JOURNAL OF MULTIVARIATE ANALYSIS 46, 309–334 (1993)

A Comparative Study of Goodness-of-Fit Tests for Multivariate Normality*

JORGE LUIS ROMEU AND AYDIN OZTURK

CASE Center of Syracuse University

A Monte Carlo power study of 10 multivariate normality goodness-of-fit tests is presented. First, multivariate goodness-of-fit methods and non-normal alternatives are classified according to their characteristics. Then, a measurement tool is defined, validated, and used to assess the performance of the methods, which are then ranked by type of alternative they best detect. Finally, Monte Carlo-derived empirical critical values for the 8 procedures, valid when samples are too small to invoke asymptotic theory, are provided.

1. INTRODUCTION AND MOTIVATION

This study describes a Monte Carlo power comparison of 10 goodness-of-fit (GOF) tests for multivariate normality (MVN). We compare 2 implementations of the recently developed multivariate Q_n procedure [22] with a selection of the 8 *best* competitor methods found in the literature.

The new Q_n procedure is a multivariate extension of the univariate Q_n test [20, 21]. It is based on performing an orthogonal transformation on the original, p-dimensional, multivariate data to obtain a new set of p independent and identically distributed univariate samples. Then, we apply the univariate Q_n procedure for the several samples case to this transformed set of multivariate data. And the multivariate hypothesis of normality can now be equivalently tested using the multisample univariate Q_n.

To assess the power of the multivariate Q_n we selected the *best* 8 MVN

Received September 27, 1991; revised October 27, 1992.

AMS 1980 Mathematics Subject Classification (1985 Revision): Applied Statistics.

Key words and phrases: multivariate normality, goodness-of-fit, performance evaluation, Monte Carlo studies.

* J. L. Romeu is with the Department of Mathematics, SUNY-Cortland and A. Ozturk is with the Computer Engineering Department, Ege University, Izmir, Turkey. Research was performed under a Dr. Nuala McGann Drescher UUP/NYS SUNY Faculty Award and NSF Grant. Critical values were computed under a grant from the Cornell National Supercomputer Facility, at Cornell's vector processor. Support from the CASE Center of Syracuse University is gratefully acknowledged.

0047-259X/93 $5.00

GOF methods. We first developed a classification scheme, then divided the existing methods into six groups, and finaly selected at least one representative from each class for the comparison.

In Section 2, we justify and describe the new *classification scheme* of multivariate normality tests, derived for this study. Then, based on this classification scheme and on each method's theoretical advantages and disadvantages, we select the *best* 8 procedures. In Section 3, we discuss and justify the statistical distributions selected as *alternatives* to the *null hypothesis of multivariate normality*. And we describe their implementation. Section 4 discusses a merit scale devised to (i) assess the performance of the 10 procedures compared and (ii) classify the alternative distributions selected. Section 5 presents the experimental results of our power study. We compare all the methods with respect to the effect of sample size, number of p-variates, tail probability, covariance structure (not all methods are covariance invariant), and alternative chosen. Then, using the merit scale defined in Section 4 as a measurement tool, the multivariate tests are classified and ranked. Section 6 discusses and provides empirical critical values for the 8 best methods compared, valid when samples are not large enough to invoke asymptotic theory. Finally, in Section 7 we summarize our results.

2. Rationale for Method Selection

Currently, there are not many GOF tests for multivariate normality and those that exist are highly constrained. Some are extremely complex to implement, so much so that they can handle only three or four p-variates. Others require extensive numerical work to solve sets of nonlinear equations, posing serious convergence problems. Still other have unknown or unsatisfactory statistical properties. Finally, several exhibit more than one problem at the same time. It was not feasible to include them all in our study. To select objectively among them, we reviewed three classification schemes in the literature. The first one, by Gnanadesikan [6], considers three categories: essentially univariate, joint normality, and *views* methods. The second, by Cox and Small [3], considers two: coordinate dependent and invariant methods. Finally, Koziol [13] also considers two catagories: properly multivariate procedures and generalizations of univariate procedures.

However, we found too much overlap among these classes to be of use in our selection process. Therefore, we devised a new classification scheme by dividing the existing methods into six classes, each with certain common characteristics:

2.1. *Essentially Multivariate Procedures*

These include methods *based on general multivariate theoretical concepts*, among them our multivariate Q_n procedure, Mardia's methods of skewness and kurtosis, and Foutz's Equivalent Statistical Blocks (ESB) method.

The multivariate Q_n procedure is defined in the following way: Let $\{\mathbf{X}_i\}$, $i=1, ..., n$, be a p-variate sample of n identically distributed and independent random vectors, having mean vector μ and covariance matrix Σ. The multivariate Q_n procedure tests whether this sample comes from a multivariate normal distributuion in two steps:

(1) Performs a linear transformation,

$$\{(X_1, ..., X_p) \sim MVN_p(\mu, \Sigma)\} \mapsto \{(Z_1, ..., Z_p) \sim MVN_p(0, I_p)\},$$

where

$$\mathbf{X} = CZ + \mu \qquad \text{and} \qquad CC' = \Sigma.$$

(2) Tests these resulting p *univariate independent samples*, Z_j, $j=1, ..., p$ for joint normality using the multisample univariate Q_n procedure.

That is, first the test transforms the p-variate sample $\mathbf{X}_i$, $i=1, ..., n$, to p independent and identically distributed normal standard samples of size n, via the matrix C (equal to the Cholesky decomposition of Σ for our Cholesky version and to $\Sigma^{1/2}$ for our Sigma Inverse version).

Then, the test rejects $H_0\colon \mathbf{X} \sim MVN(\mu, \Sigma)$, at level α if

$$\max_{1 \leqslant j \leqslant p} \exp\left\{-\frac{1}{2}\left\{\frac{(U_{n,j} - \mathscr{E}(U_j))^2}{\sigma_U^2} + \frac{(V_{n,j} - \mathscr{E}(V_j))^2}{\sigma_V^2}\right\}\right\} < 1 - (1-\alpha)^{1/p}, \quad (1)$$

where

$$U_{n,j} = \frac{1}{n}\sum_{i}^{n} \cos\theta_i\, |Z_{i,j}|$$

and

$$V_{n,j} = \frac{1}{n}\sum_{i}^{n} \sin\theta_i\, |Z_{i,j}|$$

with

$$\theta_i = \pi \int_{-\infty}^{m_{i:n}} \frac{1}{\sqrt{2n}} \exp\frac{t^2}{2}\, dt$$

and $m_{i:n}$ is the ith order statistic from the standard normal distribution.

For, under the null hypothesis $H_0 : \mathbf{X} \sim MVN(\mu, \Sigma)$, the statistics

$$g(Q_n^{(j)}) = g(U_{n,j}, V_{n,j}) = \frac{U_{n,j}^2}{\sigma_U^2} + \frac{V_{n,j} - \mathscr{E}(V_j))^2}{\sigma_V^2} \qquad 1 \leqslant j \leqslant p \qquad (2)$$

are approximately distributed as a Chi Squared with 2 degrees of freedom [30]. And $g(Q_n^{(j)}) = -2 \ln \alpha$ (or $\exp\{-\frac{1}{2} g(Q_n^{(j)})\} = \alpha$) provides an approximate, fast converging, test of size α.

Foutz's method [4] is a generalization of the Chi Squared GOF test. It is based on ESB, a multivariate concept analogous to that of class intervals in univariate statistics. Hence, it perpetuates the Chi Squared GOF test problem of class definition and dependence.

Mardia [15] and Mardia *et al.* [18] established a multivariate analytical equivalent of univariate skewness and kurtosis by analyzing $\mathrm{Corr}(\bar{\mathbf{X}}, S^2) \sim \{\beta_1/(\beta_2 - 1)\}^{1/2}$. Mardia regards this correlation as a measure of univariate skewness and extends β_1 to the multivariate case: $\beta_{1,p} = n^2 \sum_{ijk}^{p} \{\mathrm{Cov}(\bar{\mathbf{X}}_i, S_{jk})\}^2$.

