Available online at www.sciencedirect.com

SCIENCE DIRECT®

Economics Letters 90 (2006) 304–309

economics letters

www.elsevier.com/locate/econbase

More on the correct use of omnibus tests for normality

Geoffrey Poitras

Simon Fraser University, Burnaby, B.C., Canada VSA lS6

Received 29 January 2005; received in revised form 1 July 2005; accepted 16 August 2005
Available online 24 January 2006

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.

Keywords: Normality test; Omnibus test

JEL classification: C10; C20

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. 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., D'Agostino and Pearson (1973), and 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 (Jarque and Bera, 1980, 1987). Some confusion over the appropriate method of specifying the sampling distribution persists, e.g., Urzúa (1996), and the

E-mail address: poitras@sfu.ca.
URL: www.sfu.ca/~poitras.

0165-1765/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2005.08.016

appropriate testing procedure to use in specific situations remains unclear, e.g., Atwood et al. (2003). In particular, when employing $\sqrt{b_1}$ and b_2 tests, the finite sample implications of using a specific omnibus or directional test are generally unrecognized.

2. Moment $\sqrt{b_1}$ and b_2 tests

The appropriate construction of $\sqrt{b_1}$ and b_2 tests for normality has gradually been recognized in econometrics, e.g., Dufour et al. (1998), and Önder and Zaman (2005). Under general circumstances, the power of tests in this category compare favorably with empirical distribution function tests and correlation tests based on the order statistics (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. 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.54–58), D'Agostino and Stephens (1986, ch. 7 and 9), and Urzúa (1996). A number of approaches have been suggested. The simplest construction is found with the Jarque–Bera (JB) test 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:

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

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 $m_j = \sum(x_j - m_1)^j/T$ and m_1 is the sample mean. Urzúa (1996) improves on this formulation by using results 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), and 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 Khalaf, 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 candidate 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 to 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 was 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 SU curve was appropriate for transforming $\sqrt{b_1}$. Similar work on kurtosis led eventually to an effective and simple transform of b_2 to normality based on the Wilson–Hilferty transformation (Anscombe and Glynn, 1983). The omnibus D'Agostino's K^2 test combines the appropriately transformed directional components $\sqrt{b_1}$ and b_2, e.g., Thode (2002, ch. 3).

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) omnibus $\sqrt{b_1}$ and b_2 tests 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-corrected 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

Table 1
Tabulated critical values for $\sqrt{b_1}$, b_2 taken from Bowman and Shelton (ch. 7 in D'Agostino and Stephens, 1986)

	$\sqrt{b_1}$	b_2	K^2	SR	JB	JB*	JBSC
Estimated normality test powers (α = 10%) for population residuals when the null hypothesis of normality is true: T = 25, 50							
$T=25$	.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							
True H_1: stable ($\alpha'=1.9$, $\delta=0$)							
$T=25$	.228	.209	.228	.184	.167	.216	.227
$T=50$	.646	.651	.657	.640	.638	.665	.671
True H_1: stable ($\alpha'=1.9$, $\delta=1$)							
$T=25$	.255	.218	.251	.164	.182	.237	.253
$T=50$	.677	.656	.672	.624	.653	.672	.687
True H_1: stable ($\alpha'=1.6$, $\delta=0$)							
$T=25$	.526	.544	.562	.491	.499	.577	.573
$T=50$	.837	.888	.893	.862	.882	.898	.901
True H_1: stable ($\alpha'=1.6$, $\delta=1$)							
$T=25$	.649	.522	.603	.327	.542	.594	.637
$T=50$	.952	.876	.934	.776	.514	.717	.939
True H_1: stable ($\alpha'=1.0$, $\delta=0$)							
$T=25$	.848	.926	.930	.839	.916	.945	.941
$T=50$	.959	.999	.999	.990	.999	.999	.999
True H_1: stable ($\alpha'=1.0$, $\delta=1$)							
$T=25$	.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

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 $T=50$.

