

Dörr, Philip; Ebner, Bruno; Henze, Norbert

Article — Published Version

## A new test of multivariate normality by a double estimation in a characterizing PDE

Metrika

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Dörr, Philip; Ebner, Bruno; Henze, Norbert (2020) : A new test of multivariate normality by a double estimation in a characterizing PDE, Metrika, ISSN 1435-926X, Springer, Berlin, Heidelberg, Vol. 84, Iss. 3, pp. 401-427, <https://doi.org/10.1007/s00184-020-00795-x>

This Version is available at:

<https://hdl.handle.net/10419/288495>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# A new test of multivariate normality by a double estimation in a characterizing PDE

Philip Dörr<sup>1</sup> · Bruno Ebner<sup>2</sup> · Norbert Henze<sup>2</sup>

Received: 27 November 2019 / Published online: 31 August 2020  
© The Author(s) 2020

## Abstract

This paper deals with testing for nondegenerate normality of a  $d$ -variate random vector  $X$  based on a random sample  $X_1, \dots, X_n$  of  $X$ . The rationale of the test is that the characteristic function  $\psi(t) = \exp(-\|t\|^2/2)$  of the standard normal distribution in  $\mathbb{R}^d$  is the only solution of the partial differential equation  $\Delta f(t) = (\|t\|^2 - d)f(t)$ ,  $t \in \mathbb{R}^d$ , subject to the condition  $f(0) = 1$ , where  $\Delta$  denotes the Laplace operator. In contrast to a recent approach that bases a test for multivariate normality on the difference  $\Delta\psi_n(t) - (\|t\|^2 - d)\psi(t)$ , where  $\psi_n(t)$  is the empirical characteristic function of suitably scaled residuals of  $X_1, \dots, X_n$ , we consider a weighted  $L^2$ -statistic that employs  $\Delta\psi_n(t) - (\|t\|^2 - d)\psi_n(t)$ . We derive asymptotic properties of the test under the null hypothesis and alternatives. The test is affine invariant and consistent against general alternatives, and it exhibits high power when compared with prominent competitors. The main difference between the procedures are theoretically driven by different covariance kernels of the Gaussian limiting processes, which has considerable effect on robustness with respect to the choice of the tuning parameter in the weight function.

**Keywords** Test for multivariate normality · Affine invariance · Weighted  $L^2$ -statistic · Consistency · Laplace operator · Harmonic oscillator

## 1 Introduction

A useful tool for assessing the fit of data to a family of distributions are empirical counterparts of distributional characterizations. Such characterizations often emerge

---

✉ Bruno Ebner  
Bruno.Ebner@kit.edu

Norbert Henze  
Norbert.Henze@kit.edu

<sup>1</sup> Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik,  
Universitätsplatz 2, 39106 Magdeburg, Germany

<sup>2</sup> Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, 76128 Karlsruhe,  
Germany

as solutions of an equation of the type  $\rho(Df, f) = 0$ . Here,  $\rho(\cdot, \cdot)$  is some distance on a suitable function space, while  $f$  may be either the moment generating function, the Laplace transform, or the characteristic function. Moreover,  $D$  denotes a differential operator, i.e., this operator can be regarded as ordinary differentiation if  $f$  is a function of only one variable or, for instance, the Laplace operator in the multivariate case. Such (partial) differential equations have been used to test for multivariate normality, see Dörr et al. (2020), Henze and Visagie (2020), exponentiality, see Baringhaus and Henze (1991a), the gamma distribution, see Henze et al. (2012), the inverse Gaussian distribution, see Henze and Klar (2002), the beta distribution, see Riad and Mabood (2018), the univariate and multivariate skew-normal distribution, see Meintanis (2010) and Meintanis and Hlávka (2010), and the Rayleigh distribution, see Meintanis and Iliopoulos (2003). In all these references, the authors propose a goodness-of-fit test by plugging in an empirical counterpart  $f_n$  for  $f$  into  $\rho(Df, f)$ , and by measuring the deviation from the zero function in a suitable function space. If, under the hypothesis to be tested, the function  $f$  has a closed form and is known, there are two options for obtaining an empirical counterpart to the characterizing equation, namely  $\rho(Df_n, f) = 0$ , or  $\rho(Df_n, f_n) = 0$ . To the best of our knowledge, the effect of considering both options for the *same testing problem* and to study the consequences on the performance of the resulting test statistics has not yet been considered, neither from a theoretical point of view, nor in a simulation study. In this spirit, the purpose of this paper is to investigate the effect on the power of a recent test for multivariate normality based on a characterization of the multivariate normal law in connection with the harmonic oscillator, see Dörr et al. (2020).

In what follows, let  $d \geq 1$  be a fixed integer, and let  $X, X_1, \dots, X_n, \dots$  be independent and identically distributed (i.i.d.)  $d$ -dimensional random (column) vectors, that are defined on a common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . We write  $\mathbb{P}^X$  for the distribution of  $X$ , and we denote the  $d$ -variate normal law with expectation  $\mu$  and nonsingular covariance matrix  $\Sigma$  by  $N_d(\mu, \Sigma)$ . Moreover,  $\mathcal{N}_d = \{N_d(\mu, \Sigma) : \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d} \text{ positive definite}\}$  stands for the class of all nondegenerate  $d$ -variate normal distributions. To check the assumption of multivariate normality means to test the hypothesis

$$H_0 : \mathbb{P}^X \in \mathcal{N}_d, \quad (1)$$

against general alternatives. The starting point of this paper is Theorem 1 of Dörr et al. (2020). To state this result, let  $\Delta$  denote the Laplace operator,  $\|\cdot\|$  the Euclidean norm in  $\mathbb{R}^d$ , and  $I_d$  the identity matrix of size  $d$ . Then Theorem 1 of Dörr et al. (2020) states that the characteristic function  $\psi(t) = \exp(-\|t\|^2/2)$ ,  $t \in \mathbb{R}^d$ , of the  $d$ -variate standard normal distribution  $N_d(0, I_d)$  is the unique solution of the partial differential equation

$$\begin{cases} \Delta f(x) - (\|x\|^2 - d)f(x) = 0, & x \in \mathbb{R}^d, \\ f(0) = 1. \end{cases} \quad (2)$$

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  $X_1, \dots, X_n$ , respectively, where

the superscript  $\top$  means transposition, the standing tacit assumptions that  $\mathbb{P}^X$  is absolutely continuous with respect to Lebesgue measure and  $n \geq d + 1$  guarantee that  $S_n$  is invertible almost surely, see Eaton and Perlman (1973). The test statistic is based on the so-called *scaled residuals*

$$Y_{n,j} = S_n^{-1/2}(X_j - \bar{X}_n), \quad j = 1, \dots, n.$$

Here,  $S_n^{-1/2}$  is the unique symmetric positive definite square root of  $S_n^{-1}$ . Letting  $\psi_n(t) = n^{-1} \sum_{j=1}^n \exp(it^\top Y_{n,j}), t \in \mathbb{R}^d$ , denote the empirical characteristic function (ecf) of  $Y_{n,1}, \dots, Y_{n,n}$ , the test statistic proposed in Dörr et al. (2020) is

$$T_{n,a} = n \int_{\mathbb{R}^d} |\Delta\psi_n(t) - \Delta\psi(t)|^2 w_a(t) dt, \tag{3}$$

where

$$w_a(t) = \exp(-a\|t\|^2), \quad t \in \mathbb{R}^d, \tag{4}$$

and  $a > 0$  is a fixed constant. The statistic  $T_{n,a}$  has a nice closed-form expression as a function of  $Y_{n,i}^\top Y_{n,j}, i, j \in \{1, \dots, n\}$  (see display (10)-(12) of Dörr et al. (2020)) and is thus invariant with respect to full-rank affine transformations of  $X_1, \dots, X_n$ . Theorems 2 and 3 of Dörr et al. (2020) show that, elementwise on the underlying probability space, suitably rescaled versions of  $T_{n,a}$  have limits as  $a \rightarrow \infty$  and  $a \rightarrow 0$ , respectively. In the former case, the limit is a measure of multivariate skewness, introduced in Móri et al. (1993), whereas Mardia’s time-honored measure of multivariate kurtosis (see Mardia 1970) shows up as  $a \rightarrow 0$ . As  $n \rightarrow \infty$ , the statistic  $T_{n,a}$  has a nondegenerate limit null distribution (Theorem 7 of Dörr et al. 2020), and a test of (1) that rejects  $H_0$  for large values of  $T_{n,a}$  is able to detect alternatives that approach  $H_0$  at the rate  $n^{-1/2}$ , irrespective of the dimension  $d$  (Corollary 10 of Dörr et al. 2020). Under an alternative distribution satisfying  $\mathbb{E}\|X\|^4 < \infty, n^{-1}T_{n,a}$  converges almost surely to a measure of distance  $\Delta_a$  between  $\mathbb{P}^X$  and the class  $\mathcal{N}_d$  (Theorem 11 of Dörr et al. 2020). As a consequence, the test for multinormality based on  $T_{n,a}$  is consistent against any such alternative. By Theorem 14 of Dörr et al. (2020), the sequence  $\sqrt{n}(n^{-1}T_{n,a} - \Delta_a)$  converges in distribution to a centered normal law. Since the variance of this limit distribution can be estimated consistently from  $X_1, \dots, X_n$  (Theorem 16 of Dörr et al. 2020), we have an asymptotic confidence interval for  $\Delta_a$ .