Mardia shows that such $\beta_{1,p}$ is invariant under orthogonal transformations, and that its sample statistic

$$b_{1p} = \frac{1}{n^2} \sum_{rs}^{n} g_{rs}^3, \qquad 1 \leqslant r, \quad s \leqslant n, \qquad (3)$$

where $g_{rs} = (\mathbf{X}_r - \bar{\mathbf{X}})' S^{-1} (\mathbf{X}_s - \bar{\mathbf{X}})$, is asymptotically distributed: $\{nb_{1p}/6\} \to \chi_v^2$, as $n \to \infty$ with $v = \{p(p+1)(p+2)\}/6$ degrees of freedom. In a similar way, Mardia redefines the measure of kurtosis as $\beta_{2,p} = \mathscr{E}\{(\mathbf{X}_i - \mu)' \Sigma^{-1} (\mathbf{X}_i - \mu)\}^2$, $i = 1, \ldots, n$, which is the expected value of the squares of Mahalanobis distances to the mean. Mardia shows that $\beta_{2,p}$ is also invariant under orthogonal transformations and that the statistic

$$b_{2p} = \frac{1}{n} \sum_{i}^{n} \{(\mathbf{X}_i - \bar{\mathbf{X}})' S^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})\}^2 \qquad (4)$$

tends asymptotically to $\mathcal{N}(\mu, \sigma^2)$ with $\mu = \{p(p+2)(n-1)\}/\{n+1)\}$ and $\sigma^2 = \{8p(p+2)\}/n$. Based on these measures, Mardia proposed two tests (skewness and kurtosis) for multivariate normality. We selected both tests for our Monte Carlo comparison.

2.2 *Multivariate Methods of Marginal Analysis*

These essentially multivariate procedures are *based on some properties of the marginal distributions.* They include Royston's extension of Shapiro and Wilk, Small's skewness/kurtosis, and the Box–Cox transformation methods of marginal analysis.

Small's method [28] requires performing a Johnson [8] S_U transformation on each p-variate. After such transformation, the resulting marginal skewness and kurtosis become, respectively, quasi independent and identically distributed as normal standard. Such transformations require lengthy estimation and search processes for each iteration of the simulation. Box–Cox marginal power transformation [6] poses the same inefficiency problems. Individual searches on each p-variate would yield the initial values for implementing one of the known stepwise optimization procedures, to find the overall or joint optimal parameter Λ.

Royston's W, on the other hand, offered several advantages [25]. First, it is an extension of the *univariate* Shapiro and Wilk W [26] test, which is known to be an omnibus test and has an excellent track record. Then the optimization process was performed by Royston, who furnished a sort table for the necessary parameters. Royston's W is based on performing, first, p marginal Shapiro and Wilk test statistics, say W_j, $j=1, ..., p$ (which are not independent), and then a Box–Cox type of transformation f, to these W_j (say $z_j = f(W_j)$, $j = 1, ..., p$) to make them approximately standard normal. Then the statistic $K_j = \{\Phi^{-1}\{\frac{1}{2}\Phi(-z_j)\}\}^2$, $j = 1, ..., p$, where Φ is the standard normal *cdf*, is calculated.

Royston shows that his multivariate statistic

$$G = \frac{1}{p}\sum_{i}^{p} K_j \sim \chi_e^2, \qquad 0 < e \leqslant p, \tag{5}$$

where e represents the *equivalent* degrees of freedom, $e = p/\{1 + (p-1)\bar{c}\}$, and $\bar{c} = \{1/p(p-1)\}\{\sum_{i \neq j}^{n} c_{ij}\}$. Royston provides a correction coefficient for the correlation c_{ij} between X_i and X_j, with tabulated parameters λ, μ, and v, dependent on the sample size. We also chose Royston's method for our comparison.

2.3. *Regression Methods*

Include procedures that *regress some functions of the p-variate observations*. One such procedure is the Andrews method [6], which requires, first, a probability integral transformation for each coordinate element; then, a search for each p-dimensional point of the hypercube, to find the minimum distance to the nearest neighbor; finally, the computation of an exponential function of this distance and the regression of this function on the coordinate values of the unit hypercube.

The other is Cox and Small's procedure [3], where each p-variate X_j is regressed directly on all remaining others. For $p = 2$, the statistics Q_{jk}, $j, k = 1, 2, j \neq k$, are a function of the regression statistics for the coefficients of the squared terms in the regression:

$$\begin{gathered} X_{kl} = \delta_0 + \delta_{1j} X_{jl} + \delta_{2j} X_{jl}^2 + \eta_{kl}, \qquad \eta_{kl} \sim N(0, \sigma_\eta); \\ l = 1, ..., n; \quad j, k = 1, 2; \quad j \neq k. \end{gathered} \tag{6}$$

For $p=2$ Cox and Small's statistic $(Q_{21}, Q_{12}) \sim MVN_2(0, \Sigma)$. Hence, asymptotically,

$$(Q_{21}, Q_{12}) \begin{pmatrix} 1 & r_{12}(2-3r_{12}^2) \\ r_{12}(2-3r_{12}^2) & 1 \end{pmatrix}^{-1} \begin{pmatrix} Q_{21} \\ Q_{12} \end{pmatrix} \to \chi_2^2. \tag{7}$$

To implement this method for, say, $p=4$ or $p=8$ requires the solution of 4 and 8 regressions, of 9 and 35 terms, respectively. Mardia [19] states that for the case $p>2$, these statistics can be easily extended and are given in Cox and Small [3]. We compared it for $p=2$ only.

2.4. *Methods Based on the Union–Intersection Principle*

These multivariate procedures are *based on the extension of Roy's univariate principle.* They include tests based on extensions of skewness, kurtosis, and Shapiro/Wilk W, developed by Malkovich and Afifi [14] and generalized in the following way. If (i) $X_i \sim MVN_p(\mu, \Sigma)$, $i=1, ..., n$, and (ii) the two inequalities

$$\beta_1(c) = \frac{[\mathscr{E}\{c'X - c'\mathscr{E}(X)\}^3]^2}{(V(c'(X))^3} > 0 \qquad \text{and}$$
$$[\beta_2(c)]^2 = \left[\frac{\mathscr{E}\{c'X - c'\mathscr{E}(X)\}^4}{\{V(c'X)\}^2}\right]^2 > 9 \tag{8}$$

hold for some p-dimensional vector $c \neq 0$, then the corresponding *union–intersection* multivariate measures are defined:

$$\beta_1^* = \max_c \beta_1(c) > 0 \qquad \text{and} \qquad (\beta_2^*)^2 = \max_c [\beta_2(c) - 3]^2. \tag{9}$$

The generalized union–intersection Shapiro and Wilk W procedure was selected for our power comparison,

$$W^* = \frac{[\sum_i^n a_i u_{(i)}]^2}{(\mathbf{X}_m - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_m - \bar{\mathbf{X}})}, \tag{10}$$

where $\mathbf{X}_m$ is the observation vector for which $(\mathbf{X}_m - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}) = \max_{1 \leqslant i \leqslant n}\{(\mathbf{X}_i - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$, the a_i's are the Shapiro and Wilk tabulated coefficients, $u_{(i)} = (\mathbf{X}_m - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$, $i = 1, ..., n$, and $(\mathbf{X}_m - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}) \mapsto (\mathbf{X}_m - \bar{\mathbf{X}})' A^{-1}(\mathbf{X}_m - \bar{\mathbf{X}})$.

The hypothesis of multivariate normality is rejected if $\min_c W(c) < K_w$, where the vector c satisfies the two conditions $c'(\mathbf{X}_1 - \bar{\mathbf{X}}) = (n-1)/(na_1)$ and $c'(\mathbf{X}_i - \bar{\mathbf{X}}) = -1/na_1$, $i=2, ..., n$.

2.5. *Geometrical Methods*

These methods are *based on a geometric transformation of the coordinate system*. They include Koziol's radii method, Koziol's angles method, and Andrews' graphical method. The first two were chosen for our Monte Carlo study since they were based on complementary criteria.

Andrews' method [6] is informal and purely graphical and serves as the conceptual basis for subsequent, more analytical work. It is constrained in the number of variables it can handle. For the case ($p=2$), each bivariate observation, (X_{1i}, X_{2i}), $i=1, ..., n$, is transformed to polar coordinates: *radius* r_i and *angle* ϕ_i. Angles are measured with respect to a fixed, arbitrary line, taken as the axis of the abscissa, and are distributed uniformly on $(0, 2\pi)$. Radii are approximately distributed as χ_2^2. Several types of graphs can be plotted and informal tests then performed.

