values of  $\gamma$ .

### 3 Consistency

In this section, let  $X$  have an absolutely continuous distribution, and suppose that  $m(t) = \mathbb{E}[\exp(t^{\top}X)]$  exists for each  $t \in \mathbb{R}^d$ . Notice that, by Lebesgue's dominated convergence theorem, this (strong) assumption implies that  $m$  is differentiable on  $\mathbb{R}^d$ , with a continuous derivative. In view of affine invariance, let without loss of generality  $\mathbb{E}[X] = 0$  and  $\mathbb{E}[XX^{\top}] = I_d$ .

**Theorem 3** We have

$$\liminf_{n \rightarrow \infty} \frac{T_{n,\gamma}}{n} \geq \int_{\mathbb{R}^d} \|m'(t) - tm(t)\|^2 w_{\gamma}(t) dt \quad \mathbb{P}\text{-almost surely.} \quad (22)$$



PROOF. Fix  $K > 0$ , and put  $S(K) := \{t \in \mathbb{R}^d : \|t\| \leq K\}$ . Let  $B(K)$  be the Banach space of continuous functions  $f : S(K) \rightarrow \mathbb{R}$ , equipped with the norm  $\|f\|_\infty = \max_{\|t\| \leq K} |f(t)|$ . Recall  $M_n(t)$  from (4), and put

$$M_{n,0}(t) = \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j}, \quad t \in \mathbb{R}^d.$$

Let  $\xi_n = \max_{j=1, \dots, n} \|\Delta_{n,j}\|$ , where  $\Delta_{n,j}$  is given in (15). From (15) we obtain

$$\xi_n \leq \|S_n^{-1/2} - \text{Id}\|_2 \cdot \max_{1 \leq j \leq n} \|X_j\| + \|S_n^{-1/2}\|_2 \cdot \|\bar{X}_n\|. \quad (23)$$

Since the existence of  $m(t)$  implies  $\mathbb{E}\|X\|^4 < \infty$ , Theorem 5.2 of Barndorff-Nielsen (1963) yields

$$n^{-1/4} \max_{1 \leq j \leq n} \|X_j\| \rightarrow 0 \quad \mathbb{P}\text{-almost surely.} \quad (24)$$

From  $S_n - \text{Id} = n^{-1} \sum_{j=1}^n (X_j X_j^\top - \text{Id}) - \bar{X}_n \bar{X}_n^\top$  and Kolmogorov's variance criterion for averages (see Kallenberg (2002), p. 73), we have  $\lim_{n \rightarrow \infty} n^{1/2-\varepsilon} \|S_n - \text{Id}\|_2 = 0$   $\mathbb{P}$ -a.s. for each  $\varepsilon > 0$ . Hence Proposition 5 implies

$$\lim_{n \rightarrow \infty} \xi_n = 0 \quad \mathbb{P}\text{-a.s.} \quad (25)$$

Using the notation  $\|f\|_\infty = \max_{\|t\| \leq K} \|f(t)\|$  also for a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , (16) gives

$$\|M_n - M_{n,0}\|_\infty \leq \|M_{n,0}\|_\infty \cdot \xi_n \cdot K \cdot \left(1 + \frac{K\xi_n}{2} e^{K\xi_n}\right). \quad (26)$$

By the strong law of large numbers (SLLN) in  $B(K)$ , the first factor on the right-hand side converges almost surely to  $\|m\|_\infty$ , and thus (25) entails

$$\lim_{n \rightarrow \infty} \|M_n - M_{n,0}\|_\infty = 0 \quad \mathbb{P}\text{-a.s.} \quad (27)$$

Putting

$$F_n = \max_{j=1, \dots, n} \|X_j\|, \quad (28)$$

the triangle inequality gives

$$\|M'_n - M'_{n,0}\|_\infty \leq (\xi_n + F_n) \cdot \|M_n - M_{n,0}\|_\infty + \xi_n \cdot \|M_{n,0}\|_\infty.$$

Invoking (26), (23), (24) and Proposition 5, we have

$$\lim_{n \rightarrow \infty} \|M'_n - M'_{n,0}\|_\infty = 0 \quad \mathbb{P}\text{-a.s.} \quad (29)$$

Writing  $\text{id}$  for the identity function on  $\mathbb{R}^d$ , the triangle inequality yields

$$\|M'_n - \text{id} \cdot M_n\|_\infty \leq \|M'_n - M'_{n,0}\|_\infty + \|M'_{n,0} - \text{id} \cdot M_{n,0}\|_\infty + K \cdot \|M_{n,0} - M_n\|_\infty.$$

From (27), (29) and the SLLN in  $B(K)$  it follows that

$$\limsup_{n \rightarrow \infty} \|M'_n - \text{id} \cdot M_n\|_\infty \leq \|m' - \text{id} \cdot m\|_\infty \quad \mathbb{P}\text{-a.s.} \quad (30)$$

Likewise, we have

$$\|M'_{n,0} - \text{id} \cdot M_{n,0}\|_\infty \leq \|M'_{n,0} - M'_n\|_\infty + \|M'_n - \text{id} \cdot M_n\|_\infty + K \cdot \|M_n - M_{n,0}\|_\infty$$

and thus

$$\|m' - \text{id} \cdot m\|_\infty \leq \liminf_{n \rightarrow \infty} \|M'_n - \text{id} \cdot M_n\|_\infty \quad \mathbb{P}\text{-a.s.} \quad (31)$$

Upon combining (30) and (31), and using Fatou's lemma, we obtain

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{T_{n,\gamma}}{n} &\geq \liminf_{n \rightarrow \infty} \int_{S(K)} \|M'_n(t) - tM_n(t)\|^2 w_\gamma(t) dt \\ &\geq \int_{S(K)} \liminf_{n \rightarrow \infty} \|M'_n(t) - tM_n(t)\|^2 w_\gamma(t) dt \\ &= \int_{S(K)} \|m'(t) - tm(t)\|^2 w_\gamma(t) dt \quad \mathbb{P}\text{-a.s.} \end{aligned}$$

Since  $K$  was arbitrary, the assertion follows.  $\square$

*Remark 1* If  $X$  has some non-degenerate non-normal distribution with existing moment generating function, then  $m'(t) \neq tm(t)$  for at least one  $t$ . Since the weight function  $w_\gamma$  is strictly positive on  $\mathbb{R}^d$  and the function  $m'(t) - tm(t)$ ,  $t \in \mathbb{R}^d$ , is continuous, the right-hand side of (22) is strictly positive, and thus

$$\lim_{n \rightarrow \infty} T_{n,\gamma} = \infty \quad \mathbb{P}\text{-a.s.} \quad (32)$$

Hence, the test is consistent against each such alternative. We conjecture that (32) holds for *any* non-normal alternative distribution. A proof of such a result, however, remains an open problem.

*Remark 2* Regarding consistency, one may relax the assumption that  $\mathbb{E}e^{t^\top X} < \infty$  for each  $t \in \mathbb{R}^d$ . If  $\mathbb{E}e^{t^\top X} < \infty$  for each  $t \in N$ , where  $N$  is some open neighborhood of the origin in  $\mathbb{R}^d$ , the proof given above yields

$$\liminf_{n \rightarrow \infty} \frac{T_{n,\gamma}}{n} \geq \int_N \|m'(t) - tm(t)\|^2 w_\gamma(t) dt \quad \mathbb{P}\text{-a.s.},$$

since, mutatis mutandis, this inequality holds if  $N$  is replaced by any compact subset of  $N$ . If  $m'(t) \neq tm(t)$  for at least one  $t \in N$ , then (32) continues to hold.

#### 4 The limit $\gamma \rightarrow \infty$

In this section, we will show that, for fixed  $n$ , the statistic  $T_{n,\gamma}$ , after a suitable scaling, approaches a linear combination of two well-known measures of multivariate skewness as  $\gamma \rightarrow \infty$ .

