



Testing normality of a large number of populations

M. D. Jiménez-Gamero¹

Received: 8 March 2022 / Revised: 10 October 2022 / Accepted: 12 December 2022 /
Published online: 12 February 2023
© The Author(s) 2023

Abstract

This paper studies the problem of simultaneously testing that each of k independent samples come from a normal population. The means and variances of those populations may differ. The proposed procedures are based on the BHEP test and they allow k to increase, which can be even larger than the sample sizes.

Keywords Goodness-of-fit · Normality · Large number of populations

1 Introduction

Testing for normality is a topic of interest that has generated and is still generating a vast literature. Some recent contributions are Ebner et al. (2022), Henze and Jiménez-Gamero (2019), Henze et al. (2019), Henze and Koch (2020), and Jelíto and Pitera (2021); see the paper by Ebner and Henze (2020) for a review on normality tests. Most papers on this issue deal with testing normality for a sample, and the properties of the proposed procedures are stated as the sample size increases. This paper studies the problem of simultaneously testing normality of k univariate samples, where k can increase with the sample sizes. Moreover, k will be allowed to be even larger than the sample sizes. Specifically, we will consider the following general setting:

Let $\mathbf{X}_1 = \{X_{1,1}, \dots, X_{1,n_1}\}, \dots, \mathbf{X}_k = \{X_{k,1}, \dots, X_{k,n_k}\}$ be k independent samples with sizes n_1, \dots, n_k , which may be different, coming from $X_1, \dots, X_k \in \mathbb{R}$, with continuous distribution functions F_1, \dots, F_k , respectively, and $E(X_j^2) < \infty, 1 \leq j \leq k$. (1)

The elements of each sample in (1) are assumed to be independent. In this setting, we deal with the problem of testing

✉ M. D. Jiménez-Gamero
dolores@us.es

¹ Department of Statistics and Operations Research, University of Seville, Seville, Spain

$$H_0 : F_1, \dots, F_k \in \mathcal{N},$$

against general alternatives,

$$H_1 : F_j \notin \mathcal{N}, \text{ for some } 1 \leq j \leq k,$$

where \mathcal{N} is the set of univariate normal populations, $\mathcal{N} = \{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0\}$ and, as said before, k is allowed to be large (the precise meaning of “large” will be stated in the following sections).

The main motivation for testing normality comes from the fact that, under this distributional assumption, many statistical procedures become simpler and more efficient than their non-parametric counterparts, mainly due to the good properties of the normal law. Nevertheless, the efficiency of those procedures may decrease, or even disappear, if the normality assumption fails. For example, if one can assume that the populations are normal, then the classical k -sample problem becomes that of testing the equality of variances and the equality of means, for which more tests can be found in the statistical literature than for testing the equality of the k distributions when k is large (Zhan and Hart 2014; Jiménez-Gamero et al. 2022). As another instance, let us consider the problem of testing the equality of the means of a large number k of univariate normal populations, that may have different variances. Park and Park (2012) proposed two tests for this problem, whose associated statistics, conveniently normalized, are asymptotically normal. Here by asymptotic it is meant as $k \rightarrow \infty$. The assumption that the populations have all of them a normal distribution is crucial in order to derive the asymptotic distribution of those test statistics. In fact, the simulations in Jiménez-Gamero and Franco-Pereira (2021) show that, when the data meets the normality assumption, these tests can be more powerful than nonparametric competitors but, when data come from non-normal populations, the empirical type I errors of the tests in Park and Park (2012) can be far apart from the nominal value.

The problem of simultaneously testing goodness-of-fit for k populations has been studied in Gaigall (2021) by using test statistics based on comparing the empirical distribution function of each sample with a parametric estimator derived under the null hypothesis. The asymptotic properties studied in Gaigall (2021) are for fixed k and increasing sample sizes. Jiménez-Gamero et al. (2005) studied the problem of testing normality of the errors in multivariate, homoscedastic linear models. The test statistic in Jiménez-Gamero et al. (2005) is based on comparing the empirical characteristic function (ECF) of the studentized residuals with the characteristic function (CF) of a standard normal law. The asymptotic properties studied in Jiménez-Gamero et al. (2005) allow k to increase with the sample sizes in such a way that $k^2/n = o(1)$, where n is the sample size. Specifically, they show that the asymptotic null distribution of the considered test statistic coincides with that derived for independent, identically distributed (iid) data in Baringhaus and Henze (1988) and Henze and Wagner (1997). The normality test in Baringhaus and Henze (1988) and Henze and Wagner (1997) is usually called the BHEP test, since it was first proposed for the univariate case by Epps and Pulley (1983), and then extended to the multivariate case by Baringhaus and Henze (1988); moreover, due to its nice properties, it has been extended in several directions: to testing normality of the errors in homoscedastic linear models in Jiménez-Gamero

et al. (2005), as explained before; to testing normality of the errors in nonparametric regression in Hušková and Meintanis (2010) and Rivas-Martínez and Jiménez-Gamero (2018); to testing normality of the innovations in GARCH models in Jiménez-Gamero (2014), Klar et al. (2012); and to testing Gaussianity of random elements taking values in a Hilbert space in Henze and Jiménez-Gamero (2021), just to cite a few.

In this paper we first study the test in Jiménez-Gamero et al. (2005) for testing H_0 , without assuming that the populations are homoscedastic. Not assuming homoscedasticity greatly complicates theoretical derivations, since instead of estimating one variance from the pooled data, now we must deal with k variance estimators. With this aim, it will be assumed that the sample sizes are comparable in the following sense:

$$n_i = c_i m, \quad 0 < c_0 \leq c_i \leq C_0 < \infty, \quad \forall i, \quad \text{for some fixed constants } c_0 \text{ and } C_0. \quad (2)$$

It is shown that the asymptotic null distribution of the test statistic also coincides with that derived for iid data, whenever $k/m = o(1)$. Notice that if the sample sizes satisfy (2), then the condition $k/m \rightarrow 0$ is equivalent to $k^2/n \rightarrow 0$.

Since the practical calculation of the BHEP test statistic involves $O(n^2)$ sums (see Section 2), its computation can be rather time-consuming for large k . So, for the case $k/m \rightarrow \ell \in (0, \infty]$, we try other strategies for testing H_0 . First, inspired by the random projection procedure in Cuesta-Albertos et al. (2006), we could test H_0 using not all data sets but a "small" number k_0 of samples (small in the sense that $k_0/m = o(1)$) randomly selected from all of the k population samples. Finally, we consider a test statistic which combines the BHEP test statistics calculated in each sample.

As said before, this paper studies BHEP-based statistics for testing H_0 . Other test statistics could be considered. The main reason for our choice is the good properties enjoyed by this test. This is why, Section 2 starts by reviewing its definition and some properties. This section also derives new properties that will be used in Section 5. Specifically, it is shown that the two first moments of the null distribution of the BHEP test statistic converge to those of the asymptotic null distribution, and sufficient conditions are given for such convergence to hold under alternatives. Section 3 studies the test that compares the ECF of the studentized data with the CF of the standard normal law which, in our view, is the natural extension of the BHEP test statistic to the setting in (1). Section 4 studies the test in the previous section when it is calculated in a subset of randomly selected samples. Section 5 studies a test whose test statistic is based on the sum of the BHEP test statistics calculated in each sample. The properties investigated in Sects. 3–5 are asymptotic. In order to assess the finite sample performance of the proposals, a simulation study was carried out, whose results are reported in Sect. 6. Section 7 summarizes the paper and comments on extensions and further research. All proofs are deferred to the last section.

Throughout the paper we will make use of the following standard notation: $i = \sqrt{-1}$ is the imaginary unit; for any complex number $x = a + ib \in \mathbb{C}$, with $a, b \in \mathbb{R}$, $\Re x = a$ denotes the real part and $\Im x = b$ denotes the imaginary part; all random variables and random elements will be defined on a sufficiently rich probability space (Ω, \mathcal{A}, P) ;

the symbols E and V denote expectation and variance, respectively; P_0, E_0 and V_0 denote probability, expectation and variance under the null hypothesis, respectively; $\xrightarrow{\mathcal{D}}$ means convergence in distribution of random vectors and random elements.

2 The BHEP test

2.1 The test

This section revisits the BHEP test for univariate data. Let X_1, \dots, X_n ($n \geq 2$) be a sample from a random variable X with continuous distribution function F and $E(X^2) < \infty$. For testing the hypothesis $H_{0,1} : F \in \mathcal{N}$, the rationale of the BHEP test is as follows: write $\bar{X} = n^{-1} \sum_{j=1}^n X_j$ and $S^2 = n^{-1} \sum_{j=1}^n (X_j - \bar{X})^2$ for the sample mean and the sample variance, respectively, and let $Y_j = (X_j - \bar{X})/S$, $1 \leq j \leq n$, be the so-called *scaled residuals* of X_1, \dots, X_n , which provide an empirical standardization of X_1, \dots, X_n . Notice that, under the assumptions made, $P(S > 0) = 1$, and thus Y_1, \dots, Y_n are well defined. Since, under $H_{0,1}$ and for large n , the distribution of the scaled residuals should be close to the standard normal distribution, it is tempting to compare the ECF of Y_1, \dots, Y_n ,

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j), \quad t \in \mathbb{R},$$

with $\varphi_0(t) = \exp(-t^2/2)$, which is the CF of the standard normal distribution. The BHEP test rejects $H_{0,1}$ for large values of the weighted L^2 -statistic

$$\mathcal{T}_{n,\beta} = \int |\varphi_n(t) - \varphi_0(t)|^2 w_\beta(t) dt, \tag{3}$$

where an unspecified integral stands for an integral over the whole real line, $w_\beta(t)$ is the probability density function of the normal distribution $N(0, \beta^2)$, and $\beta > 0$ is a parameter, that must be fixed by the user. The test statistic $\mathcal{T}_{n,\beta}$ may be written as

$$\begin{aligned} \mathcal{T}_{n,\beta} = & \frac{1}{n^2} \sum_{j,k=1}^n \exp\left(-\frac{\beta^2}{2}(Y_j - Y_k)^2\right) \\ & - \frac{2}{n(1 + \beta^2)^{1/2}} \sum_{j=1}^n \exp\left(-\frac{\beta^2}{2(1 + \beta^2)} Y_j^2\right) + \frac{1}{(1 + 2\beta^2)^{1/2}}, \end{aligned} \tag{4}$$

which is a useful expression for the practical computation of $\mathcal{T}_{n,\beta}$. Notice that the computation of $\mathcal{T}_{n,\beta}$ involves a double sum, so the number of required calculations is of order $O(n^2)$.

Representation (4) also shows that $\mathcal{T}_{n,\beta}$ is a function of the products $Y_j Y_k = (X_j - \bar{X})(X_k - \bar{X})/S^2$, $1 \leq j, k \leq n$, and thus it is invariant with respect to affine

transformations of X_1, \dots, X_n . This property implies that the null distribution of $\mathcal{T}_{n,\beta}$ only depends on the sample size n , and on the value of β .

Critical points for several sample sizes, $\beta = 1$ and the usual values for the probability of type I error (level) can be found in Baringhaus and Henze (1988) and Henze (1990). The function `cv.quan` of the package `mnt` (Butsch and Ebner 2020) of the R language (Core Team 2020) can be used to calculate critical points of the null distribution of $n\mathcal{T}_{n,\beta}$ for any sample size, any value of β and any level. The critical points can be also approximated by those of the asymptotic null distribution of $n\mathcal{T}_{n,\beta}$. Under the null hypothesis, $n\mathcal{T}_{n,\beta}$ is asymptotically (when $n \rightarrow \infty$) distributed as $W_\beta = \sum_{\ell=1}^{\infty} \lambda_{\beta,\ell} Z_\ell^2$, where $\lambda_{\beta,1}, \lambda_{\beta,2}, \dots$ are the descending sequence of positive eigenvalues of certain integral operator, and Z_1, Z_2, \dots are independent standard normal random variables. Since those eigenvalues can be estimated (see, for example, Ebner and Henze 2021, 2022; Meintanis et al. 2022) and hence the cumulants of W_β , one could approximate the asymptotic critical values by using the Pearson system of distributions (with the help of the package `PEARSONDS` (Becker and Klößner 2022) of the R language (Core Team 2020)). This idea was proposed by Henze (1990), who (exactly) calculated the first four cumulants of W_β with $\beta = 1$.

The BHEP test is consistent against any fixed alternative and it is able to detect continuous alternatives converging to the null at the rate $n^{-1/2}$, see (Henze and Wagner 1997). Ebner and Henze (2021) and Meintanis et al. (2022) have obtained approximate Bahadur efficiencies of the BHEP test, showing that it outperforms tests based on the empirical distribution function over certain close alternatives to normality. All these properties are for $n \rightarrow \infty$.

