Forthcoming in: Advances in Econometrics, vol. 12, 1997, pp. 341-358

OMNIBUS TESTS FOR MULTIVARIATE NORMALITY BASED ON A CLASS OF MAXIMUM ENTROPY DISTRIBUTIONS

Carlos M. Urzúa

Omnibus Tests for Multivariate Normality Based on a Class of Maximum Entropy Distributions

Carlos M. Urzúa

December, 1996

Abstract

This paper provides omnibus tests for multivariate normality of both observations and residuals. They are derived by considering as the alternatives to the multivariate normal a class of maximum-entropy distributions studied elsewhere by the author. The tests, being Lagrange multiplier statistics, have optimum local asymptotic power among those alternatives. Furthermore, they coincide in the univariate case with the popular Jarque-Bera test for normality. They also include as special cases several multivariate tests available in the literature. Finally, the paper also suggests simple adjustments that can significantly improve the power of the tests in the case of small and medium size samples, even for the univariate case.

KEY WORDS: Tests for Multivariate Normality, Maximum Entropy

1 Introduction

Since the pioneering work earlier this century by, among others, K. Pearson, R. Fisher and J. Wishart, the assumption of multivariate normality has played a key role in many methods of multivariate analysis. Handy as that assumption is, however, the consequences of departure from multivariate normality are documented to be quite serious for several methods (e.g., linear discriminant analysis). Judgment is still pending on other methods, but, in principle, the consequences may also be serious as well. This can be surmised in cases such as simultaneous equation models, where the violation of the multivariate normality assumption may lead to inefficient estimators and invalid inferences.

Given the obvious importance of the multivariate normality assumption, it is thus somewhat surprising that for many years most researchers either ignored it, or were content with the evaluation of marginal normality (which, of course, does not necessarily imply joint normality). Only when Mardia (1970) introduced a simple test based on multivariate measures of skewness and kurtosis, did the issue of testing for multivariate normality gain some favor among researchers. That this favor has grown since then is attested by the burgeoning literature on the subject (see, e.g., the surveys by Mardia, 1987, and Small, 1985).

The purpose of this paper is to provide readily computable tests for multivariate normality of both observations and residuals of multivariate equation models. They are derived by considering as alternatives to the multivariate normal a class of "likely", maximum entropy, multivariate distributions introduced in Urzúa (1988).

The tests, being derived by the Lagrange multiplier procedure, have optimum locally asymptotic power among those alternatives. Thus, they distinguish themselves from other ad-hoc tests in the literature that are simply patterned as extensions of tests for univariate normality. This is not meant to deny the practical advantage of having multivariate tests with such a property, for in fact some of the tests proposed here are the multivariate counterparts to the popular Jarque-Bera test for univariate normality (Jarque and Bera, 1980 and 1987). This paper also presents simple adjustments to the LM tests that can significantly improve its performance in the case of small and medium size samples, even for the univariate case.

This paper is organized as follows: Section 2 reviews several of the properties that characterize the "likely" multivariate distributions, and presents some basic results for use in later sections. Section 3 derives the Lagrange multiplier test for multivariate normality under the premise that the alternatives to the normal are other maximum entropy distributions. It also corrects the test to improve its performance in the case of non-large samples. Finally, Section 4 presents test statistics for multivariate normality of residuals of simultaneous equation models, and of vector autoregression models for time series.

2 Likely alternatives to the multivariate normal

In his authoritative paper on significance tests written in the seventies, D. R. Cox complained about the non-existence of "a simple and general family of distributions to serve as alternatives [to the multivariate normal]" (1977, p. 56). This section reviews a class of distributions, an exponential family studied in Urzúa (1988), which could play that role.

2.1 Some definitions

The distributions to be considered in this paper are multivariate generalizations of the, as yet, relatively unknown distributions introduced by R. A. Fisher (1922). Defined over the real line, Fisher's univariate densities are of the form

$$f(x) = \tau(\alpha)\exp(-Q(x)), \quad Q(x) = \alpha_1 x + \alpha_2 x^2 + ... + \alpha_k x^k \tag{1}$$

where k is an even number, $\alpha_k > 0$, and $\tau(\alpha)$ is the constant of normalization given the vector of parameters α. Aside of course from the normal (obtained when $k = 2$), the densities in (1) were considered to be of little interest for many years. More recently, however, there has been an increasing interest on them since they play a key role in stochastic catastrophe theory (see Urzúa, 1990, and references therein).

Furthermore, as Zellner and Highfield (1988) have strikingly illustrated in the case of the quartic exponential (obtained setting $k = 4$ in equation (1) above), Fisher's distributions are flexible enough, and simple enough, to act as *bona fide* approximations to other univariate distributions.

We now turn to their multivariate counterparts. Let x denote the real column vector $(x_1, ..., x_p)'$. If $Q(x)$ is a polynomial of degree k in the p variables, then it can always be written, ignoring the constant term, as

$$Q(x) = \sum_{q=1}^{k} Q^{(q)}(x), \tag{2}$$

where $Q^{(q)}(x)$ is a homogeneous polynomial (a form) of degree q. Namely,

$$Q^{(q)}(x) = \sum \alpha^{(q)}_{j_1 ... j_p} \prod_{i=1}^{p} x_i^{j_i}, \tag{3}$$

with the summation taken over all nonnegative integer p-tuples $(j_i, ..., j_p)$ such that $j_1 + ... + j_p = q$. The polynomial Q will be assumed to be such that $g(x) = \exp(-Q(x))$ is integrable on the entire Euclidean space R^p (a necessary condition for this to happen is that the degree of $Q(x)$ relative to each x_i is an even integer).

Following Urzúa (1988), the continuous random vector $X = (X_1, ..., X_p)$ is said to have a *p-variate Q-exponential* distribution with support R^p if its density is given by

$$f(x) = \tau(\alpha)\exp(-Q(x)), -\infty < x_i < \infty, i = 1, ..., p, \tag{4}$$

where $\tau(\alpha)$ is the constant of normalization.

