The Canadian Journal of Statistics
Vol. 50, No. 3, 2022, Pages 992–1033
La revue canadienne de statistique

Testing normality in any dimension by Fourier methods in a multivariate Stein equation

Bruno EBNER[1*], Norbert HENZE[1], and David STRIEDER[2]

[1]*Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, D-76133 Karlsruhe, Germany*
[2]*Chair of Mathematical Statistics, Technical University of Munich, Boltzmannstr. 3, D-85748 Garching, Germany*

Key words and phrases: Affine invariance; characteristic function; consistency; multivariate Stein equation; test for multivariate normality; weighted L^2-statistic.

MSC 2020: Primary 62H15; secondary 62G20.

Abstract: We study a novel class of affine-invariant and consistent tests for multivariate normality. The tests are based on a characterization of the standard d-variate normal distribution by way of the unique solution of an initial value problem connected to a partial differential equation, which is motivated by a multivariate Stein equation. The test criterion is a suitably weighted L^2-statistic. We derive the limit distribution of the test statistic under the null hypothesis as well as under contiguous and fixed alternatives to normality. A consistent estimator of the limiting variance under fixed alternatives, as well as an asymptotic confidence interval of the distance of an underlying alternative with respect to the multivariate normal law, is derived. In simulation studies, we show that the tests are strong in comparison with prominent competitors and that the empirical coverage rate of the asymptotic confidence interval converges to the nominal level. We present a real data example and also outline topics for further research. *The Canadian Journal of Statistics* 50: 992–1033; 2022 © 2021 The Authors. The Canadian Journal of Statistics/La revue canadienne de statistique published by Wiley Periodicals LLC on behalf of Statistical Society of Canada.

Résumé: Nous étudions une nouvelle classe de tests de la normalité multivariée qui sont consistants et affines équivariants. Les tests en question reposent sur une caractérisation de la distribution normale standard multivariée, en tant que solution unique d'un problème à valeur initiale associé à une équation aux dérivées partielles qui, elle-même, est motivée par une équation de Stein multivariée. Le critère du test est une statistique L^2 convenablement pondérée. Nous déterminons la distribution limite de la statistique du test sous l'hypothèse nulle et sous des contre-hypothèses fixes et contiguës à la normalité. Nous construisons, d'une part, un estimateur de la variance limite convergent sous des hypothèses alternatives fixes et un intervalle de confiance asymptotique de la distance d'une alternative sous-jacente et une loi normale multivariée. Nos simulations numériques montrent que les tests proposés sont puissants comparativement à d'importants tests existants et que le taux de couverture empirique de l'intervalle de confiance asymptotique converge vers le seuil nominal. Nous présentons un exemple de données réelles et décrivons des questions de recherches ultérieures. *La revue canadienne de statistique* 50: 992–1033; 2022 © 2021 Les auteurs. La revue canadienne de statistique/The Canadian Journal of Statistics, publiée par Wiley Periodicals LLC au nom de la Société statistique du Canada.

* *Corresponding author: Bruno.Ebner@kit.edu*

1. INTRODUCTION

Statistical inference for a dataset starts with assumptions on the underlying stochastic mechanism that determines the generation of the data. In most classical models for multidimensional data, such as multivariate linear regression models or multivariate analysis of variance, the assumption of multivariate normality of the underlying random vectors is inherent. Hence, before making any serious statistical inference, one should check this assumption. To be specific, let $X, X_1, X_2, \dots$ be a sequence of independent and identically distributed (i.i.d.) d-dimensional (column) vectors that are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. As is common in the context of testing for multivariate normality, we make the basic standing assumption that the distribution $\mathbb{P}^X$ of X is absolutely continuous with respect to the d-dimensional Lebesgue measure; see also the discussion before Equation (3). In what follows, we denote by $N_d(\mu, \Sigma)$ the d-variate normal distribution with expectation vector μ and covariance matrix Σ, and we write

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

for the class of all non-degenerate d-variate normal distributions. The unit matrix of order d will be denoted by I_d. The problem at hand is testing the hypothesis

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

based on $X_1, \dots, X_n$ against general alternatives. The purpose of this article is to introduce and study a novel class of affine-invariant and consistent tests based on a partial differential equation (PDE) that determines the characteristic function (CF) of the multivariate standard normal law. We write ∇ for the gradient operator and consider for $f \in L^2(\mathbb{R}^d)$ the initial value problem of the PDE

$$\begin{cases}(t+\nabla)f(t) = 0, & t \in \mathbb{R}^d,\\ f(0) = 1.\end{cases} \tag{1}$$

Note that the multivariate Stein operator $Af(x) = (x - \nabla)f(x)$ is connected to the initial value problem (1) in the following sense: For a centred random vector X with $\mathbb{E}[XX^\intercal] = I_d$, which has a differentiable density with full support $\mathbb{R}^d$, we have $\mathbb{E}[Af(X)] = \mathbb{E}[Xf(X) - \nabla f(X)] = 0$ for each function f with existing derivatives in every direction, and for which all occurring expectations exist, if and only if X has the normal distribution $N_d(0, I_d)$; see Theorem 3.5 in Mijoule, Reinert & Swan (2018) as well as Stein (1981), Liu (1994) and Landsman, Vanduffel & Yao (2013) for more information on the multivariate Stein lemma. Here and in the following, the symbol $^\intercal$ means transposition of column vectors and matrices. In the spirit of the Stein–Tikhomirov method, see Formanov & Formanova (2013) and Arras et al. (2016), and hence using the CFs $\left\{\exp(\mathrm{i}t^\intercal x), t \in \mathbb{R}^d\right\}$ as test functions, a simple calculation shows the equivalence of the Stein equation to the initial value problem in (1). In the case $d = 1$, the same initial value problem was motivated by a fixed point of the zero-bias transform in Ebner (2021). For more information on the zero-bias transform, see Goldstein & Reinert (1997) and Shevtsova (2013).

Theorem 1. *The CF*

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

of the d-variate standard normal distribution $N_d(0, I_d)$ is the only solution of (1).

Proof. If $f \in L^2(\mathbb{R}^d)$ is an arbitrary solution of (1), the product rule yields

$$\nabla\left(\exp\left(\frac{\|t\|^2}{2}\right)f(t)\right) = \exp\left(\frac{\|t\|^2}{2}\right)\big(tf(t) + \nabla f(t)\big) = 0.$$

Considering that $f(0) = 1$, we have $\exp(\|t\|^2/2)f(t) = 1$, and the assertion follows. ■

According to Theorem 1, the CF of the d-variate standard normal distribution is the only CF satisfying $\nabla\psi(t) = -t\psi(t)$. Our test statistic will be based on this equation. To achieve affine invariance of the test statistic with respect to full-rank affine transformations of $X_1, \dots, X_n$, let

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

denote the so-called scaled residuals, where $\overline{X} = n^{-1}\sum_{j=1}^n X_j$ and $S_n := n^{-1}\sum_{j=1}^n \big(X_j - \overline{X}_n\big)\big(X_j - \overline{X}_n\big)^\top$ stand for the sample mean and the sample covariance matrix of $X_1, \dots, X_n$, respectively. The matrix $S_n^{-1/2}$ is the unique symmetric positive definite square root of S_n^{-1}. The almost sure invertibility of S_n follows from the absolute continuity of $\mathbb{P}^X$ and the henceforth tacit assumption $n \geq d + 1$, see Eaton & Perlman (1973). In particular, the condition that $\mathbb{P}(X_1 \in F) = 0$ for each proper flat F of $\mathbb{R}^d$, which follows directly from the absolute continuity of $\mathbb{P}^X$, is necessary and sufficient for the non-singularity with probability 1 of the sample covariance matrix, see p. 715 of Eaton & Perlman (1973). Writing

$$\psi_n(t) = \frac{1}{n}\sum_{j=1}^n \exp\left(\mathrm{i}t^\top Y_{n,j}\right), \quad t \in \mathbb{R}^d \tag{3}$$

for the empirical CF of $Y_{n,1}, \dots, Y_{n,n}$, our test statistic is

$$T_{n,a} = n\int_{\mathbb{R}^d} \|\nabla\psi_n(t) + t\psi(t)\|_{\mathbb{C}}^2\, w_a(t)\,dt. \tag{4}$$

Here, $w_a(t) = \exp\left(-a\|t\|^2\right)$, $a > 0$, is a suitable weight function that depends on a positive parameter a, and $\|\cdot\|_{\mathbb{C}}$ denotes the complex Euclidean vector norm. Rejection of H_0 is for large values of $T_{n,a}$. With this approach, we obtain a flexible class of genuine tests for multivariate normality, all of which are motivated by the result of Theorem 1.

Clearly, we propose a new approach to a well-known and widely studied problem. For a survey of affine-invariant tests of multivariate normality, see Henze (2002), and for recent developments with an emphasis on L^2 type statistics, see Ebner & Henze (2020). We list a short overview of different approaches: Henze & Wagner (1997), Pudelko (2005), Tenreiro (2009) and Dörr, Ebner & Henze (2021a, 2021b) consider tests connected to the empirical CF, while Henze & Jiménez-Gamero (2019), Henze, Jiménez-Gamero & Meintanis (2019) and Henze & Visagie (2020) are based on the empirical moment-generating function. The most classical approach is to consider measures of multivariate skewness and kurtosis; see, e.g., Mardia (1970), Móri, Rohatgi & Székely (1994), Kankainen, Taskinen & Oja (2007) and Doornik & Hansen (2008), although inconsistency of those measures with regard to elliptically symmetric alternatives are known, see Baringhaus & Henze (1991, 1992) and Henze (1994a, 1994b). Generalizations of tests for univariate normality (as in Sürücü, 2006; Villaseñor Alva & González Estrada, 2009; Kim & Park, 2018), the examination of nonlinearity of dependence (see Cox & Small, 1978; Ebner, 2012), canonical correlations (see Thulin, 2014), and the notion of energy (see Székely

& Rizzo, 2005) are other approaches to this testing problem. Note that Bontemps & Meddahi (2005) use the univariate Stein equations to test marginal normal distributional assumptions. Empirical competitive Monte Carlo studies can be found in Voinov et al. (2016) and Ebner & Henze (2020).

The rest of this article unfolds as follows: in Section 2, we give a representation of $T_{n,a}$ that is amenable to computational purposes. Moreover, we derive limits of $T_{n,a}$, after suitable affine transformations, as $a \to \infty$ and $a \to 0$, that hold element-wise on the underlying probability space. Section 3 deals with the limit distribution of $T_{n,a}$ under the null hypothesis, and Section 4 considers the limit behaviour of $T_{n,a}$ both under contiguous and fixed alternatives to H_0. Section 5 presents the results of a simulation study, and Section 6 exhibits a real data example. Section 7 contains a brief summary and indicates topics for further research. For the sake of readability, some of the proofs have been deferred to the Appendix.

Throughout the article, we use the following notation: the symbol $\stackrel{\mathcal{D}}{=}$ means equality in distribution, and $\stackrel{\mathbb{P}}{\longrightarrow}$ and $\stackrel{a.s.}{\longrightarrow}$ stand for convergence in probability and almost sure convergence, respectively. Moreover, $\stackrel{\mathcal{D}}{\longrightarrow}$ is shorthand for convergence in distribution for random elements in whatever space is relevant (which will be clear from the context). If not stated otherwise, each limit refers to $n \to \infty$, and each unspecified integral is over $\mathbb{R}^d$. The stochastic Landau symbols $o_{\mathbb{P}}(1)$ and $O_{\mathbb{P}}(1)$ refer to convergence to zero in probability and stochastic boundedness, respectively.

2. BASIC PROPERTIES OF THE TEST STATISTIC

In this section, we provide some information on the test statistic $T_{n,a}$ defined in (4). The first result shows that $T_{n,a}$ allows for a simple representation that is amenable to computational purposes. Moreover, since this representation shows that $T_{n,a}$ depends on $X_1, \ldots, X_n$ only via $Y_{n,i}^\top Y_{n,j}$, $i, j \in \{1, \ldots, n\}$, the statistic $T_{n,a}$ is affine-invariant.

Theorem 2. *We have*

$$T_{n,a} = n\left(\frac{\pi}{a+1}\right)^{\frac{d}{2}} \frac{d}{2(a+1)} - 2\left(\frac{2\pi}{2a+1}\right)^{\frac{d}{2}} \sum_{j=1}^{n} \frac{\|Y_{n,j}\|^2}{2a+1} \exp\left(-\frac{\|Y_{n,j}\|^2}{4a+2}\right)$$
$$+ \frac{1}{n}\left(\frac{\pi}{a}\right)^{\frac{d}{2}} \sum_{i,j=1}^{n} Y_{n,i}^\top Y_{n,j} \exp\left(-\frac{\|Y_{n,i} - Y_{n,j}\|^2}{4a}\right). \tag{5}$$

Note that this representation is implemented in the `R` package `mnt`, see Butsch & Ebner (2020). The proof of Theorem 2 is given in the Appendix.

We now consider the element-wise limits (on the underlying probability space) of $T_{n,a}$ for fixed n as $a \to \infty$ and $a \to 0$. It will be seen that the class of tests based on $T_{n,a}$ is "closed at the boundaries" $a \to \infty$ and $a \to 0$ in the sense that, after suitable affine transformations, there are well-defined "limit statistics." Our first result refers to the limit $a \to \infty$.

Theorem 3. *Element-wise on the underlying probability space* $(\Omega, \mathcal{A}, \mathbb{P})$, *we have*

$$\lim_{a\to\infty} \frac{a^{\frac{d}{2}+2}}{n\pi^{\frac{d}{2}}} 16 T_{n,a} = \tilde{b}_{1,d} + 2b_{1,d}. \tag{6}$$

Here, $b_{1,d} = n^{-2} \sum_{i,j=1}^{n} \left(Y_{n,i}^\top Y_{n,j}\right)^3$ *is Mardia's celebrated measure of multivariate skewness, see Mardia (1970), and* $\tilde{b}_{1,d} = n^{-2} \sum_{i,j=1}^{n} Y_{n,i}^\top Y_{n,j} \|Y_{n,i}\|^2 \|Y_{n,j}\|^2$ *is a measure of multivariate skewness introduced by Móri, Rohatgi & Székely (1994).*

Proof. Invoking (5), it follows that

$$\frac{a^{\frac{d}{2}+2}}{n\pi^{\frac{d}{2}}}T_{n,a} = \left(\frac{a}{a+1}\right)^{\frac{d}{2}+1}\frac{ad}{2} - \frac{a}{n}\left(\frac{a}{a+\frac{1}{2}}\right)^{\frac{d}{2}+1}\sum_{j=1}^{n}\|Y_{n,j}\|^2\exp\left(-\frac{\|Y_{n,j}\|^2}{4a+2}\right)$$
$$+\frac{a^2}{n^2}\sum_{i,j=1}^{n}Y_{n,i}^{\top}Y_{n,j}\exp\left(-\frac{\|Y_{n,i}-Y_{n,j}\|^2}{4a}\right)$$
$$=: A_n - B_n + C_n$$

(say). We now use

$$\left(\frac{a}{a+1}\right)^{\frac{d}{2}+1} = \left(1+\frac{1}{a}\right)^{-\frac{d}{2}-1} = 1-\left(\frac{d}{2}+1\right)\frac{1}{a}+O(a^{-2}) \tag{7}$$

as $a \to \infty$ and

$$\exp(-x) = 1-x+\frac{1}{2}x^2+O(x^3) \tag{8}$$

as $x \to 0$, and we employ the identities $\sum_{j=1}^{n} Y_{n,j} = 0$, $\sum_{j=1}^{n}\|Y_{n,j}\|^2 = nd$ as well as

$$\sum_{i,j=1}^{n}Y_{n,i}^{\top}Y_{n,j}\|Y_{n,i}-Y_{n,j}\|^2 = -2\sum_{i,j=1}^{n}\left(Y_{n,i}^{\top}Y_{n,j}\right)^2 = -2n^2d,$$
$$\sum_{i,j=1}^{n}Y_{n,i}^{\top}Y_{n,j}\|Y_{n,i}-Y_{n,j}\|^4 = 2n^2\tilde{b}_{1,d}+4n^2b_{1,d}-8\sum_{i,j=1}^{n}\left(Y_{n,i}^{\top}Y_{n,j}\right)^2\|Y_{n,j}\|^2,$$
$$\sum_{i,j=1}^{n}\left(Y_{n,i}^{\top}Y_{n,j}\right)^2\|Y_{n,j}\|^2 = n\sum_{j=1}^{n}\|Y_{n,j}\|^4$$

to obtain $A_n = ad/2 - d^2/4 - d/2 + o(1)$ as $a \to \infty$. Likewise

$$B_n = \frac{1}{n}\left(a-\left(\frac{d}{2}+1\right)\frac{1}{2}\right)\sum_{j=1}^{n}\|Y_{n,j}\|^2\left(1-\frac{\|Y_{n,j}\|^2}{4a+2}\right)+o(1)$$
$$= \left(da-\frac{d^2}{4}-\frac{d}{2}\right)-\frac{1}{4n}\sum_{j=1}^{n}\|Y_{n,j}\|^4+o(1),$$
$$C_n = \frac{a^2}{n^2}\sum_{i,j=1}^{n}Y_{n,i}^{\top}Y_{n,j}\left(1-\frac{\|Y_{n,i}-Y_{n,j}\|^2}{4a}+\frac{\|Y_{n,i}-Y_{n,j}\|^4}{32a^2}\right)+o(1)$$
$$= \frac{da}{2}+\frac{1}{16}\left(\tilde{b}_{1,d}+2b_{1,d}-\frac{4}{n}\sum_{j=1}^{n}\|Y_{n,j}\|^4\right)+o(1).$$

Upon combining, the assertion follows. ■

Notice that the right-hand side of (6) is a linear combination of two time-honoured measures of multivariate skewness. Notably, the same linear combination shows up not only for the class of Baringhaus–Henze–Epps–Pulley (BHEP) tests (see Theorem 2.1 of Henze, 1997), but also as a limit of a related test statistic in connection with a test for multivariate normality based on a PDE for the *moment-generating function* of the normal distribution, see Henze & Visagie (2020).

