Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit test. Comput. Statist. Data Anal. 53, 1038–1053.

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3001–454 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt
URL: http://www.mat.uc.pt/∼tenreiro/

Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 10–36

AN AFFINE INVARIANT MULTIPLE TEST PROCEDURE FOR ASSESSING MULTIVARIATE NORMALITY

CARLOS TENREIRO

Abstract: A multiple test procedure for assessing multivariate normality (MVN) that combines a finite set of affine invariant test statistics for MVN is proposed. This combination is based on a method introduced by Fromont and Laurent (Ann. Statist., 680–720, 34, 2006) that can be viewed as an improvement of the classical Bonferroni's method. The usefulness of such approach is illustrated through a multiple test involving Mardia's and BHEP (Baringhaus-Henze-Epps-Pulley) tests that are among the most recommended procedures to test a MVN hypothesis. A simulation study carried out for a wide range of alternative distributions to analyze the finite sample power behaviour of the proposed multiple test procedure indicates that the new test presents a good overall performance against other highly recommended MVN tests.

Keywords: Tests for multivariate normality, affine invariance, multiple testing, consistency, Mardia's tests, BHEP tests, Monte Carlo power comparison.
AMS Subject Classification (2010): 62G10, 62H15.

1. Introduction

Let $X_1, \dots, X_n, \dots$ be a sequence of independent copies of the d-dimensional absolutely continuous random vector X with unknown probability density function f, also denoted by f_X, and probability distribution P_f, and $\mathcal{N}_d$ the class of d-variate normal probability density functions. The problem of assessing multivariate normality (MVN) is to test, on the basis of $X_1, \dots, X_n$, the hypothesis

$$H_0 : f \in \mathcal{N}_d,$$

against a general alternative. This is a classical problem in the statistical literature and a huge amount of work has been done on this topic as stressed by Mecklin and Mundfrom (2000) that noticed the existence of about fifty procedures for testing multivariate normality. See also the bibliography given in Csörgő (1986) and the review papers of Henze (2002) and Mecklin and Mundfrom (2004). Despite this fact, there is a continued interest in this subject as attested by the recent papers of Liang et al. (2005), Mecklin and Mundfrom (2005), Székely and Rizzo (2005), Sürücü (2006), Farrel et al. (2007), Arcones (2007), Coin (2007), Tenreiro (2009) and Liang

Received September 22, 2010.
This research has been partially supported by the CMUC (Centre for Mathematics, University of Coimbra)/FCT.

et al. (2009). A strong practical motivation for this continued effort is the fact that many multivariate statistical methods, including MANOVA, multivariate regression, discriminant analysis, and canonical correlation, depend on the acceptance of the MVN hypothesis.

Among the existent wide class of MVN test procedures the Mardia's (1970) tests, based on the Mardia's empirical measures of multivariate skewness and kurtosis, play an important role being among the most recommended and widely used test procedures for assessing MVN (see Romeu and Ozturk, 1993; Mecklin and Mundfrom, 2005; and references therein). Denoting by $\bar{X}_n = n^{-1}\sum_{j=1}^n X_j$ and $S_n = n^{-1}\sum_{j=1}^n (X_j - \bar{X}_n)(X_j - \bar{X}_n)'$ the sample mean and the sample covariance matrix, respectively, Mardia's MS (multivariate skewness) and MK (multivariate kurtosis) test statistics are given by

$$\mathrm{MS} = n b_{1,d} \tag{1}$$

and

$$\mathrm{MK} = \sqrt{n}\,|\, b_{2,d} - d(d+2)| \tag{2}$$

with

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

where $Y_j = S_n^{-1/2}(X_j - \bar{X}_n), j = 1, \ldots, n$, are the scaled residuals and $S_n^{-1/2}$ is the symmetric positive definite square root of S_n^{-1}. Under the null hypothesis of MVN, we have $nb_{1,d} \xrightarrow{d} 6\chi^2_{d(d+1)(d+2)/6}$ and $\sqrt{n}\,(\, b_{2,d} - d(d+2)) \xrightarrow{d} N(0, 8d(d+2))$ (cf. Mardia, 1970). The MS test rejects H_0 for large values of $b_{1,d}$ and the MK test rejects H_0 for both small and large values of $b_{2,d}$. Mardia's test statistics are affine invariant but, similarly to almost all the MVN tests proposed in the literature, they are not consistent against each alternative distribution. Denoting by $\beta_{1,d} = \mathbb{E}\left((X_1-\mu)'\Sigma^{-1}(X_2-\mu)\right)^3$ and $\beta_{2,d} = \mathbb{E}\left((X_1-\mu)'\Sigma^{-1}(X_1-\mu)\right)^2$ the population counterparts to the previous sample skewness and kurtosis measures, where μ is the mean vector and Σ the covariance matrix of X, Baringhaus and Henze (1992) showed that if $\mathrm{E}(X'X)^3 < \infty$ the MVN test based on $b_{1,d}$ is consistent if and only if $\beta_{1,d} > 0$, and Henze (1994) proved that if $\mathrm{E}(X'X)^4 < \infty$ the MVN test based on $b_{2,d}$ is consistent if and only if $\beta_{2,d}$ differs from $d(d+2)$. Therefore, despite these tests can present a high power for an alternative in skewness or kurtosis they can also show a very poor performance especially when the alternative distribution has MVN values of skewness and kurtosis. This problem can also be found in some other test statistics that combine the previous measures of multivariate

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

Mardia, K.V., Foster, K. (1983). Omnibus tests of multinormality based on skewness and kurtosis. Comm. Statist. Theory Methods 12, 207–221.

Mardia, K.V., Kent, J.T. (1991). Rao score tests for goodness of fit and independence. Biometrika 78, 355–363.

Mecklin, C.J., Mundfrom, D.J. (2000). Comparing of the power of classical and newer tests of multivariate normality. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 24–28, 2000.

Mecklin, C.J., Mundfrom, D.J. (2004). An appraisal and bibliography of tests for multivariate normality. Int. Stat. Rev. 72, 123–138.

Mecklin, C.J., Mundfrom, D.J. (2005). A Monte Carlo comparison of Type I and Type II error rates of tests of multivariate normality. J. Stat. Comput. Simul. 75, 93–107.

R Development Core Team (2009). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Romeu, J.L., Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. J. Multivariate Anal. 46, 309–334.

Royston, J.P. (1983). Some techniques for assessing multivarate normality based on the Shapiro–Wilk W. J. Roy. Statist. Soc. Ser. C 32, 121–133.

Royston, J.P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. Stat. Comput. 2, 117–119.

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

Shorack, G.R., Wellner, J.A. (1986). Empirical Processes with Applications to Statistics, Wiley, New York.

Sürücü, B. (2006). Goodness-of-fit tests for multivariate distributions. Comm. Statist. Theory Methods 35, 1319–1331.

Székely, G.J., Rizzo, M.L. (2005). A new test for multivariate normality. J. Multivariate Anal. 93, 58–80.

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

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

Farrel, P.J., Salibian-Barrera, M., Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. J. Stat. Comput. Simul. 77, 1065–1080.

Filliben, J.J. (1975). The probability plot correlation coefficient test for normality Technometrics 17, 111–117.

Fromont, M., Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. Ann. Statist. 34, 680–720.

Henze, N., Zirkler, B. (1990). A class of invariante consistent tests for multivariate normality. Comm. Stat. Theory Methods 19, 3595–3617.

Henze, N. (1994). On Mardia's kurtosis test for multivariate normality. Comm. Statist. Theory Methods 23, 1047–1061.

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

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

Henze, N., Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. J. Multivariate Anal. 62, 1–23.

Horswell, R.L., Looney, S.W. (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. J. Stat. Comput. Simul. 42, 21–38.

Johnson, M.E. (1987). Multivariate Statistical Simulation, Wiley, New York.

Kotz, S., Kozubowski, T., Podgorski, K. (2001). The Laplace Distribution and Generalizations, Birkhauser, Boston.

Liang, J., Pan, W.S.Y., Yang, Z.-H. (2005). Characterization-based Q-Q plots for testing multinormality. Statis. Probab. Lett. 70, 183–190.

