Checking the adequacy of the multivariate semiparametric location shift model

N. Henze [a,1], B. Klar [a], L.X. Zhu [b,2]

[a] *Institute of Mathematical Stochastics, University of Karlsruhe, Germany*

[b] *Dept. of Statistics and Actuarial Sciences, University of Hong Kong, and Chinese Academy of Sciences*

Abstract

Let $X, X_1, \ldots, X_m, \ldots, Y, Y_1, \ldots, Y_n, \ldots$ be independent d-dimensional random vectors, where the X_j are i.i.d. copies of X, and the Y_k are i.i.d. copies of Y. We study a class of consistent tests for the hypothesis that Y has the same distribution as $X + \mu$ for some unspecified $\mu \in \mathbb{R}^d$. The test statistic L is a weighted integral of the squared modulus of the difference of the empirical characteristic functions of $X_1 + \hat{\mu}, \ldots, X_m + \hat{\mu}$ and $Y_1, \ldots, Y_n$, where $\hat{\mu}$ is an estimator of μ. An alternative representation of L is given in terms of an L^2-distance between two nonparametric density estimators. The finite-sample and asymptotic null distribution of L is independent of μ. Carried out as a bootstrap or permutation procedure, the test is asymptotically of a given size, irrespective of the unknown underlying distribution. A large scale simulation study shows that the permutation procedure performs better than the bootstrap.

Key words: Goodness-of-fit test, multivariate location model, empirical characteristic function, permutational principle, bootstrap
1991 MSC: 62H15, 62G10, 62G20

---

[1] Research started while the first author was on leave at the University of Hong Kong
[2] The third author's work was supported by a grant from the Chinese Academy of Sciences and a grant from the University Grant Council of Hong Kong, HKSAR, Hong Kong, China

1 Introduction

Let $X_1,\ldots,X_m,\ldots;Y_1,\ldots,Y_n,\ldots$ be independent d-dimensional random (column) vectors, which are defined on a common probability space $(\Omega,\mathcal{A},P)$. The X_j are i.i.d. copies of a random vector X, and the Y_k are i.i.d. copies of a random vector Y. The distributions of X and Y are assumed to be continuous. Within this framework of the so-called *general nonparametric two-sample model*, an important submodel is the multivariate *two-sample location model*, which states that

$$Y \ \sim \ X \ + \ \mu \quad \text{for some unspecified } \mu \in \mathbb{R}^d. \tag{1.1}$$

Here and in what follows, '$\sim$' denotes equality in distribution. Although many statistical procedures are tailored to the situation of the two-sample location model (and even make further distributional assumptions, such as Hotelling's T^2-test which assumes the underlying distributions to be normal), there has not been any attempt to check the validity of (1.1) within the *general* setting stated at the beginning of this section, at least to the authors' knowledge. [16] considers the testing problem (1.1) under the unnatural restrictive assumption that the underlying distributions are diagonally symmetric.

This paper introduces and studies a class of goodness-of-fit tests of the hypothesis (1.1). The test statistics are based on the empirical characteristic function, which has proved to be a powerful tool in statistical inference (see, e.g. [11], [3], [4], [7]). To be specific, let

$$\varphi(t) \ = \ E\left[\exp(it'X)\right], \quad \psi(t) = E\left[\exp(it'Y)\right]$$

denote the characteristic functions of X and Y, respectively, where the prime stands for transpose of vectors and matrices. Since (1.1) is equivalent to

$$\psi(t) \ = \ \varphi(t)\,\exp(it'\mu) \ (t \in \mathbb{R}^d) \quad \text{for some } \mu \in \mathbb{R}^d,$$

it nearly suggests itself to base a test of (1.1) on some suitable measure of deviation of the random function $\psi_n(t) - \varphi_m(t)\exp(it'\hat{\mu})$, $t \in \mathbb{R}^d$, from the zero function. Here,

$$\varphi_m(t) \ = \ \frac{1}{m}\sum_{j=1}^{m}\exp(it'X_j), \quad \psi_n(t) \ = \ \frac{1}{n}\sum_{k=1}^{n}\exp(it'Y_k)$$

are the empirical characteristic functions of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively, and $\hat{\mu} = \hat{\mu}_{m,n}(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ is an estimator of μ that is based on $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. We assume that $\hat{\mu}$ is *location equivariant* in the sense that

$$\hat{\mu}(X_1, \ldots, X_m, Y_1 + a, \ldots, Y_n + a) = \hat{\mu}(X_1, \ldots, X_m, Y_1, \ldots, Y_n) + a \tag{1.2}$$

and

$$\hat{\mu}(X_1 + a, \ldots, X_m + a, Y_1, \ldots, Y_n) = \hat{\mu}(X_1, \ldots, X_m, Y_1, \ldots, Y_n) - a \tag{1.3}$$

for each $a \in \mathbb{R}^d$. These conditions are natural since, in view of (1.1), μ measures the difference in location between the distributions of Y and X. Thus, shifting the Y_k by the vector a should increase the difference in location by a. Likewise, translating the X_j by the same amount should decrease the difference in location by a. A consequence of (1.2) and (1.3) is the invariance of $\hat{\mu}$ with respect to translations of the pooled sample by the same vector. A further regularity condition on $\hat{\mu}$ will be specified later.

A simple example of a location equivariant estimator is $\hat{\mu} = \bar{Y}_n - \bar{X}_m$, where $\bar{X}_m = m^{-1} \sum_{j=1}^m X_j$, $\bar{Y}_n = n^{-1} \sum_{k=1}^n Y_k$ denote the sample means of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively.

In the spirit of a class of tests for multivariate normality (see [10], [9]), the test statistic we propose is the weighted L^2-distance

$$L = L_{m,n,\beta} = \int_{\mathbb{R}^d} \left| \psi_n(t) - \varphi_m(t) \exp(it'\hat{\mu}) \right|^2 w_\beta(t) | \, dt, \tag{1.4}$$

where

$$w_\beta(t) = \frac{1}{(2\pi\beta^2)^{d/2}} \exp\left(-\frac{\|t\|^2}{2\beta^2} \right) \tag{1.5}$$

is the density of the centered d-dimensional normal distribution $\mathcal{N}(0, \beta^2 I_d)$ with independent components and marginal variances β^2. With the exception of Section 3, $\beta > 0$ will be fixed in what follows.

Using the relation

$$\int_{\mathbb{R}^d} \cos(z't)\, w_\beta(t)\, dt \;=\; \exp\left(-\frac{\beta^2\|z\|^2}{2}\right)$$

and symmetry arguments, straightforward algebra shows that L takes the simple form

$$L = \frac{1}{m^2}\sum_{j,k=1}^{m}\exp\left(-\frac{\beta^2}{2}\|X_j - X_k\|^2\right) - \frac{2}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n}\exp\left(-\frac{\beta^2}{2}\|X_j + \hat{\mu} - Y_k\|^2\right)$$
$$+\frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\|Y_j - Y_k\|^2\right). \tag{1.6}$$

Thus, a computer routine for implementing the test statistic is readily available.

Notice further that, by (1.2) and (1.3), the value of L remains unchanged under translations $X_j \mapsto X_j + a$ $(j = 1, \dots, m)$ or $Y_k \mapsto Y_k + a$ $(k = 1, \dots, n)$. Consequently, the distribution of L under (1.1) does not depend on the nuisance parameter μ.

