31/1/05

More on the Correct Use of Omnibus Tests for Normality

Geoffrey Poitras
Faculty of Business Administration
Simon Fraser University
Burnaby, B.C.
CANADA VSA lS6
email: poitras@sfu.ca
website: www.sfu.ca/~poitras

ABSTRACT

This Monte Carlo study compares the small sample properties of some commonly used omnibus and directional tests, based on the standardized third and fourth moments, for assessing the normality of random variables: the omnibus D'Agostino K^2 test and the directional components, and three versions of the Jarque-Bera test. The simulation results demonstrate that the selection of the directional tests used to form the omnibus tests can have a significant impact on normality test power in finite samples.

Keywords: Normality test; Omnibus Test

JEL Classification: C10, C20

* Geoffrey Poitras is a Professor in the Faculty of Business Administration, Simon Fraser University.
The author would like to thank Peter Kennedy, Michael Stephens, John Heaney and members of the Mathematical Statistics workshop at SFU for useful comments.

More on the Correct Use of
Omnibus Tests for Normality

1. Introduction

Few subjects in applied statistics have received as much attention as the problem of testing whether a sample is from a normal population. For more than a century, the subject has attracted the attention of leading figures in statistics. Though K. Pearson made a seminal contribution more than a quarter of a century earlier, Fisher (1928, 1930), E. Pearson (1930) and Wishart (1930) provide essential early results on tests using the standardized sample moments of skewness ($\surd b_1$) and kurtosis (b_2). The absence of exact solutions for the sampling distributions generated a large number of simulation studies exploring the power of these statistics as both directional and omnibus tests, e.g., Pearson (1963), D'Agostino and Pearson (1973), Bowman and Shelton (1975). Despite convincing evidence from these studies that convergence of the sampling distributions to asymptotic results was very slow, especially for b_2, many studies of normality testing in econometrics have emphasized the asymptotic Jarque-Bera test, e.g., Jarque and Bera (1980, 1987), Rilstone (1992). Some confusion over the appropriate method of specifying the sampling distribution persists, e.g., Urzúa (1996), and the appropriate testing procedure to use in specific situations remains unclear, e.g., Atwood, et al. (2003). In particular, when employing $\surd b_1$ and b_2 tests, the finite sample implications of using a specific omnibus or directional test is generally unrecognized.

2. Moment $\surd b_1$ and b_2 tests

The appropriate construction of $\surd b_1$ and b_2 tests for normality has gradually been recognized in econometrics, e.g., Dufour et al. (1998), Önder and Zaman (2005). Under general circumstances, the power of tests in this category compare favorably with empirical distribution function (EDF) tests, e.g., the Anderson-Darling test, and correlation tests based on the order statistics, e.g., the Shapiro-

Wilk test (Thode 2002). In practice, the null hypothesis of normality is usually specified in composite form where μ and σ are unknown. Specification of the alternative hypothesis can involve: a specific distribution, e.g., the uniform; a class of distributions with identifiable shape properties, e.g., fat-tailed symmetric; or a general non-normal alternative. While an omnibus test combining the information in $\sqrt{b_1}$ and b_2 is indicated for a general non-normal alternative, increases in power can be obtained where the class of alternatives can be narrowed. Where a specific alternative distribution is indicated, the use of an appropriately specified likelihood ratio test makes it is possible that a uniformly most powerful, location and scale invariant test may be available. However, due to analytical complexity of the solutions, the number of cases where such directional tests are available for practical application is limited (Uthoff 1970). In cases where some information about the alternative distribution is available, a directional test using only $\sqrt{b_1}$ or b_2 is indicated.