For simplicity, it will be implicitly assumed below that the polynomial Q is of degree k relative to all of its components. In such a case, several important distributions emerge: If $k = 2$, then the p-variate normal is obtained; while when $k = 4$ and $k = 6$ the p-variate *quartic* and *sextic exponentials* are obtained.

Note also that as k is increased the number of coefficients required by the corresponding Q-exponential increases at an increasing rate. In fact, as can be readily shown (see Urzúa, 1988), if $K(p, k)$ denotes the maximum possible number of parameters of a p-variate Q-exponential, then

$$K(p,k) = C(p+k,k) - 1, \tag{5}$$

where $C(p+k,k)$ is the binomial coefficient $(p+k)!/(p!k!)$. In particular, the number of possible coefficients in the homogeneous polynomial of degree q given in (3) is $C(p+q-1,q)$.

2.2 Some characterizations

It is now time to introduce a key characteristic of the Q-exponential distributions. Consider all densities f relative to Lebesgue measure that have support $\Omega = R^p$ and have finite population moments of some predetermined orders. That is, they satisfy constraints of the form

$$E\left\{\prod_{i=1}^{p} X_i^j\right\} = c_m, \quad m = 1, ..., r, \tag{6}$$

where each j is a nonnegative integer, and $c_1, ..., c_r$, is a sequence of real numbers. For each density we define, following Shannon (1948), the entropy of f as

$$H(f) = -\int f(x) log\left[f\left(x\right)\right] dx. \tag{7}$$

It can be shown (see Urzúa, 1988) that, among the densities satisfying 6, if there is a density that maximizes Shannon's entropy, it is necessarily a Q-exponential of the form

$$f(x) = \tau(\alpha) exp(-Q(x)), \quad Q(x) = \sum_{m=1}^{r} \alpha_m \prod_{i=1}^{p} X_i^{j_m}$$

.

For instance, the p-variate quartic exponential maximizes the entropy among the distributions with support R^p that are known to have finite moments up to order four. Likewise, as Shannon (1948) in his influential paper first proved, the multivariate normal maximizes the entropy among the distributions that have second order moments.

Thus, when the only known information about a distribution is the existence of population moments of some orders, the Q-exponentials can be considered to be the "most likely to be true". This according to the maximum entropy principle, which states that "in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known" (Jaynes, 1957, p. 623). This principle, as remarked by Klir and Folger (1988, p. 214), can be rephrased using the following two sentences of the Chinese philosopher Lao Tsu (who lived in the sixth century B.C.): "Knowing ignorance is strength. Ignoring knowledge is sickness."

There is a second characteristic of the Q-exponentials that is also relevant for our purposes. It can be shown that, near the multivariate normal, the quartic exponential is capable of approximating as close as needed the van Uven-Steyn multivariate Pearson family (see Urzúa, 1988). This is interesting because the latter family, although made of distributions more complex (and suspect) than the Q-exponentials, could be thought by some to constitute a class of possible alternatives to multivariate normality (in fact, Bera and John, 1983, have used such a family to derive tests for multivariate normality).

Before concluding this section, it is worth briefly mentioning other interesting properties exhibited by the Q-exponential distributions that, although not directly relevant to this paper, help to illustrate the generality of the distributions (see Urzúa, 1988, for details): First, they can exhibit several modes, and they do so with a relatively small number of parameters (as compared to mixtures of multinormals). Second, they are the stationary distributions of certain multivariate diffusion processes. Third, the maximum likelihood estimators of their population moments are the sample moments (as can be directly seen from 9 below). And fourth, using the method of moments one can easily obtain consistent estimators for the parameters of the Q-exponential distributions. This last result is particularly useful given the large number of parameters in the case of high-dimension distributions.

3 Test for multivariate normality of observations

Let X be a $p \times 1$ random vector following a Q-exponential distribution. Consider a set of n observations $x_1, ..., x_n$ on X. The corresponding log-likelihood function $L(\alpha)$ can then be easily shown to be

$$L(\alpha) = -n\log[\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} exp(-Q(x))dx] - \sum_{r=1}^{n} Q(x_r). \tag{8}$$

Furthermore, the components of the gradient (score) of $L(\alpha)$ are of the form

$$\frac{\partial L}{\partial \alpha^{(q)}_{j_1 \ldots j_p}} = nE\left\{\prod_{i=1}^{p} X_i^{j_i}\right\} - \sum_{r=1}^{n}\prod_{i=1}^{p} X_{ir}^{j_i} \tag{9}$$

while the elements of the Hessian of $L(\alpha)$ are of the form

$$\frac{\partial^2 L}{\partial \alpha^{(q)}_{j_1 \ldots j_p} \partial \alpha^{(r)}_{k_1 \ldots k_p}} = -nE\left\{\prod_{i=1}^{p} X_i^{j_i+k_i}\right\} + nE\left\{\prod_{i=1}^{p} X_i^{j_i}\right\} E\left\{\prod_{i=1}^{p} X_i^{k_i}\right\} \tag{10}$$

Consequently, Fisher's information matrix is simply made of the covariances of products of the random components multiplied by n.

It will prove useful to transform the random vector X to a vector Y having zero mean and the identity matrix as the covariance. Let μ and Σ be the mean vector and the covariance matrix of X. Let Γ denote the orthogonal matrix whose columns are the standardized eigenvectors of Σ, and Λ denote the diagonal matrix of the eigenvalues of Σ. Define $\Sigma^{-1/2}$ as the inverse of the square root decomposition of Σ; that is, $\Sigma^{-1/2} = \Gamma\Lambda^{-1/2}\Gamma'$. Then the random vector

$$Y = \Sigma^{-1/2}(X - \mu) \tag{11}$$

follows a p-variate Q-exponential, with $Q(y)$ as in (2) and $Q^{(q)}(y)$ as in (3). It has a zero mean vector, and an identity matrix as its covariance matrix.