The novel approach taken in this paper is to replace *both* of the functions  $f$  occurring in (2) by the ecf  $\psi_n$ . Since, under  $H_0, \Delta\psi_n(t)$  and  $(\|t\|^2 - d)\psi_n(t)$  should be close to each other for large  $n$ , it is tempting to see what happens if, instead of  $T_{n,a}$  defined in (3), we base a test of  $H_0$  on the weighted  $L^2$ -statistic

$$U_{n,a} = n \int_{\mathbb{R}^d} \left| \Delta\psi_n(t) - (\|t\|^2 - d)\psi_n(t) \right|^2 w_a(t) dt \tag{5}$$

and reject  $H_0$  for large values of  $U_{n,a}$ .

Since  $\Delta\psi_n(t) = -n^{-1} \sum_{j=1}^n \|Y_{n,j}\|^2 \exp(it^\top Y_{n,j})$ , the relation

$$\begin{aligned} & \int_{\mathbb{R}^d} (\|t\|^2 - d)^2 \cos(t^\top c) \exp(-a\|t\|^2) dt \\ &= \left(\frac{\pi}{a}\right)^{d/2} \frac{1}{16a^4} \exp\left(-\frac{\|c\|^2}{4a}\right) \left(16d^2a^3(a-1) + 4d(d+2)a^2\right. \\ & \quad \left.+ (8da^2 - 4(d+2)a)\|c\|^2 + \|c\|^4\right), \end{aligned} \tag{6}$$

valid for  $c \in \mathbb{R}^d$  and  $a > 0$ , and tedious but straightforward calculations yield the representation

$$\begin{aligned} U_{n,a} &= \left(\frac{\pi}{a}\right)^{d/2} \frac{1}{n} \sum_{j,k=1}^n \exp\left(-\frac{\|Y_{n,j} - Y_{n,k}\|^2}{4a}\right) \left[ \|Y_{n,j}\|^2 \|Y_{n,k}\|^2 \right. \\ & \quad - (\|Y_{n,j}\|^2 + \|Y_{n,k}\|^2) \frac{1}{4a^2} (\|Y_{n,j} - Y_{n,k}\|^2 + 2ad(2a-1)) \\ & \quad + \frac{1}{16a^4} \left(16d^2a^3(a-1) + 4d(d+2)a^2 + \|Y_{n,j} - Y_{n,k}\|^4\right. \\ & \quad \left. \left. + (8da^2 - 4(d+2)a)\|Y_{n,j} - Y_{n,k}\|^2\right) \right], \end{aligned} \tag{7}$$

which is amenable to computational purposes. Moreover,  $U_{n,a}$  turns out to be affine invariant.

The rest of the paper is organized as follows. In Sect. 2, we derive the elementwise limits of  $U_{n,a}$ , after suitable transformations, as  $a \rightarrow 0$  and  $a \rightarrow \infty$ . Section 3 deals with the limit null distribution of  $U_{n,a}$  as  $n \rightarrow \infty$ . In Sect. 4, we show that, under the condition  $\mathbb{E}\|X\|^4 < \infty$ ,  $n^{-1}U_{n,a}$  has an almost sure limit as  $n \rightarrow \infty$  under a fixed alternative to normality. As a consequence, the test based on  $U_{n,a}$  is consistent against any such alternative. Moreover, we prove that the asymptotic distribution of  $U_{n,a}$ , after a suitable transformation, is a centered normal distribution. In Sect. 5, we present the results of a simulation study that compares the power of the test for normality based on  $U_{n,a}$  with that of prominent competitors. Section 6 shows a real data example, and Sect. 7 contains some conclusions and gives an outlook on potential further work.

## 2 The limits $a \rightarrow 0$ and $a \rightarrow \infty$

This section considers the (elementwise) limits of  $U_{n,a}$  as  $a \rightarrow 0$  and  $a \rightarrow \infty$ . The results shed some light on the role of the parameter  $a$  that figures in the weight function  $w_a$  in (4). Notice that, from the definition of  $U_{n,a}$  given in (5), we have  $\lim_{a \rightarrow \infty} U_{n,a} = 0$  and  $\lim_{a \rightarrow 0} U_{n,a} = \infty$ , since  $\int |\Delta\psi_n(t) - (\|t\|^2 - d)\psi_n(t)|^2 dt = \infty$ . Suitable transformations of  $U_{n,a}$ , however, yield well-known limit statistics as  $a \rightarrow 0$  and  $a \rightarrow \infty$ .

**Theorem 1** *Elementwise on the underlying probability space, we have*

$$\lim_{a \rightarrow 0} \left[ \left( \frac{a}{\pi} \right)^{d/2} U_{n,a} - \frac{d(d+2)}{4a^2} \right] = \frac{1}{n} \sum_{j=1}^n \|Y_{n,j}\|^4 - d^2. \tag{8}$$

**Proof** Starting with (7),  $(a/\pi)^{d/2}U_{n,a}$  is, apart from the factor  $1/n$ , a double sum over  $j$  and  $k$ . Since each summand for which  $j \neq k$  vanishes asymptotically as  $a \rightarrow 0$ , we have

$$\left( \frac{a}{\pi} \right)^{d/2} U_{n,a} = \frac{1}{n} \sum_{j=1}^n \left[ \|Y_{n,j}\|^4 - \frac{d(2a-1)}{a} \|Y_{n,j}\|^2 + \frac{d^2(a-1)}{a} + \frac{d(d+2)}{4a^2} \right] + o(1)$$

as  $a \rightarrow 0$ , and the result follows from the fact that  $\sum_{j=1}^n \|Y_{n,j}\|^2 = nd$ . □

Theorem 1 means that a suitable affine transformation of  $U_{n,a}$  has a limit as  $a \rightarrow 0$ , and that this limit is — apart from the additive constant  $d^2$  — the time-honored measure of multivariate kurtosis in the sense of Mardia, see Mardia (1970). The same measure — without the subtrahend  $d^2$  — shows up as a limit of  $(a/\pi)^{d/2}T_{n,a}$  as  $a \rightarrow 0$ , see Theorem 3 of Dörr et al. (2020). The next result shows that  $U_{n,a}$  and  $T_{n,a}$ , after multiplication with the same scaling factor, converge to the same limit as  $a \rightarrow \infty$ , cf. Theorem 2 of Dörr et al. (2020).

**Theorem 2** *Elementwise on the underlying probability space, we have*

$$\lim_{a \rightarrow \infty} \frac{2}{n\pi^{d/2}} a^{d/2+1} U_{n,a} = \frac{1}{n^2} \sum_{j,k=1}^n \|Y_{n,j}\|^2 \|Y_{n,k}\|^2 Y_{n,j}^\top Y_{n,k}. \tag{9}$$

**Proof** The proof follows the lines of the Proof of Theorem 2 of Dörr et al. (2020) and is thus omitted. □

The limit figuring on the right hand side of (9) is a measure of multivariate skewness, introduced by Móri et al. (1993). Theorems 1 and 2 show that the class of tests for  $H_0$  are in a certain sense “closed at the boundaries”  $a \rightarrow 0$  and  $a \rightarrow \infty$ . However, in contrast to the test for multivariate normality based on  $U_{n,a}$  for fixed  $a \in (0, \infty)$ , tests for  $H_0$  based on measures of multivariate skewness and kurtosis lack consistency against general alternatives, see, e.g., (Baringhaus and Henze 1991b, 1992; Henze 1994).

### 3 The limit null distribution of $U_{n,a}$

In this section, we assume that the distribution of  $X$  is some nondegenerate  $d$ -variate normal law. In view of affine invariance, we may further assume that  $\mathbb{E}(X) = 0$  and  $\mathbb{E}(XX^\top) = I_d$ . By symmetry, it is readily seen that  $U_{n,a}$  defined in (5) takes the form

$$U_{n,a} = \int_{\mathbb{R}^d} S_n^2(t) w_a(t) dt, \tag{10}$$

where

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (\|Y_{n,j}\|^2 + \|t\|^2 - d)(\cos(t^\top Y_{n,j}) + \sin(t^\top Y_{n,j})), \quad t \in \mathbb{R}^d. \tag{11}$$

In view of (10), our setting for asymptotics will be the separable Hilbert space  $\mathbb{H}$  of (equivalence classes of) measurable functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  that satisfy  $\int f^2(t)w_a(t) dt < \infty$ . Here and in the sequel, each unspecified integral will be over  $\mathbb{R}^d$ . The scalar product and the norm in  $\mathbb{H}$  are given by  $\langle f, g \rangle_{\mathbb{H}} = \int f(t)g(t)w_a(t) dt$  and  $\|f\|_{\mathbb{H}} = \langle f, f \rangle_{\mathbb{H}}^{1/2}$ , respectively. Notice that, in this notation, (10) takes the form  $U_{n,a} = \|S_n\|_{\mathbb{H}}^2$ , where  $S_n$  is given in (11).

Putting  $\psi(t) = \exp(-\|t\|^2/2)$  as before, and writing  $\xrightarrow{\mathcal{D}}$  for convergence in distribution, the main result of this section is as follows.