Considerable attention has been given to the construction of omnibus tests that combine information from $\sqrt{b_1}$ and b_2, e.g., Thode.(2002, p.54-8), Jarque-Bera (1980), D'Agostino and Stephens, (1986, ch.7 and 9), Urzúa (1986). A number of approaches have been suggested. The simplest construction is found with the Jarque-Bera (*JB)* test (Jarque and Bera,1980) where the asymptotic normal values for skewness and kurtosis are used to construct a $\chi^2(2)$ test involving the first two moments of the asymptotic distributions:

$$JB = T\left[\frac{[\sqrt{b_1}]^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2(2)$$

where T is the sample size, $\sqrt{b_1} = m_3 / (m_2)^{3/2}$, $b_2 = m_4 / (m_2)^2$ and the central moments are defined as $m_j = \Sigma\,(x_j - m_1)^j / T$ and m_1 is the sample mean. Urzúa (1996) improves on this formulation by using results from Fisher (1930) for the mean and variance of the sampling distribution to formulate a chi-squared test with better small sample properties. Even with this improvement, neither formulation recognizes that the sampling distributions may not be close to normal in finite samples. Due to the slow convergence to the asymptotic results, this form of the omnibus test will have problems even in moderately sized samples, e.g., $T = 100$.

The sampling distribution problems associated with the *JB* test are confirmed by various simulation studies that have found the *JB* test to be incorrectly sized in small and moderate samples, e.g., Poitras (1992), Dufour et al. (1998). The size distortions are so significant that various studies recommend against the use of the *JB* test, in favor of other omnibus moment tests such as D'Agostino's K^2 test. The Monte Carlo results given below confirm this result and demonstrate that, while having slightly improved size compared to *JB*, the Urzúa form of the test still is incorrectly sized. An obvious solution to this problem is to use Monte Carlo methods (e.g., Dufour and Kahalf 2001) to correctly size the *JB* test by simulating the correct critical values. Being formed from the location and scale invariant statistics, $\sqrt{b_1}$ and b_2, makes the *JB* test an excellent candidates for such resizing. This raises the question about the power of the correctly sized *JB* test relative to alternative omnibus $\sqrt{b_1}$ and b_2 tests. Presumably, the sampling distribution problems that led to the incorrect sizing of the *JB* test will persist when the test is correctly sized. This intuition is not supported in the Monte Carlo simulations reported below where the size-corrected *JB* test is found to have marginally higher power than the K^2 test against a range of (true) alternative distributions with fat-tails.

The asymptotic-omnibus approach employed in econometrics is in stark contrast to the approach developed in statistics where the distributional properties of the directional tests were formulated well before efforts were expended on an omnibus test. Early theoretical work determined the moments of the distribution of population kurtosis (and skewness) in normal samples as well as the associated properties of the sampling distribution. Series solutions were used to fit moment approximations to the actual distributions. In order the accurately approximate the finite sample distributions for $\sqrt{b_1}$ and b_2, information from the third and fourth moment of the sampling distributions for these statistics had to be incorporated. From this point, one line of research led to the tabulation of critical values of the distributions for $\sqrt{b_1}$ (Pearson and Hartley 1966) and b_2, (D'Agostino and Pearson 1973). Knowledge of critical values permitted properly sized directional tests to be implemented. Subsequent results determined the appropriate transformation of the sampling distributions to achieve normality of $\sqrt{b_1}$ and b_2 in finite samples. In particular, while the distributions for $\sqrt{b_1}$ and b_2 are asymptotically normal, large sample sizes are needed before the distributions are well behaved. D'Agostino (1970) determined that Johnson's unbounded S_U curve was appropriate for transforming $\sqrt{b_1}$. Similar work on kurtosis (e.g., D'Agostino and Pearson 1973, Bowman and Shelton 1975) led eventually to an effective and simple transform of b_2 to normality based on the Wilson-Hilferty transformation (Anscombe and Glynn 1983).

The advantage of having transforms for the sampling distributions of $\sqrt{b_1}$ and b_2 to normality is that a correctly sized omnibus χ^2 test can be specified to combine the information from the two sample moments. Recognizing that a number of possible avenues can be used to construct such an omnibus test, the primary advantage of this form of the test is the simplicity provided by the χ^2 framework. The precise construction of the test depends on the specific transformations selected (see Thode 2002, p.54-8 for other variations). The calculations involved in D'Agostino's K^2 test are based on Johnson's S_u approximation for $\sqrt{b_l}$ and Anscombe and Glynn's approximation to b_2 (D'Agostino and Stephens 1986). For the S_u approximation for $\sqrt{b_l}$ first calculate:

$$Y = \sqrt{b_1} \left[\frac{(T + 1)\ (T + 3)}{6\ (T - 2)} \right]^{1/2}$$

$$\beta_2(\sqrt{b_1}) = B_2 = \frac{3\ (T^2 + 27\ T - 70)\ (T + 1)\ (T + 3)}{(T - 2)\ (T + 5)\ (T + 7)\ (T + 9)}$$

$$W^2 = -1 + (2\ (B_2 - 1))^{1/2} \qquad \delta = \frac{1}{\sqrt{\log\ W}} \qquad \phi = \sqrt{\frac{2}{W^2 - 1}}$$

It follows that:

$$Z_1 = \delta\ \log \left[\frac{Y}{\phi} + \sqrt{\left(\frac{Y}{\phi}\right)^2 + 1} \right]$$

where Z_1 is approximately normally distributed under the null hypothesis of normality.

For the Anscombe and Glynn (1983) transformation of the sampling distribution of b_2 use:

$$E[b_2] = \frac{3\ (n - 1)}{(n + 1)} \qquad var[b_2] = \frac{24\ T\ (T - 2)\ (T - 3)}{(T + 1)^2\ (T + 3)\ (T + 5)}$$

Computing the standardized value of b_2:

$$x = \frac{b_2 - E[b_2]}{\sqrt{var[b_2]}}$$

The third standardized moment of the sampling distribution of b_2 is calculated as:

$$\sqrt{\beta_1[b_2]} = \frac{6\ (T^2 - 5T + 2)}{(T + 7)\ (T + 9)} \sqrt{\frac{6\ (T + 3)\ (T + 5)}{T\ (T-2)\ (T - 3)}}$$

It is observed that a reciprocal chi-squared variable fits the distribution for $T > 30$. Anscombe and Glynn then let A denote the degrees of freedom for the chi-squared distribution and solve for A to equate the third moment of the theoretical and sampling distributions. It is now possible to compute:

$$A = 6 + \frac{8}{\sqrt{\beta_1[b_2]}} \left[\frac{2}{\sqrt{\beta_1[b_2]}} + \sqrt{1 + \frac{4}{\beta_1[b_2]}} \right]$$

Then:

$$Z_2 = \frac{\left(1 - \frac{2}{9A}\right) - \left[\frac{1 - \frac{2}{A}}{1 + x\sqrt{\frac{2}{A-4}}}\right]^{1/3}}{\sqrt{\frac{2}{9A}}}$$

is approximately standard normal under the null hypothesis of normality. From these results D'Agostino's K^2 statistic is calculated as:

$$K^2 = (Z_1)^2 + (Z_2)^2$$

which is chi-squared with two degrees of freedom.

3. Directional versus Omnibus Testing: Monte Carlo results

While it would seem desirable to test against as many possible alternative distributions as possible, it is well known that this can result in a significant loss of power. For example, at a given significance level (α) an omnibus $\sqrt{b_1}$ and b_2 test would require higher values of b_1 to reject a skewed alternative than would be required for a directional test using just b_1 at the same α. This issue is complicated for omnibus $\sqrt{b_1}$ and b_2 tests because, though $\sqrt{b_1}$ and b_2 are uncorrelated, the sampling distributions are not independent. Even for a (true) symmetric non-normal alternative, $\sqrt{b_1}$ will eventually reject the null, just as b_2 will eventually reject a (true) skewed distribution. As such, the Monte Carlo results given in Table 1 provide evidence on the loss of power associated with choosing an omnibus test over a directional test at a specific finite sample size, as well as the speed of convergence to the asymptotic results. In addition, because the sampling distributions are not well-behaved in finite samples, it is an open question as to whether a size-correctly *JB* test will have the same finite sample power as the D'Agostino K^2 statistic due to the reliance on the asymptotic normality of $\sqrt{b_1}$ and b_2

to formulate the $\chi^2(2)$ test. This issue is also addressed in Table 1.