(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. The nine distributions examined are the normal, uniform and lognormal distributions, together with six versions of the stable distribution. The eight alternative distributions selected all feature prominently in empirical studies. The uniform distribution arises in testing situations involving, for example, unit roots (e.g., Goldman and Tsurumi, 2003) and probability integral transforms (e.g., Clements, 2004). In addition to being widely used to model security prices, the lognormal has been used for the distribution of other variables such as firm size (e.g., Cabral and Mata, 2003). Stable distributions have a long history in both theoretical and applied distribution theory involving variables as diverse as stock prices and motion picture profits (e.g., De Vany and Walls, 2004). 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: $-1 \leq \delta \leq 1$). 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 $\alpha = 0.100$ can be due to Monte Carlo sampling variation of the statistics. Larger deviations 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 should not be used without appropriately sized critical values. 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 some prior information that the alternative distribution may be thin-tailed, then D'Agostino's K^2 is the preferred omnibus test.

Acknowledgements

The author would like to thank the anonymous referee, Peter Kennedy, Michael Stephens, John Heaney and members of the Mathematical Statistics workshop at SFU for useful comments.

References

Anscombe, F., Glynn, W., 1983. Distribution of the kurtosis statistic b_2 for normal statistics. Biometrika 70, 227–234.

Atwood, J., Shaik, S., Watts, M., 2003. Are crop yields normally distributed? A reexamination. American Journal of Agricultural Economics 85, 888–901.

Bowman, K., Shelton, L., 1975. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62, 243–250.

Cabral, L., Mata, J., 2003. On the evolution of the firm size distribution: facts and theory. American Economic Review 93, 1075–1091.

Clements, M., 2004. Evaluating the bank of England density forecasts of inflation. Economic Journal 114, 844–877.

D'Agostino, R., 1970. Transformation to normality of the null distribution of g_1. Biometrika 58, 679–681.

D'Agostino, R., Pearson, E., 1973. Tests for departures from normality. Empirical results for the distribution of $\sqrt{b_1}$ and b_2. Biometrika 60, 613–622.

D'Agostino, R., Stephens, M., 1986. Goodness-of-fit Techniques. Marcel Dekker, New York.
De Vany, A., Walls, D., 2004. Motion picture profit, the stable Paretian hypothesis, and the curse of the superstar. Journal of Economic Dynamics and Control 28, 1035–1058.
Dufour, J., Khalaf, L., 2001. Monte Carlo test methods in econometrics. In: Baltagi, Badi (Ed.), Companion to Theoretical Econometrics. Blackwell, Oxford, UK. Chapter 23.
Dufour, J., Farhat, A., Gardiol, L., Khalaf, L., 1998. Simulation-based finite sample normality tests in linear regressions. Econometrics Journal 1, C154–C173.
Goldman, E., Tsurumi, H., 2003. Asymptotic distribution of a unit root process under double truncation. Communications in Statistics. Theory and Methods 32, 2059–2072.
Jarque, C., Bera, A., 1980. Efficient tests for normality, homoskedasticity and serial independence of regression residuals. Economics Letters 6, 255–259.
Jarque, C., Bera, A., 1987. A test for normality of observations and regression residuals. International Statistical Review 55, 163–172.
Önder, A., Zaman, A., 2005. Robust tests for normality of errors in regression models. Economics Letters 86, 63–68.
Pearson, E., Hartley, H., 1966. Tables for statisticians, 3rd ed. Biometrika, vol. I. Cambridge University Press, Cambridge, UK.
Pearson, E., Stephens, M., 1964. The ratio of range to standard deviation in the same normal sample. Biometrika 51, 484–487.
Poitras, G., 1992. Testing regression disturbances for normality with stable alternatives: further Monte Carlo evidence. Journal of Statistical Computation and Simulation 41, 109–123.
Thode, H., 2002. Testing for Normality. Marcel Dekker, New York.
Urzúa, C., 1996. On the correct use of omnibus tests for normality. Economics Letters 53, 247–251.