**Theorem 3** *If  $X$  has some nondegenerate normal distribution, we have the following:*

a) *There is a centered Gaussian random element  $\mathcal{S}$  of  $\mathbb{H}$  having covariance kernel*

$$K(s, t) = \psi(s - t) \left\{ 2d + \|s\|^2 \|t\|^2 - 2s^\top t \|s - t\|^2 - 4\|s - t\|^2 \right\} + 2\psi(s)\psi(t) \left\{ 2\|s\|^2 + 2\|t\|^2 - d - 2s^\top t - 4(s^\top t)^2 \right\}, \quad s, t \in \mathbb{R}^d,$$

*such that, with  $S_n$  defined in (11),  $S_n \xrightarrow{\mathcal{D}} \mathcal{S}$  as  $n \rightarrow \infty$ .*

b) *We have*

$$U_{n,a} \xrightarrow{\mathcal{D}} \int \mathcal{S}^2(t)w_a(t) dt \quad \text{as } n \rightarrow \infty. \tag{12}$$

**Proof** Since the proof is analogous to the proof of Proposition 5 of Dörr et al. (2020), it will only be sketched. If  $S_n^0(t)$  stands for the modification of  $S_n(t)$  that results if we replace  $Y_{n,j}$  with  $X_j$ , then a Hilbert space central limit theorem holds for  $S_n^0$ , since the summands of  $S_n^0$  are square-integrable centered random elements of  $\mathbb{H}$ . The idea is thus to find a random element  $\tilde{S}_n$  of  $\mathbb{H}$  such that  $\tilde{S}_n \xrightarrow{\mathcal{D}} \mathcal{S}$  and  $\|S_n - \tilde{S}_n\|_{\mathbb{H}} = o_{\mathbb{P}}(1)$ . Putting  $Y_{n,j} = X_j + \Delta_{n,j}$  in (11) and using the fact that  $\cos(t^\top Y_{n,j}) = \cos(t^\top X_j) - \sin(\Theta_j)t^\top \Delta_{n,j}$ ,  $\sin(t^\top Y_{n,j}) = \sin(t^\top X_j) + \cos(\Gamma_j)t^\top \Delta_{n,j}$ , where  $\Theta_j, \Gamma_j$  depend on  $X_1, \dots, X_n$  and  $t$  and satisfy  $|\Theta_j - t^\top X_j| \leq |t^\top \Delta_{n,j}|, |\Gamma_j - t^\top X_j| \leq |t^\top \Delta_{n,j}|$ , some algebra and Proposition 18 of Dörr et al. (2020) show that a choice of  $\tilde{S}_n$  is given by

$$\tilde{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n h(X_j, t),$$

where

$$h(x, t) = (\|x\|^2 + \|t\|^2 - d)(\cos(t^\top x) + \sin(t^\top x))$$

$$-\psi(t) \left\{ 2\|t\|^2 + \|x\|^2 - d + 2t^\top x - 2(t^\top x)^2 \right\}.$$

Tedious calculations then show that the covariance kernel of  $\mathcal{S}$ , which is  $\mathbb{E}[h(X, s)h(X, t)]$ , is equal to  $K(s, t)$  given above.  $\square$

Notice that the covariance kernel figuring in Theorem 3 is much shorter than the corresponding kernel given in Theorem 7 in Dörr et al. (2020) for the related test statistic  $T_{n,a}$  defined in (3). Thus, the double estimation leads to a simpler covariance kernel. Let  $U_{\infty,a}$  denote a random variable having the limit distribution of  $U_{n,a}$  given in (12). Since the distribution of  $U_{\infty,a}$  is that of  $\|\mathcal{S}\|_{\mathbb{H}}^2$ , where  $\mathcal{S}$  is the Gaussian random element of  $\mathbb{H}$  figuring in Theorem 3, it is the distribution of  $\sum_{j \geq 1} \lambda_j N_j^2$ , where  $N_1, N_2, \dots$  is a sequence of i.i.d. standard normal random variables, and  $\lambda_1, \lambda_2, \dots$  are the positive eigenvalues corresponding to normalized eigenfunctions of the integral operator  $f \mapsto Af$  on  $\mathbb{H}$ , where  $(Af)(s) = \int K(s, t)f(t)w_a(t)dt$ . It seems to be hopeless to obtain closed-form expressions of these eigenvalues. However, in view of Fubini’s theorem, we have

$$\mathbb{E}[U_{\infty,a}] = \int \mathbb{E}[S^2(t)]w_a(t)dt = \int K(t, t)w_a(t)dt,$$

and thus straightforward manipulations of integrals yield the following result.

**Theorem 4** *Putting  $\gamma = (a/(a + 1))^{d/2}$ , we have*

$$\mathbb{E}[U_{\infty,a}] = 2d \left( \frac{\pi}{a} \right)^{d/2} \left\{ 1 - \gamma + \frac{\gamma}{a + 1} - \frac{(d + 2)\gamma}{(a + 1)^2} + \frac{d + 2}{8a^2} \right\}.$$

From this result, one readily obtains

$$\lim_{a \rightarrow 0} \left[ \left( \frac{a}{\pi} \right)^{d/2} \mathbb{E}[U_{\infty,a}] - \frac{d(d + 2)}{4a^2} \right] = 2d. \tag{13}$$

It is interesting to compare this limit relation with (8). If the underlying distribution is standard normal, i.e., if  $\mathbb{P}^X = N_d(0, I_d)$ , we have  $\mathbb{E}\|X\|^4 = 2d + d^2$ . Now, writing  $Y_{n,j} = X_j + \Delta_{n,j}$  and using Proposition 18 of Dörr et al. (2020), the right hand side of (8) turns out to converge in probability to  $\mathbb{E}\|X\|^4 - d^2$  as  $n \rightarrow \infty$ , and this expectation is the right hand side of (13). Regarding the case  $a \rightarrow \infty$ , the representation of  $\mathbb{E}[U_{\infty,a}]$  easily yields

$$\lim_{a \rightarrow \infty} \left[ \frac{2a^{d/2+1}}{\pi^{d/2}} \mathbb{E}[U_{\infty,a}] \right] = 2d(d + 2).$$

This result corresponds to (9), since, by Theorem 2.2 of Henze (1997), the right hand side of (9), after multiplication with  $n$ , converges in distribution to  $2(d + 2)\chi_d^2$  as  $n \rightarrow \infty$  if  $\mathbb{P}^X = N_d(0, I_d)$ . Here,  $\chi_d^2$  is a random variable having a chi square distribution with  $d$  degrees of freedom.



### 4 Limits of $U_{n,a}$ under alternatives

In this section we assume that  $X, X_1, X_2, \dots$  are i.i.d., and that  $\mathbb{E}\|X\|^4 < \infty$ . Moreover, let  $\mathbb{E}(X) = 0$  and  $\mathbb{E}(XX^\top) = I_d$  in view of affine invariance, and recall the Laplace operator  $\Delta$  from Sect. 1. The characteristic function of  $X$  will be denoted by  $\psi(t) = \mathbb{E}[\exp(it^\top X)]$ ,  $t \in \mathbb{R}^d$ . Letting

$$\psi^\pm(t) = \mathbb{E}[\cos(t^\top X)] \pm \mathbb{E}[\sin(t^\top X)], \quad t \in \mathbb{R}^d,$$

we first present an almost sure limit for  $n^{-1}U_{n,a}$ , which is the same limit as for  $T_{n,a}$ , see Theorem 11 in Dörr et al. (2020).

**Theorem 5** *We have*

$$\frac{U_{n,a}}{n} \xrightarrow{\text{a.s.}} \Gamma_a := \int_{\mathbb{R}^d} z^2(t)w_a(t) dt = \|z\|_{\mathbb{H}}^2,$$

where

$$z(t) = -\Delta\psi^+(t) + (\|t\|^2 - d)\psi^+(t). \tag{14}$$

**Proof** In what follows, we write  $\text{CS}^\pm(\xi) = \cos(\xi) \pm \sin(\xi)$ , and we put  $Y_j = Y_{n,j}$ ,  $\Delta_j = \Delta_{n,j}$  for the sake of brevity. From (10) and (11), we have  $n^{-1}U_{n,a} = \|V_n + W_n\|_{\mathbb{H}}^2$ , where

$$V_n(t) = \frac{1}{n} \sum_{j=1}^n \|Y_j\|^2 \text{CS}^+(t^\top Y_j), \quad W_n(t) = (\|t\|^2 - d) \frac{1}{n} \sum_{j=1}^n \text{CS}^+(t^\top Y_j).$$

Putting

$$V_n^0(t) = \frac{1}{n} \sum_{j=1}^n \|X_j\|^2 \text{CS}^+(t^\top X_j), \quad W_n^0(t) = (\|t\|^2 - d) \frac{1}{n} \sum_{j=1}^n \text{CS}^+(t^\top X_j),$$

the strong law of large numbers in Hilbert spaces (see, e.g., Theorem 7.7.2 of Hsing and Eubank (2015)) yields  $\|V_n^0 + W_n^0\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} \Gamma_a$  as  $n \rightarrow \infty$ , and thus it suffices to prove  $\|V_n + W_n\|_{\mathbb{H}}^2 - \|V_n^0 + W_n^0\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} 0$ . From

$$\|V_n + W_n\|_{\mathbb{H}}^2 - \|V_n^0 + W_n^0\|_{\mathbb{H}}^2 = \langle V_n - V_n^0 + W_n - W_n^0, V_n + W_n + V_n^0 + W_n^0 \rangle_{\mathbb{H}},$$

the Cauchy–Schwarz inequality, the fact that  $\max(|W_n(t)|, |W_n^0(t)|) \leq 2(d + \|t\|^2)$ ,  $|V_n(t)| \leq 2d$ ,  $|V_n^0(t)| \leq 2n^{-1} \sum_{j=1}^n \|X_j\|^2$  and Minkowski’s inequality, it suffices to prove  $\|V_n - V_n^0\|_{\mathbb{H}} \xrightarrow{\text{a.s.}} 0$  and  $\|W_n - W_n^0\|_{\mathbb{H}} \xrightarrow{\text{a.s.}} 0$  as  $n \rightarrow \infty$ . As for  $W_n - W_n^0$ , the inequalities  $|\cos(t^\top Y_j) - \cos(t^\top X_j)| \leq \|t\| \|\Delta_j\|$ ,  $|\sin(t^\top Y_j) - \sin(t^\top X_j)| \leq \|t\| \|\Delta_j\|$  and the Cauchy–Schwarz inequality yield  $|W_n(t) - W_n^0(t)| \leq (\|t\|^2 + d)2\|t\|(n^{-1} \sum_{j=1}^n \|\Delta_j\|^2)^{1/2}$ . In view of Proposition 18 b) of Dörr et al. (2020), we

have  $\|W_n - W_n^0\|_{\mathbb{H}} \xrightarrow{\text{a.s.}} 0$ . Regarding  $V_n - V_n^0$ , we decompose this difference according to

$$\begin{aligned}
 V_n(t) - V_n^0(t) &= \frac{1}{n} \sum_{j=1}^n (\|Y_j\|^2 - \|X_j\|^2) \text{CS}^+(t^\top Y_j) \\
 &\quad + \frac{1}{n} \sum_{j=1}^n \|X_j\|^2 (\text{CS}^+(t^\top Y_j) - \text{CS}^+(t^\top X_j)).
 \end{aligned}$$