Let the $K(p,k) \times 1$ vector of parameters of $Q(y)$ be denoted as α where $K(p,k)$ is given in (5) above. Suppose now that α is partitioned as $\alpha =$

$(\theta_1', \theta_2')'$, where θ_1 is the $C(p+1,2) \times 1$ vector of parameters of the homogeneous polynomial $Q^{(2)}(y)$. The hypothesis of multinormality can be then assessed by testing the null hypothesis $H_0 : \theta_2 = 0$. There are several asymptotic tests available for that purpose (see, e.g., the survey by Engle, 1984). Given the complexity of the alternatives considered here, the Lagrange multiplier (LM) test of Rao (1948), and Aitchison and Silvey (1958) will be used below, for it only requires the estimation of the restricted model under the null hypothesis.

3.1 The Lagrange multiplier test

In order to give an expression for the LM statistic, it is necessary to introduce some notation. Let $s(\alpha)$ be the gradient (score) of the log-likelihood function, and let I be the information matrix. Given the partition of α as $\alpha = (\theta_1', \theta_2')'$ the score can be written as $s(\alpha) = (s_1', s_2')'$, with $s_j = \partial L(\alpha)/\partial\theta_j, j = 1, 2$; while the information matrix can be partitioned into four submatrices of the form $I_{ij} = E[-\partial^2 L(\alpha)/\partial\theta_i\partial\theta_j'], i, j = 1, 2$.

Let $(\tilde{\theta}_1, 0)$ denote the restricted maximum likelihood estimator for $\alpha = (\theta_1', \theta_2')'$; that is, $\tilde{\theta}_1$ is the maximum likelihood estimator for θ_1 after imposing the constraint $\theta_2 = 0$. Let also $\tilde{s} = s(\tilde{\theta}_1, 0)$ and $\tilde{I} = I(\tilde{\theta}_1, 0)$. Then the LM statistic is defined as $LM = \tilde{s}\tilde{I}^{-1}\tilde{s}/n$, or, taking advantage of the fact that $\tilde{s}_1 = 0$,

$$LM = \tilde{s}_2' \left(\tilde{I}_{22} - \tilde{I}_{21}\tilde{I}_{11}^{-1}\tilde{I}_{12}^{-1} \right) \tilde{s}_2/n \tag{12}$$

LM is under H_0 asymptotically distributed as a χ^2_ν, a Chi-square with degrees of freedom ν equal to the dimension of the vector θ_2.

3.2 A test for multivariate normality of observations

Given the complexity of the alternatives to the multivariate normal to be considered here, the computation of the LM statistic would appear to be a daunting task. Fortunately, as will soon be seen, such will not be the case. In what follows we will assume that the quartic exponential is the alternative distribution to the multivariate normal. This is enough since, as noted earlier, (i) it is the most likely distribution when moments up to the fourth order are assumed to exist, and (ii) near the multivariate normal, it can approximate the multivariate Pearson family as close as needed. Furthermore, the results given below can be trivially extended to all possible Q-exponentials.[1]

Before finding the LM test statistic, it is convenient to introduce further notation. Let us first transform the original observations on X: Let $\bar{x}$ and S be the sample mean vector and the sample variance-covariance matrix found using the set of observations $x_1, ..., x_n$. Let G denote the orthogonal matrix whose columns are the standardized eigenvectors of S, and D denote the diagonal matrix of the eigenvalues. Using $S^{-1/2} = GD^{-1/2}G'$, transform the observations as follows:

$$y_t = S^{-1/2}\left(x_t - \bar{x}\right), t = 1, ..., n. \tag{13}$$

After defining

$$Q_{ijk} = \sum_{t=1}^{n} \frac{y_{ti} y_{tj} y_{tk}}{n}, \quad R_{ijkq} = \sum_{t=1}^{n} \frac{y_{ti} y_{tj} y_{tk} y_{tq}}{n}, \tag{14}$$

the main result of the paper can be now stated as follows:

Proposition 1 .

Under the above conditions, the Lagrange multiplier test statistic for multivariate normality of observations is given by:

$$\begin{aligned} LM_p = \sum_{i=1}^{p} \frac{nQ_{iii}^2}{6} + \sum_{i,j=1, i\neq j}^{p} \frac{nQ_{iij}^2}{2} + \sum_{i,j,k=1, i<j<k}^{p} nQ_{ijk}^2 + \\ \sum_{i=i}^{p} \frac{n\left(R_{iii}-3\right)^2}{24} + \sum_{i,j=i, i<j}^{p} \frac{n\left(R_{iijj}-1\right)^2}{4} + \sum_{i,j=i, i\neq j}^{p} \frac{nR_{iiij}^2}{6} + \\ \sum_{i,j,k=i, i\neq j, i\neq k, j<k}^{p} \frac{nR_{iijk}^2}{2} + \sum_{i,j,k,q=i, i<j<k<q}^{p} nR_{ijkq}^2, \end{aligned}$$

where the statistic LM_p is asymptotically distributed as a χ^2_ν, with $\nu = p(p+1)(p+2)(p+7)/24$.

Since the proof of the Proposition 1, although conceptually simple, is algebraically involved, it is relegated to the Appendix. Yet, the interpretation of the test is quite straightforward: It is an omnibus test involving all possible third and fourth moments (pure and mixed). Furthermore, it is constructed in the obvious way: Summing the squares of standardized normals (under the null), after using in each standardization the corresponding asymptotic mean and variance (the asymptotic covariance between any two terms in the expression is zero). For instance, each element in the first summation has zero mean and an asymptotic variance of $6/n$ (see subsection 3.3 below for more examples).

It should also be noted that another justification of Proposition 1 may be given indirectly through an elegant result derived by Gart and Tarone (1983) for all exponential families. For any family of that type, the corresponding likelihood has to be of the form $\tau(\beta,\alpha)\exp(\beta' u + \alpha' v)h(u,v)$, where $u = (u_1, ..., u_r)$' and $v = (v_1, ..., v_s)'$ are sufficient statistics, and β and α are the parameters of the distribution. Then, following those authors, it can be shown that the Lagrange multiplier (score) test for the null hypothesis $H_0 : \beta = \beta_0$ is simply given by the statistic

$$LM = (u - E\{u \mid v\})' Var\{u \mid v\}^{-1}((u - E\{u \mid v\})) \tag{15}$$

where $E\{u \mid v\}$ and $Var\{u \mid v\}$ are the asymptotic conditional mean and variance matrix of u given v, under the null.