2.2 The mean and the variance of the BHEP test statistic

Henze and Wagner (1997) (see also Henze (1990) for $\beta = 1$) have (exactly) calculated the mean and the second a third centred moments of W_β . Specifically, for $\beta = 1$, the mean and the variance of the asymptotic null distribution of $n\mathcal{T}_{n,1}$ are $\mu_0 = 1 - \sqrt{3}/2 \approx 0.13397$ and $\sigma_0^2 = 2/\sqrt{5} + 5/6 - 155/64\sqrt{2} \approx 0.015236$, respectively. So, it is tempting to approximate $\mu_{0,n} = E_0(n\mathcal{T}_{n,1})$ by means of μ_0 and $\tau_{0,n}^2 = V_0(n\mathcal{T}_{n,1})$ by σ_0^2 , where E_0 and V_0 are both understood under $H_{0,1}$. Next proposition shows that those approximations are asymptotically valid, which implies that $\{(n\mathcal{T}_{n,\beta})^2\}$ is uniformly integrable.

Proposition 1 *Let X_1, \dots, X_n be a random sample from $X \sim N(0, 1)$. Let $\mathcal{T}_{n,\beta}$ be as defined in (3). Let $\mu_{0,\beta}$ and $\sigma_{0,\beta}^2$ denote the mean and the variance of the asymptotic distribution (as $n \rightarrow \infty$) of $n\mathcal{T}_{n,\beta}$, respectively. Then, $E_0(n\mathcal{T}_{n,\beta}) \rightarrow \mu_{0,\beta}$ and $V_0(n\mathcal{T}_{n,\beta}) \rightarrow \sigma_{0,\beta}^2$, as $n \rightarrow \infty$.*

As in the first paragraph of this subsection, when $\beta = 1$ we denote $\mu_{0,\beta}$ and $\sigma_{0,\beta}^2$ as μ_0 and σ_0^2 , respectively. We have numerically checked the approximations for $\beta = 1$, $\mu_{0,n} \approx \mu_0$ and $\tau_{0,n}^2 \approx \sigma_0^2$, in finite sample sizes. For each n , the true values of $\mu_{0,n}$ and $\tau_{0,n}^2$ were calculated by simulation, based on 100,000 samples of size n from a standard normal law: $n\mathcal{T}_{n,1}$ were computed at each sample, and then the sample mean

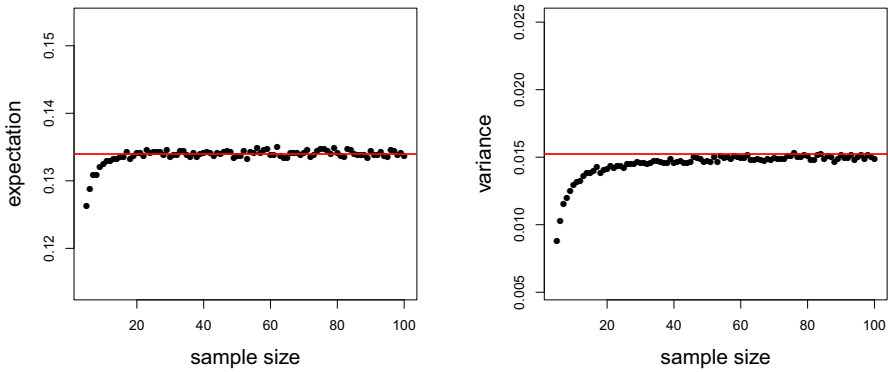


Fig. 1 True values for $\mu_{0,n}$ (left panel) and $\tau_{0,n}^2$ (right panel), $5 \leq n \leq 100$. The line $y = \mu_0$ (left panel) and the line $y = \sigma_0^2$ (right panel) are in red

and the sample variance of these 100,000 values were used to approximate $\mu_{0,n}$ and $\tau_{0,n}^2$, respectively. Figure 1 displays $\mu_{0,n}$, joint with the line $y = \mu_0$ in red, and $\tau_{0,n}^2$, joint with the line $y = \sigma_0^2$ in red, for $5 \leq n \leq 100$. Looking at this figure one can see that the approximation for the mean, $\mu_{0,n} \approx \mu_0$, is almost exact for $n \geq 20$, and the approximation for the variance, $\tau_{0,n}^2 \approx \sigma_0^2$, works really well for $n \geq 50$.

Now assume that $E(X) = 0$, $V(X) = 1$, and that the CF of X is $\varphi_X \neq \varphi_0$, which is tantamount to $\Delta_{X,\beta} = \int |\varphi_X(t) - \varphi_0(t)|^2 w_\beta(t) dt > 0$. In this setting, if $E(X^2) < \infty$, the proof of Theorem 3.1 in Ebner and Henze (2021) shows that $\mu_{n,\beta} = E(\mathcal{T}_{n,\beta}) \rightarrow \Delta_{X,\beta}$ and $V(\mathcal{T}_{n,\beta}) \rightarrow 0$, as $n \rightarrow \infty$. So, we can approximate $\mu_{n,\beta} \approx \Delta_{X,\beta}$. Example 1 in Baringhaus et al. (2017) shows that if $E(X^4) < \infty$, then $\sqrt{n}(\mathcal{T}_{n,\beta} - \Delta_{X,\beta}) \xrightarrow{D} N(0, \sigma_{X,\beta}^2)$, as $n \rightarrow \infty$. The expression of $\sigma_{X,\beta}^2 > 0$ is given in Baringhaus et al. (2017). If, in addition, $\{(\sqrt{n}(\mathcal{T}_{n,\beta} - \Delta_{X,\beta}))^2\}$ is uniformly integrable, then we can approximate $\tau_{n,\beta}^2 = V(\sqrt{n}\mathcal{T}_{n,\beta}) \approx \sigma_{X,\beta}^2$. Similar steps to those given in the proof of Proposition 1 show that a sufficient condition for the uniform integrability of $\{(\sqrt{n}(\mathcal{T}_{n,\beta} - \Delta_{X,\beta}))^2\}$ is that

$$E\{|\sqrt{n}(S - 1)|^{4+\delta}\} < \infty \quad \text{and} \quad E\{|\sqrt{n}\bar{X}|^{4+\delta}\} < \infty, \tag{5}$$

for some $\delta > 0$.

3 The BHEP test for H_0

The tests in Baringhaus and Henze (1988), Epps and Pulley (1983) chose $\beta = 1$. In order to simplify notation, in our developments we will also choose $\beta = 1$, although all results keep on being true for arbitrary (but fixed) β .

Let $Y_{j,r} = (X_{j,r} - \bar{X}_j)/S_j$, $1 \leq r \leq n_j$, $1 \leq j \leq k$, where \bar{X}_j and S_j^2 stand for the sample mean and the sample variance of the sample from X_j , respectively, $1 \leq j \leq k$. As in the one-sample case, under H_0 , the distribution of the scaled residuals should

be close to the standard normal distribution. So we consider as test statistic

$$T_{k,n} = \int |\varphi_n(t) - \varphi_0(t)|^2 w(t) dt,$$

where now φ_n is the ECF of $Y_{1,1}, \dots, Y_{1,n_1}, \dots, Y_{k,1}, \dots, Y_{k,n_k}$,

$$\varphi_n = \frac{1}{n} \sum_{j=1}^k \sum_{r=1}^{n_j} \exp(itY_{j,r}), \quad t \in \mathbb{R},$$

and $w(t) = w_1(t)$ is the probability density function of the standard normal law. Notice that, to be more precise, the proposed test statistic should be denoted as T_{n_1, \dots, n_k} but, to simplify notation, we just write $T_{k,n}$. Let $n = \sum_{i=1}^k n_i$. With this notation, the one-sample test statistic is $T_{1,n}$.

Since we are assuming that $E(X_j^2) < \infty, 1 \leq j \leq k$, we can write $X_{j,r} = \mu_j + \sigma_j W_{j,r}$, where $\mu_j = E(X_j), \sigma_j^2 = V(X_j), E(W_{j,r}) = 0, V(W_{j,r}) = 1, 1 \leq r \leq n_j, 1 \leq j \leq k$; moreover, the scaled residuals $Y_{j,r}$ calculated from $X_{j,1}, \dots, X_{j,n_j}$ coincide with those calculated from $W_{j,1}, \dots, W_{j,n_j}$. Because of this reason, we can assume that $E(X_j) = 0$ and $V(X_j) = 1, 1 \leq j \leq k$. Accordingly, from now on, instead of (1), it will be assumed that the setting is as follows:

Let $\mathbf{X}_1 = \{X_{1,1}, \dots, X_{1,n_1}\}, \dots, \mathbf{X}_k = \{X_{k,1}, \dots, X_{k,n_k}\}$ be k independent samples with sizes n_1, \dots, n_k , which may be different, coming from X_1, \dots, X_k , with continuous distribution functions F_1, \dots, F_k , respectively,
 $E(X_j) = 0$ and $V(X_j) = 1, 1 \leq j \leq k.$ (6)

Another consequence of assuming that $E(X_j) = 0$ and $V(X_j) = 1, 1 \leq j \leq k$, is that the null distribution of $T_{k,n}$ only depends on the sample sizes n_1, \dots, n_k .

As in the one-sample case, rejection of the null hypothesis H_0 is for large values of $T_{k,n}$, say $T_{k,n} > t_{k,n,\alpha}$ where $t_{k,n,\alpha}$ is the α upper percentile of the null distribution of $T_{k,n}$. So, to test H_0 we must calculate upper percentiles of the null distribution of $T_{k,n}$. Although the critical points can be calculated by simulation, from a practical point of view it would be nice if they could be approximated in some fashion. The next result is useful in that sense.

Before stating it, we introduce some notation. Since $w(t) = w(-t)$ we can write $T_{k,n} = \int Z_{k,n}^2(t)w(t) dt$, with

$$Z_{k,n}(t) = \frac{1}{n} \sum_{j,r} \cos(tY_{j,r}) + \frac{1}{n} \sum_{j,r} \sin(tY_{j,r}) - \varphi_0(t), \quad t \in \mathbb{R}. \quad (7)$$

Let L_w^2 denote the separable Hilbert space of (equivalence classes of) measurable functions $f : \mathbb{R} \mapsto \mathbb{C}$ satisfying $\int |f(t)|^2 w(t) dt < \infty$. The scalar product and the

resulting norm in L^2_w will be denoted by $\langle f, g \rangle_w = \int f(t)\overline{g(t)}w(t)dt$ and $\|f\|_w^2 = \int |f(t)|^2w(t)dt$, respectively. With this notation, $T_{k,n} = \|Z_{k,n}\|_w^2$.

Theorem 1 *Suppose that (6) holds, that H_0 is true, that the sample sizes satisfy (2), and that $k/m \rightarrow 0$, as $m \rightarrow \infty$. Then $nT_{k,n} \xrightarrow{D} \|Z\|_w^2$, as $m \rightarrow \infty$, where Z is a centred Gaussian random element of L^2_w having covariance kernel*

$$C(t, s) = \exp\{-0.5(t - s)^2\} - (1 + st + 0.5s^2t^2) \exp\{-0.5(t^2 + s^2)\}, \quad t, s \in \mathbb{R}. \tag{8}$$

Theorem 1 says that, under certain assumptions, the asymptotic null distribution of $T_{k,n}$ coincides with that of the one-sample test statistic, $T_{1,n}$ (see Henze and Wagner 1997). Therefore, at least for large n , we can approximate the percentiles of $T_{k,n}$ either by those of $T_{1,n}$ or by those of the asymptotic null distribution of $T_{1,n}$. In both cases, as commented in Sect. 2, the percentiles can be calculated using packages of the R language.

Next we study the behavior of the test under alternatives. To this end, we first state the following result that gives the a.s. behavior of $T_{k,n}$.

Theorem 2 *Suppose that (6) holds, that X_1, \dots, X_k have CFs $\varphi_1, \dots, \varphi_k$, respectively, that the sample sizes satisfy (2) and that $k/m \rightarrow 0$, as $m \rightarrow \infty$, then*

$$T_{k,n} - \|\varphi_{0,k} - \varphi_0\|_w^2 \xrightarrow{a.s.} 0,$$

as $m \rightarrow \infty$, where $\varphi_{0,k} = (1/n) \sum_{j=1}^k n_j \varphi_j$.

Notice that if X_1, \dots, X_k have all of them the same CF, say φ , then $\varphi_{0,k} = \varphi$ and, from Theorem 2, it follows that

$$T_{k,n} \xrightarrow{a.s.} \|\varphi - \varphi_0\|_w^2, \tag{9}$$

as $m \rightarrow \infty$. In particular, under H_0 we have that $T_{k,n} \xrightarrow{a.s.} 0$.

