JOURNAL OF MULTIVARIATE ANALYSIS **62**, 1–23 (1997)
ARTICLE NO. MV971684

A New Approach to the BHEP Tests for Multivariate Normality

Norbert Henze and Thorsten Wagner

Universität Karlsruhe, D-76128 Karlsruhe, Germany and
Ruhr-Universität Bochum, D-44780 Bochum, Germany

Let $X_1, ..., X_n$ be i.i.d. random d-vectors, $d \geqslant 1$, with sample mean $\bar{X}$ and sample covariance matrix S. For testing the hypothesis H_d that the law of X_1 is some non-degenerate normal distribution, there is a whole class of practicable affine invariant

$\exp(-\|t\|^2/2)$ under H_d. The test statistics have an alternative interpretation in terms of L^2-distances between a nonparametric kernel density estimator and the parametric density estimator under H_d, applied to $Y_1, ..., Y_n$. By working in the Fréchet space of continuous functions on $\mathbb{R}^d$, we obtain a new representation of the limiting null distributions of the test statistics and show that the tests have asymptotic power against sequences of contiguous alternatives converging to H_d at the rate $n^{-1/2}$, independent of d. © 1997 Academic Press

1. INTRODUCTION

There is a continued interest in the problem of testing for multivariate normality, as evidenced by the recent papers of Ahn [1], Bowman and Foster [5], Henze [12], Horswell and Looney [14], Kariya and George [16], Koziol [17], Mudholkar, McDermott, and Srivastava [20], Mudholkar, Srivastava, and Lin [21], Naito [22], Ozturk and Romeu [23], Rayner, Best, and Matthews [24], Romeu and Ozturk [25], Singh [26], Versluis [27], and Zhu, Wong, and Fang [29] on this subject.

The purpose of this paper is not to review the huge literature on tests for multivariate normality (for short; MVN tests), but to present a new approach to a whole class of MVN tests studied by Baringhaus and Henze

Received October 9, 1996; revised April 9, 1997.

AMS 1991 subject classification numbers: primary 62G10; secondary 62H15.

Key words and phrases: test for multivariate normality, empirical characteristic function, Gaussian process, contiguous alternatives.

0047-259X/97 $25.00

[3] and Henze and Zirkler [11]. These tests have the desirable properties of

- affine invariance,
- consistency against each fixed nonnormal alternative distribution,
- asymptotic power against contiguous alternatives of order $n^{-1/2}$,
- feasibility for any dimension and any sample size,

which no other MVN test shares, at least to our knowledge.

To state the testing problem and the tests under discussion, let $X_1, X_2, \ldots$ be a sequence of independent copies of a random d-dimensional column vector X, where $d \geqslant 1$ is a fixed integer. The distribution of X will be denoted by P^X. The problem is to test, on the basis of $X_1, \ldots, X_n$, the hypothesis

$$H_d \colon P^X \in \mathcal{N}_d,$$

where $\mathcal{N}_d$ is the class of all nondegenerate d-variate normal distributions. Since $\mathcal{N}_d$ is closed with respect to full rank affine transformations and the alternatives to H_d are rarely known or given in practice, we are interested in (affine) invariant and consistent tests. Such a test may be based on the test statistic

$$T_{n,\beta} := n(4 \quad \mathbf{1}\{S_n \text{ is singular}\} + W_{n,\beta} \quad \mathbf{1}\{S_n \text{ is nonsingular}\}). \tag{1.1}$$

Here, $\beta > 0$ is a parameter (the role of which is discussed later), $\mathbf{1}\{\cdot\}$ stands for the indicator function and

$$S_n := \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X}_n)(X_j - \bar{X}_n)'$$

is the empirical covariance matrix of $X_1, \ldots, X_n$, where $\bar{X}_n := n^{-1} \sum_{j=1}^{n} X_j$ and the prime denotes transpose. $W_{n,\beta}$ is the weighted L^2-distance

$$W_{n,\beta} := \int_{\mathbb{R}^d} \left| \Psi_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right)\right|^2 \varphi_\beta(t)\, dt \tag{1.2}$$

between the empirical characteristic function

$$\Psi_n(t) := \frac{1}{n} \sum_{k=1}^{n} \exp(it' Y_k) \qquad (i^2 = -1)$$

of the scaled residuals

$$Y_k := S_n^{-1/2}(X_k - \bar{X}_n) \qquad (k = 1, \ldots, n)$$

and the characteristic function $\exp(-\|t\|^2/2)$ of the standard d-variate normal distribution. $S_n^{-1/2}$ is the symmetric positive definite square root of S_n^{-1}. Note that Y_k and, hence, $W_{n,\beta}$ are only defined if S_n is nonsingular. According to (1.1), $W_{n,\beta}$ is replaced by its maximum possible value 4 in the case when S_n is not invertible (this suggestion is due to Csörgő [7]). The weight function φ_β figuring in (1.2) is

$$\varphi_\beta(t) := (2\pi\beta^2)^{-d/2} \exp\left(-\frac{\|t\|^2}{2\beta^2}\right).$$

The extremely appealing feature of this choice is that $W_{n,\beta}$ takes the simple form

$$W_{n,\beta} = \frac{1}{n^2} \sum_{j,k=1}^{n} \exp\left(-\frac{\beta^2}{2} \|Y_j - Y_k\|^2\right) - 2(1+\beta^2)^{-d/2} \frac{1}{n} \sum_{j=1}^{n} \exp\left(-\frac{\beta^2 \|Y_j\|^2}{2(1+\beta^2)}\right) + (1+2\beta^2)^{-d/2}.$$

This representation shows that $T_{n,\beta}$ is invariant; i.e., we have

$$T_{n,\beta}(AX_1 + b, ..., AX_n + b) = T_{n,\beta}(X_1, ..., X_n)$$

for each nonsingular $A \in \mathbb{R}^{d \times d}$ and each $b \in \mathbb{R}^d$. Moreover, since the computation of $\|Y_j - Y_k\|^2$ and $\|Y_j\|^2$ involves only S_n^{-1}, not even the square root $S_n^{-1/2}$ of S_n^{-1} is needed.

The statistic $W_{n,\beta}$ was proposed by Epps and Pulley [9] in the special case $d = 1$. Baringhaus and Henze [3] extended the approach of Epps and Pulley to the case $d > 1$ and obtained the limit distribution of $n \cdot W_{n,\beta}$ under H_d for the case $\beta = 1$. S. Csörgő [7] coined the term BHEP test with reference to these four authors.

Using a theorem of de Wet and Randles [8], Henze and Zirkler [11] proved that the limiting distribution of $T_{n,\beta}$ under H_d is that of $\sum_{j \geqslant 1} \delta_j(\beta) N_j^2$, where $N_1, N_2, ...$ are i.i.d. standard normal random variables and $(\delta_j(\beta))_{j \geqslant 1}$ is the sequence of eigenvalues associated with an integral operator given in Theorem 3.1 of Henze and Zirkler [11]. If the distribution of X is not in $\mathcal{N}_d$ we have

$$\liminf_{n \to \infty} n^{-1} T_{n,\beta} \geqslant C(P^X, \beta) > 0, \qquad \text{almost surely}$$

for some constant $C(P^X, \beta)$ (Csörgő [7]). This shows that rejecting H_d for large values of $T_{n,\beta}$ yields an affine invariant and universally consistent MVN test (cf. the first two bullets).