In our case the vector of sufficient statistics v is given by all possible second order moments, while u is made of all possible third and fourth moments (and first moments, but their contribution vanishes as shown in the first step of the proof in the Appendix). Thus, using (15), Proposition 1 is justified.[2]

3.3 Some special cases of the LM_p Test

The LM statistic derived above includes as special cases several tests for multivariate normality that have been proposed in the literature.

To start with, in the univariate case the LM_1 statistic can be expressed in terms of the standardized third and fourth moments of the original observations. Defining $\sqrt{b_1} = m_3/m_2^{3/2}$ and $b_2 = m_4/m_2^2$, where the i-th central moment m_i equals $\sum(x_j - \bar{x})^i/n$, then the test becomes:

$$LM_1 = n \left[\frac{\left(\sqrt{b_1}\right)^2}{6} + \frac{(b_2 - 3)^2}{24} \right] \sim^A \chi^2_2 \tag{16}$$

This statistic has been proposed by Bowman and Shenton (1975), and by Jarque and Bera (1980 and 1987). The former authors suggested the use of this statistic as the simplest possible omnibus test for normality since, under the null, the asymptotic means of $\sqrt{b_1}$ and b_2 are respectively 0 and 3, their asymptotic variances are $6/n$ and $24/n$, and their asymptotic covariance is zero. While the latter authors found (16) to be the LM test statistic obtained when the alternatives to normality are in the Pearson family.

In the more general multivariate case, the omnibus LM_p statistic contains as special cases several tests already available in the literature. Independently, Bera and John (1983) and Lütkepohl and Theilen (1991) have considered the possibility of using as tests for multivariate normality the sum of squares of the standardized pure third and fourth moments, i.e., the terms in the first and fourth summation signs appearing in the expression for LM_p in Proposition 1. Bera and John have also considered the possibility of using the mixed fourth moments that are obtained multiplying the squares of any pair of components (i.e., the elements in the fifth summation sign appearing in LM_p).

It is also interesting to note that the omnibus LM_p test (or some of its components) somewhat resembles the omnibus tests proposed by Jarque and McKenzie (1983) and Mardia and Kent (1991) using Mardia's measures of multivariate skewness and kurtosis. Note, however, that Mardia's statistics are not derived after orthogonalizing the observations, but rather after using quadratic forms of the type $(x_t - \bar{x})S^{-1}(x_r - \bar{x})$.

3.4 Adjusted Lagrange multiplier tests

Given the very large number of degrees of freedom of the LM_p statistic in Proposition 1, it should not be a surprise to learn that the author has found, after a few Monte Carlo exercises, that it does not behave well for small and medium size samples (with the problem getting worse as p is increased)[3] Although for large samples, of course, the hypothesis of p-variate normality of observations can be safely rejected at some significant level (usually taken to be 10%) if the value of LM_p exceeds the corresponding critical value of the χ^2_ν.

To solve that shortcoming, this subsection presents simpler test statistics made of some of the elements of LM_p; most of those test statistics, by the way, will continue to be LM tests, since they would arise after choosing some particular quartic exponentials. Furthermore, and more interestingly, this subsection will also show how to adjust those LM tests to substantially improve their asymptotic convergence.

Let us start first with the univariate case. As noted earlier, LM_1 is none other than the popular Bowman-Shenton-Jarque-Bera omnibus test given in (16) above. It is important to realize, however, that even for such a simple functional form the speed of convergence to the χ^2_v is quite slow (something that is often ignored when the test is used in applied econometrics).

Luckily enough, there is a very simple adjustment that can be used to improve its convergence. First note that, based on a straightforward extension of a result in Fisher (1930), it is possible to derive exactly the means and variances of $\sqrt{b_1}$ and b_2 under normality (the null). As in the asymptotic case, $E\left\{\sqrt{b_1}\right\} = 0$, but for the other mean and for the two variances it can be easily shown (see Urzúa 1996) that the exact values are:

$$E\{b_2\} = 3(n-1)/(n+1), \tag{17}$$

$$Var\left\{\sqrt{b_1}\right\} = 6(n-2)/(n+1)(n+3), \tag{18}$$

$$Var\{b_2\} = 24n(n-2)(n-3)/(n+1)^2(n+3)(n+5). \tag{19}$$

Thus, as it was first suggested in Urzúa (1996), we can now substitute the asymptotic values for the exact values to obtain a new adjusted LM_1 statistic, with the hope of speeding up the convergence of the omnibus test:

$$ALM_1 = \frac{\left(\sqrt{b_1}\right)^2}{Var\left\{\sqrt{b_1}\right\}} + \frac{(b_2 - E\{b_2\})^2}{Var\{b_2\}} \sim^A \chi^2_2 \tag{20}$$

It is shown in Urzúa (1996) that this new adjusted LM test indeed behaves better in the case of small and medium size samples. Furthermore, it is also shown there that the power of the new test is even slightly greater than the power of the Bowman-Shenton-Jarque-Bera test statistics.

Given those encouraging results, it is natural to consider also in this paper the individual counterparts to the omnibus univariate test ALM_1. That is, using the same adjustments as before, let us now introduce an adjusted skewness measure test defined as

$$ALM_{1,1} = \frac{\left(\sqrt{b_1}\right)^2}{Var\left\{\sqrt{b_1}\right\}} \sim^A \chi_1^2 \tag{21}$$

and, likewise, an adjusted kurtosis measure test defined as

$$ALM_{2,1} = \frac{\left(b_2 - E\left\{b_2\right\}\right)^2}{Var\left\{b_2\right\}} \sim^A \chi_1^2 \tag{22}$$

Note that (22) could be truly derived as a Lagrange multiplier test in the case of a quartic exponential without a cubic term, while (21) could not be (since the corresponding cubic exponential would not exist). But for purposes of consistency of notation, all the tests in this paper are called LM tests.