Next we see that the power of the test goes to 1 for alternatives so that $\|\varphi_{0,k} - \varphi_0\|_w^2 > 0$. With this aim, it will be assumed w.l.o.g. that

$$\begin{aligned} &X_1, \dots, X_r \text{ have alternative distributions with CFs } \varphi_1, \dots, \varphi_r, \\ &\text{while the other } k - r \text{ populations obey } H_0, \text{ for some } 1 \leq r \leq k. \end{aligned} \tag{10}$$

In the above setting, r is allowed to vary with m , $r = r_m$, but such dependence on m will be skipped. Let

$$f_r = \frac{1}{n} \sum_{i=1}^r n_i, \quad \varphi_{0,r} = \frac{1}{nf_r} \sum_{i=1}^r n_i \varphi_i.$$

We have that

$$\varphi_{0,k} - \varphi_0 = f_r(\varphi_{0,r} - \varphi_0). \tag{11}$$

Assume that the CFs of the alternative distributions satisfy

$$\inf_{\substack{\alpha_1 + \dots + \alpha_r = 1, \\ \alpha_1, \dots, \alpha_r \geq 0}} \left\| \sum_{i=1}^r \alpha_i \varphi_i - \varphi_0 \right\|_w \geq \tau > 0, \quad \forall r \geq 1, \tag{12}$$

and that the sample sizes satisfy (2), then from (11) it follows that

$$\frac{r}{k} M_1 \leq \|\varphi_{0,k} - \varphi_0\|_w \leq \frac{r}{k} M_2, \tag{13}$$

where M_1 and M_2 are two positive constants (depending on τ , c_0 and C_0).

As a consequence of Theorem 1, Theorem 2 and (13), we have the following result.

Corollary 1 *Suppose that (6), (10) and (12) hold, that the sample sizes satisfy (2), and that $k/m \rightarrow 0$ and $r/k \rightarrow p \in (0, 1]$, as $m \rightarrow \infty$. Then the power of the test that rejects H_0 when $T_{k,n} \geq t_{k,n,\alpha}$ goes to 1, as $m \rightarrow \infty$.*

The result in Corollary 1 remains true if $t_{k,n,\alpha}$ is replaced with a consistent estimator.

Remark 1 Assumption (12) may not be satisfied under alternatives. To see this fact, let us recall that if X has CF $\varphi_X = \Re\varphi_X + i\Im\varphi_X$ then $-X$ has CF $\varphi_{-X} = \Re\varphi_X - i\Im\varphi_X$. On the other hand, if X is a continuous random variable with probability density function (pdf)

$$2\phi(x)\pi(x), \tag{14}$$

where $\phi(x)$ is the pdf of a standard normal law and π is a skewing function (i.e. a function satisfying $0 \leq \pi(x) \leq 1$ and $\pi(-x) = 1 - \pi(x)$), then the real part of the CF of X is equal to φ_0 (see e.g. Jiménez-Gamero et al. 2016), and therefore $0.5\varphi_X + 0.5\varphi_{-X} = \varphi_0$. An example of a continuous law having a pdf of the form (14) is the skew normal law, for which $\pi(x) = \Phi(\lambda x)$, where Φ denotes the cumulative distribution function of a standard normal law and $\lambda \in \mathbb{R}$ is a constant. Thus, if $X_1 = X$, $X_2 = -X$ (say) and the pdf of X satisfies (14), then

$$\inf_{\substack{\alpha_1 + \dots + \alpha_r = 1, \\ \alpha_1, \dots, \alpha_r \geq 0}} \left\| \sum_{i=1}^r \alpha_i \varphi_i - \varphi_0 \right\|_w = 0, \quad \forall r \geq 2.$$

The same problem arises in the one-sample case if the assumption that the data are identically distributed is dropped. If, as in the one-sample case, we assume that X_1, \dots, X_r have the same alternative distribution, then (12) is satisfied.

Remark 2 Theorem 1 states that, under certain assumptions on k and the sample sizes, the asymptotic null distribution of $nT_{k,n}$ coincides with that of the BHEP test statistic. On the other hand, if X_1, \dots, X_k have all of them the same CF, say φ , then we saw that (9) holds. As a consequence of these two facts, in this setting, the Bahadur efficiencies computed in Ebner and Henze (2021) and Meintanis et al. (2022) for the BHEP test, also apply to the test proposed in this section.

Remark 3 As explained before, the main motivation for considering $T_{k,n}$ is that it can be seen as the natural extension of the BHEP test statistic. Nevertheless, other test statistics can be used for testing H_0 . For example, if we denote by \mathcal{T}_i to the BHEP test statistic calculated on the sample from X_i , following the approach in Gaigall (2021), other possible choices are $\sum_{i=1}^k \mathcal{T}_i$ or $(1/n) \sum_{i=1}^k n_i \mathcal{T}_i$. As for the studied proposal, the null distribution of these two test statistics only depends on n_1, \dots, n_k . We will come back to a test statistic of this type in Sect. 5.

The results in this section allow k to increase with the sample size, but at a lower rate. A key result to prove Theorem 1 is Lemma 2 in Sect. 8. If $k/m \rightarrow \ell > 0$ then, it can be checked that the results in Lemma 2 are no longer true. Moreover, since in such a case, even the practical calculation of $T_{k,n}$ can be very time-consuming, Sects. 4 and 5 explore other strategies to build a test of H_0 .

4 Random selection

Because, as observed before, for large k the calculation of $T_{k,n}$ can be very time-consuming, here we study a more efficient way (from a computational point of view) of testing H_0 , which consists in randomly selecting a subset of samples, and then applying the test studied in the previous section to the selected data. Specifically, the method proceeds as follows: for some (fixed) $k_0 < k$ (the precise order of k_0 will be specified later), select randomly (without replacement) I_1, \dots, I_{k_0} from $1, \dots, k$ and then apply the test in Sect. 3 to the samples $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_{k_0}}$. Let $n_0 = n_{I_1} + \dots + n_{I_{k_0}}$ be the total size of the selected data. Notice that when not all sample sizes are equal (unbalanced samples), n_0 is a random quantity. Let $T_{k_0,n_0} = T_{k_0,n_0}(\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_{k_0}})$ denote the test statistic calculated on the selected samples. Then, H_0 is rejected if $T_{k_0,n_0} \geq t_{k_0,n_0,\alpha}$, where $t_{k_0,n_0,\alpha}$ is the α upper percentile of the null distribution of T_{k_0,n_0} for n_0 fixed (non-random) and equal to its observed value.

To study properties of this procedure we first introduce some notation. For each $\mathbf{i} = (i_1, \dots, i_{k_0})$, with $1 \leq i_1 < \dots < i_{k_0} \leq k$, let

$$a(\mathbf{i}) = P(T_{k_0,n_0}(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_{k_0}}) \geq t_{k_0,n_0,\alpha} \mid I_1 = i_1, \dots, I_{k_0} = i_{k_0}),$$

and $a_0(\mathbf{i}) = P_0(T_{k_0,n_0}(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_{k_0}}) \geq t_{k_0,n_0,\alpha} \mid I_1 = i_1, \dots, I_{k_0} = i_{k_0}).$ (15)

By construction and the definition of $t_{k_0,n_0,\alpha}$, $a_0(\mathbf{i}) = \alpha, \forall \mathbf{i}$, and thus the test has level α because

$$P_0(T_{k_0,n_0} \geq t_{k_0,n_0,\alpha}) = \frac{1}{\binom{k}{k_0}} \sum_{1 \leq i_1 < \dots < i_{k_0} \leq k} a_0(\mathbf{i}) = \alpha.$$

Now we study the power.

Theorem 3 *Suppose that (6), (10) and (12) hold, that the sample sizes satisfy (2), and that $k_0 \rightarrow \infty$, $k_0/m \rightarrow 0$, $k_0/k \rightarrow \rho_0 \in [0, 1]$, and $r/k \rightarrow p \in (0, 1]$, as $m \rightarrow \infty$. Then the power of the test that rejects H_0 when $T_{k_0,n_0} \geq t_{k_0,n_0,\alpha}$ goes to 1, as $m \rightarrow \infty$.*

The result in Theorem 3 remains true if $t_{k_0,n_0,\alpha}$ is replaced with a consistent estimator.

The consistency result in Theorem 3 is very similar to that in Corollary 1, in the sense that, besides some assumptions on k_0 , both tests are consistent under the same assumptions, namely, (12) and $\frac{r}{k} \rightarrow p \in (0, 1]$. The comments in Remarks 1 and 2 also apply here.

Notice that two different random selections of I_1, \dots, I_{k_0} from $1, \dots, k$ could induce to opposite conclusions. To avoid this inconvenience, we follow Cuesta-Albertos and Febrero-Bande (2010) for tests based on random projections with functional data. These authors have proposed to take several random projections, calculate the p -value for each projection, and then apply some correction, as for example the procedure in Benjamini and Yekutieli (2001), which controls the false discovery rate. The same approach can be applied here.

5 Sum of BHEP test statistics

As observed in Remark 3, we could consider test statistics based on the sum of test statistics calculated on each sample. Here we study a test whose test statistic is of that type. Recall from Sect. 2.2 the definition of $\mu_{0,n}$ and $\tau_{0,n}^2$, and that $0 < \mu_{0,n} \rightarrow \mu_0$ and $0 < \tau_{0,n}^2 \rightarrow \sigma_0^2$, as $n \rightarrow \infty$.

Let $\mathfrak{T}_i = n_i \mathcal{T}_i$, where $\mathcal{T}_i = \mathcal{T}_{n_i,1}$ (see (3)), $1 \leq i \leq k$, and let

$$\mathbb{T}_{0,k} = \frac{\sum_{i=1}^k (\mathfrak{T}_i - \mu_{0,n_i})}{\sqrt{\sum_{i=1}^k \tau_{0,n_i}^2}}.$$

As seen in Sect. 2.2, if $F_i \notin \mathcal{N}$ then $E(\mathfrak{T}_i) = n_i \Delta_{X_i, n_i}$, with $\Delta_{X_i, n_i} \rightarrow \Delta_{X_i} = \|\varphi_i - \varphi_0\|_w^2 > 0$, as $n_i \rightarrow \infty$, where φ_i is the CF of X_i . Hence, for large enough n_i , $E(\mathfrak{T}_i)$ is bigger than $E_0(\mathfrak{T}_i) = \mu_{0,n_i}$. Thus, it seems reasonable to reject H_0 for large values of $\mathbb{T}_{0,k}$.

To test H_0 we must calculate upper percentiles of the null distribution of $\mathbb{T}_{0,k}$, which depends on n_1, \dots, n_k . Although the critical points can be calculated by simulation, from a practical point of view it would be nice if they could be approximated in some fashion, at least for large k , since n_1, \dots, n_k can take many values. The next result shows that, under H_0 , $\mathbb{T}_{0,k}$ converges in law to a standard normal law, as $k \rightarrow \infty$, no matter how large (or small) are the sample sizes n_1, \dots, n_k , it only assumes that $n_i \geq 3$. If $n_i = 2$ then the two scaled residuals take the values -1 and 1, for any possible values of $X_{i,1}, X_{i,2}$ (whenever they are different, which happens with probability 1 as X_i is assumed to be continuous), and thus \mathfrak{T}_i is a degenerate random variable.

Theorem 4 Suppose that (6) holds, that $n_i \geq 3$, and that H_0 is true. Then $\mathbb{T}_{0,k} \xrightarrow{\mathcal{D}} Z \sim N(0, 1)$, as $k \rightarrow \infty$.

From Theorem 4, the test that rejects H_0 when $\mathbb{T}_{0,k} \geq z_{1-\alpha}$, for some $\alpha \in (0, 1)$, where $\Phi(z_{1-\alpha}) = 1 - \alpha$, has (asymptotic) level α , where here asymptotic means for large k .

Now we study the power of this test. As in the previous sections, we will suppose that (10) holds. The r alternative distributions will be assumed to satisfy the following assumption.

- Assumption 1** (a) $\{(\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i, n_i}))^2\}$ is uniformly integrable, $1 \leq i \leq r$.
 (b) There exist $0 < \varsigma_1 < \varsigma_2 < \infty$ such that $\varsigma_1 \leq \inf_{1 \leq i \leq r} \tau_{X_i, n_i}^2 \leq \sup_{1 \leq i \leq r} \tau_{X_i, n_i}^2 \leq \varsigma_2$,
 $\forall r$, where $\tau_{X_i, n_i}^2 = V(\sqrt{n_i} \mathcal{T}_i)$, $1 \leq i \leq r$.
 (c) There exists $0 < \eta$ such that $\frac{1}{r} \sum_{i=1}^r (\Delta_{X_i, n_i} - \mu_{0, n_i} / n_i) \geq \eta, \forall r$.