That the role of β figuring in the weight function φ_β is that of a smoothing parameter may be seen from the representation

$$W_{n,\beta} = (2\pi)^{d/2}\,\beta^{-d} \int_{\mathbb{R}^d} \left[g_{n,\beta}(x) - (2\pi\tau^2)^{-d/2} \exp\left(-\frac{\|x\|^2}{2\tau^2} \right) \right]^2 dx, \tag{1.3}$$

where $\tau^2 = (2\beta^2+1)/(2\beta^2)$, and

$$g_{n,\beta}(x) = \frac{1}{nh^d} \sum_{k=1}^{n} (2\pi)^{-d/2} \exp\left(-\frac{\|x - Y_k\|^2}{2h^2} \right)$$

is a nonparametric kernel density estimator, applied to $Y_1, ..., Y_n$, with a standard Gaussian kernel and bandwidth $h = 1/(\beta\sqrt{2})$ (see Henze and Zirkler [8, p. 3600]). In the spirit of density estimation, Bowman and Foster [4, p. 1535] proposed to base a MVN test on

$$BF_n := \beta_n^d (2\pi)^{-d/2}\, W_{n,\beta_n},$$

where $\beta_n := (h_n\sqrt{2})^{-1}$ and $h_n := [4/(n(d+2))]^{1/(d+4)}$. We conjecture that, in contrast to the case of fixed β (see Section 3), the MVN test based on BF_n is not able to discriminate between H_0 and alternatives which are $n^{-1/2}$-apart.

Interestingly, the class of MVN tests based on $T_{n,\beta}$ is "closed at the boundaries $\beta \to 0$ and $\beta \to \infty$" which correspond to "infinite and zero smoothing," respectively. More precisely, we have (Henze [13])

$$\lim_{\beta\to 0} \beta^{-6} W_{n,\beta} = \tfrac{1}{6} \cdot b_{1,d} + \tfrac{1}{4} \cdot \tilde{b}_{1,d}, \tag{1.4}$$

where

$$b_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} (Y_j' Y_k)^3$$

is Mardia's time-honored measure of multivariate sample skewness (Mardia [18]) and

$$\tilde{b}_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} Y_j' Y_k \,\|Y_j\|^2\, \|Y_k\|^2$$

is the sample version of the population skewness measure $\tilde{\beta}_{1,d} = \|E[\,\|X\|^2 X]\|^2$ (this notation assumes X to be standardized), introduced by Móri, Rohatgi, and Székely [19]. On the other hand, we have

$$\lim_{\beta\to\infty} \beta^d \left(W_{n,\beta} - \frac{1}{n} \right) = 2^{-d/2} - \frac{2}{n} \sum_{j=1}^{n} \exp\left(-\frac{\|Y_j\|^2}{2} \right) \tag{1.5}$$

which shows that, in the limit $\beta \to \infty$, rejection of H_d is equivalent to rejecting H_d for small values of $n^{-1} \sum_{j=1}^{n} \exp(-\frac{1}{2}\|Y_j\|^2)$. Note that this statistic is similar to Mardia's measure $n^{-1} \sum_{j=1}^{n} \|Y_j\|^4$ of multivariate kurtosis (Mardia [18]) in the sense that it investigates an aspect of the "radial part" of the standardized underlying distribution.

In view of the four properties featured at the beginning of this section, we stress that standard MVN tests based on multivariate skewness and kurtosis in the sense of Mardia lack the property of universal consistency (see Baringhaus and Henze [4] and Henze [12]).

The purpose of this paper is to provide a new approach to the class of BHEP tests. The reasoning utilizes the theory of weak convergence in the Fréchet space $C(\mathbb{R}^d)$ of continuous functions on $\mathbb{R}^d$ and is presented in Section 2. By means of this approach, we obtain

- a new representation of the limiting null distribution of $T_{n,\beta}$ in terms of a Gaussian process in $C(\mathbb{R}^d)$
- the joint limiting null distribution of $T_{n,\beta}$ for several values of β
- the asymptotic power of the test based on $T_{n,\beta}$ against contiguous alternatives.

Interestingly, the MVN test based on $T_{n,\beta}$ is able to detect contiguous alternatives which converge to the normal distribution at the rate $n^{-1/2}$ (see Section 3). The final section presents some empirical results and concluding discussion.

2. THE LIMIT DISTRIBUTION OF $T_{n,\beta}$ UNDER H_d

Throughout this section, we assume that the distribution of X is a centered d-variate normal with unit covariance matrix I_d, for short, $X \sim \mathcal{N}_d(\mathbf{0}, I_d)$. Since $T_{n,\beta}$ is affine invariant, this assumption means no loss of generality when studying the distribution of $T_{n,\beta}$ under H_d. Our starting point is the observation that for $n \geqslant d+1$,

$$T_{n,\beta} = \int_{\mathbb{R}^d} Z_n^2(t)\, \varphi_\beta(t)\, dt, \tag{2.1}$$

where

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[\cos(t' Y_j) + \sin(t' Y_j) - \exp\left(-\frac{\|t\|^2}{2} \right) \right], \tag{2.2}$$

$t \in \mathbb{R}^d$. Note that Z_n is a random element in the Fréchet space $C(\mathbb{R}^d)$ of continuous functions on $\mathbb{R}^d$, endowed with the metric

$$\rho(x, y) = \sum_{k=1}^{\infty} 2^{-k} \cdot \frac{\rho_k(x, y)}{1 + \rho_k(x, y)},$$

where

$$\rho_k(x, y) = \max_{\|t\| \leqslant k} |x(t) - y(t)|.$$

THEOREM 2.1. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\mathcal{N}_d(\mathbf{0}, I_d)$ distributed random d-vectors, and let Z_n be defined in* (2.2). *There exists a centered Gaussian process Z in $C(\mathbb{R}^d)$ having covariance kernel*

$$K(s, t) = \exp\left(-\frac{\|s-t\|^2}{2}\right) - \left\{1 + s't + \frac{(s't)^2}{2}\right\} \exp\left(-\frac{\|s\|^2 + \|t\|^2}{2}\right) \tag{2.3}$$

($s, t \in \mathbb{R}^d$) *such that*

$$Z_n \xrightarrow{\mathcal{D}} Z \qquad \text{in} \quad C(\mathbb{R}^d),$$

where "$\xrightarrow{\mathcal{D}}$" *denotes convergence in distribution.*

THEOREM 2.2. *Under the conditions of Theorem* 2.1, *we have*

$$\int_{\mathbb{R}^d} Z_n^2(t)\, \varphi_\beta(t)\, dt \xrightarrow{\mathcal{D}} \int_{\mathbb{R}^d} Z^2(t)\, \varphi_\beta(t)\, dt.$$

Note that this result is not trivial since the functional

$$x \mapsto |[x]|^2 := \int x^2(t)\, \varphi_\beta(t)\, dt \tag{2.4}$$

is not continuous on $C(\mathbb{R}^d)$ (it is not even defined on $C(\mathbb{R}^d)$ but only on the subset of square-integrable functions with respect to φ_β). Here and in what follows, an unspecified integral denotes integration over the whole space $\mathbb{R}^d$.