We now turn to the multivariate case. In order to reduce the number of degrees of freedom present in the general LM_p test given in Proposition 1, it is natural to focus only on the pure third and fourth moments. Although, of course, there is no reason a priori to prefer pure moments over some mixed moments, and future work will try to explore other combinations.

Hence, using (14) and (17), an adjusted LM omnibus test for multivariate normality of observations will be defined as:

$$ALM_p = \sum_{i=1}^{p} \frac{Q_{iii}^2}{Var\left\{\sqrt{b_1}\right\}} + \sum_{i=1}^{p} \frac{\left(R_{iiii} - E\left\{b_2\right\}\right)^2}{Var\left\{b_2\right\}} \sim^A \chi_{2p}^2 \tag{23}$$

Also, generalizing the univariate case, we can define the adjusted skewness measure test as

$$ALM_{1,p} = \sum_{i=1}^{p} \frac{Q_{iii}^2}{Var\left\{\sqrt{b_1}\right\}} \sim^A \chi_p^2 \tag{24}$$

and the adjusted kurtosis measure test as

$$ALM_{2,p} = \sum_{i=1}^{p} \frac{\left(R_{iiii} - E\left\{b_2\right\}\right)^2}{Var\left\{b_2\right\}} \sim^{A} \chi_p^2 \tag{25}$$

In the case of each of these three statistics, Table 1 reports significance points for the typical confidence levels used to test for multivariate normality of observations. Each cell in the table was generated through 10000 replications from a multivariate standard normal.

The results, as evident from the table, are quite encouraging. It is noteworthy, for instance, how near to the corresponding asymptotic value are the generated critical values for the skewness-based test $ALM_{1,p}$ when $\alpha = 10\%$, even for rather small sample sizes. As will be noted in the next section, this fact should encourage the use of this test in the case of residuals of statistical models.

Comparatively, the kurtosis-based test $ALM_{2,p}$, converges more slowly, but its behavior is still quite remarkable given the well known bad convergence properties of the typical, unadjusted, statistics based on fourth moments.

Finally, the omnibus test ALM_p, the one that in principle should have the best power (although this issue will have to be answered in future work), also behaves quite well (compare, for instance, the critical values for $p = 1$ with the corresponding values for the Jarque-Bera test for univariate normality (Jarque and Bera, 1987, Table 2).

4 Tests for multivariate normality of residuals

The tests for multivariate normality of observations introduced in the last section can be easily extended, as we will do next, to the case of residuals of the typical linear structural models used in Economics. What cannot be borrowed from the last section, however, are the significance points in Table 1. This is so because the sample properties of the tests will depend in general on the particular design matrices of each structural model. It is for this same reason that in what follows we restrict our attention to the adjusted

TABLE 1

Significance Points for the ALM Tests for Multivariate Normality of Observations

		ALM1,p				ALM2,p				ALMp			
p	n	0.90	0.95	0.975	0.99	0.90	0.95	0.975	0.99	0.90	0.95	0.975	0.99
	10	2.70	4.14	5.68	7.59	2.07	4.06	6.91	11.29	4.12	7.79	12.32	18.61
	20	2.67	3.96	5.45	7.66	1.90	3.65	6.53	12.03	3.94	6.84	11.25	18.59
1	50	2.67	3.85	5.18	7.10	1.98	3.26	5.51	10.00	3.90	6.42	9.33	15.56
	100	2.61	3.89	5.19	7.46	2.12	3.41	5.39	9.02	4.06	6.13	9.02	14.44
	200	2.63	3.83	5.07	6.82	2.31	3.50	5.44	8.51	4.21	6.07	8.50	12.55
	800	2.67	3.85	4.88	6.35	2.47	3.62	4.97	7.23	4.39	5.90	7.64	10.25
	∞	2.71	3.84	5.02	6.64	2.71	3.84	5.02	6.64	4.61	5.99	7.38	9.21
	10	4.77	6.42	8.04	9.94	4.30	7.34	11.11	15.62	8.65	13.36	18.82	24.99
	20	4.79	6.34	8.21	10.64	4.22	7.23	11.33	18.61	8.39	12.63	18.66	27.68
2	50	4.64	6.34	7.93	10.41	4.07	6.80	10.82	18.08	7.83	11.57	16.96	26.17
	100	4.59	6.15	7.78	10.04	4.19	6.96	10.30	16.13	7.98	11.20	15.96	23.03
	200	4.65	6.10	7.77	10.15	4.42	6.40	9.38	14.55	7.88	10.94	14.34	21.42
	800	4.63	6.06	7.44	9.21	4.53	6.03	8.02	11.98	7.85	10.00	12.56	16.35
	∞	4.61	5.99	7.38	9.21	4.61	5.99	7.38	9.21	7.78	9.49	11.14	13.28
	10	6.57	8.47	10.28	12.69	6.88	10.33	14.28	19.93	12.96	18.39	23.94	32.33
	20	6.49	8.46	10.27	13.14	6.24	10.18	14.96	22.28	11.94	17.35	24.03	33.93
3	50	6.38	8.33	10.42	13.23	6.01	9.63	14.20	22.34	11.35	16.62	22.80	33.48
	100	6.27	8.01	9.77	12.28	6.07	9.20	13.69	21.59	11.12	15.24	21.47	30.57
	200	6.27	7.78	9.50	11.82	6.13	8.53	11.80	17.67	11.01	14.37	18.53	24.14
	800	6.14	7.75	9.26	11.11	6.34	8.29	10.55	14.75	10.83	13.24	15.97	20.51
	∞	6.25	7.82	9.35	11.35	6.25	7.82	9.35	11.35	10.65	12.59	14.45	16.81
	10	8.17	10.36	12.26	15.09	8.55	12.43	17.06	23.20	16.27	22.32	28.83	37.07
	20	8.07	10.38	12.68	15.58	8.67	13.04	19.15	27.77	15.70	22.35	30.18	40.82
4	50	8.04	10.18	12.19	15.28	8.03	12.20	17.29	26.65	14.81	20.43	27.49	39.50
	100	7.95	9.86	11.77	14.27	8.09	11.68	15.92	24.45	14.69	19.16	24.63	34.01
	200	7.93	9.70	11.42	13.98	7.93	10.93	15.29	21.53	14.00	18.21	23.03	30.97
	800	7.81	9.57	11.09	13.37	7.81	9.99	12.46	16.03	13.56	16.30	19.17	23.46
	∞	7.78	9.49	11.14	13.28	7.78	9.49	11.14	13.28	13.36	15.51	17.54	20.09
	10	9.83	12.16	14.10	17.42	10.69	15.19	20.00	26.16	19.98	26.99	33.45	42.49
	20	9.57	11.90	14.42	17.88	10.16	15.09	21.41	31.82	18.67	25.71	34.01	48.24
5	50	9.45	11.83	14.03	16.99	9.89	14.37	20.12	29.09	18.06	23.95	31.24	44.01
	100	9.44	11.48	13.64	16.70	9.77	13.66	18.15	26.33	17.60	22.91	28.68	38.34
	200	9.31	11.31	13.18	15.88	9.53	12.97	16.79	22.28	17.00	21.25	26.75	33.21
	800	9.27	11.10	12.93	15.50	9.25	11.70	14.25	17.82	16.21	19.29	22.24	26.42
	∞	9.24	11.07	12.83	15.09	9.24	11.07	12.83	15.09	15.99	18.31	20.48	23.21