Liang, J., Tang, M.-L., Chan, P.S. (2009). A generalized Shapiro–Wilk W statistic for testing high-dimensional normality. Comput. Statist. Data Anal. 53, 3883–3891.

skewness and kurtosis in order to obtain a single "omnibus" test procedure such as those proposed by Mardia and Foster (1983), Mardia and Kent (1991), Horswell and Looney (1992) or Doornik and Hansen (1994).

In order to avoid the lack of consistency for some alternative distributions, a different test for MVN can be used such as a test from the BHEP (Baringhaus–Henze–Epps–Pulley) family introduced by Baringhaus and Henze (1988) and Henze and Zirkler (1990), which extends the Epps and Pulley (1983) procedure to the multivariate context. The BHEP test statistic is a weighted L_2-distance between the empirical characteristic function of the scaled residuals

$$\Psi_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(i\, t' Y_j\right), \quad t \in \mathbb{R}^d,$$

and the characteristic function Φ of the d-dimensional Gaussian standard density $\phi(x) = (2\pi)^{-d/2}\exp(-x'x/2), x \in \mathbb{R}^d$, with weight function $t \to |\Phi_h(t)|^2 = \exp(-h^2 t' t)$, where Φ_h is the characteristic function of $\phi_h(\cdot) = \phi(\cdot/h)/h^d$ and h is a strictly positive real number that needs to be chosen by the user. Therefore the BHEP test statistic is given by

$$\begin{aligned} \mathrm{B}(h) &= n\int |\Psi_n(t) - \Phi(t)|^2 |\Phi_h(t)|^2 dt \\ &= \frac{1}{n}\sum_{i,j=1}^{n} Q(Y_i, Y_j; h), \end{aligned}$$

with $Q(u,v;h) = \phi_{(2h^2)^{1/2}}(u-v) - \phi_{(1+2h^2)^{1/2}}(u) - \phi_{(1+2h^2)^{1/2}}(v) + \phi_{(2+2h^2)^{1/2}}(0)$, for $u, v \in \mathbb{R}^d$. The simplicity of the previous expression shows the appealing feature of the considered weight function. As remarked by Henze and Zirkler (1990) and Fan (1998), the statistic $\mathrm{B}(h)$ can also be interpreted as the L_2-distance between the Parzen-Rosenblatt kernel estimator based on the scaled residuals with kernel $K = \phi$ and smoothing parameter (bandwidth) h, and the convolution $K_h * \phi$, which can be seen as an approximation of the standardized null density when h is taken small. In this form the statistic $\mathrm{B}(h)$ was firstly considered by Bowman and Foster (1993). In some of the previous cited references an alternative smoothing parameter β is considered which is connected with h through the relation $h = 1/(\beta\sqrt{2})$. A theoretical description of the asymptotic behaviour of $\mathrm{B}(h)$ under the null hypothesis, a fixed alternative distribution and a sequence of local alternatives, can be obtained from the work of several authors such as Baringhaus and Henze (1988), Csörgő (1989), Henze and Zirkler (1990), Henze (1997), and Henze and Wagner (1997). In particular, for each $h > 0$, $\mathrm{B}(h)$ has a weighted sum of χ^2 independent random

variables as limiting null distribution and the associated test procedure is consistent against each fixed alternative distribution. Extreme choices of h, $h \to 0$ and $h \to +\infty$, have been studied by Henze (1997), that shows that $\mathrm{B}(h)$ is, in some sense, related to the Mardia's measures $b_{2,d}$ and $b_{1,d}$, respectively.

From a practical point of view, it is well-known that the finite sample performance of the BHEP test is very sensitive to the choice of h. In the multivariate case the standard choice for h, as proposed by Henze and Zirkler (1990), is given by $h = h_{\mathrm{HZ}} := 1.41$. This was the choice of h considered in the above mentioned comparative studies of Mecklin and Mundfrom (2005) and Farrel et al. (2007) that leaded to the recommendation of the Henze–Zirkler test as a formal test of MVN. Despite these good overall comparative results, especially for heavy tailed distributions, these studies also identify some extremely poor results of the Henze–Zirkler test for some alternatives. In a recent paper, Tenreiro (2009) examines the previous standard choice of the smoothing parameter h. As a result of a large Monte Carlo study two distinct behaviour patterns for the BHEP empirical power as a function of h are identified. This leads the author to propose two distinct choices of the bandwidth, depending on the data dimension ($2 \leq d \leq 15$), which are suitable for short tailed or high moment alternatives and for long tailed or moderately skewed alternative distributions, respectively:

$$h = h_{\mathrm{S}} := 0.448 + 0.026\, d \tag{3}$$

and

$$h = h_{\mathrm{L}} := 0.928 + 0.049\, d. \tag{4}$$

These choices agree with a heuristic interpretation of the test performance in terms of the bandwidth h. For large values of h the weight function $t \to \exp(-h^2 t't)$, puts most of its mass near the origin, and then, as the tail behaviour of a probability distribution is reflected by the behaviour of its characteristic function at the origin, it is natural to expect that the test can be sensitive against alternative distributions with long tails. For small values of h it is expectable to obtain a test sensitive to short tailed or high moments alternative distributions. Taking into account the fact that the formulation of a specific alternative hypothesis is in general impossible in a real situation, the author strongly recommended the use of the combined bandwidth

$$h = \bar{h} := \frac{1}{2}\, h_{\mathrm{S}} + \frac{1}{2}\, h_{\mathrm{L}}, \tag{5}$$

which has shown to lead to a powerful test against a wide range of alternatives.

Proof of Theorem 3: First note that the statistics $T_{n,h}$ are defined and continuous on the open subset of $(\mathbb{R}^d)^n$ given by $\mathcal{D} = \{x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : S_n(x) \text{ is positive definite}\}$ for which $\mathrm{P}_\phi(\mathcal{D}) = 1$, where $S_n(x) = n^{-1}\sum_{j=1}^n (x_j - \bar{x}_n)(x_j - \bar{x}_n)'$, $\bar{x}_n = n^{-1}\sum_{j=1}^n x_j$ and $n > d$ (see Dykstra, 1970). Using the continuity of $T_{n,h}$, for all $s < t$ with $0 < F_{T_{n,h}}(s) \leq F_{T_{n,h}}(t) < 1$, we conclude that $T_{n,h}^{-1}(]s,t[)$ is a nonempty open subset of $(\mathbb{R}^d)^n$. Therefore, we get $P_\phi(T_{n,h}^{-1}(]s,t[)) > 0$ which enables us to conclude that $F_{T_{n,h}}$ is strictly increasing. From Theorem 1 we finally get that the multiple test MB has a level of significance inferior or equal to α. The consistency of MB follows from Theorem 2 since at least one of the test statistics involved in the combination, $\mathrm{B}(h_{\mathrm{S}})$ (but the same is true for $\mathrm{B}(h_{\mathrm{L}})$), has a weighted sum of χ^2 independent random variables as limiting null distribution (cf. Baringhaus and Henze, 1988) and the associated test procedure is consistent against each fixed alternative distribution (cf. Csörgő, 1989).

□

References

Arcones, M.A. (2007). Two tests for multivariate normality based on the characteristic function. Math. Methods Statist. 16, 177–201.

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

Baringhaus, L., Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. Ann. Statist. 20, 1889–1902.

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

Coin, D. (2007). A goodness-of-fit test for normality based on polynomial regression. Comput. Statist. Data Anal. 52, 2185–2198.

Csörgő, S. (1986). Testing for normality in arbitrary dimension. Ann. Statist. 14, 708–723.

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

Doornik, J.A., Hansen, H. (1994). An omnibus test for univariate and multivariate normality. Working Paper, Nuffield College, Oxford.