Regarding the limit of $T_{n,a}$ as $a \to 0$, we have the following result:

Theorem 4. *Element-wise on the underlying probability space, we have*

$$\lim_{a\to 0} \frac{1}{na^{\frac{d}{2}}}\left(\left(\frac{a}{\pi}\right)^{\frac{d}{2}} T_{n,a} - d\right) = \frac{d}{2} - 2^{\frac{d}{2}+1}\frac{1}{n}\sum_{j=1}^{n} \|Y_{n,j}\|^2 \exp\left(-\frac{\|Y_{n,j}\|^2}{2}\right).$$

Proof. From the representation (5), it follows that

$$\begin{aligned}\frac{T_{n,a}}{\pi^{\frac{d}{2}}} &= \frac{nd}{2(a+1)^{\frac{d}{2}+1}} - \left(\frac{2}{2a+1}\right)^{\frac{d}{2}+1}\sum_{j=1}^{n} \|Y_{n,j}\|^2 \exp\left(-\frac{\|Y_{n,j}\|^2}{4a+2}\right) \\ &\quad + \frac{1}{na^{\frac{d}{2}}}\sum_{i,j=1}^{n} Y_{n,i}^{\top} Y_{n,j} \exp\left(-\frac{\|Y_{n,i}-Y_{n,j}\|^2}{4a}\right) \\ &= A_{n,a} - B_{n,a} + C_{n,a}\end{aligned}$$

(say). Now, $\lim_{a\to 0} A_{n,a} = nd/2$ and $\lim_{a\to 0} B_{n,a} = 2^{\frac{d}{2}+1}\sum_{j=1}^{n} \|Y_{n,j}\|^2 \exp\left(-\|Y_{n,j}\|^2/2\right)$, element-wise on the underlying probability space. To tackle $C_{n,a}$, the relation $\sum_{j=1}^{n} \|Y_{n,j}\|^2 = nd$ yields

$$\begin{aligned}C_{n,a} &= \frac{1}{na^{\frac{d}{2}}}\sum_{j=1}^{n} \|Y_{n,j}\|^2 + \frac{1}{na^{\frac{d}{2}}}\sum_{i\neq j}^{n} Y_{n,i}^{\top} Y_{n,j} \exp\left(-\frac{\|Y_{n,i}-Y_{n,j}\|^2}{4a}\right) \\ &= \frac{d}{a^{\frac{d}{2}}} + \frac{1}{na^{\frac{d}{2}}}\sum_{i\neq j}^{n} Y_{n,i}^{\top} Y_{n,j} \exp\left(-\frac{\|Y_{n,i}-Y_{n,j}\|^2}{4a}\right),\end{aligned}$$

and the assertion follows. ■

Interestingly, Theorem 4 means that for (very) small values of a, rejection of H_0 for large values of $T_{n,a}$ is essentially equivalent to the rejection of H_0 for *small* values of

$$\frac{1}{n}\sum_{j=1}^{n} \|Y_{n,j}\|^2 \mathrm{e}^{-\|Y_{n,j}\|^2/2}.$$

This statistic, upon expanding the exponential function, comprises even powers of $\|Y_{n,j}\|$ and is thus related to Mardia's measure of multivariate kurtosis, which is defined by $b_{2,d} = n^{-1}\sum_{j=1}^{n} \|Y_{n,j}\|^4$, see Mardia (1970).

3. THE LIMIT NULL DISTRIBUTION

In this section we derive the limit distribution of $T_{n,a}$ under the hypothesis H_0. Because of affine invariance, we assume without loss of generality that X has the standard normal distribution $N_d(0, I_d)$ in what follows. The starting point is an alternative representation of $T_{n,a}$, namely

$$T_{n,a} = \int \|Z_n(t)\|^2 w_a(t)\, dt, \tag{9}$$

where

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Y_{n,j} \left(\cos\left(t^\top Y_{n,j}\right) + \sin\left(t^\top Y_{n,j}\right) \right) - t\psi(t) \right). \tag{10}$$

This assertion follows from straightforward calculations using

$$\int \cos\left(t^\top Y_{n,j}\right) \sin\left(t^\top Y_{n,i}\right) w_a(t)\, dt = 0, \quad \int \cos\left(t^\top Y_{n,j}\right) t^\top Y_{n,j} w_a(t)\, dt = 0. \tag{11}$$

Writing $L^2 := L^2\left(\mathbb{R}^d, \mathcal{B}^d, w_a(t)dt\right)$ for the separable Hilbert space of (equivalence classes of) functions $f : \mathbb{R}^d \to \mathbb{R}$ that are square-integrable with respect to $w_a(t)dt$, we regard Z_n as a random element of the Hilbert space $\mathbb{H} = L^2 \otimes \cdots \otimes L^2$. Putting $f = (f_1, \dots, f_d)$, $g = (g_1, \dots, g_d)$, the space $\mathbb{H}$ is equipped with the inner product $\langle f, g \rangle_{\mathbb{H}} := \langle f_1, g_1 \rangle_{L^2} + \cdots + \langle f_d, g_d \rangle_{L^2}$ and the norm $\|f\|_{\mathbb{H}} = \langle f, f \rangle_{\mathbb{H}}^{1/2}$. Notice that we have

$$T_{n,a} = \int \|Z_n(t)\|^2 w_a(t)\, dt = \|Z_n\|_{\mathbb{H}}^2.$$

The main theorem of this section is as follows:

Theorem 5. *Under H_0, there is a centred Gaussian random element Z of $\mathbb{H}$ having covariance matrix kernel*

$$\begin{aligned} K(s,t) &= \left(I_d - (s-t)(s-t)^\top\right)\psi(s-t) \\ &\quad + \left((s-t)s^\top + (t-s)t^\top - I_d + s^\top t(ss^\top + tt^\top - st^\top - I_d) - \frac{(s^\top t)^2}{2} st^\top \right) \psi(s)\psi(t), \end{aligned} \tag{12}$$

$s, t \in \mathbb{R}^d$, such that $Z_n \xrightarrow{\mathcal{D}} Z$ in $\mathbb{H}$, where Z_n is the random element defined in (10).

Since the proof of Theorem 5 is long and tedious, it is deferred to the Appendix. A crucial role will be played by the quantities

$$\Delta_{n,j} = Y_{n,j} - X_j = \left(S_n^{-\frac{1}{2}} - I_d \right) X_j - S_n^{-\frac{1}{2}} \overline{X}_n, \quad j = 1, \dots, n. \tag{13}$$

From Theorem 5 and the continuous mapping theorem, we obtain the following result:

Corollary 6. *Under H_0, we have*

$$T_{n,a} \xrightarrow{\mathcal{D}} \|Z\|_{\mathbb{H}}^2 = \int \|Z(t)\|^2 w_a(t)\, dt.$$

It is well known that the distribution of $T_{\infty,a} := \|Z\|_{\mathbb{H}}^2$ is that of $T_{\infty,a} \overset{\mathcal{D}}{=} \sum_{j=1}^{\infty} \lambda_j(a) N_j^2$, where $N_1, N_2, \dots$ is a sequence of i.i.d. standard normal random variables, and $\lambda_1(a), \lambda_2(a), \dots$ are the positive eigenvalues associated with the integral operator

$$\mathbb{K}f(s) := \int K(s,t) f(t) w_a(t)\, dt, \quad s \in \mathbb{R}^d, \tag{14}$$

$f \in \mathbb{H}$. Because of the complexity of $K(s,t)$, we did not succeed in obtaining closed-form expressions for these eigenvalues. In our simulation study presented in Section 5, we use approximate critical values for $T_{n,a}$ that have been obtained by way of simulations. Some information on the limit null distribution, however, is given by the following result:

Theorem 7. *We have*

$$\mathbb{E}[T_{\infty,a}] = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} d - \left(\frac{\pi}{a+1}\right)^{\frac{d}{2}} \frac{\left(16a^3 + (8d+48)a^2 + (12d+40)a + d^2 + 10d + 16\right)d}{16(a+1)^3}.$$

Proof. From Fubini's theorem, it follows that $\mathbb{E}[T_{\infty,a}] = \int \mathbb{E}\|Z(t)\|^2 w_a(t)\,dt$. Moreover, writing tr for trace, we have

$$\begin{aligned}\mathbb{E}\|Z(t)\|^2 &= \mathbb{E}[Z(t)^{\intercal}Z(t)] = \mathrm{tr}\left(\mathbb{E}[Z(t)Z(t)^{\intercal}]\right)\\ &= \mathrm{tr}\left(K(t,t)\right) = d - \left(d + d\|t\|^2 - \|t\|^4 + \frac{\|t\|^6}{2}\right)\exp\left(-\|t\|^2\right).\end{aligned}$$

Since

$$\int \|t\|^4 \mathrm{e}^{-a\|t\|^2}\,dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \frac{d}{4a^2}(d+2) \quad \text{and} \quad \int \|t\|^6 \mathrm{e}^{-a\|t\|^2}\,dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \frac{d}{8a^3}(d^2+6d+8),$$

the assertion follows by straightforward computations. ■

In the univariate case, which is deliberately included in our study, we have been able to calculate the first four cumulants of $T_{\infty,a}$. By the methods presented in Chapter 5 of Shorack & Wellner (1986), the mth cumulant of $T_{\infty,a}$ is derived by

$$\kappa_m(a) = 2^{m-1}(m-1)! \int_{\mathbb{R}} h_m(t,t) w_a(t)\,dt.$$

Here, $h_1(s,t) = K(s,t)$, and $h_m(s,t) := \int_{\mathbb{R}} h_{m-1}(s,u)K(u,t)w_a(u)\,du$ if $m \geq 2$. In order to calculate $\kappa_m(a)$, $m \in \{1,2,3,4\}$, we used the computer algebra system Maple, see Maplesoft (2019).

For the first two cumulants we obtain

$$\begin{aligned}\kappa_1(a) &= \int_{\mathbb{R}} \left(1 - \left(1 + t^2 - t^4 + \frac{t^6}{2}\right)\exp\left(-t^2\right)\right)\exp\left(-at^2\right)dt\\ &= \frac{\left(-16a^3 - 56a^2 - 52a - 27\right)\sqrt{\frac{\pi}{a+1}} + 16\sqrt{\frac{\pi}{a}}(a+1)^3}{16(a+1)^3}\end{aligned}$$

and

$$\begin{aligned}\kappa_2(a) = &\frac{7260811\pi}{8\,(a+2)^{5/2}\left(4a^2+8a+3\right)^{5/2}\sqrt{a}\,(2a+3)^2\,(a+1)^7}\\ &\times\left(\left(\left(\frac{1024}{7260811}a^{\frac{29}{2}} + \frac{15360}{7260811}a^{\frac{27}{2}} + \frac{108032}{7260811}a^{\frac{25}{2}} + \frac{473856}{7260811}a^{\frac{23}{2}}\right.\right.\right.\\ &+ \frac{1449216}{7260811}a^{\frac{21}{2}} + \frac{3263232}{7260811}a^{\frac{19}{2}} + \frac{5559908}{7260811}a^{\frac{17}{2}} + \frac{7254348}{7260811}a^{\frac{15}{2}}\end{aligned}$$

$$+ a^{\frac{13}{2}} + \frac{5535906}{7260811}a^{\frac{11}{2}} + \frac{160113}{367636}a^{\frac{9}{2}} + \frac{5253759}{29043244}a^{\frac{7}{2}} + \frac{6017409}{116172976}a^{\frac{5}{2}}$$

$$+ \frac{266733}{29043244}a^{\frac{3}{2}} + \frac{22113}{29043244}\sqrt{a}\Bigg)\sqrt{a+2}$$

$$+\frac{1024\,(a+3/2)^4\,(a+1)^7\,(a+1/2)^2\,\left(a^2+2a+3\right)}{7260811}\Bigg)\sqrt{4a^2+8a+3}$$

$$-\frac{51420992\sqrt{a+2}}{7260811}\Bigg(\frac{64}{803453}a^{\frac{31}{2}} + \frac{1024}{803453}a^{\frac{29}{2}} + \frac{1104}{114779}a^{\frac{27}{2}} + \frac{36544}{803453}a^{\frac{25}{2}}$$

$$+\frac{121054}{803453}a^{\frac{23}{2}} + \frac{297018}{803453}a^{\frac{21}{2}} + \frac{556163}{803453}a^{\frac{19}{2}} + \frac{807017}{803453}a^{\frac{17}{2}} + \frac{912747}{803453}a^{\frac{15}{2}} + a^{\frac{13}{2}}$$

$$+\frac{545801}{803453}a^{\frac{11}{2}} + \frac{281319}{803453}a^{\frac{9}{2}} + \frac{106779}{803453}a^{\frac{7}{2}} + \frac{28293}{803453}a^{\frac{5}{2}} + \frac{96}{16397}a^{\frac{3}{2}} + \frac{372}{803453}\sqrt{a}\Bigg)\Bigg).$$

The formulae for $\kappa_3(a)$ and $\kappa_4(a)$ are known but are too extensive to be stated here explicitly. From these cumulants, we calculate the expectation, the variance, the skewness β_1 and the kurtosis β_2 of $T_{\infty,a}$ for the case $d = 1$ (see Table 1), since

$$\mathbb{E}[T_{\infty,a}] = \kappa_1(a), \quad \mathrm{Var}[T_{\infty,a}] = \kappa_2(a), \quad \beta_1(a) = \frac{\kappa_3(a)}{\kappa_2(a)^{3/2}}, \quad \beta_2(a) = 3 + \frac{\kappa_4(a)}{\kappa_2(a)^2}.$$

Analogously to Henze (1990) and Ebner (2021), we can now approximate the distribution of $T_{\infty,a}$ by that of a member of the system of Pearson distributions which has the same first four moments as $T_{\infty,a}$. To this end, we used the statistical software `R`, see R Core Team (2019), and the package `PearsonDS`, see Becker & Klößner (2017). Table 2 shows the quantiles of the fitted Pearson distribution, which serve as approximations to the corresponding quantiles of the distribution of $T_{\infty,a}$. Here, the symbol $\star$ stands for negative values of the approximate quantiles. These are omitted, since $T_{\infty,a}$ is always positive and the fit of the Pearson family having support on $\mathbb{R}$ is obviously not suited to approximate the lower quantiles for $a = 10$.

4. LIMIT BEHAVIOUR UNDER ALTERNATIVES

In this section, we assume that H_0 does not hold, and we will derive limit distributions for $T_{n,a}$ both under contiguous and fixed alternatives to H_0. To define the setting for a triangular array

TABLE 1: Expectation, variance, skewness and kurtosis of $T_{\infty,a}$, $d = 1$.

a	0.1	0.5	1	2	5	10
$\mathbb{E}[T_{\infty,a}]$	3.0040	0.6574	0.2939	0.1092	0.0207	0.0047
$\mathrm{Var}[T_{\infty,a}]$	2.8028	0.2686	0.0742	0.0133	0.0006	0.0000
$\beta_1(a)$	1.3737	1.9098	2.1996	2.4619	2.7090	2.7938
$\beta_2(a)$	6.0366	8.8662	10.7047	12.5510	14.3071	19.4464

TABLE 2: Approximate quantiles of $T_{\infty,a}$ in the case $d = 1$.

	a					
q	0.1	0.5	1	2	5	10
0.01	0.6857	0.0903	0.0331	0.0110	0.0018	⋆
0.05	0.9970	0.1299	0.0435	0.0130	0.0020	⋆
0.1	1.2382	0.1712	0.0573	0.0165	0.0023	⋆
0.5	2.6510	0.5137	0.2091	0.0700	0.0115	0.0030
0.9	5.2211	1.3283	0.6405	0.2529	0.0511	0.0119
0.95	6.2138	1.6743	0.8329	0.3384	0.0705	0.0162
0.99	8.4485	2.4904	1.2956	0.5470	0.1182	0.0275

of contiguous alternatives, we assume that, for each $n \geq d+1$, $X_{n,1}, \ldots, X_{n,n}$ are i.i.d. d-variate random vectors having Lebesgue density

$$f_n(x) = \varphi(x)\left(1 + \frac{g(x)}{\sqrt{n}}\right), \; x \in \mathbb{R}^d.$$

Here, $\varphi(x) = (2\pi)^{-d/2}\exp(-\|x\|^2/2)$, $x \in \mathbb{R}^d$, is the density of the distribution $N_d(0, I_d)$, and g is a bounded measurable function satisfying $\int g(x)\varphi(x)\,dx = 0$. Notice that f_n is non-negative for sufficiently large n because of the boundedness of g. To derive the limit distribution of $T_{n,a}$ under this sequence of alternatives, we employ the representation (9), which comprises the random element Z_n as defined in (10). For repeated later use, we define

$$\mathrm{CS}^+(s,t) = \cos(s^\intercal t) + \sin(s^\intercal t), \quad \mathrm{CS}^-(s,t) = \cos(s^\intercal t) - \sin(s^\intercal t), \quad s,t \in \mathbb{R}^d. \tag{15}$$

Theorem 8. *Under the sequence of alternatives $(X_{n,1}, \ldots, X_{n,n})_{n\geq d+1}$, we have*

$$Z_n \xrightarrow{\mathcal{D}} Z + c \text{ in } \mathbb{H}.$$

Here, Z_n is defined in (10), Z is the centred Gaussian random element of $\mathbb{H}$ figuring in Theorem 5, and the shift function $c(\cdot)$ is given by

$$c(t) = \int Z^{**}(x,t)g(x)\varphi(x)\,dx, \quad t \in \mathbb{R}^d, \tag{16}$$

where

$$Z^{**}(x,t) = x\mathrm{CS}^+(t,x) - \left(t + x + \left(2I_d - tt^\intercal\right)\frac{1}{2}\left(xx^\intercal - I_d\right)t - t^\intercal xt\right)\psi(t), \;\; x,t \in \mathbb{R}^d.$$