Koziol's radius and angles methods are geometrically inspired by the one above. However, they are analytically derived and have no constraints with respect to the number of p variates they can handle. Koziol's [11] radii method is based on Andrews informal method. Invoking weak convergence, Koziol provided a formal distribution theory, $\mathscr{F}(r_i^2) \rightarrow \chi_p^2$ as $n \rightarrow \infty$, and defined a Cramer–von Mises type of statistic $Y_i = (\mathbf{X}_i - \bar{\mathbf{X}})' S^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$, $i=1, ..., n$, where $z_i = \mathscr{F}(Y_i)$ for, $\mathscr{F} \equiv \chi_p^2$, and

$$J_n = \sum_i^n \left\{ z_{(i)} - \left(\frac{(i-1/2)}{n} \right) \right\}^2 + (12n)^{-1}. \tag{11}$$

Via a Monte Carlo study, Koziol shows that the χ_p^2 critical values are reasonable for p small or n large. Koziol's angles method [12] is also based on the weak convergence principle, where the underlying empirical process can be approximated by a p dimensional Gaussian stochastic process. Koziol then modifies the *Rayleigh* angles test (Mardia *et al.* [18]) in the following way to assess the uniformity of the angles in the hypersphere. Let

$$l_i = \frac{\Sigma^{-1/2}(\mathbf{X}_i - \mu)}{[(\mathbf{X}_i - \mu)' \Sigma^{-1}(\mathbf{X}_i - \mu)]^{1/2}}, \qquad i = 1, ..., n. \tag{12}$$

Then, l_i, $i=1, ..., n$ are i.i.d. with mean vector 0 and covariance matrix $p^{-1}I_p$. Hence, $R = n^{-1/n} \sum_i^n l_i \rightarrow MVN_p(0, p^{-1}I_p)$, as $n \rightarrow \infty$. Therefore, the *Rayleigh statistic*, $pR'R$, tends to χ_p^2 as $n \rightarrow \infty$. And, if μ and Σ are replaced by their respective estimate $\bar{\mathbf{X}}$, S in the above equations, Koziol shows (via Monte Carlo for p small or n large) how the estimate of Rayleigh's R, say R^* exhibits the property

$$R^* \to MVN_p(0, V), \qquad \text{as} \quad n \to \infty \qquad \text{with}$$
$$V = p^{-1}\left[1 - \left(\frac{2}{p}\right)\left\{\frac{\Gamma((p+1)/2)}{\Gamma(p/2)}\right\}^2\right]. \tag{13}$$

It is worth noting that Koziol's method is a *conceptualization* of Andrew's graphical geometrical approach. Koziol actually derived, using weak convergence properties and distributional theory, an asymptotic distribution for his statistic. These issues differentiate Koziol's radii from those discussed next.

2.6. *Projection Methods*

Projection methods *use a reduction approach* (e.g., Mahalanobis distance) *to project the multivariate observations into a one-dimensional space*. They include Malkovich and Afifi's Cramer–von Mises (CVM) and Kolmogorov–Smirnoff (K–S) projection methods, Andrews' directional method and Hawkins', Anderson–Darling projection method.

Malkovich and Afifi [14] use the sample Mahalanobis distance to reduce the dimensionality of the problem from p to 1. Then they apply either the K–S or the CVM univariate GOF tests. Since the asymptotic distributions of the resultant empirical statistics are unknown, their critical values are obtained by Monte Carlo.

Andrews' directional projection method also reduced the problem from p dimensions to one dimension via a projection. However, the direction of the unidimensional projection is not known but is selected to maximize the non-normality of the data. The key issue remains finding the direction where non-normality will be apparent.

The Hawkins [7] method, chosen for our power comparison, builds on the work of Malkovich and Afifi [14] and proposes an exact rather than an empirical distribution as a basis for the test. In addition, the Hawkins test is conceived not only as a MVN GOF procedure, but also as one that can be used to test for homoscedasticity, when in the presence of several (say g) multivariate groups.

Hawkins defines the statistic $V_{jk} = (\mathbf{X}_{jk} - \bar{\mathbf{X}}_k)' \mathcal{S}_k^{-1}(\mathbf{X}_{jk} - \bar{\mathbf{X}}_K)$, where k runs on the h groups of size n_k and $\mathcal{S}_k$ is the sample covariance matrix of the kth group. Let $\mathcal{S}_*$, $\bar{\mathbf{X}}_*$ denote the covariance matrix and mean vector of group k, resulting from the removal of X_{jk}, from the sample. Since we are concerned only with the GOF of *a single multivariate sample* let $g = 1$ and $v = n - g - 1 = n - 2$. Following Hawkins, we have that $\mathbf{X}_{jk} - \bar{\mathbf{X}}_* \sim N(0, (n_k/(n_k - 1))\Sigma)$ is indepenent of $v\mathcal{S}_* \sim W(\Sigma, v)$, the Wishart distribution with v degrees of freedom.

Then (since $g = 1$), dropping the second subindex, the variable $T^2 = (n-1)(\mathbf{X}_i - \bar{\mathbf{X}}_*)' S_*^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}_*)/n$ follows a Hotelling distribution, and

$F_i = \{(\nu - p - 1)\, T^2\}/\nu p\} \sim F(p, \nu - p + 1)$, $i = 1, ..., n$. Hawkins then shows, after some algebraic manipulations, that

$$F_i = \{(n - p - 1)\, nV_i\}/\{p\{(n-1)^2 - nV_i\}\}. \tag{14}$$

Then, letting $A_i = \mathscr{P}\{F > F_i\}$ denote the tail area of F_i under this $F_{p,\, \nu - p + 1}$ distribution, Hawkins states that, under H_0, A_i is distributed as a Uniform variate, say $U(0, 1)$, and uses the Anderson–Darling statistic to test for multivariate normality.

3. The Selection of Alternative Distributions

Careful selection of statistical alternatives is a key issue in GOF power studies because (i) there are different ways in which a distribution can depart from multivariate normality and (ii) nor every GOF approach may be equally apt to discover all types of departures. For examples, the alternative distribution can be skewed, instead of symmetrical, and can be more peaked or flatter than the standard normal. The distribution can be both skewed and peaked or flat. The sample can come from more than a single population (data contamination) or the distribution can be non-normal but closely related to the normal (quasi-normal). We considered such situations in our selection of statistical alternatives, in the following way:

To assess the effect of pure skewness on the power of the test the *Generalized Lambda Distribution* (GLD) families, with lambda parameters corresponding to $\sqrt{\beta_1} = 0.89$, $\beta_2 = 3.2$ (for severely) and $\sqrt{\beta_1} = 0.4$, $\beta_2 = 3.0$ (for moderately skewed), were selected. To assess the effect of flat distributions (platykurtic), we selected the uniform $U(0, 1)$ (severe) and the GLD for $\sqrt{\beta_1} = 0$, $\beta_2 = 2.4$ (moderate). For long tailed distributions (leptokurtic) we used the t_8, Student t with 8 degrees of freedom. To assess the joint effect of skewness and kurtosis on the power of the test, we selected the χ^2_{10}, commensurate with the parameters of the first GLD and the t_8 defined above. All these five multivariate distributions were obtained by (i) generating the above defined marginal distributions directly (using the IMSL routines) and (ii) combining them via the correlation coefficient ρ.

To assess the effect of data contamination on power, we selected parameters p_0 and ρ from the model:

$$p_0 MVN_p(\mu_1, \Sigma_1) + (1 - p_0)\, MVN_p(\mu_2, \Sigma_2). \tag{15}$$

This model yields many possible combinations, even with this restricted form. Based on a study by Johnson [8] qnd seeking a mildly versus a severely contaminated alternative, we chose $\mu_1' = (0, ..., 0)$ and $\mu_2' = (1, ..., 1)$ and coveriance matrices as

$$\Sigma_i = \begin{pmatrix} 1 & \rho_i & \cdots & \rho i \\ \cdots & \cdots & \cdots & \cdots \\ \rho i & \cdots & \rho i & 1 \end{pmatrix} \qquad i = 1, 2, \tag{16}$$

where $\rho_1 = 0.5$ and $\rho_2 = 0.9$. The mixing parameter p_0 was chosen as 0.5 for severely, versus 0.9 for mildly contaminated data.

To assess the effect of an alternative with normal marginals, we selected Morgenstern and Kinchine distributions. A necessary though not sufficient condition for multivariate normality is that all marginals be univariate normally distributed. Moregenstern and Kinchine distributions with parameters $\gamma = 0.5$, 1.0, were generated, for the bivariate case, by a three step algorithm (Johnson *et al.* [9]). For the case $p > 2$, elliptically contoured distributions [1], which tend rapidly to marginal and joint normality, were used to assess a similar effect. To assess quasi normality, we selected two elliptically contoured distributions: Pearson's types II and VII [2, 8, 10], with parameter $m = 10$. Both these distributions are close to being multivariate and marginally normal, for $p \geqslant 4$.

Bivariate regression, a specialized alternative, was included to study the Cox and Small procedure. If $Y_1, Y_2 \sim N(0, 1)$ and $d \in \mathcal{R}$, and we let $X_1 = Y_1$ and $X_2 = Y_1 + dY_1^2 + Y_2$, then (X_1, X_2) is distributed as a *bivariate regression* with parameter d.

To assess the effect of scale, location, and covariance on power, we used the same covariance matrices defined in (16). And to compare these methods on a true parity basis, we also simulated the null (i.e., multivariate normal) with covariance matrix as in (16). We obtained empirical 90th, 95th, and 99th percentiles for each test statistic (two sided for Mardia's Kurtosis) for selected sample sizes n and number of p-variates. These empirical critical values were, in general, quite different from those of the tabulated asymptotic distributions and will be discussed in Section 6.

