Received: 1 October 2019 | Revised: 28 April 2020 | Accepted: 31 May 2020

DOI: 10.1111/sjos.12477

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces

Philip Dörr | **Bruno Ebner** | **Norbert Henze**

Institute of Stochastics, Karlsruhe Institute of Technology (KIT),

Correspondence
Bruno Ebner, Institut für Stochastik, Karlsruher Institut für Technologie (KIT), Englerstr. 2, D-76133 Karlsruhe, Germany.
Email: Bruno.Ebner@kit.edu

Abstract

We study a novel class of affine invariant and consistent tests for normality in any dimension in an i.i.d.-setting. The tests are based on a characterization of the standard d-variate normal distribution as the unique solution of an initial value problem of a partial differential equation motivated by the harmonic oscillator, which is a special case of a Schrödinger operator. We derive the asymptotic distribution of the test statistics under the hypothesis of normality as well as under fixed and contiguous alternatives. The tests are consistent against general alternatives, exhibit strong power performance for finite samples, and they are applied to a classical data set due to R.A. Fisher. The results can also be used for a neighborhood-of-model validation procedure.

KEYWORDS

affine invariance, consistency, empirical characteristic function, harmonic oscillator, neighborhood-of-model validation, *test for multivariate normality*

1 | INTRODUCTION

The multivariate normal distribution plays a key role in classical and hence widely used procedures, such as multivariate linear regression models with fixed effects and multivariate analysis of variance, see Eaton (1983), but it is also widespread in image analysis, see Ghaziasgar, Bagula, Thron, and Ajayi (2020), or in anomaly detection in machine learning, see Luo and Zhong (2017). Serious statistical inference that involves the assumption of multivariate normality should therefore start with a test of fit to this model. There is a continuing interest in this testing problem, as evidenced by a multitude of papers. The proposed tests may be roughly classified as follows: Arcones (2007), Baringhaus and Henze (1988), Henze and Wagner (1997), Henze and Zirkler (1990), Pudelko (2005), and Tenreiro (2009) consider tests based on the empirical characteristic function, while Henze and Jiménez-Gamero (2019), Henze, Jiménez-Gamero, and Meintanis (2019), and Henze and Visagie (2019) employ the empirical moment generating function. A classical (and still popular) approach is to consider measures of multivariate skewness and kurtosis (see Doornik and Hansen, 2008; Kankainen, Taskinen, and Oja, 2007; Malkovich and Afifi, 1973; Mardia, 1970; Mardia, 1974; Móri, Rohatgi, and Székely, 1993), as supposedly diagnostic tools with regard to the kind of deviation from normality when this hypothesis has been rejected, but the deficiencies of those measures in this regard have been clearly demonstrated (see Baringhaus and Henze, 1991; Baringhaus and Henze, 1992; Henze, 1994a, 1994b, 1997b). Other approaches involve generalizations of tests for univariate normality (see Kim and Park, 2018; Sürücü, 2006; Villaseñor Alva and González Estrada, 2009), the examination of nonlinearity of dependence (see Cox and Small, 1978; Ebner, 2012), canonical correlations (see Thulin, 2014), and the notion of energy (see Székely and Rizzo, 2005). For a survey of affine invariant tests for multivariate normality (see Henze, 2002). Monte Carlo studies can be found in Farrell, Salibian-Barrera, and Naczk (2007), Mecklin and Mundfrom (2005), and Voinov, Pya, Makarov, and Voinov (2016).

To be specific, let $X, X_1, \dots, X_n, \dots$ be a sequence of independent identically distributed (i.i.d.) d-dimensional random (column) vectors, which are defined on some common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Here, $d \geq 1$ is a fixed integer, which means that the univariate case is deliberately not excluded. We write $\mathbb{P}^X$ for the distribution of X. The d-variate normal distribution with expectation μ and nonsingular covariance matrix Σ will be denoted by $\mathrm{N}_d(\mu, \Sigma)$. Furthermore,

$$\mathcal{N}_d = \{\mathrm{N}_d(\mu, \Sigma) : \mu \in \mathbb{R}^d,\ \Sigma \in \mathbb{R}^{d\times d} \text{positive definite}\}$$

stands for the family of nondegenerate d-variate normal distributions. A check of the assumption of multivariate normality means to test the hypothesis

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

against general alternatives.

Writing I_d for the unit matrix of order d, our novel idea for testing H_0 is to use a characterization of the Fourier transform of $\mathrm{N}_d(0,\mathrm{I}_d)$ as the unique solution of an initial value problem of a partial differential equation motivated by the harmonic oscillator, which is a special case of a Schrödinger operator. The proposed test statistic is based on the squared norm of a functional of the empirical characteristic function in a suitably weighted L^2-space. This statistic is close to zero under the hypothesis (1), and rejection will be for large values of the test statistic.

Let $L^2(\mathbb{R}^d)$ be the space of square integrable functions, equipped with the usual norm and scalar product $\langle \cdot, \cdot \rangle$. Consider for sufficiently regular $f \in L^2(\mathbb{R}^d)$ the partial differential equation

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

Here, Δ stands for the Laplace operator, and $\|\cdot\|$ denotes the Euclidean norm. Notice that we can rewrite (2) as $(-\Delta + \|x\|^2 - d)f(x) = 0$ or, equivalently, as

$$\sum_{j=1}^{d}\left(-\frac{\partial^2}{\partial x_j^2} + x_j^2 - 1\right)f(x) = 0, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d. \tag{3}$$

The operator $-\Delta + \|x\|^2 - d$ is known as the multidimensional harmonic oscillator (see Gustafson, 2011). The harmonic oscillator plays a fundamental role in quantum mechanics, since it describes the behavior of a particle in an attractive electrostatic potential. To solve the PDE in (2) means to search for the null state of the particle. In the univariate case, (2) reduces to a fixed point problem or, equivalently, to the problem of finding the eigenfunction that corresponds to the eigenvalue 1, of the Hermite operator, see equation (1.1.9) in Thangavelu (1993). The solution of this problem is the 0th Hermite function, which coincides with the solution given in the following theorem.

Theorem 1. *The characteristic function*

$$\psi(t) = \exp\left(-\frac{\|t\|^2}{2}\right), \quad t \in \mathbb{R}^d, \tag{4}$$

of the d-variate standard normal distribution $N_d(0, \mathrm{I}_d)$ is the unique solution of (2).

Proof. Let f be an arbitrary solution of (2). Writing i for the imaginary unit, we introduce the creation and annihilation operators $a_j = x_j + \mathrm{i}p_j$ and $a_j^\star = x_j - \mathrm{i}p_j$, $j = 1, \dots, d$, where $p_j = -\mathrm{i}\frac{\partial}{\partial x_j}$, $j = 1, \dots, d$. For each $j \in \{1, \dots, d\}$ we have

$$a_j^\star a_j = \left(x_j - \frac{\partial}{\partial x_j}\right)\left(x_j + \frac{\partial}{\partial x_j}\right) = x_j^2 + x_j\frac{\partial}{\partial x_j} - \frac{\partial}{\partial x_j}x_j - \frac{\partial^2}{\partial x_j^2} = -\frac{\partial^2}{\partial x_j^2} + x_j^2 - 1.$$

So we can rewrite (3) as $\sum_{j=1}^d a_j^\star a_j f = 0$, which implies

$$\left\langle f, \sum_{j=1}^{d} a_j^\star a_j f \right\rangle = \sum_{j=1}^{d}\langle f, a_j^\star a_j f\rangle = \sum_{j=1}^{d}\langle a_j f, a_j f\rangle = \sum_{j=1}^{d}\|a_j f\|^2 = 0$$

and thus $a_j f = 0$ for each $j \in \{1, \dots, d\}$. We therefore have $(x + \nabla)f = 0$, where ∇ denotes the gradient operator. By the last statement and the product rule, it follows that

$$\nabla\left(\exp\left(\frac{\|x\|^2}{2}\right)f\right) = x\exp\left(\frac{\|x\|^2}{2}\right)f - x\exp\left(\frac{\|x\|^2}{2}\right)f = 0,$$

which, in view of the condition $f(0) = 1$, completes the proof. ■

Remark 1. The operator $H = -\Delta + \|x\|^2$ is the Hermite operator in $\mathbb{R}^d$, and ψ is the product of the one-dimensional 0th Hermite functions. Therefore, since $H\psi = d\psi$ (as we have shown in Theorem 1), ψ is the eigenfunction associated with the eigenvalue d. For details on the Hermite operator in $\mathbb{R}^d$ and corresponding eigenfunctions (see p. 5 of Thangavelu, 1993).

In this article, we study a family of affine invariant test statistics for H_0 that is based on the characterization given in Theorem 1. Since the class $\mathcal{N}_d$ is closed under full-rank affine transformations, Theorem 1 does not restrict the scope of the testing problem. We make the tacit standing assumption that $\mathbb{P}^X$ is absolutely continuous with respect to Lebesgue measure, and that $n \geq d+1$. Let $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j$ denote the sample mean and $S_n = n^{-1}\sum_{j=1}^n (X_j - \overline{X}_n)(X_j - \overline{X}_n)^\top$ the sample covariance matrix of $X_1, \dots, X_n$, where $x^\top$ means transposition of a column vector x. The assumptions made above guarantee that S_n is invertible almost surely (see Eaton and Perlman, 1973). The test statistic will be based on the so-called *scaled residuals*

$$Y_{n,j} = S_n^{-1/2}(X_j - \overline{X}_n), \quad j = 1, \dots, n, \tag{5}$$

which represent an empirical standardization of the data. Here, $S_n^{-1/2}$ is the unique symmetric positive definite square root of S_n^{-1}. The test statistic will be based on the empirical characteristic function

$$\psi_n(t) = \frac{1}{n}\sum_{j=1}^n \exp(it^\top Y_{n,j}), \quad t \in \mathbb{R}^d,$$

of $Y_{n,1}, \dots, Y_{n,n}$. Notice that an application of the Laplace operator Δ to ψ_n yields $\Delta\psi_n(t) = -\frac{1}{n}\sum_{j=1}^n \|Y_{n,j}\|^2 \exp(it^\top Y_{n,j})$, $t \in \mathbb{R}^d$. Motivated by (2) and Theorem 1, we propose the weighted L^2-statistic

$$\begin{aligned} T_{n,a} &= n\int |\Delta\psi_n(t) - \Delta\psi(t)|^2 w_a(t)\,\mathrm{d}t \\ &= n\int \left|\frac{1}{n}\sum_{j=1}^n \|Y_{n,j}\|^2 \exp(it^\top Y_{n,j}) + (\|t\|^2 - d)\exp\left(-\frac{\|t\|^2}{2}\right)\right|^2 w_a(t)\,\mathrm{d}t, \end{aligned}$$

where

$$w_a(t) = \exp(-a\|t\|^2), \quad t \in \mathbb{R}^d, \tag{6}$$

and $a > 0$ is a fixed constant. Moreover, $|z|$ is the modulus of a complex number z, and integration is, unless otherwise specified, over $\mathbb{R}^d$. In principle, other weight functions than w_a are conceivable in the definition of $T_{n,a}$. Since, for $c \in \mathbb{R}^d$ (see Henze and Zirkler, 1990, p. 3601),

$$\int \cos(t^\top c)\exp(-a\|t\|^2)\mathrm{d}t = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \exp\left(-\frac{\|c\|^2}{4a}\right), \tag{7}$$

as well as

$$\begin{aligned} &\int (\|t\|^2 - d)\cos(t^\top c)\exp\left(-\left(a + \frac{1}{2}\right)\|t\|^2\right)\mathrm{d}t \\ &\quad = -\frac{(2\pi)^{\frac{d}{2}}(\|c\|^2 + 2da(2a+1))}{(2a+1)^{2+\frac{d}{2}}}\exp\left(-\frac{\|c\|^2}{2(2a+1)}\right), \end{aligned} \tag{8}$$

$$\int (\|t\|^2 - d)^2 \exp(-(a+1)\|t\|^2)\mathrm{d}t = \frac{\pi^{\frac{d}{2}}}{(a+1)^{2+\frac{d}{2}}} \left(a(a+1)d^2 + \frac{d(d+2)}{4} \right), \tag{9}$$

an attractive feature of the choice of w_a is that the test statistic takes the simple form

$$T_{n,a} = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \frac{1}{n} \sum_{j,k=1}^{n} \|Y_{n,j}\|^2 \|Y_{n,k}\|^2 \exp\left(-\frac{1}{4a}\|Y_{n,j} - Y_{n,k}\|^2\right) \tag{10}$$

$$- \frac{2(2\pi)^{\frac{d}{2}}}{(2a+1)^{2+\frac{d}{2}}} \sum_{j=1}^{n} \|Y_{n,j}\|^2 (\|Y_{n,j}\|^2 + 2da(2a+1)) \exp\left(-\frac{1}{2}\frac{\|Y_{n,j}\|^2}{2a+1}\right) \tag{11}$$

$$+ n \frac{\pi^{\frac{d}{2}}}{(a+1)^{2+\frac{d}{2}}} \left(a(a+1)d^2 + \frac{d(d+2)}{4} \right), \tag{12}$$

which is amenable to computational purposes. Moreover, $T_{n,a}$ depends only on the scalar products $Y_{n,i}^\top Y_{n,j} = (X_i - \overline{X}_n)^\top S_n^{-1}(X_j - \overline{X}_n)$, where $i,j \in \{1, \dots, n\}$. This shows that $T_{n,a}$ is invariant with respect to full rank affine transformations of $X_1, \dots, X_n$, that is, we have $T_{n,a}(AX_1 + b, \dots, AX_n + b) = T_{n,a}(X_1, \dots, X_n)$ for each regular $(d \times d)$-matrix A and each $b \in \mathbb{R}^d$. Moreover, not even the square root $S_n^{-1/2}$ of S_n^{-1} is needed. The three-line-numbering above will become clear later when we consider the limit of $T_{n,a}$ as $a \to 0$.

The rest of the article is organized as follows: In Section 2, we show that, as the tuning parameter a tends to infinity, the test statistic $T_{n,a}$, after a suitable scaling, converges to a certain measure of multivariate skewness. On the other hand, a time-honoured measure of multivariate kurtosis emerges as $a \to 0$. Section 3 presents a basic Hilbert space central limit theorem, which proves beneficial for obtaining the limit distribution of $T_{n,a}$ both under H_0 and under fixed alternatives to normality. In Section 4, we derive the limit null distribution of $T_{n,a}$ as $n \to \infty$. Section 5 considers the behavior of $T_{n,a}$ with respect to contiguous alternatives to H_0. In Section 6, we show that the test for multivariate normality that rejects H_0 for large values of $T_{n,a}$ is consistent against general alternatives. Moreover, the limit distribution of $T_{n,a}$ under a fixed alternative distribution is seen to be normal. Since the variance of this normal distribution can be estimated from the data, there is an asymptotic confidence interval for the measure of distance from normality under alternative distributions that is inherent in the procedure. Furthermore, there is the option for a neighborhood-of-model validation procedure. The results of a simulation study, presented in Section 7, show that the novel test is strong with respect to prominent competitors. In Section 8, the new tests are applied to the Iris flower data set due to R.A. Fisher. The article concludes with some remarks. For the sake of readability, most of the proofs and some auxiliary results are deferred to Appendix A. Finally, the following abbreviations, valid for $t, x \in \mathbb{R}^d$, will be used in several sections:

$$\mathrm{CS}^+(t,x) := \cos(t^\top x) + \sin(t^\top x), \quad \mathrm{CS}^-(t,x) = \cos(t^\top x) - \sin(t^\top x). \tag{13}$$

2 | THE LIMITS $a \to \infty$ AND $a \to 0$

The results of this section show that the class of tests for multivariate normality based on $T_{n,a}$ is "closed at the boundaries" $a \to \infty$ and $a \to 0$ and thus shed some light on the tuning parameter a, which figures in the weight function w_a given in (6). We first consider the case $a \to \infty$.

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

$$\lim_{a\to\infty} \frac{2a^{d/2+1}}{n\pi^{\frac{d}{2}}} T_{n,a} = \frac{1}{n^2}\sum_{j,k=1}^{n} \|Y_{n,j}\|^2\|Y_{n,k}\|^2 Y_{n,j}^\top Y_{n,k}. \tag{14}$$

Proof. For short, put $Y_j := Y_{n,j}$. From the representation of $T_{n,a}$ we have

$$\begin{aligned}\frac{2a^{d/2+1}T_{n,a}}{n\pi^{d/2}} &= \frac{2a}{n^2}\sum_{j,k=1}^{n}\|Y_j\|^2\|Y_k\|^2\exp\left(-\frac{1}{4a}\|Y_j-Y_k\|^2\right)\\ &-\left(\frac{2a}{2a+1}\right)^{d/2+1}\frac{2}{n(2a+1)}\sum_{j=1}^{n}\|Y_j\|^2(\|Y_j\|^2+2da(2a+1))\exp\left(-\frac{1}{2}\frac{\|Y_j\|^2}{2a+1}\right)\\ &+\left(\frac{a}{a+1}\right)^{d/2+1}\frac{2}{a+1}\left(a(a+1)d^2+\frac{d(d+2)}{4}\right)\\ &=: U_{n,1}-U_{n,2}+U_{n,3}.\end{aligned}$$

Since $\sum_{j=1}^{n}\|Y_j\|^2 = nd$, an expansion of the exponential function yields

$$U_{n,1} = 2ad^2 - \frac{d}{n}\sum_{j=1}^{n}\|Y_j\|^4 + \frac{1}{n^2}\sum_{j,k=1}^{n}\|Y_j\|^2\|Y_k\|^2 Y_j^\top Y_k + o(1)$$

as $a\to\infty$. To tackle $U_{n,2}$, we use

$$\left(\frac{2a}{2a+1}\right)^{d/2+1} = \left(1+\frac{1}{2a}\right)^{-(d/2+1)} = 1-\left(\frac{d}{2}+1\right)\frac{1}{2a}+O(a^{-2})$$

and, after some algebra, obtain $U_{n,2} = 4ad^2-(d/2+1)2d^2-dn^{-1}\sum_{j=1}^{n}\|Y_j\|^4+o(1)$. Finally, a binomial expansion of $(a/(a+1))^{d/2+1}$ yields $U_{n,3} = 2ad^2-(d/2+1)2d^2+o(1)$. Summing up, the assertion follows. ∎

Remark 2. The limit $\tilde{b}_{1,d} := n^{-2}\sum_{j,k=1}^{n}\|Y_{n,j}\|^2\|Y_{n,k}\|^2Y_{n,j}^\top Y_{n,k}$ (say), which figures on the right-hand side of (14), is a measure of multivariate (sample) skewness, introduced by Móri, Rohatgi, and Székely (see Móri et al., 1993). A much older time-honoured measure of multivariate (sample) skewness is skewness in the sense of Mardia (see Mardia, 1970), which is given by $b_{1,d} := n^{-2}\sum_{j,k=1}^{n}(Y_{n,j}^\top Y_{n,k})^3$. It is interesting to compare Theorem 2 with similar results found in connection with other weighted L^2-statistics that have been studied for testing H_0. Thus, by theorem 2.1 of Henze (1997a), the time-honoured class of BHEP-statistics for testing for multivariate normality (see Henze and Wagner, 1997), after suitable rescaling, approaches the linear combination $2b_{1,d}+3\tilde{b}_{1,d}$, as a smoothing parameter (called β in that article) tends to 0. Since β and a are related by $\beta = a^{-1/2}$, this corresponds to letting a tend to infinity. The same linear combination $2b_{1,d}+3\tilde{b}_{1,d}$ also showed up as a limit statistic in Henze and Jiménez-Gamero (2019) and Henze et al. (2019). Notice that, in the univariate case, the limit statistic $\tilde{b}_{1,d}$ figuring in Theorem 2 is nothing but three times squared sample skewness. We stress 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 Baringhaus and Henze, 1991; Baringhaus and Henze, 1992; Henze, 1994a, 1994b, 1997b).