Proof. We write λ^d for the d-dimensional Lebesgue measure, and we put $\mathbb{P}^{(n)} := \otimes\left(\varphi\lambda^d\right)$, $Q^{(n)} := \otimes\left(f_n\lambda^d\right)$. Furthermore, let $L_n := dQ^{(n)}/d\mathbb{P}^{(n)}$. The boundedness of g and a Taylor expansion then give

$$\log(L_n(X_{n,1},\dots,X_{n,n})) = \sum_{j=1}^{n}\log\left(1+\frac{g(X_{n,j})}{\sqrt{n}}\right)$$
$$= \sum_{j=1}^{n}\left(\frac{g(X_{n,j})}{\sqrt{n}} - \frac{g(X_{n,j})^2}{2n}\right) + o_{\mathbb{P}^{(n)}}(1). \quad (17)$$

In the following, we write $\sigma^2 = \int g(x)^2\varphi(x)\,dx < \infty$. Since, under $\mathbb{P}^{(n)}$, expectation and variance of the sum figuring in (17) converge to $-\sigma^2/2$ and σ^2, respectively, the Lindeberg–Feller central limit theorem and Slutsky's lemma yield

$$\log(L_n) \xrightarrow{\mathcal{D}} N\left(-\frac{\sigma^2}{2},\sigma^2\right) \text{ under } \mathbb{P}^{(n)}. \quad (18)$$

Notice that the boundedness of g ensures the validity of the Lindeberg condition. In view of Le Cam's first lemma (see, e.g., Li & Babu, 2019, p. 297), the probability measures $Q^{(n)}$ and $\mathbb{P}^{(n)}$ are mutually contiguous. According to Theorem 5, the auxiliary process Z_n^* introduced in (A4) is tight under $\mathbb{P}^{(n)}$ and thus, in view of contiguity, also under $Q^{(n)}$. Let $\{e_k,\ k\geq 1\}$, be an arbitrary complete orthonormal system of $\mathbb{H}$. It remains to show that, for each $\ell\geq 1$, we have $\Pi_\ell(Z_n)\xrightarrow{\mathcal{D}}\Pi_\ell(Z+c)$ under $Q^{(n)}$, where Π_ℓ denotes the orthogonal projection onto the linear subspace of $\mathbb{H}$ spanned by $e_1,\dots,e_\ell$. We first consider

$$\Pi_\ell\left(Z_n^*\right) = \sum_{j=1}^{\ell}\langle Z_n^*, e_j\rangle_{\mathbb{H}} e_j,$$

where Z_n^* is given in (A4), with the only difference that X_j is replaced throughout with $X_{n,j}$. In view of Theorem 5, the asymptotic distribution of Z_n^* under $\mathbb{P}^{(n)}$ is a centred Gaussian with a covariance operator $\mathbb{K}$ given by the covariance matrix kernel $K(s,t)$, whence $\langle Z_n^*, e_j\rangle_{\mathbb{H}}\xrightarrow{\mathcal{D}} N\big(0,\left\langle\mathbb{K}e_j,e_j\right\rangle_{\mathbb{H}}\big)$ under $\mathbb{P}^{(n)}$. In view of (18), we have

$$\left(\left\langle Z_n^*,e_1\right\rangle_{\mathbb{H}},\dots,\left\langle Z_n^*,e_\ell\right\rangle_{\mathbb{H}},\log(L_n)\right)^\top \xrightarrow{\mathcal{D}} N_{\ell+1}\left(\left(0,\dots,0,-\sigma^2/2\right)^\top,\begin{bmatrix}\Sigma & \tilde{c}\\ \tilde{c}^\top & \sigma^2\end{bmatrix}\right)$$

under $\mathbb{P}^{(n)}$ for each $\ell\geq 1$. Here, $\Sigma := \left(\langle\mathbb{K}e_i,e_j\rangle_{\mathbb{H}}\right)_{1\leq i,j\leq\ell}\in\mathbb{R}^{\ell\times\ell}$, and $\tilde{c}=\left(\tilde{c}_1,\dots,\tilde{c}_\ell\right)^\top\in\mathbb{R}^\ell$, where, by Fubini's theorem, $\tilde{c}_j := \lim_{n\to\infty}\mathbb{E}\left[\left\langle Z_n^*,e_j\right\rangle_{\mathbb{H}},\log(L_n)\right] = \langle c,e_j\rangle_{\mathbb{H}}$, and c is given in (16). According to Le Cam's third lemma (see, e.g., Li & Babu, 2019, p. 300), it follows that $\left(\left\langle Z_n^*,e_1\right\rangle_{\mathbb{H}},\dots,\left\langle Z_n^*,e_\ell\right\rangle_{\mathbb{H}}\right)^\top\xrightarrow{\mathcal{D}} N_\ell(\tilde{c},\Sigma)$ under $Q^{(n)}$. Since, for the centred Gaussian random element figuring in Theorem 5 we have

$$\left(\langle Z+c,e_1\rangle_{\mathbb{H}},\dots,\langle Z+c,e_\ell\rangle_{\mathbb{H}}\right)^\top \overset{\mathcal{D}}{=} N_\ell(\tilde{c},\Sigma),$$

it follows that

$$\left(\left\langle Z_n^*,e_1\right\rangle_{\mathbb{H}},\dots,\left\langle Z_n^*,e_\ell\right\rangle_{\mathbb{H}}\right)^\top \xrightarrow{\mathcal{D}} \left(\left\langle Z+c,e_1\right\rangle_{\mathbb{H}},\dots,\langle Z+c,e_\ell\rangle_{\mathbb{H}}\right)^\top \quad (19)$$

under $\mathbb{Q}^{(n)}$. Now, let $\Psi : \mathbb{R}^{\ell} \to \mathbb{H}$ be defined by $\Psi(x) := \sum_{j=1}^{\ell} x_j e_j$, $x = (x_1, \dots, x_{\ell})^\intercal$. The continuous mapping theorem and (19) then yield

$$\Pi_{\ell}(Z_n^*) = \Psi\Big(\big(\langle Z_n^*, e_1\rangle_{\mathbb{H}}, \dots, \langle Z_n^*, e_{\ell}\rangle_{\mathbb{H}}\big)^\intercal\Big) \xrightarrow{\mathcal{D}} \Psi\Big(\big(\langle Z+c, e_1\rangle_{\mathbb{H}}, \dots, \langle Z+c, e_{\ell}\rangle_{\mathbb{H}}\big)^\intercal\Big)$$
$$= \Pi_{\ell}(Z+c)$$

under $\mathbb{Q}^{(n)}$. In view of the tightness of Z_n^* under $Q^{(n)}$, we conclude $Z_n^* \xrightarrow{\mathcal{D}} Z + c$ under $Q^{(n)}$. The assertion now follows from Slutsky's lemma since, in view of (A6) and (A7), $\|Z_n - Z_n^*\|_{\mathbb{H}}$ is asymptotically negligible under $\mathbb{P}^{(n)}$ and thus, because of contiguity, also under $Q^{(n)}$. ■

As a corollary, we have the following result:

Corollary 9. *Under the conditions of Theorem 8, we have*

$$T_{n,a} \xrightarrow{\mathcal{D}} \|Z + c\|_{\mathbb{H}}^2 = \int \big\|Z(t) + c(t)\big\|^2 w_a(t)\, dt.$$

We now consider fixed alternatives to H_0, and we suppose that the underlying distribution, in addition to being absolutely continuous, satisfies $\mathbb{E}\|X\|^4 < \infty$. In view of affine invariance, we assume $\mathbb{E}[X] = 0$ and $\mathbb{E}[XX^\intercal] = I_d$. Our first result is a strong limit of $T_{n,a}/n$ as $n \to \infty$.

Theorem 10. *If $\mathbb{E}\|X\|^2 < \infty$, we have*

$$\frac{T_{n,a}}{n} \xrightarrow{a.s.} \Delta_a,$$

where

$$\Delta_a := \int \|\mu(t) - t\psi(t)\|^2 w_a(t)\, dt \tag{20}$$

and $\mu(t) = \mathbb{E}[X\mathrm{CS}^+(t, X)]$.

Proof. Invoking (9), we have $n^{-1}T_{n,a} = \|n^{-1/2}Z_n\|_{\mathbb{H}}^2$, where Z_n is given in (10). Putting $Z_n^0(t) = n^{-1/2}\sum_{j=1}^n \big(X_j\mathrm{CS}^+(t, X_j) - t\psi(t)\big)$, the strong law of large numbers in Hilbert spaces yields $\|n^{-1/2}Z_n^0\|_{\mathbb{H}}^2 \xrightarrow{a.s.} \Delta_a$, and thus it remains to prove $\big\|n^{-1/2}\big(Z_n - Z_n^0\big)\big\|_{\mathbb{H}} \xrightarrow{a.s.} 0$. To this end, notice that

$$\frac{1}{\sqrt{n}}\big(Z_n(t) - Z_n^0(t)\big) = \frac{1}{n}\sum_{j=1}^n \Big(X_j\big(\mathrm{CS}^+\big(t, Y_{n,j}\big) - \mathrm{CS}^+\big(t, X_j\big)\big) + \Delta_{n,j}\mathrm{CS}^+\big(t, Y_{n,j}\big)\Big),$$

where $\Delta_{n,j}$ is defined in (13). Since $\mathrm{CS}^+\big(t, Y_{n,j}\big) = \mathrm{CS}^+(t, X_j) + \varepsilon_{n,j}(t) + \eta_{n,j}(t)$, where $\max(|\varepsilon_{n,j}(t)|, |\eta_{n,j}(t)|) \le \|t\|\|\Delta_{n,j}\|$, it follows that

$$\left\|\frac{1}{n}\sum_{j=1}^n X_j\big(\mathrm{CS}^+\big(t, Y_{n,j}\big) - \mathrm{CS}^+(t, X_j)\big)\right\| \le \frac{2}{n}\sum_{j=1}^n \|X_j\|\|t\|\|\Delta_{n,j}\|. \tag{21}$$

Using $\|\Delta_{n,j}\| \leq \|S_n^{-1/2} - I_d\|_2 \|X_j\| + \|S_n^{-1/2}\|_2 \|\overline{X}_n\|$, where $\|\cdot\|_2$ denotes the spectral norm of a matrix, we have

$$\frac{1}{n}\sum_{j=1}^{n} \|X_j\|\|\Delta_{n,j}\| \leq \|S_n^{-1/2} - I_d\|_2 \frac{1}{n}\sum_{j=1}^{n} \|X_j\|^2 + \|S_n^{-1/2}\|_2 \|\overline{X}_n\| \frac{1}{n}\sum_{j=1}^{n} \|X_j\|.$$

The strong law of large numbers and the continuity of the map $A \mapsto A^{-1/2}$ now yield $\overline{X}_n \xrightarrow{a.s.} 0$, $n^{-1}\sum_{j=1}^n \|X_j\| \xrightarrow{a.s.} \mathbb{E}\|X\|$, $n^{-1}\sum_{j=1}^n \|X_j\|^2 \xrightarrow{a.s.} \mathbb{E}\|X\|^2$ and $S_n^{-1/2} \xrightarrow{a.s.} I_d$. Thus, the right-hand side of (21) converges to 0 almost surely. Likewise, $n^{-1}\sum_{j=1}^n \|\Delta_{n,j}\| \xrightarrow{a.s.} 0$, and the remaining assertion $\|n^{-1/2}(Z_n - Z_n^0)\|_{\mathbb{H}} \xrightarrow{a.s.} 0$ now follows from the triangle inequality. ■

As a corollary, we obtain the following result:

Corollary 11. *The test for multivariate normality based on $T_{n,a}$ is consistent against each alternative distribution satisfying $\mathbb{E}\|X\|^2 < \infty$.*

Proof. Let $\psi_X(t) = \mathbb{E}[\exp(\mathrm{i}t^\top X)]$ be the CF of X. By straightforward calculations, we have

$$\Delta_a = \int \|\nabla\psi_X(t) - \nabla\psi(t)\|_{\mathbb{C}}^2 w_a(t)\,dt,$$

where Δ_a is given in (20). Since $\Delta_a = 0$ if and only if $X \stackrel{\mathcal{D}}{=} N_d(0, I_d)$ (recall the standing assumptions that $\mathbb{E}[X] = 0$ and $\mathbb{E}[XX^\top] = I_d$), the assertion follows. ■

Notice that, for each $a > 0$, Δ_a may be regarded as a measure of deviation from normality. The following result sheds some more light on Δ_a:

Theorem 12. *If $E\|X\|^6 < \infty$, then, under the standing assumptions $\mathbb{E}[X] = 0$ and $\mathbb{E}[XX^\top] = I_d$, we have*

$$\lim_{a\to\infty} 16a^2\left(\frac{a}{\pi}\right)^{\frac{d}{2}}\Delta_a = \mathbb{E}\left[X_1^\top X_2\|X_1\|^2\|X_2\|^2\right] + 2\mathbb{E}\left[\left(X_1^\top X_2\right)^3\right], \tag{22}$$

as well as

$$\lim_{a\to 0}\pi^{-\frac{d}{2}}\Delta_a = \frac{d}{2} - 2^{\frac{d}{2}+1}\mathbb{E}\left[\|X_1\|^2\exp\left(-\frac{\|X_1\|^2}{2}\right)\right].$$

Proof. Straightforward calculations give $\Delta_a = I_{a,1} - I_{a,2} + I_{a,3}$, where

$$I_{a,1} = \int \mathbb{E}[X_1\mathrm{CS}^+(t, X_1)]^\top \mathbb{E}[X_2\mathrm{CS}^+(t, X_2)] w_a(t)\,dt,$$

$$I_{a,2} = 2\int \mathbb{E}[X_1\mathrm{CS}^+(t, X_1)]^\top t\psi(t)w_a(t)\,dt, \quad I_{a,3} = \int t^\top t\psi(t)^2 w_a(t)\,dt.$$

Using addition theorems for the sine and the cosine function as well as (11) and (A1)–(A3), it follows that

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

$$I_{a,2} = 2\left(\frac{2\pi}{2a+1}\right)^{\frac{d}{2}} \mathbb{E}\left[\frac{\|X_1\|^2}{2a+1}\exp\left(-\frac{\|X_1\|^2}{4a+2}\right)\right], \qquad I_{a,3} = \left(\frac{\pi}{a+1}\right)^{\frac{d}{2}} \frac{d}{2a+2}.$$

Taylor expansions (7) and (8), together with $\mathbb{E}[X] = 0$, $\mathbb{E}[XX^\top] = I_d$ and $\mathbb{E}\|X\|^6 < \infty$ then yield

$$\begin{aligned} a^2\left(\frac{a}{\pi}\right)^{\frac{d}{2}} I_{a,1} &= a^2\mathbb{E}\left[X_1^\top X_2\right] - a\mathbb{E}\left[X_1^\top X_2 \frac{\|X_1 - X_2\|^2}{4}\right] + \mathbb{E}\left[X_1^\top X_2 \frac{\|X_1 - X_2\|^4}{32}\right] + O(a^{-1}) \\ &= \frac{ad}{2} + \frac{1}{16}\mathbb{E}\left[X_1^\top X_2\|X_1\|^2\|X_2\|^2\right] + \frac{2}{16}\mathbb{E}\left[\left(X_1^\top X_2\right)^3\right] - \frac{1}{4}\mathbb{E}\|X_1\|^4 + O(a^{-1}), \\ a^2\left(\frac{a}{\pi}\right)^{\frac{d}{2}} I_{a,2} &= a\left(\frac{a}{a+\frac{1}{2}}\right)^{\frac{d}{2}+1} \mathbb{E}\left[\|X_1\|^2 \exp\left(-\frac{\|X_1\|^2}{4a+2}\right)\right] \\ &= \left(ad - \frac{d^2}{4} - \frac{d}{2}\right) - \frac{1}{4}\mathbb{E}\|X_1\|^4 + O(a^{-1}), \\ a^2\left(\frac{a}{\pi}\right)^{\frac{d}{2}} I_{a,3} &= a\left(\frac{a}{a+1}\right)^{\frac{d}{2}+1}\frac{d}{2} = \frac{ad}{2} - \frac{d^2}{4} - \frac{d}{2} + O(a^{-1}). \end{aligned}$$

Upon summarizing, the assertion follows. The second statement is proved following similar arguments. ■

We remark in passing that the first term on the right-hand side of (22) is the population measure of multivariate skewness in the sense of Móri, Rohatgi & Székely (1994), and $\mathbb{E}\left[\left(X_1^\top X_2\right)^3\right]$ is the population skewness in the sense of Mardia (1970). Thus, Theorem 12 can be regarded as the "population counterpart" of Theorems 3 and 4.

Baringhaus, Ebner & Henze (2017) observed that, in the context of goodness-of-fit testing of a general parametric hypothesis $\tilde{H}_0$ (say), weighted L^2-statistics have a normal limit under fixed alternatives to $\tilde{H}_0$. To state such a theorem in our case, we first introduce some notation. Again, we write $\psi_X(t) = \mathbb{E}[\exp(\mathrm{i}t^\top X)]$ for the CF of X and put $\psi_X^{\pm}(t) := \operatorname{Re}\psi_X(t) \pm \operatorname{Im}\psi_X(t)$,

$$\begin{aligned} w(t,X) &= X\mathrm{CS}^+(t,X) - X\psi_X^+(t) - t^\top X\nabla\psi_X^+(t) \\ &\quad + \frac{1}{2}\left(\left(XX^\top + I_d\right)\nabla\psi_X^-(t) - \mathbb{E}[XX^\top\mathrm{CS}^-(t,X)]\left(XX^\top - I_d\right)t\right). \end{aligned} \tag{23}$$

Moreover, let

$$L(s,t) := \mathbb{E}\left[w(s,X)w(t,X)^\top\right], \quad s,t \in \mathbb{R}^d. \tag{24}$$

We then have the following result:

Theorem 13. *If* $\mathbb{E}\|X\|^4 < \infty$, *we have*

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

where

$$\sigma_a^2 := 4 \iint z(s)^\intercal L(s,t) z(t) w_a(s) w_a(t)\, ds\, dt. \tag{25}$$

Here

$$z(t) := \mu(t) - t\psi(t), \tag{26}$$

and $L(s,t)$ is defined in (24).