**Theorem 4** We have (pointwise on the underlying probability space)

$$\lim_{\gamma \rightarrow \infty} \gamma^{2+d/2} \frac{16T_{n,\gamma}}{n\pi^{d/2}} = 2b_{1,d} + \tilde{b}_{1,d},$$

where

$$b_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^n (Y_{n,j}^\top Y_{n,k})^3 \quad (33)$$

is nonnegative invariant sample skewness in the sense of Mardia (1970), and

$$\tilde{b}_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^n Y_{n,j}^\top Y_{n,k} \|Y_{n,j}\|^2 \|Y_{n,k}\|^2$$

denotes sample skewness in the sense of Móri, Rohatgi and Székely, as defined in Móri et al. (1993).

PROOF. We start with (9) and use

$$\exp\left(\frac{\|Y_{n,j} + Y_{n,k}\|^2}{4\gamma}\right) = 1 + \frac{\|Y_{n,j} + Y_{n,k}\|^2}{4\gamma} + \frac{\|Y_{n,j} + Y_{n,k}\|^4}{32\gamma^2} + O(\gamma^{-3})$$

as  $\gamma \rightarrow \infty$ . Multiplying this expression with the term within the rightmost bracket of (9), and using the relations  $\sum_{j=1}^n Y_{n,j} = 0$ ,  $\sum_{j=1}^n \|Y_{n,j}\|^2 = nd$ ,  $\sum_{j,k=1}^n \|Y_{n,j} + Y_{n,k}\|^2 = 2n^2d$ ,

$$\begin{aligned} \sum_{j,k=1}^n \|Y_{n,j} + Y_{n,k}\|^4 &= 2n^2 \left( \frac{1}{n} \sum_{j=1}^n \|Y_{n,j}\|^4 + d^2 + 2d \right), \\ \sum_{j,k=1}^n \|Y_{n,j} + Y_{n,k}\|^4 Y_{n,j}^\top Y_{n,k} &= 8 \sum_{j,k=1}^n (Y_{n,j}^\top Y_{n,k})^2 \|Y_{n,j}\|^2 + 4n^2 b_{1,d} + 2n^2 \tilde{b}_{1,d} \end{aligned}$$

as well as

$$\begin{aligned}
\sum_{j,k=1}^n (Y_{n,j}^\top Y_{n,k})^2 \|Y_{n,j}\|^2 &= \sum_{j,k=1}^n \text{tr} (Y_{n,j}^\top Y_{n,k} Y_{n,k}^\top Y_{n,j} \|Y_{n,j}\|^2) \\
&= \text{tr} \left( \sum_{k=1}^n Y_{n,k} Y_{n,k}^\top \sum_{j=1}^n Y_{n,j} Y_{n,j}^\top \|Y_{n,j}\|^2 \right) \\
&= \text{tr} \left( n \mathbf{I}_d \sum_{j=1}^n Y_{n,j} Y_{n,j}^\top \|Y_{n,j}\|^2 \right) \\
&= n \sum_{k=1}^n \text{tr} (Y_{n,j}^\top Y_{n,j} \|Y_{n,j}\|^2) \\
&= n \sum_{j=1}^n \|Y_{n,j}\|^4,
\end{aligned}$$

the result follows by tedious but straightforward algebra.  $\square$

*Remark 3* It is interesting to note a similarity between Theorem 4 and Theorem 2.1 of Henze (1997), who showed that the BHEP-statistic for testing for multivariate normality, after suitable rescaling, approaches the linear combination  $2b_{1,d} + 3\tilde{b}_{1,d}$  as a smoothing parameter (called  $\beta$  in that paper) tends to 0. Since  $\beta$  and  $\gamma$  are related by  $\beta = \gamma^{-1/2}$ , this corresponds to letting  $\gamma$  tend to infinity. The same linear combination  $2b_{1,d} + 3\tilde{b}_{1,d}$  also showed up as a limit statistic in Henze et al. (2018). Notice that, in the univariate case, the limit statistic figuring in Theorem 4 is nothing but three times squared sample skewness. It should be stressed that tests for multivariate normality based on  $b_{1,d}$  or  $\tilde{b}_{1,d}$  (or on related measures of multivariate skewness and kurtosis) lack consistency against general alternatives, see, e.g., Baringhaus and Henze (1991), Baringhaus and Henze (1992), Henze (1994a), and Henze (1994b).

## 5 Monte Carlo results

In this section we compare the finite-sample power performance of the newly proposed test to those of several competing tests for normality, both for the univariate and the multivariate case. In the case  $d = 1$  the competing procedures are

- a) the Cramér-von Mises (*CvM*) test,
- b) the Anderson-Darling (*AD*) test,
- c) the Shapiro-Wilk (*SW*) test,
- d) the Jarque-Bera (*JB*) test,
- e) the Zghoul (*Z*) test.

The R package *nortest* contains the functions `cvm.test` and `ad.test`, which can be used in order to calculate the test statistic and the associated p-value

for each of the first two tests mentioned above, see Gross and Ligges (2015). The Shapiro-Wilk test can be performed using the *Shapiro.test* function included in the *stats* package. The R package *tseries* contains a function called *jarque.bera.test*, which can be used in order to calculate the test statistic and p-value associated with the fourth test mentioned above, see Trapletti and Hornik (2017). Each of these tests are well-known and will not be discussed further.

The test of Zghoul (see Zghoul (2010)) is based on the empirical moment generating function. Zghoul (2010) includes a Monte Carlo study which indicates that the finite-sample power performance of the test compares favorably to that of its competitors, especially against symmetric alternatives with kurtosis higher than that of the normal distribution. However, the mentioned paper fails to provide the mathematical theory underlying this test. Henze and Koch (2017) more recently provided this theory, including a proof that the test is consistent against general alternatives.

The test statistic of Zghoul is a weighted  $L^2$ -statistic, given by

$$Z_n(\gamma) = n \int_{\mathbb{R}} (M_n(t) - m(t))^2 \exp(-\gamma t^2) dt,$$

where  $\gamma > 2$  is a smoothing parameter. Based on the finite-sample performance reported in Zghoul (2010), the author recommended setting  $\gamma$  equal to either 3 or 15. The numerical results presented below include the powers obtained using both of these values for the smoothing parameter; the resulting tests are denoted by  $Z_3$  and  $Z_{15}$  respectively. The test statistic  $Z_n(\gamma)$  can be rewritten in the computationally amenable form

$$Z_n(\gamma) = \sqrt{\pi} \left[ \frac{n}{\sqrt{\beta}} - \frac{2}{\sqrt{\gamma - \frac{1}{2}}} \sum_{i=1}^n \exp\left(\frac{Y_{n,i}^2}{4\beta + 2}\right) + \frac{1}{n\sqrt{\gamma}} \sum_{i,j=1}^n \exp\left(\frac{(Y_{n,i} + Y_{n,j})^2}{4\gamma}\right) \right],$$

where  $\beta = \gamma - 1$ . This test rejects normality for large values of the test statistic.

Turning our attention to the multivariate case, we consider the finite-sample power performance of the newly proposed test to those of some prominent competing tests, and to two very recent tests for multinormality. These procedures are

- a) Mardia's tests based on skewness and kurtosis,
- b) the Henze-Zirkler test,
- c) the *energy test* of Székely and Rizzo,
- d) a recent test of Henze, Jiménez-Gamero and Meintanis,
- e) a recent test of Henze and Jiménez-Gamero.

#### a): Mardia's tests based on skewness and kurtosis

Mardia's test for multinormality based on sample skewness rejects  $H_0$  for large values of  $b_{1,d}$ , where  $b_{1,d}$  is given in (33). Notice that  $b_{1,d}$  is a consistent estimator of  $\beta_{1,d} = E(X_1^T X_2)^3$ . Under normality we have  $\beta_{1,d} = 0$ , and the limit distribution of  $nb_{1,d}$  as  $n \rightarrow \infty$  is  $6\chi_{d(d+1)(d+2)/6}^2$ , see Mardia (1970).

The limit distribution of  $nb_{1,d}$  under elliptical symmetry has been derived by Baringhaus and Henze (1992).

Sample kurtosis in the sense of Mardia is given by

$$b_{2,d} = \frac{1}{n} \sum_{j=1}^n \|Y_{n,j}\|^4,$$

which is an estimator of  $\beta_{2,d} = \mathbb{E}\|X_1\|^4$ .