We now consider the case $a \to 0$. Since, elementwise on the underlying probability space, the expressions in (11) and (12) have finite limits as $a \to 0$, and since the double sum figuring in (10) converges to $\sum_{j=1}^{n} \|Y_{n,j}\|^4$ as $a \to 0$, we have the following result.

Theorem 3. *Elementwise on the underlying probability space, we have*

$$\lim_{a \to 0} \left(\frac{a}{\pi} \right)^{d/2} T_{n,a} = \frac{1}{n} \sum_{j=1}^{n} \|Y_{n,j}\|^4. \tag{15}$$

Remark 3. The limit statistic on the right-hand side of (15) is Mardia's celebrated measure $b_{2,d}$ of multivariate sample kurtosis (see Mardia, 1970). Together with Theorem 2, this result shows that, just like the class of BHEP tests for multivariate normality (see Henze, 1997a), also the class of tests based on $T_{n,a}$ is "closed at the boundaries" $a \to \infty$ and $a \to 0$. Notably, Mardia's measure of kurtosis shows up for the first time in connection with limits of weighted L^2-statistics for testing for multivariate normality. The corresponding limit statistic for the class of BHEP tests is, up to a linear transformation, $n^{-1} \sum_{j=1}^{n} \exp(-\|Y_{n,j}\|^2/2)$, see theorem 3.1 of Henze (1997a).

3 | A BASIC HILBERT SPACE CENTRAL LIMIT THEOREM

In this chapter, we present a basic Hilbert space central limit theorem. This theorem implies the limit distribution of $T_{n,a}$ under the null hypothesis (1), but it is also beneficial for proving a limit normal distribution of $T_{n,a}$ under fixed alternatives to H_0. Throughout this section, we assume that the underlying distribution satisfies $\mathbb{E}\|X\|^4 < \infty$. Moreover, in view of affine invariance of $T_{n,a}$, we may (and do) w.l.o.g. assume that $\mathbb{E}(X) = 0$ and $\mathbb{E}(XX^\top) = \mathrm{I}_d$, since both the finite-sample and the limit null distribution of $T_{n,a}$ do not depend on the mean and covariance matrix of the underlying normal distribution. To motivate the benefit of a Hilbert space setting and for later purposes, it will be convenient to represent $T_{n,a}$ in a different way.

Proposition 1. *Recall $\psi(t)$ from (4), and let*

$$m(t) := (d - \|t\|^2)\psi(t), \quad t \in \mathbb{R}^d, \tag{16}$$

$$Z_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{\|Y_{n,j}\|^2(\cos(t^\top Y_{n,j}) + \sin(t^\top Y_{n,j})) - m(t)\}, \quad t \in \mathbb{R}^d. \tag{17}$$

We then have

$$T_{n,a} = \int Z_n^2(t)\, w_a(t)\, \mathrm{d}t. \tag{18}$$

Proof. The proof follows by straightforward algebra using the addition theorems for the sine function and the cosine function and the fact that $\int \sin(t^\top y) m(t) w_a(t)\, \mathrm{d}t = 0, y \in \mathbb{R}^d$. ■

A convenient setting for asymptotics will be the separable Hilbert space $\mathbb{H}$ of (equivalence classes of) measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\int f^2(t) w_a(t)\, \mathrm{d}t < \infty$. The scalar product and the norm in $\mathbb{H}$ will be denoted by

$$\langle f, g\rangle_{\mathbb{H}} = \int f(t)g(t)\, w_a(t)\,\mathrm{d}t, \quad \|f\|_{\mathbb{H}} = \langle f, f\rangle_{\mathbb{H}}^{1/2}, \quad f, g \in \mathbb{H},$$

respectively. Notice that $T_{n,a} = \|Z_n\|_{\mathbb{H}}^2$. Recalling $\mathrm{CS}^{\pm}(t,x)$ from (13), and putting

$$\mu(t) := \mathbb{E}[\|X\|^2 \mathrm{CS}^+(t, X)], \quad t \in \mathbb{R}^d, \tag{19}$$

the main object of this section is the random element V_n of $\mathbb{H}$, defined by

$$V_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\|Y_{n,j}\|^2 \mathrm{CS}^+(t, Y_{n,j}) - \mu(t)), \quad t \in \mathbb{R}^d. \tag{20}$$

Observe that $V_n = Z_n$ if the distribution of X is $\mathrm{N}_d(0,\mathrm{I}_d)$, since then the functions μ and m coincide. We will show that, as $n \to \infty$, V_n converges in distribution to a centered Gaussian random element V of $\mathbb{H}$. The only technical problem in proving such a result is the fact that V_n is based on the scaled residuals $Y_{n,1}, \ldots, Y_{n,n}$ and not on $X_1, \ldots, X_n$. If $V_n^0(t)$ denotes the modification of $V_n(t)$ that results from replacing $Y_{n,j}$ with X_j, a Hilbert space central limit theorem holds for V_n^0, since the summands comprising $V_n^0(t)$ are i.i.d. square-integrable centered random elements of $\mathbb{H}$. Writing $\overset{\mathcal{D}}{\to}$ for convergence in distribution of random elements of $\mathbb{H}$ and random variables, the basic idea to prove $V_n \overset{\mathcal{D}}{\to} V$ is to find a random element $\tilde{V}_n$ of $\mathbb{H}$, such that $\tilde{V}_n \overset{\mathcal{D}}{\to} V$ and $\tilde{V}_n - V_n = o_{\mathbb{P}}(1)$ as $n \to \infty$. In what follows, the stochastic Landau symbol $o_{\mathbb{P}}(1)$ refers to convergence to zero in probability in $\mathbb{H}$, that is, we have to show

$$\|\tilde{V}_n - V_n\|_{\mathbb{H}}^2 = \int (\tilde{V}_n(t) - V_n(t))^2 w_a(t)\,\mathrm{d}t = o_{\mathbb{P}}(1) \quad \text{as } n \to \infty. \tag{21}$$

To state the main result of this section, let $\psi_X(t) = \mathbb{E}[\exp(\mathrm{i}t^\top X)]$, $t \in \mathbb{R}^d$, denote the characteristic function of X, and put

$$\psi_X^+(t) := \mathrm{Re}\ \psi_X(t) + \mathrm{Im}\ \psi_X(t), \quad \psi_X^-(t) := \mathrm{Re}\ \psi_X(t) - \mathrm{Im}\ \psi_X(t),$$

where $\mathrm{Re}\, w$ and $\mathrm{Im}\, w$ stand for the real and the imaginary part of a complex number w, respectively. For a twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, let $\mathrm{H}f(t)$ denote the Hessian matrix of f, evaluated at t. Furthermore, recall the gradient operator ∇ and the Laplace operator Δ from Section 1.

Proposition 2. *Let*

$$\tilde{V}_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} v(t, X_j), \quad t \in \mathbb{R}^d, \tag{22}$$

where

$$v(t,x) = v_1(t,x) + v_2(t,x) + v_3(t,x) + v_4(t,x), \tag{23}$$

$$v_1(t,x) = \|x\|^2 \mathrm{CS}^+(t,x),\ \ v_2(t,x) = \frac{1}{2} t^\top (xx^\top - \mathrm{I}_d)\nabla\Delta\psi_X^+(t), \tag{24}$$

$$v_3(t,x) = (2\nabla\psi_X^-(t) + \Delta\psi_X^-(t)t)^\top x,\ \ v_4(t,x) = x^\top \mathrm{H}\psi_X^+(t)x. \tag{25}$$

We then have (21).

The proof of Proposition 2 is given in Appendix A.

Since $\mathbb{E}(X) = 0$, $\mathbb{E}(XX^\top) = \mathrm{I}_d$ and $\mathbb{E}[\|X\|^2 \mathrm{CS}^+(t,X)] = -\Delta\psi_X^+(t)$, we have (writing tr for trace) $\mathbb{E}v(t,X) = -\Delta\psi_X^+(t) + \mathrm{tr}(H\psi_X^+(t)) = 0$. Thus, $v(\cdot,X_1),\dots,v(\cdot,X_n)$ are i.i.d. centered square-integrable random elements of $\mathbb{H}$, and the central limit theorem in Hilbert spaces gives $\tilde{V}_n \xrightarrow{\mathcal{D}} V$ for some centered Gaussian element V of $\mathbb{H}$. In view of (21) and Slutsky's lemma, we therefore can state the main result of this section.

Theorem 4. *Let $X,X_1,X_2,\dots$ be i.i.d. random vectors satisfying $\mathbb{E}\|X\|^4 < \infty$, $\mathbb{E}(X) = 0$ and $\mathbb{E}(XX^\top) = \mathrm{I}_d$. For the sequence of random elements V_n defined in (20) we have*

$$V_n \xrightarrow{\mathcal{D}} V \quad \text{as } n \to \infty,$$

where V is a centered Gaussian element of $\mathbb{H}$ having covariance kernel

$$L(s,t) = \mathbb{E}[v(s,X)v(t,X)], \quad s,t \in \mathbb{R}^d, \tag{26}$$

where $v(t,x)$ is given in (23).

4 | THE LIMIT NULL DISTRIBUTION OF $T_{n,a}$

In this section we derive the limit distribution of $T_{n,a}$ under the null hypothesis (1). In view of affine invariance, we assume that X has a d-variate standard normal distribution. Since the process $Z_n(t)$ given in (17) is nothing but V_n, as defined in (20), in this special case, we have the following result.

Theorem 5. *Suppose that X has some non-degenerate normal distribution. Putting $d_2 := d+2$, $d_4 := d+4$, we have the following:*

(a) There is a centered Gaussian random element Z of $\mathbb{H}$ with covariance kernel

$$\begin{aligned} K(s,t) = {} & \psi(s-t)((\|s-t\|^2 - d_2)^2 - 2d_2) + \psi(s)\psi(t)\Big\{ -\frac{(s^\top t)^2}{2}(\|s\|^2 - d_4)(\|t\|^2 - d_4) \\ & + 2d_2(\|s\|^2 + \|t\|^2) - \|s\|^4 - \|t\|^4 - \|s\|^2\|t\|^2 - s^\top t(\|s\|^2 - d_2)(\|t\|^2 - d_2) - dd_2 \Big\}, \end{aligned}$$

$s,t \in \mathbb{R}^d$, such that, with Z_n defined in (17), we have $Z_n \xrightarrow{\mathcal{D}} Z$ in $\mathbb{H}$ as $n \to \infty$.

(b) We have

$$T_{n,a} \xrightarrow{\mathcal{D}} \int Z^2(t)\, w_a(t)\, \mathrm{d}t.$$

Notice that (b) follows from (a) and the continuous mapping theorem. Part (a) follows from Theorem 4. For the special case $X \sim \mathrm{N}_d(0,\mathrm{I}_d)$, we have $\psi_X^+(t) = \psi_X^-(t) = \exp(-\|t\|^2/2) = \psi(t)$, which entails $\nabla\psi(t) = -t\psi(t)$, $\Delta\psi(t) = (\|t\|^2 - d)\psi(t)$, $\mathrm{H}\psi(t) = (tt^\top - \mathrm{I}_d)\psi(t)$, and $\nabla\Delta\psi(t) = t\psi(t)(2 + d - \|t\|^2)$. Thus, the function $v(t,x)$ figuring in the statement of Proposition 2 takes the special form

$$h(x,t) = \|x\|^2\mathrm{CS}^+(t,x) - (2\psi(t) + m(t))t^\top x - \psi(t)\|x\|^2 + \left(2\psi(t) + \frac{m(t)}{2}\right)(t^\top x)^2 - \left(\psi(t) + \frac{m(t)}{2}\right)\|t\|^2. \tag{27}$$

Long but straightforward computations, using symmetry arguments and the identities

$$\begin{aligned}
\mathbb{E}[\|X\|^2(s^\top X)^2] &= (d+2)\|s\|^2,\\
\mathbb{E}[\|X\|^4\cos(t^\top X)] &= ((d+2-\|t\|^2)(d-\|t\|^2) - 2\|t\|^2)\psi(t),\\
\mathbb{E}[\|X\|^2 s^\top X\sin(t^\top X)] &= s^\top t(d+2-\|t\|^2)\psi(t),\\
\mathbb{E}[\|X\|^2(s^\top X)^2\cos(t^\top X)] &= (d+2-\|t\|^2)(\|s\|^2 - (s^\top t)^2)\psi(t) - 2(s^\top t)^2\psi(t),
\end{aligned}$$

$s, t \in \mathbb{R}^d$, show that the covariance kernel $K(s,t) = \mathbb{E}[h(s,X)h(t,X)]$ takes the form given above.

Let $T_{\infty,a}$ be a random variable with the limit null distribution of $T_{n,a}$, that is, with the distribution of $\int Z^2(t)w_a(t)\,\mathrm{d}t$. Since $\mathbb{E}(T_{\infty,a}) = \int K(t,t)w_a(t)\,\mathrm{d}t$, the following result may be obtained by straightforward but tedious manipulations of integrals.

Theorem 6. *Putting* $c_j(a,d) := \pi^{d/2}d(a+1)^{-d/2-j}$, $j = 1,2,3,4$, *we have*

$$\begin{aligned}
\mathbb{E}(T_{\infty,a}) = d(d+2)&\left(\left(\frac{\pi}{a}\right)^{d/2} - \left(\frac{\pi}{a+1}\right)^{d/2}\right)\\
&- c_4(a,d)\frac{(d+2)(d+4)(d+6)}{32} + c_3(a,d)\frac{(d+2)(d+3)(d+4)}{8}\\
&- c_2(a,d)\frac{(d+2)(d^2+4d+14)}{8} + c_1(a,d)\frac{(d-2)(d+2)}{2}.
\end{aligned}$$

The quantiles of the distribution of $T_{\infty,a}$ can be approximated by a Monte Carlo method, see Section 7.

5 | CONTIGUOUS ALTERNATIVES

In this section, we consider a triangular array $(X_{n1}, \dots, X_{nn})$, $n \geq d+1$, of rowwise i.i.d. random vectors with Lebesgue density $f_n(x) = \varphi(x)(1 + g(x)/\sqrt{n})$, $x \in \mathbb{R}^d$, where $\varphi(x) = (2\pi)^{-d/2}\exp(-\|x\|^2/2)$, $x \in \mathbb{R}^d$, is the density of the standard normal distribution $\mathrm{N}_d(0,\mathrm{I}_d)$, and g is some bounded measurable function satisfying $\int g(x)\varphi(x)\,\mathrm{d}x = 0$. We assume that n is sufficiently large to render g nonnegative. Recall $Z_n(t)$ from (17).

Theorem 7. *Under the triangular* $X_{n,1}, \dots, X_{n,n}$ *given above, we have* $Z_n \xrightarrow{\mathcal{D}} Z + c$ as $n \to \infty$, *where* Z *is the centered random element of* $\mathbb{H}$ *figuring in Theorem 5, and* $c(t) = \int h(x,t)g(x)\varphi(x)\,\mathrm{d}x$, *with* $h(x,t)$ *given in (27).*

Proof. Since the reasoning uses standard LeCam theory on contiguous probability measures and closely parallels that given in section 3 of Henze and Wagner (1997), it will be omitted. ∎

Corollary 1. *Under the conditions of Theorem 7, we have* $T_{n,a} \xrightarrow{\mathcal{D}} \int (Z(t) + c(t))^2 w_a(t)\,\mathrm{d}t$.

From Theorem 7 and the above corollary, we conclude that the test for multivariate normality based on $T_{n,a}$ is able to detect alternatives which converge to the normal distribution at the rate $n^{-1/2}$, irrespective of the underlying dimension d. The test of Bowman and Foster (see Bowman and Foster, 1993) is a prominent example of an affine invariant tests for multivariate normality that is consistent against each fixed non-normal alternative distribution but nevertheless lacks this property of $n^{-1/2}$-consistency, see section 7 of Henze (2002).

6 | FIXED ALTERNATIVES AND CONSISTENCY

In this section we assume that $X, X_1, X_2, \ldots$ are i.i.d. with a distribution that is absolutely continuous with respect to Lebesgue measure, and that $\mathbb{E}\|X\|^4 < \infty$. In view of affine invariance, we make the additional assumptions $\mathbb{E}(X) = 0$ and $\mathbb{E}(XX^\top) = \mathrm{I}_d$. Recall $m(t)$ from (16) and $\mu(t)$ from (19). Writing $\xrightarrow{\text{a.s.}}$ for $\mathbb{P}$-almost sure convergence, our first result is a strong limit for $T_{n,a}/n$.

Theorem 8. *If* $\mathbb{E}\|X\|^4 < \infty$, *we have*

$$\frac{T_{n,a}}{n} \xrightarrow{\text{a.s.}} \Delta_a = \int (\mu(t) - m(t))^2 w_a(t)\,\mathrm{d}t \quad \text{as} \quad n \to \infty. \tag{28}$$

The proof of Theorem 8 is given in Appendix A.

Remark 4. Let $\psi_X(t) = \mathbb{E}[\exp(\mathrm{i}t^\top X)]$, $t \in \mathbb{R}^d$, denote the characteristic function of X. We have $\Delta\psi_X(t) = -\mathbb{E}[\|X\|^2 \exp(\mathrm{i}t^\top X)]$. Since $\Delta\psi(t) = (\|t\|^2 - d)\psi(t)$, some algebra gives

$$\Delta_a = \int |\Delta\psi_X(t) - \Delta\psi(t)|^2 w_a(t)\,\mathrm{d}t.$$

By Theorem 1, we have $\Delta_a = 0$ if and only if X has the normal distribution $\mathrm{N}_d(0,\mathrm{I}_d)$. In view of Theorem 8, $T_{n,a} \to \infty$ $\mathbb{P}$-almost surely under any alternative distribution satisfying $\mathbb{E}\|X\|^4 < \infty$. Since, according to Theorem 5, the sequence of critical values of a level-α-test of H_0 that rejects H_0 for large values of $T_{n,a}$ stays bounded, we have the following result.

Corollary 2. *The test for multivariate normality based on* $T_{n,a}$ *is consistent against any fixed alternative distribution satisfying* $\mathbb{E}\|X\|^4 < \infty$.

By analogy with Theorem 2, the next result shows that the (population) measure of multivariate skewness in the sense of Móri, Rohatgi, and Székely (see Móri et al., 1993) emerges as the limit of Δ_a, after a suitable normalization, as $a \to \infty$.