Proof. The basic observation is that, with Z_n defined in (10) and $z(t) := \mu(t) - t\psi(t)$, we have

$$\begin{aligned}\sqrt{n}\left(\frac{T_{n,a}}{n} - \Delta_a\right) &= \sqrt{n}\big(\|n^{-1/2}Z_n\|_{\mathbb{H}}^2 - \|z\|_{\mathbb{H}}^2\big)\\ &= \sqrt{n}\langle n^{-1/2}Z_n - z, 2z + n^{-1/2}Z_n - z\rangle_{\mathbb{H}} \qquad (27)\\ &= 2\langle Z_n - \sqrt{n}z, z\rangle_{\mathbb{H}} + n^{-1/2}\|Z_n - \sqrt{n}z\|_{\mathbb{H}}^2. \qquad (28)\end{aligned}$$

Letting $V_n(t) := Z_n(t) - \sqrt{n}z(t) = n^{-1/2}\sum_{j=1}^n \big(Y_{n,j}\mathrm{CS}^+(t, Y_{n,j}) - \mu(t)\big)$, the next step is to show that

$$V_n \xrightarrow{\mathcal{D}} V \text{ in } \mathbb{H} \tag{29}$$

for some centred Gaussian random element V of $\mathbb{H}$ having covariance matrix kernel $L(s,t)$ given in (24). The proof of (29) is completely analogous to that of Theorem 5 and is therefore omitted. In view of (29), the second summand in (28) is $o_{\mathbb{P}}(1)$, and the first converges in distribution to $2\langle V, z\rangle_{\mathbb{H}}$ by the continuous mapping theorem. The distribution of $2\langle V, z\rangle_{\mathbb{H}}$ is the normal distribution $N\big(0, \sigma_a^2\big)$. ∎

Using Slutsky's lemma, Theorem 13 yields the following asymptotic confidence interval for Δ_a:

Corollary 14. *For $\alpha \in (0,1)$, let $z_{1-\alpha/2}$ denote the $(1-\alpha/2)$-quantile of the standard normal distribution. If $\widehat{\sigma}_{n,a}^2$ is a consistent sequence of estimators for σ_a^2, and if $\sigma_a^2 > 0$, then*

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

is an asymptotic confidence interval with level $1-\alpha$ for Δ_a.

A necessary and sufficient condition for $\sigma_a^2 > 0$ is that the function $\mathbb{R}^d \ni s \mapsto \int L(s,t)z(t)w_a(t)\, dt$ does not vanish λ^d-almost everywhere, see Remark 1 of Baringhaus, Ebner & Henze (2017).

To construct a consistent sequence of estimators for σ_a^2, we replace $z(s)$, $z(t)$ and $L(s,t)$ figuring in (25) with suitable empirical counterparts. In view of (23) and (24) and the fact that $\nabla\psi_X^+(t) = \mathbb{E}[X\mathrm{CS}^-(t,X)]$, $\nabla\psi_X^-(t) = -\mathbb{E}[X\mathrm{CS}^+(t,X)]$, let

$$L_n(s,t) := \frac{1}{n}\sum_{j=1}^n W_{n,j}(s)W_{n,j}(t)^\intercal, \tag{30}$$

where

$$W_{n,j}(t) := Y_{n,j}\mathrm{CS}^{+}(t, Y_{n,j}) - Y_{n,j}\Psi_{1,n}(t) - t^{\intercal} Y_{n,j}\Psi_{2,n}(t) - \tfrac{1}{2}\left(Y_{n,j}Y_{n,j}^{\intercal} + I_d\right)\Psi_{3,n}(t) - \tfrac{1}{2}\Psi_{4,n}(t)\left(Y_{n,j}Y_{n,j}^{\intercal} - I_d\right), \tag{31}$$

and

$$\Psi_{1,n}(t) := \frac{1}{n}\sum_{j=1}^{n} \mathrm{CS}^{+}(t, Y_{n,j}), \quad \Psi_{2,n}(t) := \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^{-}(t, Y_{n,j}), \tag{32}$$

$$\Psi_{3,n}(t) := \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^{+}(t, Y_{n,j}), \quad \Psi_{4,n}(t) := \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}Y_{n,j}^{\intercal}\mathrm{CS}^{-}(t, Y_{n,j}). \tag{33}$$

Furthermore, let

$$z_n(t) := \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^{+}(t, Y_{n,j}) - t\psi(t). \tag{34}$$

We then have the following result:

Theorem 15. *Let*

$$\widehat{\sigma}_{n,a}^{2} := 4\iint z_n(s)^{\intercal} L_n(s,t) z_n(t) w_a(s) w_a(t)\, ds\, dt,$$

where $L_n(s,t)$ and $z_n(t)$ are as defined in (30) and (34), respectively. If $\mathbb{E}\|X\|^4 < \infty$, then $\left(\widehat{\sigma}_{n,a}^{2}\right)$ is a consistent sequence of estimators for σ_a^2, i.e., we have $\widehat{\sigma}_{n,a}^{2} \xrightarrow{\mathbb{P}} \sigma_a^2$. Moreover

$$\widehat{\sigma}_{n,a}^{2} = \sum_{i,j=1}^{5} \widehat{\sigma}_{n,a}^{i,j}, \tag{35}$$

where $\widehat{\sigma}_{n,a}^{i,j}$ is given in (A12).

Since the proof of Theorem 15 is long and tedious, it is deferred to the Appendix. We stress that the representation (35) does not comprise any integral, which means that $\widehat{\sigma}_{n,a}^{2}$ is a feasible estimator.

We close this section with an example that illustrates the feasibility of the asymptotic confidence interval. To this end, we consider the following standardized symmetric alternatives to normality: Firstly, let $X \stackrel{\mathcal{D}}{=} \mathrm{U}\left(-\sqrt{3}, \sqrt{3}\right)^d$ have the uniform distribution on the cube $\left(-\sqrt{3}, \sqrt{3}\right)^d$. In this case, we have $\varphi_X(t) = \prod_{i=1}^{d} \sin\left(\sqrt{3}t_i\right)/\sqrt{3}t_i$,

$$\nabla\varphi_X(t)^{(j)} = \frac{3\cos\left(\sqrt{3}t_j\right)t_j - \sqrt{3}\sin\left(\sqrt{3}t_j\right)}{3t_j^2}\prod_{i\neq j}^{d}\frac{\sin\left(\sqrt{3}t_i\right)}{\sqrt{3}t_i},$$

where $\nabla\varphi(t)^{(j)}$ is the jth component of $\nabla\varphi(t)$. Secondly, we consider a Laplace distribution with i.i.d. marginals, denoted by Laplace$(0, 1/\sqrt{2})^d$, for which

$$\varphi_X(t) = \prod_{i=1}^{d} \frac{2}{2+t_i^2}, \qquad \nabla\varphi_X(t)^{(j)} = -\frac{4t_j}{\left(2+t_j^2\right)^2} \prod_{i\neq j}^{d} \frac{2}{2+t_i^2}.$$

Finally, let X have a logistic distribution with i.i.d. marginals, denoted by Logistic$(0, 3/\pi)^d$. In this case, we obtain $\varphi_X(t) = \prod_{i=1}^{d} \sqrt{3}t_i / \sinh\left(\sqrt{3}t_i\right)$,

$$\nabla\varphi_X(t)^{(j)} = \frac{\sqrt{3}\sinh\left(\sqrt{3}t_j\right) - 3t_j\cosh\left(\sqrt{3}t_j\right)}{\sinh\left(\sqrt{3}t_j\right)^2} \prod_{i\neq j}^{d} \frac{\sqrt{3}t_i}{\sinh\left(\sqrt{3}t_i\right)}.$$

In each case, Δ_a has been computed by numerical integration. The resulting values are displayed in Table 3.

By means of a Monte Carlo study, we estimated the probability of coverage of the confidence interval $I_{n,a,\alpha}$ figuring in Corollary 14 for $a \in \{0.5, 1, 2, 5\}$, $d \in \{1, 2\}$, and the sample sizes $n \in \{50, 100, 200, 500\}$. The nominal level is 0.95, and the number of replications is 10,000. Simulations have been carried out with the statistical software R, see R Core Team (2019). In particular, we used the package `extraDistr`, see Wolodzko (2019), to generate variates from the Laplace distribution. The results are displayed in Table 4. We also considered a confidence interval $I^*_{n,a,\alpha}$ for Δ_a based on the asymptotic normality of $\sqrt{n}(\log(T_{n,a}/n) - \log(\Delta_a))$ through the delta method, since Δ_a is positive if X is not normally distributed. As one can see, the empirical coverage converges to the nominal level. However, the convergence seems to be slower for higher dimensions. The empirical coverage of the confidence interval $I_{n,a,\alpha}$ seems to converge faster, especially in higher dimensions. For larger values of the tuning parameter a, the confidence interval tends to be too wide, so we conjecture that an improvement of the asymptotic interval might be found.

5. SIMULATIONS

This section presents the results of a Monte Carlo study, with the aim to compare the power of the proposed test with respect to that of prominent competitors against selected alternatives. We used the statistical software R, see R Core Team (2019), and we employed the package `MonteCarlo`,

TABLE 3: Values of Δ_a.

		a			
	d	0.5	1	2	5
$U\left(-\sqrt{3}, \sqrt{3}\right)^d$	1	0.029273	0.011432	0.002911	0.000259
	2	0.090821	0.027841	0.005709	0.000365
Laplace$\left(0, 1/\sqrt{2}\right)^d$	1	0.026076	0.013968	0.005230	0.000778
	2	0.071014	0.032525	0.010141	0.001097
Logistic$\left(0, \sqrt{3}/\pi\right)^d$	1	0.005014	0.002688	0.001005	0.000144
	2	0.013664	0.006226	0.001942	0.000202

TABLE 4: Empirical coverage probability for Δ_a (10,000 replications, nominal level 0.95).

			$I_{n,a,0.95}$				$I^*_{n,a,0.95}$			
			a							
	d	n	0.5	1	2	5	0.5	1	2	5
$U(-\sqrt{3}, \sqrt{3})^d$	1	50	94.86	95.59	98.08	99.23	89.59	90.60	92.89	95.23
		100	94.71	95.67	97.29	98.93	92.13	92.34	93.65	96.99
		200	94.77	95.30	96.72	98.52	93.77	93.63	94.59	97.16
		500	95.08	94.75	95.42	97.91	94.58	94.06	94.37	97.00
	2	50	82.81	87.97	93.61	97.99	55.44	62.70	66.53	68.37
		100	87.62	89.10	92.75	98.24	72.63	75.89	77.53	80.10
		200	90.02	90.62	92.95	97.56	82.49	83.56	84.55	86.58
		500	92.74	92.40	92.69	95.78	89.75	89.52	89.37	90.58
Laplace$(0, 1/\sqrt{2})^d$	1	50	92.98	90.01	87.60	87.25	88.50	90.24	89.16	86.12
		100	93.94	90.54	88.36	88.45	92.36	93.09	92.99	91.84
		200	94.81	93.30	90.46	89.17	94.26	94.61	94.30	93.35
		500	95.02	94.55	92.83	90.59	94.60	95.33	94.67	93.53
	2	50	84.58	94.33	96.07	95.28	39.69	65.00	72.67	67.90
		100	90.92	96.53	97.82	97.44	62.87	80.03	84.94	82.65
		200	93.00	97.06	97.50	97.56	76.35	86.99	90.34	89.69
		500	94.22	96.36	96.79	97.23	86.83	91.59	92.74	94.01
Logistic$(0, \sqrt{3}/\pi)^d$	1	50	99.15	98.24	97.32	96.85	75.99	81.99	81.95	78.82
		100	98.55	96.22	95.07	94.98	84.70	87.83	87.65	85.87
		200	96.38	94.64	93.51	93.77	89.11	90.98	90.90	90.54
		500	95.51	94.13	93.64	93.56	92.53	94.03	94.49	94.37
	2	50	69.24	89.51	94.97	95.71	1.08	15.17	31.92	36.06
		100	79.65	94.61	97.68	98.64	9.69	37.20	53.68	54.49
		200	85.90	96.54	98.78	99.32	29.73	59.16	71.20	71.24
		500	89.05	96.45	98.49	99.32	59.30	77.13	84.43	84.58

see Leschinski (2019), which allows for parallel computing. In addition, we used the package `expm`, see Goulet et al. (2019), for the standardization of the data. Critical values for the test statistic have been estimated by means of extensive simulations (100,000 replications), and they are displayed in Table 5 for the weight parameters $a \in \{0.5, 1, 2, 5, 10, \infty\}$ and the sample sizes $n \in \{20, 50, 100\}$. Throughout, the level of significance is $\alpha = 0.05$. For the sake of comparison, Table 5 displays the approximate critical values of $T_{\infty,a}$ in the special case $d = 1$, which have been obtained in Section 3 by choosing a distribution of the Pearson family by equating the first

TABLE 5: Empirical 0.95-quantiles for $a^{d/2+2}\pi^{-d/2}16T_{n,a}$ under H_0 (100,000 replications).

		a					
d	n	0.5	1	2	5	10	∞
1	20	2.57	7.12	15.90	30.72	39.98	53.38
	50	2.64	7.42	16.82	34.00	45.48	62.93
	100	2.65	7.46	17.08	34.88	47.28	65.19
	∞	2.67	7.52	17.28	35.56	46.23	—
2	20	5.77	15.94	35.47	70.27	93.10	125.90
	50	5.83	16.27	37.16	76.41	102.65	145.38
	100	5.87	16.19	37.35	77.40	106.51	151.15
3	20	9.43	27.03	61.74	125.52	167.47	230.75
	50	9.57	27.37	64.02	135.16	186.80	267.89
	100	9.58	27.47	64.38	137.79	190.30	276.76
5	20	17.89	55.55	137.20	296.36	407.65	581.08
	50	18.03	56.21	141.10	319.59	452.61	681.00
	100	18.05	56.32	141.21	323.19	462.59	704.12

four moments. As already mentioned in Section 2, the test statistic $T_{n,\infty}$ is a linear combination of skewness in the sense of Mardia (1970) and skewness in the sense of Móri, Rohatgi & Székely (1994), and it equals the statistic HV_∞ of Henze–Visagie, see Henze & Visagie (2020).

5.1. Univariate Normal Distribution

In the univariate case $d = 1$, we compared the power of our novel test statistics with several competitors, which are

- the Cramér–von Mises test (CvM),
- the Anderson–Darling test (AD),
- the Shapiro–Wilk test (SW),
- the Baringhaus–Henze–Epps–Pulley test (BHEP), and
- the Henze–Visagie test (HV).

The first three of these tests are well known. The CvM test and the AD test have been implemented with the R-package `nortest`, see Gross & Ligges (2015), which contains the functions `cvm.test` and `ad.test`. For the SW test, we used the function `shapiro.test` of the `stats`-package. The test statistics BHEP and HV will be explained in (36) and (37), respectively.

For the BHEP test and the HV test, critical values have been simulated with 100,000 replications. These values and those of Table 5 for the novel test statistics have been employed to assess the power of the various tests against several alternatives. Table 6 gives the percentages of rejection based on 100,000 replications. An asterisk denotes power of 100%, and the best performing test for each alternative is marked in boldface. The choice of alternatives orients itself towards those used in Henze & Visagie (2020). The acronym NMix1 denotes a mixture

TABLE 6: Empirical power ($d = 1$, $\alpha = 0.05$, 100,000 replications).

	n	CvM	AD	SW	BHEP_1	HV_5	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$N(0, 1)$	20	5	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5	5
NMix1	20	20	23	25	26	25	27	**28**	**28**	**28**	27	27
	50	45	50	56	55	52	58	60	**61**	**61**	60	59
	100	75	81	85	84	82	87	88	**89**	**89**	88	88
$t_3(0, 1)$	20	30	33	34	33	**36**	**36**	**36**	35	35	34	35
	50	57	61	64	61	63	**66**	65	63	59	56	52
	100	83	85	**88**	86	84	**88**	**88**	86	80	76	64
$t_5(0, 1)$	20	15	17	19	18	**22**	20	20	20	20	20	20
	50	27	30	35	31	**39**	36	36	35	34	33	32
	100	43	48	**57**	50	56	55	56	53	49	45	40
$t_{10}(0, 1)$	20	8	9	10	9	**12**	11	11	11	11	11	11
	50	11	12	15	13	**19**	15	16	16	16	16	16
	100	14	16	23	17	**27**	21	22	22	21	20	20
$\chi^2(5)$	20	34	38	**44**	42	35	42	43	43	42	41	40
	50	73	80	**89**	83	74	86	86	87	86	85	83
	100	97	99	*	99	97	99	99	*	99	99	99
$\chi^2(15)$	20	14	15	17	17	16	18	**19**	**19**	**19**	**19**	18
	50	30	33	42	39	37	40	43	**45**	**45**	**45**	44
	100	54	61	75	68	65	71	74	76	**77**	**77**	76
Logistic(0, 1)	20	10	11	11	11	**14**	13	13	13	13	13	13
	50	14	16	20	17	**23**	20	20	20	19	19	19
	100	21	24	31	25	**32**	30	30	28	26	24	23
$U(-\sqrt{3}, \sqrt{3})$	20	14	17	**20**	12	0	10	4	2	1	1	1
	50	44	58	**75**	55	0	55	33	5	1	0	0
	100	84	95	*	94	0	96	90	48	2	1	0
$\text{P}_{VII}(5)$	20	15	17	19	18	**22**	20	20	20	20	20	21
	50	27	30	35	31	**39**	36	36	35	34	33	32
	100	43	48	**57**	50	56	55	56	53	49	45	41
$\text{P}_{VII}(10)$	20	8	9	10	9	**12**	11	11	11	11	11	11
	50	11	12	16	12	**19**	15	16	16	16	16	16
	100	14	16	23	17	**27**	21	22	22	20	20	20

of the normal distributions $N(0, 1)$ and $N(3, 1)$ with weights 0.9 and 0.1, respectively. We write P_{VII} for the Pearson-type *VII* distribution, see Becker & Klößner (2017).