Remark. There seems to be a connection between $Z(t)$ and a limiting process $\tilde{Z}(t)$ obtained by Csörgő [6] in the context of MVN testing via the empirical characteristic function (see Theorem 2.2 of [6]). Both $Z(t)$ and $\tilde{Z}(t)$ have the property that their values are independent at orthogonal vectors. Whereas the process $\tilde{Z}$ is a simple transform of a complex-valued Gaussian process Y (see formula (2.5) of [6]), it is not clear whether the process Z is a transform of that same Y. If the underlying distribution is $\mathcal{N}_d(\mathbf{0}, I_d)$, the covariance kernel $\varrho(s,t)$ of Y given in (2.2) of [6] is related to the kernel $K(s,t)$ by the formula

$$K(s,t) = \varrho(s, -t) - (s't) \exp\left(-\frac{\|s\|^2 + \|t\|^2}{2}\right).$$

Proof of Theorem 2.1. Let

$$\begin{aligned} Z_n^*(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \Bigg[& \cos(t'X_j) + \sin(t'X_j) - \exp\left(-\frac{\|t\|^2}{2}\right) \\ & + \left\{\frac{1}{2}(t'X_j)^2 - \frac{\|t\|^2}{2} - t'X_j\right\} \exp\left(-\frac{\|t\|^2}{2}\right)\Bigg] \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \tilde{Z}_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \Bigg[& \cos(t'X_j) + \sin(t'X_j) - \exp\left(-\frac{\|t\|^2}{2}\right) \\ & + \{\cos(t'X_j) - \sin(t'X_j)\}\, t'\Delta_j \Bigg], \end{aligned} \tag{2.6}$$

where

$$\Delta_j = (S_n^{-1/2} - I_d)\, X_j - S_n^{-1/2} \bar{X}_n. \tag{2.7}$$

The main steps of the proof are to show that

$$Z_n^* \xrightarrow{\mathscr{D}} Z \qquad \text{in} \quad C(\mathbb{R}^d), \tag{2.8}$$

$$\rho(Z_n, \tilde{Z}_n) \xrightarrow{\mathrm{P}} 0, \tag{2.9}$$

and

$$\rho(\tilde{Z}_n, Z_n^*) \xrightarrow{\mathrm{P}} 0. \tag{2.10}$$

To prove (2.8), use elementary trigonometric identities and the formulae

$$\begin{aligned} E\cos(t'X) &= \exp(-\|t\|^2/2), \\ E\sin(t'X) &= 0, \\ E(s'Xt'X) &= s't, \\ E(t'X\sin(s'X)) &= s't\exp(-\|s\|^2/2), \\ E((s'X)^2\,(t'X)^2) &= \|s\|^2\,\|t\|^2 + 2(s't)^2, \\ E((s'X)^2\cos(t'X)) &= (\|s\|^2 - (s't)^2)\exp(-\|t\|^2/2) \end{aligned}$$

to obtain $EZ_n^*(t)=0$ and $E(Z_n^*(s)\,Z_n^*(t))=K(s,t)$, where $K(s,t)$ is given in (2.3). By the multivariate central limit theorem, the finite dimensional distributions of Z_n^* converge to centered multivariate normal distributions with covariances determined by the kernel K. To prove that the sequence $(Z_n^*)_{n\geqslant 1}$ is tight, it suffices to show that for each $k\geqslant 1$ the sequence Z_n^*, restricted to $\mathbb{R}_k^d=\{t\in\mathbb{R}^d\colon \|t\|\leqslant k\}$, is tight in the Banach space $C(\mathbb{R}_k^d)$ (adapt the reasoning given in Karatzas and Shreve [15, p. 62f.], to the present case). To this end, let $N(\mathbb{R}_k^d,\xi)$ denote the smallest number m such that $\mathbb{R}_k^d$ can be covered with m spheres of radius ξ. Since

$$N(\mathbb{R}_k^d,\xi)\leqslant\left(1+2\cdot\frac{k+1}{\xi}\right)^d$$

is a (crude) bound for $N(\mathbb{R}_k^d,\xi)$, the metric entropy condition

$$\int_0^1\{\log N(\mathbb{R}_k^d,\xi)\}^{1/2}\,d\xi<\infty$$

holds. Moreover, letting

$$\begin{aligned}g(x,t):=\;&\cos(t'x)+\sin(t'x)-\exp\left(-\frac{\|t\|^2}{2}\right)\\&+\left\{\frac{1}{2}(t'x)^2-\frac{\|t\|^2}{2}-t'x\right\}\exp\left(-\frac{\|t\|^2}{2}\right),\end{aligned}\tag{2.11}$$

straightforward calculations yield the estimate

$$|g(X,s)-g(X,t)|\leqslant M\cdot\|s-t\|\qquad(\|s\|,\|t\|\leqslant k),$$

where $M=(2k+k^3/2)(1+\|X\|^2)+(3+k^2)\,\|X\|$. Since $M\geqslant 0$ and $EM^2<\infty$, (2.8) now follows from Corollary 7.17 of Araujo and Giné [2].

To prove (2.9), note that $Y_j=X_j+\Delta_j$ with Δ_j given in (2.7). Using trigonometric formulae, we have

$$\cos(t'Y_j)=\cos(t'X_j)-t'\Delta_j\sin(t'X_j)+\varepsilon_{n,j}(t)$$

$$\sin(t'Y_j)=\sin(t'X_j)+t'\Delta_j\cos(t'X_j)+\eta_{n,j}(t),$$

where $|\varepsilon_{n,j}(t)|\leqslant\|t\|^2\,\|\Delta_j\|^2$, $|\eta_{n,j}(t)|\leqslant\|t\|^2\,\|\Delta_j\|^2$. Since

$$Z_n(t)-\tilde{Z}_n(t)=\frac{1}{\sqrt{n}}\sum_{j=1}^n[\varepsilon_{n,j}(t)+\eta_{n,j}(t)],$$

it follows that

$$\max_{\|t\| \leqslant k} |Z_n(t) - \tilde{Z}_n(t)| \leqslant \frac{2k^2}{\sqrt{n}} \cdot \sum_{j=1}^{n} \|\varDelta_j\|^2. \tag{2.12}$$

From

$$\sqrt{n}(S_n^{-1} - I_d) = -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j X_j' - I_d) + O_P(n^{-1/2})$$

and $\sqrt{n}(S_n^{-1} - I_d) = \sqrt{n}(S_n^{-1/2} - I_d)(S_n^{-1/2} + I_d)$ we obtain

$$\sqrt{n}(S_n^{-1/2} - I_d) = -\frac{1}{2\sqrt{n}} \sum_{j=1}^{n} (X_j X_j' - I_d) + O_P(n^{-1/2}). \tag{2.13}$$

Writing $\mathrm{tr}(\cdot)$ for trace, a simple calculation shows that

$$\|\varDelta_j\|^2 = X_j'(S_n^{-1/2} - I_d)^2 X_j - 2X_j' S_n^{-1/2}(S_n^{-1/2} - I_d)\, \bar{X}_n + \bar{X}_n' S_n^{-1} \bar{X}_n$$

and, thus,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\varDelta_j\|^2 = \sqrt{n}\, \mathrm{tr}\left((S_n^{-1/2} - I_d)^2 \frac{1}{n} \sum_{j=1}^{n} X_j X_j' \right) \\ - 2\bar{X}_n' S_n^{-1/2} \sqrt{n}(S_n^{-1/2} - I_d)\, \bar{X}_n + \sqrt{n}\, \bar{X}_n' S_n^{-1} \bar{X}_n. \end{aligned}$$

In view of $\bar{X}_n = O_P(n^{-1/2})$, $n^{-1} \sum_{j=1}^{n} X_j X_j' = I_d + O_P(n^{-1/2})$, (2.12), (2.13) and the definition of the metric ρ, (2.9) follows.

