A Test of Normality Using Geary's Skewness and Kurtosis Statistics*

Dong W. Cho
Wichita State University

Kyung So Im†
University of Central Florida

November 19, 2002

Abstract

This paper discusses how the widely used Bera-Jarque test of normality has weak power against short-tailed distributions, and proposes a test that uses lower sample moments studied by R.C. Geary. The proposed test is as simple to compute as the Bera-Jarque test; small sample size is closer to the asymptotic size; and the overall power is compared favorably to the Bera-Jarque test especially for moderate sample sizes.

JEL Classification: C12.

Key Words: Normal Distribution, Skewness, Kurtosis, Size, Power.

*We thank Barton School of Business for summer research grant, and Philip Hersch for the comments.

†Correspondent: Department of Economics, College of Business Administration, P.O.Box 161400, University of Central Florida, Orlando, FL 32816-1400, telephone: 407 823-1132, fax: 407 823-3269, e-mail: kim@bus.ucf.edu

1. Introduction

The most widely used normality test, at least in the profession of economics, would be the test referred to as Bera-Jarque test (BJ hereafter) (Bowman and Shenton, 1975; Shenton and Bowman, 1977; Bera and Jarque, 1982; Jarque and Bera, 1987). BJ test is simple to compute, and the power has proved comparable to other powerful tests such as Shapiro-Wilk test (Shapiro and Wilk, 1965) or Shapiro-Francia test (Shapiro and Francia, 1972) that are computationally more demanding. See Mardia (1980) for an exhaustive account of the various normality tests. Simulation results comparing the power of the BJ tests with other tests were reported by Pearson, D'Agostino and Bowman (1977), Jarque and Bera (1987), and Deb and Sefton (1996). The BJ test was uncovered by Bera and Jarque as the Lagrangian multiplier (LM) test against the Pearson family distributions. Also, Geary (1947) showed that the third the fourth sample moments used in the BJ statistic are most sensitive asymptotically in testing for skewness and kurtosis assuming Gram-Charlier density. Therefore, the BJ test shares the nice asymptotic optimal property against a wide class of non-normal distributions.

The most often discussed problem associated with the BJ test would be the slow convergence of the test statistic to its limiting distribution, which makes the test under-sized even in a reasonably large sample. Another problem, which has not drawn much attention in the literature, is that the BJ test has in general a very weak power against platykurtic distributions. Although we tend to be more concerned with the situation of heavy-tailed population in practice, there will be many situations where we cannot exclude the possibility of short-tailed distribution.

The slow convergence of the BJ statistic to its limiting distribution would be associated with the estimation of the third and fourth moments: The average of the cubed and fourth powered normal variables converges very slowly to a normal distribution. The second problem, the low power property of the BJ test against the platykurtic distributions would be explained as follows: The sample skewness and kurtosis of the BJ statistic are standardized by the variances under normality, which are usually bigger than the actual variances when the distribution is platykurtic. Dividing by a bigger value than its actual variance makes both the non-centrality and the variance of the statistic smaller than those of usual non-central chi-square distribution. Therefore, the statistic could be clustered below the critical value to make the power smaller than the size. However, as the sample size grows and the distribution gets away from the null distribution, we may expect a rapid increase of power since the variance of the statistic is small.

The opposite would be true for leptokurtic distributions; the skewness and kurtosis are divided by smaller values than the actual variance. Therefore, the distribution of the BJ statistic would be centered even farther away from the null distribution but with larger variance. We may expect a good power property especially when the sample size is relatively small.

In this paper we develop a normality test that combines the Geary's mean deviation statistic (1935) and a skewness measure based on the second moments, which also was studied by Geary (1947). Although the proposed test cannot avoid the criticism that it is an *ad hoc*, the test is constructive in the sense that it may avoid the problems associated with the BJ test by using lower moments. Also the Geary's mean deviation test is the most powerful against the double exponential alternative (Uthoff, 1973; Dumonceux, Antle and Hass, 1973). According to the simulation results reported by D'Agostino and Roseman (1974), the Geary's mean deviation test showed a well balanced power against both leptokurtic and platykurtic distributions.

In the following sections we describe details of the proposed test, which we call G test, and conduct a Monte Carlo study to compare the two tests. We conclude in the final section.