Interestingly, the statistic L has a completely different representation in terms of a measure of distance between two nonparametric density estimators.

Proposition 1.1 *We have*

$$L_{m,n,\beta} \;=\; \frac{(2\pi)^{d/2}}{\beta^d}\int_{\mathbb{R}^d}\left(\hat{f}_{m,h}(x) - \hat{g}_{n,h}(x)\right)^2\, dx, \tag{1.7}$$

where

$$\hat{f}_{m,h}(x) \;=\; \frac{1}{m}\sum_{j=1}^{m}\left(2\pi h^2\right)^{-d/2}\exp\left(-\frac{\|x-(X_j+\hat{\mu})\|^2}{2h^2}\right),$$
$$\hat{g}_{n,h}(x) \;=\; \frac{1}{n}\sum_{k=1}^{n}\left(2\pi h^2\right)^{-d/2}\exp\left(-\frac{\|x-Y_k\|^2}{2h^2}\right),$$

and

$$h \;=\; \frac{1}{2\beta^2}. \tag{1.8}$$

PROOF. Let $L^2(\mathbb{R}^d)$ denote the Hilbert space of measurable complex-valued functions on $\mathbb{R}^d$ that are square integrable with respect to Lebesgue measure. It is well known that the Fourier transform

$$\tilde{u}(x) = \int_{\mathbb{R}^d} \exp(ix't)u(t)\,dt$$

of $u \in L^2(\mathbb{R}^d)$ belongs to $L^2(\mathbb{R}^d)$, and that, by Plancherel's theorem,

$$\int_{\mathbb{R}^d} |\tilde{u}(x)|^2\,dx \;=\; (2\pi)^d \int_{\mathbb{R}^d} |u(t)|^2\,dt. \qquad (1.9)$$

¿From (1.4), we have

$$L_{m,n,\beta} = \frac{1}{(2\pi\beta^2)^{d/2}} \int_{\mathbb{R}^d} \left| \psi_n(t) \exp\left(-\frac{\|t\|^2}{4\beta^2}\right) - \varphi_m(t)\exp(it'\hat{\mu})\exp\left(-\frac{\|t\|^2}{4\beta^2}\right)\right|^2 dt.$$

Writing $\mathcal{P}_m$ for the empirical distribution that puts mass $1/m$ on each of the data $X_j+\hat{\mu}$ $(j = 1,\ldots,m)$, and letting $\mathcal{Q}_n$ be the empirical distribution of $Y_1,\ldots,Y_n$, the function $\psi_n(t)\exp(-\|t\|^2/(4\beta^2))$ is the Fourier transform of the convolution $\mathcal{Q}_n * \mathcal{N}(0,(2\beta^2)^{-1}I_d)$, and $\varphi_m(t)\exp(it'\hat{\mu})\exp(-\|t\|^2/(4\beta^2))$ is the Fourier transform of the convolution $\mathcal{P}_m * \mathcal{N}(0,(2\beta^2)^{-1}I_d)$. Since $\mathcal{P}_m * \mathcal{N}(0,(2\beta^2)^{-1}I_d)$ and $\mathcal{Q}_n * \mathcal{N}(0,(2\beta^2)^{-1}I_d)$ have densities $\hat{f}_{m,h}$ and $\hat{g}_{n,h}$, respectively, the assertion follows immediately from (1.9). ■

REMARK. Representation (1.7) reveals that the role of β figuring in the weight function (1.5) is that of a smoothing parameter, which determines the bandwidth h of the density estimators $\hat{f}_{m,h}$ and $\hat{g}_{n,h}$ via (1.8). $\hat{f}_{m,h}$ aims at estimating the distribution of $X+\mu$, which under (1.1) coincides with the distribution of Y. The latter, in turn, is estimated by $\hat{g}_{n,h}$. A similar phenomenon was observed in the context of testing for multivariate normality (see [9]). In fact, the test statistic of Bowman and Foster (see [2]), which was motivated by density estimation and involves a bandwidth depending on the sample size, turned out to be a special member of a class of tests based on an L^2-distance between characteristic functions with a fixed 'bandwidth'. Interestingly, keeping the bandwidth fixed ensures positive asymptotic power of the test against contiguous alternatives that approach the hypothesis at the rate $1/\sqrt{n}$ (see Theorem

3.1 of [9]) However, as shown by Gürtler (see [6]), the test of Bowman and Foster is not able to detect such alternatives.

Anderson et al. [1] and Fan [5] established results similar to proposition 1.1 in a one- and two-sample goodness-of-fit setting, respectively, and arrived at analogous conclusions concerning power with a fixed/decreasing bandwidth.

2 Asymptotic distribution theory

In view of the representation (1.4), a convenient setting for asymptotic distribution theory is the separable Hilbert space $\mathcal{L}^2$ of measurable real-valued functions on $\mathbb{R}^d$ that are square integrable with respect to the normal distribution $\mathcal{N}(0, \beta^2 I_d)$. The inner product and the norm in $\mathcal{L}^2$ will be denoted by

$$\langle g, h \rangle \;=\; \int_{\mathbb{R}^d} g(t)h(t)\, w_\beta(t)\, dt, \qquad \|g\|_{\mathcal{L}^2} \;=\; \langle g, g \rangle^{1/2},$$

respectively. Weak convergence of random elements of $\mathcal{L}^2$ and random variables is denoted by $\Longrightarrow$. $O_P(1)$ means a sequence of random elements that is bounded in probability, and $o_P(1)$ stands for convergence to zero in probability.

We first study the limit behavior of $L_{m,n,\beta}$ under the null hypothesis H_0 that (1.1) holds. All limits refer to the case that the sample sizes m and n tend to infinity in such a way that

$$\lim \frac{m}{m+n} \;=\; p \quad \text{for some } p \text{ satisfying } \; 0 < p < 1, \tag{2.1}$$

which we call the *usual limiting regime*. The total sample size will be denoted by $N = m + n$.

In addition to (1.2) and (1.3), we impose the following regularity condition on $\hat{\mu}$:

There is a measurable function $l : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that

$$\hat{\mu} - \mu \;=\; -\frac{1}{m}\sum_{j=1}^{m} l(X_j + \mu) \;+\; \frac{1}{n}\sum_{k=1}^{n} l(Y_k) \;+\; o_P(1/\sqrt{N}). \tag{2.2}$$

Moreover, $El(Y) = 0$ and $E\|l(Y)\|^2 < \infty$.