Under normality, we have  $\beta_{2,d} = d(d+2)$ , and the limit null distribution of  $\sqrt{n}(b_{2,d} - d(d+2))$  is the normal distribution  $N(0, 8d(d+2))$ , see Mardia (1970). The test based on  $b_{2,d}$  rejects  $H_0$  for large and small values of  $\beta_{2,d}$ .

The R package *QuantPsyc* contains a function *mult.norm*, which calculates both Mardia's skewness and kurtosis measures as well as the p-values associated with the corresponding tests from multivariate normality, see Fletcher (2012). Below we denote the tests based on Mardia's skewness and kurtosis by  $MS_n$  and  $MK_n$ , respectively.

We stress that there are several other measures of skewness and kurtosis, see Sections 3 and 4 of Henze (2002). The deficiencies of such measures as statistics for supposedly "directed tests" for multivariate normality have been addressed in Baringhaus and Henze (1991), Baringhaus and Henze (1992), Henze (1994a) as well as Henze (1994b).

#### b): The Henze-Zirkler test

Writing  $\Psi_n(t) = n^{-1} \sum_{k=1}^n \exp(it^\top Y_{n,k})$  for the empirical characteristic function of the scaled residuals  $Y_{n,1}, \dots, Y_{n,n}$ , Henze and Zirkler (1990) proposed the test statistic

$$HZ_n(\gamma) = (2\pi\gamma^2)^{-d/2} \int_{\mathbb{R}^d} \left| \Psi_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right) \right|^2 \exp\left(-\frac{\|t\|^2}{2\gamma^2}\right) dt,$$

for some fixed  $\gamma > 0$ . The test statistic can be rewritten as

$$\begin{aligned} HZ_n(\gamma) &= \frac{1}{n^2} \sum_{j,k=1}^n \exp\left(-\frac{\gamma^2}{2} \|Y_{n,j} - Y_{n,k}\|^2\right) \\ &\quad - 2(1 + \gamma^2)^{-d/2} \frac{1}{n} \sum_{j=1}^n \exp\left(-\frac{\gamma^2 \|Y_{n,j}\|^2}{2(1 + \gamma^2)}\right) + (1 + 2\gamma^2)^{-d/2}. \end{aligned}$$

The Henze-Zirkler test is obtained by setting

$$\gamma = \frac{1}{\sqrt{2}} \left( \frac{(2d+1)n}{4} \right)^{1/(d+4)}.$$

This choice of  $\gamma$  corresponds to the optimal bandwidth for a multivariate non-parametric density estimator with a Gaussian kernel. The Henze-Zirkler test is included because of its impressive power performance reported in previous studies, see, e.g., Mecklin and Mundfrom (2005).

The Henze-Zirkler test (denoted  $HZ_n$  below) is programmed in the function *hzTest* in the R package *MVN*, see Korkmaz et al. (2014). This test rejects normality for large values of the test statistic.  $HZ_n$  is equivalent to a test by Bowman and Foster (1993), as remarked in Henze (2002).

**c): The energy test**

Székely and Rizzo (2005) proposed the test statistic

$$EN_n = n \left( \frac{2}{n} \sum_{j=1}^n \mathbb{E} \|Y_{n,j} - Z\| - \frac{2\Gamma((d+1)/2)}{\Gamma(d/2)} - \frac{1}{n^2} \sum_{j,k=1}^n \mathbb{E} \|Y_{n,j} - Y_{n,k}\| \right).$$

Here, the first expectation is with respect to a random vector  $Z$ , which is independent of  $Y_{n,j}$  and has the distribution  $N_d(0, I_d)$ , and

$$\begin{aligned} \mathbb{E} \|a - Z\| &= \sqrt{2} \frac{2\Gamma((d+1)/2)}{\Gamma(d/2)} + \sqrt{\frac{2}{\pi}} \sum_{k=0}^{\infty} \left\{ \frac{(-1)^k}{k!2^k} \frac{\|a\|^{2k+2}}{(2k+1)(2k+2)} \right. \\ &\quad \left. \times \frac{2\Gamma((d+1)/2)\Gamma(k+1.5)}{\Gamma((d/2)+k+1)} \right\}. \end{aligned}$$

This test, denoted by  $EN_n$ , is known as the energy test. Rejection of  $H_0$  is for large values of  $EN_n$ . The function *mvnorm.etest* in the R package *energy* calculates this test statistic as well as the corresponding p-value, see Rizzo and Székely (2016). The energy test is also reported to have excellent power performance, see Joensuu and Vogel (2014).

**d): The Henze–Jiménez–Gamero–Meintanis test**

By generalizing a characterization of normality involving both the characteristic and the moment generating function (see Volkmer (2014)) to the multivariate case, Henze et al. (2018) proposed to base a test of  $H_0$  on the weighted  $L^2$ -statistic

$$HM_n = n \int_{\mathbb{R}^d} \left( \frac{1}{n^2} \sum_{j=1}^n \cos(t^\top Y_{n,j}) \sum_{j=1}^n \exp(t^\top Y_{n,j}) - 1 \right)^2 \exp(-\gamma \|t\|^2) dt,$$

for some parameter  $\gamma > 1$ . Putting  $Y_{jk}^\pm = Y_{n,j} \pm Y_{n,k}$ ,  $HM_n$  can be rewritten as

$$\begin{aligned} HM_n &= \left( \frac{\pi}{\gamma} \right)^{d/2} \left\{ \frac{1}{2n^3} \sum_{j,k,\ell,m=1}^n \left\{ \exp\left( \frac{\|Y_{jk}^+\|^2 - \|Y_{\ell m}^-\|^2}{4\gamma} \right) \cos\left( \frac{Y_{jk}^{+\top} Y_{\ell m}^-}{2\gamma} \right) \right. \right. \\ &\quad \left. \left. + \exp\left( \frac{\|Y_{jk}^+\|^2 - \|Y_{\ell m}^+\|^2}{4\gamma} \right) \cos\left( \frac{Y_{jk}^{+\top} Y_{\ell m}^+}{2\gamma} \right) \right\} \right. \\ &\quad \left. - \frac{2}{n} \sum_{j,k=1}^n \exp\left( \frac{\|Y_{n,j}\|^2 - \|Y_{n,k}\|^2}{4\gamma} \right) \cos\left( \frac{Y_{n,j}^\top Y_{n,k}}{2\gamma} \right) + n \right\}. \end{aligned}$$

The test based on  $HM_n$  uses an upper rejection region.

A disadvantage of this test is the need to calculate a four-fold sum in order to evaluate the test statistic. Hence, the amount of computer time required in order to perform this test is substantially more than the time required for the other tests under consideration.

**e): The Henze–Jiménez–Gamero test**

Henze and Jiménez–Gamero (2018) present a multivariate generalization of a recent class of tests for univariate normality (see Henze and Koch (2017)) based on the empirical moment generating function. The test statistic is

$$HJ_n = n \int_{\mathbb{R}^d} (M_n(t) - m(t))^2 \exp(-\beta \|t\|^2) dt,$$

where  $\beta > 1$  is a fixed parameter. An alternative representation of  $HJ_n$  is

$$HJ_n = \pi^{d/2} \left\{ \frac{1}{n} \sum_{j,k=1}^n \frac{1}{\beta^{d/2}} \exp\left(\frac{\|Y_{n,j} + Y_{n,k}\|^2}{4\beta}\right) + \frac{n}{(\beta-1)^{d/2}} \right. \\ \left. - 2 \sum_{j=1}^n \frac{1}{(\beta-1/2)^{d/2}} \exp\left(\frac{\|Y_{n,j}\|^2}{4\beta-2}\right) \right\},$$

which is amenable to computation. This test rejects  $H_0$  for large values of the test statistic.

### 5.1 Power results

This subsection contains the results of a power study that comprises several univariate and multivariate distributions, using the software package R, see R Core Team (2015). With the exception of the computer-intensive statistic  $HM_n$ , critical values are each based on  $10^6$  replications under  $H_0$  (for  $HM_n$  we used  $10^5$  replications). Each power estimate presented below is based on 10 000 replications.