The novel tests outperform the selected competitors for the t_3-distribution, the $\chi^2(15)$-distribution and the distribution NMix1, and they keep up with the other procedures against the remaining alternatives. For most of the alternatives, power does not change much with varying the weight parameter a. A notable exception is the uniform distribution $U(-\sqrt{3}, \sqrt{3})$, against which power breaks down for larger tuning parameters, a feature shared by the HV test.

5.2. Multivariate Normal Distribution

For the dimensions $d = 2$, $d = 3$ and $d = 5$, we compared the novel test statistic with the following procedures:

- the test of Baringhaus–Henze–Epps–Pulley (BHEP),
- the test of Henze–Zirkler (HZ),
- the test of Henze–Visagie (HV), and
- the energy test (EN).

A recent synopsis of tests for multivariate normality is given in Ebner & Henze (2020). Just as the novel procedure, the BHEP test (see Henze & Wagner, 1997) is based on the empirical characteristic function. More precisely, it employs the test statistic

$$\mathrm{BHEP}_a = \int |\psi_n(t) - \psi(t)|^2 \varphi_a(t)\, dt, \tag{36}$$

where $\varphi_a(t) = (2\pi a^2)^{-d/2} \exp(-\|t\|^2/(2a^2))$, and $\psi_n(t)$ and $\psi(t)$ are given in (3) and (2), respectively. An alternative representation for BHEP_a is

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

In our study, we used the special value $a = 1$.

The test HZ of Henze–Zirkler (cf. Henze & Zirkler, 1990) originates if we choose $a = 1/\sqrt{2}\,((2d+1)n/4)^{\frac{1}{d+4}}$ in the BHEP test. The R-package HZ, see Korkmaz, Goksuluk & Zararsiz (2014), contains the function mvn, which calculates the statistic of the HZ test.

The recent test of Henze–Visagie, see Henze & Visagie (2020), is the "moment-generating function analogue" of our novel test statistic. It employs the test statistic

$$\mathrm{HV}_a = n \int \|\nabla M_n(t) - t M_n(t)\|^2 w_a(t)\, dt,$$

where $M_n(t) = n^{-1} \sum_{j=1}^{n} \exp(t^\top Y_{n,j})$ is the empirical moment-generating function of the scaled residuals. An alternative representation of HV_a is

$$\mathrm{HV}_a = \frac{1}{n}\left(\frac{\pi}{a}\right)^{\frac{d}{2}} \sum_{i,j=1}^{n} \exp\left(\frac{\|Y_{n,i} + Y_{n,j}\|^2}{4a}\right)\left(Y_{n,i}^\top Y_{n,j} + \|Y_{n,i} + Y_{n,j}\|^2\left(\frac{1}{4a^2} - \frac{1}{2a}\right) + \frac{d}{2a}\right). \tag{37}$$

TABLE 7: Empirical power ($d = 2$, $\alpha = 0.05$, 100,000 replications).

	n	BHEP$_1$	HZ	HV$_5$	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$N_2(0, I_2)$	20	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5
NMix1	20	39	34	32	37	38	**41**	**41**	40	39	38
	50	83	74	68	82	85	88	**89**	88	88	86
	100	99	96	97	99	99	99	*	*	*	*
NMix2	20	20	17	**27**	20	23	24	25	25	25	25
	50	38	30	**53**	39	45	48	49	48	47	44
	100	60	47	**77**	61	68	72	72	70	66	55
$t_3(0, I_2)$	20	47	45	54	49	49	51	**53**	**53**	**53**	52
	50	83	80	**85**	84	82	84	83	83	81	78
	100	**98**	97	97	97	97	**98**	**98**	97	95	90
$t_5(0, I_2)$	20	25	22	**32**	26	27	29	30	31	31	31
	50	49	42	**59**	50	49	53	55	54	54	52
	100	75	67	**81**	76	71	76	77	75	72	66
$t_{10}(0, I_2)$	20	11	10	**16**	12	12	14	14	15	15	**16**
	50	17	14	**29**	18	19	22	24	25	25	25
	100	27	20	**43**	28	26	31	33	34	33	33
$(\chi^2(5))^2$	20	48	44	38	46	46	48	**50**	48	47	46
	50	93	87	80	92	93	94	**95**	**95**	94	93
	100	*	*	99	*	*	*	*	*	*	*
$(\chi^2(15))^2$	20	18	16	17	17	17	19	**20**	**20**	**20**	19
	50	45	35	39	42	43	49	53	**55**	54	52
	100	78	62	71	77	78	84	88	**89**	88	88
$(\chi^2(20))^2$	20	15	13	14	14	14	15	**16**	**16**	**16**	**16**
	50	34	27	31	33	33	38	41	**43**	**43**	42
	100	64	47	58	63	64	71	76	**78**	77	77
$\Gamma(5, 1)^2$	20	26	23	23	24	24	27	**28**	27	27	26
	50	64	53	53	61	62	68	71	**72**	71	69
	100	93	84	87	93	94	96	97	**98**	97	97
$\Gamma(4, 2)^2$	20	32	28	27	30	30	33	**34**	33	33	32
	50	75	64	61	73	73	79	**81**	**81**	80	79
	100	98	92	93	97	98	**99**	**99**	**99**	**99**	**99**

TABLE 7: Continued

	n	BHEP_1	HZ	HV_5	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$\text{Logistic}(0,1)^2$	20	11	10	**16**	12	13	14	15	**16**	15	**16**
	50	18	15	**29**	20	20	23	24	25	25	25
	100	29	23	**42**	31	29	34	35	34	33	31
$U(-\sqrt{3},\sqrt{3})^2$	20	12	**18**	0	11	6	3	1	1	0	0
	50	60	**67**	0	52	32	13	3	0	0	0
	100	**98**	**98**	0	96	92	80	24	1	0	0
$\text{P}_{VII}(5)^2$	20	20	18	**28**	21	22	24	26	26	26	27
	50	39	32	**51**	40	41	45	47	46	46	45
	100	63	53	**73**	64	62	67	68	66	62	58
$\text{P}_{VII}(10)^2$	20	10	8	**13**	10	11	11	12	**13**	**13**	**13**
	50	13	11	**23**	14	15	18	19	20	20	20
	100	19	14	**35**	21	20	24	26	27	27	26
$\text{P}_{VII}(20)^2$	20	7	6	**8**	7	7	7	7	**8**	**8**	**8**
	50	7	7	**12**	8	8	9	10	11	11	11
	100	8	7	**17**	9	9	11	11	13	12	13

In our comparative study, we put $a = 5$, as recommended in Henze & Visagie (2020).

The rationale of the energy test of Székely & Rizzo (2005) is based on the fact that, if X and Y are independent, integrable d-dimensional random vectors and X' and Y' denote independent copies of X and Y, respectively, then

$$2\mathbb{E}\|X-Y\| - \mathbb{E}\|X-X'\| - \mathbb{E}\|Y-Y'\| \geq 0.$$

Here, the equality holds if and only if $X \stackrel{\mathcal{D}}{=} Y$. The statistic of the energy test for multivariate normality is

$$\text{EN} = n\left(\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}\left[\|\tilde{Y}_{n,j} - Z_1\| \,|X_1,\dots,X_n\right] - \mathbb{E}\|Z_1 - Z_2\| - \frac{1}{n^2}\sum_{i,j=1}^{n}\|\tilde{Y}_{n,i} - \tilde{Y}_{n,j}\|\right).$$

Here, $\tilde{Y}_{n,j} = \sqrt{n/(n-1)}Y_{n,j}$, and Z_1, Z_2 are i.i.d. with the normal distribution $N_d(0, I_d)$, which are also independent of $Y_{n,1},\dots,Y_{n,n}$. To calculate EN, notice that $\mathbb{E}\|Z_1 - Z_2\| = 2\Gamma\left(\frac{d+1}{2}\right)/\Gamma\left(\frac{d}{2}\right)$ and

$$\mathbb{E}\|a - Z\| = \sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} + \sqrt{\frac{2}{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!2^k}\frac{\|a\|^{2k+2}}{(2k+1)(2k+2)}\frac{2\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(k+\frac{3}{2}\right)}{\Gamma\left(k+\frac{d}{2}+1\right)}.$$

TABLE 8: Empirical power ($d = 3$, $\alpha = 0.05$, 100,000 replications).

	n	BHEP_1	HZ	HV_5	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$N_3(0, I_3)$	20	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5
NMix1	20	39	35	33	41	40	43	**44**	43	41	40
	50	89	81	66	91	91	94	**95**	**95**	93	92
	100	*	98	95	*	*	*	*	*	*	*
NMix2	20	28	24	**43**	33	34	38	40	41	41	41
	50	59	49	**80**	66	65	72	75	75	75	73
	100	85	74	**96**	88	87	92	93	94	92	87
$t_3(0, I_3)$	20	56	53	65	62	58	63	65	**66**	65	65
	50	93	90	**94**	**94**	89	93	93	93	92	91
	100	*	*	*	98	99	*	*	99	99	98
$t_5(0, I_3)$	20	29	26	**41**	35	32	37	39	**41**	40	**41**
	50	62	54	**73**	67	57	67	70	70	70	69
	100	90	83	**92**	91	80	88	90	89	88	84
$t_{10}(0, I_3)$	20	12	11	**20**	15	14	17	18	19	19	**20**
	50	22	17	**38**	26	22	28	32	34	35	35
	100	37	28	**57**	42	30	40	46	48	48	47
$(\chi^2(5))^3$	20	48	43	38	49	46	50	**51**	50	49	48
	50	95	89	82	96	94	**97**	**97**	**97**	**97**	96
	100	*	*	99	*	*	*	*	*	*	*
$(\chi^2(15))^3$	20	17	15	17	18	16	18	**19**	**19**	**19**	**19**
	50	45	34	38	48	44	51	56	**58**	57	56
	100	82	64	69	84	81	88	92	**93**	**93**	92
$(\chi^2(20))^3$	20	13	12	14	14	13	14	**16**	15	15	15
	50	34	25	30	36	31	39	43	**45**	44	44
	100	67	48	56	70	65	75	81	**83**	**83**	82
$\Gamma(5, 1)^3$	20	25	22	23	25	23	26	**28**	27	27	26
	50	65	53	53	68	64	71	**76**	**76**	75	74
	100	96	86	87	97	96	98	**99**	**99**	**99**	**99**
$\Gamma(4, 2)^3$	20	30	27	27	32	29	32	**34**	33	33	32
	50	77	65	62	79	76	82	85	**86**	85	83
	100	99	94	93	99	99	*	*	*	*	*

TABLE 8: Continued

	n	BHEP_1	HZ	HV_5	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$\text{Logistic}(0,1)^3$	20	11	10	**17**	13	13	15	16	**17**	**17**	**17**
	50	18	14	**32**	22	19	24	27	29	29	29
	100	31	23	**48**	36	27	35	39	39	39	38
$U(-\sqrt{3},\sqrt{3})^3$	20	11	**15**	0	6	5	2	1	0	0	0
	50	58	**65**	0	39	20	8	2	0	0	0
	100	**98**	**98**	0	94	79	51	12	1	0	0
$\text{P}_{VII}(5)^3$	20	20	17	**30**	24	23	27	29	**30**	**30**	**30**
	50	41	34	**58**	47	42	50	54	55	54	53
	100	69	57	**81**	73	63	72	76	75	73	69
$\text{P}_{VII}(10)^3$	20	9	8	**14**	11	11	12	13	**14**	**14**	**14**
	50	13	10	**26**	16	14	18	21	23	23	23
	100	20	14	**39**	24	18	24	29	31	31	31
$\text{P}_{VII}(20)^3$	20	6	6	**9**	7	7	7	7	8	8	8
	50	7	6	**13**	8	8	9	10	11	12	12
	100	8	7	**17**	10	8	10	12	13	14	14

The R-package `energy` Rizzo & Székely (2019) contains the function `mvnorm.etest` to calculate EN. Note that all of the mentioned procedures are also implemented in the R-package `mnt`, see Butsch & Ebner (2020).

Just as was done in the case $d = 1$, we first simulated critical values with 100,000 replications. With the same number of replications, we then simulated the power of the tests under discussion against selected alternatives. Again, the choice of alternatives orients itself towards those used in Henze & Visagie (2020). Tables 7–9 display the percentages of rejection of H_0 for dimensions $d = 2$, $d = 3$ and $d = 5$, respectively, and an asterisk again denotes power 100%. To generate pseudo-random numbers, we used the R-packages `mvtnorm`, see Genz et al. (2019), and `PearsonDS`, see Becker & Klößner (2017). Suppressing the dimension d, the distribution NMix1 is a mixture of the normal distributions $N_d(0, I_d)$ and $N_d(3, I_d)$ with mixing proportions 0.9 and 0.1, respectively. Here, 3 stands for the d-dimensional vector that contains 3 in each component. Likewise, NMix2 denotes a mixture of the normal distributions $N_d(0, I_d)$ and $N_d(0, B_d)$ with mixing proportions 0.1 and 0.9, respectively. Here, B_d is a $d \times d$-matrix with 1 for each diagonal entry and 0.9 for each off-diagonal entry.

The novel tests outperform the competitors for some alternatives, notably for the χ^2-, the Γ- and the NMix-distribution, but they can also keep up for the other alternatives. However, just as in the univariate case, power is extremely low against the uniform distribution $U(-\sqrt{3}, \sqrt{3})$, a feature shared by the HV test. Based on the results of this simulation study, we recommend as an omnibus choice $a = 1$ for the tuning parameter, since it leads to competitive power against nearly all of the alternatives considered. In particular, it also has power against alternatives like the uniform distribution.

TABLE 9: Empirical power ($d = 5$, $\alpha = 0.05$, 100,000 replications).

	n	BHEP_1	HZ	HV_5	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
$N_5(0, I_5)$	20	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5
NMix1	20	25	22	31	32	27	33	**36**	34	34	33
	50	85	74	50	94	87	94	**95**	92	90	86
	100	*	98	77	*	*	*	*	*	*	*
NMix2	20	32	27	**62**	48	40	51	56	58	59	59
	50	76	67	**96**	89	79	89	93	94	94	94
	100	96	92	*	99	96	99	99	*	*	*
$t_3(0, I_5)$	20	62	59	79	76	67	76	79	**81**	**81**	80
	50	98	97	99	99	99	*	*	99	99	99
	100	*	*	*	*	*	*	*	*	*	*
$t_5(0, I_5)$	20	31	28	54	47	37	48	52	54	54	**55**
	50	77	71	**89**	88	68	82	88	**89**	**89**	**89**
	100	98	96	**99**	**99**	88	96	98	**99**	98	98
$t_{10}(0, I_5)$	20	12	11	**26**	20	15	21	24	25	**26**	**26**
	50	28	23	**55**	44	26	39	48	52	54	53
	100	54	44	**78**	69	36	54	67	72	73	72
$(\chi^2(5))^5$	20	39	35	36	**48**	39	46	**48**	**48**	47	45
	50	94	87	80	**98**	94	97	**98**	**98**	**98**	97
	100	*	*	99	*	*	*	*	*	*	*
$(\chi^2(15))^5$	20	13	12	15	16	13	15	**17**	**17**	**17**	**17**
	50	38	29	35	52	37	49	56	**58**	**58**	56
	100	78	60	64	90	77	89	94	**95**	**95**	94
$(\chi^2(20))^5$	20	11	9	12	13	11	12	13	**14**	13	13
	50	28	22	28	39	27	36	42	**45**	44	43
	100	61	43	51	77	60	74	83	**86**	**86**	85
$\Gamma(5, 1)^5$	20	18	16	21	24	18	22	24	**25**	24	24
	50	59	47	20	74	58	71	78	**79**	78	76
	100	95	85	83	99	95	99	99	*	*	99
$\Gamma(4, 2)^5$	20	23	20	25	29	23	28	**30**	**30**	**30**	29
	50	72	60	59	84	71	83	87	**88**	87	85
	100	99	94	91	*	99	*	*	*	*	*

TABLE 9: Continued

	n	BHEP$_1$	HZ	HV$_5$	EN	$T_{0.5}$	T_1	T_2	T_5	T_{10}	T_∞
Logistic(0, 1)^5	20	9	8	**17**	13	11	14	16	**17**	**17**	**17**
	50	15	13	**34**	26	17	24	30	33	**34**	**34**
	100	29	22	**53**	42	23	34	43	47	47	47
$U(-\sqrt{3}, \sqrt{3})^5$	20	9	**11**	0	2	4	2	1	0	0	0
	50	50	**51**	0	12	12	4	1	0	0	0
	100	**96**	95	0	75	49	20	5	0	0	0
$P_{VII}(5)^5$	20	16	14	**33**	25	20	27	30	**33**	32	32
	50	39	32	**67**	56	39	54	62	65	65	65
	100	71	60	**89**	83	59	77	84	86	85	83
$P_{VII}(10)^5$	20	8	7	**14**	11	9	11	13	**14**	**14**	**14**
	50	11	9	**28**	19	12	18	23	26	27	27
	100	18	13	**44**	28	16	24	32	37	38	38
$P_{VII}(20)^5$	20	6	5	**8**	7	7	7	**8**	**8**	**8**	**8**
	50	7	6	**13**	9	8	9	11	12	12	12
	100	7	7	**19**	11	8	10	13	15	16	16

5.3. High Dimensions

To assess the power of the proposed test for higher dimensions, we performed a Monte Carlo study. We first generated critical values with 10,000 replications and then simulated the power of the test with the same number of replications against the selected alternatives. Table 10 displays percentages of rejection of H_0 for the dimensions $d \in \{50, 100, 200, 500\}$ and the sample sizes $n \in \{500, 700, 1000, 2000\}$. Again, an asterisk $*$ denotes power of 100%. The proposed test is applicable in high-dimensional settings given there is a reasonably large amount of data available. The test performs well even in high dimensions, and especially so for the t-distributions. The choice of a larger weight parameter a seems to be beneficial for higher dimensional cases. For the uniform distribution $U\left(-\sqrt{3}, \sqrt{3}\right)$, the proposed test performs notably better than in the low-dimensional cases.