Theorem 9. *If* $\mathbb{E}\|X\|^6 < \infty$, *then*

$$\lim_{a\to\infty} \frac{2a^{d/2+1}}{\pi^{d/2}} \Delta_a = \|\mathbb{E}(\|X\|^2 X)\|^2.$$

The proof of Theorem 9 is given in Appendix A. We now show that the limit distribution of $\sqrt{n}(T_{n,a}/n - \Delta_a)$ as $n \to \infty$ is centered normal. This fact is essentially a consequence of theorem 1 in Baringhaus, Ebner, and Henze (2017). The reasoning is as follows: By (18), we have $T_{n,a} = \|Z_n\|^2_{\mathbb{H}}$, where Z_n is given in (17). Putting $z(t) := \mu(t) - m(t)$, $t \in \mathbb{R}^d$, display (28) shows that $\Delta_a = \|z\|^2_{\mathbb{H}}$. Now, defining $Z_n^*(t) := n^{-1/2}Z_n(t)$, it follows that

$$\begin{aligned}\sqrt{n}\left(\frac{T_{n,a}}{n} - \Delta_a\right) &= \sqrt{n}(\|Z_n^*\|^2_{\mathbb{H}} - \|z\|^2_{\mathbb{H}}) = \sqrt{n}\langle Z_n^* - z, Z_n^* + z\rangle_{\mathbb{H}} \\ &= \sqrt{n}\langle Z_n^* - z, 2z + Z_n^* - z\rangle_{\mathbb{H}} \\ &= 2\langle \sqrt{n}(Z_n^* - z), z\rangle_{\mathbb{H}} + \frac{1}{\sqrt{n}}\|\sqrt{n}(Z_n^* - z)\|^2_{\mathbb{H}}. \end{aligned} \tag{29}$$

A little thought shows that $\sqrt{n}(Z_n^*(t) - z(t)) = V_n(t)$, where V_n is given in (20). By Theorem 4, $V_n \xrightarrow{\mathcal{D}} V$ for a centered Gaussian element V of $\mathbb{H}$. As a consequence, the second summand in (29) is $o_{\mathbb{P}}(1)$ as $n \to \infty$, and, by the continuous mapping theorem, the first summand converges in distribution to $2\langle V, z\rangle_{\mathbb{H}}$. The latter random variable has a centered normal distribution with variance $\sigma_a^2 := 4\mathbb{E}[\langle V, z\rangle^2_{\mathbb{H}}]$. The following theorem summarizes our findings.

Theorem 10. *For a fixed alternative distribution satisfying* $\mathbb{E}\|X\|^4 < \infty$, $\mathbb{E}(X) = 0$ *and* $\mathbb{E}(XX^\top) = \mathrm{I}_d$, *we have*

$$\sqrt{n}\left(\frac{T_{n,a}}{n} - \Delta_a\right) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma_a^2) \qquad \text{as } n \to \infty,$$

where

$$\sigma_a^2 = 4\iint L(s,t)z(s)z(t)\, w_a(s)w_a(t)\, \mathrm{d}s\mathrm{d}t \tag{30}$$

and $L(s,t)$ *is given in (26).*

Proof. To complete the proof, notice that, by Fubini's theorem,

$$\begin{aligned}\sigma_a^2 = 4\mathbb{E}[\langle V, z\rangle^2_{\mathbb{H}}] &= 4\mathbb{E}\left[\left(\int V(s)\, z(s)\, w_a(s)\, \mathrm{d}s\right)\left(\int V(t)\, z(t)\, w_a(t)\, \mathrm{d}t\right)\right] \\ &= 4\iint \mathbb{E}[V(s)V(t)]\, z(s)\, z(t)\, w_a(s)w_a(t)\, \mathrm{d}s\mathrm{d}t. \end{aligned}$$

■

Under slightly stronger conditions on $\mathbb{P}^X$, there is a consistent estimator of σ_a^2. To obtain such an estimator, we replace $L(s,t)$ as well as $z(s)$ and $z(t)$ figuring in (30) with suitable empirical counterparts. To this end, notice that, by (26), we have

$$L(s,t) = \sum_{i=1}^{4}\sum_{j=1}^{4} L^{i,j}(s,t), \tag{31}$$

where

$$L^{i,j}(s,t) = \mathbb{E}[v_i(s,X)v_j(t,X)] \tag{32}$$

and $v_j(t,x)$, $j \in \{1,2,3,4\}$, are given in (24), (25). Since $\nabla\Delta\psi_X^{\pm}(t) = \mp\mathbb{E}[\mathrm{CS}^{\mp}(t,X)\|X\|^2X]$, $\nabla\psi_X^{\pm}(t) = \pm\mathbb{E}[\mathrm{CS}^{\mp}(t,X)X]$, $\Delta\psi_X^{\pm}(t) = -\mathbb{E}[\mathrm{CS}^{\pm}(t,X)\|X\|^2]$ and $\mathrm{H}\psi_X^{+}(t) = -\mathbb{E}[\mathrm{CS}^{+}(t)(t,X)\,XX^\top]$, parts a) – d) of the following lemma show that the unknown quantities $\nabla\Delta\psi_X^{+}(t)$, $\nabla\psi_X^{-}(t)$, $\Delta\psi_X^{-}(t)$ and $\mathrm{H}\psi_X^{+}(t)$ that figure in the expressions of $v_2(t,x)$, $v_3(t,x)$ and $v_4(t,x)$ can be replaced with consistent estimators that are based on the scaled residuals $Y_{n,1},\dots,Y_{n,n}$ defined in (5).

Lemma 1. *If $\mathbb{E}\|X_1\|^6 < \infty$, we have*

(a) $\Psi_{1,n}(t) := n^{-1}\sum_{j=1}^n \mathrm{CS}^{+}(t,Y_{n,j})Y_{n,j} \xrightarrow{\text{a.s.}} -\nabla\psi_X^{-}(t)$,
(b) $\Psi_{2,n}(t) := n^{-1}\sum_{j=1}^n \mathrm{CS}^{+}(t,Y_{n,j})Y_{n,j}Y_{n,j}^\top \xrightarrow{\text{a.s.}} -\mathrm{H}\psi_X^{+}(t)$,
(c) $\Psi_{3,n}^{\pm}(t) := n^{-1}\sum_{j=1}^n \mathrm{CS}^{\pm}(t,Y_{n,j})\|Y_{n,j}\|^2 \xrightarrow{\text{a.s.}} -\Delta\psi_X^{\pm}(t)$,
(d) $\Psi_{4,n}^{\pm}(t) := n^{-1}\sum_{j=1}^n \mathrm{CS}^{\pm}(t,Y_{n,j})\|Y_{n,j}\|^2Y_{n,j} \xrightarrow{\text{a.s.}} \pm\nabla\Delta\psi_X^{\mp}(t)$,
(e) $\Psi_{5,n}(t) := n^{-1}\sum_{j=1}^n \mathrm{CS}^{+}(t,Y_{n,j})\|Y_{n,j}\|^2Y_{n,j}Y_{n,j}^\top \xrightarrow{\text{a.s.}} \mathrm{H}\Delta\psi_X^{+}(t)$.

The proof of Lemma 1 is given in Appendix A. In view of Lemma 1, a suitable estimator of $L(s,t)$ defined in (31) is

$$L_n(s,t) = \sum_{i=1}^4\sum_{j=1}^4 L_n^{i,j}(s,t), \tag{33}$$

where

$$L_n^{i,j}(s,t) = \frac{1}{n}\sum_{k=1}^n v_{n,i}(s,Y_{n,k})v_{n,j}(t,Y_{n,k}), \tag{34}$$

and

$$v_{n,1}(s,Y_{n,k}) = \|Y_{n,k}\|^2\mathrm{CS}^{+}(s,Y_{n,k}), \quad v_{n,2}(s,Y_{n,k}) = -\frac{1}{2}s^\top(Y_{n,k}Y_{n,k}^\top - \mathrm{I}_d)\Psi_{4,n}^{-}(s),$$
$$v_{n,3}(s,Y_{n,k}) = -(2\Psi_{1,n}(s) + \Psi_{3,n}^{-}(s)s)^\top Y_{n,k}, \quad v_{n,4}(s,Y_{n,k}) = -Y_{n,k}^\top\Psi_{2,n}(s)Y_{n,k}.$$

By straightforward algebra we have

$$L_n^{1,1}(s,t) = n^{-1}\sum_{j=1}^n\|Y_{n,j}\|^4\cos((t-s)^\top Y_{n,j}) + n^{-1}\sum_{j=1}^n\|Y_{n,j}\|^4\sin((t+s)^\top Y_{n,j}),$$
$$L_n^{1,2}(s,t) = -\frac{1}{2}\Psi_{4,n}^{-}(s)^\top\Psi_{5,n}(t)s + \frac{1}{2}\Psi_{3,n}^{+}(t)\Psi_{4,n}^{-}(s)^\top s,$$
$$L_n^{1,3}(s,t) = (-2\Psi_{1,n}(s) - \Psi_{3,n}^{-}(s)s)^\top\Psi_{4,n}^{+}(t)$$
$$L_n^{1,4}(s,t) = -n^{-1}\sum_{j=1}^n\|Y_{n,j}\|^2\mathrm{CS}^{+}(t,Y_{n,j})Y_{n,j}^\top\Psi_{2,n}(s)Y_{n,j}$$

$$L_n^{2,2}(s,t) = \frac{1}{4}\Psi_{4,n}^{-}(t)^\top n^{-1}\sum_{j=1}^{n} Y_{n,j}Y_{n,j}^\top t\Psi_{4,n}^{-}(s)^\top (Y_{n,j}Y_{n,j}^\top - \mathrm{I}_d)s,$$

$$L_n^{2,3}(s,t) = \Psi_{4,n}^{-}(t)^\top n^{-1}\sum_{j=1}^{n} Y_{n,j}Y_{n,j}^\top t\left(\Psi_{1,n}(s) + \frac{1}{2}\Psi_{3,n}^{-}(s)s\right)^\top Y_{n,j}, \tag{35}$$

as well as

$$L_n^{2,4}(s,t) = \frac{1}{2}\Psi_{4,n}^{-}(t)^\top n^{-1}\sum_{j=1}^{n}(Y_{n,j}Y_{n,j}^\top - \mathrm{I}_d)tY_{n,j}^\top \Psi_{2,n}(s)Y_{n,j},$$

$$L_n^{3,3}(s,t) = (2\Psi_{1,n}(t) + \Psi_{3,n}^{-}(t)t)^\top (2\Psi_{1,n}(s) + \Psi_{3,n}^{-}(s)s),$$

$$L_n^{3,4}(s,t) = n^{-1}\sum_{j=1}^{n}(2\Psi_{1,n}(s) + \Psi_{3,n}^{-}(s)s)^\top Y_{n,j}Y_{n,j}^\top \Psi_{2,n}(t)Y_{n,j},$$

$$L_n^{4,4}(s,t) = n^{-1}\sum_{j=1}^{n} Y_{n,j}^\top \Psi_{2,n}(t)Y_{n,j}Y_{n,j}^\top \Psi_{2,n}(s)Y_{n,j}.$$

Notice that, by symmetry, $L_n^{i,j} = L_n^{j,i}$ if $i \neq j$. Since $z(s) = \mu(s) - m(s) = \mathbb{E}[\|X\|^2 \mathrm{CS}^+(s,X)] - m(s)$, a natural estimator of $z(s)$ is

$$z_n(s) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{CS}^+(s, Y_{n,k})\|Y_{n,k}\|^2 - m(s). \tag{36}$$

Writing $\xrightarrow{\mathbb{P}}$ for convergence in probability, we then have the following result.

Theorem 11. *Suppose* $\mathbb{E}\|X\|^6 < \infty$, $\mathbb{E}(X) = 0$ *and* $\mathbb{E}(XX^\top) = \mathrm{I}_d$. *Let*

$$\hat{\sigma}_{n,a}^2 = 4\iint L_n(s,t)\, z_n(s)\, z_n(t)\, w_a(s)w_a(t)\, \mathrm{d}s\mathrm{d}t, \tag{37}$$

where $L_n(s,t)$ *and* $z_n(s)$ *are defined in (33) and (36), respectively. We then have* $\hat{\sigma}_{n,a}^2 \xrightarrow{\mathbb{P}} \sigma_a^2$.

The proof of Theorem 11 is given in Appendix A. Under the conditions of Theorem 11, Theorem 10 and Sluzki's lemma yield

$$\frac{\sqrt{n}}{\hat{\sigma}_{n,a}}\left(\frac{T_{n,a}}{n} - \Delta_a\right) \xrightarrow{\mathcal{D}} \mathrm{N}(0,1) \qquad \text{as } n \to \infty, \tag{38}$$

provided that $\sigma_a^2 > 0$. We thus obtain the following asymptotic confidence interval for Δ_a.

Corollary 3. *Let* $\alpha \in (0,1)$, *and write* $\Phi_{1-\alpha/2}$ *for the* $(1-\alpha/2)$*-quantile of the standard normal law. Then*

$$I_{n,a,\alpha} := \left[\frac{T_{n,a}}{n} - \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}\Phi_{1-\alpha/2}, \frac{T_{n,a}}{n} + \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}\Phi_{1-\alpha/2}\right]$$

is an asymptotic confidence interval for Δ_a.

TABLE 1 Percentages of coverage of $\Delta_{0.1}$ by $I_{n,0.1,\alpha}$ for different distributions (5,000 replications)

n	$\mathrm{U}(-\sqrt{3}, \sqrt{3})$	$\mathrm{L}(0, 1/\sqrt{2})$	$\mathrm{Lo}(0, \sqrt{3}/\pi)$
20	91.5	87.5	83.4
50	93.6	96.9	95.2
100	94.4	97.8	98.4
200	94.8	97.9	99.0
300	94.5	97.5	98.9

Example 1. We consider the case $d = 1$, $a = 0.1$ and the following standardized symmetric alternative distributions: the uniform distribution $\mathrm{U}(-\sqrt{3}, \sqrt{3})$, the Laplace distribution $\mathrm{L}(0, 1/\sqrt{2})$, and the logistic distribution $\mathrm{Lo}(0, \sqrt{3}/\pi)$. The corresponding characteristic functions and their second derivatives are given by

$$\varphi_{\mathrm{U}}(t) = \frac{\sin(\sqrt{3}t)}{\sqrt{3}t}, \quad \varphi''_{\mathrm{U}}(t) = \frac{\sqrt{3}(2-3t^2)\sin(\sqrt{3}t) - 6t\cos(\sqrt{3}t)}{3t^3},$$
$$\varphi_{\mathrm{L}}(t) = \frac{2}{2+t^2}, \quad \varphi''_{\mathrm{L}}(t) = \frac{12t^2-8}{(2+t^2)^3},$$
$$\varphi_{\mathrm{Lo}}(t) = \frac{\sqrt{3}t}{\sinh(\sqrt{3}t)}, \quad \varphi''_{\mathrm{Lo}}(t) = \frac{3}{2}\frac{3\sqrt{3}t - 2\sinh(2\sqrt{3}t) + \sqrt{3}t\cosh(2\sqrt{3}t)}{\sinh(\sqrt{3}t)^3}.$$

The pertaining values of $\Delta_{0.1}$ are $\Delta_{0.1,\mathrm{U}} \approx 0.3322$, $\Delta_{0.1,\mathrm{L}} \approx 0.127$ and $\Delta_{0.1,\mathrm{Lo}} \approx 0.033$. Table 1 shows the percentages of coverage of the confidence intervals $I_{n,0.1,\alpha}$ for $\alpha = .05$ and several sample sizes. Each entry is based on 5,000 Monte-Carlo-replications. The results are quite satisfactory for $n \geq 50$.

Remark 5. A further application of (38) addresses a fundamental drawback inherent in any goodness of fit test (see Baringhaus et al., 2017). If a level-α-test of H_0 does not reject H_0, the hypothesis H_0 is by no means *validated* or *confirmed*, since each model is wrong, and there is probably only not enough evidence to reject H_0. However, if we define a certain "essential distance" $\delta_0 > 0$, we can consider the inverse testing problem $H_{\delta_0} : \Delta_a(F) \geq \delta_0$ against $K_{\delta_0} : \Delta_a(F) < \delta_0$. Here, the dependence of Δ_a on the underlying distribution has been made explicit.

From (38), we obtain an asymptotic level-α-*neighborhood-of-model-validation* test of H_{δ_0} against K_{δ_0}, which rejects H_{δ_0} if and only if $n^{-1}T_{n,a} \leq \delta_0 - \hat{\sigma}_{n,a}\Phi^{-1}(1-\alpha)/\sqrt{n}$. Indeed, from (38) we have for each $F \in H_{\delta_0}$

$$\limsup_{n\to\infty} \mathbb{P}_F\left(\frac{T_{n,a}}{n} \leq \delta_0 - \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right) = \limsup_{n\to\infty} \mathbb{P}_F\left(\frac{\sqrt{n}}{\hat{\sigma}_{n,a}}\left(\frac{T_{n,a}}{n} - \delta_0\right) \leq -\Phi^{-1}(1-\alpha)\right) \leq \alpha.$$

Moreover, it follows that

$$\lim_{n\to\infty} \mathbb{P}_F\left(\frac{T_{n,a}}{n} \leq \delta_0 - \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right) = \alpha$$

for each F such that $\Delta_a(F) = \delta_0$. Finally, we have

$$\lim_{n\to\infty} \mathbb{P}_F\left(\frac{T_{n,a}}{n} \leq \delta_0 - \frac{\hat{\sigma}_{n,a}}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right) = 1$$

if $\Delta_a(F) < \delta_0$. Thus, the test is consistent against each alternative.