Power results are reported for a sample size of  $n = 50$  and  $d \in \{1, 2, 3, 5\}$ . A nominal significance level of  $\alpha = 0.05$  is used throughout. As is pointed out in Mecklin and Mundfrom (2005), the maximum possible standard error for each power estimate is 0.005. Thus, the 99% confidence interval for each reported power estimate is contained in the interval obtained by subtracting 1% from, and adding 1% to, the stated power.

The results for the univariate case are given in Table 2. The entries are percentages of rejection of  $H_0$ , rounded to the nearest integer. The power against the standard normal distribution shows that the nominal level is maintained very closely. As for the alternative distributions, NMIX1 and NMIX2 denote mixtures of the normal distributions  $N(0, 1)$  and  $N(0, 4)$ . The mixture NMIX1 gives equal weight to these distributions, while NMIX2 is obtained when the



probability of sampling from the standard normal distribution is 0.75. The remaining alternative distributions considered are  $t$ -distributions with 3, 5 and 10 degrees of freedom, the lognormal distributions with parameters  $(0, 1/2)$  and  $(0, 1/4)$  (denoted by  $LN(\cdot)$ ), the  $\chi^2$ -distributions with 5 and 15 degrees of freedom, the standard logistic distribution, the Weibull distributions with shape parameters 10 and 20, the Pearson type VII distributions with 5 and 10 degrees of freedom (denoted by  $P_{VII}(\cdot)$ ), and the skew-normal law with skewness parameters 3 and 5 (denoted by  $SN(\cdot)$ ), see Azzalini (1985).

The R Package *PearsonDS* contains the function *rpearsonVII*, which can be employed to simulate random variables from this distribution, see Becker and Klößner (2017). The R Package *sn* disposes of the function *rsn*, which can be used to simulate random variates from the skew normal distribution, see Azzalini (2017).

Tables 2, 3, 4 and 5 report the powers calculated in the cases where  $d$  equals 1, 2, 3 and 5, respectively. Note that the subscript  $n$  in the names of the test statistics is omitted in the tables in order to save space.

	<i>CvM</i>	<i>AD</i>	<i>SW</i>	<i>JB</i>	$Z_3$	$Z_{15}$	$T_{2.5}$	$T_5$	$T_{10}$	$T_\infty$
$N(0, 1)$	5	5	5	5	5	5	5	5	5	5
NMIX1	18	20	21	28	24	20	24	23	21	18
NMIX2	19	22	28	37	34	28	34	32	30	26
$t(3)$	58	61	63	69	65	57	65	63	60	52
$t(5)$	28	31	37	44	41	35	41	39	37	32
$t(10)$	12	13	16	21	20	17	20	20	19	17
$LN(0, \frac{1}{2})$	83	87	93	85	80	89	76	85	88	91
$LN(0, \frac{1}{4})$	31	35	44	39	37	45	34	40	44	47
$\chi^2(5)$	74	81	89	75	69	82	62	74	80	83
$\chi^2(15)$	30	34	43	36	34	43	31	38	41	45
<i>Logistic</i> (0, 1)	14	16	19	26	24	20	24	23	21	19
<i>Weibull</i> (10)	25	28	34	30	28	36	25	31	35	37
<i>Weibull</i> (20)	39	44	53	46	44	53	40	48	52	55
$P_{VII}(5)$	27	30	36	43	40	35	41	39	37	32
$P_{VII}(10)$	10	12	16	21	20	17	20	19	18	16
$SN(3)$	31	34	40	32	30	39	25	33	37	41
$SN(5)$	53	59	67	49	43	58	36	49	55	61

**Table 2:** Monte Carlo power estimates in the univariate case,  $\alpha = 0.05$

The results shown in Table 2 indicate that the newly proposed class of tests exhibit substantial power against the distributions considered.  $T_\infty$  outperforms each of the competing tests in terms of the estimated powers against the  $LN(0, 1/4)$ ,  $\chi^2(15)$  and  $SN(3)$  laws, as well as both of the Weibull distributions considered. The newly proposed tests also proves to be serious competitors against each of the remaining distributions considered, especially for small values of  $\gamma$  (in which case  $T_{n,\gamma}$  is often only outperformed by the Jarque-Bera test).

We now turn our attention to the case  $d > 1$ . As was the case for the univariate tests, the powers against 16 alternative distributions are reported

for each of the data dimensions considered. The powers of each of the tests against the standard normal distribution are also included in the relevant tables. The alternative distributions considered include mixtures of normal laws, distributions with independent marginals, distributions where each marginal is normal except for one, a spherically symmetric distribution, and a distribution for which every marginal is normal, but the joint distribution does not follow the normal law.

The parameter combinations used for the mixtures of normal distributions were taken from Székely and Rizzo (2005). Let  $pN_d(\mu_1, \Sigma_1) + (1-p)N_d(\mu_2, \Sigma_2)$  denote a normal mixture, where the probability of sampling from  $N_d(\mu_1, \Sigma_1)$  is  $p$  and the probability of sampling from  $N_d(\mu_2, \Sigma_2)$  is  $1-p$ . Let  $\mu = 0$  and  $\mu = 3$  denote  $d$ -dimensional column vectors of 0's and 3's, respectively, and let  $B_d$  denote a  $(d \times d)$ -matrix containing 1's on the main diagonal and 0.9's on each off diagonal entry. The normal mixtures are constructed by combining  $N_d(0, I_d)$ ,  $N_d(3, I_d)$  and  $N_d(0, B_d)$ . The first mixture, denoted by NMIX1, is  $0.9N_d(0, I_d) + 0.1N_d(3, I_d)$ . This distribution is skewed with heavy tails. The second mixture, denoted by NMIX2, is  $0.9N_d(0, B_d) + 0.1N_d(0, I_d)$ . This is a symmetric, heavy-tailed distribution. In addition, we included two multivariate  $t$ -distributions; the  $t_\nu(0, I_d)$ -distribution for  $\nu = 5$  and  $\nu = 10$ . Next, we included distributions with independent marginals, the latter being the  $\chi^2$ -distribution with 15 and 20 degrees of freedom respectively, the *logistic*(0, 1) distribution, the gamma distributions with parameters (5, 1) and (4, 2), as well as the Pearson Type VII distributions with 10 and 20 degrees of freedom.

Three  $d$ -dimensional distributions are obtained by combining  $d - 1$  independent standard normal marginals with one non-normal distribution. This distribution is denoted by  $N(0, 1)^{d-1} \otimes F$ , where  $F$  denotes the non-normal marginal distribution. The three alternatives considered for  $F$  are the  $\chi^2$ -distributions with 5 and 10 degrees of freedom, respectively, as well as the  $t$ -distribution with 3 degrees of freedom.

Spherically symmetric distributions can be defined in R using the *Elliptical Distribution* function from the R package *distrEllipse*, see Ruckdeschel et al. (2006). Tables 3–5 display the estimated powers of the various tests considered against the  $d$ -dimensional spherically symmetric distribution, where the radius of the distribution follows a lognormal distribution with parameters 0 and 0.5. This distribution is denoted by  $\mathcal{S}^d(LN(0, 1/2))$ .

Let  $\rho_d$  and  $\rho'_d$  denote positive definite  $(d \times d)$ -matrices with 1's on the main diagonal, where  $\rho_d$  has the constant  $\rho$  and  $\rho'_d$  the constant  $-\rho$  on each off diagonal entry. The final distribution considered is the mixture  $0.5N_d(0, \rho_d) + 0.5N_d(0, \rho'_d)$ . This distribution is a non-normal  $d$ -variate distribution with normal marginals.