If $\hat{\mu} = \bar{Y}_n - \bar{X}_m$, then, assuming $E\|Y\|^2 < \infty$, (2.2) holds setting $l(z) = z - E[Y]$. Putting

$$Z_{m,n}(t) \;=\; \sqrt{\frac{n}{N}}\, U_m(t) \;-\; \sqrt{\frac{m}{N}}\, V_n(t),$$

where

$$U_m(t) \;=\; \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\Big(\cos(t'(X_j + \hat{\mu})) + \sin(t'(X_j + \hat{\mu})) \;-\; \Psi(t)\Big), \tag{2.3}$$

$$V_n(t) \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\cos(t'Y_j) + \sin(t'Y_j) \;-\; \Psi(t)\Big),$$

$$\Psi(t) \;=\; E\Big(\cos(t'Y) + \sin(t'Y)\Big), \tag{2.4}$$

it is readily seen that

$$\frac{mn}{m+n}\, L_{m,n,\beta} \;=\; \int\limits_{\mathbb{R}^d} Z^2_{m,n}(t)\, w_\beta(t)\, dt. \tag{2.5}$$

We will show that the process $Z_{m,n}(\cdot)$, regarded as a random element of $\mathcal{L}^2$, will converge in distribution to some centered Gaussian process $W(\cdot)$. Since the right-hand side of (2.5) is $\|Z_{m,n}\|^2_{\mathcal{L}^2}$, the continuous mapping theorem then yields the convergence in distribution of $mn/(m+n)L_{m,n,\beta}$ to $\int_{\mathbb{R}^d} W^2(t)w_\beta(t)dt$.

Performing a second-order Taylor expansion of $\cos(t'(X_j + \hat{\mu})) = \cos(t'(X_j + \mu) + t'(\hat{\mu} - \mu))$ around $t'(X_j + \mu)$ (and likewise for the sine term figuring in (2.3)), we obtain

$$U_m(t) = U_{m,1}(t) + U_{m,2}(t) + U_{m,3}(t),$$

where

$$U_{m,1}(t) = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\Big(\cos(t'(X_j + \mu)) + \sin(t'(X_j + \mu)) - \Psi(t)\Big),$$

$$U_{m,2}(t) = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \Big(\cos(t'(X_j+\mu)) - \sin(t'(X_j+\mu)) \Big) \cdot t'(\hat{\mu}-\mu),$$

and

$$|U_{m,3}(t)| \le \frac{1}{2}\sqrt{m}\ (t'(\hat{\mu}-\mu))^2 \ \le\ \frac{1}{2}\sqrt{m}\ \|\hat{\mu}-\mu\|^2\|t\|^2.$$

Since $\sqrt{N}(\hat{\mu}-\mu) = O_P(1)$ by (2.2) and the Lindeberg-Feller central limit theorem, it follows that

$$\|U_{m,3}\|_{\mathcal{L}^2} \ =\ o_P(1). \tag{2.6}$$

We next consider $U_{m,2}(t)$. Putting

$$\bar{\Psi}_m(t) = \frac{1}{m} \sum_{j=1}^{m} \Big(\cos(t'(X_j+\mu)) - \sin(t'(X_j+\mu)) \Big),$$

note that $E\bar{\Psi}_m(t) = \bar{\Psi}(t)$ where, by analogy with (2.4), we define

$$\bar{\Psi}(t) \ =\ E\Big(\cos(t'Y) - \sin(t'Y) \Big). \tag{2.7}$$

Now,

$$\|U_{m,2}(\cdot) - \bar{\Psi}(\cdot)\cdot{}'\sqrt{m}(\hat{\mu}-\mu)\|_{\mathcal{L}^2} \ \le\ \sqrt{m}\|\hat{\mu}-\mu\| \int_{\mathbb{R}^d} \|t\|^2 \Big(\bar{\Psi}_m(t) - \Psi(t)\Big)^2 w_\beta(t)dt,$$

which is $o_P(1)$ since, by Fubini's theorem, the expectation of the integral converges to zero, and since $\sqrt{m}\|\hat{\mu}-\mu\| = O_P(1)$ because of (2.2) and (2.1). Combining this result with (2.6) and using the representation (2.2), we obtain

$$Z_{m,n}(t) \ =\ \sqrt{\frac{n}{N}}\ W_m^{(1)}(t) \ -\ \sqrt{\frac{m}{N}}\ W_n^{(2)}(t) + o_P(1), \tag{2.8}$$

where

$$W_m^{(1)}(t) \ =\ \sum_{j=1}^{m} W_{m,j}^{(1)}(t), \quad W_n^{(2)}(t) \ =\ \sum_{j=1}^{n} W_{n,j}^{(2)}(t),$$

and

$$W^{(1)}_{m,j}(t) = \frac{1}{\sqrt{m}}\Big(\cos(t'(X_j+\mu)) + \sin(t'(X_j+\mu)) - \Psi(t) - \bar{\Psi}(t)t'l(X_j+\mu)\Big),$$
$$W^{(2)}_{m,j}(t) = \frac{1}{\sqrt{n}}\Big(\cos(t'Y_j) + \sin(t'Y_j) - \Psi(t) - \bar{\Psi}(t)t'l(Y_j)\Big).$$

By a standard central limit theorem for i.i.d. random elements in Hilbert spaces, $W^{(1)}_m$ converges to some centered Gaussian process $W^{(1)}$ on $\mathcal{L}^2$ with covariance kernel

$$c(s,t) = E\Big[\Big(\cos(s'Y) + \sin(s'Y) - \Psi(s) - \bar{\Psi}(s)s'l(Y)\Big)\times \\ \times\Big(\cos(t'Y) + \sin(t'Y) - \Psi(t) - \bar{\Psi}(t)t'l(Y)\Big)\Big] \tag{2.9}$$

(recall $X+\mu \sim Y$ under H_0). By independence of the two samples, $W^{(2)}_n$ converges in distribution to some independent copy $W^{(2)}$ of $W^{(1)}$. Since, for constants a, b satisfying $a^2+b^2=1$, the (centered) process $W = aW^{(1)} + bW^{(2)}$ has the covariance kernel (2.9), and since the constants $a_{m,n} = (n/N)^{1/2}$, $b_{m,n} = (m/N)^{1/2}$ figuring in (2.8) satisfy $a^2_{m,n} + b^2_{m,n} = 1$ and converge to $1-p$ and p, respectively (recall the limiting regime (2.1)), $Z_{m,n} \Longrightarrow W$ under H_0. ■

The results obtained so far may be summarized as follows:

Theorem 2.1 *Assume that (1.1) holds, and that the estimator $\hat{\mu}$ allows for a representation of the form (2.2), where $El(Y) = 0$ and $E\|Y\|^2 < \infty$. Then, under the limiting regime (2.1),*

$$\frac{mn}{m+n}\, L_{m,n,\beta} \quad \Longrightarrow \quad \int_{\mathbb{R}^d} W^2(t)\, w_\beta(t)\, dt,$$

where $W(\cdot)$ is a centered Gaussian process on $\mathcal{L}^2$ having covariance kernel (2.9).

We now consider the problem of consistency of the test that rejects (1.1) for large values of $L_{m,n,\beta}$.

Theorem 2.2 *Assume that (1.1) does not hold. If, under the usual limiting regime,*

$$\hat{\mu}_{m,n} \longrightarrow a \quad \textit{almost surely} \tag{2.10}$$

for some $a \in \mathbb{R}^d$, then

$$\frac{mn}{m+n}\, L_{m,n,\beta} \longrightarrow \infty$$

almost surely. Thus, a test that rejects (1.1) for large values of $L_{m,n,\beta}$ is consistent against such alternatives.

