ELSEVIER

Economics Letters 53 (1996) 247–251

economics letters

On the correct use of omnibus tests for normality

Carlos M. Urzúa[a,b,*]

[a]El Colegio de México, Camino al Ajusco No. 20, México, DF 01000, Mexico
[b]ITESM-CCM

Received 6 July 1996; accepted 7 November 1996

Abstract

This letter warns about the incorrect use of the popular Jarque–Bera test for normality of residuals in the case of small- and medium-size samples. It also provides a natural modification of the test that mitigates the problem.

Keywords: Normality test; Omnibus test

JEL classification: C10; C20

1. Introduction

The test for univariate normality of observations and residuals introduced by Jarque and Bera (1980, 1987) has gained great acceptance among economists. It is an omnibus test based on the standardized third and fourth moments:

$$LM = n\left(\frac{(b_1^{1/2})^2}{6} + \frac{(b_2 - 3)^2}{24}\right), \tag{1}$$

where n is the number of observations, $b_1^{1/2} = m_3/m_2^{3/2}$, $b_2 = m_4/m_2^2$ and m_i is the ith central moment of the observations (i.e. $m_i = \Sigma\,(x_j - \bar{x})^i/n]$. Asymptotically, the hypothesis of normality is rejected at some significance level if the value of LM exceeds the critical value of a chi-squared distribution with two degrees of freedom. In the more usual case of a regression, (1) is calculated using the estimated residuals.

As shown by Jarque and Bera (1987), the test performs quite well compared with others available in the literature. This is not surprising since they proved that, if the alternatives to the normal distribution are in the Pearson family, LM is the corresponding Lagrange

* Correspondence address: Centro de Estudios Económicos, El Colegio de México, Camino Al Ajusco No. 20, México, DF 01000, Mexico. Tel: (52-5) 645-5955, ext. 4171; fax: (52-5) 645-0464; e-mail: curzua@colmex.mx.

PII S0165-1765(96)00923-8

multiplier test for normality. Urzúa (1997) shows the same when the alternatives are the maximum-entropy ("most likely") distributions with finite moments defined by Urzúa (1988).

However, the good performance of the test is highly dependent on the use, through a Monte Carlo simulation, of empirical significance points (something, by the way, that is almost never done in studies where (1) is used). This is so because of the slow convergence in the distribution to the chi-squared value.

Interestingly enough, (1) has been known among statisticians since the work of Bowman and Shenton (1975). They derived it after noting that, under normality, the asymptotic means of $b_1^{1/2}$ and b_2 are 0 and 3, the asymptotic variances are $6/n$ and $24/n$, and the asymptotic covariance is zero. Thus, LM is just the sum of squares of two asymptotically independent standardized normals.

Yet, there are few (if any) instances in the statistics literature where the Bowman–Shenton–Jarque–Bera test has been used. As one author flatly states in a comprehensive survey of tests for normality: "Due to the slow convergence of b_2 to normality this test is not useful." (D'Agostino, 1986, p. 391).

2. A new better-behaved test statistic

This section presents a new better-behaved omnibus test for normality that is a natural extension of the Jarque–Bera test. The idea is straightforward; instead of the asymptotic means and variances of the standardized third and fourth moments, use their exact means and variances. Under normality, the latter can be easily computed using results already known to Fisher (1930).

Fisher's results were stated in terms of the so-called k statistics, which can be expressed in terms of moments as (Stuart and Ord, 1987, pp. 392, 422):

$$k_2 = \frac{nm_2}{n-1}, \quad k_3 = \frac{n^2 m_3}{(n-1)(n-2)}, \quad k_4 = \frac{n^2[(n+1)m_4 - 3(n-1)m_2^2]}{(n-1)(n-2)(n-3)}.$$

For our purposes, his relevant derivations are that, under normality, k_2 is independent of $k_p/k_2^{p/2}$ for $p = 3, 4, \ldots$, and that

$$\mathrm{var}\left(\frac{k_3}{k_2^{3/2}}\right) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)}, \quad \mathrm{var}\left(\frac{k_4}{k_2^2}\right) = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}.$$