Dykstra, R.L. (1970). Establishing the positive definiteness of the sample covariance matrix. Ann. Math. Statist. 41, 2153–2154.

a) For a fixed $u \in]0,1[$ let u_m be a sequence with $u_m \uparrow u$. Using the right continuity of $F^{-1}_{T_{n,h}}$ for each $h \in H$ that comes from the fact that the distribution function of $T_{n,h}$ under H_0 is strictly increasing for all $h \in H$ (see Shorack and Wellner, 1986; p. 8, Proposition 5), we get $c_{n,h}(u_m) = F^{-1}_{T_{n,h}}(1-u_m) \downarrow F^{-1}_{T_{n,h}}(1-u) = c_{n,h}(u)$, for all $h \in H$. Therefore, $\mathbf{T}_n(u_m) \uparrow \mathbf{T}_n(u)$ and $\psi(u_m) = \mathrm{P}_\phi(\mathbf{T}_n(u_m) > 0) \uparrow \mathrm{P}_\phi(\mathbf{T}_n(u) > 0) = \psi(u)$.
b) For a fixed $u \in]0,1[$ let u_m be a sequence with $u_m \downarrow u$. From the left continuity of $F^{-1}_{T_{n,h}}$ we have $c_{n,h}(u_m) = F^{-1}_{T_{n,h}}(1-u_m) \uparrow F^{-1}_{T_{n,h}}(1-u) = c_{n,h}(u)$, for all $h \in H$. Therefore, $\mathbf{T}_n(u_m) \downarrow \mathbf{T}_n(u)$ and $\{\mathbf{T}_n(u) > 0\} \subset \bigcap_m \{\mathbf{T}_n(u_m) > 0\} \subset \{\mathbf{T}_n(u) \geq 0\}$. Finally, $\psi(u) \leq \lim_m \psi(u_m) \leq \psi(u) + \mathrm{P}_\phi(\mathbf{T}_n(u) = 0)$, where $\mathrm{P}_\phi(\mathbf{T}_n(u) = 0) \leq \sum_{h\in H} \mathrm{P}_\phi(T_{n,h} = c_{n,h}(u)) = 0$ from the continuity of $F_{T_{n,h}}$ under H_0 for all $h \in H$.

□

Proof of Theorem 1: Using the fact that ψ is an increasing function, we deduce that $I_{n,\alpha}$ is an interval of the type $I_{n,\alpha} =]0,\beta[$ or $I_{n,\alpha} =]0,\beta]$ with $\beta = u_{n,\alpha}$ by definition of $u_{n,\alpha}$. Taking $u_m \in I_{n,\alpha}$ such that $u_m \uparrow u_{n,\alpha}$, from part a) of Lemma 1 we conclude that $\psi(u_{n,\alpha}) = \lim_m \psi(u_m) \leq \alpha$, which proves that the level of significance of the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ is at most α, whenever the distribution function of $T_{n,h}$ under H_0 is strictly increasing for all $h \in H$. Additionally, assuming that the distribution function of $T_{n,h}$ under H_0 is continuous for all $h \in H$, from part b) of Lemma 1 and for a sequence u_m such that $u_m \downarrow u_{n,\alpha}$ we have $\psi(u_m) > \alpha$, because $u_{n,\alpha}$ is the supreme of $I_{n,\alpha}$, and $\psi(u_{n,\alpha}) = \lim_m \psi(u_m) \geq \alpha$. Therefore, $\psi(u_{n,\alpha}) = \alpha$ which proves that the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ has a level of significance equal to α. Finally we will prove that $u_{n,\alpha} \leq \alpha$. For that we only need to assume that there exists $h \in H$ such that $F_{T_{n,h}}$ is continuous under H_0. For such an h and for $u \in]0,1[$ we have $\{T_{n,h} > c_{n,h}(u)\} \subset \{\max_{h\in H}(T_{n,h} - c_{n,h}(u)) > 0\} = \{\mathbf{T}_n(u) > 0\}$ and then $\{u \in]0,1[: \mathrm{P}_\phi(\mathbf{T}_n(u) > 0) \leq \alpha\} \subset \{u \in]0,1[: \mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) \leq \alpha\}$. From the continuity of $F_{T_{n,h}}$ under H_0 we get $u_{n,\alpha} \leq \sup\{u \in]0,1[: F_{T_{n,h}}(F^{-1}_{T_{n,h}}(1-u)) \geq 1-\alpha\} = \alpha$.

□

Proof of Theorem 2: Let f be a non-normal density and take $h \in H$ such that $T_{n,h} \xrightarrow{p} +\infty$ under f. We have $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \geq \mathrm{P}_f(T_{n,h} > c_{n,h}(u_{n,\alpha})) \geq \mathrm{P}_f(T_{n,h} > c_{n,h}(\alpha/|H|))$, since $c_{n,h}(u_{n,\alpha}) \leq c_{n,h}(\alpha/|H|)$. Moreover, from the continuity of $F^{-1}_{T_{\infty,h}}$ and the convergence $F^{-1}_{T_{n,h}}(t) \to F^{-1}_{T_{\infty,h}}(t)$ for all $0 < t < 1$ (see Shorack and Wellner, 1986; p. 10), we get $c_{n,h}(\alpha/|H|) = F^{-1}_{T_{n,h}}(1-\alpha/|H|) \to F^{-1}_{T_{\infty,h}}(1-\alpha/|H|)$, and then $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \geq \mathrm{P}_f(T_{n,h} > \sup_{n\in\mathbb{N}} c_{n,h}(\alpha/|H|)) \to 1$.

□

Despite this good property, for several alternative distributions the BHEP test based on $\mathrm{B}(\bar{h})$ is clearly outperformed by one of the Mardia's tests. The main propose of this paper is to show that it is not mandatory to choose between one of the previous approaches to test a MVN hypothesis. Using the method introduced in Fromont and Laurent (2006), which can be viewed as an improvement of the classical Bonferroni's method, it is possible to propose a multiple test procedure that combines the previous MVN tests in a single test procedure that inherits the good properties of each one of the involved tests. Given a finite set of affine invariant statistics, $T_{n,h}, h \in H$, the multiple test procedure rejects the null hypothesis of MVN if one of the statistics is larger than its $(1-u_{n,\alpha})$ quantile under the null hypothesis, $u_{n,\alpha}$ being calibrated so that the final test has a α-level of significance.

The paper is organized as follows. Sufficient conditions for the exact α-level property and the consistency of the multiple test procedure are given in Section 2. In Section 3 the previous approach is used to propose a MVN test that combines both Mardia's tests and the BHEP tests based on $\mathrm{B}(h_\mathrm{S})$ and $\mathrm{B}(h_\mathrm{L})$. A simulation study is carried out in Section 4 to analyze its finite sample power performance in comparison with other highly recommended MVN tests. The proposed multiple test procedure reveals a good performance for a wide range of alternative distributions showing that it may be considered a benchmark MVN test. Finally, in Section 5 we provide some overall conclusions. All the proofs are deferred to Section 6. The simulations and plots in this paper were carried out using the R software (cf. R Development Core Team, 2009).

2. A multiple test procedure for MVN

Given a finite family of statistics $T_{n,h} = T_{n,h}(X_1,\ldots,X_n)$, $h \in H$, to test the MVN hypothesis $H_0 : f \in \mathcal{N}_d$, and a preassigned level of significance $\alpha \in]0,1[$, the standard Bonferroni's method enables us to define a multiple test procedure which leads to the rejection of H_0 if at least one of the test statistics $T_{n,h}$ is larger than its quantile of order $1-\alpha/|H|$, where $|H|$ denotes the cardinality of H and the large values of the different test statistics are considered significants. However, this is in general a too conservative procedure that lacks power especially when several highly correlated test statistics under H_0 are considered.

Assuming that $T_{n,h}$, $h \in H$, are affine invariant statistics, that is,

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

for all $b \in \mathbb{R}^d$ and nonsingular matrix A, we consider in the following an alternative method proposed by Fromont and Laurent (2006) to define an affine invariant multiple test for assessing MVN with an exact α-level of significance. Note that the results presented in this section do not depend on the considered null hypothesis of normality. They are also valid if other affine invariant null family of probability density functions is considered.