6. A REAL DATA EXAMPLE

The Black–Scholes–Merton model is a stochastic model for the dynamics of a financial market that contains derivative investment instruments. One of the basic assumptions of this model is the normality of the log returns of stocks and indexes. To test the hypothesis of joint normality of log returns of several indexes, we consider the following five stock indexes: Standard & Poor 500 (^GSPC), Dow Jones Industrial Average (^DJI), NASDAQ Composite (^IXIC), DAX Perfomance Index (^GDAXI) and EURO STOXX 50 (^STOXX50E), over a period of 50 trading days, starting 1 July 2017. The data (daily closing prices of the stocks) were obtained by means of the R-package `quantmod`, see Ryan & Ulrich (2019). To model the independence assumption between the realizations, we ignored a time span of 10 trading days between each of the five-dimensional observations. Figure 1 shows a plot of the two-dimensional projections

TABLE 10: Empirical power for high dimensions ($\alpha = 0.05$, 10,000 replications).

	d	50			100			200			300		
		a											
	n	2	5	10	2	5	10	2	5	10	2	5	10
$N(0, I_d)$	500	5	5	5	6	5	5	5	6	5	5	4	5
	700	5	5	5	5	5	5	5	5	5	5	6	5
	1000	5	5	5	5	5	5	4	5	5	7	5	5
	2000	5	5	5	5	5	5	5	6	5	5	5	5
NMix1	500	52	89	89	6	17	26	0	5	7	0	4	5
	700	78	99	99	7	31	45	0	4	9	0	0	4
	1000	98	*	*	10	52	71	0	5	13	2	4	5
	2000	*	*	*	25	98	*	0	14	37	0	0	10
NMix2	500	*	*	*	*	*	*	0	79	91	0	0	0
	700	*	*	*	*	*	*	*	*	*	0	79	51
	1000	*	*	*	*	*	*	0	*	*	0	*	*
	2000	*	*	*	*	*	*	*	*	*	0	*	*
$t_5(0, I_d)$	500	*	*	*	*	*	*	*	*	*	22	*	*
	700	*	*	*	*	*	*	*	*	*	14	*	*
	1000	*	*	*	*	*	*	*	*	*	*	*	*
	2000	*	*	*	*	*	*	*	*	*	*	*	*
$t_{10}(0, I_d)$	500	*	*	*	*	*	*	*	*	*	1	*	8
	700	*	*	*	*	*	*	*	*	*	0	*	*
	1000	*	*	*	*	*	*	99	*	*	91	*	*
	2000	*	*	*	*	*	*	*	*	*	74	*	*
$(\chi^2(15))^d$	500	*	*	*	8	97	*	0	0	14	*	0	0
	700	*	*	*	26	*	*	*	0	78	0	0	0
	1000	*	*	*	71	*	*	0	1	*	*	0	0
	2000	*	*	*	*	*	*	0	*	*	0	0	*
$\Gamma(5, 1)^d$	500	*	*	*	16	*	*	0	0	22	0	0	0
	700	*	*	*	62	*	*	0	0	98	0	0	0
	1000	*	*	*	99	*	*	0	4	*	0	0	0
	2000	*	*	*	*	*	*	*	*	*	0	0	*

TABLE 10: Continued

	d	50			100			200			300		
		a											
	n	2	5	10	2	5	10	2	5	10	2	5	10
Logistic(0, 1)^d	500	13	74	98	2	1	83	*	0	0	0	0	0
	700	26	80	99	3	2	89	*	0	0	95	95	0
	1000	51	87	*	8	6	92	0	0	0	*	0	0
	2000	99	98	*	69	44	97	0	0	0	0	0	0
$U(-\sqrt{3}, \sqrt{3})^d$	500	97	0	0	*	88	0	4	*	98	4	99	99
	700	*	0	0	*	96	0	3	*	*	4	*	*
	1000	*	0	0	*	99	0	3	*	*	6	*	*
	2000	*	14	0	*	*	0	3	*	*	3	*	*
$P_{VII}(5)^d$	500	93	*	*	14	8	*	0	0	0	*	*	0
	700	*	*	*	70	52	*	4	0	0	0	0	0
	1000	*	*	*	*	98	*	0	0	0	1	0	0
	2000	*	*	*	*	*	*	95	*	70	0	52	0

of the log returns. For each value $a \in \{0.5, 1, 2, 5, 10\}$ of the weight parameter a, we performed a Monte Carlo simulation based on 100,000 replications in order to estimate the P-value of the observations. The empirical P-values are displayed in Table 11. As can be seen, the hypothesis of multivariate normality of the log returns of the selected stock prices is rejected at the 1% level, for each choice of the weight parameter a. The hypothesis of univariate normality of the marginal distributions of the data, however, is not rejected at the 5% level for most of the choices of the weight parameter a.

7. SUMMARY AND OUTLOOK

We proposed a novel class of tests of normality based on an initial value problem connected to a multivariate Stein equation, which characterizes the multivariate standard normal law. We derived asymptotic theory under the null hypothesis, as well as under contiguous and fixed alternatives. Moreover, we proved consistency against each alternative distribution that satisfies a weak moment condition, and we provided insights into the structure of the behaviour of the test statistic under fixed alternatives by calculating asymptotic confidence intervals for Δ_a and by providing a consistent estimator for the limiting variance σ_a^2. Monte Carlo simulations showed that the methods operated as expected and that the new family of tests is a strong class of competitors to established procedures.

A first open question for further research is to find explicit formulae or numerically stable approximations for the eigenvalues $\lambda_j(a)$, $j = 1, 2, \ldots$ connected to the integral operator $\mathbb{K}$ in (14). We also leave as an open problem the calculation of higher cumulants of $T_{\infty,a}$ for dimensions $d > 1$. Results of this kind would open ground to efficient approximation methods for the computation of critical values that avoid Monte Carlo simulations and efficiency statements, since the largest eigenvalue has a crucial influence on the approximate Bahadur efficiency, see

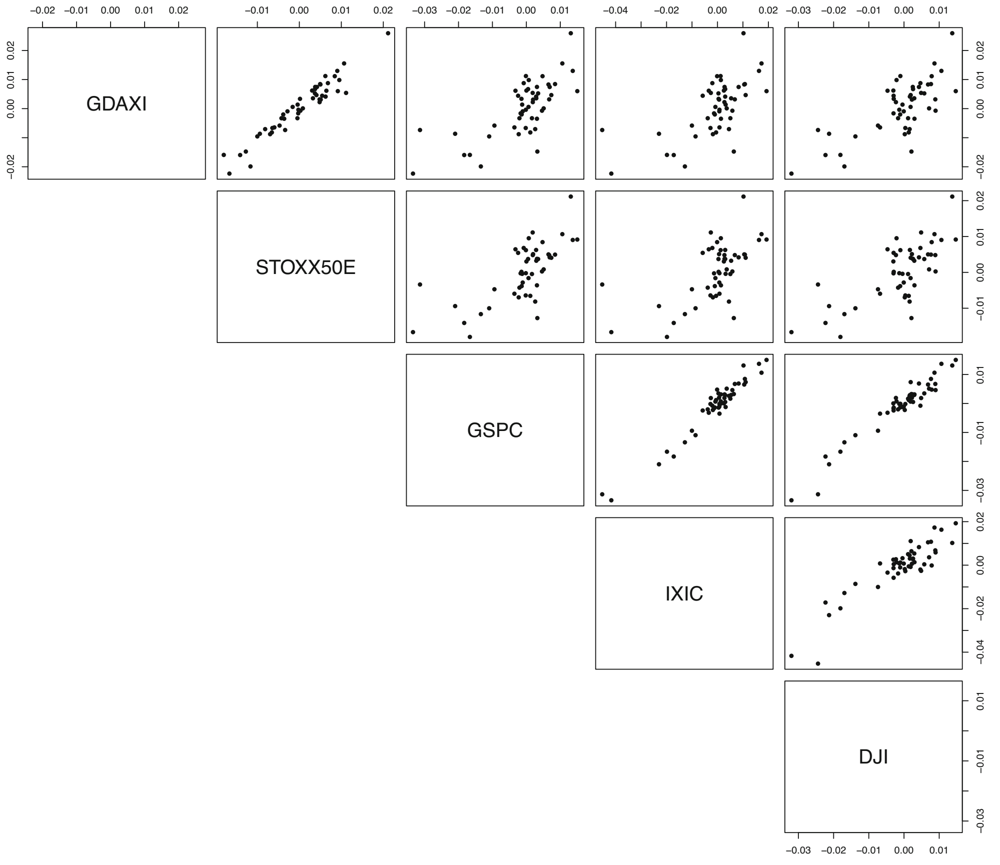

FIGURE 1: 2D projections of the log returns of the indexes.

TABLE 11: Empirical P-value (100,000 replications).

	a	0.5	1	2	5	10
Univariate	^GSPC	0.1025	0.0721	0.0535	0.0457	0.0436
	^DJI	0.1750	0.1324	0.1212	0.1225	0.1231
	^IXIC	0.2226	0.2100	0.2093	0.2236	0.2297
	^GDAXI	0.1391	0.1062	0.0690	0.0491	0.0434
	^STOXX50E	0.0991	0.0930	0.0677	0.0488	0.0424
Multivariate		0.0002	0.0001	0.0002	0.0003	0.0003

Bahadur (1960) and Nikitin (1995). A promising new field of interest in connection with tests of multivariate normality is to consider their behaviour in high-dimensional settings, that is, to find a suitable rescaling and shifting of the test statistic to obtain a non-trivial limit distribution under a suitable limiting regime, under which, for example, $n, d \to \infty$ such that $d/n \to \tau \in [0, \infty]$. For initial results, see Chen & Xia (2019). As a starting point, we conjecture that for a sequence

$(n_d)_{d\in\mathbb{N}}$, where $n_d \geq d+1$ and $n_d = o\left(\left(\frac{2a}{2a+1}\right)^{-\frac{d}{2}}\right)$, we have under H_0 as $d \to \infty$

$$\left(\frac{a}{\pi}\right)^{\frac{d}{2}} \frac{T_{n_d,a}}{d} \xrightarrow{a.s.} 1.$$

Finally, it would be of interest to consider a related family of test statistics, which is given by

$$S_{n,a} = n \int_{\mathbb{R}^d} \|\nabla\psi_n(t) + t\psi_n(t)\|_{\mathbb{C}}^2 \, w_a(t)\, dt.$$

Thus, the theoretical CF in $T_{n,a}$ has been replaced by the empirical counterpart. Note that in the univariate case, this family is extensively studied in Ebner (2021), but the generalization to higher dimensions is still open. We conjecture that similar results as derived in Sections 2–4 hold for $S_{n,a}$.

ACKNOWLEDGEMENTS

The authors would like to thank two anonymous referees for many valuable remarks.

REFERENCES

Arras, B., Mijoule, G., Poly, G., & Swan, Y. (2016). A new approach to the Stein–Tikhomirov method: With applications to the second Wiener chaos and Dickman convergence. arXiv preprint, arXiv:160506819, pp. 1–45.

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. *Annals of the Institute of Statistical Mathematics*, 69, 969–995.

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 behaviour of extreme order statistics. *Annals of Mathematical Statistics*, 34, 992–1002.

Becker, M. & Klößner, S. (2017). PearsonDS: Pearson distribution system. https://CRAN.R-project.org/package=PearsonDS, R package version 1.1.

Bontemps, C. & Meddahi, N. (2005). Testing normality: A GMM approach. *Journal of Econometrics*, 124, 149–186.

Bosq, D. (2000). Linear processes in function spaces: Theory and applications. *Lecture Notes in Statistics*, Vol. 149, Springer, New York.

Butsch, L. & Ebner, B. (2020). mnt: Affine invariant tests of multivariate normality. https://CRAN.R-project.org/package=mnt, R package version 1.3.

Chen, H. & Xia, Y. (2019). A nonparametric normality test for high-dimensional data. arXiv preprint, arXiv:190405289, pp. 1–23.

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

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

Dörr, P., Ebner, B., & Henze, N. (2021a). A new test of multivariate normality by a double estimation in a characterizing PDE. *Metrika*, 84, 401–427.

Dörr, P., Ebner, B., & Henze, N. (2021b). Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. *Scandinavian Journal of Statistics*, 48, 456–501.

Eaton, M. L. & Perlman, M. D. (1973). The non-singularity of generalized sample covariance matrices. *Annals of Statistics*, 1, 710–717, 10.1214/aos/1176342465.

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

Ebner, B. (2021). On combining the zero bias transform and the empirical characteristic function to test normality. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, 18, 1029–1045.

Ebner, B. & Henze, N. (2020). Tests for multivariate normality—A critical review with emphasis on weighted L^2-statistics. *TEST*, 29, 845–892.

Formanov, S. K. & Formanova, T. A. (2013). The Stein–Tikhomirov method and Berry–Esseen inequality for sampling sums from a finite population of independent random variables. In Shiryaev, A. N., Varadhan, S. R. S., & Presman, E. L. (Eds.) *Prokhorov and Contemporary Probability Theory*, Springer Berlin Heidelberg, Berlin, Heidelberg, 261–273.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2019). mvtnorm: Multivariate normal and *t* distributions. https://CRAN.R-project.org/package=mvtnorm, R package version 1.0-11.

Goldstein, L. & Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *The Annals of Applied Probability*, 7, 935–952.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., & Stadelmann, M. (2019). expm: Matrix exponential, log, 'etc'. https://CRAN.R-project.org/package=expm, R package version 0.999-4.

Gross, J. & Ligges, U. (2015). nortest: Tests for normality. https://CRAN.R-project.org/package=nortest, R package version 1.0-4.

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

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

Henze, N. & 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. & Visagie, J. (2020). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *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*, 62, 1–23.

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

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.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). Mvn: An R package for assessing multivariate normality. *The R Journal*, 6, 151–162.

Landsman, Z., Vanduffel, S., & Yao, J. (2013). A note on Stein's lemma for multivariate elliptical distributions. *Journal of Statistical Planning and Inference*, 143, 2016–2022.

Leschinski, C. H. (2019). MonteCarlo: Automatic parallelized Monte Carlo simulations. https://CRAN.R-project.org/package=MonteCarlo, R package version 1.0.6.

Li, B. & Babu, G. J. (2019). A graduate course on statistical inference. *Springer Texts in Statistics. Springer eBooks*, Springer, New York, NY.

Liu, J. S. (1994). Siegel's formula via Stein's identities. *Statistics & Probability Letters*, 21, 247–251.

Maplesoft. (2019). *Maple 2019*. Maplesoft, Waterloo, ON.

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

Mijoule, G., Reinert, G., & Swan, Y. (2018). Stein operators, kernels and discrepancies for multivariate continuous distributions. arXiv preprint, arXiv:180603478, pp. 1–43.
Móri, T. F., Rohatgi, V. K., & Székely, G. J. (1994). On multivariate skewness and kurtosis. *Theory of Probability & its Applications*, 38, 547–551.
Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press, Cambridge.
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*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rizzo, M. & Székely, G. (2019). energy: E-statistics: Multivariate inference via the energy of data. https://CRAN.R-project.org/package=energy, R package version 1.7-7.
Ryan, J. A. & Ulrich, J. M. (2019). quantmod: Quantitative financial modelling framework. https://CRAN.R-project.org/package=quantmod, R package version 0.4-15.
Shevtsova, I. G. (2013). A square bias transformation of probability distributions: Some properties and applications. *Doklady Mathematics*, 88, 388–390.
Shorack, G. R. & Wellner, J. A. (1986). Empirical processes with applications to statistics. *Wiley Series in Probability and Mathematical Statistics*, Wiley, New York.
Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9, 1135–1151.
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.
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.
Wolodzko, T. (2019). extraDistr: Additional univariate and multivariate distributions. https://CRAN.R-project.org/package=extraDistr, R package version 1.8.11.