Remark 6. The double integral figuring (37) may be calculated explicitly. To this end, notice that $\hat{\sigma}_{n,a}^2 = \sum_{i,j=1}^4 \hat{\sigma}_{n,a}^{i,j}$, where

$$\hat{\sigma}_{n,a}^{i,j} = 4\iint L_n^{i,j}(s,t) z_n(s) z_n(t) w_a(s) w_a(t)\,\mathrm{d}s\mathrm{d}t \tag{39}$$

and $\hat{\sigma}_{n,a}^{i,j} = \hat{\sigma}_{n,a}^{j,i}$, $i,j \in \{1,\dots,4\}$, by symmetry. We put

$$\begin{aligned}
q_{1,a}(y) &:= \int m(t)\mathrm{CS}^+(t,y)w_a(t)\mathrm{d}t = \frac{(2\pi)^{d/2}}{(2a+1)^{d/2+2}}(\|y\|^2 + 2da(2a+1))\exp\left(-\frac{1}{2}\frac{\|y\|^2}{2a+1}\right),\\
p_{1,a}(y,z) &:= \int \mathrm{CS}^+(t,y)\mathrm{CS}^+(t,z)w_a(t)\mathrm{d}t = \left(\frac{\pi}{a}\right)^{d/2}\exp\left(-\frac{\|y-z\|^2}{4a}\right),\\
p_{2,a}(y,z) &:= \int \mathrm{CS}^+(t,y)\mathrm{CS}^-(t,z)tw_a(t)\mathrm{d}t = \left(\frac{\pi}{a}\right)^{d/2}\frac{1}{2a}\exp\left(-\frac{\|y-z\|^2}{4a}\right)(y-z),\\
q_{2,a}(y) &:= \int m(t)\mathrm{CS}^-(t,y)tw_a(t)\mathrm{d}t\\
&= \frac{(2\pi)^{d/2}}{(2a+1)^{d/2+3}}(2(2a+1)(1-ad) - \|y\|^2)\exp\left(-\frac{1}{2}\frac{\|y\|^2}{2a+1}\right)y
\end{aligned}$$

for $y, z \in \mathbb{R}^d$ and

$$\begin{aligned}
P_n^{1,a,1} &:= \frac{1}{n^2}\sum_{j,k=1}^n \|Y_j\|^2\|Y_k\|^2 p_{1,a}(Y_j,Y_k) - \frac{1}{n}\sum_{j=1}^n \|Y_j\|^2 q_{1,a}(Y_j),\\
\tilde{P}_n^{1,a,1} &:= \frac{1}{n^2}\sum_{j,k=1}^n \|Y_j\|^2\|Y_k\|^2 p_{1,a}(Y_j,Y_k)Y_k - \frac{1}{n}\sum_{j=1}^n \|Y_j\|^2 q_{1,a}(Y_j)Y_j,\\
\tilde{P}_n^{1,a,2} &:= \frac{1}{n^2}\sum_{j,k=1}^n \|Y_j\|^2 p_{1,a}(Y_j,Y_k)Y_k - \frac{1}{n}\sum_{j=1}^n q_{1,a}(Y_j)Y_j,\\
\overline{P}_n^{1,a} &:= \frac{1}{n^2}\sum_{j,k=1}^n \|Y_j\|^2\|Y_k\|^2 p_{1,a}(Y_j,Y_k)Y_kY_k^\top - \frac{1}{n}\sum_{j=1}^n \|Y_j\|^2 Y_jY_j^\top q_{1,a}(Y_j),\\
P_n^{1,a,2}(Y_j) &:= \frac{1}{n^2}\sum_{k,\ell=1}^n \|Y_k\|^2 (Y_j^\top Y_\ell)^2 p_{1,a}(Y_k,Y_l) - \frac{1}{n}\sum_{k=1}^n (Y_j^\top Y_k)^2 q_{1,a}(Y_k),\\
P_n^{1,a,3}(Y_j) &:= \frac{1}{n}\sum_{k=1}^n \|Y_k\|^2 p_{1,a}(Y_j,Y_k) - q_{1,a}(Y_j),\\
P_n^{2,a,1} &:= \frac{1}{n^2}\sum_{j,k=1}^n \|Y_j\|^2\|Y_k\|^2 Y_k^\top p_{2,a}(Y_j,Y_k) - \frac{1}{n}\sum_{j=1}^n \|Y_j\|^2 Y_j^\top q_{2,a}(Y_j),
\end{aligned}$$

$$
\begin{aligned}
P_n^{2,a,2} &:= \frac{1}{n^2}\sum_{j,k=1}^{n}\|Y_j\|^2\|Y_k\|^2 Y_k^\top \overline{P}_{1,a} p_{2,a}(Y_j,Y_k) - \frac{1}{n}\sum_{j=1}^{n}\|Y_j\|^2 Y_j^\top \overline{P}_{1,a} q_{2,a}(Y_j),\\
\tilde{P}_n^{2,a} &:= \frac{1}{n^2}\sum_{j,k=1}^{n}\|Y_j\|^2\|Y_k\|^2 p_{2,a}(Y_j,Y_k) - \frac{1}{n}\sum_{j=1}^{n}\|Y_j\|^2 q_{2,a}(Y_j),\\
P_n^{2,a,3}(Y_j) &:= \frac{1}{n^2}\sum_{k,\ell=1}^{n}\|Y_k\|^2\|Y_\ell\|^2 Y_k^\top Y_j Y_j^\top p_{2,a}(Y_k,Y_\ell) - \frac{1}{n}\sum_{k=1}^{n}\|Y_k\|^2 Y_k^\top Y_j Y_j^\top q_{2,a}(Y_k),
\end{aligned}
$$

where Y_j is shorthand for $Y_{n,j}$. Notice that $\tilde{P}_n^{1,a,1}$, $\tilde{P}_n^{1,a,2}$, and $\tilde{P}_n^{2,a}$ are vectors and $\overline{P}_n^{1,a}$ is a matrix. By straightforward but tedious manipulations of the integrals in (39), each $\hat{\sigma}_n^{i,j}$ is seen to be an arithmetic mean of functions of the scaled residuals, namely,

$$
\begin{aligned}
\hat{\sigma}_{n,a}^{1,1} &= \frac{4}{n}\sum_{j=1}^{n}\|Y_j\|^4 P_n^{1,a,3}(Y_j)^2, \quad \hat{\sigma}_{n,a}^{1,2} = 2P_n^{1,a,1}P_n^{2,a,1} - 2P_n^{2,a,2},\\
\hat{\sigma}_{n,a}^{1,3} &= -4(2\tilde{P}_n^{1,a,2} + \tilde{P}_n^{2,a})^\top \tilde{P}_n^{1,a,1}, \quad \hat{\sigma}_{n,a}^{1,4} = -\frac{4}{n}\sum_{j=1}^{n}\|Y_j\|^2 P_n^{1,a,2}(Y_j)P_n^{1,a,3}(Y_j),\\
\hat{\sigma}_{n,a}^{2,2} &= \frac{1}{n}\sum_{j=1}^{n}P_{2,a,3}(Y_j)^2 - (P_n^{2,a,1})^2, \quad \hat{\sigma}_{n,a}^{2,3} = \frac{4}{n}\sum_{j=1}^{n}P_{2,a,3}(Y_j)\left(\tilde{P}_n^{1,a,2} + \frac{1}{2}\tilde{P}_n^{2,a}\right)^\top Y_j,\\
\hat{\sigma}_{n,a}^{2,4} &= \frac{2}{n}\sum_{j=1}^{n}(P_n^{2,a,3}(Y_j) - P_n^{2,a,1})P_n^{1,a,2}(Y_j), \quad \hat{\sigma}_{n,a}^{3,3} = 4\cdot\|2\tilde{P}_n^{1,a,2} + \tilde{P}_n^{2,a}\|^2,\\
\hat{\sigma}_{n,a}^{3,4} &= \frac{4}{n}\sum_{j=1}^{n}P_n^{1,a,2}(Y_j)(2\tilde{P}_n^{1,a,2} + \tilde{P}_n^{2,a})^\top Y_j, \quad \hat{\sigma}_{n,a}^{4,4} = \frac{4}{n}\sum_{j=1}^{n}P_n^{1,a,2}(Y_j)^2.
\end{aligned}
$$

7 | SIMULATIONS

In this section, we present the results of a Monte Carlo simulation study on the finite-sample power performance of the test based on $T_{n,a}$ with that of several competitors. Since different procedures have been used for univariate and multivariate data, we distinguish the cases $d = 1$ and $d \geq 2$. In the univariate case, the sample sizes are $n \in \{20,50,100\}$, and in the multivariate case we consider $n \in \{20,50,100,200\}$. The nominal level of significance is fixed throughout all simulations to 0.05. Critical values for $T_{n,a}$ (in fact, for a scaled version of $T_{n,a}$ in order to obtain values of a similar magnitude across the range of values for d and a considered) have been simulated under H_0 with 100,000 replications (see Table 2). The critical values in the rows in Table 2 denoted by "∞" represent approximations of quantiles of the distribution of the limit random element $T_{\infty,a} = \int Z^2(t)\, w_a(t)\, \mathrm{d}t$ in Theorem 5(b). To obtain these values we simulate (say) m i.i.d. random supporting points $U_1, \dots, U_m$ where $U_1 \sim \mathrm{N}_d(0,(2a)^{-1}\mathrm{I}_d)$, which are adapted to the integration over $\mathbb{R}^d$ with respect to the weight function $w_a(t) = \exp(-a\|t\|^2)$ and a large number (say) ℓ of random variables $Z_j = \|X_j\|^2/d^2m, j = 1, \dots, \ell$, with i.i.d. $X_j \sim \mathrm{N}_m(0,\Sigma_K)$, where the covariance matrix Σ_k is given by $\Sigma_K = (K(U_{k_1}, U_{k_2}))_{k_1,k_2\in\{1,\dots,m\}}$ and K is the covariance kernel in Theorem 5(a). Next, we calculate the empirical 95% quantile of $Z_1, \dots, Z_\ell$. Each approximation was simulated with $\ell = 100{,}000$ and $m = 1{,}000$ for $d \in \{2,3,5,10\}$ and each entry in Tables 3–6, which exhibit percentages of rejection of H_0 of the tests under consideration against various alternative distributions,

TABLE 2 Empirical and approximate quantiles for $d^{-2}(a/\pi)^{d/2}T_{n,a}$ and $\alpha = .05$ (100,000 replications)

d	$n\setminus a$	0.1	0.25	0.5	0.75	1	2	3	5	10
1	20	3.489	2.754	2.099	1.777	1.603	1.337	1.190	0.971	0.643
	50	3.651	2.953	2.291	1.954	1.756	1.445	1.294	1.064	0.721
	100	3.662	2.967	2.336	1.995	1.797	1.482	1.329	1.104	0.753
	200	3.616	2.927	2.301	1.978	1.795	1.486	1.330	1.106	0.759
	∞	3.518	2.839	2.348	1.915	1.765	1.495	1.289	1.151	0.736
2	20	2.207	1.916	1.482	1.172	0.957	0.622	0.529	0.434	0.298
	50	2.263	1.978	1.551	1.255	1.039	0.689	0.592	0.494	0.349
	100	2.245	1.963	1.543	1.241	1.044	0.707	0.608	0.514	0.364
	200	2.220	1.941	1.519	1.217	1.024	0.698	0.605	0.512	0.368
	∞	2.235	2.018	1.489	1.187	1.013	0.674	0.598	0.491	0.362
3	20	1.769	1.610	1.304	1.042	0.842	0.462	0.352	0.277	0.193
	50	1.814	1.662	1.360	1.105	0.908	0.525	0.407	0.325	0.234
	100	1.799	1.649	1.347	1.092	0.899	0.525	0.412	0.334	0.244
	200	1.777	1.636	1.330	1.074	0.883	0.518	0.409	0.336	0.247
	∞	1.777	1.617	1.314	1.049	0.857	0.505	0.408	0.348	0.243
5	20	1.407	1.360	1.201	1.011	0.835	0.417	0.263	0.169	0.110
	50	1.471	1.420	1.260	1.073	0.902	0.482	0.315	0.210	0.144
	100	1.469	1.420	1.259	1.072	0.901	0.485	0.321	0.218	0.154
	∞	1.498	1.437	1.260	1.028	0.862	0.485	0.312	0.221	0.161
10	20	1.130	1.129	1.108	1.049	0.959	0.569	0.336	0.153	0.060
	50	1.207	1.205	1.181	1.121	1.032	0.652	0.409	0.201	0.090
	100	1.221	1.219	1.194	1.132	1.043	0.664	0.422	0.211	0.098
	200	1.223	1.220	1.195	1.133	1.044	0.665	0.424	0.214	0.102
	∞	1.289	1.284	1.254	1.194	1.061	0.696	0.427	0.215	0.106

is based on 10,000 replications. Entries that are marked with $*$ denote 100% and the tests with the highest rejection rate are highlighted for easy reference. The simulations have been performed using the statistical computing environment R (see R Core Team, 2019).

7.1 | Testing univariate normality

For testing the hypothesis H_0 that the distribution of X is univariate normal (i.e., $\mathbb{P}^X \in \mathcal{N}_1$), we considered the following competitors to the new test statistic:

1. the Shapiro–Wilk test (SW), see Shapiro and Wilk (1965),
2. the Shapiro–Francia test (SF), see Shapiro and Francia (1972),

TABLE 3 Empirical power of $T_{n,a}$ against competitors ($d = 1$, $\alpha = .05$, 10,000 replications)

		$T_{n,a}$														
Alt.	$n \setminus a$	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**BE**	**SW**	**BCMR**	**BHEP**	**AD**	**SF**
N(0,1)	20	5	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	5
	100	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
NMix(0.3,1,0.25)	20	12	14	15	17	18	20	19	19	18	23	28	28	27	**30**	25
	50	28	33	38	42	44	46	45	43	40	54	60	60	62	**68**	57
	100	61	67	72	74	76	78	76	73	69	85	89	89	90	**94**	88
NMix(0.5,1,4)	20	36	37	38	38	38	37	35	33	32	36	40	43	42	46	**48**
	50	65	67	69	69	69	65	60	54	48	64	78	80	80	**86**	83
	100	90	93	94	94	94	92	88	79	68	92	97	98	98	**99**	98
t_3	20	**40**	**40**	**40**	**40**	39	38	37	37	36	34	35	37	34	33	**40**
	50	68	**69**	**69**	68	68	66	64	61	58	57	64	65	61	60	**69**
	100	87	89	89	89	88	87	86	83	78	81	88	89	86	85	**91**
t_5	20	**23**	**23**	**23**	22	22	21	21	21	21	20	19	20	18	17	22
	50	**41**	**41**	**41**	**41**	40	38	37	35	34	31	35	37	32	31	**41**
	100	58	61	61	61	60	57	55	51	47	46	56	58	50	48	**63**
t_{10}	20	**12**	**12**	**12**	**12**	**12**	11	11	11	11	11	10	11	9	9	**12**
	50	19	19	19	19	18	17	17	16	16	14	16	17	13	12	**20**
	100	26	27	**28**	**28**	27	25	24	23	21	19	22	24	16	15	**28**
$U(-\sqrt{3}, \sqrt{3})$	20	1	1	1	1	1	2	1	1	1	3	**21**	17	13	17	8
	50	40	25	14	13	12	5	2	1	0	8	**75**	70	55	58	47
	100	98	96	89	83	80	59	24	3	1	45	*	99	95	95	97

(Continues)

TABLE 3 (Continued)

		$T_{n,a}$														
Alt.	$n \setminus a$	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**BE**	**SW**	**BCMR**	**BHEP**	**AD**	**SF**
χ^2_5	20	33	35	37	38	40	42	42	42	41	**44**	**44**	**44**	42	38	42
	50	75	79	82	83	84	85	85	85	84	87	**89**	88	84	80	85
	100	99	99	99	*	*	*	*	99	99	*	*	*	99	99	*
χ^2_{15}	20	16	16	17	18	18	**19**	**19**	**19**	**19**	**19**	18	18	17	16	18
	50	33	37	40	42	43	45	45	**46**	45	45	42	42	39	33	40
	100	60	67	71	73	74	76	76	**77**	**77**	**77**	75	74	68	61	71
$\Gamma(1,5)$	20	60	64	67	70	71	73	73	72	71	77	**83**	**83**	77	77	80
	50	99	99	99	99	*	*	*	99	99	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\Gamma(5,1)$	20	19	21	22	23	24	24	**25**	**25**	**25**	**25**	24	24	23	20	24
	50	45	50	53	55	57	59	59	60	59	**61**	59	59	55	49	56
	100	79	85	88	88	89	90	90	90	90	**91**	90	90	85	81	88
W(1,0.5)	20	61	65	68	70	72	74	74	73	72	*	84	83	78	77	80
	50	99	99	99	99	99	*	*	99	99	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Gum(1,2)	27	29	30	31	31	33	33	33	33	33	**34**	31	32	31	28	32
	50	57	63	66	68	69	70	71	71	71	**72**	69	69	66	60	67
	100	88	92	94	94	95	95	95	95	95	**96**	94	94	91	89	93

(Continues)

TABLE 4 Empirical power of $T_{n,a}$ against competitors ($d = 2$, $\alpha = .05$, 10,000 replications)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	**HV$_5$**	**HV$_\infty$**	**HJG$_{1.5}$**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**
N(0,I$_2$)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**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**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	100	**5**	**5**	**5**	**5**	**5**	**4**	**4**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**4**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
N(μ_2, Σ_1)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**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**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	100	**5**	**4**	**5**	**5**	**4**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**4**	**4**	**5**	**4**	**4**	**5**	**5**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
NMix(0.1,3,I$_2$)	**20**	**33**	**39**	**27**	**39**	**40**	**42**	**42**	**39**	**24**	**17**	**12**	**9**	**32**	**34**	**35**	**35**	**36**	**39**	**39**	**39**	**38**
	50	**69**	**86**	**36**	**86**	**87**	**89**	**87**	**83**	**62**	**47**	**30**	**16**	**72**	**80**	**82**	**83**	**84**	**87**	**86**	**85**	**84**
	100	**97**	*	**45**	*	*	*	**99**	**99**	**93**	**84**	**63**	**32**	**98**	**99**	**99**	*	*	*	*	**99**	**99**
	200	*	*	**56**	*	*	*	*	*	*	**99**	**96**	**69**	*	*	*	*	*	*	*	*	*
NMix(0.5,0,B$_2$)	**20**	**17**	**17**	**16**	**17**	**17**	**16**	**16**	**15**	**12**	**10**	**9**	**6**	**20**	**19**	**19**	**18**	**19**	**18**	**17**	**16**	**15**
	50	**25**	**21**	**23**	**21**	**22**	**26**	**30**	**31**	**28**	**24**	**17**	**11**	**40**	**40**	**38**	**35**	**34**	**31**	**28**	**24**	**21**
	100	**37**	**24**	**31**	**24**	**27**	**42**	**54**	**60**	**59**	**49**	**34**	**18**	**68**	**69**	**65**	**61**	**57**	**49**	**43**	**34**	**27**
	200	**55**	**26**	**40**	**26**	**38**	**71**	**88**	**92**	**92**	**86**	**70**	**39**	**95**	**96**	**93**	**91**	**88**	**78**	**69**	**56**	**39**
NMix(0.9,0,B$_2$)	**20**	**28**	**27**	**27**	**26**	**26**	**25**	**24**	**20**	**12**	**10**	**8**	**6**	**28**	**28**	**28**	**28**	**28**	**27**	**26**	**25**	**25**
	50	**54**	**43**	**53**	**43**	**45**	**47**	**43**	**37**	**22**	**17**	**12**	**9**	**54**	**55**	**56**	**55**	**55**	**53**	**51**	**48**	**45**
	100	**77**	**55**	**76**	**57**	**65**	**69**	**66**	**59**	**39**	**30**	**20**	**12**	**76**	**78**	**79**	**79**	**79**	**76**	**74**	**70**	**66**
	200	**94**	**62**	**93**	**69**	**85**	**90**	**89**	**85**	**68**	**54**	**37**	**20**	**93**	**94**	**95**	**95**	**95**	**94**	**92**	**90**	**87**

(Continues)

TABLE 4 (Continued)