Recall from Sect. 2.2 that, for alternative distributions with $E(X_i^4) < \infty$, $\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i})$ converges in law to a zero-mean normal distribution. Thus, it makes sense to consider Assumption 1 (a) and (b). Moreover, (5) (for each i) was seen to be a sufficient condition for Assumption 1 (a) to hold. Since Δ_{X_i, n_i} is positive and close to Δ_{X_i} , Assumption 1 (c) is saying that n_i must large enough so that $E(\mathbb{T}_{0,k}) > 0$. Table 1 below displays the values of $\mu_{0, n}$ and $n\Delta_{X, n}$, for some alternative distributions and some small values of n , that were calculated by simulation based on 100,000 samples in each case. The considered alternative distributions are:

- The beta distribution with parameters (2,2) ($b(2, 2)$), whose pdf is $f(x) = 6x(1 - x)$ if $x \in [0, 1]$,
- the Laplace distribution (Lap) with pdf $f(x) = 0.5 \exp(-|x|)$, $x \in \mathbb{R}$,
- the uniform distribution (unif) with pdf $f(x) = 1$ if $x \in [0, 1]$,
- the logistic distribution (log) with pdf $f(x) = \exp(-x) / \{1 + \exp(-x)\}^2$, $x \in \mathbb{R}$,
- the Student t-distribution with ν degrees of freedom (t_ν), whose pdf is given by $f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} (1 + x^2/\nu)^{-(\nu+1)/2}$, $x \in \mathbb{R}$, where Γ is the gamma function,
- a scale mixture of two normal populations (SMN): $pN(0, \sigma^2) + (1 - p)N(0, 1)$, with $p = 0.2$ and $\sigma = 3$,
- the negative exponential distribution (exp) with pdf $f(x) = \exp(-x)$, $x \in [0, \infty)$, and
- the chi-squared distribution with ν degrees of freedom (χ_ν^2), whose pdf is given by $f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} \exp(-x/2)$, $x \in [0, \infty)$.

Looking at Table 1 we see that Assumption 1(c) is not restrictive at all: it suffices to take $n_i \geq 5$ for all considered alternatives.

To derive the asymptotic null distribution of $\mathbb{T}_{0,k}$ no assumption was made on the sample sizes n_1, \dots, n_k (except that all of them are greater than or equal to 3). To study the power, we will assume that the sample sizes are comparable, in the sense that they satisfy (2). Nevertheless, in contrast to the results for the power in the previous sections (see Corollary 1 and Theorem 3), here m is not assumed to increase, it only must be large enough so that Assumption 1(c) holds true.

Table 1 Values of $\mu_{0,n}$ and $n\Delta_{X,n}$, for $3 \leq n \leq 10$ and some alternative distributions for X

n	$\mu_{0,n}$	$n\Delta_{X,n}$											
		$b(2, 2)$	Lap	unif	log	t_7	t_{20}	t_{30}	SMN	exp	χ^2_4	χ^2_{20}	χ^2_{40}
3	0.33	0.34	0.34	0.34	0.33	0.33	0.33	0.33	0.34	0.36	0.35	0.34	0.33
4	0.49	0.49	0.52	0.51	0.49	0.50	0.49	0.49	0.53	0.61	0.54	0.50	0.49
5	0.13	0.63	0.73	0.68	0.66	0.66	0.64	0.64	0.74	0.93	0.78	0.66	0.65
6	0.13	0.77	0.94	0.85	0.82	0.83	0.79	0.78	0.99	1.33	1.06	0.83	0.80
7	0.13	0.91	1.18	1.02	0.99	1.00	0.94	0.93	1.26	1.79	1.37	1.01	0.96
8	0.13	1.04	1.44	1.21	1.16	1.18	1.09	1.07	1.56	2.33	1.73	1.18	1.12
9	0.13	1.18	1.70	1.40	1.34	1.37	1.23	1.21	1.89	2.93	2.11	1.38	1.28
10	0.13	1.33	1.97	1.61	1.52	1.55	1.37	1.36	2.25	3.60	2.54	1.58	1.45

Theorem 5 Suppose that (6), (10) and Assumption 1 hold, that the sample sizes satisfy (2), and that $r/k \rightarrow p \in (0, 1]$, as $k \rightarrow \infty$. Then the power of the test that rejects H_0 when $\mathbb{T}_{0,k} \geq z_{1-\alpha}$ goes to 1, as $k \rightarrow \infty$.

6 Simulation results

This section presents the results of several simulation experiments designed to study the finite sample performance of the three tests studied in this paper. We first study the goodness of the approximations given to the null distribution of the proposed test statistics, and then their powers, which are also compared with the following procedures for testing H_0 :

- Compute the BHEP test for testing $H_{0i} : X_i \in \mathcal{N}, 1 \leq i \leq k$. Then one can apply either the Bonferroni method, which controls the family-wise error rate, or the Benjamini-Hochberg method (see Benjamini and Yekutieli 2001), which controls the false discovery rate when the k tests are independent. Both procedures agree in rejecting H_0 if $\min_{1 \leq i \leq k} p_i \leq \alpha/k$, where p_1, \dots, p_k are the p -values obtained when testing H_{01}, \dots, H_{0k} , respectively. The results for this procedure are headed in the tables by BH (and we will also refer to that test as the test BH).
- The tests in Gaigall (2021).

6.1 Simulations for the level

We first consider the test in Sect. 3, which rejects H_0 when $T_{k,n} > t_{k,n,\alpha}$. By construction, if one uses that critical region, then the test will have exactly level α . So, to check the actual level of this test has no interest. Recall that Theorem 1 states that, under certain assumptions, the asymptotic null distribution of $T_{k,n}$ coincides with that of the one-sample test statistic, $T_{1,n}$. Therefore, we could approximate the percentiles of $T_{k,n}$ by those of $T_{1,n}$. Here we study that approximation by simulation, and hence we consider the test that rejects H_0 when $T_{k,n} > t_{1,\alpha}$ where $t_{1,\alpha}$ is the α upper percentile

of the null distribution of $T_{1,n}$. The results for such test are headed in the tables by $T_{k,n}^{as}$ (and we will also refer to that test as the test $T_{k,n}^{as}$).

The critical point $t_{k,n,\alpha}$ can be also approximated by the α upper percentile of the asymptotic null distribution of $T_{k,n}$, say t_α . As explained by the end of Sect. 2.1, t_α can be estimated by using the Pearson system of distributions, as proposed in Henze (1990). The results for such test are headed in the tables by $T_{k,n}^{Pe}$ (and we will also refer to that test as the test $T_{k,n}^{Pe}$).

We also consider the test that rejects H_0 when $\mathbb{T}_{0,k} \geq z_{1-\alpha}$. The results for such test are headed in the tables by $\mathbb{T}_{0,k}$ (and we will also refer to that test as the test $\mathbb{T}_{0,k}$).

The test statistics in Gaigall (2021) are sum of the Kolmogorov-Smirnov test statistics and the Cramér-von-Mises test statistics in each sample, and so, it is expected that, conveniently normalized (subtracting the mean and dividing by the square root of their variances, as we did in Sect. 5 to obtain $\mathbb{T}_{0,k}$) those statistics are asymptotically normal as $k \rightarrow \infty$. Notice that the null distribution of the Kolmogorov-Smirnov and the Cramér-von-Mises test statistics in each sample does not depend on the values of the population mean and variance, but only on the sample size. We calculated by simulation, the means and the variances of these statistics in a sample, and considered the test that rejects H_0 when $KS \geq z_{1-\alpha}$ and the test that rejects H_0 when $CM \geq z_{1-\alpha}$, where KS and CM are the Kolmogorov-Smirnov and the Cramér-von-Mises analogues of $\mathbb{T}_{0,k}$, respectively. The results for such tests are headed in the tables by KS and CM (and we will also refer to those tests as the test KS and the test CM), respectively.

In each case we did the following experiment: k random samples with size $n_i = m$, $1 \leq i \leq k$, were generated from a standard normal law and the tests KS , CM , $T_{k,n}^{as}$, $T_{k,n}^{Pe}$, $\mathbb{T}_{0,k}$ and BH were applied with $\alpha = 0.05$. The experiment was repeated 10,000 times (all simulations for the level in this paper are based on 10,000 samples). Table 2 displays the proportion of times that H_0 was rejected for $k = 2, 3, 5, 10, 20$ and $n_i = 5, 10, 15, \dots, 45$. Looking at Table 2 we conclude that: (a) for the test $T_{k,n}^{as}$: as expected, the approximation works when k is small in relation to n_i , the distortion of the level may be important if such relation is not met; (b) the same can be concluded for the test $T_{k,n}^{Pe}$, whose performance is quite close to $T_{k,n}^{as}$; (c) for the test $\mathbb{T}_{0,k}$: the approximation works better for larger values of k (as expected), nevertheless, the empirical levels are not far apart from the nominal level even for $k = 2$; its behavior does not seem to be influenced by the sample sizes; (d) the same can be concluded for the tests KS and CM ; (e) for the test BH: the levels are reasonably close to the nominal level.

The above experiment was repeated for larger values of k (specifically, for $k = 100, 200$) and $n_i = 5, 10, 15, 20$, but instead of the test $T_{k,n}^{as}$ (and $T_{k,n}^{Pe}$) we considered the random selection test studied in Sect. 4 with $k_0 = 10, 20$, headed in the tables by RP (and we will also refer to that test as the test RP). We tried that test with one and more than one random selections. In view of the results of Table 2 for the test $T_{k,n}^{as}$ (and $T_{k,n}^{Pe}$), and since the sample sizes considered are not very large compared to k_0 , we used the exact critical values of the null distribution of T_{k_0,n_0} ; when several random selections are taken into account (5, 10, 20 and 30), we proceed as explained

before for the test BH. The results obtained are displayed in Table 3. In all cases, the empirical levels match quite closely the nominal value.

6.2 Simulations for the power

As said at the beginning of Sect. 3, in our developments we chose $\beta = 1$, although all results keep on being true for arbitrary (but fixed) β . It is well-known that, for finite sample sizes, the power of the BHEP test strongly depends on the value of the parameter β and on the alternative. Tenreiro (2009) observed that for short-tailed alternatives, high power is obtained for large (but not too large) values of β , and for long-tailed (symmetric or asymmetric) alternatives a small value for β should be chosen. Since the tests studied in this paper are all of them based on the BHEP test, it is expected that they inherit their characteristics. So, in the power study, we tried the proposed tests for several values of β . As in Ebner and Henze (2021), we considered $\beta \in \{0.25, 0.5, 0.75, 1, 2, 3, 5, 10\}$.

To examine the power we repeated the experiments in the previous section, but now r samples were taken from an alternative distribution and the other $k - r$ samples were generated from a standard normal distribution. r is taken so that the percentage of alternative distributions equals 20%, 40%, 60% and 80%. All simulations for the power in this paper are based on 2,000 samples. For the test that rejects H_0 when $T_{k,n} > t_{k,n,\alpha}$ (the dependence on β is skipped to simplify notation), we calculated the exact critical values (headed in the tables by $T_{k,n}^{ex}$, we will also refer to that test as the test $T_{k,n}^{ex}$). The columns headed as RP_1 and RP_2 display the results for the random selection method with $k_0 = 10$ and $k_0 = 20$, respectively, both with 30 random selections because, in most cases, that number of selections gave the higher power (taken among 1, 3, 5, 10, 20 and 30 random selections).

As for the alternatives, we considered several short-tailed and some long-tailed distributions. The picture in each case is very similar and agree with Tenreiro (2009) observations. Due to space limitations, next we just summarize the experiment results. All tables and a detailed description of the numerical results can be found in the Supplementary Material.

Looking at the tables for the power in the Supplementary Material, in general, it can be concluded that: (a) the power of all tests increases with the percentage of alternative distributions, with the sample sizes and with k ; (b) the test BH gives the poorest results; (c) there is no test giving the highest power for all alternatives; (d) among the BHEP-based tests, we again see that there is no one giving the highest power for all alternatives; nevertheless, $\mathbb{T}_{0,k}$ gives powers (for adequate choices of β) that are either optimal or reasonably close to the optimal; (e) KS is less powerful than CM ; their powers are smaller than those of $T_{k,n}^{ex}$ and $\mathbb{T}_{0,k}$ (for adequate choices of β), for small k , and of RP_2 and $\mathbb{T}_{0,k}$ (for adequate choices of β), for larger k .

From a computational point of view, among BHEP-based tests, the test $\mathbb{T}_{0,k}$ is the best choice, since the number of required computations for the calculation of its test statistic is of order $O(km^2)$, and its application does not involve the calculation of critical points.