2.1. Description of the multiple test procedure. For $u \in]0,1[$ and $h \in H$, denote by $c_{n,h}(u)$ the quantile of order $1-u$ of the test statistic $T_{n,h}$ under the hypothesis H_0 and take the corrected statistic

$$\mathbf{T}_n(u) = \max_{h\in H} \left(T_{n,h} - c_{n,h}(u)\right). \tag{6}$$

Since $f_X \in \mathcal{N}_d$ if and only if $f_{AX+b} \in \mathcal{N}_d$, the quantile $c_{n,h}(u)$ does not depend on the distribution considered under the null hypothesis. Moreover, the affine invariance of each one of the statistics $T_{n,h}$, $h \in H$, implies the affine invariance of $\mathbf{T}_n(u)$, for every $u \in]0,1[$. The idea is now to consider the test procedure that rejects the null hypothesis whenever

$$\mathbf{T}_n(u_{n,\alpha}) > 0$$

where

$$u_{n,\alpha} = \sup I_{n,\alpha} \tag{7}$$

with

$$I_{n,\alpha} = \{u \in]0,1[: \mathrm{P}_\phi(\mathbf{T}_n(u) > 0) \le \alpha\},$$

and ϕ the d-dimensional Gaussian standard density. In practice, the value $u_{n,\alpha}$, the level at which each one of the tests $T_{n,h}$, $h \in H$, is performed, is estimated by Monte Carlo experiments under the null hypothesis as described in Fromont and Laurent (2006) and explained later.

Denoting by $F_{T_{n,h}}$ and $F^{-1}_{T_{n,h}}$ the probability distribution function and the quantile function of $T_{n,h}$ under H_0, respectively, we have

$$\begin{aligned}\mathrm{P}_\phi(\mathbf{T}_n(\alpha/|H|) > 0) &\le \sum_{h\in H} \mathrm{P}_\phi(T_{n,h} > c_{n,h}(\alpha/|H|))\\ &= \sum_{h\in H}\Big(1 - F_{T_{n,h}}(F^{-1}_{T_{n,h}}(1-\alpha/|H|))\Big)\\ &\le \sum_{h\in H}(1-(1-\alpha/|H|)) = \alpha.\end{aligned}$$

Therefore $\alpha/|H| \in I_{n,\alpha}$ and $\alpha/|H| \le u_{n,\alpha}$, which shows that the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ is at least as powerful as the Bonferroni's procedure $I(\mathbf{T}_n(\alpha/|H|) > 0)$ whenever its level of significance is at most α as we establish in the following.

defined by $L_n(s_0) = \inf\{\beta \in]0,1[: s_0 \in R_\beta\}$ and it is easy to see that $s_0 \in R_\alpha$ whenever $L_n(s_0) < \alpha$ and that $\mathrm{P}_\phi(\{s : L_n(s) < \alpha\}) \le \alpha$. Thus, if we compare the P-value with the preassigned level α and rejects the null hypothesis if $L_n < \alpha$, we get a test procedure that has an error of first kind inferior or equal to α. An approximation of $L_n(s_0)$ can be easily obtained if for each dimension d and sample size n we get Monte Carlo estimates of the quantiles $c_{n,h}(u_{n,\alpha})$ for each one of the involved statistics and for α varying on a grid of the interval $]0,1[$. Such estimates and a R function to evaluate (an approximation of) the P-value associated to an observation s_0 may be obtained from the author.

5. Conclusions

In this paper, by using a improved Bonferroni's method introduced by Fromont and Laurent (2006), a multiple test procedure that enables the combination of a finite set of affine invariant tests for MVN is considered. Its usefulness is illustrated through a multiple test combining Mardia's and BHEP tests which are among the most recommended procedures to test a MVN hypothesis. The proposed multiple test procedure reveals a good empirical power for a wide range of alternative distributions, showing an overall good performance against the most recommended MVN tests in the literature.

6. Proofs

In the next result we establish some useful properties of the function $\psi(u) = \mathrm{P}_\phi(\mathbf{T}_n(u) > 0)$ defined on $]0,1[$, where $\mathbf{T}_n(u)$ is given by (6) and ϕ is the d-dimensional Gaussian standard density.

Lemma 1. *For $n \in \mathbb{N}$, the function ψ is increasing with $\lim_{u\downarrow 0}\psi(u) = 0$ and $\lim_{u\uparrow 1}\psi(u) = 1$. Additionally, it satisfies: a) If the distribution function of $T_{n,h}$ under H_0 is strictly increasing for all $h \in H$, ψ is left continuous; b) If the distribution function of $T_{n,h}$ under H_0 is continuous for all $h \in H$, ψ is right continuous.*

Proof: Let $u, v \in]0,1[$ be such that $u < v$. For all $h \in H$ we have $c_{n,h}(u) \ge c_{n,h}(v)$, and then $\mathbf{T}_n(u) \le \mathbf{T}_n(v)$ which entails that $\psi(u) = \mathrm{P}_\phi(\mathbf{T}_n(u) > 0) \le \mathrm{P}_\phi(\mathbf{T}_n(v) > 0) = \psi(v)$. Moreover, $\mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) = 1 - F_{T_{n,h}}(c_{n,h}(u)) = 1 - F_{T_{n,h}}(F^{-1}_{T_{n,h}}(1-u)) \le 1-(1-u) = u$, (see Shorack and Wellner, 1986; p. 5, Proposition 1) and $\lim_{u\downarrow 0}\mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) = 0$, $\lim_{u\uparrow 1}\mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) = 1$, for all $h \in H$. Therefore, $\lim_{u\downarrow 0}\psi(u) \le \sum_{h\in H}\lim_{u\downarrow 0}\mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) = 0$ and, for $h \in H$, $\lim_{u\uparrow 1}\psi(u) \ge \lim_{u\uparrow 1}\mathrm{P}_\phi(T_{n,h} > c_{n,h}(u)) = 1$.

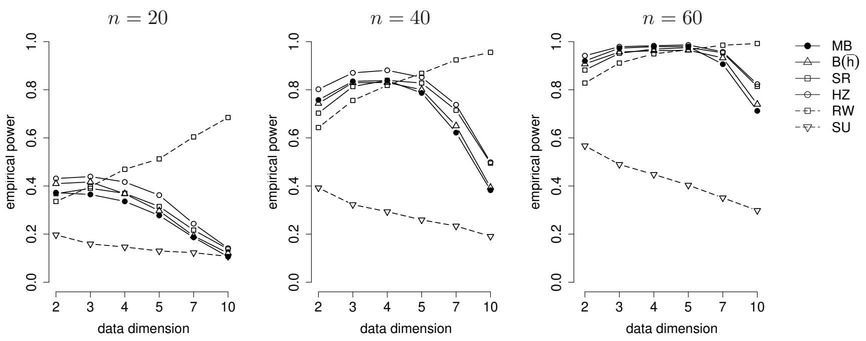

FIGURE 11. *Normal location mixture distribution with* $p = 0.9$.

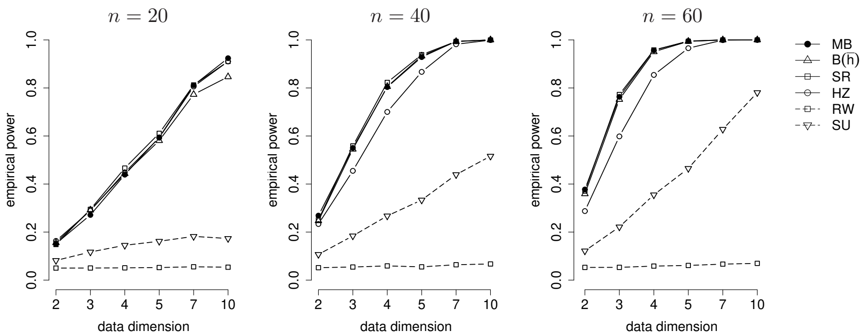

FIGURE 12. *Normal scale mixture distribution with* $p = 0.5$.