					$BHEP_a$									$T_{n,a}$								
Alt.	n	HV_5	HV_∞	$HJG_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$t_3(0,I_2)$	20	54	53	53	52	52	52	51	48	36	30	23	15	57	56	56	56	57	56	55	52	51
	50	85	78	83	77	80	84	85	84	75	67	55	36	89	88	88	89	89	89	88	85	81
	100	98	90	97	91	95	98	98	98	96	93	85	66	99	99	99	99	99	99	99	98	96
	200	*	96	*	98	*	*	*	*	*	*	99	94	*	*	*	*	*	*	*	*	*
$t_5(0,I_2)$	20	32	32	32	31	31	30	29	25	16	13	11	8	34	34	34	34	34	33	32	31	30
	50	60	53	58	52	53	55	53	49	35	27	20	12	64	63	64	64	65	64	62	58	53
	100	82	67	79	66	71	78	78	75	61	50	36	21	84	84	85	86	87	87	85	80	74
	200	96	77	94	80	89	95	96	95	90	82	66	40	96	97	98	98	98	99	98	97	93
$t_{10}(0,I_2)$	20	16	16	16	15	15	14	13	11	7	6	6	6	16	16	16	16	17	16	15	15	14
	50	29	25	28	24	24	23	20	17	11	9	8	7	30	30	30	30	31	30	28	26	24
	100	44	32	42	32	32	33	31	27	17	13	10	7	42	42	44	45	46	45	42	38	34
	200	62	39	58	39	44	51	51	46	31	23	15	10	58	58	62	65	66	67	63	57	48
$C^2(0,1)$	20	94	93	94	93	93	95	96	96	95	94	90	82	97	97	97	97	97	96	96	94	93
	50	*	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$L^2(0,1)$	20	17	16	16	16	16	15	13	12	8	7	6	6	17	17	17	17	17	17	16	15	15
	50	29	25	27	24	24	23	21	18	12	10	8	7	31	30	31	31	32	30	28	26	24
	100	42	31	40	30	31	34	33	29	20	15	11	8	46	45	47	48	48	47	43	38	33
	200	61	37	56	36	42	52	55	52	37	27	18	11	63	65	68	70	71	70	65	57	47

(Continues)

TABLE 4 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	***n***	**HV$_5$**	**HV$_\infty$**	**HJG$_{1.5}$**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**
$\Gamma^2(0.5,1)$	20	90	96	77	96	97	98	99	99	98	97	95	87	92	93	92	93	94	96	96	96	95
	50	*	*	97	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	97	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	97	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\Gamma^2(5,1)$	20	23	27	20	28	28	29	28	25	15	12	9	7	20	22	22	23	24	26	27	27	27
	50	53	69	36	70	71	73	69	63	40	29	19	11	47	54	58	59	61	67	68	68	68
	100	87	97	53	97	97	98	96	94	75	58	38	19	82	89	92	94	95	96	96	96	96
	200	*	*	74	*	*	*	*	*	98	93	76	41	99	*	*	*	*	*	*	*	*
$P^2_{VII}(5)$	20	28	27	27	27	27	25	24	20	13	10	9	7	29	29	29	29	29	29	27	26	25
	50	51	44	49	44	45	45	43	39	26	20	14	10	55	55	56	55	56	55	51	48	44
	100	73	57	70	57	61	66	66	62	46	37	26	15	77	77	78	79	79	78	75	69	63
	200	92	68	89	71	81	90	91	89	77	65	47	27	94	95	96	96	96	96	94	91	85
$P^2_{VII}(10)$	20	13	13	13	13	13	12	11	10	7	6	6	6	14	14	14	14	14	13	13	13	12
	50	23	21	22	20	20	19	16	14	9	8	7	7	23	23	24	24	25	24	23	21	20
	100	35	26	33	25	26	25	23	19	12	10	8	7	34	34	35	36	36	35	33	29	26
	200	49	31	46	30	32	36	35	31	20	15	11	8	46	47	50	53	54	52	48	42	35
$S^2(\mathrm{Exp}(1))$	20	67	66	64	65	65	69	74	76	78	77	72	59	77	75	75	76	77	76	73	68	64
	50	92	83	89	83	89	96	99	99	*	*	99	97	98	98	98	98	99	99	98	96	91
	100	99	90	98	92	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	93	*	98	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

(Continues)

TABLE 4 (Continued)

Alt.	n	HV_5	HV_∞	$HJG_{1.5}$	$BHEP_a$ 0.1	0.25	0.5	0.75	1	2	3	5	10	$T_{n,a}$ 0.1	0.25	0.5	0.75	1	2	3	5	10
$S^2(B(1,2))$	20	15	16	14	16	16	18	21	24	34	36	35	28	25	22	22	22	23	22	19	17	15
	50	10	12	7	12	13	22	38	53	75	79	77	67	35	33	33	33	35	33	26	19	14
	100	7	10	2	9	12	34	66	85	97	98	98	94	52	53	55	58	59	53	42	26	15
	200	6	9	0	8	15	65	95	*	*	*	*	*	81	84	89	91	92	88	76	51	22
$S^2(\chi^2_5)$	20	22	21	21	21	20	20	18	16	10	9	8	6	23	23	23	23	23	23	21	20	20
	50	38	32	36	32	32	33	32	30	20	16	12	9	43	41	42	43	44	43	40	36	32
	100	55	40	51	39	42	49	51	50	37	29	19	12	61	60	62	64	65	64	60	52	44
	200	76	47	68	48	59	76	81	81	70	58	40	22	80	81	84	87	88	89	86	79	66

TABLE 5 Empirical power of $T_{n,a}$ against competitors ($d = 3$, $\alpha = .05$, 10,000 replications)

					BHEP$_a$									$T_{n,a}$								
Alt.	***n***	**HV$_5$**	**HV$_\infty$**	**HJG$_{1.5}$**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**
N(0,I$_3$)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**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**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	100	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
N(μ_3, Σ_1)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**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**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	100	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**6**	**5**	**5**	**6**	**6**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
NMix(0.1,3,I$_3$)	**20**	**32**	**39**	**28**	**40**	**41**	**44**	**44**	**39**	**21**	**14**	**10**	**7**	**31**	**34**	**35**	**35**	**35**	**38**	**39**	**39**	**38**
	50	**65**	**91**	**35**	**91**	**93**	**95**	**93**	**89**	**64**	**42**	**22**	**10**	**71**	**84**	**87**	**86**	**86**	**89**	**91**	**89**	**87**
	100	**95**	*	**35**	*	*	*	*	*	**95**	**81**	**47**	**17**	**99**	*	*	*	*	*	*	*	*
	200	*	*	**38**	*	*	*	*	*	*	*	**89**	**37**	*	*	*	*	*	*	*	*	*
NMix(0.5,0,B$_3$)	**20**	**27**	**28**	**25**	**27**	**28**	**29**	**29**	**28**	**21**	**16**	**12**	**9**	**37**	**37**	**35**	**34**	**33**	**32**	**30**	**27**	**25**
	50	**45**	**40**	**37**	**39**	**42**	**53**	**64**	**68**	**62**	**50**	**29**	**12**	**76**	**77**	**73**	**69**	**67**	**63**	**58**	**50**	**40**
	100	**62**	**45**	**46**	**43**	**55**	**82**	**94**	**97**	**95**	**88**	**63**	**24**	**97**	**98**	**96**	**94**	**93**	**90**	**85**	**75**	**57**
	200	**87**	**48**	**61**	**51**	**80**	**99**	*	*	*	*	**96**	**54**	*	*	*	*	*	*	**99**	**97**	**85**
NMix(0.9,0,B$_3$)	**20**	**43**	**42**	**42**	**41**	**41**	**39**	**34**	**28**	**15**	**11**	**8**	**7**	**42**	**43**	**43**	**43**	**43**	**43**	**42**	**40**	**38**
	50	**79**	**72**	**78**	**71**	**73**	**73**	**67**	**59**	**35**	**25**	**15**	**9**	**80**	**80**	**80**	**80**	**80**	**80**	**78**	**75**	**72**
	100	**96**	**87**	**95**	**88**	**92**	**93**	**90**	**85**	**63**	**46**	**27**	**13**	**96**	**96**	**96**	**96**	**97**	**96**	**96**	**95**	**92**
	200	*	**93**	*	**96**	**99**	*	**99**	**98**	**90**	**77**	**51**	**21**	*	*	*	*	*	*	*	*	**99**

(Continues)

TABLE 5 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	HV$_5$	HV$_\infty$	HJG$_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$t_3(0,I_3)$	20	65	65	62	64	64	64	61	56	40	31	21	11	70	70	69	69	69	69	68	64	61
	50	94	91	91	90	91	94	94	93	85	75	56	28	97	97	97	96	97	97	96	95	92
	100	*	98	99	98	99	*	*	*	99	97	89	56	*	*	*	*	*	*	*	*	99
	200	*	*	*	*	*	*	*	*	*	*	*	90	*	*	*	*	*	*	*	*	*
$t_5(0,I_3)$	20	40	40	38	39	39	37	34	29	17	13	10	8	44	43	43	42	42	43	41	38	36
	50	73	69	69	67	68	69	67	62	44	32	20	10	81	79	79	79	79	79	78	73	68
	100	92	83	88	82	86	91	91	89	75	60	38	17	96	95	95	95	96	96	96	94	89
	200	99	93	98	94	98	*	*	*	97	91	71	33	*	*	*	*	*	*	*	*	99
$t_{10}(0,I_3)$	20	19	20	18	19	19	18	15	12	7	7	6	6	21	21	21	21	21	21	20	18	17
	50	38	34	35	33	33	31	27	22	13	10	8	6	43	41	41	41	41	42	40	36	32
	100	56	46	51	44	46	46	42	36	21	15	11	7	64	60	60	61	62	64	62	55	46
	200	79	56	70	55	62	70	69	64	41	28	16	9	84	79	80	83	84	87	85	79	68
$C^3(0,1)$	20	98	98	97	97	98	98	98	99	97	96	91	76	99	99	99	99	99	99	99	98	97
	50	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$L^3(0,1)$	20	17	17	16	17	16	16	14	11	7	7	6	6	18	18	18	18	18	18	17	16	15
	50	31	28	29	27	27	25	21	17	11	9	8	6	37	35	35	35	35	35	33	29	26
	100	48	37	42	35	37	38	35	31	19	14	10	7	58	55	55	55	56	57	53	46	38
	200	68	44	57	43	49	59	60	55	35	24	14	8	79	76	78	79	80	81	77	69	55

(Continues)

TABLE 5 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	HV$_5$	HV$_\infty$	HJG$_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$\Gamma^3(0.5, 1)$	20	91	98	80	98	98	99	*	*	98	96	89	66	94	96	95	95	95	97	98	98	97
	50	*	*	98	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\Gamma^3(5, 1)$	20	22	27	20	27	28	29	28	24	13	10	8	6	19	21	22	22	22	25	26	27	26
	50	53	74	37	74	76	77	74	65	37	23	13	8	47	55	59	60	62	68	71	72	71
	100	87	99	53	98	99	99	98	95	74	50	25	11	80	91	94	95	96	98	98	97	97
	200	*	*	74	*	*	*	*	*	98	89	55	20	99	*	*	*	*	*	*	*	*
$P^3_{VII}(5)$	20	29	29	28	29	28	27	23	20	11	9	7	6	32	32	32	32	32	31	30	28	26
	50	58	53	55	51	52	52	47	41	25	18	12	8	66	64	64	65	65	64	61	56	51
	100	82	70	76	69	72	76	74	68	48	34	21	11	89	87	87	88	88	88	86	81	74
	200	96	81	92	82	90	95	95	94	80	64	39	16	99	98	98	99	99	99	98	97	93
$P^3_{VII}(20)$	20	14	14	14	14	13	13	11	9	7	6	6	5	15	15	15	15	15	15	14	13	12
	50	26	23	24	23	22	20	16	13	8	7	6	6	29	28	28	28	28	28	27	23	21
	100	40	31	36	30	30	28	24	19	12	9	8	6	44	41	42	43	44	44	41	36	30
	200	58	39	51	37	39	42	39	34	19	14	10	8	62	58	59	62	63	65	61	52	43

(Continues)

TABLE 5 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	HV$_5$	HV$_\infty$	HJG$_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$\mathcal{S}^3$(Exp(1))	20	87	89	83	87	88	91	94	95	96	95	89	75	96	95	94	94	94	95	93	90	85
	50	*	98	98	98	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\mathcal{S}^3$(B(1, 2))	20	44	48	36	45	47	52	59	64	73	72	65	49	67	63	61	61	61	61	58	50	44
	50	56	53	39	49	59	82	94	97	99	99	98	91	93	91	90	91	92	92	90	80	63
	100	68	53	34	52	75	98	*	*	*	*	*	*	*	99	*	*	*	*	*	98	88
	200	89	53	27	58	96	*	*	*	*	*	*	*	*	*	*	*	*	*	100	*	*
$\mathcal{S}^3(\chi^2_5)$	20	42	43	38	41	41	42	40	37	29	23	16	11	53	51	50	49	49	49	47	42	39
	50	72	67	65	65	68	75	78	79	72	63	43	20	88	85	84	84	85	86	84	78	68
	100	91	79	84	78	86	96	98	98	97	93	79	42	99	98	98	98	99	99	99	97	91
	200	99	87	96	89	98	*	*	*	*	*	99	80	*	*	*	*	*	*	*	*	*

TABLE 6 Empirical power of $T_{n,a}$ against competitors ($d = 5$, $\alpha = .05$, 10,000 replications)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	HV$_5$	HV$_\infty$	HJG$_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
N(0,I$_5$)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	50	**5**	**5**	**5**	**5**	**5**	**5**	**6**	**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**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
N(μ_5, Σ_1)	**20**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	50	**5**	**5**	**5**	**5**	**5**	**5**	**6**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	100	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**4**	**4**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	200	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**
NMix(0.1,3,I$_5$)	**20**	**30**	**33**	**28**	**34**	**34**	**35**	**31**	**24**	**12**	**9**	**8**	**3**	**25**	**26**	**28**	**28**	**28**	**30**	**32**	**33**	**31**
	50	**50**	**86**	**30**	**86**	**90**	**95**	**94**	**85**	**38**	**18**	**9**	**8**	**48**	**65**	**78**	**78**	**76**	**75**	**81**	**84**	**82**
	100	**78**	*	**25**	*	*	*	*	*	**84**	**43**	**14**	**9**	**76**	**98**	*	*	*	*	*	*	*
	200	**99**	*	**21**	*	*	*	*	*	*	**89**	**29**	**9**	**99**	*	*	*	*	*	*	*	*
NMix(0.5,0,B$_5$)	**20**	**55**	**60**	**45**	**57**	**57**	**60**	**59**	**55**	**38**	**25**	**19**	**5**	**72**	**73**	**71**	**69**	**67**	**65**	**64**	**58**	**50**
	50	**77**	**77**	**60**	**73**	**80**	**92**	**98**	**98**	**95**	**76**	**35**	**20**	**98**	**99**	**98**	**98**	**97**	**95**	**95**	**91**	**80**
	100	**93**	**84**	**73**	**83**	**95**	*	*	*	*	*	**71**	**26**	*	*	*	*	*	*	*	*	**97**
	200	*	**89**	**84**	**92**	*	*	*	*	*	*	**99**	**36**	*	*	*	*	*	*	*	*	*
NMix(0.9,0,B$_5$)	**20**	**62**	**60**	**61**	**59**	**58**	**55**	**44**	**33**	**14**	**10**	**9**	**3**	**57**	**57**	**57**	**58**	**59**	**60**	**60**	**58**	**55**
	50	**96**	**94**	**95**	**94**	**94**	**92**	**86**	**75**	**43**	**25**	**13**	**10**	**95**	**95**	**95**	**95**	**95**	**95**	**95**	**95**	**93**
	100	*	*	*	**99**	*	**99**	**98**	**96**	**78**	**51**	**19**	**12**	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	**98**	**86**	**40**	**13**	*	*	*	*	*	*	*	*	*

(Continues)

TABLE 5 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	**HV$_5$**	**HV$_\infty$**	**HJG$_{1.5}$**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**	**0.1**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**5**	**10**
$t_3(0,I_5)$	20	79	80	74	79	79	77	71	62	38	26	19	5	84	84	83	83	83	83	82	79	74
	50	99	99	98	98	99	99	99	98	91	76	42	24	*	*	*	*	*	*	*	*	99
	100	*	*	*	*	*	*	*	*	*	98	79	36	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	99	56	*	*	*	*	*	*	*	*	*
$t_5(0,I_5)$	20	54	55	49	54	53	49	41	30	16	12	10	3	59	59	58	58	58	57	57	53	48
	50	90	89	84	87	87	87	85	77	51	30	15	11	96	96	95	94	94	95	95	92	87
	100	99	98	97	97	98	99	99	98	88	64	26	13	*	*	*	*	*	*	*	*	99
	200	*	*	*	*	*	*	*	*	*	95	58	17	*	*	*	*	*	*	*	*	*
$t_{10}(0,I_5)$	20	26	26	23	25	25	22	17	12	8	7	7	2	29	28	28	28	28	28	27	25	22
	50	55	54	48	50	50	46	40	28	14	11	8	7	67	65	63	62	62	62	62	57	48
	100	78	71	68	67	69	70	63	54	28	16	9	7	91	88	86	86	86	87	87	83	73
	200	95	85	87	82	89	93	92	87	56	31	13	9	99	99	98	98	98	99	99	98	94
$C^5(0,1)$	20	*	*	99	*	*	*	*	99	98	94	87	61	*	*	*	*	*	*	*	*	99
	50	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$L^5(0,1)$	20	16	17	15	16	16	14	12	9	6	6	6	2	18	18	18	18	18	18	17	16	14
	50	35	33	30	31	30	27	24	15	9	7	6	6	45	44	42	42	42	41	40	35	29
	100	54	48	45	44	45	44	36	29	15	10	7	6	73	70	67	67	67	68	67	59	47
	200	75	58	60	54	60	66	63	55	27	16	9	7	94	90	88	88	88	90	89	83	68

(Continues)

TABLE 5 (Continued)

					BHEP$_a$									$T_{n,a}$								
Alt.	n	HV$_5$	HV$_\infty$	HJG$_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$\Gamma^5(0.5, 1)$	20	93	98	84	98	99	*	*	99	93	81	69	34	93	96	97	96	96	97	98	98	97
	50	*	*	99	*	*	*	*	*	*	*	99	91	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\Gamma^5(5, 1)$	20	20	23	18	24	25	26	23	18	9	7	7	4	17	18	19	20	20	21	22	24	24
	50	50	76	37	77	79	80	76	59	23	12	8	7	44	52	60	62	62	64	69	73	72
	100	83	99	55	99	99	*	98	95	54	24	9	7	73	87	94	96	96	97	98	98	98
	200	*	*	72	*	*	*	*	*	93	57	17	7	97	*	*	*	*	*	*	*	*
$P^5_{VII}(5)$	20	32	33	29	32	31	28	21	15	8	7	7	2	35	34	34	34	34	34	34	30	28
	50	66	64	60	62	62	58	52	39	20	13	9	7	77	76	74	74	74	73	72	68	60
	100	89	83	82	81	83	84	80	72	40	23	11	9	96	96	95	95	95	95	95	92	85
	200	99	93	95	92	96	98	98	96	76	47	19	9	*	*	*	*	*	*	*	*	98
$P^5_{VII}(10)$	20	14	14	13	14	13	12	9	7	6	6	5	3	14	14	14	14	14	14	14	13	12
	50	28	26	25	25	24	21	18	11	7	6	6	6	34	32	32	32	32	31	31	27	23
	100	45	38	39	36	36	31	23	18	10	8	6	6	58	53	51	51	51	53	52	45	36
	200	66	49	54	46	48	48	41	32	16	10	8	6	82	75	71	72	73	76	76	68	54
$\mathcal{S}^5(\mathrm{Exp}(1))$	20	99	99	96	99	99	99	99	99	99	98	95	83	*	*	*	*	*	*	*	99	98
	50	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

(Continues)

TABLE 5 (Continued)

					BHEP_a									$T_{n,a}$								
Alt.	n	HV_5	HV_∞	$\mathrm{HJG}_{1.5}$	0.1	0.25	0.5	0.75	1	2	3	5	10	0.1	0.25	0.5	0.75	1	2	3	5	10
$\mathcal{S}^5(\mathrm{B}(1,2))$	20	87	91	75	89	89	91	93	94	94	91	86	66	97	97	96	96	95	95	94	91	84
	50	98	98	90	97	99	*	*	*	*	*	*	99	*	*	*	*	*	*	*	*	*
	100	*	99	96	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	200	*	*	99	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\mathcal{S}^5(\chi_5^2)$	20	77	81	68	78	78	77	75	71	58	45	36	12	89	88	87	86	85	85	84	80	72
	50	97	98	93	96	98	99	*	*	99	95	75	51	*	*	*	*	*	*	*	*	98
	100	*	*	99	*	*	*	*	*	*	*	98	73	*	*	*	*	*	*	*	*	*
	200	*	*	*	*	*	*	*	*	*	*	*	95	*	*	*	*	*	*	*	*	*

3. the Anderson–Darling test (AD), see Anderson and Darling (1952),
4. the Baringhaus–Henze–Epps–Pulley test (BHEP), see Henze and Wagner (1997),
5. the del Barrio–Cuesta–Albertos–Mátran–Rodríguez-Rodríguez test (BCMR), see del Barrio, Cuesta-Albertos, Matran, and Rodriguez-Rodriguez (1999),
6. the Betsch–Ebner test (BE), see Betsch and Ebner (2020).

