J. King Saud Univ., Vol. 20, Admin. Sci. (1), pp. 18-25, Riyadh (2008/1429H.)

A Modification of the Jarque-Bera Test for Normality

Moawad El-Fallah Abd El-Salam
Dept. of Psychology, College of Social Sciences,
Imam Muhammad bin Saud Islamic University, Riyadh, Saudi Arabia

(Received 30/12/1426H.; accepted for publication on 8/4/1428H.)

Abstract. Many statistical tests have been proposed to find out whether a sample is drawn from a normal distribution or not. These statistical tests are typically constructed using OLS residuals. However, since diagnostic tests for normality are very sensitive to outliers, they have a zero breakdown value. In this paper, we attempt to investigate the effects of using residuals from robust regression replacing OLS residuals in test statistics for normality. We study a modified Jarque-Bera test statistic for normality based on robust regression residuals. The asymptotic distribution of this robustified normality test is derived, and its breakdown property is discussed.

Keywords: Normality; Jarque-Bera test; Robust regression; Least Trimmed Squares; Breakdown point.

1. Introduction

Significant deviations from normality of the regression residuals can substantially affect the performance of usual inference techniques. Thus, diagnostic tests for normality are important for validating inferences made from regression models. Several such tests have been proposed [1-4]. In general, these tests are conducted by testing the ordinary least squares (OLS) residuals for normality. However, since OLS estimates themselves are obtained by maximizing the likelihood function that the residuals come from some normal distributions, OLS residuals tend to look as normal as they can possibly be. Thus, these tests fail to detect any lack of normality in the OLS analysis.

Consider we have independently sampled n observations $X_n=(x_1, x_2,\ldots, x_n)$ with mean μ and standard deviation σ. The coefficients of skewness and kurtosis are defined as:

$$b_1 = \frac{\mu_3}{\sigma^3} = \frac{E\,[(x-\mu)^3]}{[E(x-\mu)^2]^{3/2}} \qquad (1)$$

$$b_2 = \frac{\mu_4}{\sigma^4} = \frac{E\,[(x-\mu)^4]}{[E(x-\mu)^2]^2} \tag{2}$$

where $\mu_r = E[(x-\mu)^r]$ be the $r\underline{th}$ centeral moment of X_n with $\mu_2=\sigma^2$. Kendall and Stuart [5] pointed out that if X_n is i.i.d. and normally distributed, then:

$$\sqrt{n}b_1 \xrightarrow{d} N(0,6) \text{ and } \sqrt{n}(b_2-3) \xrightarrow{d} N(0,24) \tag{3}$$

Regarding univariate normality tests amongst the several alternatives, we have selected Jarque and Bera (JB) as being one of the popular tests for testing normality in regression applications. It is well known that the JB test is based on a weighted average of the skewness and kurtosis as:

$$JB = n\left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2-3)^2}{24}\right] \tag{4}$$

where n is the sample size, $\sqrt{b_1}$ is the sample skewness and b_2 is the sample kurtosis. Jarque and Bera [4] proved that if the alternatives to the normal distribution belong to the Pearson family, JB is a score test for normality. However, because the classical skewness and kurtosis coefficients $\sqrt{b_1}$ and b_2 are based on OLS residuals, they are very sensitive to outliers in the data. Therefore, the main idea of this paper is to use the same JB test using residuals from a robust regression estimator instead of OLS residuals, and denote this by JB*. The particular estimator we study is called least trimmed squares (LTS), introduced by Rousseeuw [6].

The paper is organized as follows. Section 2, first, introduces the basic idea of the least trimmed squares robust procedure. In Section 3, we derive the asymptotic distribution of the robustified normality test, and the high breakdown property of the test statistic is discussed. Section 4 includes simulation results, while Section 5 concludes the paper.

2. The Least Trimmed Squares Method

The least trimmed squares (LTS) method has the property of being highly resistant to a relatively large proportion of outliers. Thus, LTS has a high breakdown value. For the details of this technique and its properties, see Rousseeuw and Leroy [7], and also Chapter 5 of Zaman [8].

Consider the following regression model:

$$y_i = x_i\beta + u_i\,, \qquad i = 1,2,\ldots, n, \tag{5}$$

where x_i is a (1xk) vector of regressors, β is a (kx1) vector of regression coefficients, y_i are observed responses, and u_i represent the error terms which are assumed to be independent and identically distributed with mean 0 and standard deviation σ.

The LTS estimate is given by:

$$\text{Min} \sum_{i=1}^{h} e_{(i)}^2, \tag{6}$$

where $e^2_{(1)} \leq e^2_{(2)} \leq, \ldots, \leq e^2_{(n)}$ are the ordered squared residuals $e^2_i = (y_i - x_i \beta)^2$, i=1,…,n, and the value of h must be determined based on trimming the data values. For example, the 20% trimmed LTS estimator is defined to be the value of β minimizing $\sum_{i=1}^{(4n/5)} e_{(1)}^2(\beta)$. Chen [9] defined h in the range $[(n/2) + 1] \leq h \leq [(3n+k+1)/4]$, and recommended that the breakdown value for the LTS estimator is n-h/n, when h=[(3n+k+1)/4].