PROOF. Let

$$\Delta \;=\; \inf_{\mu\in\mathbb{R}^d} \int_{\mathbb{R}^d} |\varphi(t)\exp(it'\mu) - \psi(t)|^2 w_\beta(t)\,dt.$$

Use compactness and tightness arguments to show that $\Delta > 0$. Let $m(s)$, $n(s)$, $s \geq 1$ be increasing sequences of integers that follow the usual limiting regime. By Fatou's lemma,

$$\liminf_{s\to\infty} L_{m(s),n(s),\beta} \geq \int_{\mathbb{R}^d} \liminf_{s\to\infty} |\varphi_{m(s)}(t)\exp(it'\hat{\mu}_{m(s),n(s)}) - \psi_{n(s)}(t)|^2 w_\beta(t)\,dt.$$

¿From (2.10), the integrand converges almost surely to $|\varphi(t)\exp(it'a) - \psi(t)|^2$ whence, by dominated convergence,

$$\begin{aligned}\liminf_{s\to\infty} L_{m(s),n(s),\beta} &\geq \int_{\mathbb{R}^d} |\varphi(t)\exp(it'a) - \psi(t)|^2 w_\beta(t)\,dt \\ &\geq \Delta > 0,\end{aligned}$$

proving the assertion. ■

3 The cases $\beta \to 0$ and $\beta \to \infty$

This section sheds more light on the role of the weight function w_β figuring in (1.5). We will show that the test statistic $L_{m,n,\beta}$, when suitably transformed, converges to some limit statistic as $\beta \to 0$ or $\beta \to \infty$. Notice that, in view of (1.7) and (1.8), these cases correspond to 'infinite' and 'zero' smoothing, respectively (see [8] for a similar observation in connection with testing for multivariate normality). Thus, the class of tests based on $L_{m,n,\beta}$, $\beta > 0$, is 'closed at the boundaries' $\beta \to 0$ and $\beta \to \infty$. To state the first result, let

$$S_m \;=\; \frac{1}{m}\sum_{j=1}^{m}(X_j - \bar{X}_m)(X_j - \bar{X}_m)', \quad T_n \;=\; \frac{1}{n}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)(Y_k - \bar{Y}_n)'$$

be the sample covariance matrices of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively. The trace of a square matrix A will be denoted by $tr(A)$.

Proposition 3.1 *If $\hat{\mu} = \bar{Y}_n - \bar{X}_m$, then*

$$\lim_{\beta \to 0} \beta^{-4} L_{m,n,\beta} \;=\; \frac{1}{4}\left(\{tr(S_m - T_n)\}^2 \;+\; 2tr(S_m - T_n)^2\right). \tag{3.1}$$

PROOF. An expansion of the exponential terms in (1.6) yields

$$L_{m,n,\beta} \;=\; -\frac{\beta^2}{2} A_1 \;+\; \frac{\beta^4}{8} A_2 \;+\; O(\beta^6) \;\text{ as }\; \beta \to 0,$$

where

$$A_1 = \frac{1}{m^2}\sum_{j,k=1}^{m} \|X_j - X_k\|^2 + \frac{1}{n^2}\sum_{j,k=1}^{n} \|Y_j - Y_k\|^2 - \frac{2}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} \|X_j + \hat{\mu} - Y_k\|^2,$$

$$A_2 = \frac{1}{m^2}\sum_{j,k=1}^{m} \|X_j - X_k\|^4 + \frac{1}{n^2}\sum_{j,k=1}^{n} \|Y_j - Y_k\|^4 - \frac{2}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n} \|X_j + \hat{\mu} - Y_k\|^4.$$

By replacing X_j, X_k and Y_j, Y_k in the first two sums occuring in the definition of A_1 by their 'centered values' $X_j - \bar{X}_m$, $X_k - \bar{X}_m$, $Y_j - \bar{Y}_n$, $Y_k - \bar{Y}_n$, respectively, and then using

$$\sum_{j=1}^{m}(X_j - \bar{X}_m) = 0, \qquad \sum_{k=1}^{n}(Y_k - \bar{Y}_n) = 0, \tag{3.2}$$

it is readily seen that $A_1 = 0$. To tackle A_2, apply the centering as above to $\|X_j - X_k\|^4$ and $\|Y_j - Y_k\|^4$ and then use (3.2) and the fact that the operation $tr(\cdot)$ is a linear functional on the set of square matrices of given order satisfying $tr(AB) = tr(BA)$. For example, we have

$$\begin{aligned}
&\frac{1}{m^2}\sum_{j,k=1}^{m}\|X_j-\bar{X}_m-(X_k-\bar{X}_m)\|^4\\
&=\frac{1}{m^2}\sum_{j,k=1}^{m}\Big\{\|X_j-\bar{X}_m\|^4+2\|X_j-\bar{X}_m\|^2\|X_k-\bar{X}_m\|^2+\\
&\qquad\qquad+\|X_k-\bar{X}_m\|^4+4\Big[(X_j-\bar{X}_m)'(X_k-\bar{X}_m)\Big]^2\Big\}\\
&=\frac{1}{m^2}\left(2m\sum_{j=1}^{m}\|X_j-\bar{X}_m\|^4+2\left[\sum_{j=1}^{m}\|X_j-\bar{X}_m\|^2\right]^2+\right.\\
&\qquad\qquad\left.+4\sum_{j,k=1}^{m}\Big[(X_j-\bar{X}_m)'(X_k-\bar{X}_m)\Big]^2\right),
\end{aligned}$$

$$\begin{aligned}
\sum_{j,k=1}^{m}\Big[(X_j-\bar{X}_m)'(X_k-\bar{X}_m)\Big]^2 &=\sum_{j=1}^{m}(X_j-\bar{X}_m)'mS_m(X_j-\bar{X}_m)\\
&=m\sum_{j=1}^{m}tr\left((X_j-\bar{X}_m)'S_m(X_j-\bar{X}_m)\right)\\
&=m\sum_{j=1}^{m}tr\left(S_m(X_j-\bar{X}_m)(X_j-\bar{X}_m)'\right)\\
&=m\ tr\left(S_m\sum_{j=1}^{m}(X_j-\bar{X}_m)(X_j-\bar{X}_m)'\right)\\
&=m^2\ tr\left(S_m^2\right)
\end{aligned}$$

etc. The details are omitted. ■

REMARK. Proposition 3.1 shows that, as $\beta \to 0$, $L_{m,n,\beta}$ degenerates to a functional of the difference of sample covariance matrices. Notice that the right-hand side of (3.1) is always nonnegative, and is zero if, and only if, $S_m = T_n$. Thus, in the limit $\beta \to 0$, $L_{m,n,\beta}$ provides a test for the equality of covariance matrices. Interestingly, the time-honored normal theory test for covariance matrices uses a completely different criterion, namely, apart from a factor, $|S_m|^{(m-1)/2}|T_n|^{(n-1)/2}/|mS_m+nT_n|^{(N-2)/2}$ (see e.g. [12], p. 526). Here, $|A|$ stands for the determinant of a square matrix A. Other test statistics for testing the equality of two covariance matrices can be found in [14], Sec. 8.2.8.

We state the next result only for the balanced case $m = n$; the expression looks more complicated in the general case.