To verify (2.10), note that

$$\begin{aligned} &\max_{\|t\| \leqslant k} |\tilde{Z}_n(t) - Z_n^*(t)| \\ &= \max_{\|t\| \leqslant k} \left| U_n(t) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\frac{(t'X_j)^2}{2} - \frac{\|t\|^2}{2} - t'X_j \right) \exp\left(-\frac{\|t\|^2}{2} \right) \right|, \end{aligned} \tag{2.14}$$

where

$$\begin{aligned} U_n(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} t'\varDelta_j (\cos(t'X_j) - \sin(t'X_j)) \\ &= t' \sqrt{n}(S_n^{-1/2} - I_d)(A_n(t) - \bar{X}_n B_n(t)) - t' \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j B_n(t) \end{aligned}$$

and

$$A_n(t) = \frac{1}{n} \sum_{j=1}^{n} X_j (\cos(t'X_j) - \sin(t'X_j)), \tag{2.15}$$

$$B_n(t) = \frac{1}{n} \sum_{j=1}^{n} (\cos(t'X_j) - \sin(t'X_j)). \tag{2.16}$$

From the compactness of $\{t: \|t\| \leqslant k\}$, the continuity of $A_n(\cdot)$ and $B_n(\cdot)$ and the strong law of large numbers, it is straightforward (although a little tedious) to show that

$$\max_{\|t\| \leqslant k} \left\| A_n(t) + t \exp\left(-\frac{\|t\|^2}{2}\right) \right\| \xrightarrow[n \to \infty]{} 0 \qquad \text{almost surely}$$

and

$$\max_{\|t\| \leqslant k} \left| B_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right) \right| \xrightarrow[n \to \infty]{} 0 \qquad \text{almost surely.}$$

On combining (2.15), (2.16) with (2.13) and plugging into (2.14), we obtain

$$\max_{\|t\| \leqslant k} |\tilde{Z}_n(t) - Z_n^*(t)| \xrightarrow{\mathrm{P}} 0.$$

This proves (2.10) and concludes the proof of Theorem 2.1. ∎

Proof of Theorem 2.2. Use $\int K(t, t)\, \varphi_\beta(t)\, dt < \infty$ and Tonellis theorem to conclude that $\int Z^2(t)\, \varphi_\beta(t)\, dt$ is finite almost surely. The main steps of the proof are to show that

$$|[Z_n^*]|^2 \xrightarrow{\mathscr{D}} |[Z]|^2 \tag{2.17}$$

and

$$|[\tilde{Z}_n - Z_n^*]|^2 \xrightarrow{\mathrm{P}} 0 \tag{2.18}$$

(recall the notation $|[x]|^2$ from (2.4)). Note that, from (2.17) and the continuous mapping theorem (CMT), we have $|[Z_n^*]| \xrightarrow{\mathscr{D}} |[Z]|$, which, in view of the triangle inequality $|\,|[Z_n^*]| - |[\tilde{Z}_n]|\,| \leqslant |[Z_n^* - \tilde{Z}_n]|$ and (2.18), implies $|[\tilde{Z}_n]| \xrightarrow{\mathscr{D}} |[Z]|$ and, thus,

$$|[\tilde{Z}_n]|^2 \xrightarrow{\mathscr{D}} |[Z]|^2. \tag{2.19}$$

From the proof of Theorem 2.1 (cf. (2.12)) we obtain

$$
\begin{aligned}
|Z_n(t) - \tilde{Z}_n(t)| &\leqslant \frac{2\,\|t\|^2}{\sqrt{n}} \sum_{j=1}^{n} \|\Delta_j\|^2 \\
&= \|t\|^2 \cdot o_P(1)
\end{aligned}
$$

which, in turn, implies $|[\tilde{Z}_n - Z_n]|^2 \xrightarrow{\mathrm{P}} 0$. The assertion of Theorem 2.2 now follows from (2.19), the CMT and the inequality $|\,|[\tilde{Z}_n]| - |[Z_n]|\,| \leqslant |[\tilde{Z}_n - Z_n]|$.

To prove (2.17), recall that Z_n^* and Z have the same covariance kernel K defined in (2.3). For fixed $\varepsilon > 0$, we may thus choose a compact set C such that

$$
E(V_{n,1}) = E(V_1) \leqslant \varepsilon^2, \tag{2.20}
$$

where

$$
V_{n,1} = \int_{\mathbb{R}^d \setminus C} Z_n^*(t)^2\, \varphi_\beta(t)\, dt, \qquad n \geqslant 1,
$$

$$
V_1 = \int_{\mathbb{R}^d \setminus C} Z^2(t)\, \varphi_\beta(t)\, dt.
$$

Put

$$
V_{n,2} = \int_C Z_n^*(t)^2\, \varphi_\beta(t)\, dt, \qquad V_2 = \int_C Z^2(t)\, \varphi_\beta(t)\, dt,
$$

and write F_n and F for the distribution functions of $|[Z_n^*]|^2$ $(= V_{n,1} + V_{n,2})$ and $|[Z]|^2$ $(= V_1 + V_2)$, respectively. Using

$$
\begin{aligned}
\{V_{n,1} + V_{n,2} \leqslant t\} &\subseteq \{V_{n,2} \leqslant t\}, \\
\{V_2 \leqslant t\} &\subseteq \{V_1 + V_2 \leqslant t + \varepsilon\} \cup \{V_1 \geqslant \varepsilon\}, \\
\{V_{n,1} + V_{n,2} \leqslant t\} \cup \{V_{n,1} \geqslant \varepsilon\} &\supseteq \{V_{n,2} \leqslant t - \varepsilon\}, \\
\{V_2 \leqslant t - \varepsilon\} &\supseteq \{V_1 + V_2 \leqslant t - \varepsilon\},
\end{aligned}
$$

together with $V_{n,2} \xrightarrow{\mathscr{D}} V_2$, (2.20), and the continuity of the distribution function of V_2, we have

$$\begin{aligned}
F(t-\varepsilon)-\varepsilon &\leqslant P(V_2 \leqslant t-\varepsilon)-\varepsilon \\
&= \lim_{n\to\infty} P(V_{n,2} \leqslant t-\varepsilon)-\varepsilon \\
&\leqslant \liminf_{n\to\infty} F_n(t) \\
&\leqslant \limsup_{n\to\infty} F_n(t) \\
&\leqslant \lim_{n\to\infty} P(V_{n,2} \leqslant t) \\
&= P(V_2 \leqslant t) \\
&\leqslant F(t+\varepsilon)+\varepsilon
\end{aligned}$$

and, thus, (2.17) by letting ε tend to zero.