2. A Normality Test Using Geary's Skewness and Kurtosis Statistics

Suppose we have a random sample $\{x_1, x_2, ..., x_n\}$, obtained from an independent population. Let the j-th central sample moment be:

$$m_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j\,, \text{ for } j = 2, 3, ..., \tag{2.1}$$

where $\bar{x}$ is the sample mean. Therefore, $\hat{\sigma}^2 = m_2$. The Pearson's skewness and kurtosis statistics are obtained as

$$\sqrt{b_1} = \frac{m_3}{\hat{\sigma}^3},\ b_2 = \frac{m_4}{\hat{\sigma}^4}, \tag{2.2}$$

and we have the BJ statistic:

$$BJ = n\left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24}\right]. \tag{2.3}$$

BJ approaches to the chi-square distribution with two degrees of freedom as n grows when $x's$ are from iid normal population.

Define the moments of the product of the sign and the power of $(x - \mu)$;

$$s\mu_j = E\left[sgn(x-\mu)\cdot(x-\mu)^j\right],\ j = 1, 2, ..., \tag{2.4}$$

where μ is the population mean, and $sgn(x - \mu)$ denotes the sign of $(x - \mu)$. When j is odd, $s\mu_j = E\left|x - \mu\right|^j$. We have, for symmetric distribution of x, $s\mu_j = 0$ when j is even. We propose a test based on the first two moments; $s\mu_1$ and $s\mu_2$.

Define the sample counterparts of $s\mu_j$:

$$sm_j = \frac{1}{n}\sum_{i=1}^{n}\left[sgn(x-\bar{x})\cdot(x-\bar{x})^j\right],\ j=1,2,.... \tag{2.5}$$

If we let

$$a_j = \frac{sm_j}{\hat{\sigma}^j},\ j=1,2,..., \tag{2.6}$$

we obtain a statistic:

$$G = n\left[\frac{\left(a_1-\sqrt{\frac{2}{\pi}}\right)^2}{\left(1-\frac{3}{\pi}\right)} + \frac{a_2^2}{\left(3-\frac{8}{\pi}\right)}\right]. \tag{2.7}$$

As we will show, under normality, this G statistic approaches to the chi-square distribution with two degrees of freedom as the sample size grows.

The two statistics a_1 and a_2 have been studied, among others, by Geary (1935, 1936, 1947) and Gastwirth and Owens (1977). For example, $sm_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$. Therefore, a_1 is the ratio of the mean deviation and standard deviation proposed by Geary (1935) as a statistic for normality test. Under normality a_1 converges strongly to the expected value of the mean deviation of standard normal, which is $\sqrt{\frac{2}{\pi}}$. Therefore, deviation of a_1 from $\sqrt{\frac{2}{\pi}}$ is viewed as an evidence of non-normality. Geary (1935) showed that when x is iid normal;

$$E|x-\mu|^j = \begin{cases} (j-1)(j-3)...3\cdot 1\sigma^j \text{ when } j \text{ is even,} \\ (j-1)(j-3)...4\cdot 2\sqrt{\frac{2}{\pi}}\sigma^j \text{ when } j \text{ is odd.} \end{cases} \tag{2.8}$$

Using this result, it is easy to deduce that

$$Avar\sqrt{n}a_1 = 1-\frac{3}{\pi}. \tag{2.9}$$

Therefore, we have

$$\sqrt{n}\left(a_1-\sqrt{\frac{2}{\pi}}\right)\xrightarrow{d} N\left(0,1-\frac{3}{\pi}\right). \tag{2.10}$$

It is obvious that a_2 converges to zero under normality. Geary (1947) proved that

$$Avar(\sqrt{n}a_2) = 3-\frac{8}{\pi}. \tag{2.11}$$

Therefore, we have

$$\sqrt{n}a_2 \xrightarrow{d} N\left(0,3-\frac{8}{\pi}\right). \tag{2.12}$$

Combining the results of (2.10) and (2.12), and noting that, as was shown by Gastwirth and Owens (1977), $\sqrt{n}a_1$ and $\sqrt{n}a_2$ are independent asymptotically under normality, the G statistic follows the chi-square distribution with two degrees of freedom in the limit.