APPENDIX

Proof of Theorem 2. Putting $t = (t_1, \dots, t_d)^\intercal \in \mathbb{R}^d$ and $Y_{n,j} = \left(Y_{n,j}^{(1)}, \dots, Y_{n,j}^{(d)}\right)^\intercal$, some algebra (using symmetry and the addition theorem for the cosine function) yields

$$\begin{aligned}
T_{n,a} &= n \int \left\| \nabla \psi_n(t) + t\psi(t) \right\|_{\mathbb{C}}^2 w_a(t)\, dt \\
&= n \int \left\| \frac{1}{n} \sum_{j=1}^n \mathrm{i} Y_{n,j} \exp\left(\mathrm{i} t^\intercal Y_{n,j}\right) + t\psi(t) \right\|_{\mathbb{C}}^2 w_a(t)\, dt \\
&= n \int \left\| \frac{1}{n} \sum_{j=1}^n \left\{ t\psi(t) - Y_{n,j} \sin\left(t^\intercal Y_{n,j}\right) + \mathrm{i} Y_{n,j} \cos\left(t^\intercal Y_{n,j}\right) \right\} \right\|_{\mathbb{C}}^2 w_a(t)\, dt \\
&= n \int \sum_{k=1}^d \left\{ \left(\frac{1}{n} \sum_{j=1}^n t^{(k)}\psi(t) - Y_{n,j}^{(k)} \sin\left(t^\intercal Y_{n,j}\right) \right)^2 + \left(\frac{1}{n} \sum_{j=1}^n Y_{n,j}^{(k)} \cos\left(t^\intercal Y_{n,j}\right) \right)^2 \right\} w_a(t)\, dt
\end{aligned}$$

$$= n \int \sum_{k=1}^{d} \left\{ t^{(k)} t^{(k)} \exp\left(-\|t\|^2\right) - \frac{2}{n} \sum_{j=1}^{n} t^{(k)} \psi(t) Y_{n,j}^{(k)} \sin\left(t^\intercal Y_{n,j}\right) \right.$$
$$\left. + \frac{1}{n^2} \sum_{i,j=1}^{n} Y_{n,j}^{(k)} Y_{n,i}^{(k)} \cos\left(t^\intercal \left(Y_{n,i} - Y_{n,j}\right)\right) \right\} w_a(t)\, dt.$$

We thus have

$$T_{n,a} = n \int \left\{ \|t\|^2 \exp\left(-(a+1)\|t\|^2\right) - \frac{2}{n} \sum_{j=1}^{n} t^\intercal Y_{n,j} \sin\left(t^\intercal Y_{n,j}\right) \exp\left(- \left(a + \frac{1}{2}\right) \|t\|^2 \right) \right.$$
$$\left. + \frac{1}{n^2} \sum_{i,j=1}^{n} Y_{n,i}^\intercal Y_{n,j} \cos\left(t^\intercal \left(Y_{n,i} - Y_{n,j}\right)\right) \exp\left(-a\|t\|^2\right) \right\} dt.$$

Using

$$\int \|t\|^2 \exp\left(-a\|t\|^2\right) dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \frac{d}{2a}, \tag{A1}$$

$$\int \cos(t^\intercal c) \exp\left(-a\|t\|^2\right) dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \exp\left(-\frac{\|c\|^2}{4a}\right), \tag{A2}$$

$$\int t^\intercal c \sin(t^\intercal c) \exp\left(-a\|t\|^2\right) dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}} \frac{\|c\|^2}{2a} \exp\left(-\frac{\|c\|^2}{4a}\right), \tag{A3}$$

the assertion follows readily. ■

Proof of Theorem 5. Recall that, in view of invariance, there is no loss of generality if we assume $X \overset{\mathcal{D}}{=} N_d(0, I_d)$. With the notation in (15), Z_n defined in (10) takes the form

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Y_{n,j} \mathrm{CS}^+\left(t, Y_{n,j}\right) - t\psi(t)\right).$$

To prove Theorem 5, we use the central limit theorem for Hilbert space valued random elements, see, for example, Theorem 2.7 of Bosq (2000). Since Z_n does not comprise independent summands, we approximate Z_n by a sum of i.i.d. random elements of $\mathbb{H}$. To this end, we introduce the auxiliary random elements

$$\tilde{Z}_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\left(X_j + \Delta_{n,j}\right) \mathrm{CS}^+\left(t, X_j\right) - t\psi(t) + X_j \mathrm{CS}^-\left(t, X_j\right) t^\intercal \Delta_{n,j}\right),$$

$$Z_n^*(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(X_j \mathrm{CS}^+\left(t, X_j\right) - \left(t + X_j + \left(2I_d - tt^\intercal\right) \frac{1}{2}\left(X_j X_j^\intercal - I_d\right) t - t^\intercal X_j t\right) \psi(t)\right) \tag{A4}$$

$$=: \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j^{**}(t),$$

where $\Delta_{n,j}$ is defined in (13).

The proof of Theorem 5 comprises three steps. We show

$$Z_n^* \xrightarrow{\mathcal{D}} Z \text{ in } \mathbb{H}, \tag{A5}$$

$$\|Z_n - \tilde{Z}_n\|_{\mathbb{H}} \xrightarrow{\mathbb{P}} 0, \tag{A6}$$

$$\|\tilde{Z}_n - Z_n^*\|_{\mathbb{H}} \xrightarrow{\mathbb{P}} 0. \tag{A7}$$

The assertion then follows from Slutsky's lemma. To prove (A5), notice that $Z_1^{**}, Z_2^{**}, \ldots$ is a sequence of i.i.d. random elements of $\mathbb{H}$. These elements are centred, since

$$\begin{aligned}\mathbb{E}[Z_1^{**}(t)] &= \mathbb{E}\left[X\mathrm{CS}^+(t,X) - \left(t + X + (2I_d - tt^\intercal)\frac{1}{2}(XX^\intercal - I_d)t - t^\intercal Xt\right)\psi(t)\right]\\ &= \mathbb{E}\big[X\mathrm{CS}^+(t,X) - t\psi(t)\big] = 0, \quad t \in \mathbb{R}^d.\end{aligned}$$

The covariance matrix kernel $\mathbb{E}\big[Z_n^*(s)Z_n^*(t)^\intercal\big] = \mathbb{E}\big[Z_1^{**}(s)Z_1^{**}(t)^\intercal\big] = K(s,t)$ (say), where $s, t \in \mathbb{R}^d$, is given by

$$\begin{aligned}K(s,t) = \mathbb{E}\Big[&\Big(X\mathrm{CS}^+(s,X) - \Big(s + X + (2I_d - ss^\intercal)\frac{1}{2}(XX^\intercal - I_d)s - s^\intercal Xs\Big)\psi(s)\Big)\\ &\Big(X\mathrm{CS}^+(t,X) - \Big(t + X + (2I_d - tt^\intercal)\frac{1}{2}(XX^\intercal - I_d)t - t^\intercal Xt\Big)\psi(t)\Big)^\intercal\Big].\end{aligned}$$

In view of $\mathbb{E}[X] = 0$ and $\mathbb{E}[XX^\intercal] = I_d$, tedious but straightforward calculations yield

$$\begin{aligned}K(s,t) =\ & \mathbb{E}\big[XX^\intercal\mathrm{CS}^+(s,X)\mathrm{CS}^+(t,X)\big] - s\psi(s)\mathbb{E}\big[X^\intercal\mathrm{CS}^+(t,X)\big] - \psi(s)\mathbb{E}\big[XX^\intercal\mathrm{CS}^+(t,X)\big]\\ &- \psi(s)\mathbb{E}\left[\left((2I_d - ss^\intercal)\frac{1}{2}(XX^\intercal - I_d) - s^\intercal X\right)sX^\intercal\mathrm{CS}^+(t,X)\right]\\ &- \mathbb{E}\big[X\mathrm{CS}^+(s,X)\big]t^\intercal\psi(t) + st^\intercal\psi(s)\psi(t) - \mathbb{E}\big[XX^\intercal\mathrm{CS}^+(s,X)\big]\psi(t) + I_d\psi(s)\psi(t)\\ &+ \mathbb{E}\left[\left((2I_d - ss^\intercal)\frac{1}{2}(XX^\intercal - I_d) - s^\intercal X\right)sX^\intercal\right]\psi(s)\psi(t)\\ &- \mathbb{E}\left[X\mathrm{CS}^+(s,X)t^\intercal\left((2I_d - tt^\intercal)\frac{1}{2}(XX^\intercal - I_d) - t^\intercal X\right)^\intercal\right]\psi(t)\\ &+ \mathbb{E}\left[Xt^\intercal\left((2I_d - tt^\intercal)\frac{1}{2}(XX^\intercal - I_d) - t^\intercal X\right)^\intercal\right]\psi(s)\psi(t)\\ &+ \mathbb{E}\left[\left((2I_d - ss^\intercal)\frac{1}{2}(XX^\intercal - I_d) - s^\intercal X\right)st^\intercal\left((2I_d - tt^\intercal)\frac{1}{2}(XX^\intercal - I_d) - t^\intercal X\right)^\intercal\right]\psi(s)\psi(t).\end{aligned}$$

Since the occurring expectations are given by

$$\begin{aligned}\mathbb{E}\big[\mathrm{CS}^+(t,X)\big] &= \psi(t),\\ \mathbb{E}\big[X\mathrm{CS}^+(t,X)\big] &= t\psi(t),\\ \mathbb{E}\big[X\mathrm{CS}^-(t,X)\big] &= -t\psi(t),\\ \mathbb{E}\big[XX^\intercal\mathrm{CS}^+(t,X)\big] &= (I_d - tt^\intercal)\psi(t),\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[XX^\intercal \mathrm{CS}^-(t,X)\right] &= (I_d - tt^\intercal)\psi(t),\\
\mathbb{E}\left[s^\intercal XXX^\intercal \mathrm{CS}^+(t,X)\right] &= \left(s^\intercal t(I_d - tt^\intercal) + st^\intercal + ts^\intercal\right)\psi(t),\\
\mathbb{E}\left[XX^\intercal \mathrm{CS}^+(s,X)\mathrm{CS}^+(t,X)\right] &= \mathbb{E}\left[XX^\intercal(\sin(s+t) + \cos(s-t))\right]\\
&= \left(I_d - (s-t)(s-t)^\intercal\right)\psi(s-t),\\
\mathbb{E}\left[s^\intercal XXX^\intercal\right] &= 0 \in \mathbb{R}^{d\times d},\\
\mathbb{E}\left[(XX^\intercal - I_d)st^\intercal(XX^\intercal - I_d)\right] &= ts^\intercal + s^\intercal t I_d,
\end{aligned}$$

some algebra shows that $K(s,t)$ takes the form given in (12). Thus, by the central limit theorem in Hilbert spaces, (A5) follows. To prove (A6), notice that

$$\begin{aligned}
\cos(t^\intercal Y_{n,j}) &= \cos(t^\intercal X_j) - \sin(t^\intercal X_j)t^\intercal \Delta_{n,j} + \varepsilon_{n,j}(t),\\
\sin(t^\intercal Y_{n,j}) &= \sin(t^\intercal X_j) + \cos(t^\intercal X_j)t^\intercal \Delta_{n,j} + \eta_{n,j}(t),
\end{aligned}$$

where

$$\max(|\varepsilon_{n,j}(t)|, |\eta_{n,j}(t)|) \le \|t\|^2\|\Delta_{n,j}\|^2. \tag{A8}$$

Hence

$$\mathrm{CS}^+(t, Y_{n,j}) = \mathrm{CS}^+(t, X_j) + \mathrm{CS}^-(t, X_j)t^\intercal \Delta_{n,j} + \varepsilon_{n,j}(t) + \eta_{n,j}(t),$$

and some algebra gives

$$Z_n(t) - \tilde{Z}_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \left((X_j + \Delta_{n,j})(\varepsilon_{n,j}(t) + \eta_{n,j}(t)) + \Delta_{n,j}\mathrm{CS}^-(t, X_j)t^\intercal \Delta_{n,j}\right).$$

Putting

$$A_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n 2\|X_j\|\|\Delta_{n,j}\|^2, \quad B_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n 2\|\Delta_{n,j}\|^2, \quad C_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n 2\|\Delta_{n,j}\|^3,$$

(A8) and the Cauchy–Schwarz inequality yield

$$\|Z_n(t) - \tilde{Z}_n(t)\| \le A_n\|t\|^2 + B_n\|t\| + C_n\|t\|^2.$$

By Theorem 5.2 of Barndorff-Nielsen (1963), we have $n^{-1/4}\max_{j=1,\dots,n}\|X_j\| \xrightarrow{a.s.} 0$. Invoking Proposition A.1 of Dörr, Ebner & Henze (2021b) gives us $n^{1/4}\max_{j=1,\dots,n}\|\Delta_{n,j}\| \xrightarrow{a.s.} 0$ and $\sum_{j=1}^n \|\Delta_{n,j}\|^2 = O_{\mathbb{P}}(1)$, from which it is readily seen that each of the expressions A_n, B_n and C_n converges to zero in probability as $n \to \infty$. In view of

$$\|Z_n - \tilde{Z}_n\|_{\mathbb{H}}^2 \le \int \left(A_n\|t\|^2 + B_n\|t\| + C_n\|t\|^2\right)^2 w_a(t)\,dt,$$

the proof of (A6) is finished. To prove (A7), we put

$$A_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(\Delta_{n,j} \mathrm{CS}^+(t, X_j) + \Big(X_j + \frac{1}{2}\big(X_j X_j^\intercal - I_d\big) t \Big)\, \psi(t) \Big),$$

$$B_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(X_j \mathrm{CS}^-(t, X_j) t^\intercal \Delta_{n,j} + \Big(\big(I_d - tt^\intercal\big) \frac{1}{2}\big(X_j X_j^\intercal - I_d\big) t - t^\intercal X_j t \Big)\, \psi(t) \Big).$$

Using the triangle inequality, some calculations give $\|\tilde{Z}_n - Z_n^*\|_{\mathbb{H}} \le \|A_n\|_{\mathbb{H}} + \|B_n\|_{\mathbb{H}}$, and thus (A7) follows provided we can show that $\|A_n\|_{\mathbb{H}} = o_{\mathbb{P}}(1)$ and $\|B_n\|_{\mathbb{H}} = o_{\mathbb{P}}(1)$. We only prove $\|A_n\|_{\mathbb{H}} = o_{\mathbb{P}}(1)$, since the reasoning for $\|B_n\|_{\mathbb{H}} = o_{\mathbb{P}}(1)$ is similar. From the definition of $\Delta_{n,j}$ in (13), we have

$$\begin{aligned} A_n(t) &= \Big(S_n^{-\frac{1}{2}} - I_d \Big) \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big(X_j \mathrm{CS}^+(t, X_j) - t\psi(t) \big) - S_n^{-\frac{1}{2}} \overline{X}_n \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big(\mathrm{CS}^+(t, X_j) - \psi(t) \big) \\ &\quad - \psi(t) \Big(S_n^{-\frac{1}{2}} - I_d \Big) \sqrt{n} \overline{X}_n + \Big(\sqrt{n} \Big(S_n^{-\frac{1}{2}} - I_d \Big) + \frac{1}{2\sqrt{n}} \sum_{j=1}^{n} \big(X_j X_j^\intercal - I_d \big) \Big) t\psi(t) \\ &= A_{n,1}(t) - A_{n,2}(t) - A_{n,3}(t) + A_{n,4}(t), \end{aligned}$$

say, and thus it remains to prove that each of $\|A_{n,k}\|_{\mathbb{H}}$, $k \in \{1, 2, 3, 4\}$, is $o_{\mathbb{P}}(1)$. Letting $\|\cdot\|_2$ denote the spectral norm, it follows that

$$\|A_{n,1}\|_{\mathbb{H}}^2 \le \Big\| \sqrt{n} \Big(S_n^{-\frac{1}{2}} - I_d \Big) \Big\|_2^2 \, \Big\| \frac{1}{n} \sum_{j=1}^{n} \big(X_j \mathrm{CS}^+(t, X_j) - t\psi(t) \big) \Big\|_{\mathbb{H}}^2 .$$

Here, the first factor on the right-hand side is $O_{\mathbb{P}}(1)$, and the second converges to zero almost surely because of the strong law of large numbers in $\mathbb{H}$. As for $\|A_{n,2}\|_{\mathbb{H}}^2$, it holds that

$$\|A_{n,2}\|_{\mathbb{H}}^2 \le \Big\| S_n^{-\frac{1}{2}} \Big\|_2^2 \Big\| \sqrt{n} \overline{X}_n \Big\|^2 \, \Big\| \frac{1}{n} \sum_{j=1}^{n} \big(\mathrm{CS}^+(t, X_j) - \psi(t) \big) \Big\|_{L^2}^2 .$$

Here, each of the first two factors on the right-hand side are $O_{\mathbb{P}}(1)$, and the last one converges to zero almost surely because of the strong law of large numbers in L^2. The term $\|A_{n,3}\|_{\mathbb{H}}^2$ is bounded from above by

$$\|A_{n,3}\|_{\mathbb{H}}^2 \le \Big\| \sqrt{n} \Big(S_n^{-\frac{1}{2}} - I_d \Big) \Big\|_2^2 \|\overline{X}_n\|^2 \int \exp(-\|t\|^2) w_a(t)\, dt.$$

Hence $\|A_{n,3}\|_{\mathbb{H}}^2 = o_{\mathbb{P}}(1)$ since $\|\overline{X}_n\|^2 = o_{\mathbb{P}}(1)$. Finally, we have

$$\|A_{n,4}\|_{\mathbb{H}}^2 \le \Big\| \sqrt{n} \Big(S_n^{-\frac{1}{2}} - I_d \Big) + \frac{1}{2\sqrt{n}} \sum_{j=1}^{n} \big(X_j X_j^\intercal - I_d \big) \Big\|_2^2 \int \|t\|^2 \exp(-\|t\|) w_a(t)\, dt.$$

From Display (2.13) of Henze & Wagner (1997), the factor preceding the integral is $o_{\mathbb{P}}(1)$, and thus $\|A_{n,4}\|_{\mathbb{H}}^2 = o_{\mathbb{P}}(1)$. The proof of Theorem 5 is completed. ■

Proof of Theorem 15

Proof. Since the proof is analogous to that given in Dörr, Ebner & Henze (2021b), it will only be sketched here. The first observation is that the quantities $\Psi_{\ell,n}(t)$, $\ell \in \{1,2,3,4\}$, defined in (32), (33) have the following almost sure limits:

$$\Psi_{1,n}(t) \xrightarrow{a.s.} \psi_X^+(t),\ \Psi_{2,n}(t) \xrightarrow{a.s.} \nabla\psi_X^+(t),\ \Psi_{3,n}(t) \xrightarrow{a.s.} -\nabla\psi_X^-(t),\ \Psi_{4,n}(t) \xrightarrow{a.s.} \mathbb{E}[XX^\intercal \mathrm{CS}^-(t,X)].$$