Table 2 Empirical levels for small to moderate values of k at the nominal level $\alpha = 5.0\%$

n_i	$k = 2$						$k = 3$						$k = 5$					
	KS	CM	$T_{k,n}^{as}$	$T_{k,n}^{Pe}$	$\mathbb{T}_{0,k}$	BH	KS	CM	$T_{k,n}^{as}$	$T_{k,n}^{Pe}$	$\mathbb{T}_{0,k}$	BH	KS	CM	$T_{k,n}^{as}$	$T_{k,n}^{Pe}$	$\mathbb{T}_{0,k}$	BH
5	5.9	6.7	5.1	3.8	7.3	4.8	5.9	6.6	6.4	5.6	6.7	4.8	5.7	6.3	9.8	9.4	6.3	5.2
10	6.1	6.6	5.1	4.9	6.7	4.6	5.9	6.7	5.8	5.6	6.5	4.7	5.7	6.5	6.7	6.6	6.3	4.3
15	6.3	7.2	5.2	5.0	7.0	5.2	5.8	6.5	5.2	5.0	6.6	4.9	5.6	6.5	6.3	6.0	6.5	4.6
20	5.8	6.3	5.0	4.7	6.5	4.7	5.7	6.2	5.1	4.9	6.5	4.4	5.6	6.3	5.9	5.7	6.4	4.5
25	6.2	6.4	4.8	4.6	6.7	4.6	6.1	6.7	5.2	5.0	6.7	4.3	5.8	6.6	6.2	6.0	6.8	4.3
30	6.1	6.5	4.7	4.5	6.7	4.6	5.9	6.4	5.2	5.1	6.3	4.5	5.7	5.8	5.8	5.8	6.3	4.5
35	6.2	6.7	4.8	4.6	6.9	5.0	5.6	6.6	5.2	5.2	6.6	4.9	6.0	6.5	6.0	5.8	6.7	4.8
40	6.5	6.8	5.1	5.0	7.3	5.1	6.0	6.8	5.4	5.4	7.2	5.4	5.9	6.4	5.8	5.8	6.7	5.2
45	6.2	6.5	5.2	5.0	6.9	4.9	5.9	6.6	5.3	5.3	6.8	4.8	6.1	6.5	5.5	5.5	6.8	4.8

n_i	$k = 10$						$k = 20$					
	KS	CM	$T_{k,n}^{as}$	$T_{k,n}^{Pe}$	$\mathbb{T}_{0,k}$	BH	KS	CM	$T_{k,n}^{as}$	$T_{k,n}^{Pe}$	$\mathbb{T}_{0,k}$	BH
5	5.4	5.9	24.2	23.6	6.0	5.1	5.2	5.9	64.5	64.0	5.8	4.6
10	5.7	6.3	10.8	10.7	6.1	4.3	5.3	6.0	21.3	21.2	6.4	4.1
15	5.1	5.5	8.9	8.8	6.0	4.9	5.0	5.5	13.7	13.5	5.9	5.1
20	5.4	6.0	7.6	7.5	6.5	4.7	5.3	5.3	11.1	10.9	5.9	4.4
25	5.7	6.2	7.0	6.9	6.3	4.5	5.5	6.0	9.6	9.5	5.9	4.3
30	5.7	5.9	6.6	6.5	6.0	4.8	5.2	5.7	8.9	8.8	6.1	5.4
35	5.7	6.1	6.4	6.3	6.0	4.9	5.5	5.6	8.5	8.4	5.7	5.3
40	5.6	5.8	6.2	6.1	6.3	5.0	5.4	5.5	7.8	7.7	5.7	4.5
45	5.8	6.3	6.1	6.0	6.2	5.0	5.4	5.7	7.6	7.5	5.8	5.0

Table 3 Empirical levels for large values of k at the nominal level $\alpha = 5.0\%$

k	n_i	KS	CM	RP, $k_0 = 10$										$T_{0,k}$	BH						
				Number of random selections																	
				1	3	5	10	20	30	1	3	5	10			20	30				
100	5	4.9	5.4	5.2	5.0	4.8	4.8	5.0	4.8	4.8	4.8	4.9	4.9	4.9	5.0	5.2	4.8	5.5	5.3		
	10	5.0	5.4	5.0	4.9	5.4	5.0	4.6	4.7	4.7	4.7	5.1	4.7	4.4	5.0	5.1	5.2	5.9	4.1	4.1	
	15	4.7	4.8	4.9	4.8	5.0	4.9	4.9	4.9	4.8	4.8	4.8	4.8	4.7	4.3	4.3	4.3	4.1	5.3	4.5	4.5
200	20	4.7	4.8	5.4	4.9	4.8	4.5	5.2	5.3	5.3	5.3	4.9	4.8	4.4	5.0	5.6	6.2	5.4	5.5	5.5	
	5	5.3	5.7	4.9	5.0	4.8	4.6	5.2	5.1	5.1	5.1	4.9	5.1	5.0	5.1	5.3	4.8	5.2	5.2	5.2	5.2
	10	4.8	4.7	5.0	5.1	5.1	4.8	5.0	4.7	4.7	4.7	5.3	5.0	5.3	5.2	5.3	5.6	5.1	5.1	3.2	3.2
	15	4.5	4.3	5.2	5.2	5.1	5.0	4.9	5.0	5.0	5.0	4.7	4.7	4.9	4.4	4.2	4.1	5.3	4.7	4.7	4.7
	20	4.8	4.7	5.6	4.9	4.7	5.0	5.2	5.1	5.1	5.1	4.7	4.7	4.8	5.0	5.6	5.9	4.8	4.8	6.6	6.6

6.3 Further simulation results

In the above experiments all samples have the same size $n_1 = n_2 = \dots = n_k := m$, so the data can be also seen as

$$Y_1 = \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{k,1} \end{pmatrix}, Y_2 = \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{k,2} \end{pmatrix}, \dots, Y_m = \begin{pmatrix} X_{1,m} \\ X_{2,m} \\ \vdots \\ X_{k,m} \end{pmatrix} \text{ which are iid from } Y = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} \in \mathbb{R}^k.$$

An anonymous referee asked us to apply the BHEP test to the Y -data. With this aim we need $m \geq k + 1$, so we only considered the cases $k = 10$ with $n_i = 15, 20, 25, 30$ and $k = 20$ with $n_i = 25, 30$. Tenreiro (2009) performed an extensive simulation study on the power of the BHEP test for a wide range of data dimension. He recommends using $\beta_k = \sqrt{2}/(1.376 + 0.075k)$. We repeated the power simulation study for the Y -data using β_k and the critical points of the BHEP test for each value of k (that now becomes the dimension) and m (the sample size). The obtained results are displayed in the Supplementary Material. Comparing them with those yielded by the tests proposed in this paper, CM and KS, applied to the X -data, we see that when the BHEP is applied to the Y -data the power is really poor.

The above Y -data description assumes that the components of Y are independent. To numerically study the effect on the level of the tests considered in Sect. 6.1 when the independence assumption is dropped, the following simulation experiment was carried out: we generated data from $Y \sim N_k(0, \Sigma_\rho)$, where Σ_ρ is the equicorrelation matrix, and then applied the tests in Table 2 to the associated X -data. The obtained results are displayed in the Supplementary Material. Looking at them we see that as the dependence between the components of Y becomes stronger, the empirical levels move further away from the nominal value 0.05. As a consequence, the case of Y -data with correlated components requires the development of new procedures that take into account such dependence.

7 Concluding remarks and further research

This paper proposes and studies three procedures for simultaneously testing that k independent samples come from normal populations, which can have different means and variances. All of them are based on the BHEP test and allow k to increase. The first test, based on the test statistic $T_{k,n}$, can be seen as the direct extension of the BHEP test. Its null distribution only depends on k and the sample sizes, so that exact critical points can be calculated by simulation, but when the sample sizes are large relative to k , one can use the critical points of the BHEP test. If k is very large, the practical calculation of the test statistic $T_{k,n}$ is very time-consuming, so one can randomly select k_0 samples (one or more times) and apply the previous test to the selected samples. One can also calculate the BHEP test statistic in each sample and then sum the obtained values, which conveniently centred and scaled $(\mathbb{T}_{0,k})$ converges in law

to a standard normal law when the null hypothesis is true. The normal approximation works reasonably well even for not too large k , so its practical use does not require the calculation of critical points. All test are consistent against alternatives where the fraction of samples not obeying the null goes to a positive constant. The test based on $\mathbb{T}_{0,k}$ is, from a computational point of view, the best choice.

This paper is centered in studying BHEP-based procedures for simultaneously testing that k independent samples come from normal populations. Other normality tests could be used to build similar procedures to those developed in Sects. 3–5. Moreover, parallel approaches could be used for simultaneously testing that k independent samples come from any location-scale family.

As observed in Remark 2, in certain specific settings, the Bahadur efficiencies calculations made in Ebner and Henze (2021) and Meintanis et al. (2022) for the BHEP test, also apply to the tests proposed in Sects. 3 and 4. It would be interesting to study Bahadur efficiencies in more general settings, and also for the test in Sect. 5. Those calculations could help to determine optimal values of β and k_0 .

Finally, in simulations we saw that the tests studied in this paper are not valid for dependent data. New procedures that take into account such dependence should be developed.

8 Proofs

Along this section M is a generic positive constant taking many different values throughout the proofs.

8.1 Auxiliary results

Lemma 1 *Suppose that (6) holds and that H_0 is true. Then*

- (a) $E\{n_j(1 - S_j)^2\} = 0.5 + O(1/n_j)$,
- (b) $E\{(1 - 1/S_j)^2\} = 1/n_j + O(1/n_j^2)$,
- (c) $E\{(1 - 1/S_j)^4\} = O(1/n_j^2)$,

as $n_j \rightarrow \infty, 1 \leq j \leq k$.

Proof (a) Under $H_0, n_j S_j^2 \sim \chi_{n_j-1}^2$, thus $E(n_j S_j^2) = n_j - 1$ and $E(n_j S_j) = \sqrt{n_j} \sqrt{2} \Gamma(n_j/2) / \Gamma((n_j - 1)/2)$, where Γ stands for the gamma function. Using Legendre duplication formula (see display 5.5.5 of Olver et al. 2010) and Stirling’s formula (see display 5.11.3 of Olver et al. 2010), one gets that, for large n_j ,

$$E(n_j S_j) = \sqrt{n_j(n_j - 1)} \left(1 + \frac{1}{4n_j} + O\left(\frac{1}{n_j^2}\right) \right).$$

Thus

$$E\{n_j(1 - S_j)^2\} = 2n_j - 1 - 2\sqrt{n_j(n_j - 1)} \left(1 + \frac{1}{4n_j} + O\left(\frac{1}{n_j^2}\right) \right).$$

Finally, taking into account that $2n_j - 1 - 2\sqrt{n_j(n_j - 1)} = 2(1 + \sqrt{1 - 1/n_j})^{-1} - 1$, $\sqrt{n_j(n_j - 1)}/n_j = \sqrt{1 - 1/n_j}$ and $\sqrt{1 - 1/n_j} = 1 - 1/(2n_j) + O(1/n_j^2)$, the result follows.

(b, c) The proof is similar to that of part (a). □

Remark 4 The result in Lemma 1 (a) can be also derived from pages 421–422 of Johnson et al. (1994).

$$\text{Let } \Delta_{j,r} = Y_{j,r} - X_{j,r} = X_{j,r}(1/S_j - 1) - \bar{X}_j/S_j, 1 \leq r \leq n_j, 1 \leq j \leq k.$$

Lemma 2 Suppose that (6) holds, that H_0 is true, that the sample sizes satisfy (2), and that $k/m \rightarrow 0$, as $m \rightarrow \infty$. Then,

- (a) $\frac{1}{\sqrt{n}} \sum_{j,r} \Delta_{j,r}^2 = o_P(1)$,
- (b) $\|W_n\|_w = o_P(1)$, for
 - (b.1) $W_n(t) = t \frac{1}{\sqrt{n}} \sum_{j=1}^k (\bar{X}_j/S_j) \sum_{r=1}^{n_j} \sin(tX_{j,r}), t \in \mathbb{R}$,
 - (b.2) $W_n(t) = t \frac{1}{\sqrt{n}} \sum_{j=1}^k (\bar{X}_j/S_j) \sum_{r=1}^{n_j} \{\cos(tX_{j,r}) - \varphi_0(t)\}, t \in \mathbb{R}$,
 - (b.3) $W_n(t) = t \frac{1}{\sqrt{n}} \sum_{j=1}^k (1 - 1/S_j) \sum_{r=1}^{n_j} X_{j,r} \cos(tX_{j,r}), t \in \mathbb{R}$,
 - (b.4) $W_n(t) = t \frac{1}{\sqrt{n}} \sum_{j=1}^k (1 - 1/S_j) \sum_{r=1}^{n_j} \{X_{j,r} \sin(tX_{j,r}) + \varphi'_0(t)\}, t \in \mathbb{R}$,
 - (b.5) $W_n(t) = t\varphi_0(t) \frac{1}{\sqrt{n}} \sum_{j=1}^k n_j \bar{X}_j (1 - 1/S_j), t \in \mathbb{R}$,
 - (b.6) $W_n(t) = t\varphi'_0(t) \frac{1}{\sqrt{n}} \sum_{j=1}^k \sum_{r=1}^{n_j} \{(1/S_j - 1) + 0.5(X_{j,r}^2 - 1)\}, t \in \mathbb{R}$,

as $m \rightarrow \infty$, where $\varphi'_0(t) = \frac{d}{dt} \varphi_0(t)$.