The squared norm in  $\mathbb{H}$  of the second summand on the right hand side converges to zero almost surely, see the treatment of  $U_{n,1}$  in the Proof of Theorem 11 of Dörr et al. (2020). The same holds for the first summand, since its modulus is bounded from above by  $4\|t\|n^{-1} \sum_{j=1}^n \|\Delta_j\| + 2n^{-1} \sum_{j=1}^n \|\Delta_j\|^2$ , and the inequality  $n^{-1} \sum_{j=1}^n \|\Delta_j\| \leq (n^{-1} \sum_{j=1}^n \|\Delta_j\|^2)^{1/2}$ , together with Proposition 18 b) of Dörr et al. (2020), yield the assertion.  $\square$

Since, under the conditions of Theorem 5,  $\Gamma_a$  is strictly positive if the underlying distribution does not belong to  $\mathcal{N}_d$ ,  $U_{n,a}$  converges almost surely to  $\infty$  under such an alternative, and we have the following result.

**Corollary 1** *The test which reject the hypothesis  $H_0$  for large values of  $U_{n,a}$  is consistent against each fixed alternative satisfying  $\mathbb{E}\|X\|^4 < \infty$ .*

The next result, which corresponds to Theorem 13 of Dörr et al. (2020), shows that the (population) measure of multivariate skewness in the sense of Móri, Rohatgi and Székely emerges as the limit of  $\Gamma_a$ , after a suitable scaling, as  $a \rightarrow \infty$ .

**Theorem 6** *Under the condition  $\mathbb{E}\|X\|^6 < \infty$ , we have*

$$\lim_{a \rightarrow \infty} 2a \left(\frac{a}{\pi}\right)^{d/2} \Gamma_a = \left\| \mathbb{E} \left( \|X\|^2 X \right) \right\|^2.$$

**Proof** By definition,

$$\begin{aligned}
 \Gamma_a &= \int (\|t\|^2 - d)^2 \psi^+(t)^2 w_a(t) dt - 2 \int (\|t\|^2 - d) \psi^+(t) \Delta \psi^+(t) w_a(t) dt \\
 &\quad + \int (\Delta \psi^+(t))^2 w_a(t) dt \\
 &= \Gamma_{a,1} + \Gamma_{a,2} + \Gamma_{a,3} \text{ (say)}.
 \end{aligned}$$

In what follows, let  $Y, Z$  be independent copies of  $X$ . Since  $\psi^+(t)^2 = \mathbb{E}[\text{CS}^+(t^\top Y)\text{CS}^+(t^\top Z)]$ , the addition theorems for the cosine and the sine function and symmetry yield

$$\Gamma_{a,1} = \mathbb{E} \left[ \int (\|t\|^2 - d)^2 \cos(t^\top (Y - Z)) w_a(t) dt \right].$$

Putting  $c = Y - Z$ , display (6) then gives

$$\Gamma_{a,1} = \left(\frac{\pi}{a}\right)^{d/2} \frac{1}{16a^4} \mathbb{E} \left[ \left( 16d^2 a^3 (a - 1) + 4d(d + 2)a^2 + \|Y - Z\|^4 + (8da^2 - 4(d + 2)a)\|Y - Z\|^2 \right) \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right].$$

Likewise, it follows that  $\psi^+(t) \Delta \psi^+(t) = -\mathbb{E}[\|Y\|^2 \cos(t^\top(Y - Z))]$ , whence

$$\begin{aligned} \Gamma_{a,2} &= 2\mathbb{E} \left[ \|Y\|^2 \int (\|t\|^2 - d) \cos(t^\top(Y - Z)) w_a(t) dt \right] \\ &= -2 \left(\frac{\pi}{a}\right)^{d/2} \mathbb{E} \left[ \|Y\|^2 \left( \frac{\|Y - Z\|^2}{4a^2} + d - \frac{d}{2a} \right) \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right]. \end{aligned}$$

Finally,

$$\Gamma_{a,3} = \left(\frac{\pi}{a}\right)^{d/2} \mathbb{E} \left[ \|Y\|^2 \|Z\|^2 \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right],$$

and it follows that

$$\begin{aligned} 2a \left(\frac{a}{\pi}\right)^{d/2} \Gamma_a &= 2a \mathbb{E} \left[ \|Y\|^2 \|Z\|^2 \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right] \\ &\quad - 4a \mathbb{E} \left[ \|Y\|^2 \left( \frac{\|Y - Z\|^2}{4a^2} + d - \frac{d}{2a} \right) \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right] \\ &\quad + \frac{1}{8a^3} \mathbb{E} \left[ \left( 16d^2 a^3 (a - 1) + 4d(d + 2)a^2 + \|Y - Z\|^4 + (8da^2 - 4(d + 2)a)\|Y - Z\|^2 \right) \exp \left( -\frac{\|Y - Z\|^2}{4a} \right) \right]. \end{aligned}$$

Now, dominated convergence yields

$$\begin{aligned} 2a \left(\frac{a}{\pi}\right)^{d/2} \Gamma_a &= 2ad^2 - \frac{1}{2} \mathbb{E}[\|Y\|^2 \|Z\|^2 \|Y - Z\|^2] - 4ad^2 + d \mathbb{E}[\|Y\|^2 \|Y - Z\|^2] \\ &\quad + 2d^2 + 2d^2(a - 1) - d^3 + o(1) \end{aligned}$$

as  $a \rightarrow \infty$ . Since  $\mathbb{E}\|Y\|^2 = d = \mathbb{E}\|Z\|^2$  and  $\mathbb{E}(Y) = \mathbb{E}(Z) = 0$ , we have

$$\begin{aligned} \mathbb{E}[\|Y\|^2 \|Z\|^2 \|Y - Z\|^2] &= 2d \mathbb{E}\|Y\|^4 - 2\mathbb{E}\|X\|^2 X\|^2, \\ \mathbb{E}[\|Y\|^2 \|Y - Z\|^2] &= \mathbb{E}\|Y\|^4 + d^2, \end{aligned}$$

and the assertion follows. □

We close this section with a result on the asymptotic normality of  $U_{n,a}$  under fixed alternatives. That such a result holds in principle follows from Theorem 1 of Baringhaus et al. (2017). To state the main idea, write again  $\text{CS}^\pm(\xi) = \cos(\xi) \pm \sin(\xi)$  and notice that, by (10),  $U_{n,a} = \|\mathcal{S}_n\|_{\mathbb{H}}^2$ , where  $\mathcal{S}_n(t)$  is given in (11). Putting

$$\mathcal{S}_n^*(t) = \frac{\mathcal{S}_n(t)}{\sqrt{n}} = \frac{1}{n} \sum_{j=1}^n (\|Y_{n,j}\|^2 + \|t\|^2 - d)\text{CS}^+(t^\top Y_{n,j}), \quad t \in \mathbb{R}^d,$$

Theorem 5 and (14) show that

$$\begin{aligned} \sqrt{n} \left( \frac{U_{n,a}}{n} - \Gamma_a \right) &= \sqrt{n} (\|\mathcal{S}_n^*\|_{\mathbb{H}}^2 - \|z\|^2) = \sqrt{n} \langle \mathcal{S}_n^* - z, \mathcal{S}_n^* + z \rangle_{\mathbb{H}} \\ &= \sqrt{n} \langle \mathcal{S}_n^* - z, 2z + \mathcal{S}_n^* - z \rangle_{\mathbb{H}} \\ &= 2 \langle \mathcal{V}_n^*, z \rangle_{\mathbb{H}} + \frac{1}{\sqrt{n}} \|\mathcal{V}_n^*\|_{\mathbb{H}}^2, \end{aligned} \tag{15}$$

where  $\mathcal{V}_n^*(t) = \sqrt{n}(\mathcal{S}_n^*(t) - z(t))$ ,  $t \in \mathbb{R}^d$ . In the sequel, let  $\nabla(f)(t)$  denote the gradient of a differentiable function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , evaluated at  $t$ , and write  $\text{H}f(t)$  for the Hessian matrix of  $f$  at  $t$  if  $f$  is twice continuously differentiable. By proceeding as in the Proof of Theorem 6 of Dörr et al. (2020), there is a centered Gaussian element  $\mathcal{V}^*$  of  $\mathbb{H}$  having covariance kernel

$$K^*(s, t) = \mathbb{E}[h^*(X, s)h^*(X, t)], \quad s, t \in \mathbb{R}^d,$$

where

$$\begin{aligned} h^*(x, t) &= (\|x\|^2 + \|t\|^2 - d)\text{CS}^+(x, t) + 2\nabla\psi^-(t)^\top x - (\|t\|^2 - d)\psi^+(t) \\ &\quad + \left( \frac{1}{2}\nabla\Delta\psi^+(t)^\top - \frac{1}{2}(\|t\|^2 - d)\nabla\psi^+(t)^\top \right) (xx^\top - \text{I}_d)t \\ &\quad + (\Delta\psi^-(t) - (\|t\|^2 - d)\psi^-(t))t^\top x + x^\top \text{H}\psi^+(t)x, \end{aligned}$$

such that  $\mathcal{V}_n^* \xrightarrow{\mathcal{D}} \mathcal{V}^*$  as  $n \rightarrow \infty$ . In view of (15) and the fact that the distribution of  $2\langle \mathcal{V}^*, z \rangle_{\mathbb{H}}$  is centered normal, we have the following result.