Although there exist other high breakdown estimators, the LTS has many advantages to recommend itself. A recently developed algorithm has permitted fast computation of the LTS (see Rousseeuw and Van Driessen [10]). The main advantages of the LTS method are as follows. First, LTS are simple to understand and easy to motivate. Second, it is more efficient than the LMS (least median of squares) introduced by Rousseeuw [6] with which it shares these advantages. Many estimators commonly regarded as 'robust' in the econometrics literature have low breakdown values and cannot deal with any significant number of outliers. For example, the bounded influence estimator of Krasker and Welsch [11], and the least absolute deviation method, both suffer heavily from the presence of a small subgroup of outliers (see Yohai [12]).

3. Asymptotic and Robustness Properties

In this section, we discuss the asymptotic distribution and the breakdown property of the robustified normality test statistic.

Theorem 1: Under the null hypothesis of normality of the error terms, JB* is distributed asymptotically chi-squared with 2 degrees of freedom.

Proof: As can be seen from Hossjer [13] that:

$$\sqrt{n}(\hat{\beta} - \beta) = O_P(1), \tag{7}$$

Then, using Eq. (7) and Davidson and Mackinnon [14, pp. 568-570], we obtain:

$$(6n)^{-1/2} \sum_{i=1}^{n} \frac{u_i^3}{\sigma^3} - (6n)^{-1/2} \sum_{i=1}^{n} \frac{e_i^3}{\hat{\sigma}^3} \xrightarrow[p]{} 0, \tag{8}$$

$$(6n)^{-1/2}\sum_{i=1}^{n}\frac{u_i^3}{\sigma^3}\xrightarrow{d}N(0,1)\text{'} \tag{9}$$

where e_i are the residuals from LTS regression and $\hat{\sigma}$ is the LTS estimate of the standard deviation. So, we have:

$$(6n)^{-1/2}\sum_{i=1}^{n}\frac{e_i^3}{\hat{\sigma}^3}\xrightarrow{d}N(0,1) \tag{10}$$

Similarly for

$$(24)^{-1/2}\left(\sum_{i=1}^{n}\frac{e_i^4}{\hat{\sigma}^4}-3\right)\xrightarrow{d}N(0,1) \tag{11}$$

Then, combining Eqs. (10) and (11) in the test statistic, we have:

$$JB^* \xrightarrow{d} X_{(2)}^2 \tag{12}$$

The breakdown point of a statistical function is the smallest fraction of data contamination that can produce arbitrarily extreme results. So, a power breakdown point is the least amount of data contamination that can drive the test statistic to its null value regardless of the true alternative value (see He et al. [15] and Markatou and He [9] for a detailed discussion).

On the other hand, the weighted likelihood estimator [17] may reduce to the LTS estimator with the following weights:

$$w_i = \begin{bmatrix} 1 & \text{if } i \le h \\ 0 & \text{if } i > h \end{bmatrix}, \tag{13}$$

where h is the trimming point, and i=1, …, n are ordered. Following Agostinelli and Markatou [18], we define the score type test function as:

$$JB^* = S_w^{/}(\theta)\left[\sum_{i=1}^{n} w_i I(\theta)\right]^{-1} S_w(\theta)\text{'} \tag{14}$$

where θ is the vector of parameters, the superscript $'$ denotes transpose and I(θ) is the Fisher information matrix evaluated at θ ,$S_w(\theta) = \sum_{i=1}^{n} w_i u(y_i, \theta)$ is the score function evaluated at θ, and $u(y_i, \theta) = \frac{\partial}{\partial\theta} L(y_i, \theta)$, where $L(y_i, \theta)$ is the log-likelihood function.

Following the same technique as Agostinelli and Markatou [18, p. 507], but making minor modifications required to allow for the fact that we have a Pearson family alternative, it may be possible to establish that the power breakdown point of the JB* test is the same as the breakdown point of LTS estimator, if the alternative belongs to the Pearson family.

4. Numerical Results

Rousseeuw and Leroy [7] used several data sets for their analysis related to the robust regression. As a preliminary test for the validity of our basic idea, we conducted a study of tests for normality using the same five data sets. The results, which are presented in Table 1, are highly encouraging. In each of the five data sets studied, we can say that the JB test fails to detect any lack of normality in the OLS residuals, while JB* picks up the lack of normality of the errors very clearly.