To prove (2.18), we deduce from the proof of Theorem 2.1 (cf. the reasoning following (2.14)) that

$$\begin{aligned}
(\tilde{Z}_n(t)-Z_n^*(t))^2 \leqslant \Bigg[& \frac{\|t\|}{2} \left\| n^{-1/2} \sum_{j=1}^{n} (X_j X_j' - I_d) \right\| \left\| A_n(t) + t \exp\left(-\frac{\|t\|^2}{2}\right) \right\| \\
& + \|t\| \, \|A_n(t)\| \, o_P(1) \\
& + \|t\| \sqrt{n} \, \|\bar{X}_n\| \left| B_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right) \right| \\
& + \|t\| \, \|\sqrt{n}(S_n^{-1/2} - I_d)\| \, \|\bar{X}_n\| \, |B_n(t)| \Bigg]^2,
\end{aligned} \tag{2.21}$$

where $A_n(t)$ and $B_n(t)$ are given in (2.15) and (2.16), respectively. Letting

$$W_n := \int \left(B_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right) \right)^2 \|t\|^2 \, \varphi_\beta(t) \, dt,$$

Tonellis theorem and some algebra give

$$E(W_n) = \frac{1}{n} \int (1 - \exp(-\|t\|^2)) \, \|t\|^2 \, \varphi_\beta(t) \, dt$$

and, thus,

$$W_n = o_P(1). \tag{2.22}$$

In the same way, straightforward calculations yield

$$E(\widetilde{W}_n) = \frac{1}{n} \int (1 - \|t\|^2 \exp(-\|t\|^2)) \, \|t\|^2 \, \varphi_\beta(t) \, dt$$

and, thus,

$$\widetilde{W}_n = o_P(1), \tag{2.23}$$

where

$$\widetilde{W}_n := \int \left\| A_n(t) + t \exp\left(-\frac{\|t\|^2}{2}\right) \right\|^2 \|t\|^2 \, \varphi_\beta(t) \, dt.$$

Since by (2.21), we have

$$\begin{aligned} |[\tilde{Z}_n - Z_n^*]|^2 \leqslant O_P(1) \cdot \int \|t\|^2 \left\| A_n(t) + t \exp\left(-\frac{\|t\|^2}{2}\right) \right\|^2 \varphi_\beta(t) \, dt \\ + o_P(1) \cdot \int \|t\|^2 \, \|A_n(t)\|^2 \, \varphi_\beta(t) \, dt \\ + O_P(1) \cdot \int \|t\|^2 \left(B_n(t) - \exp\left(-\frac{\|t\|^2}{2}\right) \right)^2 \varphi_\beta(t) \, dt \\ + o_P(1) \cdot \int \|t\|^2 \, B_n(t)^2 \, \varphi_\beta(t) \, dt, \end{aligned}$$

the proof now follows from (2.22), (2.23), and the fact that both the second and the last integral form a tight sequence (we have convergence of expectations). ∎

It is well known that the distribution of

$$T_\beta(d) := \int Z^2(t) \, \varphi_\beta(t) \, dt$$

is that of $\Sigma_{j \geqslant 1} \lambda_j(\beta) \, N_j^2$, where $N_1, N_2, ...$ is a sequence of independent unit normal random variables, and $(\lambda_j(\beta))_{j \geqslant 1}$ is the sequence of nonzero eigenvalues of the integral operator A defined by

$$Aq(s) = \int K(s, t) \, q(t) \, \varphi_\beta(t) \, dt.$$

Although K given in (2.3) looks much simpler than the kernel h_β^* figuring in Theorem 3.1 of Henze and Zirkler [11], we did not succeed in solving the equation $Aq(s) = \lambda q(s)$ and thus getting an explicit form for $\lambda_j(\beta)$.

However, some valuable information on the distribution of $T_\beta(d)$ may be obtained from the relations

$$E(T_\beta(d)) = \int K(t, t)\, \varphi_\beta(t)\, dt,$$

$$\mathrm{Var}(T_\beta(d)) = 2 \iint K^2(s, t)\, \varphi_\beta(s)\, \varphi_\beta(t)\, ds\, dt,$$

and

$$E(T_\beta(d) - ET_\beta(d))^3$$
$$= 8 \iiint K(s, t)\, K(t, u)\, K(u, s)\, \varphi_\beta(s)\, \varphi_\beta(t)\, \varphi_\beta(u)\, ds\, dt\, du.$$

By tedious manipulations of integrals we obtain the following result (for $ET_\beta(d)$ and $\mathrm{Var}(T_\beta(d))$; see also Theorem 3.2 of Henze and Zirkler [11]):

THEOREM 2.3. *We have*

$$E(T_\beta(d)) = 1 - \gamma^{-d/2} \left[1 + \frac{d\beta^2}{\gamma} + \frac{d(d+2)\, \beta^4}{2\gamma^2} \right],$$

$$\mathrm{Var}(T_\beta(d)) = 2(1 + 4\beta^2)^{-d/2} + 2\gamma^{-d} \left[1 + \frac{2d\beta^4}{\gamma^2} + \frac{3d(d+2)\, \beta^8}{4\gamma^4} \right]$$
$$- 4\delta^{-d/2} \left[1 + \frac{3d\beta^4}{2\delta} + \frac{d(d+2)\, \beta^8}{2\delta^2} \right],$$

$$E(T_\beta(d) - ET_\beta(d))^3$$
$$= 8(1 + 3\beta^2)^{-d} - 12(\gamma\omega)^{-d/2} \left[2 + \frac{d\beta^4}{\gamma^2} + \frac{2d\beta^6}{\gamma\omega} + \frac{2d\beta^8}{\gamma^2\omega} + \frac{d(d+2)\, \beta^{12}}{\gamma^2\omega^2} \right]$$
$$+ 6(\gamma\delta)^{-d/2} \left[4 + \frac{4d\beta^4}{\delta} + \frac{4d\beta^6}{\gamma\delta} + \frac{d(d+2)\, \beta^8}{\gamma^2\delta} + \frac{3d(d+2)\, \beta^{12}}{\gamma^2\delta^2} \right]$$
$$- \gamma^{-3d/2} \left[8 + \frac{12d\beta^4}{\gamma^2} + \frac{8d\beta^6}{\gamma^3} + \frac{6d(d+2)\, \beta^8}{\gamma^4} + \frac{d(d+2)(d+8)\, \beta^{12}}{\gamma^6} \right],$$

where

$$\gamma = \gamma(\beta) = 1 + 2\beta^2,$$
$$\delta = \delta(\beta) = 1 + 4\beta^2 + 3\beta^4,$$
$$\omega = \omega(\beta) = 1 + 4\beta^2 + 2\beta^4.$$

3. CONTIGUOUS ALTERNATIVES

In this section, we consider a triangular array $X_{n1}, ..., X_{nn}$, $n \geqslant d+1$, of rowwise independent and identically distributed random vectors having Lebesgue density,

$$f_n(x) = \varphi(x) \cdot (1 + n^{-1/2} h(x)),$$

where $\varphi = \varphi_1$ is the density of $\mathcal{N}_d(\mathbf{0}, I_d)$ and h is a bounded measurable function such that $\int h(x)\, \varphi(x)\, dx = 0$. To guarantee that f_n is nonnegative, we tacitly assume n to be large enough.

In what follows, we retain the notation adopted in the previous sections; i.e., we write $\bar{X}_n = n^{-1} \sum_{j=1}^n X_{nj}$, $S_n = n^{-1} \sum_{j=1}^n (X_{nj} - \bar{X}_n)(X_{nj} - \bar{X}_n)'$, $Y_j := S_n^{-1/2}(X_{nj} - \bar{X}_n)$, etc.

THEOREM 3.1. *Under the triangular array $X_{n1}, ..., X_{nn}$ and the standing assumptions, we have*

$$Z_n \xrightarrow{\mathcal{D}} Z + c$$

in $C(\mathbb{R}^d)$, where Z_n is defined in (2.2), *Z is the Gaussian process figuring in Theorem* 2.1 *and the shift function c is given by*

$$c(t) = \int g(x, t)\, h(x)\, \varphi(x)\, dx,$$

where $g(x, t)$ is defined in (2.11).