Geary (1947) studied the statistics of a more general form

$$h(c) = \frac{\frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^c}{\hat{\sigma}^c} \tag{2.13}$$

and

$$g(d) = \frac{1}{n}\left\{ \sum_{x_i - \bar{x} > 0} |x_i - \bar{x}|^d - \sum_{x_i - \bar{x} < 0} |x_i - \bar{x}|^d \right\} / \hat{\sigma}^d. \tag{2.14}$$

$h(c)$ was studied as a potential kurtosis measure, and $g(d)$ was studied as a skewness measure. Note that $h(1) = a_1$, $h(4) = b_2$, $g(3) = \sqrt{b_1}$, and $g\,(2) = a_2$. Geary showed that $h(c)$ is the most sensitive asymptotically at $c = 4$ in testing whether $h(c)$ deviates from the $h(c)$ of the normal distribution assuming the density is:

$$f_1(x) = \frac{1}{\sqrt{2\pi}}\left\{1 + \frac{\lambda_4}{24}\left(x^4 - 6x^2 + 3\right)\right\} e^{-\frac{x^2}{2}}. \tag{2.15}$$

Also Geary showed that $g(d)$ is the most sensitive at $d = 3$ assuming the density:

$$f_2(x) = \frac{1}{\sqrt{2\pi}}\left\{1 + \frac{\lambda_3}{6}\left(x^3 - 3x\right)\right\} e^{-\frac{x^2}{2}}. \tag{2.16}$$

However, as Pearson (1935) noted, a_1 approaches to a normal distribution faster than b_2 and has a reasonable power against both platykurtic and leptokurtic alternatives. See also D'Agostino and Roseman (1974) for a simulation result comparing the power of a_1 and b_2. Uthoff (1973) and Dumonceaux, Antle and Hass (1973) proved that a_1 is the most powerful when the alternative distribution is double exponential. See Hogg (1972) for an another justification of using a_1 and b_2 as a normality test depending on the alternative distributions. Also, Geary (1947) compared the power of a_1 and b_2 against the density $f_1(x)$ with $\lambda_4 = 0.5$ in (2.15). He demonstrated an emphatic superiority of b_2 when $n = 500$, but he also discovered that b_2 is no longer superior when $n = 100$. This result suggests that a_1 may be preferred especially when the sample size is moderate to small even in the situation where b_2 is asymptotically optimal.

To the best of our knowledge, a_2 has never been used in practice, and we do not know of any study about the power of a_2 other than Geary (1947), where he showed that $\sqrt{b_1}$ is more sensitive asymptotically assuming the density is $f_2(x)$ in (2.16). But, Geary also showed that the difference in the asymptotic sensitivity of a_2 and $\sqrt{b_1}$ is quite small, which led us to conjecture that a_2 would serve better in detecting skewness when the sample size is relatively small.

3. Simulation

First we compare how quickly the small sample distributions of the BJ and G statistics approach to their limiting distribution. The 1%, 5% and 10% critical values of the several small sample distributions of the two statistics are reported in Table 1. All the figures are computed based on 600,000 sampling from the GAUSS random number generator that uses the standard rejection method.[1]

Table 1
Significant Points of Statistics

N	BJ			G		
	1%	5%	10%	1%	5%	10%
20	9.762	3.821	2.359	8.786	5.269	4.071
50	12.578	5.007	3.203	9.162	5.696	4.385
100	12.626	5.442	3.673	9.172	5.834	4.487
200	11.858	5.694	4.037	9.188	5.920	4.537
∞	9.210	5.991	4.605	9.210	5.991	4.605

A faster convergence of the G statistic is obvious. Another interesting point is that the 1% significance points of the BJ statistic are larger than their limiting value. Therefore, it is expected that the BJ test will be over-sized at $\alpha = 0.01$, which contrasts to the well known under-size property of the BJ test at $\alpha = 0.05$ and at $\alpha = 0.1$. On the other hand, the G test is expected to be under-sized slightly at $\alpha = 0.01$.

Table 2 reports the rejection ratio of the BJ and G tests for the 15 distributions at $\alpha = 0.05$. All the figures are computed based on 10,000 replications. The numbers in parentheses are the size-adjusted rejection ratio using the critical values reported in Table 1. Also, for example, .2N(0,1)+.8N(0,9) denotes the mixture of two independent normal distributions with variances one and nine. Here, 0.2 and 0.8 are the proportions.

As is seen in Table 2, both the BJ and G tests tend to under-size at $\alpha = 0.05$, but the problem is much less serious for the G test. A similar under-size tendency was observed when the test were conducted at $\alpha = 0.1$ (not reported), whereas, as was predicted from Table 1, an opposite over-size tendency was observed for the BJ test when the tests were conducted at $\alpha = 0.01$.