Table 1. The values of the normality test statistics from Rousseeuw and Leroy data sets

Series	n	JB	JB*
Brain	28	1.93	25.52
Cloud	19	2.23	7.14
Salinity	28	0.03	165.32
Aircraft	23	0.17	57.88
Delivery	25	0.01	54.22

Rousseeuw and Leroy [7] from pages 57, 96, 82, 154, 155.

In addition, since the presence of many outliers (corresponding to lack of normality) in these series is well known, we observe that the JB test, conducted to the basis of OLS residuals, can lead to misleading outcomes. On the other hand, in order to understand the reasons for the superiority of JB* test on the Rousseeuw and Leroy data sets, we carry out a Monte Carlo simulation to obtain the power comparisons of JB and JB* using the GAUSS program.

To generate random variates from normal and other distributions considered throughout the study, we used the random number generator of GAUSS 3.1.1. We considered a linear model with a constant term of one and one additional regressor. Every experiment in this simulation allows some alternative distributions than just i.i.d. with some non-normal density. To accomplish this, we started by using the mixture of two normal distributions ($\alpha N(0, 1) + (1 - \alpha) N(\mu, \sigma^2)$) as an alternative to the null of normality. In this way, α allowed us to choose the percentage of outliers and μ and σ^2 allowed us to control the behavior of the outliers.

In the examples related to the failure of OLS series, we observed that the outliers lie in clusters (see, for example, Rousseeuw and Leroy [7, p. 58]). Therefore, we study the effects of clustered outliers (referred to as case A below) as opposed to randomly generated outliers (referred to as case B below). The nature of the outliers also turns out to be crucial. High variance outliers from N (0, 9) and N (0, 16) compared to the base distribution N (0, 1) were studied, as well as mean-shift outliers generated from N (5, 1) and N (10, 1) distributions.

In order to make a comparison, we also generated outliers at random places in samples. The first case, where the outliers are clustered, is labeled case A, and the second case, where they are at random places, is labeled case B. The results for 20 and 50 observations are presented in Table 2. In all cases, 80% of the true errors are generated from N (0, 1) and 20% of them from normal distributions with different means and variances. So, 4 outliers are generated for the case of 20 observations and 10 outliers for the case of 50 observations.

We compute the values of JB and JB* and we find out whether null hypothesis (normality) is rejected by each individual test. The level of significance used was 0.10 throughout the simulations. We carried out 1000 replications and estimated the power of each test by computing the percentage when the null hypothesis was rejected. The results for 20 and 50 observations are presented in Table 2.

Table 2. Power comparisons of mixture of normal alternatives with 20% outliers under different conditions

N = 20		**Case A**	**Case B**
N (5 , 1)	JB	0.33	0.69
	JB*	0.45	0.71
N (10, 1)	JB	0.41	0.91
	JB*	0.67	0.97
N (0 , 9)	JB	0.55	0.54
	JB*	0.55	0.55
N (0,16)	JB	0.73	0.71
	JB*	0.73	0.71
N = 50		***Case A***	***Case B***
N (5 , 1)	JB	0.60	0.99
	JB*	0.88	1.00
N (10, 1)	JB	0.68	1.00
	JB*	0.75	1.00
N (0 , 9)	JB	0.84	0.83
	JB*	0.84	0.83
N (0,16)	JB	0.96	0.95
	JB*	0.96	0.95

The results of Table 2 show that when the outliers are clustered (case A) and generated by shifting the mean of the normal, the robust version of the test has substantially greater power than the test based on OLS residuals. In all other cases, both tests have similar performance. This can be explained as follows. With balanced outliers on both sides of the règression line, the OLS estimator is not much affected by the outliers, and hence OLS residuals are similar to robust residuals. In addition, outliers generated by high variance are likely equally affected and did not lead to improved performance for JB* as we had expected, while outliers generated by a mean shift all lie on one side of the regression line and hence are much more effective in systematically distorting OLS estimates.

Also, clustering of the outliers has a significant effect on the relative performance of these tests. Clustered outliers tend to systematically distort OLS estimators to the possible largest extent (outliers with high leverage have a similar effect and result in similar outcomes). Thus, these situations lead to maximum improvement for tests based on robust residuals. When the outliers are not clustered, their impact on the OLS estimators is reduced (equivalently, they have less leverage). In such situations, both OLS and robust residuals pick out the outliers easily, and therefore both achieve high power, as indicated in Table 2.

5. Conclusions

In this paper, we discussed the Jarque-Bera test of normality, which is not able to detect normality in the presence of outliers. Therefore, we proposed to use residuals from a robust regression technique (LTS) instead of OLS residuals as a basis for the normality test. We found that the idea was valid by using some examples from the literature, in which the robustified Jarque-Bera test (JB*) reveals lack of normality not detected by the test based on OLS residuals. We obtained the asymptotic distribution of the modified test as chi-squared with 2 degrees of freedom under the null hypothesis of normality, and the breakdown point of the test statistic was investigated. From the results of simulation, we found that the robust test yields substantially greater power for systematic and clustered outliers situations.