4. The Measurement Tool

Since the ways in which a multivariate distribution can depart from normality are many and complex, we sought to define areas where one (or more) method(s) would tend to perform better. To define such areas we classified the multivariate statistical alternatives defined in Section 3, according to their location in the bivariate skewness vs kurtosis plane (Fig. 1).

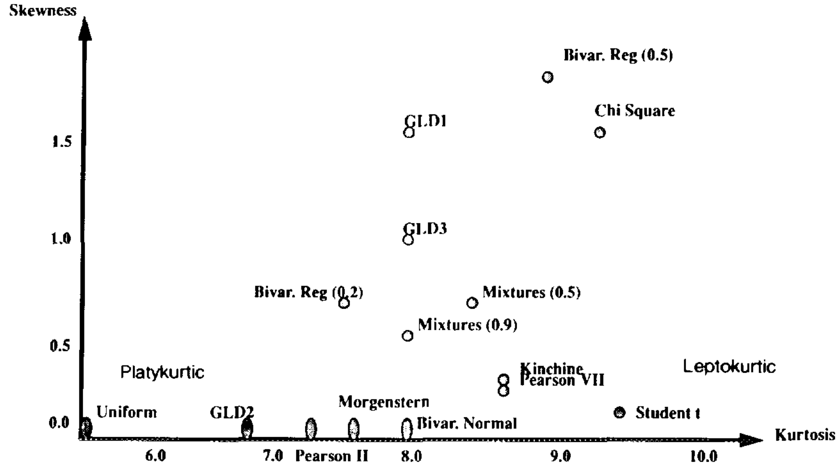

FIG. 1. Schematic of statistical alternatives to Bivariate normal, with respect to skewness and kurtosis.

4.1. *Classification of the Alternative Distributions*

Departing from coordinates (8.0, 0.0) of the skewness vs. kurtosis plane (values for the bivariate normal) we moved away in every direction (following a trajectory parallel to the coordinate axes). We then assessed each statistical alternative degree of non-normality based on the length of its trajectory to the coordinates of the bivariate normal. We grouped our 15 statistical alternatives, by the degree in which they departed from normality, into four increasing classes or levels: *quasi*, *mildly*, *moderately*, *and severely non-normal.* In addition, we considered the way in which these distributions departed from normality: *skewness*, *kurtosis*, *or a combination* of these problems (Table I).

2. *The Merit Scale*

To aid in objectively comparing the ten MVN GOF procedures selected, we defined a *merit scale*. Ordinal in nature, this measurement tool has three possible scores: (i) score 0, if the performance of a given parameter is assessed as poor, (ii) score 1, if the performance is assessed as intermediate, and (iii) score 2, if it is assessed as good. For example, to assess the empirical power levels of the tests, in each of our statistical alternatives, we gave a score 0 if the test attained a low power, score 1 if an intermediate power, and score 2 if a high power.

Since our comparison was performed under several, sometimes conflicting, qualitative and quantitative *criteria*, we divided these criteria into two groups:

TABLE I

Statistical Alternative Class Scheme

Alternative	Parameter	Severity class	Shape class	Merit ranking
Chi square	$d.f. = 10$	Severe	Combined	15
Bivar. reg.	$d = 0.5$	Severe	Combined	14
GLD-1	$\sqrt{\beta_1} = 1.5$	Severe	Purely skewed	13
Uniform	$U(0, 1)$	Severe	Purely kurtic	12
Mixtures	$p = 0.5$	Moderate	Combined	11
Student t	$d.f. = 8$	Moderate	Combined	10
Bivar. reg	$d = 0.2$	Moderate	Combined	9
GLD-3	$\sqrt{\beta_1} = 0.7$	Moderate	Purely skewed	8
GLD-2	$\beta_2 = 6.8$	Moderate	Purely kurtic	7
Pearson VII	$m = 10$	Mild	Combined	6
Kinchine	$\gamma = 0.5; 1.0$	Mild	Combined	5
Mixtures	$p = 0.9$	Mild	Purely skewed	4
Pearson II	$m = 10$	Mild	Purely kurtic	3
Morgenstern	$\gamma = 1.0$	Mild	Purely Kurtic	2
Morgenstern	$\gamma = 0.5$	Quasi-normal	Quasi-normal	1
Bivariate normal	Standard	Normal	Normal	0

(1) Those that measure general (qualitative) characteristics of a test.

(2) Those that measure performance (power of a test) under different statistical alternatives.

The following examples of general characteristics are expressed as opposites:

(1) The test can handle any number of p-variates (scoring 2) or exhibits algorithmic problems that seriously hinder the number of p-variates it can handle (scoring 0).

(2) the test exhibits a very small correlation effect (score 2) or exhibits such a dependence that the power s corrupted (score of 0).

(3) The test is practically sample size independent (i.e., the statistic converges very rapidly to its asymptotic distribution) or requires a very large sample to attain its asymptotic values.

The *merit scale* was first validated using our statistical alternative classification scheme: we reversed the roles of the merit scores matrix values, now adding them by statistical alternatives instead. The four levels of non-normality and two qualitative types of alternatives (*skewness* and *kurtosis*) defined in Section 4.1 were reproduced and the non-normal statistical alternatives were reclassified and ranked in them as in our original scheme. These results (i) validated our merit scale and (ii)

TABLE II

Classification and Ranking of MVN GOF Procedures

Procedure	Specialization	Correlation	p-variates	Convergence	Rank
M-skew	Skewness	Independent	Unconstrained	Slow	1
Q_n-Cholesky	General	Independent	Unconstrained	Fast	2
M-kurt	Kurtosis	Independent	Unconstrained	Very slow	3
Koziol χ^2	Kurtosis	Independent	Unconstrained	Empirical	4
Hawkins	Kurtosis	Independent	Unconstrained	Empirical	4
Royston W	General	Dependent	Unconstrained	Slow	6
Malkovich	Skewness	Independent	Unconstrained	Empirical	7
Q_n-Sigma	General	Dependent	Unconstrained	Fast	8
Cox–Small	Mild depart.	Dependent	Constrained	Slow	9
Koziol angles	Skewness	Dependent	Constrained	Slow	10

provided an experimental framework where the *best* procedures for each type of alternative could now be analyzed.

Then, we used the validated merit scale to compare an rank the 10 multivariate procedures in the context of this multi-criteria problem. In Table II we show the procedure rankings, obtained *overall*, by combining the *qualitative characteristics* and *power merit scores* into a single figure-of-merit. We also classified the procedures by *type of departure* they specialize in, and recorded their characteristics with respect to *correlation, p-variates, and convergence rate.*

5. Experimental Results

In this section, results are summarized by type of non-normal alternative and individual procedure. For details, see Romeu [23].

5.1. *Analysis by Type of Non-normal Departures*

Classification and ranking of the MVN GOF methods by the type of departure they best detect and specialize in:

(1) Methods that perform better under severe departures of both type (skewness and kurtosis): our *Cholesky implementation of* Q_n, *Sigma, and Royston's methods.*

(2) Methods that perform better under severe and moderately kurtic-prone alternatives: *Mardia's Kurtosis, Hawkins and Koziol's Chi Square tests.*

(3) Methods that perform better under severe and moderately skewed-prone alternatives: *Mardia's Skewness, Malkovich–Afifi, and Koziol's Angle* methods.

(4) Method that performs better under very mild skewness departures from normality: *Cox and Small.*

Table III shows our non-normal alternatives class scheme with the corresponding *best* GOF methods in each. Note the clear distinction in procedure specialization. Under purely and severely skewed (kurtic) entries of Table III, we find the MVN GOF methods which are, respectively, skewed (kurtic) prone. Their specialization is such that skewed-prone methods fail to detect severe (but *purely*) kurtic departures from normality, and viceversa. Note, however, that the multivariate Q_n detects all types of moderate and severe departures from normality with good power.