Proposition 3.2 *If $m = n$, we have*

$$\lim_{\beta\to\infty}\left(-\frac{2}{\beta^2}\log\left\{\frac{m^2}{2}\left|L_{m,m,\beta}-\frac{2}{m}\right|\right\}\right)=M,$$

where

$$M \;=\; \min\left(\min_{j<k}\|X_j-X_k\|^2,\min_{j<k}\|Y_j-Y_k\|^2,\min_{j,k}\|Y_k-X_j-\hat{\mu}\|^2\right).$$

Proof. The proof follows readily upon noting that

$$\begin{aligned}L_{m,n,\beta}-\frac{1}{m}-\frac{1}{n}=&\frac{2}{m^2}\sum_{j<k}\exp\left(-\frac{\beta^2}{2}||X_j-X_k\|^2\right)+\frac{2}{n^2}\sum_{j<k}\exp\left(-\frac{\beta^2}{2}||Y_j-Y_k\|^2\right)\\&-\frac{2}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n}\exp\left(-\frac{\beta^2}{2}\|X_j+\hat{\mu}-Y_k\|^2\right)\end{aligned}$$

and using the fact that, if $0 < a_1 < a_2 < \ldots < a_r$ and $b > 0$, then

$$\sum_{j=1}^{r}\exp\left(-ba_j\right) \;=\; \exp\left(-ba_1\right)\left(1+o(1)\right) \;\text{ as } b\to\infty. \;\blacksquare$$

Remark. The statistic M in Proposition 3.2 compares the minimum interpoint distance within the two samples with the interpoint distance of the pooled sample. The resulting test is not consistent against all alternatives. The same holds for a test based on the limiting statistic in Proposition 3.1.

4 Permutation– and Bootstrap–Tests

To perform the test of (1.1) based on $L_{m,n,\beta}$, we suggest the use of resampling procedures. One possibility is the permutation test procedure, which works as follows. Pool the values $X_1+\hat{\mu},\ldots,X_m+\hat{\mu},Y_1,\ldots,Y_n$ into a sample of size N. Then randomly divide the pooled sample into two subsamples such that one has size m and the other has size n. This is just a random permutation of the pooled sample. Denote the first sample by $Z_j^{(1)}$, $j=1,\ldots,m$, and the second by $Z_j^{(2)}$, $j=1,\ldots,n$. We estimate μ by $\hat{\mu}_p=\hat{\mu}(Z_1^{(1)},\ldots,Z_m^{(1)},Z_1^{(2)},\ldots,Z_n^{(2)})$. Putting

$$L^p_{m,n,\beta}=L_{m,n,\beta}(Z_1^{(1)},\ldots,Z_m^{(1)},Z_1^{(2)},\ldots,Z_n^{(2)}),$$

we have

$$\frac{mn}{N}L^p_{m,n,\beta}=\int\limits_{\mathbb{R}^d}(W^p_{m,n}(t))^2 w_\beta(t)dt,$$

where

$$W^p_{m,n}(t)=\sqrt{\frac{n}{N}}U^p_m(t)-\sqrt{\frac{m}{N}}V^p_n(t), \tag{4.3}$$

and

$$\begin{aligned}
U^p_m(t)&=\frac{1}{\sqrt{m}}\sum_{j=1}^m\left[\cos\left(t'\left(Z_j^{(1)}+\hat{\mu}_p\right)\right)+\sin\left(t'\left(Z_j^{(1)}+\hat{\mu}_p\right)\right)-\psi_N^{(p)}(t)\right],\\
V^p_m(t)&=\frac{1}{\sqrt{n}}\sum_{j=1}^n\left[\cos\left(t'Z_j^{(2)}\right)+\sin\left(t'Z_j^{(2)}\right)-\psi_N^{(p)}t\right],\\
\psi_N^{(p)}(t)&=\frac{1}{N}\sum_{a\in\mathcal{P}_{m,n}}\left(\cos(t'a)+\sin(t'a)\right)\ .
\end{aligned}$$

Here and in what follows, $\mathcal{P}_{m,n}=\{Z_1^{(1)},\ldots,Z_m^{(1)},Z_1^{(2)},\ldots,Z_n^{(2)}\}$ $(=\{X_1+\hat{\mu},\ldots,X_m+\hat{\mu},Y_1,\ldots,Y_n\})$ denotes the pooled sample.

Theorem 4.1 *Assume that the conditions of Theorem 2.1 hold. Then for almost all sequences* $\{X_1,\ldots,X_m,\ldots\}$ *and* $\{Y_1,\ldots,Y_n,\ldots\}$,

$$\frac{mn}{N}L^p_{m,n,\beta}\implies\int\limits_{\mathbb{R}^d}W^2(t)w_\beta(t)dt,$$

under the limiting regime (2.1), where $W(\cdot)$ *is the Gaussian process figuring in the statement of Theorem 2.1.*

PROOF. For fixed θ, let $\mathcal{P}_{m,n}(\theta) \equiv \{X_1+\theta,\ldots,X_m+\theta,Y_1,\ldots,Y_n\}$, and let $Z_{j_\nu}^{(\nu)}(\theta)$ ($\nu=1,2; j_1=1,\ldots,m; j_2=1,\ldots,n$) be a random permutation of $\mathcal{P}_{m,n}(\theta)$. Write

$$F_m(\theta,t) \;=\; \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\{X_j+\theta\le t\},$$

$$F_m^p(\theta,t) \;=\; \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}\{Z_j^{(1)}(\theta)\le t\}$$

for the empirical distribution functions of $X_1+\theta,\ldots,X_m+\theta$ and $Z_1^{(1)}(\theta),\ldots,Z_m^{(1)}(\theta)$, respectively. Likewise, let $G_n(\cdot)$ and $G_n^p(\theta,\cdot)$ denote the empirical distribution functions of $Y_1,\ldots,Y_n$ and $Z_1^{(2)}(\theta),\ldots,Z_n^{(2)}(\theta)$, respectively. Finally, write

$$H_N(\theta,t)=\frac{m}{N}F_m(\theta,t)+\frac{n}{N}G_n(t)$$

for the empirical distribution of the pooled sample $\mathcal{P}_{m,n}(\theta)$.

Notice that $H_N(\theta,t)=\frac{m}{N}F_m^p(\theta,t)+\frac{n}{N}G_n^p(\theta,t)$ and thus

$$\sqrt{\frac{nm}{N}}\left(F_m^p(\theta,t)-G_n^p(\theta,t)\right)=\sqrt{\frac{mN}{n}}\left(F_m^p(\theta,t)-H_N(\theta,t)\right)\ .$$

For a fixed constant $C>0$, let RV_N denote the centered process

$$RV_N(\theta,t) \;=\; \sqrt{\frac{mN}{n}}\left(F_m^p(\theta,t)-H_N(\theta,t)\right) \qquad (\|\theta-\mu\|\le C, t\in\mathbb{R}^d).$$

Applying Theorem 1 of [15], p. 309, for almost all sequences $\{X_1,\ldots,X_m,\ldots\}$ and $\{Y_1,\ldots,Y_n,\ldots\}$, RV_N converges in distribution to a H-Brownian bridge RV_H, where $H(\theta,t)=\lim_{m,n\to\infty}H_N(\theta,t)$ (for a definition, see [15], p. 308).

The convergence is convergence in distribution in $l^\infty(\mathcal{F})$, the space consisting of bounded, real-valued functions defined on the class $\mathcal{F}$ of indicator functions $\mathbf{1}\{\cdot\le t\}$, $t\in\mathbb{R}^d$. Note that when $\theta=\mu$ then $F_m(\mu,\cdot)$ and $G_n(\cdot)$ converge in distribution to the same limit, which is the distribution function $H(\mu,\cdot)$ of Y.