THEOREM 3.2. *Under the conditions of Theorem* 3.1, *we have*

$$\int Z_n^2(t)\, \varphi_\beta(t)\, dt \xrightarrow{\mathcal{D}} \int (Z(t) + c(t))^2\, \varphi_\beta(t)\, dt.$$

From Theorem 3.1 and Theorem 3.2, we conclude that the BHEP tests are able to detect alternatives which converge to the normal distribution at the rate $n^{-1/2}$, irrespective of the underlying dimension d.

Proof of Theorem 3.1. Consider the probability measures

$$P^{(n)} := \bigotimes_{j=1}^{n} (\varphi\lambda^d), \qquad Q^{(n)} := \bigotimes_{j=1}^{n} (f_n\lambda^d)$$

on the measurable space $(\mathscr{X}_n, \mathscr{B}_n) := \bigotimes_{j=1}^{n} (\mathbb{R}^d, \mathscr{B}^d)$, where λ^d is Lebesgue measure on the Borel sets $\mathscr{B}^d$ of $\mathbb{R}^d$. Putting $L_n := dQ^{(n)}/dP^{(n)}$, we have

$$\begin{aligned}\log L_n(X_{n1}, ..., X_{nn}) &= \sum_{j=1}^{n} \log(1+n^{-1/2}h(X_{nj}))\\ &= \sum_{j=1}^{n} \left(n^{-1/2}h(X_{nj}) - \frac{h^2(X_{nj})}{2n}\right) + o_{P^{(n)}}(1)\end{aligned}$$

and thus, by the Lindeberg–Feller theorem and the law of large numbers,

$$\log L_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathscr{N}\left(-\frac{\sigma^2}{2}, \sigma^2\right) \qquad \text{under} \quad P^{(n)},$$

where $\sigma^2 := \int h^2(x)\,\varphi(x)\,dx < \infty$. By LeCam's first lemma (see, e.g., Witting and Müller–Funk [28, p. 311]) the sequence $Q^{(n)}$ is contiguous to $P^{(n)}$. Noting that, under $P^{(n)}$,

$$\lim_{n\to\infty} \operatorname{Cov}(Z_n^*(t), \log L_n) = c(t),$$

where Z_n^* is the auxiliary process introduced in (2.5), it is straightforward to show that, for fixed k and $t_1, ..., t_k \in \mathbb{R}^d$, the joint limiting distribution of $Z_n^*(t_1), ..., Z_n^*(t_k)$ and $\log L_n$ under $P^{(n)}$ is the $(k+1)$-variate normal distribution

$$\mathscr{N}_{k+1}\left[\begin{bmatrix} 0\\ \vdots\\ 0\\ -\sigma^2/2\end{bmatrix}, \begin{pmatrix}\Sigma & \mathbf{c}\\ \mathbf{c}' & \sigma^2\end{pmatrix}\right],$$

where $\Sigma = (K(t_l, t_m))_{1\leqslant l, m\leqslant k}$ and $\mathbf{c} = (c(t_1), ..., c(t_k))'$ (recall $K(s, t)$ from (2.3)). Invoking LeCam's third lemma (see, e.g., Witting and Müller–Funk [28, p. 329]), we thus obtain that, under $Q^{(n)}$, the finite dimensional distributions of Z_n^* converge to the finite dimensional distributions of the shifted Gaussian process $Z + c$. Since tightness of Z_n^* under $P^{(n)}$ and the contiguity of $Q^{(n)}$ to $P^{(n)}$ imply tightness of Z_n^* under $Q^{(n)}$, we have

$$Z_n^* \xrightarrow{\mathscr{D}} Z + c \qquad \text{under} \quad Q^{(n)}.$$

Since, by (2.9) and (2.10), $\rho(Z_n, Z_n^*)$ tends to zero stochastically under $P^{(n)}$ and thus also under $Q^{(n)}$ (because of contiguity), the assertion of Theorem 3.1 follows. ▮

Proof of Theorem 3.2. Since the proof follows the reasoning given in the proof of Theorem 2.2, it will only be sketched. First, from the boundedness of h and the fact that $|g(x, t)| \leqslant 3 + \|t\|^2 \|x\|^2/2 + \|t\|^2/2 + \|t\| \|x\|$, we have $|c(t)| \leqslant \gamma(1 + \|t\|^2)$ for some constant γ which ensures that $\int (Z(t) + c(t))^2 \varphi_\beta(t)\, dt$ is finite almost surely. To prove

$$|[Z_n^*]|^2 \xrightarrow{\mathscr{D}} |[Z+c]|^2 \qquad \text{under} \quad Q^{(n)}, \tag{3.1}$$

note that, under $Q^{(n)}$, we have

$$E \int Z_n^*(t)^2 \varphi_\beta(t)\, dt = \int \left(K(t, t) + \frac{n-1}{n} c^2(t) \right) \varphi_\beta(t)\, dt + o(1).$$

This shows that, for fixed $\varepsilon > 0$, there is a compact set C such that, under $Q^{(n)}$, $E(V_1) \leqslant \varepsilon^2$ and $E(V_{n,1}) \leqslant \varepsilon^2$, $n \geqslant 1$, where $V_{n,1}$, V_1 are given in the proof of Theorem 2.2 with the only exception that Z is replaced by the shifted process $Z + c$. The rest of the argument for proving (3.1) runs along the lines of the proof of (2.17). Since $|[Z_n]| - |[Z_n^*]|$ converges to zero stochastically under $P^{(n)}$ (see the argument following (2.18)) and, thus, because of contiguity, also under $Q^{(n)}$, the assertion of Theorem 3.2 is a consequence of (3.1) and the CMT. ▮

4. EMPIRICAL RESULTS AND DISCUSSION

We stress that, for each fixed $\beta > 0$, rejecting the hypothesis H_d of multivariate normality for large values of $T_{n,\beta}$ yields an affine invariant and universally consistent MVN test which may be carried out easily for any number of dimensions (cf. the four bullets at the beginning of Section 1).

To determine approximate upper quantiles of the null distribution of $T_{n,\beta}$, a Monte Carlo experiment was performed. The results are given in Tables I–IV for several values of n, d, and β and the significance levels $\alpha = 0.1$ and $\alpha = 0.05$. Each tabulated value is based on 10.000 Monte Carlo replications. An entry like ${}^{-5}1.17$ stands for 1.17×10^{-5}, and $1.0_3 21$ means 1.00021. The results clearly show that the test is practically sample size independent; i.e., the critical values seem to converge rapidly to their corresponding asymptotic values. This observation (which contrasts with other nonparametric settings where a normal limit arises) is typical for situations where the statistic under discussion is a degenerate U- or

TABLE I

Empirical Percentage Points of $T_{n,\beta}$ $(d=2)$

	$\beta=0.1$		$\beta=0.5$		$\beta=1.0$		$\beta=3.0$	
$1-\alpha$	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
$n=20$	${}^{-5}1.17$	${}^{-5}1.48$	0.060	0.073	0.468	0.540	1.081	1.172
$n=50$	${}^{-5}1.40$	${}^{-5}1.77$	0.065	0.078	0.477	0.553	1.089	1.177
$n=100$	${}^{-5}1.47$	${}^{-5}1.85$	0.065	0.079	0.480	0.555	1.100	1.191
$q_{\beta,d}(\alpha)$	${}^{-5}1.44$	${}^{-5}1.84$	0.065	0.079	0.472	0.554	1.096	1.186
$q^{+}_{\beta,d}(\alpha)$	${}^{-5}1.05$	${}^{-5}1.41$	0.065	0.079	0.475	0.552	1.096	1.188