The Monte Carlo simulations in Table 1 report the power of seven normality test statistics for nine different distributional specifications of the population residuals. The tests are: the D'Agostino's K^2 test and the directional components $\sqrt{b_1}$ and b_2 ; three forms of the Jarque-Bera test, the asymptotic version (*JB)*, the Urzúa (1996) version that adjusts for the sampling distribution of the mean and variance (*JB*)*, and a version of the test that uses size-adjusted critical values (*JBSC*). Results are also reported for the studentized range, the most powerful scale and location invariant test for normality against a uniformly distributed alternative (Uthoff 1970). The nine distributions examined are the normal, uniform and lognormal distributions, together with six versions of the stable distribution. The stable distributions are characterized by four parameters; in addition to location and scale parameters, there is a dispersion parameter, the characteristic exponent (α': $0 \leq \alpha' \leq 2$), and a skewness parameter (δ: $0 \leq \delta \leq 1$). The class of stable distributions include the normal ($\alpha' = 2$) and Cauchy ($\alpha' =1$) as special cases. The stable distributions are well suited to simulation studies of normality tests due to the ability to precisely control variation in dispersion and skewness. Each cell in the table reports the rejection frequency. When the null is true, small deviations from the critical value of α=0.100 can be due to Monte Carlo sampling variation of the statistics. Deviations beyond 0.009 are due to the incorrect sizing associated with a significantly biased statistic.

The Monte Carlo results confirm an incorrect sizing of the *JB* and *JB** tests severe enough to warrant a recommendation that these tests not be used. While the correction provided by Urzúa (1996) does partially alleviate the size distortion of the asymptotic test, the improvement is insufficient to warrant use of the test in practice. In contrast, the size corrected version of the Jarque-Bera test *(JBSC)* has decided strengths. The *JBSC* has marginally higher power than the D'Agostino

K^2 against all the alternative distributions with population kurtosis larger than for the normal distribution. Only a directional $\sqrt{b_1}$ test has higher power against (true) skewed alternatives. This promising performance for *JBSC* is undermined by the strikingly poor performance when the thin-tailed uniform distribution is the true alternative. The loss in power compared to D'Agostino's K^2 is still substantial even for samples of T=50. If the alternative hypothesis is a general non-normal alternative, where there is no prior knowledge as to whether the distribution is thin or fat tailed, then the results for the uniform argue against the use of *JBSC* as an omnibus test, in favor of D'Agostino K^2. Comparing *JBSC* with K^2, the loss in power for the fat tailed alternatives is slight compared to overwhelming gains when the thin-tailed alternative is true. A directional test based on b_2 does compare well against all tests except the directional *SR* when the uniform alternative is true, but there is a loss in power compared to D'Agostino's K^2 for the other alternative distributions. On balance, for normality testing where there is no prior information that the alternative distribution is fat-tailed, then D'Agostino's K^2 is the preferred omnibus test.

Table 1*

Estimated Normality Test Powers (α=10%) for Population Residuals when the Null Hypothesis of Normality is True: T = 25, 50

	$\sqrt{b_1}$	b_2	K^2	*SR*	*JB*	*JB**	*JBSC*
T=25	.099	.098	.098	.101	.039	.074	.100
T=50	.098	.092	.096	.098	.048	.068	.100

Estimated Normality Test Powers (α=10%) for Population Residuals when the Alternative Hypothesis is True: T = 25, 50

	$\sqrt{b_1}$	b_2	K^2	*SR*	*JB*	*JB**	*JBSC*
True H_1*: Stable* (α'=1.9, δ=0)							
T=25	.228	.209	.228	.184	.167	.216	.227
T=50	.646	.651	.657	.640	.638	.665	.671
True H_1*: Stable* (α'=1.9, δ=1)							
T=25	.255	.218	.251	.164	.182	.237	.253
T=50	.677	.656	.672	.624	.653	.672	.687
True H_1*: Stable* (α'=1.6, δ=0)							
T=25	.526	.544	.562	.491	.499	.577	.573
T=50	.837	.888	.893	.862	.882	.898	.901
True H_1*: Stable* (α'=1.6, δ=1)							
T=25	.649	.522	.603	.327	.542	.594	.637
T=50	.952	.876	.934	.776	.514	.717	.939
True H_1*: Stable* (α'=1.0, δ=0)							
T=25	.848	.926	.930	.839	.916	.945	.941
T=50	.959	.999	.999	.990	.999	.999	.999
True H_1*: Stable* (α'=1.0, δ=l)							
T=25	.984	.895	.968	.477	.954	.964	.978
T=50	1.000	.998	1.000	.863	1.000	1.000	1.000
True H_1*: Uniform*							
T=25	.018	.600	.407	.701	.000	.001	.047
T=50	.502	.969	.938	.991	.513	.549	.749
True H_1*: Lognormal*							
T=25	.971	.736	.914	.240	.870	.910	.962
T=50	1.000	.962	1.000	.710	.999	1.000	1.000