**Theorem 7** *Under the standing assumptions stated at the beginning of this section, we have*

$$\sqrt{n} \left( \frac{U_{n,a}}{n} - \Gamma_a \right) \xrightarrow{\mathcal{D}} \text{N}(0, \sigma_a^2),$$

where

$$\sigma_a^2 = 4 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K^*(s, t)z(s)z(t)w_a(s)w_a(t) \, dsdt.$$

We remark that a consistent estimator of  $\sigma_a^2$  can be obtained by analogy with the reasoning given in Dörr et al. (2020), see Lemma 15, Theorem 16 and the Remark before Section 7 of that paper. Note that the above result differs from that of Theorem 6 in Dörr et al. (2020), since  $h^*(\cdot, \cdot)$ , compared to  $v(\cdot, \cdot)$  in display (23) of Dörr et al. (2020), has different components. Interestingly and in contrast to the null distribution, the formulas for  $h^*(\cdot, \cdot)$  are more involved.

## 5 Simulations

In this section, we present the results of a Monte Carlo simulation study on the finite-sample power of the tests based on  $U_{n,a}$  and  $T_{n,a}$ . This study is twofold in the sense that we consider testing for both univariate and multivariate normality, where the latter case is restricted to dimensions  $d \in \{2, 3, 5\}$ . Moreover, the study is designed to match and complement the counterparts in Dörr et al. (2020), Section 7, and Henze and Visagie (2020), since we take exactly the same setting with regard to sample size, nominal level of significance and selected alternative distributions. In this way, we facilitate an easy comparison with existing procedures. Note that the test families  $U_{n,a}$  and  $T_{n,a}$  have been implemented in the R package `mnt`, see Butsch and Ebner (2020). In the univariate case, we consider sample sizes  $n \in \{20, 50, 100\}$  and restrict the simulations to  $n \in \{20, 50\}$  in the multivariate setting. The nominal level of significance is fixed throughout all simulations to 0.05. We simulated empirical critical values under  $H_0$  for  $d^{-2} (a/\pi)^{d/2} U_{n,a}$  with 100,000 replications, see Table 1, and used Table 2 in Dörr et al. (2020) for critical values of  $T_{n,a}$ . In each table, the rows entitled '∞' give approximations of the quantiles of the limit random element  $U_{\infty,a} = \int S^2(t)w_a(t) dt$  in Theorem 3(b). The entries have been calculated by the method presented in Dörr et al. (2020), Section 7, setting  $\ell = 100,000$  and  $m = 2000$  for  $d \in \{2, 3, 5, 10\}$ . Note that this approach only relies on the structure of the covariance kernel given in Theorem 3(a), the multivariate normal distribution, and the weight function.

In the univariate case, we consider the following alternatives: symmetric distributions, like the Student  $t_\nu$ -distribution with  $\nu \in \{1, 3, 5, 10\}$  degrees of freedom (note that  $t_1$  is the standard Cauchy distribution), as well as the uniform distribution  $U(-\sqrt{3}, \sqrt{3})$ , and asymmetric distributions, such as the  $\chi_\nu^2$ -distribution with  $\nu \in \{5, 15\}$  degrees of freedom, the beta distributions  $B(1, 4)$  and  $B(2, 5)$ , and the gamma distributions  $\Gamma(1, 5)$  and  $\Gamma(5, 1)$ , both parametrized by their shape and rate parameter, the Gumbel distribution  $\text{Gum}(1, 2)$  with location parameter 1 and scale parameter 2, the Weibull distribution  $W(1, 0.5)$  with scale parameter 1 and shape parameter 0.5, and the lognormal distribution  $\text{LN}(0, 1)$ . As representatives of bimodal distributions, we simulate the mixture of normal distributions  $\text{NMix}(p, \mu, \sigma^2)$ , where the random variables are generated by  $(1 - p)N(0, 1) + pN(\mu, \sigma^2)$ ,  $p \in (0, 1)$ ,  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ . Note that these alternatives can also be found in the simulation studies presented in Betsch and Ebner (2020), Dörr et al. (2020), Romão et al. (2010). We chose these alternatives in order to ease the comparison with many other existing tests.

First we oppose the tests  $T_{n,a}$  and  $U_{n,a}$  in Table 2. Remarkably, the test based on  $U_{n,a}$  shows a better performance for the  $\text{NMix}$ -alternatives, especially for the choice of the tuning parameter  $a \in \{0.25, 0.5\}$ . On the other hand,  $U_{n,a}$  is almost uniformly

**Table 1** Empirical quantiles for  $d^{-2} (a/\pi)^{d/2} U_{n,a}$  and  $\alpha = 0.05$  (100,000 replications)

$d$	$n \setminus a$	0.1	0.25	0.5	1	2	3	5
1	20	147.99	25.86	7.14	2.46	1.47	1.27	1.02
	50	149.69	26.20	7.29	2.62	1.61	1.39	1.13
	100	150.85	26.45	7.34	2.65	1.63	1.42	1.16
	$\infty$	152.52	27.70	7.94	2.43	1.61	1.43	1.16
2	20	64.59	11.61	3.41	1.26	0.72	0.61	0.49
	50	65.63	11.87	3.50	1.33	0.79	0.68	0.56
	100	65.81	11.94	3.52	1.34	0.80	0.70	0.58
	$\infty$	66.33	12.12	3.46	1.39	0.78	0.71	0.58
3	20	46.49	8.22	2.45	0.91	0.49	0.40	0.33
	50	46.81	8.37	2.52	0.97	0.55	0.47	0.38
	100	46.88	8.41	2.53	0.97	0.56	0.48	0.40
	$\infty$	51.69	8.38	2.55	0.92	0.55	0.48	0.41
5	20	35.79	6.11	1.79	0.65	0.31	0.25	0.20
	50	36.05	6.20	1.85	0.70	0.37	0.30	0.25
	100	36.07	6.23	1.86	0.71	0.38	0.31	0.26
	$\infty$	39.38	6.27	1.90	0.68	0.38	0.32	0.27
10	20	30.13	4.93	1.33	0.42	0.17	0.12	0.09
	50	30.21	5.01	1.41	0.49	0.23	0.18	0.14
	100	30.22	5.02	1.42	0.50	0.25	0.19	0.15
	$\infty$	32.70	5.39	1.47	0.52	0.25	0.20	0.16

dominated by  $T_{n,a}$  for the  $t_\nu$ -distribution. If the underlying distribution is  $\chi^2$ , beta, gamma, Weibull, Gumbel or lognormal, both procedures have a comparable power. Table 4 in Dörr et al. (2020) also provides finite-sample powers of strong either time-honored or recent tests for normality, like the Shapiro–Wilk test, the Shapiro–Francia test, the Anderson–Darling test, the Baringhaus–Henze–Epps–Pulley test (BHEP), see Henze and Wagner (1997), the del Barrio–Cuesta–Albertos–Mátran–Rodríguez–Rodríguez test (BCMR), see del Barrio et al. (1999), and the Betsch–Ebner test, see Betsch and Ebner (2020). For a description of the test statistics and critical values, see Dörr et al. (2020) and the references therein. A comparison shows that, for suitable choice of the tuning parameter,  $U_{n,a}$  can compete with each of these tests, sometimes outperforming them, for example in case of the uniform distribution,  $n = 20$ , and  $a = 0.25$ , and the  $\chi^2_{15}$ -distribution for all sample sizes and  $a = 5$ , but mostly being on the same power level. It is interesting to see that the finite-sample power of  $U_{n,a}$  depends heavily on the choice of  $a$ . This observation is in contrast to the behavior of  $T_{n,a}$ , the power of which depends much less on  $a$ .