Several notable patterns emerge on the power property of the tests. First and most notably, the BJ test performs miserably against the uniform, Beta(2,2) and Beta(3,2) distributions for relatively small n. For example, the size-adjusted power of the BJ tests are nearly zero when $n = 50$ against all three distributions. The G test suffers a similar problem, but the problem is not as extreme.

[1] We tried to replicate the significance points reported in Deb and Sefton (1996) for the BJ test, but observed small differences in the second decimal points. Deb and Sefton did not report the 1% significance points.

However, once the BJ test begins to pick up the power, the power rises quickly. In uniform case, the power of the BJ test exceeds the power the of the G test when $n = 200$, where the power of both tests is close to one. In the case of Beta(2,2), the BJ test has little power even when $n = 100$, but becomes quite comparable to the power of the G test when $n = 200$. From these two results, we projected that the power of the BJ test, in case of Beta(2,2), would be better when the sample size is larger than 200. We conducted an additional simulation to see what happens when n is bigger than 200. The result, computed for $n = 50$ (5) 550 through 10,000 replications for each sample size, is plotted in Figure 1. The power of the BJ test begins to exceed the power of the G test at n about 220 where the power is about 0.74, and remains more powerful for larger n.

The simulation results of the uniform and Beta(2,2) distributions made us to expect that the power curves of the BJ and G tests would cross at n larger than 220 and the power lower than 0.74 when the sample is taken from the Beta(3,3) distribution. We found (not reported) that the power of BJ catches up the power of the G test at $n = 315$, where the power is 0.56. But, we note that the power of the BJ test begins to exceed the 5% nominal size only at $n = 160$.

Much of the reasoning behind these results was explained in the beginning section. For example, the variance of $\sqrt{b_1}$ of the normal sample is about three times bigger than the variance of $\sqrt{b_1}$ of the Beta(2,2) sample. Note that the non-centrality of $\sqrt{b_1}$ is zero in this case in this case because Beta(2,2) is symmetric. Dividing $n\sqrt{b_1^2}$ by 6 will make the variance of the BJ statistic much smaller than that of the chi-square distribution. Consequently, the statistic $n\sqrt{b_1^2}/6$ will be clustered at one, the mean of the chi-square distribution with one degree of freedom, regardless of the sample size.

The variance of b_2 of the Beta(2,2) sample is about one ninth of the variance of b_2 of the normal sample, and the population kurtosis, β_2, is 2.14. Therefore, we would expect that the statistic $nb_2^2/24$ tends to remain small unless n is so big that the non-centrality dominates the denominator. But, once the test begins to pick up the power, the power would rise rapidly since the statistic $nb_2^2/24$ is clustered together (the variance is about 0.025).

We are not the first to compute the power of the BJ test against the short-tailed distributions. Jarque and Bera (1987) reported that the size-adjusted power of the BJ test against Beta(3,2) distribution at $\alpha = 0.1$ are 0.116 and 0.412 when $n = 20$ and $n = 50$, respectively. This result, however, was based only on 250 replications. Deb and Sefton (1996), based on 100,000 replications, estimated the power of the BJ test, for the same cases, as 0.0640 and 0.1298. We obtained 0.043 and 0.103 for the corresponding case (not reported), which closely match with the results reported by Deb and Sefton (1996). The beta distribution belongs to the Pearson family, so that the BJ test is asymptotically optimal. However, the small sample problem seems too serious.

Second, the G test clearly dominates the BJ test when the alternative distribution is double exponential, which is well expected from the theory advanced by Uthoff (1973) and Dumonceaux, Antle and Hass (1973) that the Geary's test based on a_1 statistic is the most powerful in this case. Note that when $n = 20$, the size-adjusted powers of the two tests are almost identical, but the G test tends to dominate the BJ test as n grows.

Third, for the case of the t-distributions, the G test remains as powerful as the BJ test when the degrees of freedom is 3, but the BJ test dominates when the degrees of freedom is 6 or 10. Following our additional simulation (not reported), the G test is more powerful when the degrees of freedom is 1 and 2, but less powerful when the degrees of freedom is 5 or higher. We do not have an answer why the G test is better or comparable to the BJ test when the degrees of freedoms are 1-3, but becomes eventually dominated by the BJ test as the degrees of freedom rises. However, the overall superiority of the BJ test is hardly surprising since the t-distribution belongs to the Pearson family, and the BJ test is optimal asymptotically. Note that the G test retains a reasonable power throughout.