We can then use these results to show easily that, under normality, the exact mean and variance of the standardized third and fourth moments are

$$E(b_1^{1/2}) = 0, \quad \mathrm{var}(b_1^{1/2}) = \frac{6(n-2)}{(n+1)(n+3)}, \tag{2}$$

$$E(b_2) = \frac{3(n-1)}{(n+1)}, \quad \mathrm{var}(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}. \tag{3}$$

Table 1
Significance points for two tests for normality of observations

n	10	20	35	50	75	100	150	200	300	800	∞
ALM											
20%	2.14	2.25	2.35	2.42	2.62	2.69	2.85	2.88	2.92	3.07	3.22
15%	2.77	2.81	2.84	2.95	3.20	3.16	3.38	3.44	3.50	3.65	3.79
10%	4.23	4.01	3.81	3.89	4.17	4.05	4.26	4.31	4.32	4.47	4.61
5%	7.70	7.15	6.38	6.43	6.45	6.04	6.34	6.22	6.15	6.00	5.99
1%	17.69	19.59	16.77	16.33	16.32	13.27	13.34	12.62	12.10	10.68	9.21
Ratio[a]	1.00	1.01	0.97	0.98	1.02	0.98	1.00	1.00	1.01	1.00	1.00
LM											
20%	1.14	1.58	1.96	2.18	2.37	2.45	2.64	2.74	2.90	3.06	3.22
15%	1.30	1.85	2.32	2.60	2.83	2.87	3.10	3.25	3.43	3.62	3.79
10%	1.68	2.35	2.97	3.32	3.55	3.58	3.77	4.00	4.23	4.44	4.61
5%	2.65	3.76	4.83	5.15	5.34	5.45	5.49	5.71	5.74	5.86	5.99
1%	6.04	10.32	12.08	12.21	12.90	12.31	11.76	11.32	10.86	10.38	9.21
Ratio[a]	0.47	0.66	0.80	0.85	0.89	0.88	0.92	0.95	0.97	0.99	1.00

Source: Own simulations using 10000 replications.
[a] Ratio of the empirical mean to the mean of the asymptotic distribution.

Hence, using (2) and (3), we can finally define the new test, to be called the adjusted Lagrange multiplier test for normality, as

$$ALM = \frac{(b_1^{1/2})}{\mathrm{var}(b_1^{1/2})} + \frac{[b_2 - E(b_2)]^2}{\mathrm{var}(b_2)} . \tag{4}$$

This new test statistic behaves much better for small- and medium-size samples than the Jarque–Bera statistic, as can be assessed from the Monte Carlo simulations reported in Table 1. In the case of $\alpha = 10\%$, the significance level commonly used to test for normality, the behavior of *ALM* is particularly good. It should be noted in passing that the significance points in that table, for both *ALM* and *LM*, can be used to test for normality of observations, but they cannot be used in the case of regression residuals since, for each particular regression, the significance points depend on the design (regressor) matrix and the distribution of the residuals (e.g., Weisberg, 1980).

3. Estimated power of the tests

This section compares the power of *ALM* and *LM* when used as tests for normality of regression residuals. The Monte Carlo simulation procedures used by us were, on purpose, identical with those employed by White and MacDonald (1980) in their much-quoted paper on the subject. As in there, the five alternatives to the normal distribution of the residuals were

Table 2
Tests for normality of residuals; estimated power with 10000 replications, using as significance point $\chi^2_{2,0.10}$

n		t_5	Heteroscedastic normal	χ^2_2	Laplace	Log-normal
20	*ALM*	0.231	0.091	0.493	0.290	0.808
	LM	0.140	0.039	0.380	0.181	0.727
35	*ALM*	0.362	0.116	0.829	0.464	0.985
	LM	0.293	0.077	0.782	0.374	0.978
50	*ALM*	0.467	0.128	0.963	0.595	1.000
	LM	0.406	0.093	0.950	0.513	0.999
100	*ALM*	0.694	0.161	1.000	0.835	1.000
	LM	0.658	0.135	1.000	0.797	1.000

as follows: Student's t distribution with five degrees of freedom; heteroscedastic normal distribution; chi-squared distribution with two degrees of freedom; Laplace distribution (double exponential); log-normal distribution (all of them standardized to have mean zero and variance 25). Furthermore, for the generation of pseudo-random numbers we followed in each case the same computational procedure as in their paper.