For the implementation of the first three tests in `R` we refer to the package `nortest` by Gross and Ligges (2015). Tests based on the empirical characteristic function are represented by the BHEP-test with tuning parameter $a > 0$. The statistic is given in (40), $a = 1$ is fixed, and the critical values are taken from Henze (1990).

The BCMR-test is based on the L^2-Wasserstein distance, see section 3.3 in del Barrio et al. (2000), and is given by

$$\mathrm{BCMR} = n\left(1 - \frac{1}{S_n^2}\left(\sum_{k=1}^{n} X_{(k)} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \Phi^{-1}(t)\,\mathrm{d}t\right)^2\right) - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{(\varphi(\Phi^{-1}(t)))^2}\,\mathrm{d}t.$$

Here, $X_{(k)}$ is the kth order statistic of $X_1, \dots, X_n$, S_n^2 is the sample variance, and Φ^{-1} is the quantile function of the standard normal law. Simulated critical values can be found in Krauczi (2009), or in table 4 of Betsch and Ebner (2020).

The BE-test is based on a L^2-distance between the empirical zero-bias transformation and the empirical distribution. Since the zero-bias transformation has a unique fixed point under normality, this distance is minimal under H_0. The statistic is given by

$$\begin{aligned}\mathrm{BE}_a = \frac{2}{n}\sum_{1\le j<k\le n}\Bigg\{&\left(1 - \Phi\left(\frac{Y_{(k)}}{\sqrt{a}}\right)\right)((Y_{(j)}^2 - 1)(Y_{(k)}^2 - 1)\\ &+ aY_{(j)}Y_{(k)}) + \frac{a}{\sqrt{2\pi a}}\exp\left(-\frac{Y_{(k)}^2}{2a}\right)(-Y_{(j)}^2Y_{(k)} + Y_{(k)} + Y_{(j)})\Bigg\}\\ &+ \frac{1}{n}\sum_{j=1}^{n}\left\{\left(1 - \Phi\left(\frac{Y_j}{\sqrt{a}}\right)\right)(Y_j^4 + (a-2)Y_j^2 + 1) + \frac{a}{\sqrt{2\pi a}}\exp\left(-\frac{Y_j^2}{2a}\right)(2Y_j - Y_j^3)\right\},\end{aligned}$$

where $Y_1, \dots, Y_n$ is shorthand for the scaled residuals $Y_{n,1}, \dots, Y_{n,n}$, $Y_{(1)} \le \dots \le Y_{(n)}$ are the order statistics of $Y_1, \dots, Y_n$, and Φ stands for the distribution function of the standard normal law. The implementation employs a bootstrap procedure to find a data-driven version of the tuning parameter a, see Allison and Santana (2015). We used $B = 400$ bootstrap replications and the same grid of tuning parameters as in Betsch and Ebner (2020, p. 19).

The alternative distributions are chosen to fit the extensive power study of univariate normality tests by Romão, Delgado, and Costa (2010), in order to ease the comparison with a wide variety of other existing tests. As representatives of symmetric distributions we simulate the Student t_ν-distribution with $\nu \in \{3, 5, 10\}$ degrees of freedom, as well as the uniform distribution $U(-\sqrt{3}, \sqrt{3})$. The asymmetric distributions are represented by the χ_ν^2-distribution with $\nu \in \{5, 15\}$ degrees of freedom, the Gamma distributions $\Gamma(1, 5)$ and $\Gamma(5, 1)$, parametrized by their shape and rate parameter, the Gumbel distribution Gum(1,2) with location parameter 1 and scale parameter 2 as well as the Weibull distribution W(1,0.5) with scale parameter 1 and shape parameter 0.5. As representatives of bimodal distributions we take the mixture of normal distributions $\mathrm{NMix}(p, \mu, \sigma^2)$, where the random variables are generated by $(1-p)\ \mathrm{N}(0,1) + p\ \mathrm{N}(\mu, \sigma^2)$, $p \in (0,1),\ \mu \in \mathbb{R},\ \sigma > 0$.

Table 3 shows that the empirical power estimates of the new test $T_{n,a}$ outperform the other strong procedures for the symmetric t-distribution, and they can compete for most of the other alternatives. Interestingly, the power does not differ too much when varying the tuning parameter a, although an effect is clearly visible, especially for the uniform distribution. A data-driven choice as in Allison and Santana (2015), criticized and revised in Tenreiro (2019), might be of benefit also in connection with the new testing procedure.

7.2 | Testing multivariate normality

In this subsection we consider testing the hypothesis H_0 that the distribution of X is multivariate normal (i.e., belongs to $\mathcal{N}_d$), for the dimensions $d \in \{2,3,5\}$. As competitors to the new test statistic we chose

1. the Henze–Visagie test (HV), see Henze and Visagie (2019),
2. the Henze–Jiménez–Gamero test (HJG), see Henze and Jiménez-Gamero (2019),
3. the Baringhaus–Henze–Epps–Pulley test (BHEP), see Henze and Wagner (1997).

The HV-test uses a weighted L^2-type statistic based on a characterization of the moment generating function that employs a system of first-order partial differential equations. The statistic is defined by

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

where $\gamma > 2$. We followed the recommendation of the authors in Henze and Visagie (2019) and fixed $\gamma = 5$. Since the limiting statistic HV_∞ for $\gamma \to \infty$ is a linear combination of sample skewness in the sense of Mardia and that of Móri, Rohatgi and Székely, we also included HV_∞.

The HJG-test uses a weighted L^2-distance between the empirical moment generating function of the standardized sample and the moment generating function of the standard normal distribution. The test statistic is given by

$$\mathrm{HJG}_\beta = \frac{1}{n\beta^{\frac{d}{2}}} \sum_{j,k=1}^{n} \exp\left(\frac{\|Y_{n,j}+Y_{n,k}\|^2}{4\beta}\right) - \frac{2}{\sqrt{\beta-1/2}} \sum_{j=1}^{n} \exp\left(\frac{\|Y_{n,j}\|^2}{4\beta-2}\right) + \frac{n}{(\beta-1)^{\frac{d}{2}}},$$

with $\beta > 0$. In our simulation we fix $\beta = 1.5$. For each of the tests based on HV_5, HV_∞, and $\mathrm{HJG}_{1.5}$, critical values were simulated with 100,000 replications.

Finally, the now classical BHEP-test examines the weighted L^2-distance between the empirical characteristic function of the standardized data and the characteristic function of the d-variate standard normal distribution. The statistic has the simple form

$$\mathrm{BHEP}_a = \frac{1}{n^2}\sum_{j,k=1}^{n} \exp\left(-\frac{a^2}{2}\|Y_{n,j}-Y_{n,k}\|^2\right) - 2(1+a^2)^{-\frac{d}{2}}\frac{1}{n}\sum_{j=1}^{n}\exp\left(-\frac{a^2\|Y_{n,j}\|^2}{2(1+a^2)}\right) + (1+2a^2)^{-\frac{d}{2}}, \tag{40}$$

with a tuning parameter $a > 0$. A variety of values of a, that is, $a \in \{0.1, 0.25, 0.5, 0.75, 1, 2, 3, 5, 10\}$, has been considered. Critical values can be found in tables I and III of Henze and Wagner (1997), whereas missing critical values have been simulated separately with 100,000 replications.

In order to show that all procedures indeed have the stated Type I error and exhibit their affine invariance, we include the d-dimensional standard normal distribution $\mathrm{N}(0,\mathrm{I}_d)$ as well as the $\mathrm{N}(\mu_d, \Sigma_1)$ distribution, where $\mu_d = (1, 2, \dots, d)^\top$ and Σ_1 is a positive definite matrix with 1's on the diagonal and 0.1 on every off-diagonal entry. The alternative distributions are chosen to fit the simulation study in Henze and Visagie (2019) and are defined as follows. We denote by $\mathrm{NMix}(p, \mu, \Sigma)$ the normal mixture distributions generated by $(1 - p)\ \mathrm{N}_d(0, \mathrm{I}_d) + p\ \mathrm{N}_d(\mu, \Sigma),\ p \in (0, 1),\ \mu \in \mathbb{R}^d,\ \Sigma > 0$, where $\Sigma > 0$ stands for a positive definite matrix. We write in the notation of above $\mu = 3$ for a d-variate vector of 3's and $\Sigma = \mathrm{B}_d$ for a $(d \times d)$-matrix containing 1's on the main diagonal and 0.9's on each off-diagonal entry. We denote by $\mathrm{t}_\nu(0, \mathrm{I}_d)$ the multivariate t-distribution with ν degrees of freedom, see Genz and Bretz (2009). By $\mathrm{DIST}^d(\vartheta)$ we denote the d-variate random vector generated by independently simulated components of the distribution DIST with parameter vector (ϑ), where DIST is taken to be the Cauchy distribution C, the logistic distribution L, the Gamma distribution Γ as well as the Pearson-type VII distribution $\mathrm{P}_{\mathrm{VII}}$, with ϑ denoting in this specific case the degrees of freedom. The spherical symmetric distributions where simulated using the `R` package `distrEllipse`, see Ruckdeschel, Kohl, Stabla, and Camphausen (2006), and are denoted by $\mathcal{S}^d(\mathrm{DIST})$, where DIST stands for the distribution of the radii and was chosen to be the exponential, the beta and the χ^2-distribution.

From Tables 4 to 6, it is obvious that $T_{n,a}$ outperforms the competing tests for most of the alternatives considered, again showing that the tuning parameter has—compared to the BHEP test—little effect on the power. As in the univariate case $T_{n,a}$ has very strong power against the multivariate t-distribution. If the radial distribution of the spherical symmetric alternatives has compact support, the BHEP test exhibits a better performance than $T_{n,a}$. The HV_5- and the $\mathrm{HJG}_{1.5}$-test have a good power, but they are mostly dominated by the BHEP-test and the $T_{n,a}$-test. From our simulation results, it seems that the test performs better for greater values of a if the underlying alternative is skewed. For nonskewed alternatives, we observe a higher power for smaller values of the tuning parameter. Overall, we suggest to choose a small tuning parameter like $a = 0.25$, or, alternatively, if the practitioner suggests skewness of the data, to select $a = 3$. Again, a data-driven choice of the tuning parameter a would be beneficial for the test, but to the best of our knowledge no reliable multivariate method is yet available. It should be stressed, however, that there cannot be an "optimal" value of a, since the global power function of any nonparametric test is flat on balls of alternatives except for alternatives coming from some finite dimensional subspace, see Janssen (2000). Finally, note that all of the tests for multivariate normality under discussion are implemented in the `R` package `mnt`, see Butsch and Ebner (2020).

8 | REAL DATA EXAMPLE: THE IRIS DATA SET

In 1936 R.A. Fisher presented the classical data set called `Iris Flower`, see table I in Fisher (1936). The data consist of the four variables sepal length, sepal width, petal length, and petal width, measured on $n = 50$ specimens of each of three types or iris, namely *Iris setosa*, *Iris versicolor*, and *Iris virginica*. This data set is included in the statistical language `R`, and it can be downloaded from the UCI Machine Learning Repository, see Dua and Graff (2017). That reference provides a list of articles that use this specific data set to validate clustering methods,

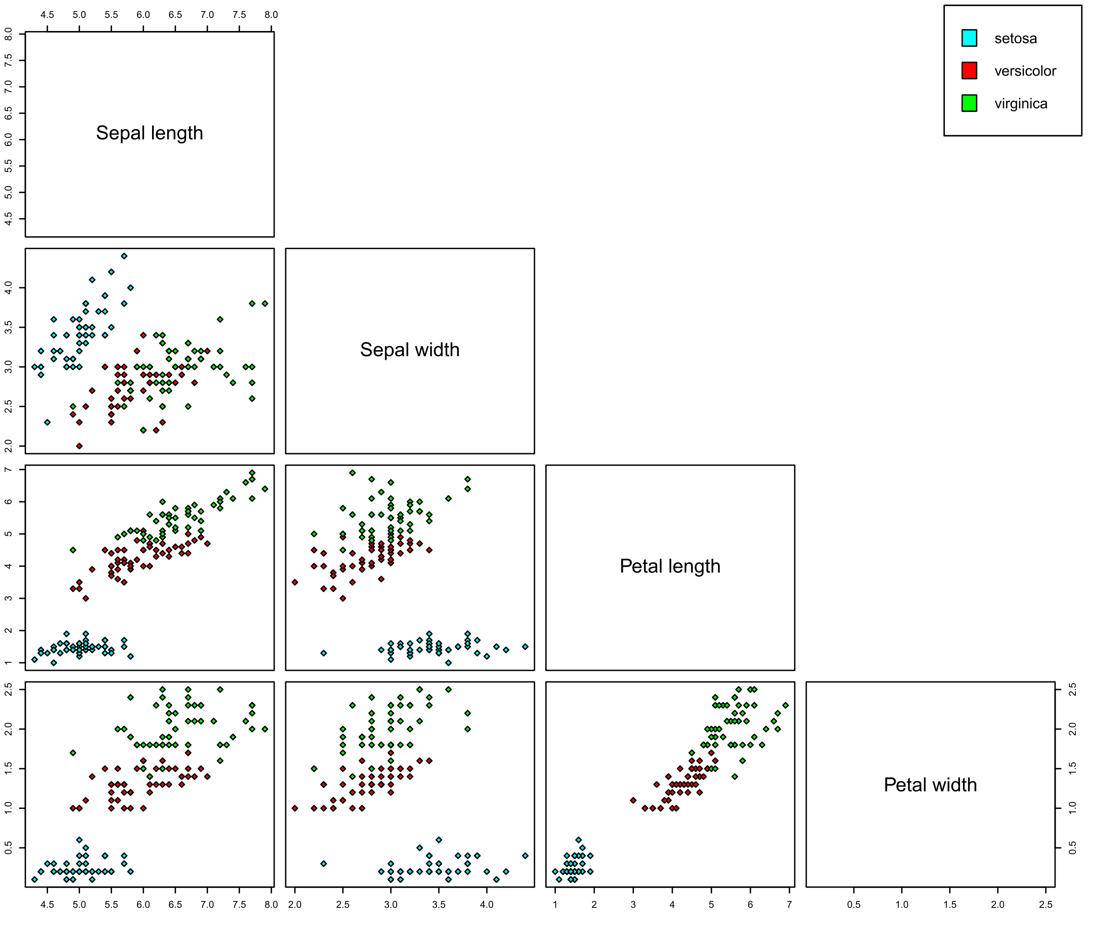

FIGURE 1 2D projections of the `iris` data set with colored species [Color figure can be viewed at wileyonlinelibrary.com]

neural networks or learning algorithms, and it presents a typical test case for statistical classification techniques in machine learning, such as support vector machines. A visualization of two-dimensional projections of the data set is given in Figure 1.

In Table 7 we present empirical p-values, that is, estimated probabilities of obtaining a value at least as large as the observed value of $T_{n,a}$ under the null hypothesis, simulated with 10,000 replications. As can be seen, the test does not reject the hypothesis of normality on a small significance level (like $\alpha = .01$) for the different species for each of the tuning parameters considered. For the *Iris setosa* data, however, an increase of the significance level to .05 results in a rejection of the hypothesis for $a = 2$ and $a = 3$. For the whole data set, we observe a small p-value due to the mixture of the three populations and consequently reject the hypothesis H_0.

9 | CONCLUDING REMARKS

We proved consistency of the test for multivariate normality based on $T_{n,a}$ against each alternative distribution that satisfies the moment condition $\mathbb{E}\|X\|^4 < \infty$. Intuitively, the test should be "all

TABLE 7 Empirical p-values for $T_{n,a}$

Species∖a	0.25	0.5	1	2	3	5	10
setosa	.0631	.0706	.0683	.0431	.0386	.0555	.0918
versicolor	.4402	.3560	.2912	.2766	.2707	.2626	.2573
virginica	.1943	.1671	.1336	.1385	.1643	.2042	.2071
Mixed populations	.0000	.0000	.0000	.0000	.0012	.0048	.0150

the more consistent" if $\mathbb{E}\|X\|^4 = \infty$. In fact, we conjecture consistency of the new test against any non-normal alternative distribution.

The limiting random element $T_{\infty,a} = \|Z\|^2_{\mathbb{H}}$ from Theorem 5(b) has the same distribution as $\sum_{j=1}^{\infty} \lambda_j N_j^2$, where the N_j are i.i.d. standard normal random variables, and the λ_j are the positive eigenvalues corresponding to eigenfunctions of the linear integral operator $Kf(s) = \int_{\mathbb{R}^d} K(s,t) f(t)\, w_a(t)\, \mathrm{d}t$ associated with the covariance kernel K from Theorem 5(a), that is, we have $\lambda f(s) = \int_{\mathbb{R}^d} K(s,t) f(t)\, w_a(t)\, \mathrm{d}t$. These positive eigenvalues clearly depend on K and the weight function w_a. It is hardly possible to obtain a closed form expression for $\lambda_1, \lambda_2, \ldots$. It would be interesting to approximate the eigenvalues either numerically or by some Monte Carlo method, since the largest eigenvalue has a crucial influence on the approximate Bahadur efficiency, see Bahadur (1960) and the monograph Nikitin (1995), as well as Henze, Nikitin, and Ebner (2009) for results on distribution-free L_p-type statistics.

ACKNOWLEDGEMENTS

The authors thank Ioannis Anapolitanos for his help in proving Theorem 1 and for sharing his knowledge of Schrödinger operators. They also express our gratitude to Michael Ruggenthaler from the Center for Free-Electron Laser Science (CFEL) at the Max Planck Institute for the Structure and Dynamics of Matter for elaborations and discussions on the harmonic oscillator. Furthermore, the authors want to thank three anonymous reviewers for their helpful suggestions and insights that led to an improvement of the article.