Proof (a) We first observe that

$$\sum_{r=1}^{n_j} \Delta_{j,r}^2 = n_j \{(1 - S_j)^2 + \bar{X}_j^2\}, \quad 1 \leq j \leq k.$$

From (2),

$$n = \sum_{j=1}^k n_j = m \sum_{j=1}^k c_j = \bar{c}km, \quad \text{with } 0 < c_0 \leq \bar{c} = (1/k) \sum_{j=1}^k c_j \leq C_0 < \infty \tag{16}$$

Taking into account (16), the result in Lemma 1(a), and that $E(n_j \bar{X}_j^2) = 1, 1 \leq j \leq k$, one gets

$$0 \leq E \left(\frac{1}{\sqrt{n}} \sum_{j,r} \Delta_{j,r}^2 \right) \leq M \sqrt{\frac{k}{m}}.$$

Since we are assuming that $k/m \rightarrow 0$, the result follows.

(b.1) We have that $E\{W_n(t)^2\} = t^2(1/n) \sum_{j,v=1}^k E_{jv}$, where

$$E_{jv} = E \left(\frac{\bar{X}_j}{S_j} \frac{\bar{X}_v}{S_v} \sum_{r=1}^{n_j} \sin(tX_{j,r}) \sum_{s=1}^{n_v} \sin(tX_{v,s}) \right).$$

If $j \neq v$, then

$$\begin{aligned} E_{jv} &= E \left(\frac{\bar{X}_j}{S_j} \sum_{r=1}^{n_j} \sin(tX_{j,r}) \right) E \left(\frac{\bar{X}_v}{S_v} \sum_{s=1}^{n_v} \sin(tX_{v,s}) \right) \\ &\leq E^{1/2} \left(\frac{\bar{X}_j^2}{S_j^2} \right) E^{1/2} \left(\frac{\bar{X}_v^2}{S_v^2} \right) E^{1/2} \left(\left\{ \sum_{r=1}^{n_j} \sin(tX_{j,r}) \right\}^2 \right) E^{1/2} \left(\left\{ \sum_{s=1}^{n_v} \sin(tX_{v,s}) \right\}^2 \right). \end{aligned}$$

Since $\sqrt{n_j - 1} \bar{X}_j/S_j \sim t_{n_j-1}$, one gets that $E \left(\bar{X}_j^2/S_j^2 \right) = 1/(n_j - 3)$. We also have that $E \left(\left\{ \sum_{r=1}^{n_j} \sin(tX_{j,r}) \right\}^2 \right) = n_j E\{\sin^2(tX_j)\} = 0.5n_j\{1 - \varphi_0(2t)\} \leq Mn_j$. Thus, $E_{jv} \leq M, 1 \leq j \neq v \leq k$. If $j = v$, then

$$E_{jj} \leq E^{1/2} \left(\frac{\bar{X}_j^4}{S_j^4} \right) E^{1/2} \left(\left\{ \sum_{r=1}^{n_j} \sin(tX_{j,r}) \right\}^4 \right).$$

Since

$$E \left(\left\{ \sum_{r=1}^{n_j} \sin(tX_{j,r}) \right\}^4 \right) = \sum_{r,s=1}^{n_j} E \left\{ \sin^2(tX_{j,r}) \sin^2(tX_{j,s}) \right\} \leq n_j^2,$$

and, for $n_j \geq 6$,

$$E \left(\frac{\bar{X}_j^4}{S_j^4} \right) = \frac{1}{(n_j - 1)^2} \left(\frac{6}{n_j - 5} \frac{(n_j - 1)^2}{(n_j - 3)^2} + 3 \right) \leq M \frac{1}{n_j^2},$$

we get that $E_{jj} \leq M, 1 \leq j \leq k$, whenever $n_j \geq 6$. Therefore, $E\{W_n(t)^2\} \leq t^2Mk^2/n$, whenever $n_j \geq 6, 1 \leq j \leq k$. From (16), it follows that $k^2/n \leq Mk/m$. Since $\int t^2w(t)dt = 1$ and $k/m \rightarrow 0$, as $m \rightarrow \infty$, it follows that $\|W_n\|_w = o_P(1)$, as $m \rightarrow \infty$.

(b.2) The proof is similar to that of (b.1).

(b.3, b.4) The proof is similar to that of (b.1) using Lemma 1 (b) and (c).

(b.5) We have that $E\{W_n(t)^2\} = t^2\varphi_0(t)^2(1/n) \sum_{j,v=1}^k E_{jv}$, where

$$E_{jv} = n_j n_v E \left\{ \bar{X}_j(1 - 1/S_j) \bar{X}_v(1 - 1/S_v) \right\}.$$

If $j \neq v$, then

$$E_{jv} = n_j n_v E(\bar{X}_j) E\{(1 - 1/S_j)\} E(\bar{X}_v) E\{(1 - 1/S_v)\} = 0.$$

If $j = v$, then

$$E_{jj} = n_j^2 E(\bar{X}_j^2) E\{(1 - 1/S_j)^2\}.$$

Taking into account that $E(n_j \bar{X}_j^2) = 1$ and Lemma 1(b), we get that, for large m , $E_{jj} \leq M$.

Therefore, for large m , $E\{W_n(t)^2\} \leq Mt^2 \varphi_0(t)^2 k/n$. Since $\int t^2 w(t) \varphi_0(t)^2 dt < \infty$, from (16), it follows that $\|W_n\|_w = o_P(1)$, as $m \rightarrow \infty$.

(b.6) We can write $W_n(t) = t \varphi'_0(t) \{A_n + 0.5B_n\}$, where $A_n = (1/\sqrt{n}) \sum_{j=1}^k n_j (1/S_j - 1)$ and $B_n = (1/\sqrt{n}) \sum_{j=1}^k \sum_{r=1}^{n_j} (X_{j,r}^2 - 1)$. Since $\int t^2 w(t) \varphi'_0(t)^2 dt < \infty$, to show the result it suffices to see that $A_n = o_P(1)$ and $B_n = o_P(1)$, as $m \rightarrow \infty$.

We have that $E\{A_n^2\} = (1/n) \sum_{j,v=1}^k E_{jv}$, where $E_{jv} = n_j n_v E\{(1/S_j - 1)^2 (1/S_v - 1)^2\}$. From Lemma 1 (b) and (c), it follows that for large m , $E_{jv} \leq M$, $1 \leq j, v \leq k$, and thus $E\{A_n^2\} \leq Mk^2/n$. Now using (16) one gets that $A_n = o_P(1)$, as $m \rightarrow \infty$.

Since

$$E\{(X_{j,r}^2 - 1)(X_{v,s}^2 - 1)\} = \begin{cases} 2 & \text{if } j = v \text{ and } r = s, \\ 0 & \text{otherwise,} \end{cases}$$

and using (16), we have that $E\{B_n^2\} = 2k/n = 2/(\bar{c}m) \rightarrow 0$, as $m \rightarrow \infty$. Therefore, $B_n = o_P(1)$, as $m \rightarrow \infty$. □

Lemma 3 *Suppose that (6) holds, that the sample sizes satisfy (2), and that $k/m \rightarrow 0$, as $m \rightarrow \infty$. Then $(1/n) \sum_{j,r} |\Delta_{j,r}| \xrightarrow{a.s.} 0$, as $m \rightarrow \infty$.*

Proof To stress the dependence on $\omega \in \Omega$ we write $\Delta_{j,r}(\omega)$. By the strong law of large numbers (SLLN) and the fact that the intersection of a countable collection of sets of probability one has probability one, there is a measurable subset Ω_0 of Ω such that $P(\Omega_0) = 1$, and for each $\omega \in \Omega_0$ we have $a_j = a_j(\omega) = (1/n_j) \sum_{r=1}^{n_j} |\Delta_{j,r}(\omega)| \rightarrow 0$, as $m \rightarrow \infty$, $1 \leq j \leq k$. Using (16) we get that

$$\frac{1}{n} \sum_{j,r} |\Delta_{j,r}| = \frac{1}{n} \sum_j n_j a_j = \frac{1}{\bar{c}} \frac{1}{k} \sum_{j=1}^k c_j a_j \leq \frac{C_0}{c_0} \frac{1}{k} \sum_{j=1}^k a_j,$$

and the result follows. □

8.2 Proof of main results

Proof of Proposition 1 We have that $\mathcal{T}_{n,\beta} = \int Z_n^2(t)w_\beta(t)dt$, with $Z_n(t) = Z_{1,n}$ and $Z_{k,n}$ is as defined in (7). By Taylor expansion we can write

$$\begin{aligned} \cos(tY_j) &= \cos(tX_j) - t \sin(t\check{X}_j)\Delta_j, \\ \sin(tY_j) &= \sin(tX_j) + t \cos(t\check{X}_j)\Delta_j, \end{aligned}$$

where $\Delta_j = X_j(1/S - 1) - \bar{X}/S$, \check{X}_j and \check{X}_j both lie between Y_j and X_j , $1 \leq j \leq n$. Thus, $Z_n(t) = \mathcal{Z}_{n,1}(t) + \mathcal{Z}_{n,2}(t) + 2t\mathcal{Z}_{n,3}(t)$, $t \in \mathbb{R}$, with

$$\begin{aligned} \mathcal{Z}_{n,1}(t) &= \frac{1}{n} \sum_{j=1}^n \{\cos(tX_j) - \varphi_0(t)\}, \\ \mathcal{Z}_{n,2}(t) &= \frac{1}{n} \sum_{j=1}^n \sin(tX_j), \\ \mathcal{Z}_{n,3}(t) &= \frac{1}{2n} \sum_{j=1}^n \{\cos(t\check{X}_j) - \sin(t\check{X}_j)\}\Delta_j. \end{aligned}$$

Let $\mathcal{T}_{n,i,\beta} = n \int Z_{n,i}^2(t)w_\beta(t)dt$, $1 \leq i \leq 3$. Since $n\mathcal{T}_{n,\beta}$ converges in law to its limit distribution, from Lemma 1.4.A and Theorem 1.4.A in Serfling (2009), a sufficient condition for the convergence of the two first moments of $n\mathcal{T}_{n,\beta}$ to those of the limit distribution is that $E\{(n\mathcal{T}_{n,\beta})^r\} < \infty, \forall n$, for some $r > 2$. With this aim, we will see that $E\{|n\mathcal{T}_{n,i,\beta}|^3\} = E\{(n\mathcal{T}_{n,i,\beta})^3\} < \infty, 1 \leq i \leq 3$.

Notice that $\mathcal{Z}_{n,1}(t)$ is the average of independent elements with zero mean and hence, routine calculations show that $E\{(n\mathcal{T}_{n,1,\beta})^3\} < \infty$. The same argument can be used to see that $E\{(n\mathcal{T}_{n,2,\beta})^3\} < \infty$. Finally, since

$$|\mathcal{Z}_{n,3}(t)| \leq \left(\frac{1}{n} \sum_{j=1}^n \Lambda_j^2\right)^{1/2} = \left\{(S - 1)^2 + \bar{X}^2\right\}^{1/2} \leq |S - 1| + |\bar{X}|,$$

to show that $E\{(n\mathcal{T}_{n,3,\beta})^3\} < \infty$, it suffices to see that $E\{[\sqrt{n}(S_n - 1)]^6\} < \infty$ and that $E\{(\sqrt{n}\bar{X})^6\} < \infty$. As $\sqrt{n}\bar{X} \sim N(0, 1)$, it follows that $E\{(\sqrt{n}\bar{X})^6\} = 15 < \infty$. On the other hand, $\sqrt{n}S_n$ has a chi-distribution with $n - 1$ degrees of freedom (see Chap. 18 of Johnson et al. (1994)). Deutler (1984) has calculated the first six cumulants of the chi-distribution. Using the results in Deutler (1984) it can be easily seen that $E\{[\sqrt{n}(S_n - 1)]^6\} < \infty$. The proof is concluded. \square

Proof of Theorem 1 By Taylor expansion we can write

$$\begin{aligned} \cos(tY_{j,r}) &= \cos(tX_{j,r}) - t \sin(tX_{j,r})\Delta_{j,r} - 0.5t^2 \sin(t\check{X}_{j,r})\Delta_{j,r}^2, \\ \sin(tY_{j,r}) &= \sin(tX_{j,r}) + t \cos(tX_{j,r})\Delta_{j,r} - 0.5t^2 \cos(t\check{X}_{j,r})\Delta_{j,r}^2, \end{aligned}$$

where $\check{X}_{j,r}$ and $\check{\check{X}}_{j,r}$ both lie between $Y_{j,r}$ and $X_{j,r}$. Thus,

$$\begin{aligned} \sqrt{n}Z_{k,n}(t) &= \frac{1}{\sqrt{n}} \sum_{j,r} \left\{ \cos(tX_{j,r}) + \sin(tX_{j,r}) - \varphi_0(t) - t \sin(tX_{j,r})\Delta_{j,r} \right. \\ &\quad \left. + t \cos(tX_{j,r})\Delta_{j,r} \right\} + r_1(t), \end{aligned}$$

where $|r_1(t)| \leq t^2(1/\sqrt{n}) \sum_{j,r} \Delta_{j,r}^2$. Since $\int t^2w(t)dt < \infty$, from Lemma 2 (a), it follows that $\|r_1\|_w = o_P(1)$, as $m \rightarrow \infty$.