TABLE III

Classification of Procedures within Best Alternatives

Severity	Skewness	Kurtosis	Combined
Mildly non-normal	Mixtures $p = 0.9$	Morgenstern $\gamma = 1.0$	Kinchine $\gamma = 1.0$
	Cox–Small	Royston W	Pearson VII
	Mardia-skew	Pearson II	Mardia-skew
		Mardia-kurt	Mardia-kurt
	Koziol χ^2	Cox-small	
		Hawkins	
Moderately non-normal	GLD-3	GLD-2	BIV.REG. $d = 0.2$
	Mardia-skew	Mardia-kurt	Mixtures $p = 0.5$
	Q_n Cholesky	Koziol χ^2	Student t_8
	Royston W	Royston W	Mardia-skew
	Koziol angle	Hawkins	Mardia-kurt
	Q_n Sigma	Q_n Cholesky	Cox–Small
			Hawkins
Severely non-normal	GLD-1	Uniform	Chi-square
	Mardia-skew	Q_n Cholesky	Mardia-skew
	Q_n Cholesky	Royston W	Q_n Cholesky
	Royston W	Mardia-kurt	Royston W
	Koziol angle	Koziol χ^2	Malkovich
	Malkovich	Hawkins	Koziol angle
	Q_n Sigma		BIV.REG $d = 0.5$
			Mardia-skew
			Cox–Small
			Q_n Sigma

5.2. *Analysis by Individual Procedures*

Based on all the above factors we summarize the power study results, for each of 10 procedures:

I. Methods that Detect General Severe/Moderate Non-normality.

(1) Cholesky-Implementation of Q_n: Handles an unconstrained number of p-variates and sample sizes; converges rapidly to its asymptotic distributution; power rises steadily with n and p and exhibits no correlation effect or algorithmic problems. Detects with good power all types of severely and moderately non normal alternatives analyzed. Cholesky implementation of Q_n rates second best in the overall ranking of the 10 methods compared.

(2) Sigma Inverse Implementation of Q_n: Similar to Cholesky, but with lower power and dependence on the underlying (and unknown) correlation structure Power increases slowly with p and n. Sigma ranks eighth in our rating of the 10 methods.

(3) Royston's W: Also handles any number of p-variates and sample sizes, but is severely affected by correlation. Works best when correlation and sample sizes are small, but these factors are usually not under the control of the investigator. Royston's ranks sixth in our overall evaluation of 10 ten tests.

II. Skewness-Prone Methods.

(1) Mardia's Skewness: Handles any number of p-variates and sample sizes; power rises steadily with n and p and converges slowly to its asymptotic distribution, requiring empirical critical values for n small. Fails to detect pure kurtosis problems, but is excellent with skewed or combined alternatives of all severity degrees and is not affected by correlation. Mardia's Skewness ranks first in our overall method classification.

(2) Malkovich and Afifi's W: Test behavior is similar to that of Mardia's Skewness, but with a lower power. Requires empirical critical values, which are very sensitive to sample size. It ranks seventh in our overall classification.

(3) Koziol's Angles Method: Has severe difficulties in handling more than three p-variates due to algorithmic problems (this difficulty increases with p) and is affected by correlation. Otherwise, it behaves like Mardia's Skewness test,with higher power for n small. However, given its algorithmic problems (see next section), its use is not recommended. Koziol's Angles test ranks last in our overall classification.

III. Kurtosis-Prone Methods.

(1) Mardia's Kurtosis: Handles any number of p-variates and sample sizes; tends very slowly to its asymptotic distribution, requiring the use of empirical critical values even when n is moderate. Fails to detect pure skewness departures; power rises steadily with n and p. As for the Skewness test, it is very stable and reliable, ranking third in our overall method classification.

(2) Koziol's Chi Square Method: Similar to Mardia's Kurtosis test in behavior, but with lower power. It requires the use of empirical critical values, ranking fourth in our overall classification.

(3) Hawkins' Method: Easy to implement projection method, also similar to Mardia's Kurtosis, with lower power. Requires empirical critical values and ranks fourth, tied with Koziol's Chi Square, in our classification.

IV. Mild Non-normality Methods.

(1) Cox and Small's Method: Difficult to implement with more than a few p-variates (two or three, feasibly) due to the large amount of computing it requires for $p > 3$; power is mildly affected by correlation, but rises slowly with n; converges slowly to its asymptotic distribution. It is excellent with mild types of non-normality. Cox–Small is a skewness-prone test with specialization on very small departures, ranking ninth among all tests analyzed.

6. Empirical Critical Values

Many of the multivariate procedures compared in our power study converged very slowly to their nominal (asymptotic) critical values. Others were affected by the underlying (and in practice unknown) correlation among the p-variates. Mardia [16, 17] performed a limited power study on such convergence problems. Malkovich and Afifi [14] compared the empirical powers of Mardia's Skewness and Kurtosis tests, among others, with their own multivariate W. They used lognormal, uniform, Student t, and mixtures of normals with various p-variates. Giorgi and Fattorini [5] compared empirical powers for both Mardia's tests, for Malkovich and Afifi's, and for several directional tests, using multivariate Chi Square and lognormal alternatives. Their results agree with ours in recommending both Mardia's skewness and kurtosis tests for n large.

6.1 *Simulation*

Since our power comparison required parity conditions under the null, we obtained, by Monte Carlo, the necessary (empirical) critical values to achieve parity.

Critical values were obtained by running 10,000 replications (for $p = 2$) and 5000 replications (for $p > 2$). Generating under the null (multivariate normal) distribution, we calculated, each time, the corresponding GOF statistics using Eqs. (3), (4), (5), (7), (10), (11), (12), and (14). Then we sorted them to obtain the corresponding percentile estimators. The simulations were implemented in Syracuse University's IBM 3090, for $p = 2, 3, 4$, and for $p = 5, 6, 8, 10$ in Cornell's National Supercomputer Facility vector processor. The values were obtained for sample sizes $n = 25(25)200$ and correlation matrices as in Eq. 16, for $\rho = 0.5, 0.9$.

Tables of results in the Appendix show the 90th, 95th, 99th percentiles (e.g. $\eta_{.90}$), by sample size n and p-variate p. At the bottom of each column, and for comparison, we also give (when available) the *asymptotic* critical value. Note how slowly most methods converge to (and how far they can be from) the asymptotic critical values for n small or medium, and/or p medium or large.

Cornell's Supercomputer made this experiment possible for $p > 4$, by reducing the increasingly large simulation time. The sets of programs submitted to both computers were indentical. There is, however, a large difference in the results for Koziol's Angles method. The reason lies in the algorithms involved (see Eq. (12)) and the accuracy of both computers in handling them.

6.2. *Discussion*

Mardia's skewness (Eq. (3)) empirical critical values (ecv) converged slowly to asmptotic values ($n = 200$), were not affected by correlation, and were hardware robust (results in both machines were similar). Mardia's kurtosis ecv (Eq. (4)) converged much more slowly ($n > 200$) but were also hardware robust and correlation independent. Cox and Small's ecv converged slowly ($n = 200$) to asymptotic values and were also harware robust.

Koziol's Angles ecv converged slowly ($n = 200$) to its asymptotic values, depended on correlation, and were severely affected by the hardware being used. Note, in Eq. (12), how eigenvalues and eigenvectors are calculated, the resulting matrices being multiplied to yield a scalar in the denominator of a fraction. The accuracy of the hardware determines the results obtained, which can greatly differ in both machines.

Royston's W (Eq. (5)) ecv were also correlation dependent, in spite of

Royston's correction for correlation. Note how, for $\rho = 0.5$, the ecv values are close to the asymptotic values (and not for $\rho = 0.9$). Apparently, some residual correlation remains, even after the correction, when ρ increases. This problem gets worse as p increases. Royston's ecv also converge slowly to asymptotic values ($n = 200$).

Malkovich and Afifi's ecv do not have asymptotic values, since the procedure is empirical. Its ecv are not affected by correlation and are machine independent. Hawkins' ecv are also correlation independent and machine robust, as are Koziol's Chi Square ecv.

Our ecv are based on 10,000 replications and are given for any correlation ρ, *except in Royston's W and Koziol's Angles procedures*. These two procedures exhibited severe algorithmic and correlation dependence problems. Their ecv are based on 5000 replications and are given individually for $\rho = 0.5$, 0.9. The accuracy of our ecv's was investigated [31, 32] using approximated 95% *confidence intervals*. Upper bounds for 10,000 replications yield 3%, 4%, 6% of the tabulated values, corresponding to the 90th, 95th, 99th percentiles, respectively.

These ecv (i) allow the practitioner to use the empirical procedures, and (ii) correct the *only serious problem* detected for both Mardia's tests: their *slow convergence* to the asymptotic values. Before using the ecv, and as an accuracy check for their validity in other hardware, a small simulation, under the null, for desired, n, p, ρ is recommended.

7. Summary

We have provided new classification schemes for the MVN GOF procedures and the multivariate nonnormal statistical alternatives. We have devised a measurement tool to compare procedure performances in this context.

We have selected, described, and empirically compared the powers of eight well known MVN GOF methods with two implementations of a newly developed statistic: the multivariate Q_n. We have established areas where each of these methods perform better or worse. And, based on all these factors, we have ranked them.

Finally, we have provided empirical (Monte Carlo) critical values, for selected (small) sample sizes and p-variates, for the eight *best* tests compared. These empirical critical values are recommended when the sample sizes and/or the number of p-variates are such that we are unable to invoke the asymptotic theory and, hence, to safely use the asymptotic distribution of the test statistics.