ORCID

Bruno Ebner https://orcid.org/0000-0003-4329-8794

REFERENCES

Allison, J. S., & Santana, L. (2015). On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, *85*, 3276–3288.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, *23*, 193–212.

Arcones, M. A. (2007). Two tests for multivariate normality based on the characteristic function. *Mathematical Methods of Statistics*, *16*, 177–201.

Bahadur, R. R. (1960). Stochastic comparison of tests. *The Annals of Mathematical Statistics*, *31*, 276–295.

Baringhaus, L., Ebner, B., & Henze, N. (2017). The limit distribution of weighted L^2-Goodness-of-Fit statistics under fixed alternatives, with applications. *The Annals of the Institute of Statistical Mathematics*, *69*, 969–995.

Baringhaus, L., & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, *35*, 339–348.

Baringhaus, L., & Henze, N. (1991). Limit distributions for measures of multivariate skewness and kurtosis based on projections. *Journal of Multivariate Analysis*, *38*, 51–69.

Baringhaus, L., & Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. *The Annals of Statistics*, *20*, 1889–1902.

Barndorff-Nielsen, O. (1963). On the limit behavior of extreme order statistics. *The Annals of Mathematical Statistics*, *34*, 992–1002.

del Barrio, E., Cuesta-Albertos, J. A., Matran, C., & Rodriguez-Rodriguez, J. M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *The Annals of Statistics*, *27*, 1230–1239.

Betsch, S., & Ebner, B. (2020). Testing normality via a distributional fixed point property in the stein characterization. *TEST*. *29*(1):105-138.

Bowman, A., & Foster, P. (1993). Adaptive smoothing and density-based tests of multivariate normality. *Journal of the American Statistical Association*, *88*, 529–537.

Butsch, L. & Ebner, B. (2020) *mnt: Affine invariant tests of multivariate normality* (R package version 1.2). Retrieved from https://CRAN.R-project.org/package=mnt

Cox, D., & Small, N. (1978). Testing multivariate normality. *Biometrika*, *65*, 263–272.

del Barrio, E., Cuesta-Albertos, J. A., Matrán, C., Csörgő, S., Cuadras, C. M., de Wet, T., … Stute, W. (2000). Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *TEST*, *9*, 1–96.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939.

Dua, D. & Graff, C. (2017) *UCI machine learning repository*. Retrieved from http://archive.ics.uci.edu/ml

Eaton, M. L. (1983). *Multivariate statistics : A vector space approach Wiley Series in Probability and Mathematical Statistics*. New York: Wiley.

Eaton, M. L., & Perlman, M. D. (1973). The non-singularity of generalized sample covariance matrices. *The Annals of Statistics*, *1*, 710–717.

Ebner, B. (2012). Asymptotic theory for the test for multivariate normality by Cox and Small. *Journal of Multivariate Analysis*, *111*, 368–379.

Farrell, P. J., Salibian-Barrera, M., & Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, *77*, 1065–1080.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179–188.

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities Lecture Notes in Statistics* (Vol. *195*). Dordrecht: Springer.

Ghaziasgar, M., Bagula, A., Thron, C., & Ajayi, O. (2020). *Automatic sign language manual parameter recognition (I): Survey* (pp. 63–91). Cham: Springer International Publishing.

Gross, J. & Ligges, U. (2015) *nortest: Tests for Normality* (R package version 1.0-4).

Gustafson, S. J. (2011). *Mathematical concepts of quantum mechanics* (2nd ed.). Heidelberg: Springer.

Henze, N. (1990). An approximation to the limit distribution of the Epps–Pulley test statistic for normality. *Metrika*, *37*, 7–18.

Henze, N. (1994a). The asymptotic behaviour of a variant of multivariate kurtosis. *Communications in Statistics - Theory and Methods*, *23*, 1047–1061.

Henze, N. (1994b). On Mardia's kurtosis test for multivariate normality. *Communications in Statistics - Theory and Methods*, *23*, 1031–1045.

Henze, N. (1997a). Extreme smoothing and testing for multivariate normality. *Statistics & Probability Letters*, *35*, 203–213.

Henze, N. (1997b). Limit laws for multivariate skewness in the sense of Móri, Rohatgi and Székely. *Statistics & Probability Letters*, *33*, 299–307.

Henze, N. (2002). Invariant tests for multivariate normality: A critical review. *Statistical Papers*, *43*, 467–506.

Henze, N., & Jiménez-Gamero, M. D. (2019). A class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. *TEST*, *28*, 499–521.

Henze, N., Jiménez-Gamero, M. D., & Meintanis, S. G. (2019). Characterizations of multinormality and corresponding tests of fit, including for GARCH models. *Econometric Theory*, *35*, 510–546.

Henze, N., & Mayer, C. (2020). More good news on the HKM test for multivariate reflected symmetry about an unknown centre. *Annals of the Institute of Statistical Mathematics*, *72*, 741–770.

Henze, N., Nikitin, Y., & Ebner, B. (2009). Integral distribution-free statistics of L_p-type and their asymptotic comparison. *Computational Statistics & Data Analysis*, *53*, 3426–3438.

Henze, N., & Visagie, J. (2020). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *The Annals of the Institute of Statistical Mathematics*, *72*, 1109–1136.

Henze, N., & Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, *61*, 1–23.

Henze, N., & Zirkler, B. (1990). A class of invariant and consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, *19*, 3595–3617.

Janssen, A. (2000). Global power functions of goodness of fit tests. *The Annals of Statistics*, *28*, 239–253.

Kankainen, K., Taskinen, S., & Oja, H. (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods & Applications*, *16*, 357–359.

Kim, I., & Park, S. (2018). Likelihood ratio tests for multivariate normality. *Communications in Statistics - Theory and Methods*, *47*, 1923–1934.

Krauczi, É. (2009). A study of the quantile correlation test for normality. *TEST*, *18*, 156–165.

Luo, H. & Zhong, S. (2017) Gas turbine engine gas path anomaly detection using deep learning with Gaussian distribution. In *2017 Prognostics and System Health Management Conference (PHM-Harbin)* (pp 1–6).

Malkovich, J., & Afifi, A. (1973). On tests of multivariate normality. *Journal of the American Statistical Association*, *68*, 176–179.

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

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya Series B*, *36*, 115–128.

Mecklin, C. J., & Mundfrom, D. J. (2005). A Monte Carlo comparison of the 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., & Székely, G. J. (1993). On multivariate skewness and kurtosis. *Teor. Veroyatn. Primen.*, *38*, 675–679.

Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge: Cambridge University Press.

Pudelko, J. (2005). On a new affine invariant and consistent test for multivariate normality. *Probability and Mathematical Statistics*, *25*, 43–54.

R Core Team. (2019). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

Romão, X., Delgado, R., & Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, *80*, 545–591.

Ruckdeschel, P., Kohl, M., Stabla, T., & Camphausen, F. (2006). S4 classes for distributions. *R News*, *6*, 2–6.

Shapiro, S. S., & Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, *67*, 215–216.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Sürücü, B. (2006). Goodness-of-fit tests for multivariate distributions. *Communications in Statistics - Theory and Methods*, *35*, 1319–1331.

Székely, G. J., & Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, *93*, 58–80.

Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Computational Statistics & Data Analysis*, *53*, 1038–1053.

Tenreiro, C. (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, *89*, 1780–1797.

Thangavelu, S. (1993). *Lectures on Hermite and Laguerre expansions Mathematical Notes* (Vol. *42*). Princeton, NJ: Princeton University Press.

Thulin, M. (2014). Tests for multivariate normality based on canonical correlations. *Statistical Methods & Applications*, *23*, 189–208.

Villaseñor Alva, J. A., & González Estrada, E. (2009). A generalization of Shapiro–Wilk's test for multivariate normality. *Communications in Statistics - Theory and Methods*, *38*, 1870–1883.

Voinov, V., Pya, N., Makarov, R., & Voinov, Y. (2016). New invariant and consistent chi-squared type goodness-of-fit tests for multivariate normality and a related comparative simulation study. *Communications in Statistics - Theory and Methods*, *45*, 3249–3263.

How to cite this article: Dörr P, Ebner B, Henze N. Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. *Scand J Statist*. 2020;1–46. https://doi.org/10.1111/sjos.12477

APPENDIX . PROOFS AND AUXILIARY RESULTS

This section contains the proofs of some of the theorems as well as some auxiliary results that are used in the main text. Recall our standing assumption that the distribution of X is absolutely continuous. In what follows, let $\Delta_{n,j} = (S_n^{-1/2} - \mathrm{I}_d)X_j - S_n^{-1/2}\overline{X}_n, \quad j = 1, \dots, n$. Notice that $\Delta_{n,j} = Y_{n,j} - X_j$, where the scaled residuals $Y_{n,j}$ are given in (5).

Proposition 3. *Let $X, X_1, X_2, \dots$ be i.i.d. random vectors satisfying $\mathbb{E}\|X\|^4 < \infty$, $\mathbb{E}(X) = 0$ and $\mathbb{E}(XX^\top) = \mathrm{I}_d$. We then have:*

(a) $\sum_{j=1}^n \|\Delta_{n,j}\|^2 = O_{\mathbb{P}}(1)$.

(b) $\frac{1}{n}X_{j=1}^n \|\Delta_{n,j}\|^2 \overset{\text{a.s.}}{\to} 0$.

(c) $\max_{j=1,\dots,n}\|\Delta_{n,j}\| = o_{\mathbb{P}}(n^{-1/4})$.

Proof. (a) Notice that $\|\Delta_{n,j}\|^2 = X_j^\top (S_n^{-1/2} - \mathrm{I}_d)^2 X_j - 2\overline{X}_n^\top S_n^{-1/2}(S_n^{-1/2} - \mathrm{I}_d)X_j + \overline{X}_n S_n^{-1}\overline{X}_n$ and thus, using $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, where $\mathrm{tr}(C)$ denotes the trace of a square matrix C,

$$\sum_{j=1}^n \|\Delta_{n,j}\|^2 = \mathrm{tr}\left((S_n^{-1/2} - \mathrm{I}_d)^2 \sum_{j=1}^n X_j X_j^\top\right) - 2n\overline{X}_n^\top S_n^{-1/2}(S_n^{-1/2} - \mathrm{I}_d)\overline{X}_n + n\overline{X}_n S_n^{-1}\overline{X}_n. \tag{A1}$$

Due to $\mathbb{E}\|X_1\|^4 < \infty$, the central limit theorem implies $\sqrt{n}(S_n^{-1} - \mathrm{I}_d) = -S_n^{-1}\sqrt{n}(S_n - \mathrm{I}_d) = O_{\mathbb{P}}(1)$. Since $\sqrt{n}(S_n^{-1/2} - \mathrm{I}_d)(S_n^{-1/2} + \mathrm{I}_d) = \sqrt{n}(S_n^{-1} - \mathrm{I}_d)$, it follows that $\sqrt{n}(S_n^{-1/2} - \mathrm{I}_d) = O_{\mathbb{P}}(1)$. In view of $\sum_{j=1}^n X_j X_j^\top = O_{\mathbb{P}}(n)$ and $\sqrt{n}\overline{X}_n = O_{\mathbb{P}}(1)$, we are done.

(b) After dividing both sides of (A1) by n, the first summand on the right hand side converges to 0 $\mathbb{P}$-almost surely because $n^{-1}X_{j=1}^n X_j X_j^\top \overset{\text{a.s.}}{\to} \mathrm{I}_d$, and $S_n^{-1/2} \overset{\text{a.s.}}{\to} \mathrm{I}_d$. The same holds for the other two terms, since $\overline{X}_n \overset{\text{a.s.}}{\to} 0$.

(c) Let $\|A\|_{\mathrm{sp}}$ be the spectral norm of a matrix A. Then

$$\|\Delta_{n,j}\| \le \|S_n^{-1/2} - \mathrm{I}_d\|_{\mathrm{sp}}\ \|X_j\| + \|S_n^{-1/2}\|_{\mathrm{sp}}\|\overline{X}_n\| \tag{A2}$$

and thus $\max_{j=1,\dots,n}\|\Delta_{n,j}\| \le \|S_n^{-1/2} - \mathrm{I}_d\|_{\mathrm{sp}} \max_{j=1,\dots,n}\|X_j\| + \|S_n^{-1/2}\|_{\mathrm{sp}}\|\overline{X}_n\|$. From theorem 5.2 of Barndorff-Nielsen (1963) we have $\max_{j=1,\dots,n}\|X_j\|/n^{1/4} \overset{\text{a.s.}}{\to} 0$. Since $\sqrt{n}\|S_n^{-1/2} - \mathrm{I}_d\|_{\mathrm{sp}} = O_{\mathbb{P}}(1)$, $\|S_n^{-1/2}\|_{\mathrm{sp}} = O_{\mathbb{P}}(1)$ and $\sqrt{n}\|\overline{X}_n\| = O_{\mathbb{P}}(1)$, we are done. ■

Proposition 4. *Let $X, X_1, X_2, \ldots$ be i.i.d. random vectors satisfying $\mathbb{E}\|X\|^6 < \infty$, $\mathbb{E}(X) = 0$ and $\mathbb{E}(XX^\top) = \mathrm{I}_d$. We then have $n^{-1}\sum_{j=1}^n \|\Delta_{n,j}\|^k \|X_j\|^\ell \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$ for each $k \in \mathbb{N}$ and $\ell \in \mathbb{N}_0$ such that $k + \ell \leq 6$.*

Proof. From (A2) we have $\|\Delta_{n,j}\|^k \, \|X_j\|^\ell \leq 2^k \|S_n^{-1/2} - \mathrm{I}_d\|_{\text{sp}}^k \, \|X_j\|^{k+\ell} + 2^k \|S_n^{-1/2}\|_{\text{sp}}^k \|\overline{X}_n\|^k \|X_j\|^\ell$. Since $\|S_n^{-1/2} - \mathrm{I}_d\|_{\text{sp}} \overset{\text{a.s.}}{\to} 0$, $S_n^{-1/2} \overset{\text{a.s.}}{\to} \mathrm{I}_d$ and $\|\overline{X}_n\| \overset{\text{a.s.}}{\to} 0$, the assertion follows from the strong law of large numbers. ∎

Proof. In what follows, we put $Y_j = Y_{n,j}$ and $\Delta_j = \Delta_{n,j}$ for the sake of brevity. Notice that $\cos(t^\top Y_j) = \cos(t^\top X_j) - \sin(\Theta_j) t^\top \Delta_j$, $\sin(t^\top Y_j) = \sin(t^\top X_j) + \cos(\Gamma_j) t^\top \Delta_j$, where Θ_j, Γ_j depend on $X_1, \ldots, X_n$ and t and satisfy

$$|\Theta_j - t^\top X_j| \leq |t^\top \Delta_j|, \quad |\Gamma_j - t^\top X_j| \leq |t^\top \Delta_j|. \tag{A3}$$

Since $\|Y_j\|^2 = \|X_j\|^2 + \|\Delta_j\|^2 + 2X_j^\top \Delta_j$, it follows that $V_n(t) = \sum_{k=1}^6 V_{n,k}(t)$, where

$$V_{n,1}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{\|X_j\|^2 \mathrm{CS}^+(t, X_j) - \mu(t)\}, \quad V_{n,2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \|X_j\|^2 t^\top \Delta_j (\cos(\Gamma_j) - \sin(\Theta_j)),$$
$$V_{n,3}(t) = \frac{2}{\sqrt{n}} \sum_{j=1}^n X_j^\top \Delta_j \mathrm{CS}^+(t, X_j), \quad V_{n,4}(t) = \frac{2}{\sqrt{n}} \sum_{j=1}^n X_j^\top \Delta_j t^\top \Delta_j (\cos(\Gamma_j) - \sin(\Theta_j)),$$
$$V_{n,5}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \|\Delta_j\|^2 \mathrm{CS}^+(t, X_j), \quad V_{n,6}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \|\Delta_j\|^2 t^\top \Delta_j (\cos(\Gamma_j) - \sin(\Theta_j)). \tag{A4}$$

We first show that $V_{n,\ell} = o_{\mathbb{P}}(1)$ if $\ell \in \{4,5,6\}$. As for $V_{n,4}$, notice that, by the Cauchy–Schwarz inequality, $|V_{n,4}(t)| \leq 4\|t\| n^{-1/2} \max_{i=1,\ldots,n} \|X_i\| \sum_{j=1}^n \|\Delta_j\|^2$. Since $\mathbb{E}\|X\|^4 < \infty$, theorem 5.2 of Barndorff-Nielsen (1963) yields $\max_{i=1,\ldots,n} \|X_i\| = o_{\mathbb{P}}(n^{1/4})$. In view of Proposition 3(a), we have $V_{n,4} = o_{\mathbb{P}}(1)$. The same proposition immediately also gives $V_{n,5} = o_{\mathbb{P}}(1)$. Since $|V_{n,6}(t)| \leq 2\|t\| n^{-1/2} \max_{i=1,\ldots,n} \|\Delta_i\| \sum_{j=1}^n \|\Delta_j\|^2$, we have $V_{n,6} = o_{\mathbb{P}}(1)$ in view of Proposition A1(a) and Proposition 3(c).