Furthermore, when $C = O(1/\sqrt{N})$, we have that $\{RV_N(\theta,t) - RV_N(\mu,t) : \|\theta - \mu\| \leq C, t \in \mathbb{R}^d\}$ converges in distribution to zero. This can be verified by computing the variance function which is asymptotically zero at rate $O(1/\sqrt{N})$. Note that the distance between $X_j - \hat{\mu}_p$ and $X_j - \mu$ equals $\hat{\mu}_p - \mu = O_P(1/\sqrt{N})$. Therefore $\{RV_N(\hat{\mu}_p,t) = \sqrt{\frac{mN}{n}}\left(F_m^p(\hat{\mu}_p,t) - H_N(\hat{\mu}_p,t)\right), t \in \mathbb{R}^d\}$ has the same limit as

$$\{RV_N(\mu,t) = \sqrt{\frac{mN}{n}}\left(F_m^p(\mu,t) - H_N(\mu,t)\right), t \in \mathbb{R}^d\}.$$

Based on this result, we can now derive the convergence of the test statistic. Write $W_{m,n}^p(t)$ figuring in (4.3) as a stochastic integral according to

$$\begin{aligned} W_{m,n}^p(t) &= \sqrt{\frac{mn}{N}}\Big(\int \left(\cos\left(t'(\bullet + \hat{\mu}_p)\right) + \sin\left(t(\bullet + \hat{\mu}_p)\right)\right) dF_m^p(\hat{\mu}_p,t) \\ &\qquad - \int \left(\cos(t'\bullet) + \sin(t'\bullet)\right) dG_n^p(\hat{\mu}_p,t)\Big) \\ &= \sqrt{\frac{mn}{N}}\int \left(\cos(t'\bullet) + \sin(t'\bullet) - \bar{\psi}(t)t'l(\bullet)\right) d\left(F_m^p(\hat{\mu}_p,t) - G_n^p(\hat{\mu}_p,t)\right) \\ &\quad + o_P(1) \\ &= \int \left(\cos(t'\bullet) + \sin(t'\bullet) - \bar{\psi}(t)t'l(\bullet)\right) d\left\{\sqrt{\frac{mN}{n}}\left(F_m^p(\hat{\mu}_p,t) - H_N(\hat{\mu}_p,t)\right)\right\} \\ &\quad + o_P(1)\,. \end{aligned}$$

The term $o_p(1)$ in the second equation can be obtained by similar arguments as between (2.5) and (2.8) for proving Theorem 2.1. $W_{m,n}^p(t)$ converges to

$$W^p(t) = \int \left(\cos(t'\bullet) + \sin(t'\bullet) - \bar{\psi}(t)\,t'l(\bullet)\right)\, d\,RV_H(\mu,t)$$

in the Skorohod space $D[-\infty,\infty]$. It is easy to see that the covariance function of W^p is identical with that of W. Therefore W^p has the same distribution as W. The convergence of $\frac{mn}{N}L_{m,n,\beta}^p$ is a direct consequence (see the proof of Theorem 2.2 in [9] for a similar case). The proof is completed. ■

According to Theorem 4.1, a permutation test based on $L_{m,n,\beta}$ (to be definded in Section 5) works for large samples. As an alternative to the permutation procedure, it is possible to resample from the combined sample $X_1 + \hat{\mu}, \ldots, X_m + \hat{\mu}, Y_1, \ldots, Y_n$

with replacement. To show the validity of this bootstrap procedure, one can proceed similar as in the proof of Theorem 4.1, replacing Theorem 1 in [15] by Theorem 3 in [15]. Indicating bootstrap quantities by an upper index b, the theorem yields that $\{\sqrt{m}\left(F_m^b(\theta,t) - H_N(\theta,t)\right) : \|\theta-\mu\| \le C, t \in \mathbb{R}^d\}$ converges in distribution to RV_H for almost all sequences $\{X_1,\ldots,X_m,\ldots,\}$ and $\{Y_1,\ldots,Y_n,\ldots\}$. Hence,

$$\sqrt{\frac{mn}{N}}(F_m^b - G_n^b) = \sqrt{\frac{n}{N}}\sqrt{m}(F_m^b - H_N) - \sqrt{\frac{m}{N}}\sqrt{n}(G_n^b - H_N)$$

converges in distribution to $\sqrt{1-p}\,RV_H - \sqrt{p}\,RV_H'$, where RV_H' is an independent copy of RV_H. As a consequence, $W_{m,n}^b(t)$ converges to

$$W^b(t) = \int \left(\cos(t'\bullet) + \sin(t'\bullet) - \bar{\psi}(t)\,t'l(\bullet)\right)\, d\left(\sqrt{1-p}RV_H(\mu,t) - \sqrt{p}RV_H'(\mu,t)\right)$$

in $D[-\infty,\infty]$ which shows that a statement analogous to Theorem 4.1 holds for the bootstrap procedure.

5 Simulation results

To assess the actual level of the tests for the location shift model based on $L_{m,n,\beta}$, a simulation study was performed for sample sizes $N = 40$ $(m = n = 20)$ and $N = 80$ $(m = n = 40)$ and dimensions $d = 2$ and $d = 5$. As estimator of μ we used $\hat{\mu} = \bar{y}_n - \bar{x}_m$. Besides $L_{m,n,0}$, $L_{m,n,0.5}$, $L_{m,n,1}$ and $L_{m,n,2}$, we included Bartlett's modified likelihood ratio test statistic

$$\Lambda_{m,n} = \frac{m}{m-1}|S_m|^{(m-1)/2} \cdot \frac{n}{n-1}|T_n|^{(n-1)/2} \Big/ \frac{1}{N-2}|mS_m + nT_n|^{(N-2)/2}.$$

The validity of the pooled bootstrap procedure for $\Lambda_{m,n}$ was proved in [17], [18].

We used the following distributions:

- MN_1: the d-variate standard Normal distribution $\mathcal{N}(0, I_d)$
- MN_2: the d-variate Normal distribution $\mathcal{N}(0, \Sigma_d^1)$, where $\Sigma_2^1 = diag(2,4)$ and $\Sigma_5^1 = diag(2,2,2,4,4)$;

- MN_3: a d-variate normal distribution with mean zero, unit variances and equal correlation $\rho = 0.5$ between components; the covariance matrix is denoted by Σ_d^2;
- MT_1: the multivariate t distribution with 5 degrees of freedom $t_d(5; 0, I_d)$, generated as $U/\sqrt{V}$, where U and V are independent, and $U \sim \mathcal{N}(0, I_d)$, $V \sim \chi_5^2/5$;
- MT_2, MT_3: multivariate t distribution $t_d(5; 0, \Sigma_5^1)$ and $t_d(5; 0, \Sigma_5^2)$;
- CN: a contaminated normal distribution, where each component is a $\mathcal{N}(0, 1)$ random variate with probability 0.9 and a χ_2^2 variate with probability 0.1, and where the components are independent.