ALM tests, rather than considering the general omnibus test LM_p derived in Proposition 1.

4.1 Simultaneous equation models

Consider the simultaneous equation model

$$By_r + \Gamma z_r = u_r, \quad r = 1, ..., n \tag{26}$$

where y_r is a $p \times 1$ vector of observed endogenous variables, z_r is a $k \times 1$ vector of observed predetermined variables, u_r is a $p \times 1$ vector of unobserved disturbances, B is a $p \times p$ nonsingular matrix of coefficients with ones in its diagonal, and Γ is a $p \times k$ matrix of coefficients. All identities are assumed to be substituted out, and the system is assumed to be identified through exclusions in B and Γ. Assume furthermore that the alternative to the possible p-variate normal distribution of ur is, as before, a p-variate quartic exponential.

Suppose first that the system is estimated using full information maximum likelihood (FIML) under the assumption of multivariate normality. Following the reasoning in the last section, one can construct tests for multivariate normality of the residuals as follows: Let $\hat{\mathrm{u}}_r$ denote the estimated FIML residuals of the structural equations (one could similarly use the estimated residuals of the reduced form). Using the transformation $e_r = S^{-1/2}\hat{\mathrm{u}}_r$, where $S^{-1/2}$ is defined as in (13), define next:

$$V_{ijk} = \sum_{t=1}^{p} \frac{e_{ti}e_{tj}e_{tk}}{n}$$

and

$$W_{ijkq} = \sum_{t=1}^{p} \frac{e_{ti}e_{tj}e_{tk}e_{tq}}{n}$$

Then the ALM's tests for multivariate normality of the residuals can be defined as:

$$ALMR_p = \sum_{i=1}^{p} \frac{V_{iii}^2}{Var\left\{\sqrt{b_1}\right\}} + \sum_{i=1}^{p} \frac{\left(W_{iiii} - E\left\{b_2\right\}\right)^2}{Var\left\{b_2\right\}} \sim^A \chi^2_{2p} \tag{27}$$

$$ALMR_{1,p} = \sum_{i=1}^{p} \frac{V_{iii}^2}{Var\left\{\sqrt{b_1}\right\}} \sim^A \chi_p^2 \tag{28}$$

$$ALMR_{1,p} = \sum_{i=1}^{p} \frac{\left(W_{iiii} - E\left\{b_2\right\}\right)^2}{Var\left\{b_2\right\}} \sim^A \chi_p^2 \tag{29}$$

But, what if, as is usually the case, the system is not estimated by FIML, but rather by some other method (e.g., 2SLS)? Provided the method renders consistent estimators, one can use the corresponding ALMR's constructed using the estimated residuals of the structural equations (or, equivalently, the estimated residuals of the reduced form equations). This is so because, following White and MacDonald (1980), one can show that the statistics constructed using the estimated residuals are consistent estimators of the true statistics.[4]

As was stressed earlier, in the case of small samples one cannot use the empirical significance points given in Table 1. But, as was also noted earlier, the asymptotic critical values can be confidently used in the case of the $ALMR_{1,p}$ test. Furthermore, it is straightforward to simulate the significance points corresponding to the $ALMR$ tests, for a given linear structural model.

4.2 Vector autoregressive models

Before closing this section, it is interesting to pose the following question arising in the case of multivariate time series that follow a vector autoregressive (VAR) process: In order to test for Gaussianity, should we use the original observations, or the estimated residuals after fitting the VAR model? Based on simulation studies, Lütkepohl and Theilen (1991) recommend the second alternative. Thus, in our context, tests based on the $ALMR$'s rather than on the ALM's should be preferred.

5 Conclusions and extensions

This paper has provided an omnibus LM test for multivariate normality which is, in effect, the most comprehensive test that one can ever devise using third and fourth (pure and mixed) moments. Being derived using the

Lagrange multiplier procedure, the test has optimum local asymptotic power among the multivariate quartic exponentials, the maximum-entropy ("most likely") multivariate distributions when it is assumed that moments up to the fourth order exist. This paper has also provided some particular adjusted LM tests that converge quite rapidly to their asymptotic distribution.

Three extensions to the results presented above are clearly called for. First, it is worth exploring the possibility of using other elements of LM_p, and comparing the power of the resulting tests with the ones for the ALM tests given here. Second, given the dozens of already available tests for multivariate normality, a complete Monte Carlo study appraising the power of each of them is much needed. And third, maintaining the hypothesis of Q-exponentials as the alternative distributions, it would be interesting to derive new tests for multivariate normality of the residuals of other common multivariate models, such as simultaneous limited dependent variable models.