We now consider $V_{n,2}(t)$. Since $|V_{n,2}(t)| \leq 2\|t\| n^{-1} \sum_{j=1}^n \|X_j\|^2 n^{1/4} \max_{i=1,\ldots,n} \|\Delta_i\| n^{1/4}$, the law of large numbers and Proposition A1(c) show that $V_{n,2} = o_{\mathbb{P}}(n^{1/4})$. In view of (A3) and Proposition A1(c), the error is thus $o_{\mathbb{P}}(1)$ if we replace both Γ_j and Θ_j with $t^\top X_j$. Moreover, plugging $\Delta_j = (S_n^{-1/2} - \mathrm{I}_d) X_j - S_n^{-1/2} \overline{X}_n$ into the definition of $V_{n,2}(t)$, the error is $o_{\mathbb{P}}(1)$ if we replace $S_n^{-1/2} \overline{X}_n$ with $\overline{X}_n$. Recalling (13), we thus obtain

$$V_{n,2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \|X_j\|^2 t^\top \{(S_n^{-1/2} - \mathrm{I}_d) X_j - \overline{X}_n\} \mathrm{CS}^-(t, X_j) + o_{\mathbb{P}}(1). \tag{A5}$$

We now use display (2.13) of Henze and Wagner (1997), according to which $2\sqrt{n}(S_n^{-1/2} - \mathrm{I}_d) = -n^{-1/2} \sum_{j=1}^n (X_j X_j^\top - \mathrm{I}_d) + O_{\mathbb{P}}(n^{-1/2})$. Plugging this expression into (A5) we obtain

$$V_{n,2}(t) = -\frac{1}{2n} \sum_{k=1}^n \|X_k\|^2 \mathrm{CS}^-(t, X_k) t^\top \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j X_j^\top$$

$$-\mathrm{I}_d)X_k - \frac{1}{n}\sum_{k=1}^{n}\|X_k\|^2\mathrm{CS}^-(t,X_k)\ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}t^\top X_j + o_{\mathbb{P}}(1). \tag{A6}$$

In (A6) we now use the fact that $\mathrm{tr}(AB)=\mathrm{tr}(BA)$, where tr denotes trace and AB is a square matrix. Furthermore, the error is $o_{\mathbb{P}}(1)$ if we replace $n^{-1}\sum_{j=1}^{n}\|X_j\|^2\mathrm{CS}^-(t,X_j)$ and $n^{-1}\sum_{j=1}^{n}\|X_j\|^2X_j\mathrm{CS}^-(t,X_j)$ with their almost sure limits $\mathbb{E}[\|X\|^2\mathrm{CS}^-(t,X)] = -\Delta\psi_X^-(t)$ and $\mathbb{E}[\|X\|^2X\mathrm{CS}^-(t,X)] = -\nabla\Delta\psi_X^+(t)$, respectively. We therefore obtain

$$V_{n,2}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{1}{2}\nabla\Delta\psi_X^+(t)(X_jX_j^\top - \mathrm{I}_d)t + \Delta\psi_X^-(t)t^\top X_j\right\} + o_{\mathbb{P}}(1). \tag{A7}$$

In the same way, we proceed with $V_{n,3}(t)$ and, using $\mathbb{E}[X\mathrm{CS}^+(t,X)X^\top] = -H\psi_X^+(t)$ as well as $\mathbb{E}[X\mathrm{CS}^+(t,X)] = -\nabla\psi_X^-(t)$, finally arrive at

$$V_{n,3}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{2X_j^\top\nabla\psi_X^-(t) + X_j^\top H\psi_X^-(t)X_j + \mu(t)\} + o_{\mathbb{P}}(1). \tag{A8}$$

By adding (A4), (A7), and (A8), we have $V_n = \tilde{V}_n + o_{\mathbb{P}}(1)$, where $\tilde{V}_n$ is given in (22). ∎

Proof of Theorem 8. By analogy with $Z_n(t)$, as defined in (17), let $Z_n^0(t) := n^{-1/2}\sum_{j=1}^{n}\|X_j\|^2\{\mathrm{CS}^+(t,X_j) - m(t)\}$, $t\in\mathbb{R}^d$. A straightforward calculation gives $n^{-1/2}(Z_n(t) - Z_n^0(t)) = \sum_{j=1}^{3}U_{n,j}(t)$, where

$$U_{n,1}(t) = \frac{1}{n}\sum_{j=1}^{n}\|X_j\|^2(\mathrm{CS}^+(t,Y_{n,j}) - \mathrm{CS}^+(t,X_j)),$$

$$U_{n,2}(t) = \frac{2}{n}\sum_{j=1}^{n}X_j^\top\Delta_{n,j}\mathrm{CS}^+(t,Y_{n,j}),\quad U_{n,3}(t) = \frac{1}{n}\sum_{j=1}^{n}\|\Delta_{n,j}\|^2\mathrm{CS}^+(t,Y_{n,j}).$$

Since $|\cos(t^\top Y_{n,j}) - \cos(t^\top X_j)| \le \|t\|\ \|\Delta_{n,j}\|$, $|\sin(t^\top Y_{n,j}) - \sin(t^\top X_j)| \le \|t\|\ \|\Delta_{n,j}\|$, the Cauchy–Schwarz inequality gives $|U_{n,1}(t)| \le 2\|t\|(n^{-1}\sum_{j=1}^{n}\|X_j\|^4)^{1/2}(n^{-1}\sum_{j=1}^{n}\|\Delta_{n,j}\|^2)^{1/2}$. By the strong law of large numbers, we have $n^{-1}\sum_{j=1}^{n}\|X_j\|^4 \xrightarrow{\text{a.s.}} \mathbb{E}\|X\|^4$. In view of Proposition 3 b), we thus obtain $\|U_{n,1}\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} 0$. Next, notice that, again by the Cauchy–Schwarz inequality, $|U_{n,2}(t)| \le 4\left(n^{-1}\sum_{j=1}^{n}\|X_j\|^2\right)^{1/2}\left(n^{-1}\sum_{j=1}^{n}\|\Delta_{n,j}\|^2\right)^{1/2}$. Hence, we have $\|U_{n,2}\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} 0$. Finally, $|U_{n,3}(t)| \le 2n^{-1}\sum_{j=1}^{n}\|\Delta_{n,j}\|^2$ which, in view of Proposition 3 b), shows that $\|U_{n,3}\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} 0$. Summarizing, we have

$$\|n^{-1/2}(Z_n(\cdot) - Z_n^0(\cdot))\|_{\mathbb{H}} \xrightarrow{\text{a.s.}} 0. \tag{A9}$$

By the strong law of large numbers in Banach spaces, it follows that $n^{-1/2}Z_n^0(\cdot) \xrightarrow{\text{a.s.}} \mu(\cdot) - m(\cdot)$ as $n\to\infty$ in $\mathbb{H}$. In view of (A9), we thus obtain $\frac{T_{n,a}}{n} = \|n^{-1/2}Z_n(\cdot)\|_{\mathbb{H}}^2 \xrightarrow{\text{a.s.}} \|\mu(\cdot) - m(\cdot)\|_{\mathbb{H}}^2 = \int(\mu(t) - m(t))^2w_a(t)\,\mathrm{d}t$. ∎

Proof of Theorem 9. Starting with (28), we have

$$\begin{aligned}\Delta_a = &\int (\mathbb{E}(\|X\|^2 \mathrm{CS}^+(t,X)))^2 \exp(-a\|t\|^2)\,\mathrm{d}t \\ &- 2\int (d-\|t\|)^2 \mathbb{E}(\|X\|^2 \mathrm{CS}^+(t,X)) \exp\left(-\frac{2a+1}{2}\|t\|^2\right) \mathrm{d}t \\ &+ \int (d-\|t\|^2)^2 \exp(-(a+1)\|t\|^2)\,\mathrm{d}t \\ =:&\ I_{1,a} - I_{2,a} + I_{3,a},\end{aligned}$$

say. Letting X_1,X_2 be independent copies of X, Fubini's theorem, the addition theorems for the sine function and the cosine function, considerations of symmetry and (7) yield

$$I_{1,a} = \left(\frac{\pi}{a}\right)^{d/2} \mathbb{E}\left[\|X_1\|^2\|X_2\|^2 \exp\left(-\frac{\|X_1 - X_2\|^2}{4a}\right)\right].$$

From (8) and (9), we have

$$\begin{aligned}I_{2,a} &= \frac{2(2\pi)^{d/2}}{(2a+1)^{d/2+2}}\ \mathbb{E}\left[\|X\|^2(\|X\|^2 + 2da(2a+1)) \exp\left(-\frac{\|X\|^2}{2(2a+1)}\right)\right], \\ I_{3,a} &= \frac{\pi^{d/2}}{(a+1)^{d/2+2}}\left(a(a+1)d^2 + \frac{d(d+2)}{4}\right)\end{aligned}$$

and thus

$$\begin{aligned}2a\left(\frac{a}{\pi}\right)^{d/2}\Delta_a = &\ 2a\ \mathbb{E}\left[\|X_1\|^2\|X_2\|^2 \exp\left(-\frac{\|X_1 - X_2\|^2}{4a}\right)\right] \\ &- \frac{2(2a)^{d/2+1}}{(2a+1)^{d/2+2}}\mathbb{E}\left[\|X\|^2(\|X\|^2 + 2da(2a+1)) \exp\left(-\frac{\|X\|^2}{2(2a+1)}\right)\right] \\ &+ \frac{2a^{d/2+1}}{(a+1)^{d/2+2}}\left(a(a+1)d^2 + \frac{d(d+2)}{4}\right) \\ =:&\ J_{1,a} - J_{2,a} + J_{3,a},\end{aligned}$$

say. An expansion of the exponential terms, dominated convergence (notice that $\exp(-u) \le 1 - u + u^2$ if $u \ge 0$) and a binomial expansion gives

$$\begin{aligned}J_{1,a} &= 2ad^2 - d\mathbb{E}\|X_1\|^4 + \mathbb{E}(\|X_1\|^2\|X_2\|^2 X_1^\top X_2) + O(a^{-1}), \\ J_{2,a} &= 4ad^2 - d^3 - 2d^2 - d\mathbb{E}\|X_1\|^4 + O(a^{-1}), \\ J_{3,a} &= 2ad^2 - d^3 - 2d^2 + O(a^{-1})\end{aligned}$$

and thus

$$\lim_{a\to\infty} 2a\left(\frac{a}{\pi}\right)^{d/2}\Delta_a = \mathbb{E}(\|X_1\|^2\|X_2\|^2 X_1^\top X_2) = \mathbb{E}(\|X_1\|^2 X_1)^\top \mathbb{E}(\|X_2\|^2 X_2) = \|\mathbb{E}(\|X\|^2 X)\|^2.$$

Notice that the condition $\mathbb{E}\|X\|^6 < \infty$ is not only needed for the existence of the final limit, but it also occurs when dealing with $J_{2,a}$. ■

Proof of Lemma 1. Putting $Y_j = Y_{n,j}$ and $\Delta_j = Y_j - X_j$, we have

$$\cos(t^\top Y_j) = \cos(t^\top X_j) - t^\top \Delta_j \sin(t^\top X_j) + \varepsilon_j(t), \quad \sin(t^\top Y_j) = \sin(t^\top X_j) + t^\top \Delta_j \cos(t^\top X_j) + \eta_j(t),$$

where $|\varepsilon_j(t)|, |\eta_j(t)| \leq \|t\|^2 \|\Delta_j\|^2$ (see Henze and Wagner 1997, p. 8) and thus

$$\mathrm{CS}^{\pm}(t, Y_j) = \mathrm{CS}^{\pm}(t, X_j) \pm t^\top \Delta_j \mathrm{CS}^{\mp}(t, X_j) + \varepsilon_j(t) + \eta_j(t). \tag{A10}$$

To prove (a), notice that $\Psi_{1,n}(t) = n^{-1} \sum_{j=1}^n \mathrm{CS}^+(t, X_j) X_j + R_{1,n}(t)$, where

$$R_{1,n}(t) = \frac{1}{n} \sum_{j=1}^n (t^\top \Delta_j \mathrm{CS}^-(t, X_j) + \varepsilon_j(t) + \eta_j(t)) X_j$$
$$+ \frac{1}{n} \sum_{j=1}^n (\mathrm{CS}^+(t, X_j) + t^\top \Delta_j \mathrm{CS}^-(t, X_j) + \varepsilon_j(t) + \eta_j(t)) \Delta_j.$$

Since $|\mathrm{CS}^{\pm}(t,X_j)| \leq 2$ and $|\varepsilon_j(t) + \eta_j(t)| \leq 2\|t\|^2 \|\Delta_j\|^2$, the Cauchy–Schwarz inequality yields

$$|R_{1,n}(t)| \leq \frac{2\|t\|}{n} \sum_{j=1}^n \|\Delta_j\| \|X_j\| + \frac{2\|t\|^2}{n} \sum_{j=1}^n \|\Delta_j\|^2 \|X_j\|$$
$$+ \frac{2}{n} \sum_{j=1}^n \|\Delta_j\| + \frac{2\|t\|}{n} \sum_{j=1}^n \|\Delta_j\|^2 + \frac{2\|t\|^2}{n} \sum_{j=1}^n \|\Delta_j\|^3.$$

In view of Proposition 4, each summand converges to zero almost surely, which proves (a). The remaining assertions (b), …, (e) are treated similarly. To tackle (b) and (e), one can show negligibility of terms by using the fact that the spectral norm $\|\cdot\|_{\mathrm{sp}}$ of a matrix satisfies $\|xy^\top\|_{\mathrm{sp}} = \|x\| \, \|y\|$, if x,y are (column) vectors in $\mathbb{R}^d$. The condition $\mathbb{E}\|X_1\|^6 < \infty$ is needed for part e), since

$$\|Y_j\|^2 \|Y_j Y_j^\top\|_{\mathrm{sp}} \leq (\|X_j\|^2 + 2\|X_j\| \|\Delta_j\| + \|\Delta_j\|^2)^2$$
$$= \|X_j\|^4 + 4\|X_j\|^3 \|\Delta_j\| + 6\|X_j\|^2 \|\Delta_j\|^2 + 4\|X_j\| \|\Delta_j\|^3 + \|\Delta_j\|^4,$$

and multiplication with $\|\Delta_j\|^2$ (which arises from an expansion of $\mathrm{CS}^+(t,Y_{n,j})$) gives monomials of order 6. ■

Proof of Theorem 11. The proof is similar to the proof of theorem 5 of Henze and Mayer (2020) and will therefore only be sketched. From (26), (23), and (30), we have $\sigma_a^2 = \sum_{i,j=1}^4 \sigma_a^{i,j}$, where

$$\sigma_a^{i,j} = 4 \iint L^{i,j}(s,t) z(s) z(t) w_a(s) w_a(t) \, \mathrm{d}s\mathrm{d}t,$$

and $L^{i,j}(s,t)$ is given in (32). It thus suffices to prove $\hat{\sigma}_{n,a}^{i,j} \xrightarrow{\mathbb{P}} \sigma_a^{i,j}$ for each pair (i,j), where $\hat{\sigma}_{n,a}^{i,j}$ is given in (39). The first step of the proof is to replace $L_n^{i,j}(s,t)$ with $L_{n,0}^{i,j}(s,t)$, which arises from $L_n^{i,j}(s,t)$ given in (34) by throughout replacing $Y_{n,k}$ with X_k in the functions $v_{n,j}(s, Y_{n,k})$, $j \in \{1, \ldots, 4\}$. Notice that this replacement also refers to the quantities $\Psi_{4,n}^-(s)$, $\Psi_{1,n}(s)$, $\Psi_{3,n}^-(s)$ and $\Psi_{2,n}(s)$ that figure in the definition of $v_{n,2}$, $v_{n,3}$ and $v_{n,4}$. Moreover, we replace $z_n(s)$ with $z_{n,0}(s) = n^{-1}\sum_{j=1}^n \mathrm{CS}^+(s,X_j)\|X_j\|^2 - m(s)$. Putting

$$\hat{\sigma}_{n,0,a}^{i,j} = 4\iint L_{n,0}^{i,j}(s,t)z_{n,0}(s)z_{n,0}(t)w_a(s)w_a(t)\,\mathrm{d}s\mathrm{d}t,$$

Fubini's theorem shows that $\hat{\sigma}_{n,0,a}^{i,j} \xrightarrow{\mathbb{P}} \sigma_a^{i,j}$. It thus remains to prove $\hat{\sigma}_{n,a}^{i,j} - \hat{\sigma}_{n,0,a}^{i,j} = o_{\mathbb{P}}(1)$ as $n \to \infty$. To tackle $\hat{\sigma}_{n,a}^{i,j} - \hat{\sigma}_{n,0,a}^{i,j}$, we put $\Lambda_n(s,t) = L_n^{i,j}(s,t)z_n(s)z_n(t) - L_{n,0}^{i,j}(s,t)z_{n,0}(s)z_{n,0}(t)$ and notice that

$$\Lambda_n(s,t) = (L_n^{i,j}(s,t) - L_{n,0}^{i,j}(s,t))z_n(s)z_n(t) + L_{n,0}^{i,j}(s,t)(z_n(s)z_n(t) - z_{n,0}(s)z_{n,0}(t)), \tag{A11}$$

$$z_n(s)z_n(t) - z_{n,0}(s)z_{n,0}(t) = (z_n(s) - z_{n,0}(s))(z_n(t) - z_{n,0}(t))$$

$$+ z_{n,0}(s)(z_n(t) - z_{n,0}(t)) + z_{n,0}(t)(z_n(s) - z_{n,0}(s)). \tag{A12}$$

From (36), we have $|z_n(s)| \le 2d + m(s)$. Moreover, $|z_{n,0}(s)| \le 2n^{-1}\sum_{j=1}^n \|X_j\|^2 + m(s)$. A Taylor expansion shows that $|z_n(s) - z_{n,0}(s)|$ is bounded from above by a finite sum of terms of the type $\|s\|^{\ell} n^{-1}\sum_{j=1}^n \|X_j\|^{\beta}\|\Delta_{n,j}\|^{\gamma}$, where $s \in \{0,1\}$, $\gamma \ge 1$ and $\beta + \gamma \le 3$. Next, by straightforward calculations we obtain that $|L_{n,0}^{i,j}(s,t)|$ is bounded from above by $\|s\|^{\ell}\,\|t\|^m$ times finitely many products of the type $n^{-1}\sum_{j=1}^n \|X_j\|^{\beta}$, where $\ell, m \in \{0,1\}$ and $\beta \in \{1,2,3,4\}$. It now follows from Proposition 4 that the term figuring in (A12), multiplied with $w_a(s)w_a(t)$ and integrated over $\mathbb{R}^d \times \mathbb{R}^d$, converges to zero in probability.

As for the term figuring in (A11), we have $|z_n(s)z_n(t)| \le (2d + m(s))(2d + m(t))$. To tackle $L_n^{i,j}(s,t) - L_{n,0}^{i,j}(s,t)$, we confine ourselves to the case $i = 1, j = 2$, since the other cases can be treated in a similar way. From (35), we have

$$L_n^{1,2}(s,t) = -\frac{1}{2}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^-(s,Y_j)\|Y_j\|^2 Y_j\}\top\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^+(t,Y_j)\|Y_j\|^2 Y_j Y_j^\top s\right\} \tag{A13}$$

$$+\frac{1}{2}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^+(t,Y_j)\|Y_j\|^2\right\}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^-(s,Y_j)\|Y_j\|^2 Y_j^\top s\right\}. \tag{A14}$$

Moreover,

$$L_{n,0}^{1,2}(s,t) = -\frac{1}{2}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^-(s,X_j)\|X_j\|^2 X_j\right\}^\top\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^+(t,X_j)\|X_j\|^2 X_j X_j^\top s\right\} \tag{A15}$$

$$+\frac{1}{2}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^+(t,X_j)\|X_j\|^2\right\}\left\{\frac{1}{n}\sum_{j=1}^n \mathrm{CS}^-(s,X_j)\|X_j\|^2 X_j^\top s\right\}. \tag{A16}$$

Using (A10), some calculations show that, for each $i \in \{1,2\}$, the norm of the difference of the ith curly bracket in (A13) and the ith curly bracket in (A15) is bounded from above by finite sum of terms of the type

$$\|s\|^{\ell}\|t\|^m\,\frac{1}{n}\sum_{j=1}^n \|X_j\|^{\beta}\|\Delta_{n,j}\|^{\gamma}, \tag{A17}$$

where $\ell, m \in \{0,1,2\}$, $\gamma \geq 1$, and $\beta + \gamma \leq 5$. Since the same holds for the norm of the difference of the ith curly bracket in (A14) and the ith curly bracket in (A16), $i \in \{1,2\}$, it follows that $|L_n^{1,2}(s,t) - L_{n,0}^{1,2}(s,t)|$ is bounded from above by a finite sum of terms, which are either products of two terms of type (A17), or a product of a term of type (A17) and one of the terms inside one of the curly brackets in (A15) or (A16). From Proposition 4, it follows that the term figuring in (A11), multiplied with $w_a(s)w_a(t)$ and integrated over $\mathbb{R}^d \times \mathbb{R}^d$, converges to zero in probability. ■