For each fixed combination of N, d and the underlying distribution as given above, the following procedure was replicated 5 000 times:

(1) generate random samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$
(2) compute $L_{m,n,\beta}(x_1, \ldots, x_m, y_1, \ldots, y_n)$
(3) compute $x'_j = x_j + \hat{\mu}$, $j = 1, \ldots, m$
(4) draw 500 samples with (without) replacement from the pooled sample $x'_1, \ldots, x'_m$, $y_1, \ldots, y_n$
(5) calculate the corresponding 500 realizations $L^B_{m,n,\beta}(j)\,(L^P_{m,n,\beta}(j)),\ 1 \leq j \leq 500$, (say) of the bootstrap (permutation) statistic $L^B_{m,n,\beta}$ $(L^P_{m,n,\beta})$
(6) reject H_0 if $L_{m,n,\beta}$, computed on $x_1, \ldots, x_m, y_1, \ldots, y_n$, exceeds the empirical 95%-quantile of $L^B_{m,n,\beta}(j)\,(L^P_{m,n,\beta}(j)),\ 1 \leq j \leq 500$.

Table 1 and Table 2 show the percentage of the number of rejections of H_0 for the bootstrap and the permutation procedure, respectively. Obviously, the bootstrap procedure is conservative to a greater or lesser extent; it performs worse with increasing values of the dimension d. In contrast, the actual level of the permutation procedure is more or less above the nominal level of 5%. In the majority of cases, the percentage of the number of rejections is less than 8% for $n = 20$ and less than 7% for $n = 40$. For larger values of β, the bootstrap procedure for $L_{m,n,\beta}$ breaks down. This effect might have been anticipated from Proposition 3.2 since, for large values of β, $L_{m,n,\beta}$ is approximately a minimum of random variables. Notice, however, that the permutation procedure works well also for larger values of β. The observations concerning the bootstrap procedure for $\Lambda_{m,n}$ are in agreement with the findings of Zhang and Boos [17].

			MN_1	MN_2	MN_3	MT_1	MT_2	MT_3	CN
$\Lambda_{m,n}$	$n=20$	$d=2$	4.3	3.8	3.9	4.0	4.3	4.6	4.1
		$d=5$	1.3	2.0	1.7	2.2	1.9	2.1	2.8
	$n=40$	$d=2$	5.3	4.6	4.7	4.3	4.4	4.9	4.9
		$d=5$	3.1	3.2	3.3	3.3	3.2	3.4	4.8
$L_{m,n,0}$	$n=20$	$d=2$	4.8	6.3	5.2	3.8	4.0	4.3	3.2
		$d=5$	3.7	4.1	5.0	3.0	2.3	4.1	1.6
	$n=40$	$d=2$	5.1	5.2	5.4	3.7	3.8	4.3	3.3
		$d=5$	3.9	4.4	5.5	2.5	2.3	3.6	2.3
$L_{m,n,0.5}$	$n=20$	$d=2$	4.9	4.0	4.6	3.7	3.4	4.0	2.7
		$d=5$	2.2	0.9	3.7	1.6	0.6	2.6	1.5
	$n=40$	$d=2$	4.9	4.8	4.8	3.4	3.8	3.3	3.1
		$d=5$	3.1	1.5	4.2	2.1	1.4	2.9	2.0
$L_{m,n,1}$	$n=20$	$d=2$	3.6	1.7	3.8	2.9	1.5	3.9	2.7
		$d=5$	0.2	0.0	0.7	0.1	0.0	0.5	0.0
	$n=40$	$d=2$	4.0	3.0	4.5	3.5	2.5	3.7	3.3
		$d=5$	0.4	0.0	1.9	0.7	0.0	1.1	0.3
$L_{m,n,2}$	$n=20$	$d=2$	1.4	0.1	1.9	1.0	0.0	1.1	0.0
		$d=5$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$n=40$	$d=2$	2.1	0.4	2.4	1.2	0.1	2.0	1.3
		$d=5$	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1

Estimated level for the bootstrap test (nominal level: 5%)

			MN_1	MN_2	MN_3	MT_1	MT_2	MT_3	CN
$\Lambda_{m,n}$	$n=20$	$d=2$	4.7	5.1	5.2	5.3	6.0	5.5	6.2
		$d=5$	5.3	4.6	5.1	6.5	6.5	5.8	7.7
	$n=40$	$d=2$	5.3	5.4	5.2	4.9	5.5	5.9	6.2
		$d=5$	5.2	5.7	5.3	5.7	6.0	5.7	5.6
$L_{m,n,0}$	$n=20$	$d=2$	6.1	6.3	5.3	6.8	6.7	5.9	8.2
		$d=5$	6.8	6.2	5.5	7.8	6.8	6.1	9.2
	$n=40$	$d=2$	5.3	6.2	5.5	5.6	6.2	5.6	7.6
		$d=5$	5.8	6.3	5.7	6.0	6.2	61	6.5
$L_{m,n,0.5}$	$n=20$	$d=2$	6.2	5.5	6.4	7.8	6.9	7.2	7.7
		$d=5$	6.4	5.7	5.8	7.9	7.2	6.6	7.5
	$n=40$	$d=2$	5.9	5.0	5.4	6.4	5.4	5.3	6.4
		$d=5$	5.5	5.1	5.3	6.1	5.5	6.4	5.9
$L_{m,n,1}$	$n=20$	$d=2$	5.5	5.9	5.7	6.1	7.2	6.4	5.9
		$d=5$	6.0	6.2	6.1	7.1	8.1	6.5	6.0
	$n=40$	$d=2$	5.1	5.1	4.8	5.4	5.5	5.6	6.5
		$d=5$	5.2	5.4	5.2	6.1	5.6	5.6	6.1
$L_{m,n,2}$	$n=20$	$d=2$	5.4	6.6	5.6	7.0	7.7	6.3	6.5
		$d=5$	6.2	8.1	6.3	8.6	9.9	7.8	7.4
	$n=40$	$d=2$	5.0	5.7	5.3	5.3	6.5	6.2	5.7
		$d=5$	5.3	7.5	5.5	7.1	9.0	6.1	6.3

Table 2

Estimated level for the permutation test (nominal level: 5%)

To assess the power of the different tests, we simulated data from the following distributions:

- A_1: MN_1 against MN_2;
- A_2: MN_1 against MN_3;
- A_3: MT_1 against MT_2;
- A_4: MT_1 against MT_3;
- A_5: MN_1 against $\sqrt{0.6}\,MT_1$;
- A_6: MN_1 against $CN/\sqrt{0.85}$;
- A_7: $\sqrt{0.6}\,MT_1$ against $CN/\sqrt{0.85}$.

The covariance of MT_1 is $5/3\,I_d$; hence, the covariances of both distributions of A_5 coincide. (But A_5 does not satisfy (2) in [17]; i.e., H_0 does not hold even for $\Lambda_{m,n}$). The same remark applies to A_6 and A_7.

Table 3 and Table 4 show the percentages of rejection of H_0. An asterisk denotes power 100%. The main conclusions that can be drawn from the power study are the following:

(1) The bootstrap and the permutation tests behave similar in all cases in which the bootstrap procedure maintains its nominal level. However, for small n and $d = 5$, the bootstrap loses power. For larger values of β, the bootstrap tests based on $L_{m,n,\beta}$ breaks down, as expected from the results of Table 1.

(2) For alternatives with mere covariance differences, the tests based on $\Lambda_{m,n}$ and $L_{m,n,0}$ outperform the other tests. In particular, $L_{m,n,0}$ performs best for the scale alternatives A_1 and A_3, whereas $\Lambda_{m,n}$ is better for A_2 and A_4.