We have that

$$-t \frac{1}{\sqrt{n}} \sum_{j,r} \sin(tX_{j,r})\Delta_{j,r} = r_2(t) - r_3(t) - r_4(t) + r_5(t),$$

where r_2 is the process in Lemma 2 (b.1), r_3 is the process in Lemma 2 (b.4), r_4 is the process in Lemma 2 (b.6), and

$$r_5(t) = t^2\varphi_0(t)0.5 \frac{1}{\sqrt{n}} \sum_{j,r} (X_{j,r}^2 - 1).$$

We also have that

$$t \frac{1}{\sqrt{n}} \sum_{j,r} \cos(tX_{j,r})\Delta_{j,r} = -r_6(t) - r_7(t) + r_8(t) + r_9(t),$$

where r_6 is the process in Lemma 2 (b.2), r_7 is the process in Lemma 2 (b.3), r_8 is the process in Lemma 2 (b.5), and

$$r_9(t) = -t\varphi_0(t) \frac{1}{\sqrt{n}} \sum_{j,r} X_{j,r}.$$

Summarizing,

$$\sqrt{n}Z_{k,n}(t) = W_{k,n}(t) + r_{10}(t), \quad t \in \mathbb{R},$$

with $\|r_{10}\|_w = o_P(1)$, as $m \rightarrow \infty$, and

$$\begin{aligned} W_{k,n}(t) &= \frac{1}{\sqrt{n}} \sum_{j,r} \left\{ \cos(tX_{j,r}) + \sin(tX_{j,r}) - \varphi_0(t) + \frac{1}{2}t^2\varphi_0(t)(X_{j,r}^2 - 1) - t\varphi_0(t)X_{j,r} \right\}, \quad t \in \mathbb{R}. \end{aligned}$$

Notice that $\{\cos(tX_{j,r}) + \sin(tX_{j,r}) - \varphi_0(t) + \frac{1}{2}t^2\varphi_0(t)(X_{j,r}^2 - 1) - t\varphi_0(t)X_{j,r}, t \in \mathbb{R}, 1 \leq r \leq n_j, 1 \leq j \leq k\}$ are n iid random elements taking values in L_w^2 with covariance kernel $C(t, s)$ in (8). By the central limit theorem in Hilbert spaces (see

Theorem 2.7 in Bosq (2000), $W_{k,n} \xrightarrow{\mathcal{D}} Z$ in L_w^2 , as $m \rightarrow \infty$. Finally, the assertion follows by applying the continuous mapping theorem. \square

Proof of Theorem 2 By Taylor expansion, proceeding as in the proof of Proposition 1, we can write

$$Z_{k,n}(t) = \frac{1}{n} \sum_{j,r} \{ \cos(tX_{j,r}) + \sin(tX_{j,r}) - \varphi_0(t) \} + r_{1n}(t),$$

where $Z_{k,n}$ is as defined in (7) and $|r_{1n}(t)| \leq t \frac{1}{n} \sum_{j,r} |\Delta_{j,r}|$. Since $\int t^2 w(t) dt = 1$, from Lemma 3, one gets that $\|r_{1n}\|_w \xrightarrow{a.s.} 0$.

Let $R_{0k}(t) = \Re \varphi_{0k}(t)$ and $W_{n,1}(t) = \frac{1}{n} \sum_{j,r} \{ \cos(tX_{j,r}) - R_{0k}(t) \}$. The reasoning in the proof of Lemma 3 can be used to prove (using now the SLLN in Hilbert spaces, see Theorem 2.4 in Bosq (2000)) that $W_{n,1} \xrightarrow{a.s.} 0$ in L_w^2 . Analogously, it can be seen that $W_{n,2} \xrightarrow{a.s.} 0$ in L_w^2 , with $W_{n,2}(t) = \frac{1}{n} \sum_{j,r} \{ \sin(tX_{j,r}) - I_{0k}(t) \}$ and $I_{0k}(t) = \Im \varphi_{0k}(t)$.

Summarizing,

$$Z_{k,n}(t) = R_{0k}(t) + I_{0k}(t) - \varphi_0(t) + r_{2n}(t), \quad t \in \mathbb{R},$$

with $\|r_{2n}\|_w \xrightarrow{a.s.} 0$. Recalling that $T_{k,n} = \|Z_{k,n}\|_w^2$ and noticing that $\|R_{0k} + I_{0k} - \varphi_0\|_w = \|\varphi_{0k} - \varphi_0\|_w < \infty$, we get that

$$\begin{aligned} \left| T_{k,n} - \|\varphi_{0k} - \varphi_0\|_w^2 \right| &= \left| \|r_{2n}\|_w^2 + 2\langle R_{0k} + I_{0k} - \varphi_0, r_{2n} \rangle_w \right| \\ &\leq \|r_{2n}\|_w^2 + 2\|r_{2n}\|_w \|\varphi_{0k} - \varphi_0\|_w \xrightarrow{a.s.} 0, \end{aligned}$$

which proves the result. \square

Proof of Theorem 3 We have that

$$\text{power} := P(T_{k_0, n_0} \geq t_{k_0, n_0, \alpha}) = \Sigma_0 + \Sigma_1 + \dots + \Sigma_{k_0},$$

where

$$\Sigma_j = \frac{1}{\binom{k}{k_0}} \sum_{\mathbf{i} \in \mathfrak{J}_j} a(\mathbf{i}),$$

$\mathfrak{J}_0 = \{ \mathbf{i} = (i_1, \dots, i_{k_0}) : r < i_1 < \dots < i_{k_0} \leq k \}$, $\mathfrak{J}_j = \{ \mathbf{i} = (i_1, \dots, i_{k_0}) : 1 \leq i_1 < \dots < i_j \leq r < i_{r+1} < \dots < i_{k_0} \leq k \}$, $1 \leq j \leq k_0$, and $a(\mathbf{i})$ is as defined in (15). If, for some j , \mathfrak{J}_j is empty, then we define $\Sigma_j = 0$.

Suppose that \mathfrak{J}_j is not empty. Let $\mathbf{i} \in \mathfrak{J}_j$, $n_0 = n_{i_1} + \dots + n_{i_{k_0}}$, $f_{\mathbf{i}} = (1/n_0) \sum_{v=1}^j n_{i_v}$,

$$\tilde{\varphi}_{\mathbf{i}} = \frac{1}{n_0} \sum_{v=1}^j n_{i_v} \varphi_{i_v} + (1 - f_{\mathbf{i}}) \varphi_0, \quad \varphi_{\mathbf{i}} = \frac{1}{n_0 f_{\mathbf{i}}} \sum_{v=1}^j n_{i_v} \varphi_{i_v}.$$

Then, $\tilde{\varphi}_{\mathbf{i}} - \varphi_0 = f_{\mathbf{i}}(\varphi_{\mathbf{i}} - \varphi_0)$, and thus, if the sample sizes satisfy (2) and $\varphi_1, \dots, \varphi_r$ satisfy (12), we obtain

$$\frac{j}{k_0} M_1 \leq \|\tilde{\varphi}_{\mathbf{i}} - \varphi_0\|_w \leq \frac{j}{k_0} M_2, \quad \forall \mathbf{i} \in \mathfrak{J}_j, \quad 1 \leq j \leq \min\{r, k_0\},$$

where M_1 and M_2 are two positive constants (depending on τ , c_0 and C_0).

Let $0 < p_0 < p$. We have that

$$\text{power} = \sum_{u: u \leq p_0 k_0} \Sigma_u + \sum_{u: u > p_0 k_0} \Sigma_u := \Sigma_{\leq p_0} + \Sigma_{> p_0}.$$

Let $H \sim H(k, r, k_0)$, where $H(k, r, k_0)$ stands for a hypergeometric distribution: from a population with k units, r of type A and $k - r$ of type B, a sample without replacement of size k_0 is selected, H is the number of units type A in the sample. Let $\mathbb{I}(p_0) = \{\mathbf{i} \in \mathfrak{J}_u, u > p_0 k_0\}$ and $N_{p_0} = \text{card}\{\mathbb{I}(p_0)\}$. We can write

$$\Sigma_{> p_0} = \sum_{u: u > p_0 k_0} P(H = u) + \frac{1}{\binom{k}{k_0}} \sum_{\mathbf{i} \in \mathbb{I}(p_0)} (a(\mathbf{i}) - 1) := W_1 + W_2.$$

We can write $W_1 = 1 - P(H \leq k_0 p_0)$. Since (see Hoeffding 1963) $P(H \leq k_0 p_0) \leq \exp\{-2(p - p_0)^2 k_0\}$, it follows that W_1 converges to 1, as $m \rightarrow \infty$. On the other hand, from Corollary 1, for each $\mathbf{i} \in \mathfrak{J}_u$, with $u > p_0 k_0$, we have that $a(\mathbf{i}) \rightarrow 1$ and thus

$$\bar{a} = \frac{1}{N_{p_0}} \sum_{\mathbf{i} \in \mathbb{I}(p_0)} a(\mathbf{i}) \rightarrow 1.$$

Since $W_2 = (\bar{a} - 1) N_{p_0} / \binom{k}{k_0}$ and $0 \leq N_{p_0} / \binom{k}{k_0} \leq 1$, it follows that $W_2 \rightarrow 0$. Therefore, $\text{power} \rightarrow 1$. □

Proof of Theorem 4 $\mathbb{T}_{0,k}$ is a sum of independent zero-mean random variables. Thus, it suffices to see that Lindeberg condition below is met,

$$h(\varepsilon) = \frac{\sum_{i=1}^k E_0 \left\{ (\mathfrak{T}_i - \mu_{0,n_i})^2 \mathbf{1}[(\mathfrak{T}_i - \mu_{0,n_i})^2 > \varepsilon \sum_{i=1}^k \tau_{0,n_i}^2] \right\}}{\sum_{i=1}^k \tau_{0,n_i}^2} \rightarrow 0, \quad \forall \varepsilon > 0, \quad \text{as } k \rightarrow \infty,$$

where $\mathbf{1}(\cdot)$ stands for the indicator function. From Proposition 1 it follows that $\sum_{i=1}^k \tau_{0,n_i}^2 \geq k\tau_0, \forall k$, for some $\tau_0 > 0$. As a consequence, we have that,

$$0 < h(\varepsilon) \leq H(\varepsilon) = \frac{1}{k\tau_0} \sum_{i=1}^k \mu(n_i, k, \varepsilon),$$

with $\mu(n_i, k, \varepsilon) = E_0 \{ (\mathfrak{X}_i - \mu_{0,n_i})^2 \mathbf{1}[(\mathfrak{X}_i - \mu_{0,n_i})^2 > \varepsilon \tau_0 k] \}$. From Proposition 1, it also follows that $\lim_{k \rightarrow \infty} \sup_n \mu(n, k, \varepsilon) = 0$. This fact implies that $H(\varepsilon) \rightarrow 0$ as $k \rightarrow \infty$, and thus the result is proven. \square

Proof of Theorem 5 We first prove that

$$\mathbb{T}_{r,k} = \frac{\sum_{i=1}^k (\mathfrak{X}_i - \mu_{n_i})}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} \xrightarrow{\mathcal{D}} Z \sim N(0, 1), \quad \text{as } k \rightarrow \infty, \tag{17}$$

where

$$\mu_{n_i} = \begin{cases} n_i \Delta_{X_i, n_i} & \text{if } 1 \leq i \leq r, \\ \mu_{0,n_i} & \text{if } r + 1 \leq i \leq k, \end{cases} \quad \tau_{n_i}^2 = \begin{cases} n_i \tau_{X_i, n_i}^2 & \text{if } 1 \leq i \leq r, \\ \tau_{0,n_i}^2 & \text{if } r + 1 \leq i \leq k. \end{cases}$$

With this aim, we prove that the Lindeberg condition below is met,

$$h(\varepsilon) = \frac{\sum_{i=1}^k E \left\{ (\mathfrak{X}_i - \mu_{n_i})^2 \mathbf{1}[(\mathfrak{X}_i - \mu_{n_i})^2 > \varepsilon \sum_{i=1}^k \tau_{n_i}^2] \right\}}{\sum_{i=1}^k \tau_{n_i}^2} \rightarrow 0, \quad \forall \varepsilon > 0, \quad \text{as } k \rightarrow \infty.$$

We can write $h(\varepsilon) = h_1(\varepsilon) + h_2(\varepsilon)$, with

$$h_1(\varepsilon) = \frac{\sum_{i=1}^r E \left\{ (\mathfrak{X}_i - \mu_{n_i})^2 \mathbf{1}[(\mathfrak{X}_i - \mu_{n_i})^2 > \varepsilon \sum_{i=1}^k \tau_{n_i}^2] \right\}}{\sum_{i=1}^k \tau_{n_i}^2},$$

and $h_2(\varepsilon) = h(\varepsilon) - h_1(\varepsilon)$.