A Appendix: Proof of Proposition 1

The proof will be developed in several steps, making continuous use of the expressions for the first and second partial derivatives of the log-likelihood given in (9) and (10), and reproduced here ($q, r = 1, 2, 3, 4$):

$$\frac{\partial L}{\partial \alpha^{(q)}_{j_1 \cdots j_p}} = nE\left\{\prod_{i=1}^{p} Y_i^{j_i}\right\} - \sum_{r=1}^{n}\prod_{i=1}^{p} y_{ir}^{j_i} \tag{30}$$

$$\frac{\partial^2 L}{\partial \alpha^{(q)}_{j_1 \cdots j_p}\partial \alpha^{(r)}_{k_1 \dots k_p}} = -nE\left\{\prod_{i=1}^{p} Y_i^{j_i+k_i}\right\} + nE\left\{\prod_{i=1}^{p} Y_i^{j_i}\right\} E\left\{\prod_{i=1}^{p} Y_i^{k_i}\right\} \tag{31}$$

Step 1: Note that, under the null, the elements of the score corresponding to the first order moments are zero:

$$\frac{\partial L}{\partial \alpha^{(1)}_{j_1 \cdots j_p}} = nE\left\{Y_i\right\} - \sum_{i=1}^{n} y_{ir}, \quad \text{for all } j_i = 1. \tag{32}$$

Hence, the expression for the Lagrange multiplier statistic given in (12) can be further simplified as:

$$LM = d'(F - C'B^{-1}C^{-1})d/n, \tag{33}$$

where, the new symbols, *all* evaluated under the null, are given by:

$$d = \left\{\partial L'/\partial\alpha^{(3)}, \partial L'/\partial\alpha^{(4)}\right\}';$$
$$F = -E\left\{\partial^2 L/\partial\alpha^{(q)}\partial\alpha^{(r)}\right\} \quad q, r = 3, 4;$$
$$B = -E\left\{\partial^2 L/\partial\alpha^{(q)}\partial\alpha^{(r)}\right\} \quad q, r = 1, 2;$$
$$C = -E\left\{\partial^2 L/\partial\alpha^{(q)}\partial\alpha^{(r)}\right\} \quad q = 1, 2; r = 3, 4.$$

Step 2: Since the components of the random vector Y are independent, and have, under the null, odd-order population moments equal to zero, it follows that the expression in (31) is equal to zero whenever q is even and r is odd, or viceversa. Thus, all matrices in (33) are block diagonal: $F = \mathrm{diag}(F_{11}, F_{22})$, $B = \mathrm{diag}(B_{11}, B_{22})$, and $C = \mathrm{diag}(C_{11}, C_{22})$.

Step 3: In fact, B is a completely diagonal matrix. This is so because all the elements of B_{11} are zero except for:

$$\frac{-\partial L}{\partial\alpha^{(1)}_{j_1\ldots j_p}\partial\alpha^{(1)}_{k_1\ldots k_p}} = nE\left\{Y_i^2\right\} = n, \text{ when } j_i = k_i = 1 \tag{34}$$

where the last equality obtains since, under the null, the second population moment of each component equals one.

Most elements of B_{22} are zero as well, except for two possible arrangements of the subindices. The first case is:

$$\frac{-\partial L}{\partial\alpha^{(2)}_{j_1\ldots j_p}\partial\alpha^{(2)}_{k_1\ldots k_p}} = n\left[E\left\{Y_i^4\right\} - E\left\{Y_i^2\right\}^2\right] = 2n, \text{ when } j_i = k_i = 2 \tag{35}$$

where in the second equality use has been made of the fact that, for a univariate standard normal,

$$E\left\{Y_i^{2t}\right\} = (2t)!E\left\{Y_i^2\right\}^t/2^t t! \qquad t = 1, 2... \tag{36}$$

The second case arises when the left-hand side of (35) evaluated under the null becomes $nE\left\{Y_i^2\right\}E\left\{Y_s^2\right\} = n$, if $j_i = k_i = 1$ and $j_s = k_s = 1$.

Step 4: The same procedure is used to find the elements of the matrices C_{ii}'s and F_{ii}'s, although now there will be more non zero elements. As in the case of the B_{ii}'s, all elements will be found evaluating, under the null, expressions of the form (31) for different values of q and r, while also making frequent use of (36). Also, in what follows one should keep in mind that all matrices are symmetric, so that it suffices to list the entries on the diagonal and on either the upper or lower triangular parts.

In the case of C_{11}, obtained when $q = 1$ and $r = 3$ (and viceversa) in (31), the non zero entries are equal to: $3n$ if $j_i = 1$ and $k_i = 3$; and n if $j_i = k_i = 1$ and $k_s = 2$.

In the case of C_{22}, arising when $q = 2$ and $r = 4$ (and viceversa) in (31), the non zero entries are equal to: $12n$ if $j_i = 2$ and $k_i = 4$; $2n$ if $j_i = k_i = 2$ and $k_s = 2$; n if $j_i = k_i = 1$, $j_s = k_s = 1$ and $k_t = 2$; and $3n$ if $j_i = 1$, $k_i = 3$ and $j_s = k_s = 1$.

In the case of F_{11}, resulting when $q = 3$ and $r = 3$ in (31), the non zero entries are equal to: $15n$ if $j_i = k_i = 3$; $3n$ if $j_i = 1$, $j_s = 2$ and $k_i = 3$, or if $j_i = k_i = 2$ and $j_s = k_s = 1$; and n if $j_i = k_i = 1$, $j_s = k_s = 1$ and $j_t = k_t = 1$, or if $j_i = k_i = 1$, $j_s = 2$ and $k_t = 2$.

Finally, in the case of F_{22}, obtained when $q = 4$ and $r = 4$ in (31), the non zero entries are equal to: $96n$ if $j_i = k_i = 4$; $12n$ if $j_i = j_s = 2$ and $k_i = 4$; $15n$ if $j_i = k_i = 3$ and $j_s = k_s = 1$; $3n$ if $j_i = 1$, $k_i = 3$, $j_s = k_s = 1$ and $j_t = 2$, or if $j_i = k_i = 2$, $j_s = k_s = 1$ and $j_t = k_t = 1$; $9n$ if $j_i = 3$, $k_i = 1$, $j_s = 1$ and $k_s = 3$; $8n$ if $j_i = k_i = 2$ and $j_s = k_s = 2$; $2n$ if $j_i = k_i = 2$, $j_s = 2$ and $k_t = 2$; and n if $j_i = k_i = 1$, $j_s = k_s = 1$, $j_t = 2$ and $k_u = 2$, or if $j_i = k_i = 1$, $j_s = k_s = 1$, $j_t = k_t = 1$ and $j_u = k_u = 1$.