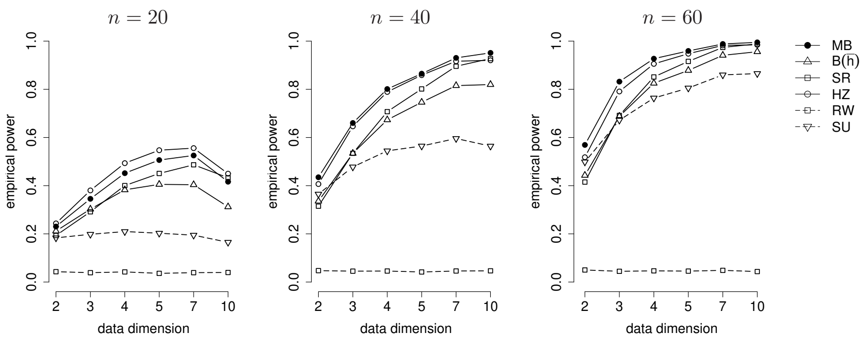

FIGURE 13. *Normal scale mixture distribution with* $p = 0.9$.

2.2. Finite sample behaviour under H_0. Under some conditions on the null distribution of the statistics $T_{n,h}$, $h \in H$, the next non-asymptotic result states that the level of significance of the test procedure $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$, with $u_{n,\alpha}$ given by (7), is at most α. As we can conclude from the proof given in Section 6, this result essentially depends on the continuity properties of the function $\psi(u) = \mathrm{P}_\phi(\mathbf{T}_n(u) > 0)$ defined on the interval $]0, 1[$.

Theorem 1. *If for all $h \in H$ the distribution function of $T_{n,h}$ under H_0 is strictly increasing (on the set $\{t : 0 < F_{T_{n,h}}(t) < 1\}$), then for all $f \in \mathcal{N}_d$ we have*

$$\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \le \alpha,$$

for $0 < \alpha < 1$. Moreover, if the distribution function of $T_{n,h}$ under H_0 is continuous for all $h \in H$, then $u_{n,\alpha} \le \alpha$ and for all $f \in \mathcal{N}_d$ we have

$$\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) = \alpha.$$

2.3. Consistency against fixed alternatives. Under the previous conditions, for a fixed alternative f the power $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0)$ of the multiple test satisfies the following double inequality that highlights its main features

$$\max_{h \in H} \mathrm{P}_f(T_{n,h} > c_{n,h}(u_{n,\alpha})) \le \mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \le \sum_{h \in H} \mathrm{P}_f(T_{n,h} > c_{n,h}(\alpha)).$$

The multiple test presents a low power for alternatives that show a low power for each one of the tests $T_{n,h}, h \in H$. However, its power is always superior to the power of the best of the involved tests performed at level $u_{n,\alpha}$. Whenever the level $u_{n,\alpha}$ is bigger than $\alpha/|H|$ we expect that the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ can show a better power performance than the standard Bonferroni's test procedure. Under the conditions of Theorem 1 note that if the test statistics $T_{n,h}, h \in H$, are independent under H_0 then $u_{n,\alpha} = 1 - (1-\alpha)^{1/|H|}$ which is close to $\alpha/|H|$ for small α. Therefore, if the test statistics $T_{n,h}, h \in H$, are highly uncorrelated the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ can be close to a Bonferroni's multiple test procedure.

Under some weak conditions the proposed multiple test procedure is consistent as stated in the next result. In particular, it is consistent for each alternative distribution if at least one of the involved tests is consistent against each alternative distribution.

Theorem 2. *Let f be a non-normal probability density function, and assume there exists $h \in H$ such that $T_{n,h} \xrightarrow{p} +\infty$, under f. If $T_{n,h} \xrightarrow{d} T_{\infty,h}$ under H_0, where the distribution function of $T_{\infty,h}$ is strictly increasing, then $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \to 1$, as $n \to +\infty$.*

3. Combining Mardia's and BHEP tests

From several simulation studies it is well-known that Mardia's skewness test MS performs well for skewed or long tailed alternatives and Mardia's kurtosis test MK is particularly good for short tailed alternatives, being among the most recommended tests for MVN (cf. Henze and Zirkler, 1990; Romeu and Ozturk, 1993). However, the Mardia's tests do not reveal any power if the alternative distribution has MVN values of skewness and kurtosis. In order to overcome this negative feature, the approach introduced in the previous section is used here to propose a MVN test that can perform well for a wide range of alternative distributions.

The multiple test we consider, labelled MB henceforth, involves both Mardia's test statistics MS and MK given by (1) and (2), and the BHEP tests with $h = h_{\mathrm{S}}$ and $h = h_{\mathrm{L}}$ given by (3) and (4). From Tenreiro (2009), we know that $\mathrm{B}(h_{\mathrm{S}})$ is suitable for short tailed or high moment alternatives and $\mathrm{B}(h_{\mathrm{L}})$ presents a relevant performance for long tailed or moderately skewed alternative distributions. Moreover, these two last tests are consistent against each alternative distribution. Therefore, for $T_{n,1} = \mathrm{MS}$, $T_{n,2} = \mathrm{MK}$, $T_{n,3} = \mathrm{B}(h_{\mathrm{S}})$ and $T_{n,4} = \mathrm{B}(h_{\mathrm{L}})$, the multiple test MB is based on

$$\mathbf{T}_n(u) = \max_{h\in H} \left(T_{n,h} - c_{n,h}(u)\right), \tag{8}$$

where $H = \{1, 2, 3, 4\}$ and $c_{n,h}(u)$ is the quantile of order $1 - u$ of the test statistic $T_{n,h}$ under the null hypothesis of MVN. The next result, which is a consequence of Theorems 1 and 2, establish that the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ based on (8) with $u_{n,\alpha}$ given by (7) is consistent against each fixed alternative and has a level of significance that is at most equal to α.

Theorem 3. *For $n > d$ and $0 < \alpha < 1$ we have $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \leq \alpha$, for all $f \in \mathcal{N}_d$. Moreover, $\mathrm{P}_f(\mathbf{T}_n(u_{n,\alpha}) > 0) \to 1$, as $n \to +\infty$, for all $f \notin \mathcal{N}_d$.*

In order to implement the MB test, 20,000 simulations under the null hypothesis of the involved test statistics and the R function quantile(·,type=7) were used for estimating the $1 - u$ quantiles $c_{n,h}(u)$ for u varying on a regular grid, $u_{i+1} = u_i + p$ with $u_1 = p$, of the interval $]0, 1[$, and other 20,000 simulations were used for estimating the probabilities $\mathrm{P}_\phi(\,\mathbf{T}_n(u) > 0\,)$. Finally, we have taken the largest value of u that satisfies $\mathrm{P}_\phi(\mathbf{T}_n(u) > 0) \leq \alpha$ as an approximation for $u_{n,\alpha}$ defined by (7). For $\alpha = 0.01$ and $\alpha = 0.05$, and several sample sizes n and data dimensions d, we present in Table 1 the estimated levels $u_{n,\alpha}$ based on a regular grid of size $p = 0.0001$. Note