As was the case in the univariate setting, the newly proposed test is associated with several high powers reported in Tables 3, 4, and 5. When comparing the results for the distributions with independent marginals, we see that  $T_\infty$  outperforms each of the competitors against both of the distributions with  $\chi^2$  marginals as well as both of the distributions with gamma marginals considered. This is also the case against the  $N(0, 1)^{d-1} \otimes \chi^2(5)$  and  $N(0, 1)^{d-1} \otimes \chi^2(10)$

	<i>MS</i>	<i>MK</i>	<i>HZ</i>	<i>EN</i>	<i>HM</i>	<i>HJ</i>	$T_{2.5}$	$T_5$	$T_{10}$	$T_\infty$
$N(0, 1)$	5	5	5	5	5	5	5	5	5	5
NMIX1	85	34	75	82	57	73	48	69	80	86
NMIX2	44	48	29	38	57	53	55	54	52	44
$t_5(0, I_2)$	53	62	42	51	67	60	60	60	58	53
$t_{10}(0, I_2)$	24	26	14	19	32	29	29	29	28	25
$(\chi^2(15))^2$	49	19	34	42	26	41	30	39	45	52
$(\chi^2(20))^2$	40	16	27	33	24	34	25	32	37	42
Logistic $(0, 1)^2$	24	27	15	19	33	28	28	29	28	25
Gamma $(5, 1)^2$	67	27	52	61	38	57	41	54	62	70
Gamma $(4, 2)^2$	76	32	64	72	42	66	48	62	71	78
$PVII(10)^2$	20	21	11	14	27	23	24	24	23	20
$PVII(20)^2$	11	10	7	8	14	12	13	13	12	12
$N(0, 1) \otimes t(3)$	47	52	42	49	61	55	56	56	54	47
$N(0, 1) \otimes \chi^2(5)$	63	25	52	60	36	52	39	49	57	65
$N(0, 1) \otimes \chi^2(10)$	38	15	26	32	21	32	24	30	35	40
$S^2(LN(0, \frac{1}{2}))$	26	25	15	21	29	30	31	31	29	26
$NM_2(\rho = 0.2)$	6	6	5	6	6	6	6	6	6	6

Table 3: Monte Carlo power estimates in the multivariate case for  $d = 2$ 

	<i>MS</i>	<i>MK</i>	<i>HZ</i>	<i>EN</i>	<i>HM</i>	<i>HJ</i>	$T_{2.5}$	$T_5$	$T_{10}$	$T_\infty$
$N(0, 1)^3$	5	5	5	5	5	5	5	5	5	5
NMIX1	89	36	81	91	59	72	43	66	82	91
NMIX2	71	76	49	66	79	79	79	80	78	72
$t_5(0, I_3)$	68	78	55	68	77	73	71	73	73	69
$t_{10}(0, I_3)$	34	38	18	27	35	38	36	38	38	34
$(\chi^2(15))^3$	52	21	35	49	27	42	31	39	47	55
$(\chi^2(20))^3$	40	16	26	37	21	33	24	30	36	44
Logistic $(0, 1)^3$	28	30	15	22	33	31	30	31	31	28
Gamma $(5, 1)^3$	72	30	53	69	39	58	41	53	65	75
Gamma $(4, 2)^3$	80	36	65	79	46	66	47	61	73	83
$PVII(10)^3$	22	22	10	16	24	25	25	26	25	23
$PVII(20)^3$	12	10	6	8	14	13	13	13	13	12
$N(0, 1)^2 \otimes t(3)$	42	43	29	40	54	48	49	49	48	43
$N(0, 1)^2 \otimes \chi^2(5)$	47	18	33	46	28	39	29	36	43	51
$N(0, 1)^2 \otimes \chi^2(10)$	26	12	17	24	16	22	17	21	24	28
$S^3(LN(0, \frac{1}{2}))$	53	58	18	43	62	58	57	58	58	54
$NM_3(\rho = 0.2)$	8	7	5	6	7	8	8	8	8	8

Table 4: Monte Carlo power estimates in the multivariate case for  $d = 3$ 

distributions. The mentioned predominance is for each of the data dimensions considered. Furthermore, the newly proposed test statistic shows high power against the remaining distributions for finite values of  $\gamma$ . Specifically, when  $d = 2$  the new test outperforms its competitors against the spherically symmetric distribution. In the case  $d = 3$ , none of the competing tests are able to outperform the newly proposed class of tests against the second normal mixture considered, the  $t$ -distribution with 10 degrees of freedom or the  $PVII(10)$  distribution. The corresponding list of distributions in the case where  $d = 5$  is obtained by substituting the  $t$ -distribution with 10 degrees of freedom for the  $t$ -distribution with 5 degrees of freedom and adding the  $PVII(20)$  distri-

	<i>MS</i>	<i>MK</i>	<i>HZ</i>	<i>EN</i>	<i>HM</i>	<i>HJ</i>	$T_{2.5}$	$T_5$	$T_{10}$	$T_\infty$
$N(0, 1)^5$	5	5	5	5	5	5	5	5	5	5
NMIX1	82	33	74	94	43	58	34	51	68	86
NMIX2	94	94	68	89	95	95	95	96	95	94
$t_5(0, I_5)$	88	94	72	88	89	90	86	89	90	89
$t_{10}(0, I_5)$	54	58	23	45	51	55	51	55	57	55
$\chi^2(15)^5$	51	22	30	52	26	39	29	36	44	56
$\chi^2(20)^5$	39	16	22	39	20	30	23	28	33	42
Logistic(0, 1) <sup>5</sup>	33	34	13	25	31	34	31	34	36	33
Gamma(5, 1) <sup>5</sup>	72	33	49	74	37	55	40	51	63	76
Gamma(4, 2) <sup>5</sup>	81	40	60	84	40	64	47	59	72	85
$P_{VII}(10)^5$	27	25	9	19	26	28	26	28	29	27
$P_{VII}(20)^5$	12	9	6	9	12	12	11	12	12	12
$N(0, 1)^4 \otimes t(3)$	35	32	16	30	42	39	38	39	39	35
$N(0, 1)^4 \otimes \chi^2(5)$	28	13	16	28	19	23	19	22	25	31
$N(0, 1)^4 \otimes \chi^2(10)$	16	8	10	15	13	14	12	13	15	18
$S^5(LN(0, \frac{1}{2}))$	89	95	77	90	90	90	86	90	91	89
$NM_5(\rho = 0.2)$	12	9	5	11	12	13	12	13	13	12

**Table 5:** Monte Carlo power estimates in the multivariate case for  $d = 5$

bution. Finally, none of the competing tests is able to outperform the newly proposed class of tests against the  $NM_d(\rho = 0.2)$  distribution.

In most of the cases considered, the power of the newly proposed class of tests is a monotone function of  $\gamma$ . Based on the numerical results presented, it is recommended that  $\gamma = 5$  be used when performing the test as this value results in reasonably high power against the majority of the alternative distributions considered.

## 6 A real data example

The payoff function of certain types of financial derivatives depends on the joint behaviour of multiple stocks or indexes; an important example is the class of basket options. When calculating the price of a basket option, it is often assumed that the log-returns of the stocks or indexes considered are realized from a multivariate normal distribution (this assumption is an extension of the celebrated Black-Merton-Scholes model for options on a single stock or index). As a result, testing the hypothesis that observed financial log-returns follow a multivariate normal law is of interest when pricing basket options. For more details regarding the pricing of these options, the interested reader is referred to Caldana et al. (2016).

As a practical application, we consider the log-returns associated with three major indexes traded in the financial market of the United States. 50 daily log-returns were calculated for the period ending 29 December 2017, the relevant prices were downloaded from <http://finance.yahoo.com>. The first index considered is the Dow Jones Industrial Average (DJIA) index; this index is comprised of a price-weighted average of 30 large publicly owned companies. The second is the Standard & Poor 500 (S&P 500); a market-capitalization

weighted index comprising 500 large companies. Finally, we consider the log-returns of the National Association of Securities Dealers Automated Quotations (NASDAQ) composite index. We are interested in testing the hypothesis that the log-returns are realized from a multivariate normal distribution using the newly proposed test.

Table 6 shows the estimated p-values associated with the newly proposed tests for various values of  $\gamma$ , the reported p-values were obtained using one million Monte Carlo simulations in each case. The results indicate that the hypothesis of multivariate normality is rejected at a 1% significance level for each value of  $\gamma$  considered.