In the multivariate case, the alternative distributions are selected to match those employed in the simulation studies in Dörr et al. (2020), Henze and Visagie (2020), and are given as follows. Let  $N\text{Mix}(p, \mu, \Sigma)$  be the normal mixture distribution generated by

$$(1 - p) N_d(0, I_d) + p N_d(\mu, \Sigma),$$

**Table 2** Empirical power of  $U_{n,a}$  and  $T_{n,a}$  ( $d = 1, \alpha = 0.05, 10,000$  replications)

Alt.	$n \setminus a$	$U_{n,a}$					$T_{n,a}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
N(0, 1)	20	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5	5	5	5	5
NMix(0.3, 1, 0.25)	20	18	28	27	20	19	19	19	12	14	15	18	20	19	19
	50	45	65	61	49	46	45	43	28	33	38	44	46	45	43
	100	79	93	90	80	77	76	73	61	67	72	76	78	76	73
NMix(0.5, 1, 4)	20	20	40	44	39	36	35	33	36	37	38	38	37	35	33
	50	53	83	84	71	64	60	54	65	67	69	69	65	60	54
	100	88	99	99	95	91	87	79	90	93	94	94	92	88	79
$t_1$	20	72	84	88	86	85	84	82	86	87	87	86	85	84	82
	50	97	99	100	100	99	99	99	100	100	100	100	99	99	99
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$t_3$	20	12	24	37	39	38	37	36	40	40	40	39	38	37	37
	50	21	49	66	68	65	63	61	68	69	69	68	66	64	61
	100	36	75	88	89	87	85	82	87	89	89	88	87	86	83
$t_5$	20	6	11	19	22	21	21	21	23	23	23	22	21	21	21
	50	8	21	36	40	38	37	35	41	41	41	40	38	37	35
	100	11	34	54	59	57	55	51	58	61	61	60	57	55	51

Table 2 continued

Alt.	$n \setminus a$	$U_{n,a}$					$T_{n,a}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$t_{10}$	20	5	7	10	11	11	11	11	12	12	12	12	11	11	11
	50	6	8	15	18	17	17	16	19	19	19	18	17	17	16
	100	6	10	21	26	25	24	23	26	27	28	27	25	24	23
$U(-\sqrt{3}, \sqrt{3})$	20	12	22	18	2	1	1	1	1	1	1	1	2	1	1
	50	30	59	64	20	4	2	1	40	25	14	12	5	2	1
	100	67	95	97	87	50	20	3	98	96	89	80	59	24	3
$\chi^2_5$	20	10	24	39	41	42	42	42	33	35	37	40	42	42	42
	50	21	60	82	85	85	85	85	75	79	82	84	85	85	85
	100	44	93	99	100	100	100	99	99	99	99	100	100	100	99
$\chi^2_{15}$	20	5	9	16	18	19	19	19	16	16	17	18	19	19	19
	50	7	18	37	44	45	45	46	33	37	40	43	45	45	46
	100	9	34	64	75	76	76	77	60	67	71	74	76	76	77
$B(1, 4)$	20	21	39	49	46	46	46	45	34	38	42	45	45	50	59
	50	55	87	95	94	92	92	91	90	91	92	93	92	94	98
	100	92	100	100	100	100	100	100	100	100	100	100	100	100	100
$B(2, 5)$	20	7	11	15	14	14	14	14	11	12	13	15	15	15	16
	50	10	26	41	41	40	40	39	31	33	37	40	40	43	50
	100	16	55	80	82	80	79	78	78	79	79	80	79	81	90
$\Gamma(1, 5)$	20	39	64	75	73	73	73	72	60	64	67	71	73	73	72
	50	86	99	100	100	100	100	99	99	99	99	100	100	100	99
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100



Table 2 continued

Alt.	$n \setminus \alpha$	$U_{n,\alpha}$					$T_{n,\alpha}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$\Gamma(5, 1)$	20	6	12	21	24	25	25	25	19	21	22	24	24	25	25
	50	9	27	51	58	59	60	60	45	50	53	57	59	59	60
	100	13	54	83	90	90	90	90	79	85	88	89	90	90	90
$W(1, 0.5)$	20	39	65	76	74	74	73	73	61	65	68	72	74	74	73
	50	86	98	100	100	100	99	99	99	99	99	99	100	100	99
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$\text{Gum}(1, 2)$	20	7	16	28	32	33	33	33	29	30	31	33	33	33	33
	50	10	37	62	70	71	71	71	57	63	66	69	70	71	71
	100	17	67	90	95	95	95	95	88	92	94	95	95	95	95
$\text{LN}(0, 1)$	20	62	82	90	89	89	89	89	84	86	88	89	89	91	93
	50	97	100	100	100	100	100	100	100	100	100	100	100	100	100
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

where  $p \in (0, 1)$ ,  $\mu \in \mathbb{R}^d$ , and  $\Sigma$  is a positive definite matrix. In this notation,  $\mu = 3$  stands for a  $d$ -variate vector of 3's, and  $\Sigma = \mathbf{B}_d$  is a  $(d \times d)$ -matrix containing 1's on the main diagonal, and each off-diagonal entry has the value 0.9. We denote by  $t_\nu(0, \mathbf{I}_d)$  the multivariate  $t_\nu$ -distribution with  $\nu$  degrees of freedom, see Genz and Bretz (2009). The acronym  $\text{DIST}^d(\vartheta)$  stands for a  $d$ -variate random vector with i.i.d. marginal laws that belong to the distribution DIST with parameter  $\vartheta$ . In the sequel, DIST is either the lognormal distribution LN, the gamma distribution  $\Gamma$ , or the Pearson Type VII distribution  $\text{P}_{VII}$ . Note that  $t_1(0, \mathbf{I}_d)$  stands for the marginal standard Cauchy distribution  $C^d(0, 1)$  in the previous notation. For the latter distribution,  $\vartheta$  denotes the number of degrees of freedom. The spherical symmetric distributions have been simulated using the R package `distrEllipse`, see Ruckdeschel et al. (2006). These are denoted by  $S^d(\text{DIST})$ , where DIST stands for the distribution of the radii, and was chosen to be the exponential, the beta and the  $\chi^2$ -distribution.

Tables 3, 4, and 5 can be contrasted to Table 5–7 in Dörr et al. (2020), and for  $n = 50$ , with Tables 3–5 in Henze and Visagie (2020). Again, we start with a comparison of  $T_{n,a}$  and  $U_{n,a}$ . For  $d = 2$  (see Table 3 and Table 5 in Dörr et al. (2020)),  $T_{n,a}$  is outperformed by  $U_{n,a}$  for  $\text{NMix}(0.1, 3, \mathbf{I}_2)$  and  $\text{NMix}(0.9, 3, \mathbf{B}_2)$ , but shows a stronger performance for  $\text{NMix}(0.5, 3, \mathbf{B}_2)$ . In case of the multivariate  $t_\nu$ -distributions, both procedures have a similar performance, as well as for the  $\text{DIST}^d(\vartheta)$  distributions. The spherical symmetric distributions are dominated by  $U_{n,a}$  for a suitable choice of the tuning parameter, except for the  $S^d(\chi_5^2)$  distribution, where a similar behaviour is asserted. Again,  $U_{n,a}$  seems to be much more sensitive to the choice of a proper tuning parameter than  $T_{n,a}$ . Competing tests of multivariate normality are the Henze–Visagie test, see Henze and Visagie (2020), the Henze–Jiménez-Gamero test, see Henze and Jiménez-Gamero (2019), the BHEP-test, the Henze–Jiménez-Gamero–Meintanis test, see Henze et al. (2019), and the energy test, see Székely and Rizzo (2005). A description of the test statistics, as well as procedures for computing critical values is found in Henze and Visagie (2020). The BHEP-test performs best for the  $\text{NMix}(0.1, 3, \mathbf{I}_2)$ -distribution (NMIX1 in Henze and Visagie 2020) but is outperformed by  $T_{n,a}$  for  $\text{NMix}(0.5, 0, \mathbf{B}_2)$ , and by  $U_{n,a}$  for the  $\text{NMix}(0.9, 3, \mathbf{B}_2)$  (NMIX2 in Henze and Visagie 2020), where these procedures show the best performance of all tests considered. A similar behavior is observed for the  $t_\nu$ - and the spherical symmetric distributions, where again  $U_{n,a}$  and  $T_{n,a}$  are strong competitors to all procedures considered.

## 6 A real data example

As a real data example, we examine the meteorological data set `weather` provided in the R package `RandomFields`, see Schlather et al. (2019), which consists of differences between forecasts and observations (forecasts minus observations) of temperature and pressure at  $n = 157$  locations in the North American Pacific Northwest. The data are pointwise realizations of a bivariate ( $d = 2$ ) error Gaussian random field, see Fig. 1. The forecasts are from the GFS member of the University of Washington regional numerical weather prediction ensemble, see Eckel and Mass (2005), and they were valid on December 18, 2003 at 4 p.m. local time, at a forecast horizon of 48 hours. We ignore the given location of measurements in this evaluation and test the

**Table 3** Empirical power of  $U_{n,a}$  and  $T_{n,a}$  ( $d = 2, \alpha = 0.05, 10,000$  replications)

Alt.	$n \setminus \alpha$	$U_{n,a}$					$T_{n,a}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$N_2(0, I_2)$	20	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5	5	5	5	5
NMix(0.5, 3, $I_2$ )	20	11	27	16	3	4	4	3	3	2	2	3	3	3	3
	50	28	78	68	9	6	5	4	6	4	3	5	5	5	3
NMix(0.9, 3, $I_2$ )	20	8	17	38	39	41	39	39	33	35	35	40	39	39	39
	50	11	47	84	88	87	86	85	73	83	84	86	87	87	85
NMix(0.5, 0, $B_2$ )	20	7	12	19	19	17	17	15	19	19	18	18	17	17	15
	50	11	29	43	36	30	26	23	40	37	34	31	27	23	23
NMix(0.9, 0, $B_2$ )	20	6	11	25	27	25	25	25	27	28	27	26	26	25	25
	50	8	23	49	55	53	50	47	54	56	55	53	51	48	48
$t_1(0, I_2)$	20	85	95	98	97	96	95	95	97	97	97	97	96	95	95
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$t_3(0, I_2)$	20	16	34	53	57	56	53	51	58	55	57	57	54	52	52
	50	31	70	88	90	88	87	84	88	89	90	89	87	84	84
$t_5(0, I_2)$	20	8	15	30	34	33	30	30	34	34	34	34	31	30	30
	50	11	31	57	64	63	59	56	63	62	64	64	60	57	57
$t_{10}(0, I_2)$	20	6	7	14	16	16	15	15	17	16	16	16	15	15	15
	50	6	11	23	30	28	27	25	30	29	30	29	28	28	26