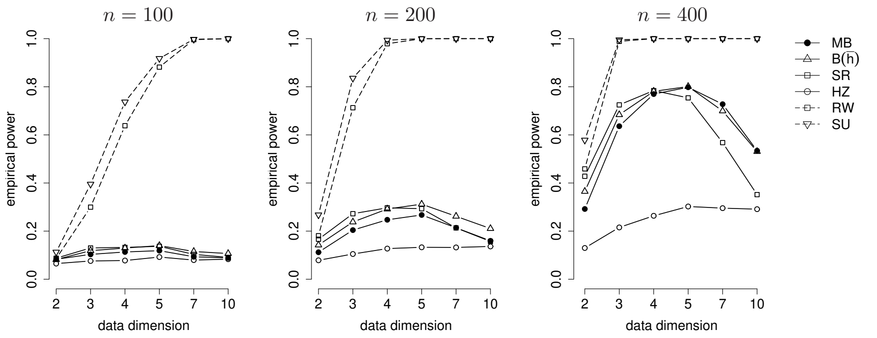

FIGURE 8. *High moment Khintchine distribution with GEP marginals.*

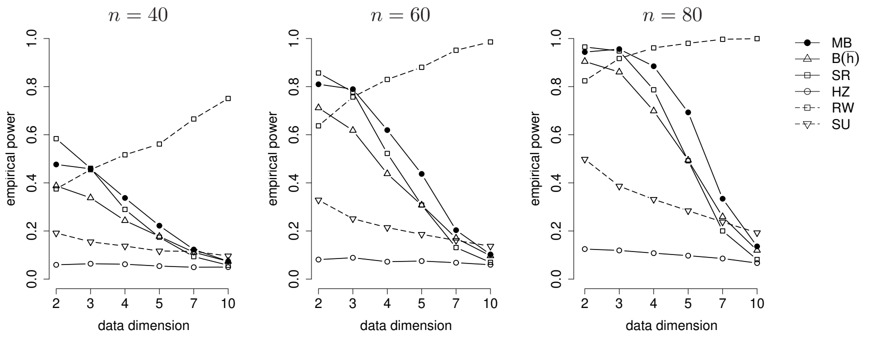

FIGURE 9. *Normal location mixture distribution with $p = 0.5$.*

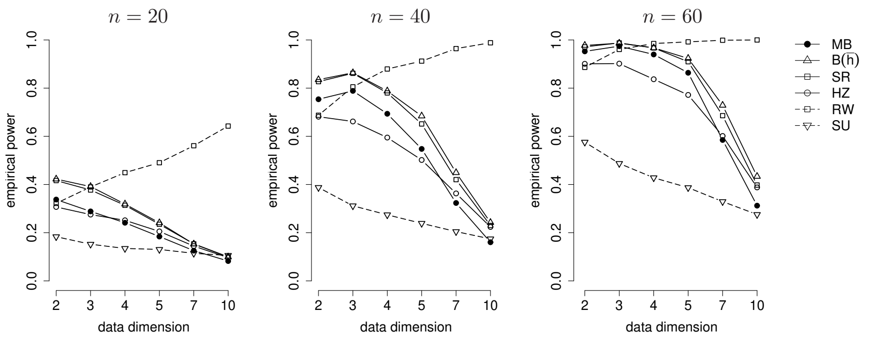

FIGURE 10. *Normal location mixture distribution with $p = 0.79$.*

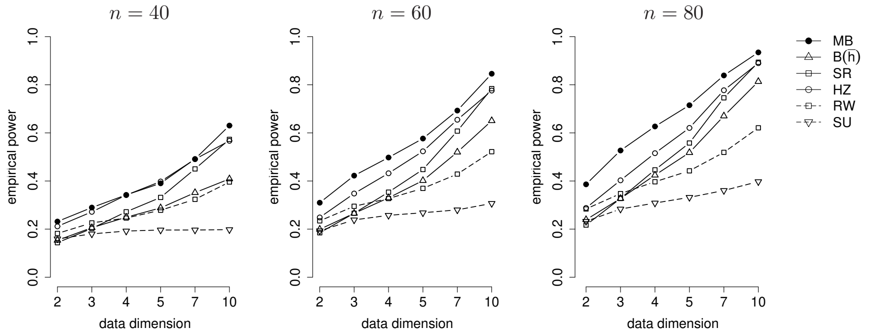

Figure 5. *Pearson Type VII distribution with $m = 10$.*

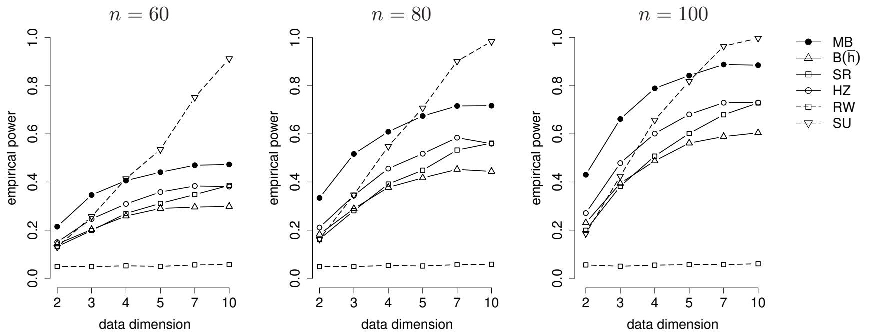

Figure 6. *Burr-Pareto-Logistic distribution with normal marginals and $\alpha = 1$.*

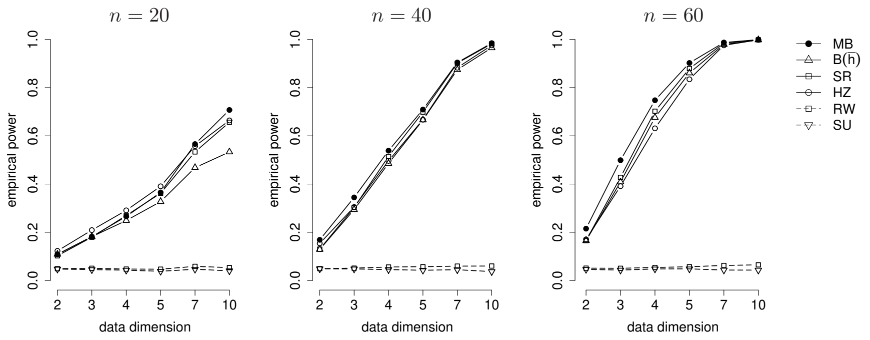

Figure 7. *Khintchine distribution with normal marginals.*

sample size	2	3	4 (data dimension)	5	7	10
			$\alpha = 0.01$			
20	3.8e-03	4.1e-03	2.7e-03	2.9e-03	2.5e-03	2.5e-03
60	3.4e-03	3.2e-03	3.1e-03	2.6e-03	3.0e-03	2.8e-03
100	4.0e-03	2.9e-03	3.3e-03	3.0e-03	3.2e-03	3.0e-03
200	3.6e-03	3.1e-03	2.6e-03	2.6e-03	2.6e-03	2.8e-03
400	3.3e-03	3.2e-03	2.7e-03	3.0e-03	3.2e-03	3.0e-03
			$\alpha = 0.05$			
20	1.8e-02	1.7e-02	1.6e-02	1.6e-02	1.5e-02	1.5e-02
60	1.9e-02	1.9e-02	1.6e-02	1.5e-02	1.4e-02	1.5e-02
100	2.1e-02	1.7e-02	1.7e-02	1.7e-02	1.5e-02	1.5e-02
200	1.8e-02	1.7e-02	1.6e-02	1.6e-02	1.5e-02	1.4e-02
400	1.8e-02	1.7e-02	1.8e-02	1.7e-02	1.6e-02	1.6e-02

Table 1. *Estimates of $u_{n,\alpha}$ for $\alpha = 0.01, 0.05$ based on a regular grid of size 0.0001 of the interval $]0, \alpha]$. The number of replications for each stage of the estimation process is $20{,}000$.*

sample size	2	3	4 (data dimension)	5	7	10
			$\alpha = 0.01$			
20	1.04e-02	9.91e-03	9.61e-03	9.98e-03	9.03e-03	1.05e-02
60	9.84e-03	9.26e-03	9.48e-03	9.16e-03	9.10e-03	9.77e-03
100	1.06e-02	9.37e-03	1.05e-02	1.04e-02	1.02e-02	1.03e-02
200	9.72e-03	9.71e-03	9.34e-03	8.92e-03	9.17e-03	9.62e-03
400	1.01e-02	9.65e-03	9.26e-03	1.07e-02	1.08e-02	9.22e-03
			$\alpha = 0.05$			
20	5.01e-02	4.99e-02	5.13e-02	5.25e-02	5.06e-02	5.17e-02
60	4.94e-02	4.87e-02	4.78e-02	4.85e-02	4.58e-02	5.10e-02
100	5.19e-02	4.88e-02	5.02e-02	5.17e-02	5.09e-02	5.26e-02
200	4.90e-02	5.03e-02	4.88e-02	4.92e-02	4.94e-02	4.74e-02
400	5.02e-02	5.06e-02	5.02e-02	4.99e-02	5.04e-02	5.03e-02