TABLE II

Empirical Percentage Points of $T_{n,\beta}$ $(d=3)$

	$\beta=0.1$		$\beta=0.5$		$\beta=1.0$		$\beta=3.0$	
$1-\alpha$	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
$n=20$	${}^{-5}2.14$	${}^{-5}2.63$	0.097	0.113	0.615	0.671	1.056	1.095
$n=50$	${}^{-5}2.55$	${}^{-5}3.10$	0.103	0.119	0.621	0.682	1.057	1.095
$n=100$	${}^{-5}2.67$	${}^{-5}3.19$	0.105	0.120	0.621	0.683	1.062	1.098
$q_{\beta,d}(\alpha)$	${}^{-5}2.72$	${}^{-5}3.26$	0.106	0.121	0.620	0.684	1.061	1.096
$q^{+}_{\beta,d}(\alpha)$	${}^{-5}2.52$	${}^{-5}3.15$	0.106	0.121	0.620	0.684	1.062	1.099

TABLE III

Empirical Percentage Points of $T_{n,\beta}$ $(d=5)$

	$\beta=0.1$		$\beta=0.5$		$\beta=1.0$		$\beta=3.0$	
$1-\alpha$	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
$n=20$	${}^{-5}5.14$	${}^{-5}5.73$	0.186	0.199	0.814	0.845	1.008	1.020
$n=50$	${}^{-5}6.24$	${}^{-5}7.13$	0.197	0.213	0.815	0.847	1.014	1.022
$n=100$	${}^{-5}6.55$	${}^{-5}7.44$	0.200	0.216	0.815	0.851	1.014	1.021
$q_{\beta,d}(\alpha)$	${}^{-5}6.75$	${}^{-5}7.52$	0.201	0.217	0.814	0.849	1.015	1.020
$q^{+}_{\beta,d}(\alpha)$	${}^{-5}6.72$	${}^{-5}7.58$	0.201	0.217	0.814	0.852	1.015	1.021

TABLE IV

Empirical Percentage Points of $T_{n,\beta}$ $(d=10)$

	$\beta=0.1$		$\beta=0.5$		$\beta=1.0$		$\beta=3.0$	
$1-\alpha$	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
$n=50$	$^{-4}2.64$	$^{-4}2.79$	0.463	0.474	0.980	0.986	0.9_55	1.0_5
$n=100$	$^{-4}2.79$	$^{-4}2.95$	0.466	0.476	0.979	0.986	1.0_41	1.0_48
$n=200$	$^{-4}2.87$	$^{-4}3.02$	0.467	0.480	0.979	0.986	1.0_48	1.0_319
$q_{\beta,d}(\alpha)$	$^{-4}2.91$	$^{-4}3.04$	0.468	0.479	0.978	0.985	1.0_321	1.0_327
$q^+_{\beta,d}(\alpha)$	$^{-4}2.91$	$^{-4}3.06$	0.468	0.481	0.978	0.985	1.0_321	1.0_327

V-statistic (see Götze [10]). That $W_{n,\beta}$ defined in (1.2) is a degenerate V-statistic with estimated parameters has been exploited by Henze and Zirkler [11].

In each table, the row denoted by $q_{\beta,d}(\alpha)$ is the $(1-\alpha)$-quantile of a lognormal distribution having expectation $\mu_{\beta,d} := E(T_\beta(d))$ and variance $\sigma^2_{\beta,d} = \mathrm{Var}(T_\beta(d))$ given in Theorem 2.3, i.e.,

$$q_{\beta,d}(\alpha) = \mu_{\beta,d}\left(1+\frac{\sigma^2_{\beta,d}}{\mu^2_{\beta,d}}\right)^{-1/2} \exp\left(\Phi^{-1}(1-\alpha)\sqrt{\log\left(1+\frac{\sigma^2_{\beta,d}}{\mu^2_{\beta,d}}\right)}\right),$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal distribution function. Likewise, the row denoted by $q^+_{\beta,d}(\alpha)$ is the $(1-\alpha)$-quantile of a

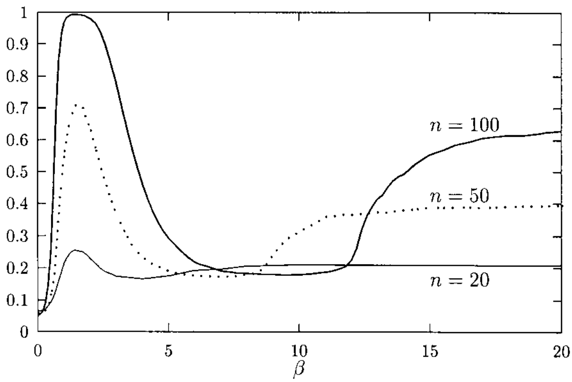

FIG. 4.1. Power of the MVN tests based on $T_{n,\beta}$ $(d=5)$ for a normal mean mixture.

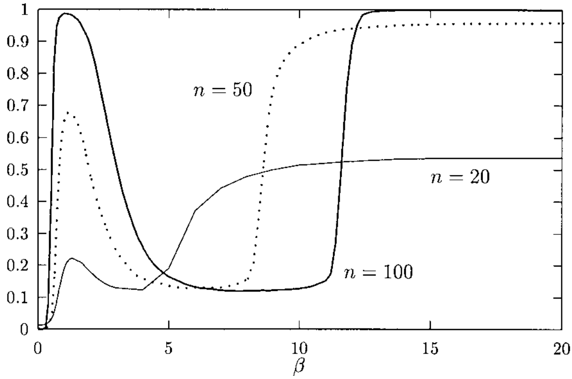

FIG. 4.2. Power of the MVN tests based on $T_{n,\beta}$ $(d=5)$ for a uniform distribution over the unit 5-cube $[0, 1]^5$.

three-parameter lognormal distribution having the first three moments as given in Theorem 2.3; i.e.,

$$q^{+}_{\beta,d}(\alpha)=\mu_{\beta,d}-\frac{\sigma_{\beta,d}}{\sqrt{a+1/a-2}}\left(1-\frac{\exp(\Phi^{-1}(1-\alpha)\sqrt{\log(a+1/a-1)})}{\sqrt{a+1/a-1}}\right),$$

where

$$a=\left(1+\frac{m^2_{\beta,d}}{2\sigma^3_{\beta,d}}+\frac{1}{2}\sqrt{4\frac{m^2_{\beta,d}}{\sigma^3_{\beta,d}}+\frac{m^4_{\beta,d}}{\sigma^6_{\beta,d}}}\right)^{1/3},$$

$$m_{\beta,d}=E(T_\beta(d)-ET_\beta(d))^3.$$

With the exception of the cases $d=2$, $\beta=0.1$ and $d=10$, $\beta=3$, both $q_{\beta,d}(\alpha)$ and $q^{+}_{\beta,d}(\alpha)$ show remarkably good agreement with the empirical quantiles. We therefore suggest to use $q_{\beta,d}(\alpha)$ or $q^{+}_{\beta,d}(\alpha)$ as an approximate critical value for a nominal level α test based on $T_{n,\beta}$ if the sample size n is not too small.