----

* Tabulated critical values for $\sqrt{b_1}$, b_2 taken from Bowman and Shelton (ch.7 in D'Agostino and Stephens, 1986). Critical values for the studentized range (*SR*) provided by Pearson and Stephens (1964). Size corrected *JB* test critical values are (2.645 for *T=25*, 3.155 for *T*=50). The number of iterations at each true alternative (null) sample size are 7000 (21,000) for *T=25,* 3500 (10,500) for *T=50* .

References

Anscombe, F. and W. Glynn (1983), “Distribution of the kurtosis statistic b_2 for normal statistics”, Biometrika 70: 227-34.

Atwood, J, S. Shaik and M. Watts (2003), “Are Crop Yields Normally Distributed? A Reexamination”, American Journal of Agricultural Economics 85: 888-901.

Bowman, K. and L. Shelton (1975), “Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2", Biometrika 62: 243-50.

D’Agostino, R. and E. Pearson (1973), “Tests for departures from normality. Empirical results for the distribution of $\sqrt{b_1}$ and b_2", Biometrika 60: 613-22.

D’Agostino, R. and M. Stephens (1986), Goodness-of-fit Techniques, New York: Marcel Dekker.

Dufour, J., A. Farhat, L. Gardiol and L. Khalaf (1998), “Simulation-based finite sample normality tests in linear regressions””, Econometrics Journal 1:C154-C173.

Dufour, J. and L. Khalaf (2001), “Monte Carlo Tests in Econometrics”, chap. 23 in B. Baltagi (ed.), Companion to Theoretical Econometrics, Oxford, UK: Blackwell.

Fisher, R. (1928), “Moments and product moments of sampling distributions”, Proceedings of the London Mathematical Society (Series 2) 30: 199-238.

Fisher, R. (1930), “The moments of the distribution for normal samples of measures of departures from normality”, Proceedings of the Royal Society A 130: 16-28.

Jarque, C. and A. Bera (1980), "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals", Economics Letters, 6: 255-259.

Jarque, C. and A. Bera (1987), “A test for normality of observations and regression residuals”, International Statistical Review 55: 163-72.

Önder, A. and A. Zaman (2005), “Robust tests for normality of errors in regression models”, Economics Letters 86: 63-8.

Pearson, E. (1930), “A further development of tests for normality”, Biometrika 22: 239-49.

Pearson, E. (1963), “Some problems arising in approximating to probability distributions using moments”, Biometrika 50: 95-111.

Pearson, E. and H. Hartley (1966), Biometrika Tables for Statisticians (Vol.I, 3rd.ed), Cambridge, UK: Cambridge University Press.

Poitras, G. (1992), "Testing Regression Disturbances for Normality with Stable Alternatives: Further Monte Carlo Evidence", Journal of Statistical Computation and Simulation 41: 109-23.

Rilstone, P. (1992), "A Simple Bera-Jarque Normality Test for Nonparametric Residuals", Econometric Reviews 11: 355-65.

Thode, H (2002), Testing for Normality, New York: Marcel Dekker.

Urzúa, C. (1996), "On the correct use of omnibus tests for normality", Economics Letters 53:247-51.

Uthoff, V. (1970), "An optimum test property of two well-known statistics", Journal of the American Statistical Association 65: 1597-1600.

Wishart, J. (1930), "The deviation of certain higher-order sampling product moments from a normal population", Biometrika 22: 224.