Table 2. *Estimates of the nominal level of significance of the multiple test MB for a preassigned level α. The number of replications for each case is $100{,}000$.*

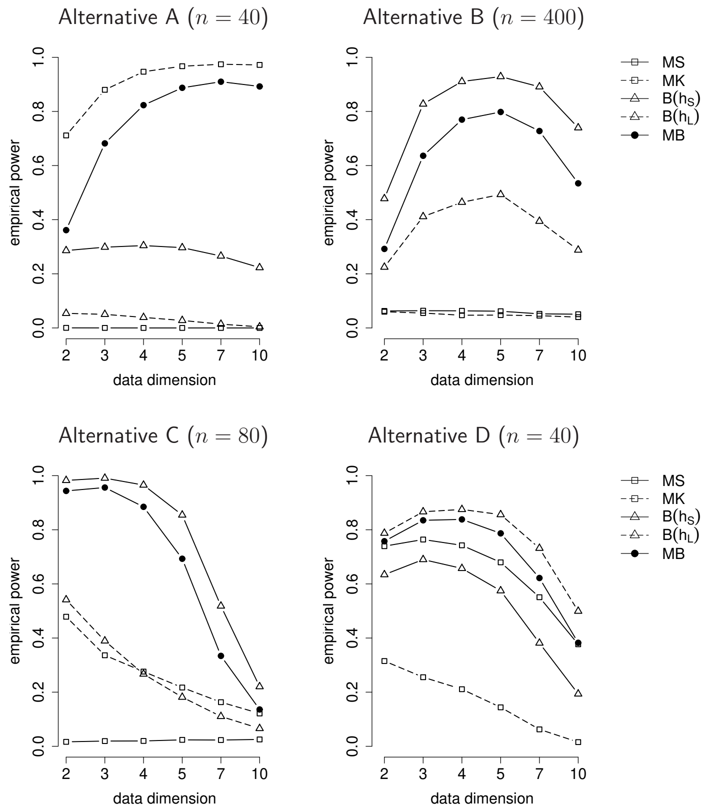

FIGURE 1. *Empirical power at level $\alpha = 0.05$ based on 10^4 replications for each distribution and data dimension d. Alternative A is a Pearson Type II distribution with $m = 0.5$ and alternative B is a high moment Khintchine distribution with GEP marginals. Alternatives C and D are mixtures of MVN distributions. Alternative C is symmetric with light tails and alternative D is skewed with heavy tails.*

that for moderate and large data dimensions and specially for $\alpha = 0.01$, the considered combination is close to the Bonferroni's test procedure. Table 2 shows estimates for the nominal levels of significance of the test MB based on 100,000 simulations under the null hypothesis. With some few

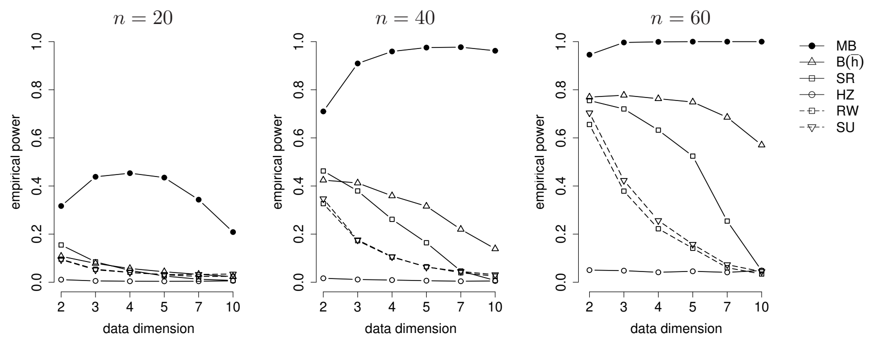

FIGURE 2. *Pearson Type II distribution with $m = 0$.*

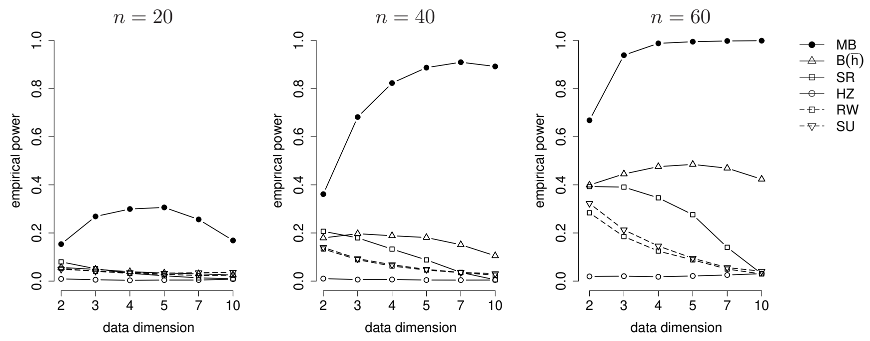

FIGURE 3. *Pearson Type II distribution with $m = 0.5$.*

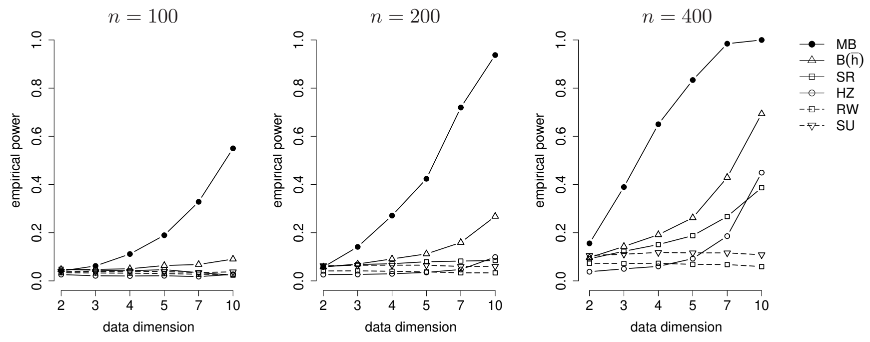

FIGURE 4. *Pearson Type II distribution with $m = 10$.*

in order to obtain an approximation of its null distribution as described in Sürücü (2006; p. 1322).

Figures 2–13 show the empirical power results for 12 typical alternatives that give us a quite complete overview of the finite sample performance of the considered tests. Some alternatives that show drastic departures from normality such as the heavily skewed alternatives or the asymmetric Laplace distributions are not considered in the following because the empirical power of the considered tests was very high, close to 1. From the figures we can clearly identify some alternative distributions where the tests HZ, RW and SU show a low empirical power. The test HZ is very sensitive against heavy tailed alternatives but it also reveals an inferior performance for distributions with light tails. The test RW seems to be especially effective when the marginal alternative distributions are far from normal, but it also shows a very poor behaviour otherwise. For some of the alternatives, its power is even inferior to the significance level of the test. Although these tests can reveal a very good power for some of the considered alternatives, the fact that they can also present a very poor performance for other alternatives is an undesirable feature particularly when no information about the alternative hypothesis is available. Hence, especially with the availability of other test procedures with better power properties, the tests HZ, RW and SU are not recommended.

A better overall performance seems to be reached by the affine invariant test procedures SR, $\mathrm{B}(\bar{h})$ and MB. The test SR is very sensitive against heavy tailed alternatives which corroborate previous research by Székely and Rizzo (2005) but it also reveals an inferior performance for distributions with light tails in comparison with the tests MB and $\mathrm{B}(\bar{h})$ especially for large data dimensions. Taking into account the excellent performance shown by the MB test for some of the considered alternatives together with the fact that this test is among the best tests for all the considered alternative distributions, if one is going to rely on one and only one of the considered test procedures the MB test is recommended.