$\gamma = 2.5$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 7$	$\gamma = 10$	$\gamma = \infty$
0.0072	0.0061	0.0044	0.0033	0.00210	0.0013	0.0002

**Table 6:** p-values associated with the newly proposed tests

## 7 Conclusion

We proposed and studied a new class of affine invariant tests for normality in any dimension that are based on a partial differential equation involving the moment generating function. Some properties of the limit null distribution of the test statistic  $T_{n,\gamma}$  have been derived, and the consistency of this class of tests against general alternatives has been proved under some mild conditions. For fixed  $n$ , the test statistic  $T_{n,\gamma}$ , after suitable scaling, approaches a linear combination of two measures of multivariate skewness as  $\gamma \rightarrow \infty$ .

A Monte Carlo study investigates the finite-sample performance of  $T_{n,\gamma}$  compared to those of competing tests in the univariate and multivariate settings. The competing tests considered for univariate normality comprise four well-known tests, while, in the multivariate case, we include four prominent classical tests for multinormality and two very recent tests. The numerical results indicate that  $T_{n,\gamma}$  often exhibits power greater than those associated with several of its competitors, both in univariate and multivariate settings. Based on the numerical results obtained, it is recommended that  $\gamma = 5$  is used when performing the test.

## 8 Appendix: Technical proofs

**Proposition 1** We have  $\|C_n\|_{\oplus} = o_{\mathbb{P}}(1)$ , where  $C_n$  is given in (20).

PROOF. Putting  $\xi_j(t) := \exp(t^\top X_j (t^\top \Delta_{n,j})^2 \exp(\Theta_{n,j} t^\top \Delta_{n,j}) / 2$ , we have

$$\|C_n(t)\|^2 = \frac{1}{n} \sum_{j,k=1}^n \xi_j(t) \xi_k(t) (X_j - t + \Delta_{n,j})^\top (X_k - t + \Delta_{n,k}).$$

Recall  $F_n$  from (28) and put  $A_n := \|S_n^{-1/2} - \mathbf{I}_d\|_2 F_n + \|S_n^{-1/2}\|_2 \|\overline{X}_n\|$ . We have  $|\Theta_{n,j} t^\top \Delta_{n,j}| \leq \|t\| A_n$ . Furthermore, using  $|t^\top (X_j + X_k)| \leq 2 \|t\| F_n$  and

$$(t^\top \Delta_{n,j})^2 \leq 2 \|t\|^2 \|S_n^{-1/2} - \mathbf{I}_d\|_2^2 \|X_j\|^2 + 2 \|t\|^2 \|S_n^{-1/2}\|_2^2 \|\overline{X}_n\|^2,$$

as well as

$$|(X_j - t + \Delta_{n,j})^\top (X_k - t + \Delta_{n,k})| \leq (\|X_j\| + \|t\| + \|\Delta_{n,j}\|) (\|X_k\| + \|t\| + \|\Delta_{n,k}\|),$$

we obtain

$$\begin{aligned} \|C_n(t)\|^2 &\leq e^{2\|t\|\Gamma_n} \frac{1}{n} \sum_{j,k=1}^n \left( \left\{ \|t\|^2 \|S_n^{-1/2} - \mathbf{I}_d\|_2^2 \|X_j\|^2 + \|t\|^2 \|S_n^{-1/2}\|_2^2 \|\overline{X}_n\|^2 \right\} \right. \\ &\quad \times \left\{ \|t\|^2 \|S_n^{-1/2} - \mathbf{I}_d\|_2^2 \|X_k\|^2 + \|t\|^2 \|S_n^{-1/2}\|_2^2 \|\overline{X}_n\|^2 \right\} \\ &\quad \left. \times \left\{ \|X_j\| + \|t\| + \|\Delta_{n,j}\| \right\} \left\{ \|X_k\| + \|t\| + \|\Delta_{n,k}\| \right\} \right), \end{aligned}$$

where  $\Gamma_n = F_n + A_n$ . Expanding the curly brackets, the leading terms are those that do not involve any of  $\Delta_{n,j}$ . We concentrate on

$$S_n(t) := e^{2\|t\|\Gamma_n} \frac{1}{n} \sum_{j,k=1}^n \left( \|t\|^4 \|S_n^{-1/2} - \mathbf{I}_d\|_2^4 \|X_j\|^3 \|X_k\|^3 \right),$$

which originates from choosing the first term within each of the curly brackets. The other terms are treated similarly. Notice that

$$\int_{\mathbb{R}^d} S_n(t) w_\gamma(t) dt = \|S_n^{-1/2} - \mathbf{I}_d\|_2^4 \cdot n \cdot \left( \frac{1}{n} \sum_{j=1}^n \|X_j\|^3 \right)^2 \cdot \int_{\mathbb{R}^d} \|t\|^4 e^{2\|t\|\Gamma_n} w_\gamma(t) dt,$$

and that  $\|S_n^{-1/2} - \mathbf{I}_d\|_2^2 = O_{\mathbb{P}}(n^{-2})$ ,  $n^{-1} \sum_{j=1}^n \|X_j\| = O_{\mathbb{P}}(1)$ . By Proposition 10.2 of Henze et al. (2018), the integral is of order  $O_{\mathbb{P}}(\Gamma_n^{d+3}) \exp(\Gamma_n^2/\gamma)$  (notice that  $1 + \gamma$  in that paper corresponds to (our)  $\gamma$ ). From display (10.6) and display (10.7) of Henze et al. (2018) we have  $\Gamma_n = O_{\mathbb{P}}(\sqrt{\log n})$  and  $\exp(\Gamma_n^2/\gamma) = n^{2/\gamma} \cdot (\log n)^{(d-2)/\gamma} O_{\mathbb{P}}(1)$ . Since  $\gamma > 2$ , it follows that

$$\int_{\mathbb{R}^d} S_n(t) w_\gamma(t) dt = O_{\mathbb{P}}\left(n^{\frac{2}{\gamma}-1}\right) O_{\mathbb{P}}\left((\log n)^{\frac{d+3}{2} + \frac{d-2}{\gamma}}\right) = o_{\mathbb{P}}(1). \quad \square$$

**Proposition 2** For  $B_n$  given in (19), we have

$$B_n(t) = -\frac{1}{\sqrt{n}} \exp\left(\frac{\|t\|^2}{2}\right) \sum_{j=1}^n \left( \frac{X_j X_j^\top t - t}{2} + X_j \right) + o_{\mathbb{P}}(1).$$

PROOF. Observe that  $B_n(t) = n^{-1/2} \sum_{j=1}^n \exp(t^\top X_j) \Delta_{n,j} + R_n(t)$ , where

$$R_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} (t^\top \Delta_{n,j}) \Delta_{n,j}.$$

Use

$$\|R_n(t)\|^2 \leq \left( \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} \right)^2 n \|t\|^2 \max_{i=1, \dots, n} \|\Delta_{n,i}\|^4$$

and  $\|\Delta_{n,j}\| \leq \|S_n^{-1/2} - \text{Id}\|_2 \|F_n\| + \|S_n^{-1/2}\| \|\bar{X}_n\|$  with  $F_n$  given in (28) together with  $F_n = O_{\mathbb{P}}(\sqrt{\log n})$  (see Prop. 10.1. of Henze et al. (2018)) and

$$\mathbb{E} \left[ \left( \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} \right)^2 \right] = \mathbb{E} \left[ \frac{1}{n^2} \sum_{j,k=1}^n e^{t^\top (X_j + X_k)} \right] \leq e^{2\|t\|^2} + e^{\|t\|^2} \quad (34)$$

to show  $R_n = o_{\mathbb{P}}(1)$ . Next,  $B_n(t) - R_n(t) = B_{n,1}(t) - B_{n,2}(t) - B_{n,3}(t)$ , where

$$B_{n,1}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} \left( S_n^{-1/2} - \text{Id} \right) X_j,$$

$$B_{n,2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} \left( S_n^{-1/2} - \text{Id} \right) \bar{X}_n, \quad B_{n,3}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} \bar{X}_n.$$