Table 3 continued

Alt.	$n \setminus \alpha$	$U_{n,\alpha}$					$T_{n,\alpha}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$L^2(0, 1)$	20	70	92	97	96	96	95	96	91	92	92	94	96	95	96
	50	99	100	100	100	100	100	100	100	100	100	100	100	100	100
$B^2(1, 2)$	20	11	19	19	10	12	13	13	6	5	5	6	10	12	12
	50	25	60	73	54	54	52	47	27	30	25	26	47	49	46
$P^2_{VIT}(5)$	20	7	12	25	28	27	26	26	29	28	29	28	28	26	26
	50	9	24	48	56	54	50	47	55	55	55	57	55	51	47
$P^2_{VIT}(10)$	20	6	7	12	14	13	13	12	13	13	14	14	13	13	12
	50	6	9	19	24	23	21	20	24	23	24	25	23	22	21
$S^2(\text{Exp}(1))$	20	60	77	82	78	76	72	68	77	75	76	77	77	73	69
	50	97	99	100	99	98	97	95	98	98	98	99	98	98	96
$S^2(B(1, 2))$	20	29	34	30	23	21	19	16	24	23	21	22	22	20	17
	50	65	73	60	39	31	25	19	34	32	33	34	31	26	19
$S^2(\chi^2_5)$	20	6	10	19	22	22	21	20	24	22	22	23	22	22	20
	50	7	18	37	42	41	38	34	42	42	42	42	43	39	35

**Table 4** Empirical power of  $U_{n,\alpha}$  ( $d = 3, \alpha = 0.05, 10,000$  replications)

Alt.	$n \setminus \alpha$	$U_{n,\alpha}$					$T_{n,\alpha}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$N_3(0, I_3)$	20	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5	5	5	5	5
NMix(0.5, 3, $I_3$ )	20	11	20	11	3	4	4	5	3	3	2	2	3	4	4
	50	20	69	51	6	5	5	4	4	5	3	2	3	4	4
NMix(0.9, 3, $I_3$ )	20	9	18	36	40	40	40	39	31	35	35	36	38	41	39
	50	12	47	89	92	92	91	89	70	84	87	86	90	91	90
NMix(0.5, 0, $B_3$ )	20	10	25	38	36	32	28	27	38	38	34	34	33	30	27
	50	19	61	82	72	62	55	46	75	76	73	66	63	58	48
NMix(0.9, 0, $B_3$ )	20	9	19	37	42	43	39	37	42	43	42	42	44	41	38
	50	13	45	75	80	79	76	75	80	80	80	80	80	78	76
$t_1(0, I_3)$	20	89	98	99	99	99	99	98	99	99	99	99	99	99	98
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$t_3(0, I_3)$	20	18	44	67	70	69	65	63	71	70	70	69	70	67	64
	50	42	85	96	97	97	96	94	97	97	97	97	97	96	95
$t_5(0, I_3)$	20	10	21	39	44	42	39	38	44	43	43	44	43	41	39
	50	16	46	75	82	79	77	72	82	80	79	80	80	79	74
$t_{10}(0, I_3)$	20	6	9	18	21	20	18	18	22	20	20	21	21	20	18
	50	8	16	34	42	41	37	34	44	41	41	42	43	40	36

Table 4 continued

Alt.	$n \setminus \alpha$	$U_{n,\alpha}$					$T_{n,\alpha}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$L^3(0, 1)$	20	69	93	98	97	97	97	97	94	95	95	95	97	97	97
	50	99	100	100	100	100	100	100	100	100	100	100	100	100	100
$B(1, 2)$	20	9	13	12	7	9	10	10	3	4	4	4	6	8	10
	50	17	45	58	38	46	49	44	7	17	16	12	25	40	42
$P^3_{VII}(5)$	20	8	14	27	32	30	28	27	33	33	32	32	32	30	28
	50	11	30	57	65	62	59	54	67	64	64	65	64	62	56
$P^3_{VII}(10)$	20	6	8	12	15	13	13	12	15	14	14	15	14	14	13
	50	6	10	22	28	27	25	22	29	28	28	28	28	27	23
$S^3(\text{Exp}(1))$	20	86	95	97	95	94	93	89	95	95	95	94	95	94	90
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$S^3(B(1, 2))$	20	61	73	73	66	60	55	48	69	64	61	62	61	57	50
	50	96	99	98	95	94	89	78	93	91	90	92	93	90	80
$S^3(\chi^2_5)$	20	14	31	48	52	49	44	41	55	51	49	50	50	47	42
	50	31	68	86	88	86	82	76	88	85	85	85	86	84	78

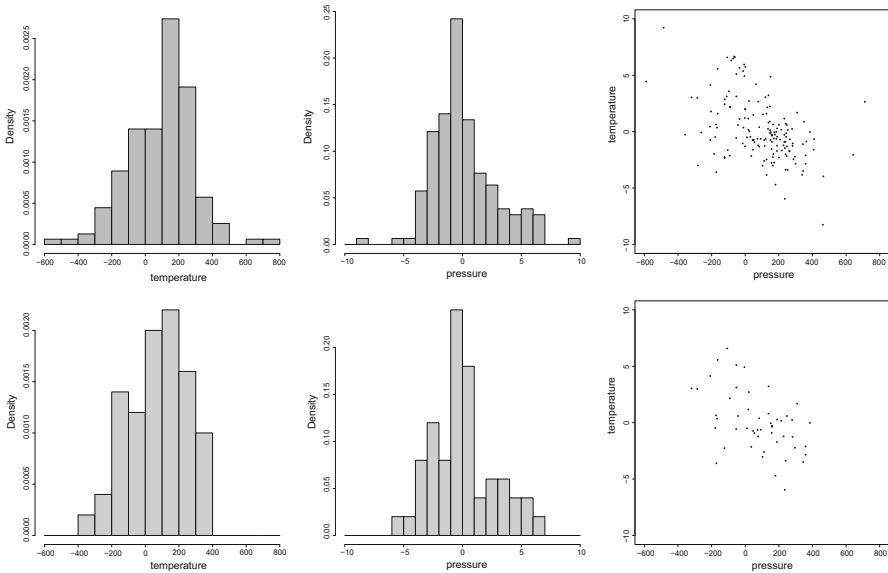
**Table 5** Empirical power of  $U_{n,a}$  and  $T_{n,a}$  ( $d = 5, \alpha = 0.05, 10,000$  replications)

Alt.	$n \setminus \alpha$	$U_{n,a}$					$T_{n,a}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$N_5(0, I_5)$	20	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	50	5	4	5	6	5	5	5	5	5	6	5	5	5	5
	20	8	9	6	3	4	4	4	2	3	3	3	3	3	4
	50	9	22	12	3	3	4	4	1	3	3	2	2	2	3
	20	9	17	28	30	31	32	31	26	27	29	29	30	33	33
$NMix(0.9, 3, I_5)$	50	13	37	75	79	83	85	84	48	65	78	76	74	80	84
	20	23	57	76	69	65	58	52	71	72	71	65	65	63	57
	50	47	97	100	99	97	94	89	98	99	99	97	95	95	92
	20	13	36	55	60	58	56	53	57	56	58	60	60	60	57
	50	43	86	94	95	95	94	93	95	95	95	95	95	95	95
$NMix(0.9, 0, B_5)$	20	92	99	100	100	100	100	99	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	20	28	66	82	84	81	79	75	84	84	84	83	82	82	79
	50	74	98	100	100	100	100	99	100	100	100	100	100	100	99
	20	14	37	56	60	55	52	47	59	59	58	59	57	57	52
$t_1(0, I_5)$	50	38	82	94	96	95	93	90	96	96	95	94	95	94	92
	20	8	16	27	28	26	23	21	28	28	28	27	28	27	24
	50	14	39	61	66	63	58	52	67	66	63	63	63	62	57
	20	64	94	98	98	98	98	98	95	97	98	97	97	97	98
	50	99	100	100	100	100	100	100	100	100	100	100	100	100	100
$t_3(0, I_5)$	20	14	37	56	60	55	52	47	59	59	58	59	57	57	52
	50	38	82	94	96	95	93	90	96	96	95	94	95	94	92
	20	8	16	27	28	26	23	21	28	28	28	27	28	27	24
	50	14	39	61	66	63	58	52	67	66	63	63	63	62	57
	20	64	94	98	98	98	98	98	95	97	98	97	97	97	98
$t_5(0, I_5)$	50	99	100	100	100	100	100	100	100	100	100	100	100	100	100

Table 5 continued

Alt.	$n \setminus \alpha$	$U_{n,\alpha}$					$T_{n,\alpha}$								
		0.1	0.25	0.5	1	2	3	5	0.1	0.25	0.5	1	2	3	5
$B^5(1, 2)$	20	6	7	5	4	6	8	8	2	3	3	3	3	4	6
	50	8	14	14	7	23	36	35	0	3	6	4	4	10	26
$P^5_{VII}(5)$	20	9	19	32	35	32	30	27	35	34	34	34	35	34	29
	50	19	50	73	76	73	70	63	77	76	75	75	74	74	69
$P^5_{VII}(10)$	20	6	10	14	15	13	13	12	15	14	15	15	14	14	13
	50	9	17	29	33	31	28	25	34	32	32	32	32	32	28
$S^5(\text{Exp}(1))$	20	97	100	100	100	100	100	99	100	100	100	100	100	100	100
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$S^5(B(1, 2))$	20	87	97	98	97	95	92	88	97	97	96	95	95	94	91
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$S^5(\chi^2_5)$	20	42	76	88	87	85	80	73	89	88	87	86	85	84	78
	50	86	100	100	100	100	100	99	100	100	100	100	100	100	100