4.4. P-value evaluation. The MB multiple test can be viewed as a test procedure based on the increasing family of critical regions $R_\alpha = \{\mathbf{T}_n(u_{n,\alpha}) > 0\}$, indexed by $\alpha \in]0,1[$, where $\mathbf{T}_n(u) = \mathbf{T}_n(u;s)$ depends on the observation $s = \{X_1,\ldots,X_n\}$, and $\mathrm{P}_\phi(R_\alpha) \leq \alpha$, for all $\alpha \in]0,1[$ (cf. Theorem 3). For a fixed level α we reject the null hypothesis of MVN on the basis of the the observation s_0 if and only if $s_0 \in R_\alpha$. In practice, it is useful to be able to evaluate the P-value associated to the observation s_0 that represents the degree to which the test procedure rejects H_0. It is

exceptions the estimated levels are inside the approximate 95% confidence interval for the preassigned level α. Although we were not able to prove that the test $I(\mathbf{T}_n(u_{n,\alpha}) > 0)$ has an exact α-level of significance, the previous implementation enables us to obtain a multiple test procedure with a attained level of significance close to α.

With the goal of getting some insight about the finite sample behaviour of the multiple test procedure in relation to each of the tests involved in the combination, we present in Figure 1 their empirical power for four alternative distributions labelled A, B, C and D. A more detailed description of these alternatives will be given in the next section. Alternative A is a Pearson Type II distribution with $m = 0.5$. It is symmetric with light tails and the test MK is the best of the considered tests for this distribution. Alternative B is a high moment Khintchine alternative with GEP marginals (see Johnson, 1987; chapter 8 and paragraph 2.4), the reason way both Mardia's tests have no power. $\mathrm{B}(h_\mathrm{S})$ is the best choice for this alternative. Alternatives C and D are mixtures of multivariate normal distributions. Alternative C is symmetric with light tails whereas alternative D is skewed with heavy tails. The BHEP tests $\mathrm{B}(h_\mathrm{S})$ and $\mathrm{B}(h_\mathrm{L})$ are, respectively, the best of the considered tests for these alternatives.

The test MB is never the best of the considered tests. However, it inherits the good properties of each of the involved tests revealing a good performance for all the referred alternatives. Having in mind that the formulation of a specific alternative hypothesis is in general impossible in a real situation, this property, not shared by any of the tests involved in the combination, is a relevant one.

4. Finite sample power analysis

In order to assess the performance of the proposed multiple test, a simulation study is conducted to compare its empirical power with other highly recommended MVN tests. In the following we describe the MVN tests and the alternative distributions included in the study and we summarize the observed empirical power results.

4.1. Tests under study. Besides the MB multiple test, five other MVN tests have been included in the study. We have chosen three affine invariant tests that are consistent against all fixed alternatives: the Henze and Zirkler's (1990) test (labelled HZ) which is based on $\mathrm{B}(h_\mathrm{HZ})$ with $h_\mathrm{HZ} = 1.41$, the BHEP test based on $\mathrm{B}(\bar{h})$ with $\bar{h}$ given by (5) and the test proposed by Székely and Rizzo (2005) (labelled SR). The HZ test was considered in the comparative studies of Henze and Zirkler (1990), Mecklin

and Mundfrom (2005) and Farrel et al. (2007) which recommend HZ as a formal test of MVN. The BHEP test based on $\mathrm{B}(\bar{h})$ was recommended by Tenreiro (2009) as a good alternative to the HZ test which usually reveals a poor performance against short tailed alternatives. The results of a Monte Carlo power study undertaken by Székely and Rizzo (2005) suggest that the SR test is a powerful competitor to existing affine invariant tests being very sensitive against heavy tailed alternatives.

Two other MVN tests that revealed a promising behaviour in some recent studies have been included in our study. The first one, labelled RW, in a revision given in Royston (1992) of the Royston's (1983) multivariate extension of the Shapiro and Wilks's (1965) goodness of fit test and has been considered in Farrel et al. (2007). The second one, labelled SU, is the test proposed by Sürücü (2006). This test is based on a d-variate version of the test statistic defined as a weighted sum of the Shapiro and Wilks's (1965) statistic and a correlation statistic due to Filliben (1975), the weights being determined by the sample skewness and kurtosis.

4.2. The alternative distributions. The considered set of alternative distributions includes a wide set of distributions previously considered in other simulations studies such as those of Henze and Zirkler (1990), Romeu and Ozturk (1993), Mecklin and Mundfrom (2005) and Székely and Rizzo (2005).

We investigate some symmetric distributions from Pearson's Types II and VII families, including the multivariate uniform or the multivariate Cauchy distributions, and the quasi normal distributions with parameter $m = 10$ from both families. The Pearson Type II distributions have tails lighter than normal whereas the Pearson Type VII distributions have tails heavier than normal. For a detailed discussion about these two types of elliptically contoured distributions see Johnson (1987; p. 110–121).

We also considered some heavily skewed distributions such as the multivariate χ^2_1 and the multivariate lognormal with independent marginals. Some members of the multivariate asymmetric Laplace family described in Kotz et al. (2001; chapter 6) were also studied. All these distributions have tails heavier than normal and express strong departures from the MVN hypothesis.

Distributions with some characteristics identical to MVN were also included in the study. These distributions include (meta-)Burr-Pareto-Logistic distributions with normal marginals (see Johnson, 1987; chapter 9), and two Khintchine distributions with generalized exponential power, GEP, marginal distributions (see Johnson, 1987; chapter 8 and paragraph 2.4).

A Khintchine distribution with GEP marginal distributions with shape parameters $\alpha, \tau > 0$ is defined by $X = Z(2U-1) = Z(2U_1-1,\ldots,2U_d-1)'$, where the U_i's are independent having a uniform distribution over the interval $[0,1]$ and $Z = (3\Gamma(\alpha)/\Gamma(\alpha+2\tau))^{1/2} W^\tau$, where W is a gamma variable independent of U with shape parameter α and scale parameter 1. Note that X has a centrally symmetric distribution about the origin. For the first alternative from this family we took $\alpha = 1.5$ and $\tau = 0.5$, which leads to a Khintchine distribution with normal marginals and Mardia's kurtosis coefficient larger than the MVN one. For the second alternative, we took again $\alpha = 1.5$ but now $\tau > 0$ is determined by $\Gamma(\alpha+4\tau)\Gamma(\alpha)/\Gamma(\alpha+2\tau)^2 = 5(d+2)/(5d+4)$. In this way we obtain an interesting departure from multivariate normality since the values of Mardia's skewness and kurtosis are equal to the MVN ones. Moreover, the marginal distributions of this high moment alternative are symmetric with mean 0, variance 1 and kurtosis coefficient given by $9(d+2)/(5d+4)$.

Finally, to assess the effect of data contamination we took five mixtures of two multivariate normals from the Székely and Rizzo's (2005) study. Three of them are location mixtures of the form $pN_d(0,I)+(1-p)N_d(\mu,I)$, where $\mu = (3,\ldots,3)'$ and $p = 0.5, 0.79, 0.9$, and the other two are scale mixtures of the form $pN_d(0,B) + (1-p)N_d(0,I)$, where B denotes a correlation matrix with all off-diagonal elements equal to 0.9 and $p = 0.5, 0.9$. The scale mixtures are symmetric with heavier tails than normal whereas the location mixture with $p = 0.5$ is symmetric with lighter tails than normal. The remaining location mixtures with $p = 0.79$ and $p = 0.9$ are skewed with normal kurtosis and heavier tails than normal, respectively. Similar normal mixtures have also been considered in Henze and Zirkler (1990), Romeu and Ozturk (1993) and Mecklin and Mundfrom (2005).

We used the algorithms described in Johnson (1987) and Kotz et al. (2001) to generate all the previous distributions.

4.3. Empirical power results. The empirical power results presented in the following for the considered MVN tests are based on 10,000 samples of different sizes ($n = 20, 40, 60, 80, 100, 200, 400$) and data dimensions ($d = 2, 3, 4, 5, 7, 10$) from the considered set of alternative distributions. The standard level of significance $\alpha = 0.05$ was used. With 10,000 repetitions the margin of error for approximate 95% confidence intervals for the proportion of rejections does not exceed 0.01. For the affine invariant tests, the evaluation of the critical values was based on 20,000 repetitions under the null hypothesis of MVN. The same number of repetitions under H_0 was used to estimate the first three moments of the SU test statistic