Appendix

TABLE A-1

Small Sample Empirical Critical Values for Mardia's Skewness Test

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	5.84	7.37	11.18	12.21	14.26	18.42	21.45	24.63	30.34	35.39	38.77	45.63	53.86	57.89	66.15	108.00	113.36	124.22	189.94	196.32	209.94
50	6.78	8.38	12.24	14.25	16.72	22.53	25.40	28.50	35.81	40.98	44.99	54.15	62.09	66.80	76.49	125.02	131.78	146.09	219.37	227.97	246.15
75	7.04	8.81	12.97	14.86	17.28	23.13	26.28	29.55	36.12	43.27	47.52	56.60	65.27	70.01	80.63	130.25	138.00	153.40	230.35	238.86	258.68
100	7.28	9.17	13.13	15.21	17.77	23.31	26.84	30.14	37.50	43.73	47.76	57.07	66.35	71.45	82.29	132.91	139.65	153.73	234.00	243.03	261.33
125	7.42	9.25	13.25	15.30	17.65	23.85	27.30	30.36	37.48	44.42	48.35	56.93	67.10	71.95	82.10	135.10	141.00	155.48	237.46	247.69	266.75
150	7.47	9.34	13.47	15.47	17.95	23.23	28.66	30.82	37.66	44.72	48.86	57.44	67.60	72.58	82.10	135.74	142.34	155.95	239.75	248.41	265.31
175	7.52	9.31	13.31	15.47	17.96	23.00	27.56	30.67	37.35	44.69	48.73	57.09	68.09	72.73	82.08	136.66	143.55	157.78	241.47	250.24	267.14
200	7.60	9.38	13.43	15.47	17.92	23.28	27.62	30.63	37.68	45.08	48.99	57.38	68.20	73.19	83.60	137.09	143.85	156.29	241.30	250.20	267.10
∞	7.78	9.49	13.28	15.99	18.31	23.21	28.41	31.41	37.57	46.06	49.80	57.34	69.92	74.47	83.51	139.56	145.98	157.66	246.60	255.19	270.48

TABLE A-2

Small Sample Empirical Critical Values for Malkovich and Afifi's Test

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	0.913	0.895	0.855	0.888	0.868	0.825	0.858	0.838	0.787	0.828	0.806	0.755	0.796	0.771	0.717	0.730	0.705	0.646	0.659	0.633	0.577
50	0.953	0.944	0.924	0.943	0.932	0.910	0.931	0.919	0.893	0.919	0.906	0.878	0.906	0.892	0.860	0.876	0.862	0.828	0.845	0.830	0.796
75	0.965	0.959	0.947	0.960	0.953	0.937	0.955	0.947	0.930	0.947	0.939	0.920	0.939	0.930	0.912	0.922	0.912	0.891	0.904	0.894	0.870
100	0.971	0.967	0.956	0.969	0.964	0.952	0.965	0.960	0.948	0.960	0.954	0.940	0.956	0.949	0.934	0.944	0.937	0.922	0.931	0.924	0.906
125	0.975	0.971	0.962	0.974	0.970	0.960	0.971	0.967	0.957	0.968	0.964	0.953	0.965	0.960	0.950	0.956	0.951	0.938	0.946	0.940	0.926
150	0.976	0.973	0.967	0.976	0.973	0.966	0.975	0.971	0.963	0.973	0.969	0.959	0.971	0.966	0.958	0.964	0.959	0.950	0.956	0.951	0.941
175	0.978	0.975	0.970	0.978	0.976	0.969	0.978	0.974	0.967	0.976	0.973	0.965	0.974	0.971	0.963	0.969	0.965	0.957	0.963	0.959	0.950
200	0.979	0.977	0.971	0.980	0.978	0.972	0.979	0.977	0.971	0.979	0.976	0.970	0.977	0.974	0.967	0.973	0.970	0.963	0.968	0.965	0.956

TABLE A-3

Small Sample Empirical Critical Values for Mardia's Kurtosis Test: Lower Values

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	−1.22	−1.33	−1.52	−1.38	−1.49	−1.67	−1.48	−1.61	−1.80	−1.61	−1.76	−2.10	−1.69	−1.79	−2.00	−1.87	−1.97	−2.16	−2.04	−2.14	−2.31
50	−1.35	−1.51	−1.75	−1.49	−1.63	−1.91	−1.58	−1.74	−2.03	−1.65	−1.79	−2.09	−1.77	−1.93	−2.20	−1.91	−2.09	−2.38	−2.04	−2.21	−2.49
75	−1.44	−1.59	−1.91	−1.55	−1.75	−2.05	−1.64	−1.84	−2.17	−1.69	−1.90	−2.27	−1.77	−1.97	−2.30	−1.90	−2.11	−2.45	−2.04	−2.23	−2.59
100	−1.44	−1.62	−1.95	−1.54	−1.75	−2.11	−1.65	−1.86	−2.23	−1.72	−1.92	−2.34	−1.78	−1.99	−2.37	−1.90	−2.10	−2.52	−2.01	−2.23	−2.61
125	−1.50	−1.67	−2.03	−1.57	−1.78	−2.15	−1.65	−1.85	−2.23	−1.73	−1.94	−2.35	−1.75	−1.96	−2.34	−1.89	−2.09	−2.53	−2.00	−2.21	−2.62
150	−1.50	−1.71	−2.12	−1.56	−1.75	−2.18	−1.64	−1.86	−2.26	−1.70	−1.92	−2.32	−1.74	−1.97	−2.37	−1.89	−2.13	−2.56	−1.96	−2.22	−2.69
175	−1.49	−1.71	−2.11	−1.59	−1.79	−2.27	−1.69	−1.90	−2.32	−1.73	−1.96	−2.39	−1.77	−2.03	−2.44	−1.86	−2.08	−2.56	−1.98	−2.19	−2.64
200	−1.52	−1.76	−2.14	−1.61	−1.83	−2.21	−1.67	−1.88	−2.33	−1.70	−1.93	−2.38	−1.78	−2.02	−2.43	−1.88	−2.13	−2.59	−1.98	−2.21	−2.67
∞	−1.65	−1.95	−2.58	−1.65	−1.95	−2.58	−1.65	−1.95	−2.58	−1.65	−1.95	−2.68	−1.65	−1.95	−2.58	−1.65	−1.95	−2.58	−1.65	−1.95	−2.58

TABLE A-4

Small Sample Empirical Critical Values for Mardia's Kurtosis Test: Upper Values

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	0.87	1.23	2.05	0.63	0.91	1.61	0.49	0.76	1.38	0.26	0.50	1.07	0.09	0.28	0.71	−0.29	−0.08	0.32	−0.65	−0.46	−0.09
50	1.21	1.60	2.58	1.06	1.45	2.38	0.94	1.28	2.05	0.77	1.11	1.85	0.63	0.93	1.57	0.39	0.69	1.27	0.14	0.40	0.96
75	1.36	1.79	2.80	1.21	1.62	2.49	1.08	1.46	2.27	1.03	1.38	2.17	0.89	1.23	1.96	0.73	1.07	1.71	0.48	0.76	1.35
100	1.43	1.85	2.91	1.35	1.78	2.63	1.25	1.64	2.48	1.16	1.57	2.31	1.06	1.41	2.08	0.84	1.16	1.83	0.65	0.96	1.53
125	1.46	1.90	2.93	1.35	1.74	2.62	1.27	1.72	2.54	1.18	1.52	2.27	1.12	1.45	2.18	0.96	1.32	1.98	0.82	1.17	1.84
150	1.51	1.94	2.81	1.46	1.85	2.74	1.35	1.70	2.59	1.28	1.66	2.44	1.18	1.53	2.26	1.02	1.41	2.30	0.90	1.25	1.93
175	1.55	2.00	2.78	1.46	1.87	2.63	1.38	1.73	2.57	1.30	1.71	2.47	1.22	1.60	2.32	1.13	1.44	2.06	0.99	1.40	1.99
200	1.54	1.95	2.99	1.48	1.89	2.77	1.41	1.75	2.55	1.32	1.66	2.44	1.30	1.69	2.38	1.13	1.49	2.12	1.01	1.38	2.06
∞	1.65	1.95	2.58	1.65	1.95	2.58	1.65	1.95	2.58	1.65	1.95	2.58	1.65	1.95	2.58	1.65	1.95	2.58	1.65	1.95	2.58