To illustrate the dependence of power of the MVN test based on $T_{n,\beta}$ on the parameter β, Figs. 4.1–4.4 exhibit plots of the empirical power (based on 10.000 Monte Carlo replications) of the $T_{n,\beta}$ test for the case $d=5$ and the sample sizes $n=20$, $n=50$, and $n=100$ as a function of β. In each case, the nominal level is 0.1. The alternatives to normality chosen are an equal mixture of the standard normal distribution $\mathcal{N}_d(\mathbf{0}, I_d)$ and the normal distribution $\mathcal{N}_d(a, I_d)$, where $a=(3, 3, ..., 3)$ (Fig. 4.1), the uniform distribution

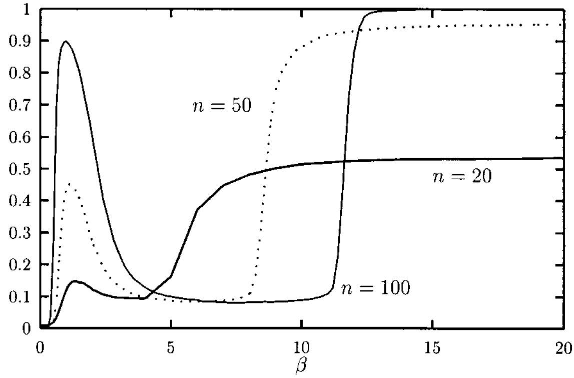

FIG. 4.3. Power of the MVN tests based on $T_{n,\beta}$ $(d=5)$ for a symmetric Pearson type II distribution given in (4.1).

over the unit 5-cube (Fig. 4.2), the spherically symmetric Pearson Type II distribution with density

$$\frac{\Gamma(9/2)}{\pi^{5/2}}(1-\|x\|^2)\cdot\mathbf{1}\{\|x\|^2<1\} \tag{4.1}$$

(Fig. 4.3) and the spherically symmetric Pearson Type VII distribution having density

$$\frac{32}{\pi^3}(1+\|x\|^2)^{-5} \tag{4.2}$$

It is striking to see that, at least for the values of n under study and a certain range of values for β, power does not increase with the sample size. We have no theoretical explanation for this weird behavior which certainly constitutes a field of future research. Taking a large value of β (which effectively amounts in rejecting H_d for large values of the right hand side of (1.5)) seems to be a powerful procedure against short-tailed symmetric alternatives. On the other hand, taking a small value of β which essentially results in computing a convex combination of two skewness measures (see (1.4)) is a good safeguard against symmetric heavy-tailed distributions. Of course, much more work needs to be done to understand the dependence of power on β in order to obtain some kind of an adaptive test for multivariate normality.

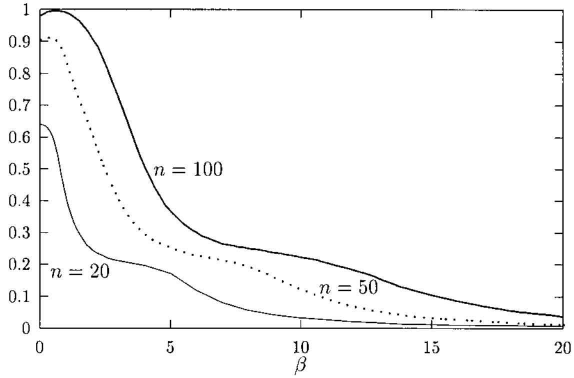

FIG. 4.4. Power of the MVN tests based on $T_{n,\beta}$ $(d=5)$ for the symmetric Pearson type VII distribution given in (4.2).

ACKNOWLEDGMENT

The authors thank the referees for their constructive comments.

REFERENCES

[1] Ahn, S. K. (1992). *F*-probability plot and its application to multivariate normality. *Commun. Statist. Theory Methods* **31** 997–1023.

[2] Araujo, A., and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables.* Wiley, New York.

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

[4] Baringhaus, L., and Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. *Ann. Statist.* **20** 1989–2002.

[5] Bowman, A. W., and Foster, P. J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.* **88** 529–537.

[6] Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.* **14** 708–732.

[7] Csörgő, S. (1989). Consistency of some tests for multivariate normality. *Metrika* **36** 107–116.

[8] De Wet, T., and Randles, R. H. (1987). On the effect of substituting parameter estimators in limiting χ^2, U, and V statistics. *Ann. Statist.* **15** 398–412.

[9] Epps, T. W., and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70** 723–726.

[10] Götze, F. (1979). Asymptotic expansions for Bivariate von Mises Functionals. *Z. Wahrsch. Verw. Gebiete* **50** 333–355.

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

[12] Henze, N. (1994). On Mardia's kurtosis test for multivariate normality. *Commun. Statist. Theory Methods* **23** 1031–1045.

[13] Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statist. Probab. Lett.*. [to appear]

[14] Horswell, R. L., and Looney, St. W. (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. *Journ. Statist. Comp. Simul.* **42** 21–38.

[15] Karatzas, I., and Shreve, S. (1991). *Brownian motion and stochastic calculus.* Springer-Verlag, New York.

[16] Kariya, T., and George, E. (1992). Locally best invariant tests for multivariate normality in curved families with μ known. In *IMS Lecture Notes* Vol. 24 pp. 311–322.

[17] Koziol, J. A. (1993). Probability plots for assessing multivariate normality. *The Statistician* **42** 161–173.

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

[19] Móri, T. F., Rohatgi, V. K., and Székely, G. J. (1993). On multivariate skewness and kurtosis. *Theor. Probab. Appl.* **38** 547–551.

[20] Mudholkar, G. S., Mc Dermott, M., and Srivastava, D. K. (1992). A test of p-variate normality. *Biometrika* **79** 850–854.

[21] Mudholkar, G. S., Srivastava, D. K., and Lin, C. Th. (1995). Some p-variate adaptations of the Shapiro–Wilk test of normality. *Commun. Statist. Theory Methods* **24** 953–985.

[22] Naito, K. (1996). On weighting the studentized empirical characteristic function for testing normality. *Commun. Statist. Simul.* **25** 201–213.

[23] Ozturk, A., and Romeu, J. L. (1992). A new method for assessing multivariate normality with graphical applications. *Commun. Statist. Simul.* **21** 15–34.

[24] Rayner, J. C. W., Best, D. J. and Matthews, K. L. (1995). Interpreting the skewness coefficient. *Commun. Statist. Theory Methods* **24** 593–600.

[25] Romeu, J. L., and Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *Journ. Multiv. Anal.* **46** 309–334.

[26] Singh, A. (1993). Omnibus robust procedures for assessment of multivariate normality and detections of multivariate outliers. In *Multivariate Environmental Statistics* (G. P. Patil and C. R. Rao, Eds.), pp. 445–488. North-Holland, Amsterdam.

[27] Versluis, C. (1996). Comparison of tests for bivariate normality with unknown parameters by transformation to an univariate statistic. *Commun. Statist. Theory Methods* **25** 647–665.

[28] Witting, H., and Müller-Funk, U. (1995). *Mathematische Statistik II.* Teubner, Stuttgart.

[29] Zhu, L., Wong, H. L., and Fang, K. T. (1995). A test for multivariate normality based on sample entropy and projection pursuit. *Journ. Statist. Plann. Inf.* **45** 373–385.