Here, the convergence of $\Psi_{3,n}(t)$ is assertion (a) of Lemma 6.6 of Dörr, Ebner & Henze (2021b), and the remaining claims follow, after some notational changes, the reasoning given in the proof of Lemma 6.6. of Dörr, Ebner & Henze (2021b). From (30) and (31), we have

$$L_n(s,t) = \sum_{i,j=1}^{5} L_n^{i,j}(s,t), \tag{A9}$$

where $L_n^{i,j}(s,t) = L_n^{j,i}(t,s)^\intercal$ and—putting $I_{n,j}^{\pm} := Y_{n,j}Y_{n,j}^\intercal \pm I_d$

$$L_n^{1,1}(s,t) = \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^+\left(s,Y_{n,j}\right)Y_{n,j}^\intercal \mathrm{CS}^+\left(t,Y_{n,j}\right),$$

$$L_n^{1,2}(s,t) = -\frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^+\left(s,Y_{n,j}\right)Y_{n,j}^\intercal \Psi_{1,n}(t),$$

$$L_n^{1,3}(s,t) = -\frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^+\left(s,Y_{n,j}\right)t^\intercal Y_{n,j}\Psi_{2,n}(t)^\intercal,$$

$$L_n^{1,4}(s,t) = -\frac{1}{2n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^+\left(s,Y_{n,j}\right)\Psi_{3,n}(t)^\intercal I_{n,j}^+,$$

$$L_n^{1,5}(s,t) = -\frac{1}{2n}\sum_{j=1}^{n} Y_{n,j}\mathrm{CS}^+\left(s,Y_{n,j}\right)t^\intercal I_{n,j}^- \Psi_{4,n}(t),$$

$$L_n^{2,2}(s,t) = \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\Psi_{1,n}(s)Y_{n,j}^\intercal \Psi_{1,n}(t),$$

$$L_n^{2,3}(s,t) = \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}\Psi_{1,n}(s)t^\intercal Y_{n,j}\Psi_{2,n}(t)^\intercal,$$

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

$$L_n^{2,5}(s,t) = \frac{1}{2n}\sum_{j=1}^{n} Y_{n,j}\Psi_{1,n}(s)t^\intercal I_{n,j}^- \Psi_{4,n}(t),$$

$$L_n^{3,3}(s,t) = \frac{1}{n}\sum_{j=1}^{n} s^\intercal Y_{n,j}\Psi_{2,n}(s)t^\intercal Y_{n,j}\Psi_{2,n}(t)^\intercal,$$

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

$$L_n^{3,5}(s,t) = \frac{1}{2n}\sum_{j=1}^{n} s^\intercal Y_{n,j}\Psi_{2,n}(s)t^\intercal I_{n,j}^{-}\Psi_{4,n}(t),$$

$$L_n^{4,4}(s,t) = \frac{1}{4n}\sum_{j=1}^{n} I_{n,j}^{+}\Psi_{3,n}(s)\Psi_{3,n}(t)^\intercal I_{n,j}^{+},$$

$$L_n^{4,5}(s,t) = \frac{1}{4n}\sum_{j=1}^{n} I_{n,j}^{+}\Psi_{3,n}(s)t^\intercal I_{n,j}^{-}\Psi_{4,n}(t),$$

$$L_n^{5,5}(s,t) = \frac{1}{4n}\sum_{j=1}^{n} \Psi_{4,n}(s)I_{n,j}^{-}st^\intercal I_{n,j}^{-}\Psi_{4,n}(t).$$

From (A9), it follows that $\widehat{\sigma}_{n,a}^2 = \sum_{i,j=1}^5 \widehat{\sigma}_{n,a}^{i,j}$, where

$$\widehat{\sigma}_{n,a}^{i,j} = 4\iint z_n(s)^\intercal L_n^{i,j}(s,t) z_n(t) w_a(s) w_a(t)\, ds\, dt. \tag{A10}$$

Notice that $\widehat{\sigma}_{n,a}^{i,j} = \widehat{\sigma}_{n,a}^{j,i}$. In view of (23) and (24), we have $L(s,t) = \sum_{i,j=1}^5 L^{i,j}(s,t)$, where $L^{i,j}(s,t) = \mathbb{E}[w_i(s,X)w_j(t,X)^\intercal]$, and

$$w_1(t,X) = X\mathrm{CS}^+(t,X), \quad w_2(t,X) = -X\psi_X^+(t), \quad w_3(t,X) = -t^\intercal X\nabla\psi_X^+(t),$$
$$w_4(t,X) = \frac{1}{2}\left(XX^\intercal + I_d\right)\nabla\psi_X^-(t), \quad w_5(t,X) = -\frac{1}{2}\mathbb{E}[XX^\intercal \mathrm{CS}^-(t,X)](XX^\intercal - I_d)t.$$

Therefore, $\sigma_a^2 = \sum_{i,j=1}^5 \sigma_a^{i,j}$, where

$$\sigma_a^{i,j} = 4\iint z(s)^\intercal L^{i,j}(s,t) z(t) w_a(s) w_a(t)\, ds\, dt$$

and, by symmetry, $L^{i,j}(s,t) = L^{j,i}(t,s)^\intercal$ and hence $\sigma_a^{i,j} = \sigma_a^{j,i}$. We thus have to prove $\widehat{\sigma}_{n,a}^{i,j} \xrightarrow{\mathbb{P}} \sigma_a^{i,j}$ for each choice of $i,j \in \{1,\dots,5\}$. To this end, we proceed in two steps. The first one is to replace $L_n^{i,j}(s,t)$ in (A10) with $L_{n,0}^{i,j}(s,t)$. Here, $L_{n,0}^{i,j}(s,t)$ originates from $L_n^{i,j}(s,t)$ by replacing each $Y_{n,j}$ with X_j, and this replacement also affects the quantities $\Psi_{\ell,n}(t)$, $\ell \in \{1,\dots,4\}$. Moreover, we replace $z_n(t)$ with $z_{n,0}(t) = n^{-1}\sum_{j=1}^n X_j \mathrm{CS}^+(t,X_j) - t\psi(t)$. Putting

$$\widehat{\sigma}_{n,0,a}^{i,j} = 4\iint z_{n,0}(s)^\intercal L_{n,0}^{i,j}(s,t) z_{n,0}(t) w_a(s) w_a(t)\, ds\, dt,$$

it follows from Fubini's theorem that $\widehat{\sigma}^{i,j}_{n,0,a} \xrightarrow{\mathbb{P}} \sigma^{i,j}_a$. The second, much more technical, step is to prove $\widehat{\sigma}^{i,j}_{n,a} - \widehat{\sigma}^{i,j}_{n,0,a} = o_{\mathbb{P}}(1)$. To this end, notice that

$$\begin{aligned} z_n(s)^\intercal L^{i,j}_n(s,t)z_n(t) - z_{n,0}(s)^\intercal L^{i,j}_{n,0}(s,t)z_{n,0}(t) = {} & z_n(s)^\intercal \left(L^{i,j}_n(s,t) - L^{i,j}_{n,0}(s,t)\right)z_n(t) \\ & + \left(z_n(s) - z_{n,0}(s)\right)^\intercal L^{i,j}_{n,0}(s,t)z_n(t) \\ & + z_{n,0}(s)^\intercal L^{i,j}_{n,0}(s,t)\left(z_n(t) - z_{n,0}(t)\right), \end{aligned} \tag{A11}$$

where

$$\left|\left(z_n(s) - z_{n,0}(s)\right)^\intercal L^{i,j}_{n,0}(s,t)z_n(t)\right| \le \|z_n(s) - z_{n,0}(s)\| \|L^{i,j}_{n,0}(s,t)\|_2 \|z_n(t)\|,$$
$$\left|z_{n,0}(s)^\intercal L^{i,j}_{n,0}(s,t)\left(z_n(t) - z_{n,0}(t)\right)\right| \le \|z_{n,0}(s)\| \|L^{i,j}_{n,0}(s,t)\|_2 \|z_n(t) - z_{n,0}(t)\|.$$

We have $\|z_{n,0}(t)\| \le 2n^{-1}\sum_{j=1}^n \|X_j\| + \|t\|\psi(t)$, and a Taylor expansion yields

$$\|z_n(t)\| \le \frac{2}{n}\sum_{j=1}^n \left(\|X_j\| + \|X_j\|\|t\|\|\Delta_{n,j}\| + \|\Delta_{n,j}\| + \|t\|\|\Delta_{n,j}\|^2\right) + \|t\|\psi(t),$$
$$\|z_n(t) - z_{n,0}(t)\| \le \frac{2}{n}\sum_{j=1}^n \|\Delta_{n,j}\| + \frac{2\|t\|}{n}\sum_{j=1}^n \|\Delta_{n,j}\|\|X_j\|.$$

Notice that each of the terms $\|L^{i,j}_{n,0}(s,t)\|_2$ is bounded from above by terms of the type $2^k\|s\|^\ell\|t\|^m$, multiplied with finitely many products of the type $n^{-1}\sum_{j=1}^n \|X_j\|^\beta$, with $k \le 2$, $\ell, m \in \{0,1\}$, and $\beta \in \{1,2,3,4\}$. In view of the condition $\mathbb{E}\|X\|^4 < \infty$ and the fact that $n^{-1}\sum_{j=1}^n \|\Delta_{n,j}\|^k\|X_k\|^\ell \xrightarrow{a.s.} 0$ (see Proposition A.2 of Dörr, Ebner & Henze, 2021b), it follows that

$$\iint \left|\left(z_n(s) - z_{n,0}(s)\right)^\intercal L^{i,j}_{n,0}(s,t)z_n(t)\right| w_a(s)w_a(t)\,ds\,dt \xrightarrow{\mathbb{P}} 0,$$
$$\iint \left|z_{n,0}(s)^\intercal L^{i,j}_{n,0}(s,t)\left(z_n(t) - z_{n,0}(t)\right)\right| w_a(s)w_a(t)\,ds\,dt \xrightarrow{\mathbb{P}} 0.$$

As a consequence, we only have to consider the first term on the right-hand side of (A11). To/ this end, notice that

$$\left|z_n(s)^\intercal \left(L^{i,j}_n(s,t) - L^{i,j}_{n,0}(s,t)\right)z_n(t)\right| \le \|z_n(s)\| \|L^{i,j}_n(s,t) - L^{i,j}_{n,0}(s,t)\|_2 \|z_n(t)\|.$$

To find an upper bound for $\|L^{i,j}_n(s,t) - L^{i,j}_{n,0}(s,t)\|_2$, we have to consider each case $i, j \in \{1, \dots, 5\}$ such that $i \le j$ separately. We will elaborate on the case $i = j = 1$; the other cases are treated similarly. Putting $\mathrm{CS}^+(s,t,\xi) = \mathrm{CS}^+(s,\xi)\mathrm{CS}^+(t,\xi)$, we have

$$\|L^{1,1}_n(s,t) - L^{1,1}_{n,0}(s,t)\|_2 = \left\| \frac{1}{n}\sum_{j=1}^n \left(Y_{n,j}Y_{n,j}^\intercal \mathrm{CS}^+\left(s,t,Y_{n,j}\right) - X_jX_j^\intercal \mathrm{CS}^+\left(s,t,X_j\right)\right)\right\|_2,$$

and a Taylor expansion yields

$$\begin{aligned}\left\|L_n^{1,1}(s,t)-L_{n,0}^{1,1}(s,t)\right\|_2 \leq \frac{4}{n}\sum_{j=1}^{n}\|X_j\|^2\big(\|t\|\|\Delta_{n,j}\|+\|s\|\|\Delta_{n,j}\|\big)+\frac{4}{n}\sum_{j=1}^{n}\|X_j\|^2\big(\|s\|\|t\|\|\Delta_{n,j}\|^2\big)\\ +\frac{8}{n}\sum_{j=1}^{n}\|X_j\|\|\Delta_{n,j}\|\big(1+\|s\|\|\Delta_{n,j}\|\big)\big(1+\|t\|\|\Delta_{n,j}\|\big)\\ +\frac{4}{n}\sum_{j=1}^{n}\|\Delta_{n,j}\|^2\big(1+\|s\|\|\Delta_{n,j}\|\big)\big(1+\|t\|\|\Delta_{n,j}\|\big).\end{aligned}$$

From Proposition A.2 of Dörr, Ebner & Henze (2021b), it follows that $\|L_n^{1,1}(s,t)-L_{n,0}^{1,1}(s,t)\|_2 \xrightarrow{a.s.} 0$.

To prove (35), we need the integrals

$$\begin{aligned}L_{1,a}(x) &:= \int t\psi(t)\mathrm{CS}^+(t,x)w_a(t)\,dt = \frac{(2\pi)^{\frac{d}{2}}}{(2a+1)^{\frac{d}{2}+1}}x\exp\left(-\frac{\|x\|^2}{4a+2}\right),\\ L_{2,a}(x) &:= \int tt^\intercal x\psi(t)\mathrm{CS}^-(t,x)w_a(t)\,dt\\ &= \frac{(2\pi)^{\frac{d}{2}}}{(2a+1)^{\frac{d}{2}+2}}\big((2a+1)x-\|x\|^2x\big)\exp\left(-\frac{\|x\|^2}{4a+2}\right),\\ I_{1,a}(x,y) &:= \int \mathrm{CS}^+(t,x)\mathrm{CS}^+(t,y)w_a(t)\,dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}}\exp\left(-\frac{\|x-y\|^2}{4a}\right),\\ I_{2,a}(x,y) &:= \int t\mathrm{CS}^+(t,x)\mathrm{CS}^-(t,y)w_a(t)\,dt = \left(\frac{\pi}{a}\right)^{\frac{d}{2}}\frac{(x-y)}{2a}\exp\left(-\frac{\|x-y\|^2}{4a}\right).\end{aligned}$$

Putting

$$\begin{aligned}P_{1,a}^{i,j} &:= Y_{n,i}^\intercal Y_{n,j}I_{1,a}(Y_{n,i},Y_{n,j})-L_{1,a}(Y_{n,j})^\intercal Y_{n,j},\\ P_{2,a}^{i,j,k} &:= Y_{n,i}^\intercal Y_{n,j}I_{1,a}(Y_{n,i},Y_{n,k})-L_{1,a}(Y_{n,k})^\intercal Y_{n,j},\\ P_{3,a}^{i,j,k} &:= Y_{n,i}^\intercal Y_{n,k}Y_{n,j}^\intercal I_{2,a}(Y_{n,i},Y_{n,k})-Y_{n,j}^\intercal L_{2,a}(Y_{n,k}),\\ P_{4,a}^{i,j,k} &:= Y_{n,i}^\intercal\big(Y_{n,j}Y_{n,j}^\intercal+I_d\big)Y_{n,k}I_{1,a}(Y_{n,i},Y_{n,k})-Y_{n,k}^\intercal\big(Y_{n,j}Y_{n,j}^\intercal+I_d\big)L_{1,a}(Y_{n,k}),\\ P_{5,a}^{i,j,k} &:= Y_{n,i}^\intercal Y_{n,k}Y_{n,k}^\intercal\big(Y_{n,j}Y_{n,j}^\intercal-I_d\big)I_{2,a}(Y_{n,i},Y_{n,k})-Y_{n,k}^\intercal\big(Y_{n,j}Y_{n,j}^\intercal-I_d\big)L_{2,a}(Y_{n,k}),\end{aligned}$$

straightforward calculations give

$$\widehat{\sigma}_{n,a}^{1,1} = \frac{4}{n^3}\sum_{i,j,k=1}^{n}P_{1,a}^{i,j}P_{1,a}^{k,j}, \quad \widehat{\sigma}_{n,a}^{1,2} = -\frac{4}{n^4}\sum_{i,j,k,\ell=1}^{n}P_{1,a}^{i,j}P_{2,a}^{\ell,j,k},$$

$$\widehat{\sigma}_{n,a}^{1,3} = -\frac{4}{n^4} \sum_{i,j,k,\ell=1}^{n} P_{1,a}^{i,j} P_{3,a}^{\ell,j,k}, \quad \widehat{\sigma}_{n,a}^{1,4} = -\frac{2}{n^4} \sum_{i,j,k,\ell=1}^{n} P_{1,a}^{i,j} P_{4,a}^{\ell,j,k},$$

$$\widehat{\sigma}_{n,a}^{1,5} = -\frac{2}{n^4} \sum_{i,j,k,\ell=1}^{n} P_{1,a}^{i,j} P_{5,a}^{\ell,j,k}, \quad \widehat{\sigma}_{n,a}^{2,2} = \frac{4}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{2,a}^{i,j,k} P_{2,a}^{m,j,\ell},$$

$$\widehat{\sigma}_{n,a}^{2,3} = \frac{4}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{2,a}^{i,j,k} P_{3,a}^{m,j,\ell}, \quad \widehat{\sigma}_{n,a}^{2,4} = \frac{2}{n^5} \sum_{i,j,k,l,m=1}^{n} P_{2,a}^{i,j,k} P_{4,a}^{m,j,\ell},$$

$$\widehat{\sigma}_{n,a}^{2,5} = \frac{2}{n^5} \sum_{i,j,k,l,m=1}^{n} P_{2,a}^{i,j,k} P_{5,a}^{m,j,\ell}, \quad \widehat{\sigma}_{n,a}^{3,3} = \frac{4}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{3,a}^{i,j,k} P_{3,a}^{m,j,\ell},$$

$$\widehat{\sigma}_{n,a}^{3,4} = \frac{2}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{3,a}^{i,j,k} P_{4,a}^{m,j,\ell}, \quad \widehat{\sigma}_{n,a}^{3,5} = \frac{2}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{3,a}^{i,j,k} P_{5,a}^{m,j,\ell},$$

$$\widehat{\sigma}_{n,a}^{4,4} = \frac{1}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{4,a}^{i,j,k} P_{4,a}^{m,j,\ell}, \quad \widehat{\sigma}_{n,a}^{4,5} = \frac{1}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{4,a}^{i,j,k} P_{5,a}^{m,j,\ell},$$

$$\widehat{\sigma}_{n,a}^{5,5} = \frac{1}{n^5} \sum_{i,j,k,\ell,m=1}^{n} P_{5,a}^{i,j,k} P_{5,a}^{m,j,\ell}. \tag{A12}$$

■

Received 4 August 2020
Accepted 1 July 2021