Since

$$\|B_{n,2}(t)\|^2 \leq \left( \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} \right)^2 \|\sqrt{n} (S_n^{-1/2} - \text{Id})\|_2^2 \cdot \|\bar{X}_n\|^2$$

and  $\|\sqrt{n} (S_n^{-1/2} - \text{Id})\|_2^2 \cdot \|\bar{X}_n\|^2 = O_{\mathbb{P}}(n^{-1})$ , one may use (34) and Fubini's theorem to show  $B_{n,2} = o_{\mathbb{P}}(1)$ . To proceed, rewrite  $B_{n,3}(t)$  in the form

$$B_{n,3}(t) = \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} \cdot \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j.$$

Since replacing the first factor with its expectation  $\exp(\|t\|^2/2)$  means adding a term that is asymptotically negligible, we have

$$B_{n,3}(t) = e^{\|t\|^2/2} \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j + o_{\mathbb{P}}(1). \quad (35)$$

To tackle  $B_{n,1}(t)$ , we rewrite its transpose in the form

$$B_{n,1}(t)^\top = \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} X_j^\top \sqrt{n} \left( S_n^{-1/2} - \text{Id} \right)$$

and use display (2.13) of Henze and Wagner (1997), according to which

$$\sqrt{n} \left( S_n^{-1/2} - I_d \right) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^\top - I_d) + D_n, \quad (36)$$

where  $D_n = O_{\mathbb{P}}(n^{-1/2})$ . Putting

$$E(t) = \mathbb{E} \left[ e^{t^\top X_1} X_1 \right] = \frac{d}{dt} e^{\|t\|^2/2} = e^{\|t\|^2/2} \cdot t \quad (37)$$

and  $Y_n(t) = n^{-1} \sum_{j=1}^n (\exp(t^\top X_j) X_j - E(t))$ , we have

$$B_{n,1}(t)^\top = (Y_n(t) + E(t))^\top \left\{ -\frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^\top - I_d) + D_n \right\}. \quad (38)$$

Abbreviating the matrix within curly brackets by  $S_n$ , we have  $\|S_n Y_n(t)\|^2 \leq \|S_n\|_2^2 \|Y_n(t)\|^2$ . Since  $\|S_n\|_2^2 = O_{\mathbb{P}}(1)$ , it follows that

$$\int_{\mathbb{R}^d} \|S_n Y_n(t)\|^2 w_\gamma(t) dt \leq O_{\mathbb{P}}(1) \int_{\mathbb{R}^d} \|Y_n(t)\|^2 w_\gamma(t) dt.$$

Now, observe that  $Y_n(t)$  is a mean of centred random vectors, and invoking Fubini's theorem the expectation of the integral is seen to be of order  $O(n^{-1})$ . Thus,  $\|S_n Y_n\|_{\oplus} = o_{\mathbb{P}}(1)$ , and hence (since the matrix  $D_n$  figuring in (38) is asymptotically negligible) we have

$$B_{n,1}(t) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^n (X_j X_j^\top - I_d) E(t) + o_{\mathbb{P}}(1).$$

Upon combing this result with (35) and (37), the assertion follows.  $\square$

**Proposition 3** For  $A_n$  given in (18), we have

$$A_n(t) = -\frac{1}{\sqrt{n}} \frac{e^{\|t\|^2/2}}{2} \sum_{j=1}^n (X_j X_j^\top t - t) + o_{\mathbb{P}}(1).$$

PROOF. Notice that  $A_n(t) = A_{n,1}(t) - A_{n,2}(t) - A_{n,3}(t)$ , where

$$A_{n,1}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} t^\top \left( S_n^{-1/2} - I_d \right) X_j (X_j - t),$$

$$A_{n,2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} t^\top \left( S_n^{-1/2} - I_d \right) \bar{X}_n (X_j - t),$$

$$A_{n,3}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e^{t^\top X_j} t^\top \bar{X}_n (X_j - t).$$



To show  $A_{n,2} = o_{\mathbb{P}}(1)$ , use  $n\|S_n^{-1/2} - \text{Id}\|_2^2 \|\bar{X}_n\|^2 = O_{\mathbb{P}}(n^{-1})$  together with

$$\|A_{n,2}(t)\|^2 \leq \left\| \frac{1}{n} \sum_{j=1}^n e^{t^\top X_j} (X_j - t) \right\|^2 \cdot \|t\|^2 \cdot n\|S_n^{-1/2} - \text{Id}\|_2^2 \cdot \|\bar{X}_n\|^2$$

and Fubini's theorem, since  $\mathbb{E}\|n^{-1} \sum_{j=1}^n \exp(t^\top X_j)(X_j - t)\|^2 = O(n^{-1})$ , due to the fact that the summands are centred random vectors. Likewise,

$$\|A_{n,3}(t)\|^2 = \frac{1}{n^3} \sum_{i,j,k,\ell=1}^n e^{t^\top X_i} (X_i - t)^\top \cdot e^{t^\top X_j} (X_j - t) \cdot t^\top X_k \cdot t^\top X_\ell.$$

Since each of the summands is a product of centred random vectors or random variables, we have  $\mathbb{E}\|A_{n,3}(t)\|^2 = O(1/n)$ , and Fubini's theorem yields  $A_{n,3} = o_{\mathbb{P}}(1)$ . To conclude the proof, observe that, by (36),

$$A_{n,1}(t) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^n \Delta_n(t) V_j(t) + \Delta_n(t) D_n t, \quad (39)$$

where  $\Delta_n(t) = n^{-1} \sum_{i=1}^n \exp(t^\top X_i)(X_i - t)X_i^\top$ ,  $V_j(t) = X_j X_j^\top t - t$ . Notice that  $\Delta_n(t)$  is a mean of i.i.d. random matrices, and that

$$\mathbb{E} \left[ e^{t^\top X_1} (X_1 - t) X_1^\top \right] = e^{\|t\|^2} \text{Id}.$$

Straightforward calculations show that replacing  $\Delta_n(t)$  with the right-hand side of the last equation means adding a term that is asymptotically negligible. Hence, the first term on the right-hand side of (39) is

$$-\frac{1}{\sqrt{n}} \frac{e^{\|t\|^2/2}}{2} \sum_{j=1}^n (X_j X_j^\top t - t) + o_{\mathbb{P}}(1).$$

The second term is  $o_{\mathbb{P}}(1)$  since  $D_n = O_{\mathbb{P}}(n^{-1/2})$ .  $\square$

**Proposition 4** (Calculation of  $K(s, t)$ )

Recall  $\tilde{Z}_1(t)$  from (21). Putting  $m(s) = \exp(\|s\|^2/2)$  and  $X_1 = X$ , we have

$$\begin{aligned} \tilde{Z}_1(s) \tilde{Z}_1(t)^\top &= e^{(s+t)^\top X} (X - s)(X - t)^\top - m(s) e^{t^\top X} (X X^\top s - s)(X - t)^\top \\ &\quad - m(s) e^{t^\top X} X (X - t)^\top - m(t) e^{s^\top X} (X - s)(t^\top X X^\top - t^\top) \\ &\quad + m(s) m(t) (X X^\top s - s)(t^\top X X^\top - t^\top) \\ &\quad + m(s) m(t) X (t^\top X X^\top - t^\top) - m(t) e^{s^\top X} (X - s) X^\top \\ &\quad + m(s) m(t) (X X^\top s - s) X^\top + m(s) m(t) X X^\top. \end{aligned}$$

Writing  $\mathbf{0}$  for the zero matrix of order  $d$ , we have

$$\begin{aligned}
\mathbb{E} \left[ e^{(s+t)^\top X} (X-s)(X-t)^\top \right] &= e^{\|s+t\|^2/2} (\mathbf{I}_d + ts^\top), \\
\mathbb{E} \left[ e^{t^\top X} (XX^\top s - s)(X-t)^\top \right] &= e^{\|t\|^2/2} (ts^\top + s^\top t \mathbf{I}_d), \\
\mathbb{E} \left[ e^{t^\top X} X(X-t)^\top \right] &= e^{\|t\|^2/2} \mathbf{I}_d, \\
\mathbb{E} \left[ e^{s^\top X} (X-s)(t^\top XX^\top - t^\top) \right] &= e^{\|s\|^2/2} (ts^\top + s^\top t \mathbf{I}_d), \\
\mathbb{E} \left[ (XX^\top s - s)(t^\top XX^\top - t^\top) \right] &= ts^\top + s^\top t \mathbf{I}_d, \\
\mathbb{E} \left[ X(t^\top XX^\top - t^\top) \right] &= \mathbf{0}, \\
\mathbb{E} \left[ e^{s^\top X} (X-s)X^\top \right] &= e^{\|s\|^2/2} \mathbf{I}_d, \\
\mathbb{E} \left[ (XX^\top s - s)X^\top \right] &= \mathbf{0}, \\
\mathbb{E} \left[ XX^\top \right] &= \mathbf{I}_d.
\end{aligned}$$

The assertion now follows from straightforward calculations.