Step 5: The elements of d in (33) are, on the other hand, easily found. They simply involve population moments that can be evaluated using (36), and sample moments, which are the ones that will appear at the end in LM_p.

Step 6: Using the above results one can now display all the vectors and matrices in (33), and calculate directly LM_p (after some algebra). The number of degrees of freedom of the Chi-square is simply equal to $C(p+3, 4)+C(p+2, 3)$, following subsection 2.1.

Acnowledgments

The first version of this paper was presented at the IX Latin American Meeting of the Econometric Society held in Santiago de Chile in 1989, and it was circulated as a working paper (Urzúa, 1989). I am indebted to Robert F. Engle, Thomas B. Fomby, Marc Nerlove, and the participants in seminars at several institutions over the last few years for helpful comments. As usual, I am the only one responsible for any errors.

Notes

1. As noted in footnote 2.

2. An analogous reasoning would apply if we had chosen to work with more general Q-exponentials. For instance, if the multivariate sextic exponential were to be considered as the likely alternative to the multivariate normal (and the quartic exponential), then the omnibus LM test would be made of quadratic terms involving all possible third, fourth, fifth and sixth moments (pure and mixed).

3. The Monte Carlo study is available upon request from the author. A $33-$line procedure written in GAUSS to compute LM_p is also freely available upon request.

4. Or one can also use a similar result to Proposition 3 in Lütkepohl and Theilen (1991).

References

[1] Aitchison, J., and S. D. Silvey (1958) "Maximum Likelihood Estimation of Parameters Subject to Restraints." Annals of Mathematical Statistics 29, 813-828.

[2] Bera, A, and S. John (1983) "Tests for Multivariate Normality with Pearson Alternatives." Communications in Statistics, Theory and Methods 12, 103-117.

[3] Bowman, K. O., and L. R. Shenton (1975) "Omnibus Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2." Biometrika 62, 243-250.

[4] Cox, D. R. (1977) "The Role of Significance Tests." Scandinavian Journal of Statistics 4, 49-70.

[5] Engle, R. F. (1984) "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." Pp. 775-826 in Handbook of Econometrics, Vol. 2, edited by Z. Griliches and M. Intriligator. Amsterdam, North Holland.

[6] Fisher, R. A. (1922) "On the Mathematical Foundations of Theoretical Statistics." Philosophical Transactions of the Royal Society of London, Series A 222, 309-368.

[7] Fisher, R. A. (1930) "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality." Proceedings of the Royal Statistical Society, Series A 130, 16-28.

[8] Gart, J. J., and R. E. Tarone (1983) "The Relation Between Score Tests and Approximate UMPU Tests in Exponential Models Common in Biometry." Biometrics 39, 781-786.

[9] Jarque, C. M., and A. K. Bera (1980) "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals." Economics Letters 6, 255-259.

[10] Jarque, C. M., and A. K. Bera (1987) "A Test for Normality of Observations and Regression Residuals." International Statistical Review 55,

163-172.

[11] Jarque, C. M., and C. R. McKenzie (1983) "Testing for Multivariate Normality in Simultaneous Equations Models." Working Paper in Economics and Econometrics No. 82, Australian National University.

[12] Jaynes, E. T. (1957) "Information Theory and Statistical Mechanics." Physical Review 106, 620-630.

[13] Klir, G. J., and T. A. Folger (1988) "Fuzzy Sets, Uncertainty, and Information." New Jersey, Prentice Hall.

[14] Lütkepohl, H., and B. Theilen (1991) "Measures of Multivariate Skewness and Kurtosis for Tests of Nonnormality." Statistical Papers 32, 179-193.

[15] Mardia, K. V. (1970) "Measures of Multivariate Skewness and Kurtosis with Applications." Biometrika 57, 519-530.

[16] Mardia, K. V. (1987) "Tests of Univariate and Multivariate Normality." Pp. 279-320 in Handbook of Statistics, Vol. 1, edited by P. R. Krishnaiah. Amsterdam, North-Holland.

[17] Mardia, K. V., and J. T. Kent (1991) "Rao Score Tests for Goodness of Fit and Independence." Biometrika 78, 355-363.

[18] Rao, C. R. (1948) "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation." Proceedings of the Cambridge Philosophical Society 44, 49-57.

[19] Shannon, C. E. (1948) "A Mathematical Theory of Communication." Bell System Technical Journal, 27, 379-423 and 623-656.

[20] Small, N. J. H. (1985) "Multivariate Normality: Testing For." Pp. 95-100 in Encyclopedia of Statistical Sciences, Vol. 6, edited by S. Kotz and N. L. Johnson. New York, John Wiley.

[21] Urzúa, C. M. (1988) "A Class of Maximum-Entropy Multivariate Distributions." Communications in Statistics, Theory and Methods 17, 4039-4057.

[22] Urzúa, C. M. (1989) "Tests for Multivariate Normality of Observations and Residuals." Working Paper No. III-89, Centro de Estudios Económicos, El Colegio de México.

[23] Urzúa, C. M. (1990) "Lending to Sovereign Borrowers: Snapshots of the Eurocurrency Market Using a Thomian Camera." International Economic Review 31, 469-490.

[24] Urzúa, C. M. (1997) "On the Correct Use of Omnibus Tests for Normality." Economics Letters, 53, 247-251.

[25] White, H., and G. M. MacDonald (1980) "Some Large-Sample Tests for Nonnormality in the Linear Regression Model." Journal of the American Statistical Association 75, 16-28.

[26] Zellner, A., and R. A. Highfield (1988) "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions." Journal of Econometrics 37, 195-209.