**Fig. 1** Histogram of  $n = 157$  (upper row) and  $n = 50$  (lower row) differences between forecasts and observations of temperature (left) and pressure (middle) and scatterplot of temperature and pressure (right) in the North American Pacific Northwest

**Table 6** Empirical  $p$ -values for  $U_{n,a}$  for univariate and bivariate cases of the complete data set  $n = 157$  and the subsample  $n = 50$  (10,000 replications)

Diff.	$n \setminus a$	0.1	0.25	0.5	1	2	3	5
Temperature	157	0.0128	0	0	0.0001	0.0001	0	0.0001
Pressure		0.1042	0.0155	0.0102	0.0235	0.0292	0.0322	0.0407
Bivariate		0.0001	0	0	0	0	0	0
Temperature	50	0.9472	0.6847	0.3675	0.3144	0.3145	0.3168	0.3337
Pressure		0.1649	0.2019	0.1822	0.2282	0.2169	0.2101	0.2109
Bivariate		0.8485	0.6694	0.5528	0.5413	0.3497	0.2998	0.2879

hypothesis that the pairs of differences can be modeled as i.i.d. copies from a bivariate normal distribution. In Table 6, we calculate empirical  $p$ -values based on 10,000 replications for  $U_{n,a}$  for the univariate differences of temperature and pressure, as well as for the bivariate data for the whole data set,  $n = 157$ , and for a random selection of  $n = 50$  points (selected in R with function `sample()` and seed fixed to '0721'). Regarding the complete data set, we reject the hypothesis of normality in nearly all cases on a 5% level of significance, while on a 1% level of significance we are not able to reject  $H_0$  for the differences in pressure. However, for the temperature and the bivariate data the hypothesis of normality is nearly always rejected. These results are not surprising, since the weather data set is an example of influence of spatial correlation, which has to be carefully modeled. In Gneiting et al. (2010), a bivariate Gaussian random field is fitted to the data, taking the mentioned spatial correlation

**Table 7** Empirical  $p$ -values for  $T_{n,a}$  for univariate and bivariate cases of the complete data set  $n = 157$  and the subsample  $n = 50$  (10,000 replications)

Diff.	$n \setminus a$	0.1	0.25	0.5	1	2	3	5
Temperature		0.0002	0.0003	0.0002	0.0001	0.0001	0.0001	0.0002
Pressure	157	0.0300	0.0231	0.0236	0.0269	0.0297	0.0334	0.0411
Bivariate		0	0	0	0	0	0	0.0001
Temperature		0.5775	0.4011	0.3100	0.2502	0.2173	0.2104	0.2118
Pressure	50	0.5871	0.4401	0.3787	0.3393	0.3164	0.3214	0.3375
Bivariate		0.7589	0.7084	0.7279	0.6455	0.4094	0.3236	0.2948

into account, for a visualization of the locations see Figure 3 in Gneiting et al. (2010). For the subsample of points we see that the structure vanishes, and we throughout do not reject the hypotheses. In Table 7 we conduct the same study using the method  $T_{n,a}$  to contrast the empirical  $p$ -values to those of  $U_{n,a}$ . As would be expected, we nearly draw the same conclusions, although we can reject the hypothesis of normality for pressure in all cases on a 5% level of significance for the full data set. In comparison, the  $p$ -values in Table 7 show a smaller fluctuation than in Table 6. Here, we have only applied the methods as a proof of principle.

## 7 Conclusions and outlook

We have introduced and studied a new affine invariant class of tests for multivariate normality that is easy to apply and consistent against general alternatives. Although consistency has only been proved under the condition  $\mathbb{E}\|X\|^4 < \infty$ , the test should be “all the more consistent” if  $\mathbb{E}\|X\|^4 = \infty$ , and we conjecture that, as is the case for the BHEP-tests, also the test based on  $U_{n,a}$  is consistent against *each* nonnormal alternative distribution. A further topic of research would be to choose the tuning parameter  $a$  in an adaptive way, similar to the bootstrap based univariate approaches in Allison and Santana (2015) and Tenreiro (2019). It would also be of interest to obtain more information on the limit null distribution of  $U_{n,a}$ . We finish the outlook by pointing out that, with respect to the references in the introduction regarding other procedures and distributions, a similar analysis can be performed, and it is of theoretical and practical relevance to study the resulting statistics in order to assess the influence of the options of estimating or not estimating certain of the pertaining functions.

After a comparison of  $U_{n,a}$  and  $T_{n,a}$  from Dörr et al. (2020), and in view of the results of the simulation study, we recommend to use  $T_{n,a}$ , since it seems to be more robust with respect to the choice of the tuning parameter  $a$ . Nevertheless,  $U_{n,a}$  is a strong competitor, and with a suitable data driven procedure for the choice of  $a$  at hand,  $U_{n,a}$  may turn out to be a favorable choice over the most classical and recent tests of uni- and multivariate normality.

**Acknowledgements** The authors thank two anonymous referees and the editor for valuable remarks.

**Funding** Open Access funding provided by Projekt DEAL.

## Compliance with ethical standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allison JS, Santana L (2015) On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *J Stat Comput Simul* 85(16):3276–3288
- Baringhaus L, Ebner B, Henze N (2017) The limit distribution of weighted  $l^2$ -statistics under fixed alternatives, with applications. *Ann Inst Stat Math* 69(5):969–995
- Baringhaus L, Henze N (1991) A class of consistent tests for exponentiality based on the empirical Laplace transform. *Ann Instit Stat Math* 43(3):551–564
- Baringhaus L, Henze N (1991) Limit distributions for measures of multivariate skewness and kurtosis based on projections. *J Multivar Anal* 38(1):51–69
- Baringhaus L, Henze N (1992) Limit distributions for Mardia's measure of multivariate skewness. *Ann Stat* 20(4):1889–1902
- del Barrio E, Cuesta-Albertos JA, Matran C, Rodríguez-Rodríguez JM (1999) Tests of goodness of fit based on the  $L_2$ -Wasserstein distance. *Ann Stat* 27(4):1230–1239
- Betsch S, Ebner B (2020) Testing normality via a distributional fixed point property in the Stein characterization. *TEST* 29(1):105–138
- Butsch L, Ebner B (2020) mnt: Affine invariant tests of multivariate normality. <https://CRAN.R-project.org/package=mnt>. R package version 1.3
- Dörr P, Ebner B, Henze N (2020) Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. *Scand J Stat* pp. 1–34. <https://doi.org/10.1111/sjos.12477>
- Eaton ML, Perlman MD (1973) The non-singularity of generalized sample covariance matrices. *Ann Stat* 1(4):710–717
- Eckel FA, Mass CF (2005) Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecast* 20(3):328–350
- Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. *Lecture notes in statistics*, 195. Springer, Dordrecht
- Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. *J Am Stat Assoc* 105(491):1167–1177
- Henze N (1994) On Mardia's kurtosis test for multivariate normality. *Commun Stat – Theory Methods* 23(4):1031–1045
- Henze N (1997) Limit laws for multivariate skewness in the sense of Móri, Rohatgi and Székely. *Stat Probab Lett* 33(3):299–307
- Henze N, Jiménez-Gamero MD (2019) A class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. *TEST* 28(2):499–521
- Henze N, Jiménez-Gamero MD, Meintanis SG (2019) Characterizations of multinormality and corresponding tests of fit, including for GARCH models. *Econom Theory* 35(3):510–546
- Henze N, Klar B (2002) Goodness-of-fit tests for the inverse Gaussian distribution based on the empirical Laplace transform. *Ann Instit Stat Math* 54(2):425–444
- Henze N, Meintanis SG, Ebner B (2012) Goodness-of-fit tests for the Gamma distribution based on the empirical Laplace transform. *Commun Stat – Theory and Methods* 41(9):1543–1556

- Henze N, Visagie J (2020) Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Ann Inst Stat Math* 72:1109–1136
- Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. *J Multivar Anal* 61(1):1–23
- Hsing T, Eubank R (2015) *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley, New York
- Mardia KV (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57(3):519–530
- Meintanis S, Iliopoulos G (2003) Tests of fit for the Rayleigh distribution based on the empirical Laplace transform. *Ann Instit Stat Math* 55(1):137–151
- Meintanis SG (2010) Testing skew normality via the moment generating function. *Math Methods Stat* 19(1):64–72
- Meintanis SG, Hlávka Z (2010) Goodness-of-fit tests for bivariate and multivariate skew-normal distributions. *Scand J Stat* 37(4):701–714
- Móri TF, Rohatgi VK, Székely GJ (1993) On multivariate skewness and kurtosis. *Teoriya Veroyatnostoni I Yeye Primeniya* 38(3):675–679
- Riad M, Mabood OFAE (2018) A new goodness of fit test for the beta distribution based on the empirical Laplace transform. *Adv Appl Stat* 53(2):165–177
- Romão X, Delgado R, Costa A (2010) An empirical power comparison of univariate goodness-of-fit tests for normality. *J Stat Comput Simul* 80(5):545–591
- Ruckdeschel P, Kohl M, Stabla T, Camphausen F (2006) S4 classes for distributions. *R News* 6(2):2–6
- Schlather M, Malinowski A, Oesting M, Boecker D, Strokorb K, Engelke S, Martini J, Ballani F, Moreva O, Auel J, Menck PJ, Gross S, Ober U, Ribeiro P, Ripley BD, Singleton R, Pfaff B, R Core Team (2019) *RandomFields: simulation and analysis of random fields*. <https://cran.r-project.org/package=RandomFields>. R package version 3.3.6
- Székely GJ, Rizzo ML (2005) A new test for multivariate normality. *J Multivar Anal* 93:58–80
- Tenreiro C (2019) On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *J Stat Comput Simul* 89(10):1780–1797