**Proposition 5** Let  $(A_n)$  be a sequence of symmetric positive definite  $(d \times d)$ -matrices and  $(b_n)$  an increasing sequence of positive real numbers satisfying  $\lim_{n \rightarrow \infty} b_n = \infty$  so that

$$\lim_{n \rightarrow \infty} b_n \|A_n - \mathbf{I}_d\|_2 = 0.$$

We then have  $\lim_{n \rightarrow \infty} b_n \|A_n^{-1/2} - \mathbf{I}_d\|_2 = 0$ .

PROOF. Let  $A_n = \text{diag}(\lambda_1, \dots, \lambda_n)$  be the diagonal matrix consisting of the positive eigenvalues of  $A_n$  so that  $\|A_n\|_2 = \max_{i=1, \dots, n} \lambda_i$  and

$$\|(A_n^{1/2} + \mathbf{I}_d)^{-1}\|_2 = \max_{i=1, \dots, n} (\lambda_i^{1/2} + 1)^{-1} < 1. \quad (40)$$

Since the assumptions imply  $\|A_n - \mathbf{I}_d\|_2 \rightarrow 0$ , choose  $n_0$  so large that  $\|A_n - \mathbf{I}_d\|_2 \leq 1/2$  for each  $n \geq n_0$ . Putting  $T_n = \mathbf{I}_d - A_n$ , we have

$$\|A_n^{-1}\|_2 = \|(\mathbf{I}_d - T_n)^{-1}\|_2 \leq \frac{1}{1 - \|T_n\|_2} \leq \frac{1}{1 - 1/2} = 2, \quad n \geq n_0,$$

and thus  $\|A_n^{-1/2}\|_2 \leq \sqrt{2}$ ,  $n \geq n_0$ . Now,  $A_n - \mathbf{I}_d = (A_n^{1/2} + \mathbf{I}_d)A_n^{1/2}(\mathbf{I}_d - A_n^{-1/2})$  implies  $\mathbf{I}_d - A_n^{-1/2} = A_n^{-1/2}(A_n^{1/2} + \mathbf{I}_d)^{-1}(A_n - \mathbf{I}_d)$ , whence

$$\|A_n^{-1/2} - \mathbf{I}_d\|_2 = \|\mathbf{I}_d - A_n^{-1/2}\|_2 \leq \|A_n^{-1/2}\|_2 \cdot \|(A_n^{1/2} + \mathbf{I}_d)^{-1}\|_2 \cdot \|A_n - \mathbf{I}_d\|_2.$$

In view of (40) and  $\|A_n^{-1/2}\|_2 \leq \sqrt{2}$ ,  $n \geq n_0$ , the assertion follows.  $\square$

**Acknowledgements** The authors thank Tobias Jahnke for providing the proof of Proposition 5 and the referees for helpful remarks.

## References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178
- Azzalini, A. (2017). The R package 'sn': The skew-normal and skew-t distributions. R package version 1.5-0. <http://azzalini.stat.unipd.it/SN>
- Barndorff-Nielsen, O. (1963). On the behaviour of extreme order statistics. *Annals of Mathematical Statistics*, 34, 992–1002
- Baringhaus, L., and Henze, N. (1991). Limit distributions for measures of multivariate skewness and kurtosis based on projections. *Journal of Multivariate Analysis*, 38, 51–69
- Baringhaus, L., and Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. *Annals of Statistics*, 20, 1889–1902
- Becker, M, and Klößner, S. (2017). PearsonDS: Pearson Distribution System. R package version 1.0. <https://CRAN.R-project.org/package=PearsonDS>
- Bowman, A.W., and Foster, P.J. (1993). Adaptive smoothing and density based tests of multivariate normality. *Journal of the American Statistical Association*, 88, 529–537
- Caldana, R., Fusai, G., Gnoatto, A., and Grasselli, M. (2016). General closed-form basket option pricing bounds. *Quantitative Finance*, 16, 535–554
- Eaton, M.L., and Perlman, M.D. (1973). The non-singularity of generalized sample covariance matrices. *Annals of Statistics*, 1, 710–717
- Farrell, P.J., Salibian-Barrera, M., and Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, 75, 93–107
- Fletcher, T.D. (2012). QuantPsyc: Quantitative Psychology Tools. R package version 1.5, <https://CRAN.R-project.org/package=QuantPsyc>.
- Gross, J., and Ligges, U. (2015). norrest: Tests for normality. R package version 1.0-4. <https://CRAN.R-project.org/package=norrest>
- Henze, N. (1994a). On Mardia's kurtosis test for multivariate normality. *Communications in Statistics – Theory and Methods*, 23, 1031–1045
- Henze, N. (1994b). The asymptotic behavior of a variant of multivariate kurtosis. *Communications in Statistics – Theory and Methods*, 23, 1047–1061
- Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statistics & Probability Letters*, 35, 203–213
- Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers*, 43, 467–506
- Henze, N., and Jiménez-Gamero, M.D. (2018). A new class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. *TEST* <https://doi.org/10.1007/s11749-018-0589-z>
- Henze, N., and Koch, S. (2017). On a test of normality based on the empirical moment generating function. *Statistical Papers*, doi:10.1007/s00362-017-0923-7.
- Henze, N., and Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, 62, 1–23
- Henze, N., and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics – Theory and Methods*, 19, 3595–3617
- Henze, N., Jiménez-Gamero, M.D., and Meintanis, S.G. (2018). A novel characterization of multinormality and corresponding tests of fit, including for Garch models. *Econometric Theory* <https://doi.org/10.1017/S0266466618000154>
- Joensuu, D.W., and Vogel, J. (2014). A power study of goodness-of-fit tests for multivariate normality implemented in R. *Journal of Statistical Computation and Simulation*, 84, 1055–1078
- Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, New York.
- Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: An R Package for Assessing Multivariate Normality. *The R Journal*, 6(2), 151–162
- Kundu, D., Majumdar, S., and Mukherjee, K. (2000). Central Limit theorems revisited. *Statistics & Probability Letters*, 47, 265–275
- Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530

- Mecklin, C.J., and Mundfrom, D.J. (2005). A Monte Carlo comparison of Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, 75, 93–107
- Móri, T.F., Rohatgi, V.K., and Székely, G.J. (1993). On multivariate skewness and kurtosis. *Theory of Probability and its Applications*, 38, 547–551
- R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rizzo, M.L., and Székely, G.J. (2016). E-Statistics: Multivariate Inference via the Energy of Data. R package version 1.7-0. <https://CRAN.R-project.org/package=energy>.
- Ruckdeschel, P., Kohl, M., Stabla, T., and Camphausen, F. (2006). S4 Classes for Distributions. *Journal of Statistical Computation and Simulation*, 35, 1–27
- Székely, G.J. and Rizzo, M.L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93, 58–80
- Trapletti, A., and Hornik, K. (2017). tseries: Time series analysis and computational finance. R package version 0.10-40. <https://CRAN.R-project.org/package=tseries>
- Volkmer, H. (2014). A characterization of the normal distribution. *Journal of Statistical Theory and Applications*, 13, 83–85
- Zghoul, A.A. (2010). A goodness of fit test for normality based on the empirical moment generating function. *Communications in Statistics – Simulation and Computation*, 39, 1929–1304