From Proposition 1 and Assumption 1(b), there exists $0 < \varsigma_1 \leq \varsigma_2 < \infty$ such that

$$\varsigma_1 \leq \min\{\inf_n \tau_{0,n}^2, \inf_i \tau_{X_i, n_i}^2\} \leq \max\{\sup_n \tau_{0,n}^2, \sup_i \tau_{X_i, n_i}^2\} \leq \varsigma_2. \tag{18}$$

As a consequence, $\sum_{i=1}^k \tau_{n_i}^2 \geq k\varsigma_1$. Now proceeding as in the proof of Theorem 4, we have that $h_2(\varepsilon) \rightarrow 0, \forall \varepsilon > 0$.

Taking into account that the sample sizes satisfy (2), we can write

$$\sum_{i=1}^k \tau_{n_i}^2 \geq \sum_{i=1}^r \tau_{n_i}^2 = \sum_{i=1}^r n_i \tau_{X_i, n_i}^2 \geq mrc_0 \varsigma_1,$$

and hence

$$h_1(\varepsilon) \leq H_1(\varepsilon) = \frac{\sum_{i=1}^r E \left\{ (\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i, n_i}))^2 \mathbf{1}[(\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i, n_i}))^2 > \varepsilon r \zeta_1 c_0 / C_0] \right\}}{m r c_0 \zeta_1},$$

Because $r/k \rightarrow p \in (0, 1]$, as $k \rightarrow \infty$, it implies that $r \rightarrow \infty$ as $k \rightarrow \infty$. From Assumption 1(a), it follows that

$$\lim_{r \rightarrow \infty} E \left\{ (\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i, n_i}))^2 \mathbf{1}[(\sqrt{n_i}(\mathcal{T}_i - \Delta_{X_i, n_i}))^2 > \varepsilon r \zeta_1 c_0 / C_0] \right\} = 0, \quad 1 \leq i \leq r.$$

As a consequence, we have that $h_1(\varepsilon) \rightarrow 0$, as $k \rightarrow \infty$, $\forall \varepsilon > 0$. Therefore, (17) has been proven.

Now we can study the power. We have that

$$\text{power} = P(\mathbb{T}_{0,k} \geq z_{1-\alpha}) = P \left(\mathbb{T}_{r,k} \geq z_{1-\alpha} \frac{\sqrt{\sum_{i=1}^k \tau_{0,n_i}^2}}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} - \frac{\sum_{i=1}^r (n_i \Delta_{X_i, n_i} - \mu_{0,n_i})}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} \right).$$

From (17), we can write

$$\text{power} \approx \Phi \left(\frac{\sum_{i=1}^r (n_i \Delta_{X_i, n_i} - \mu_{0,n_i})}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} - z_{1-\alpha} \frac{\sqrt{\sum_{i=1}^k \tau_{0,n_i}^2}}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} \right).$$

From (18) and using that the sample sizes satisfy (2), we can write

$$\frac{1}{r} \sum_{i=1}^k \tau_{n_i}^2 = \frac{1}{r} \sum_{i=1}^r n_i \tau_{X_i, n_i}^2 + \frac{1}{r} \sum_{i=r+1}^k \tau_{0,n_i}^2 \leq C_0 m \zeta_2 \frac{k}{r}.$$

Let $\varepsilon > 0$ be such that $p - \varepsilon > 0$. Since $r/k \rightarrow p$, as $k \rightarrow \infty$, we have that $p - \varepsilon \leq r/k$, $\forall k \geq k(\varepsilon)$, for some large enough $k(\varepsilon)$. Thus, $\forall k \geq k(\varepsilon)$,

$$\frac{1}{r} \sum_{i=1}^k \tau_{n_i}^2 \leq C_0 m \zeta_2 / (p - \varepsilon),$$

and hence, using Assumption 1(c), we have that

$$\frac{\sum_{i=1}^r (n_i \Delta_{X_i, n_i} - \mu_{0,n_i})}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} \geq \sqrt{r} \frac{\eta c_0 \sqrt{m}}{\sqrt{C_0 \zeta_2 / (p - \varepsilon)}}, \tag{19}$$

$\forall k \geq k(\varepsilon)$. On the other hand, from (18),

$$\frac{\sum_{i=1}^k \tau_{0,n_i}^2}{\sum_{i=1}^k \tau_{n_i}^2} = \frac{\sum_{i=1}^k \tau_{0,n_i}^2}{\sum_{i=1}^r n_i \tau_{X_i,n_i}^2 + \sum_{i=r+1}^k \tau_{0,n_i}^2} \leq \frac{\xi_2}{\xi_1} < \infty. \quad (20)$$

From (19) and (20),

$$\frac{\sum_{i=1}^r (n_i \Delta_{X_i,n_i} - \mu_{0,n_i})}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} - z_{1-\alpha} \frac{\sqrt{\sum_{i=1}^k \tau_{0,n_i}^2}}{\sqrt{\sum_{i=1}^k \tau_{n_i}^2}} \geq \sqrt{r} \frac{\eta c_0 \sqrt{m}}{\sqrt{C_0 \xi_2 / (p - \varepsilon)}} - \frac{\xi_2}{\xi_1} |z_{1-\alpha}|,$$

$\forall k \geq k(\varepsilon)$, and the right-hand side of the above inequality goes to ∞ as $k \rightarrow \infty$, implying that power $\rightarrow 1$, as $k \rightarrow \infty$. \square

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00362-022-01384-y>.

Acknowledgements The author thanks two anonymous referees for their constructive comments and suggestions which helped to improve the presentation. M.D. Jiménez-Gamero has been partially supported by grants PID2020-118101GB-I00, Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033), P18-FR-2269 (Junta de Andalucía) and US-1381178 (Junta de Andalucía).

Funding Funding for open access publishing: Universidad de Sevilla/CBUA

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

References

- Baringhaus L, Henze N (1988) A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35(6):339–348
- Baringhaus L, Ebner B, Henze N (2017) The limit distribution of weighted L^2 -goodness-of-fit statistics under fixed alternatives, with applications. *Ann Inst Stat Math* 69(5):969–995
- Becker M, Klößner S (2022) *PearsonDS: Pearson Distribution System*. R package version 1(2):2
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188
- Bosq D (2000) *Linear processes in function spaces. Lecture notes in statistics, vol. 149*. Springer, New York [Theory and applications]
- Butsch L, Ebner B (2020) *mnt: affine invariant tests of multivariate normality*. R package version 1:3
- Cuesta-Albertos JA, Febrero-Bande M (2010) A simple multiway ANOVA for functional data. *TEST* 19(3):537–557
- Cuesta-Albertos JA, Fraiman R, Ransford T (2006) Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bull Braz Math Soc (NS)* 37(4):477–501

- Deutler T (1984) A series expansion for the cumulants of the x -distribution and a Cornish-Fisher-expansion for the noncentrality parameter of the noncentral t -distribution. *Commun Stat Simul Comput* 13(4):507–513
- Ebner B, Henze N (2020) Tests for multivariate normality—a critical review with emphasis on weighted L^2 -statistics. *TEST* 29(4):845–892
- Ebner B, Henze N (2021) Bahadur efficiencies of the Epps-Pulley test for normality. *Zapiski Nauchnykh Semin* 501:302–314
- Ebner B, Henze N (2022) On the eigenvalues associated with the limit null distribution of the Epps-Pulley test of normality. *Stat Pap*. <https://doi.org/10.1007/s00362-022-01336-6>
- Ebner B, Henze N, Strieder D (2022) Testing normality in any dimension by Fourier methods in a multivariate Stein equation. *Can J Stat* 50(3):992–1033
- Epps TW, Pulley LB (1983) A test for normality based on the empirical characteristic function. *Biometrika* 70(3):723–726
- Gaigall D (2021) On a new approach to the multi-sample goodness-of-fit problem. *Commun Stat Simul Comput* 50(10):2971–2989
- Henze N (1990) An approximation to the limit distribution of the Epps-Pulley test statistic for normality. *Metrika* 37(1):7–18
- Henze N, Jiménez-Gamero MD (2019) A new class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. *TEST* 28(2):499–521
- Henze N, Jiménez-Gamero MD (2021) A test for Gaussianity in Hilbert spaces via the empirical characteristic functional. *Scand J Stat* 48(2):406–428
- Henze N, Koch S (2020) On a test of normality based on the empirical moment generating function. *Stat Pap* 61(1):17–29
- Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. *J Multivar Anal* 62(1):1–23
- Henze N, Jiménez-Gamero MD, Meintanis SG (2019) Characterizations of multinormality and corresponding tests of fit, including for Garch models. *Econom Theory* 35(3):510–546
- Hoeffding W (1963) Probability inequalities for sums of bounded random variables. *J Am Stat Assoc* 58:13–30
- Hušková M, Meintanis SG (2010) Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *TEST* 19(1):92–112
- Jelito D, Pitera M (2021) New fat-tail normality test based on conditional second moments with applications to finance. *Stat Pap* 62(5):2083–2108
- Jiménez-Gamero MD (2014) On the empirical characteristic function process of the residuals in GARCH models and applications. *TEST* 23(2):409–432
- Jiménez-Gamero MD, Franco-Pereira AM (2021) Testing the equality of a large number of means of functional data. *J Multiv Anal* 185:Paper No. 104778, 18
- Jiménez-Gamero MD, Muñoz García J, Pino-Mejías R (2005) Testing goodness of fit for the distribution of errors in multivariate linear models. *J Multivar Anal* 95(2):301–322
- Jiménez-Gamero MD, Alba-Fernández MV, Jodrá P, Chalco-Cano Y (2016) Testing for the symmetric component in skew distributions. *Math Methods Appl Sci* 39(16):4713–4722
- Jiménez-Gamero MD, Cousido-Rocha M, Alba-Fernández MV, Jiménez-Jiménez F (2022) Testing the equality of a large number of populations. *TEST* 31(1):1–21
- Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley series in probability and mathematical statistics: applied probability and statistics, 2nd edn. Wiley, New York
- Klar B, Lindner F, Meintanis SG (2012) Specification tests for the error distribution in GARCH models. *Comput Stat Data Anal* 56(11):3587–3598
- Meintanis SG, Milošević B, Obradović M (2022) Bahadur efficiency for certain goodness-of-fit tests based on the empirical characteristic function. *arXiv preprint*. [arXiv:2201.08604](https://arxiv.org/abs/2201.08604)
- Olver FWJ, Lozier DW, Boisvert RF, Clark CW (eds) (2010) NIST handbook of mathematical functions. U.S. Department of Commerce, National Institute of Standards and Technology/Cambridge University Press, Washington, DC/Cambridge
- Park J, Park D (2012) Testing the equality of a large number of normal population means. *Comput Stat Data Anal* 56(5):1131–1149
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

- Rivas-Martínez GI, Jiménez-Gamero MD (2018) Computationally efficient goodness-of-fit tests for the error distribution in nonparametric regression. *REVSTAT* 16(1):137–166
- Serfling RJ (2009) *Approximation theorems of mathematical statistics*, vol 162. Wiley, New York
- Tenreiro C (2009) On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Comput Stat Data Anal* 53(4):1038–1053
- Zhan D, Hart JD (2014) Testing equality of a large number of densities. *Biometrika* 101(2):449–464

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.