TABLE A-5

Small Sample Empirical Critical Values for Royston's W Test: $\rho = 0.5$

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	4.56	5.98	9.28	5.55	7.19	11.15	6.55	8.45	12.41	8.25	10.07	14.26	8.75	10.56	15.26	10.47	12.80	18.66	11.92	14.56	21.01
50	4.50	5.93	9.29	5.82	7.40	10.97	6.52	8.28	12.69	8.45	10.37	15.09	8.85	10.72	16.09	10.92	13.21	18.97	12.32	14.71	19.46
75	4.07	5.43	8.61	5.72	7.39	11.26	6.81	8.42	12.65	8.37	10.42	15.34	9.32	11.18	15.63	11.04	13.24	18.17	12.65	14.98	21.22
100	4.63	5.98	9.43	5.64	7.13	11.05	6.67	8.41	12.96	8.36	10.31	14.74	9.08	11.18	15.90	11.19	13.42	19.07	13.01	15.72	21.35
125	4.31	5.60	8.93	5.76	7.41	10.87	6.90	8.57	12.33	8.17	9.92	14.48	9.26	11.50	16.26	11.39	13.63	17.91	12.66	15.30	21.37
150	4.25	5.80	9.05	5.71	7.33	11.19	6.99	8.76	13.20	8.53	10.75	15.20	9.27	11.24	15.86	11.26	13.68	18.75	12.98	15.28	22.38
175	4.41	5.88	9.25	5.95	7.48	11.03	7.05	8.99	13.53	8.83	10.97	15.79	9.49	11.80	16.91	11.43	13.70	18.76	13.01	15.59	20.66
200	4.73	6.31	9.94	5.83	7.48	11.55	6.88	8.78	12.92	8.97	11.09	15.80	9.26	11.35	16.33	11.59	14.02	19.09	13.20	15.51	21.50
∞[a]	4.61	5.99	9.21	6.25	7.81	11.34	7.80	9.49	13.28	9.24	11.07	15.09	10.65	12.59	16.81	13.36	15.51	20.09	15.99	18.31	23.21

[a] Asymptotic values calculated for $\rho = 0.0$.

TABLE A-6

Small Sample Empirical Critical Values for Royston's W Test: $\rho = 0.9$

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	3.58	4.84	8.31	4.70	6.02	9.91	5.92	7.49	11.34	6.11	7.79	11.65	6.87	8.78	13.24	7.24	9.09	13.82	7.57	9.39	14.46
50	3.53	4.84	7.76	4.93	6.62	11.01	5.93	7.41	10.99	6.41	8.30	12.53	6.88	8.63	13.23	7.45	9.22	13.62	7.81	9.62	14.52
75	4.31	5.61	8.79	4.90	6.50	10.71	5.99	7.64	11.51	6.39	8.19	12.00	6.91	8.69	12.79	7.59	9.49	13.55	7.71	9.58	14.25
100	3.65	5.01	8.08	4.84	6.24	9.39	6.13	7.83	11.90	6.49	8.43	12.31	7.14	8.93	13.81	7.67	9.35	13.37	7.79	9.77	13.95
125	4.11	5.60	8.95	4.90	6.26	9.66	6.04	7.81	12.32	6.29	8.22	12.85	7.15	9.12	12.99	7.72	9.57	14.31	7.83	9.82	13.88
150	4.17	5.59	9.02	5.05	6.65	10.47	6.08	7.84	11.80	6.60	8.29	12.09	7.31	9.15	13.49	7.66	9.61	13.76	7.91	9.87	14.39
175	4.17	5.52	9.14	4.98	6.48	10.19	6.37	8.06	12.64	6.44	8.31	12.95	7.19	8.90	13.81	7.66	9.68	14.34	7.96	9.90	14.74
200	3.56	4.91	7.89	4.86	6.25	9.99	5.98	7.87	11.90	6.62	8.54	12.81	7.25	8.80	13.26	7.73	9.41	14.13	8.25	10.29	14.85
∞[a]	3.38	4.61	7.57	3.70	4.98	8.01	3.89	5.19	8.26	4.01	5.33	8.42	4.10	5.43	8.54	4.21	5.55	8.69	4.28	5.63	8.78

[a] Asymptotic values calculated for $\rho = 0.5$. Note how the empirical critical values for this table are affected by residual correlation.

TABLE A-7

Small Sample Empirical Critical Values for Koziol's Chi Squared Test

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	0.167	0.211	0.318	0.158	0.193	0.281	0.154	0.193	0.274	0.157	0.193	0.282	0.166	0.207	0.290	0.191	0.235	0.323	0.230	0.276	0.367
50	0.173	0.218	0.336	0.156	0.199	0.296	0.155	0.193	0.287	0.150	0.188	0.272	0.152	0.191	0.284	0.159	0.198	0.291	0.166	0.214	0.301
75	0.172	0.220	0.331	0.158	0.199	0.287	0.153	0.193	0.276	0.152	0.187	0.273	0.150	0.187	0.270	0.154	0.192	0.278	0.162	0.195	0.285
100	0.169	0.214	0.325	0.158	0.203	0.306	0.156	0.195	0.290	0.152	0.192	0.275	0.148	0.188	0.275	0.152	0.190	0.275	0.155	0.193	0.278
125	0.178	0.222	0.333	0.159	0.200	0.296	0.153	0.188	0.275	0.152	0.188	0.278	0.147	0.181	0.261	0.148	0.186	0.267	0.152	0.188	0.277
150	0.174	0.219	0.331	0.160	0.200	0.313	0.156	0.194	0.282	0.151	0.190	0.274	0.151	0.188	0.267	0.150	0.186	0.269	0.149	0.186	0.271
175	0.170	0.215	0.333	0.162	0.205	0.308	0.155	0.194	0.290	0.152	0.192	0.280	0.150	0.187	0.281	0.147	0.183	0.270	0.147	0.185	0.271
200	0.173	0.219	0.342	0.162	0.203	0.299	0.154	0.191	0.278	0.148	0.183	0.267	0.149	0.187	0.267	0.148	0.183	0.268	0.147	0.183	0.267

TABLE A-8

Small Sample Empirical Values for Hawkins' Test

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	1.040	1.301	1.952	0.977	1.202	1.765	0.981	1.205	1.844	0.943	1.171	1.752	0.919	1.148	1.716	0.914	1.125	1.667	0.920	1.131	1.705
50	1.066	1.332	2.004	0.990	1.236	1.840	0.986	1.219	1.777	0.951	1.147	1.664	0.936	1.158	1.636	0.928	1.134	1.630	0.913	1.123	1.711
75	1.054	1.316	1.950	0.993	1.235	1.787	0.980	1.192	1.721	0.961	1.186	1.682	0.944	1.161	1.653	0.943	1.150	1.667	0.914	1.126	1.646
100	1.044	1.299	1.917	1.014	1.242	1.847	0.989	1.224	1.799	0.967	1.175	1.687	0.950	1.162	1.700	0.935	1.146	1.638	0.924	1.115	1.577
125	1.065	1.325	1.930	0.996	1.229	1.821	0.961	1.179	1.726	0.955	1.175	1.714	0.935	1.141	1.662	0.928	1.150	1.629	0.925	1.140	1.673
150	1.051	1.298	1.955	1.002	1.255	1.888	0.987	1.238	1.761	0.965	1.192	1.707	0.950	1.181	1.675	0.942	1.141	1.648	0.918	1.143	1.648
175	1.049	1.322	1.942	1.017	1.244	1.868	0.981	1.212	1.712	0.967	1.187	1.735	0.958	1.184	1.729	0.939	1.149	1.649	0.928	1.134	1.615
200	1.057	1.317	1.957	1.000	1.242	1.784	0.975	1.188	1.722	0.942	1.165	1.638	0.956	1.177	1.705	0.936	1.152	1.617	0.924	1.115	1.595

TABLE A-9

Small Sample Empirical Critical Values for Cox and Small's Test (for $p = 2$)

Size	$n = 25$	$n = 50$	$n = 75$	$n = 100$	$n = 125$	$n = 150$	$n = 175$	$n = 200$	∞
$\eta_{.90}$	5.12	4.82	4.73	4.74	4.72	4.68	4.64	4.65	4.61
$\eta_{.95}$	6.79	6.39	6.19	6.23	6.11	6.07	6.20	6.07	5.99
$\eta_{.99}$	11.04	9.94	9.92	9.55	9.48	9.44	9.17	9.50	9.21

TABLE A-10

Small Sample Empirical Critical Values for Koziol's Angles Test: for $\rho = 0.5$

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	4.53	5.87	9.07	6.42	8.35	14.30	9.98	18.01	19.75	11.54	17.64	*	9.92	11.83	16.63	11.09	13.50	18.92	12.14	14.11	18.94
50	4.45	5.94	9.15	6.50	8.30	13.15	8.30	10.48	15.75	13.67	24.30	*	10.96	13.55	18.24	13.02	15.49	21.06	14.93	17.39	24.10
75	4.59	5.93	8.76	6.73	8.76	14.85	9.88	13.81	66.17	12.23	19.98	*	11.30	14.06	19.72	13.67	16.27	22.50	15.58	18.64	23.92
100	4.62	6.06	9.48	6.64	8.37	14.01	8.59	10.85	16.52	11.49	15.89	*	11.57	14.13	19.68	13.77	16.39	22.92	16.01	18.63	24.72
125	4.52	5.82	8.74	6.75	8.90	14.88	10.20	14.65	82.65	11.79	16.35	*	11.56	14.36	20.20	14.06	16.65	22.26	16.38	19.13	24.74
150	4.64	5.85	9.06	6.96	8.99	15.78	10.85	17.92	68.56	12.66	18.43	*	11.69	14.22	19.80	13.78	16.27	22.56	16.34	19.38	25.42
175	4.67	6.12	9.37	6.57	8.65	12.98	12.38	33.18	871.5	11.86	15.80	*	11.94	14.46	20.26	13.96	16.49	21.45	16.35	19.01	25.12
200	4.61	5.99	9.07	6.61	8.67	14.10	17.60	*	*	12.96	22.23	*	11.59	14.25	20.35	14.15	16.56	22.83	16.40	19.26	25.37
∞	4.61	5.99	9.21	6.25	7.81	11.34	7.80	9.49	13.28	9.24	11.07	15.09	10.65	12.59	16.81	13.36	15.51	20.09	15.99	18.31	23.21

Note. * indicates extremely large empirical critical values due to algorithmic problems. Note the difference with $p = 6$, 8, 10, calculated at Cornell Supercomputer Facility.