Also, following White and MacDonald (1980), the design matrices for the regressions were constructed adding to a column of ones three columns of uniform random numbers with mean zero and variance 25. The number of rows in each design matrix (i.e. the sample size) was given by $n = 20, 35, 50, 100$.

As a first exercise, we estimated the power of both tests when, as is incorrectly done in almost all empirical studies, the significance point is taken to be $\chi^2_{2,0.10} = 4.61$, even though the sample sizes are not large. The number of replications in each Monte Carlo simulation was 10 000 (instead of 200 in the work of White and MacDonald (1980)), and the results are presented in Table 2.

As can be appreciated there, the results are overwhelmingly in favor of the new *ALM* test. It comes first in all the distributions considered and all the sample sizes. Furthermore, in the case of the smaller samples the power of *ALM* is significantly larger than the power of *LM*.

Naturally, the next question to ask is whether the same relative performance is obtained when, prior to applying the tests, "correct" significance points are found for each test using Monte Carlo simulations (this is actually the procedure that was explicitly suggested by Jarque and Bera (1987). The results obtained in that way are presented in Table 3. Once again, *ALM* outperforms *LM*. Interestingly enough, the only five cases (out of 20) where the *LM* test comes first correspond to the distributions that are farther apart from the normal.

4. Concluding remarks

This paper has presented a new omnibus test for normality of residuals and observations: the adjusted Lagrange multiplier test *ALM*. As shown here, the *ALM* test outperforms in

Table 3
Tests for normality of residuals; estimated power with 10000 replications, using estimated significance point ($\alpha = 0.10$)

n		t_5	Heteroskedastic normal	χ_2^2	Laplace	Log-normal
20	*ALM*	0.254	0.477	0.533	0.317	0.831
	LM	0.247	0.456	0.586	0.306	0.856
35	*ALM*	0.391	0.724	0.864	0.494	0.989
	LM	0.376	0.697	0.896	0.470	0.992
50	*ALM*	0.493	0.852	0.973	0.624	1.000
	LM	0.474	0.831	0.982	0.595	1.000
100	*ALM*	0.712	0.984	1.000	0.849	1.000
	LM	0.698	0.981	1.000	0.836	1.000

terms of power the traditional Jarque–Bera *LM* test, both when significance points are directly taken from a chi-square distribution, or when the "correct" significance points are obtained through simulations. Thus, the use of *ALM* over *LM* seems warranted in both circumstances. As a final point, a similar adjustment can be extended to the multivariate tests for normality that are also based on third and fourth standardized moments, as shown by Urzúa (1997).

References

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.

D'Agostino, R.B., 1986, Tests for the normal distribution, in: R.B. D'Agostino and M.A. Stephens, eds., Goodness of fit techniques (Marcel Dekker, New York), 367–419.

Fisher, R.A., 1930, The moments of the distribution for normal samples of measures of departure from normality, Proceedings of the Royal Statistical Society A 130, 16–28.

Jarque, C.M. and A.K. Bera, 1980, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Economics Letters 6, 255–259.

Jarque, C.M. and A.K. Bera, 1987, A test for normality of observations and regression residuals, International Statistical Review 55, 163–172.

Stuart, A. and J.K. Ord, 1987, Kendall's advanced theory of statistics, Vol. 1 (Oxford University Press, Oxford).

Urzúa, C.M., 1988, A class of maximum-entropy multivariate distributions, Communications in Statistics, Theory and Methods 17, 4039–4057.

Urzúa, C.M., 1997, Omnibus tests for multivariate normality based on a class of maximum entropy distributions, in: T.B. Fomby and R.C. Hill, eds., Advances in econometrics: applying maximum entropy to econometric problems (JAI Press, Connecticut), in press.

Weisberg, S., 1980, Comment to some large-sample tests for nonnormality in the linear regression model, Journal of the American Statistical Association 75, 28–31.

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.