(3) The permutation test based on $L_{m,n,1}$ dominates all other tests in the remaining cases where the covariance matrix is the identity. The power of $\Lambda_{m,n}$ and $L_{m,n,0}$ does not increase with larger sample size.

(4) Over the whole range of alternatives considered, the permutation test based on $L_{m,n,0.5}$ seems to be a good compromise.

			A_1	A_2	A_3	A_4	A_5	A_6	A_7
$\Lambda_{m,n}$	$n=20$	$d=2$	67.1	21.8	47.4	16.8	5.4	6.9	4.3
		$d=5$	47.4	32.5	36.5	25.1	2.3	3.6	2.4
	$n=40$	$d=2$	97.6	51.5	80.2	31.0	7.1	8.9	5.2
		$d=5$	99.5	92.0	86.1	71.0	4.9	5.3	3.6
$L_{m,n,0}$	$n=20$	$d=2$	82.4	13.2	47.7	7.4	7.1	10.0	4.0
		$d=5$	98.3	41.7	56.4	13.2	7.0	8.5	2.1
	$n=40$	$d=2$	99.1	26.4	76.5	10.7	6.8	11.5	3.7
		$d=5$	$*$	88.1	80.8	30.2	6.9	7.7	2.0
$L_{m,n,0.5}$	$n=20$	$d=2$	74.0	12.0	43.2	6.4	7.1	10.2	3.7
		$d=5$	93.2	32.2	51.1	9.8	8.8	9.8	2.2
	$n=40$	$d=2$	98.1	24.5	80.5	10.6	9.6	15.7	4.3
		$d=5$	$*$	81.6	94.0	30.6	18.3	18.7	3.1
$L_{m,n,1}$	$n=20$	$d=2$	46.9	9.2	25.2	6.1	7.6	8.8	4.1
		$d=5$	9.9	4.1	1.6	1.0	4.4	2.7	0.9
	$n=40$	$d=2$	88.2	19.1	62.0	9.7	14.7	19.0	5.1
		$d=5$	83.4	34.6	34.2	9.3	20.4	15.4	2.8
$L_{m,n,2}$	$n=20$	$d=2$	8.0	2.7	3.3	1.3	3.7	3.6	1.8
		$d=5$	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$n=40$	$d=2$	42.8	7.5	19.3	4.2	8.2	10.3	3.6
		$d=5$	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 3

Estimated power for the bootstrap test (nominal level: 5%)

			A_1	A_2	A_3	A_4	A_5	A_6	A_7
$\Lambda_{m,n}$	$n = 20$	$d = 2$	71.0	24.8	52.2	18.5	6.7	7.8	6.3
		$d = 5$	68.1	51.9	52.3	43.3	7.1	6.8	6.2
	$n = 40$	$d = 2$	98.0	50.8	81.6	35.1	7.3	10.8	5.9
		$d = 5$	99.8	94.9	90.0	78.8	8.2	8.4	6.1
$L_{m,n,0}$	$n = 20$	$d = 2$	87.3	14.8	65.1	12.0	8.5	11.9	7.5
		$d = 5$	99.6	53.0	80.2	26.4	10.6	13.2	9.2
	$n = 40$	$d = 2$	99.5	27.2	86.9	15.6	7.2	13.0	7.3
		$d = 5$	$*$	91.3	93.7	47.8	9.7	11.0	6.6
$L_{m,n,0.5}$	$n = 20$	$d = 2$	81.3	15.4	56.1	12.2	9.5	13.0	8.5
		$d = 5$	97.8	49.7	77.0	26.0	17.2	18.5	9.5
	$n = 40$	$d = 2$	98.6	28.4	86.0	17.5	11.8	20.2	9.2
		$d = 5$	$*$	88.2	97.5	49.7	25.9	31.9	9.7
$L_{m,n,1}$	$n = 20$	$d = 2$	59.8	11.9	41.0	11.2	11.8	14.2	8.0
		$d = 5$	79.4	31.0	56.1	19.7	22.4	19.8	9.7
	$n = 40$	$d = 2$	91.7	21.7	72.9	14.4	16.5	23.3	8.7
		$d = 5$	98.9	64.7	86.9	36.8	41.2	34.1	10.3
$L_{m,n,2}$	$n = 20$	$d = 2$	36.6	9.4	28.6	9.3	10.5	10.9	7.1
		$d = 5$	38.6	15.9	32.9	15.6	16.7	13.3	8.9
	$n = 40$	$d = 2$	69.7	14.4	52.0	11.5	15.1	16.5	7.9
		$d = 5$	64.8	26.2	48.7	21.4	28.2	18.6	9.6

Table 4

Estimated power for the permutation test (nominal level: 5%)

References

[1] Anderson, N.H., Hall, P., and Titterington, D.M. Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multiv. Anal.* **50**, 41–54.

[2] Bowman, A.W., and Foster, P.J. (1993). Adaptive smoothing and testing for multivariate normality. *J. Amer. Statist. Ass.* **88**, 529–537.

[3] Csörgő, S. (1984). Testing by the empirical characterstic function: A survey. *Asymptotic Statistics, Proc. 3rd Prague Symp. 1983, Asymptotic Stat.*, **2**, 45 – 56.

[4] Fan, Y. (1997). Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. *J. Multiv. Anal.* **62**, 36 –63.

[5] Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* **14**, 604–621.

[6] Gürtler, N. (2000). Asymptotic results on the BHEP class of tests for multivariate normality with fixed and variable smoothing parameter (in German). *Doctoral Dissertation, University of Karlsruhe, Germany.*

[7] Gürtler, N., and Henze, N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.* **52**, 267 – 286.

[8] Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statist. & Probab. Lett.*, **35**, 203 – 213.

[9] Henze, N, and Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *J. Multiv. Anal.*, **62**, 1, 1–23.

[10] Henze, N., and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Commun. Statist. – Th. Meth.*, **19**, 3595 – 3617.

[11] Kellermeier, J. (1980). The empirical characteristic function and large sample hypothesis testing. *J. Multiv. Anal.* **10**, 78 – 87.

[12] Krishnaiah, P.R., and Lee, J.C. (1980). Likelihood ratio tests for mean vectors and covariance matrices. *In: Handbook of Statistics Vol.1 (P.R. Krishnaiah, ed.), North-Holland, Amsterdam, New York.*, 513–570.

[13] Kundu, S., Majumdar, S., and Mukherjee, K. (2000). Central limit theorems revisited. *Statist. & Probab. Lett.*, **47**, 265–275.

[14] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

[15] Praestgaard, J.T. (1995). Permutation and bootstrap Kolmogorov–Smirnov tests for the equality of two distributions, *Scand. J. Statist.* **22**, 305–322.

[16] Sen, P.K. (1984). An aligned goodness-of-fit test for the multivariate two-sample model: Locations unknown. *In: Colloquia Mathematica Societatis János Bolyai 45. Goodness-of-fit, Debrecen, Hungary*, 493–510.

[17] Zhang, J., and Boos, D.D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *J. Amer. Statist. Ass.* **87**, 425–429.

[18] Zhang, J., and Boos, D.D. (1993). Testing hypotheses about covariance matrices using bootstrap methods. *Commun. Statist. – Th. Meth.* **22**, 723–739.