Fourth, for chi-square alternatives, the G test is better in all of the three cases we report. We extended our simulation for degrees of freedoms 1-15 to observe a similar pattern throughout. This result is a little surprising since the chi-square belongs to the Pearson family so that the BJ test shares the asymptotic optimal property. To clarify the source of power, we conducted an additional simulation to compare the four statistics, $\sqrt{b_1}$, a_2, b_2, a_1, separately. It turns out that a_2 clearly dominates $\sqrt{b_1}$, while a_1 is dominated by b_2. Therefore, it is a_2 that makes the G test more powerful in this case.

Fifth, for the lognormal distribution, the power of both tests quickly approach to one, but we still see that the G test is more powerful.

Sixth and last, we simulate the mixture of normal distributions as a neutral ground since the mixture normal is not a member of the Pearson family. The G test shows a stronger performance in our simulation. But, it seems too hasty to draw a general conclusion from the reported results.

We also compared the performance of the BJ and G tests using regression residuals. We replicated the four different types of regressors considered in Deb and Sefton (1996). We simulated the same 15 error distributions and sample cases as we have in Table 2 for each type of regressors. We did not find any remarkable difference from the results reported in Table 2. These results are not reported, but available upon request.

4. Conclusion

We proposed a new normality test using lower moments studied earlier by R.C. Geary (1935, 1936, 1947). The proposed test has a better size property compared to the widely used Bera-Jarque test. The power comparison depends on the shape of the distribution under test. However, the BJ test suffers an extreme power deficit when the population distribution has short-tails. The G test has more stable power against various directions of the deviation from normality. All in all, it seems that the G test is a good alternative to the BJ test especially when the sample size is moderate to small, unless we have a strong *a priori* belief that the population distribution of the sample we put to a normality test cannot be platykurtic.

We also considered a test that uses the proportion of the signs of $(x - \bar{x})$ as a skewness measure, which, however, turned out not nearly competitive with a_2 or $\sqrt{b_1}$.

References

[1] Bera, A.K. and C.M. Jarque, 1982, "Model specification tests: A simultaneous approach," *Journal of Econometrics*, 20, 59-82.

[2] Bowman, K.O. and L.R. Shenton, 1975, "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2," *Biometrika*, 1975, 62, 243-250.

[3] D'Agostino and B. Roseman, 1974, "The power of Geary's test of normality," 61, *Biometrika*, 181-184.

[4] Deb, P. and M. Sefton, 1996, "The distribution of Lagrangian multiplier test of normality," *Economics Letters*, 51, 123-130.

[5] Dumonceaux, R., C. Antle, and G. Haas, 1973, "Likelihood ratio test for discrimination between two models with unknown location and scale parameter," *Technometrics*, 15, 19-27.

[6] Gastwirth, J.L. and M.G.B. Owens, 1977, "On classical test of normality," *Biometrika*, 64, 135-139.

[7] Geary, R.C., 1935, "The ratio of the mean deviation to standard deviation as a test of normality," *Biometrika,* 27, 310-332.

[8] Geary, R.C., 1936, "Moments of the ratio of mean deviation to the standard deviation as a test of normality," *Biometrika,* 28, 295-305.

[9] Geary, R.C., 1947, "Testing for normality," *Biometrika,* 34, 209-242.

[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] Hogg, R.V., 1972, "More lights on the kurtosis and related statistics," *Journal of the American Statistical Association, 67, 422-424.*

[12] Mardia, K.V., 1980, "Test of univariate and multivariate normality," Handbook of Statistics, P. R. Krishnaiah, edition, 1, 279-320.

[13] Pearson, E.S., 1935, "A comparison of β_2 and Mr. Geary's w_n criteria," *Biometrika*, 27, 333-347.

[14] Shapiro, S.S. and R.S. Francia, 1972, "An approximation analysis of variance test for normality, *Journal of the American Statistical Association,* 67, 215-216.

[15] Shapiro, S.S. and M.B. Wilk, 1965, "An analysis of variance test for normality (complete samples)," *Biometrika*, 52, 591-611.

[16] Shenton, L.R. and K.O. Bowman, 1977, "A bivariate model for the distribution of $\sqrt{b_1}$ and b_2," *Journal of the American Statistical Association*, 72, 206-211.

[17] Uthoff, V.A. 1973, "The most powerful scale and location invariant test of the normal versus double exponential," *Annals of Statistics* 1, 170-174.