TABLE A-11

Small Sample Empirical Critical Values for Koziol's Angles Test for $\rho = 0.9$

Size n	$p=2$ $\eta_{.90}$	$p=2$ $\eta_{.95}$	$p=2$ $\eta_{.99}$	$p=3$ $\eta_{.90}$	$p=3$ $\eta_{.95}$	$p=3$ $\eta_{.99}$	$p=4$ $\eta_{.90}$	$p=4$ $\eta_{.95}$	$p=4$ $\eta_{.99}$	$p=5$ $\eta_{.90}$	$p=5$ $\eta_{.95}$	$p=5$ $\eta_{.99}$	$p=6$ $\eta_{.90}$	$p=6$ $\eta_{.95}$	$p=6$ $\eta_{.99}$	$p=8$ $\eta_{.90}$	$p=8$ $\eta_{.95}$	$p=8$ $\eta_{.99}$	$p=10$ $\eta_{.90}$	$p=10$ $\eta_{.95}$	$p=10$ $\eta_{.99}$
25	4.53	5.79	8.86	7.09	9.71	21.79	10.32	16.80	21.12	13.17	20.30	*	10.06	12.26	18.08	12.52	15.22	22.72	14.50	15.75	24.98
50	4.71	6.20	9.09	7.11	9.73	18.17	9.02	12.77	46.74	15.22	24.70	*	10.82	13.06	18.14	14.63	17.87	26.21	18.10	22.27	32.72
75	4.48	5.82	9.13	7.35	10.18	25.05	10.92	16.21	422.8	13.99	23.52	*	11.23	13.55	18.24	15.01	18.63	26.53	19.18	24.12	35.88
100	4.53	5.90	9.04	6.72	8.67	14.12	9.94	14.37	341.9	17.75	47.27	*	11.13	13.51	18.76	15.33	18.77	26.67	20.37	24.84	36.82
125	4.58	6.04	9.58	7.00	9.08	17.61	10.86	16.12	158.5	17.28	48.13	*	11.40	13.87	19.95	15.64	19.09	27.56	19.89	24.57	36.74
150	4.57	6.09	9.17	7.23	9.54	16.24	10.86	15.56	90.23	15.34	31.40	*	11.23	13.64	18.85	15.78	19.39	27.64	19.87	24.04	34.76
175	4.50	5.78	8.64	7.62	10.58	30.79	10.59	16.36	255.5	13.42	19.71	*	11.65	14.15	19.21	15.35	18.70	27.50	20.78	25.60	36.71
200	4.62	6.06	9.06	7.41	9.97	30.60	9.60	12.00	*	13.30	20.82	*	11.53	13.76	19.42	15.69	18.92	25.57	20.54	25.91	38.43
∞	4.61	5.99	9.21	6.25	7.81	11.34	7.80	9.49	13.28	9.24	11.07	15.09	10.65	12.59	16.81	13.36	15.51	20.09	15.99	18.31	23.21

Note. * indicates extremely large empirical critical values due to algorithmic problems. Note the difference with $p = 6, 8, 10$, calculated at Cornell Supercomputer Facility.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Editor and one reviewer for their useful comments, which enhanced the presentation of this paper, and for singling out an important background reference.

REFERENCES

[1] CAMBANIS, S., HUANG, S., AND SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385.

[2] CHMIELEWSKI, M. A. (1981). A re-appraisal of tests for normality. *Comm. Statist. Theory Methods* **10** (20) 2005–2014.

[3] COX, D. R., AND SMALL, N. J. H. (1978). Testing multivariate normality. *Biometrica* **65** 263–272.

[4] FOUTZ, R. V. (1980). A test for goodness-of-fit based on an empirical probability measure. *Ann. Statist.* **8** (5) 989–1001.

[5] GIORGI, G. M., AND FATTORINI L., (1976). An empirical study of some tests for multivariate normality. *Quad. Inst. Statist.* **20** 1–8.

[6] GNANADESIKAN, R. (1977). *Methods of Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

[7] HAWKINS, D. M. (1981). A new test for multivariate normality and homoscedasticity. *Technometrics* **23** 105–110.

[8] JOHNSON, M. E. (1987). *Multivariate Statistical Simulation.* Wiley, New York.

[9] JOHNSON, M. E., CHIANG, W. AND RAMBERG, J. S. (1984). Generation of continuous multivariate distributions for statistical applications. *Amer. J. Math. Sci. Management* **4** 225–248.

[10] JOHNSON, M., AND RAMBERG, J. The Johnson translation system in Monte Carlo studies. *Comm. Statist. Simulation Comput.* **11** (5) 521–525.

[11] KOZIOL, J. A. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika* **69** 423–427.

[12] KOZIOL, J. A. (1983). On assessing multivariate normality. *J. Roy. Statist. Soc. Ser. 8* **45** 358–361.

[13] KOZIOL, J. A. (1986). Assessing multivariate normality: A compendium. *Comm. Statist.* **15** 2763–2683.

[14] MALKOVICH, J. F., AND AFIFI, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68** 176–179.

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

[16] MARDIA, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness study. *Sankhyā Ser B* **36**(2) 114–118.

[17] MARDIA, K. V. (1975). Assessment of multinormality and the robustness of Hotelling *T* test. *Appl. Statist.* **24** 163–171.

[18] MARDIA, K. V., KENT, J. T., AND BIBBY, M. (1979). *Multivariate Analysis.* Academic Press, New York.

[19] MARDIA, K. V. (1980). Tests of univariate and multivariate normality. In *Handbook of Statistics* (P. R. Krishnaiah, Ed.), Vol. 1, pp. 279–320.

[20] OZTURK, A. (1992). A general algorithm for univariate and multivariate GOF tests based on graphical representation. *Comm. Statist. Theory Methods* **20** (10) 3111–3137.

[21] OZTURK, A., AND DUDEWICZ, E. J. (1992). A new statistical goodness-of-fit based on graphical representation. *Biometrical J.* **34** (4) 403–427.

[22] Ozturk, A., and Romeu J. L. (1992). A new multivariate goodness of fit procedure with graphical applications. *Comm. Statist. Simulation* **21** (1) 15–34.

[23] Romeu, J. L. (1990). *Development and Evaluation of a New Multivariate Goodness-of-Fit Procedure.* Ph.D. Dissertation, Syracuse University (also published as Technical Report 9022, October 1990, by the CASE Center of Syracuse University).

[24] Romeu, J. L. (1991). A new multivariate normality goodness of fit test graphical applications. In *Proceedings of the 13th Annual Conference on Computers and Industrial Engineering*, pp. 191–195.

[25] Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro Wilk *W*. *Appl. Statist.* **32** 121–133.

[26] Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika* **52** 591–611.

[27] Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc.* **63** 1343–1372.

[28] Small, N. J. H. (1980). Marginal skewness and kurtosis in testing multivariate normality. *Appl. Statist.* **29** 85–87.

[29] Tong, Y. L. (1990). *The Multivariate Normal Distribution.* Springer-Verlag, Berlin/New York.

[30] Johnson, N., and Kotz, S. (1976). *Continuous, Univariate Distributions*, Vols. 1 and 2. Wiley, New York.

[31] Romeu, J. L. Validation of multivariate Monte Carlo studies. In *Proceedings of the IV International Meeting of Statistics in the Basque Country (IMSIBAC-IV)*, San Sebastian, Spain, August 1992.

[32] Romeu, J. L. Small sample empirical critical values as a tool for the comparison of multivariate normality GOF tests. In *Proceedings of the 1992 Interface Symposium, College Station, TX*, pp. 524–528.