References

[1] Pearson, E.S.; D'Agostina, R.B. and Bowman, K.O. "Tests for Departure from Normality: Comparison of Powers." *Biometrika,* 64 (1977), 231-246.

[2] White, H. and Macdonald, G.M. "Some Large Sample Tests for Nonnormality in the Linear Regression Model." *J. AM. Statist. Assoc.*, 75 (1980), 16-28.

[3] Pierce, D.A. and Gray, R.J. "Testing Normality of Errors in Regression Models." *Biometrika*, 69 (1982), 233-236.

[4] Jarque, C.M. and Bera, A.K. "A Test for Normality of Observations and Regression Residuals." *Int. Statistical Review*, 55 (1987), 163–172.

[5] Kendall, M. and Stuart, A. *The Advanced Theory of Statistics*. McGraw Hill, 1969.

[6] Rousseeuw, P.J. "Least Median of Squares Regression." *J. AM. Statist. Assoc.*, 79 (1984), 871-880.

[7] Rousseeuw, P.J. and Leroy, A.M. *Robust Regression and Outlier Detection.* New York: Wiley, 1987.

[8] Zaman, A. *Statistical Foundations for Econometric Techniques.* New York: Academic Press, 1996.

[9] Chen, C. "Robust Regression and Outlier Detection with the ROBUSTRER Procedure." SUGI paper, *SAS Institute,* (2002), 265-277.

[10] Rousseeuw, P.J. and Van Driessen, K. "A Fast Algorithm for the Minimum Covariance Determinant Estimator." *Technometrics*, 41 (1999), 212-223.

[11] Krasker, W.S. and Welsch, R.E. "Efficient Bounded Influence Regression Estimation." *J. AM. Statist. Assoc.*, 77 (1982), 595-604.

[12] Yohai, V.J. "High Breakdown-point and High Efficiency Robust Estimates for Regression." *Ann. Statist.,* 15 (1987), 642-656.

[13] Hossjer, O. "Rank-based Estimates in the Linear Model with High Breakdown Point." *J. AM. Statist. Assoc.*, 89 (1994), 149-158.

[14] Davison, R. and Mackinnon, J. *Estimation and Inference in Econometrics.* New York: Oxford Univ. Press, 1993.

[15] He, X.; Simpson, D.G. and Portnoy, S.L. "Breakdown Robustness of Tests." *J. AM. Statist. Assoc.*, 85 (1990), 446-452.

[16] Markatou, M. and He, X. "Bounded Influence and High Breakdown Point Testing Procedures in Linear Models." *J. AM. Statist. Assoc.*, 89 (1994), 543-549.

[17] Markatou, M.; Basu, A. and Lindsay, B.G. "Weighted Likelihood Estimating Equations with a Bootstrap Root Search." *J. AM. Statist. Assoc.*, 93, (1998), 740-750.

[18] Agostinelli, C. and Markatou, M. "Tests of Hypotheses Based on the Weighted Likelihood Methodology." *Statistica Sinica*, 11 (2001), 499-514.

تعديل اختبار جارك بيرا للكشف عن اعتدالية توزيع البواقي

معوض بن الفلاح عبد السلام

أستاذ مساعد إحصاء، قسم علم النفس، كلية العلوم الاجتماعية،

جامعة الإمام محمد بن سعود الإسلامية، الرياض، المملكة العربية السعودية

(قدم للنشر في ١٤٢٦/١٢/٣٠هـ، وقبل للنشر في ١٤٢٨/٤/٨هـ)

ملخص البحث. يعد اختبار جارك بيرا أحد الاختبارات الهامة التي تستخدم للكشف عن مدى توافر شرط اعتدالية توزيع البواقي لنموذج الانحدار الخطي. وحيث إن الاختبار يعتمد على البواقي المقدَّرة باستخدام طريقة المربعات الصغرى، لذا فإن الاختبار كأداة فحص للكشف عن مدى اعتدالية توزيع الأخطاء يتأثر كثيرًا لوجود القيم المتطرفة في البيانات، ومن هنا كان التفكير في إجراء تعديل على اختبار جارك بيرا ليعتمد على إحدى طرق الانحدارالمقاومة (robust) لتقليل أثر وجود القيم المتطرفة بدلاً من الاعتماد على طريقة المربعات الصغرى. حيث تم استخدام طريقة المربعات الصغرى المرتبة (The Least Trimmed Squares Method).

ولقد تبين من نتائج استخدام بعض البيانات الحقيقية وبيانات المحاكاة أن التعديل المقترح للاختبار يتميز ببعض الخصائص الإحصائية الجيدة في حالة وجود قيم متطرفة في البيانات.