Normality Testing- A New Direction

Islam, Tanweer ul

International Islamic University, Islamabad, Pakistan

2008

Normality Testing- A new Direction

Tanweer-ul-Islam

IIIE, International Islamic University, Islamabad, Pakistan

Email: tanweer.ul.islam@gmail.com

Abstract

This paper is concerned with the evaluation of the performance of the normality tests to ensure the validity of the t-statistics used for assessing significance of regressors in a regression model. For this purpose, we have explored 40 distributions to find the most damaging distribution on the t-statistic. Power comparisons are conducted to find the best performing test against these distributions. It is found that Anderson-Darling statistic is the best option among the five normality tests, Jarque-Bera, Shapiro-Francia, D'Agostino & Pearson, Anderson-Darling & Lilliefors.

Key words: Normality test, power of the test, t-statistic,

JEL Classification: C01, C12, C15

1. Introduction

The normality of error terms is a basic assumption of the linear regression model. Most of the inferential procedures currently used are based on this assumption (Bartolucci & Scaccia, 2005). Zaman *et al.* (2001) give several examples of published regression results where testing reveals lack of normality of errors, and this results changes the findings of these papers. Thus, diagnostic tests for normality are important for validating inferences made from regression models (Onder & Zaman, 2003). Several such tests have been devised (see, for example Geary, 1947; Hogg, 1972; D'Agostino & Pearson, 1973; Pearson *et al*., 1977; Jarque and Bera, 1987; Urzua, 1996; Cho & Im, 2002, Bonett & Seier, 2002; Bry *et. al.*, 2004; Onder and Zaman, 2005, Gel *et. al.*, 2007). Availability of such a large number of normality tests has generated a large number of simulation studies to find a best performing test (see, for example Shapiro *et al*., (1968); Pearson *et al*., (1977); Thadewald *et al*., (2004) and Yazici & Yolacan (2007). However, normality tests are based on different characteristics of the normal distribution and the power of these tests differs depending on the nature of non-normality (Seier, 2002). For any two good tests, we can find alternatives to normality such that either one outperforms the other. See, for example, Shapiro *et al.*, 1968 Thadewald & Büning, 2004, Yazici & Yolacan, 2007. This leads to a dilemma: how can we choose a best test in practical situations?

We propose to solve this dilemma by focusing on the purpose of testing. In regression model, one important goal of testing normality is to make sure that our t-statistic is giving us the right message (i.e. whether the independent variable is a significant explanatory variable or not?). Similarly there are many other goals such as forecast encompassing, general validity of confidence intervals, inference, etc. By focusing on a goal one may be able to find a best test for that goal. We evaluated different tests and alternative distributions used in major simulation studies with respect to how well they "protect' the t-statistic. In contrast to the simulation studies which lead to inconclusive results, we find that the Anderson-Darling test is the unique best test, over the entire range of alternatives and tests studied,

2. Distributions which Damage the t-statistic

To protect t-statistic in the best way, we should know how much a distribution can damage our t-statistic. We used the asymptotic expansion of T by Yanagihara, (2003) to calculate how much a distribution can damage the t-statistic. So, based on the probability formula:

$$P(T \le x) = G_h(x) - \frac{2x}{nh} g_h(x)\left\{b_1 + b_2 + b_3 + \frac{(b_2 + b_3)x}{h+2} + \frac{b_3 x^2}{(h+2)(h+4)}\right\} + o(n^{-1})$$

where n is number of observations, h is number of restrictions, $G_h(x)$ is the distribution function and $g_h(x)$ is the density function of a central chi-squared distribution with h degrees of freedom and the coefficients b_j are given in Yanagihara (2003, p.234).

By using this asymptotic expansion formula, we calculated the following deviations:

$$\text{DEVIATION} = P(\,T \le x \mid \varepsilon_t \text{ i.i.d Normal}\,) - P(\,T \le x \mid \varepsilon_t \text{ i.i.d } K\,)$$

where, K is any i.i.d non-normal distribution. K is a less damaging distribution if the deviation is small, and K is a more damaging distribution if the deviation is large. If the errors are exactly normal, deviation will be zero.

Table: 1 Deviation from normal probabilities

Distributions	n=30		n=50		n=100	
	Probability	Deviations	Probability	Deviations	Probability	Deviations
Normal(0,1)	0.9478	------	0.9489	-------	0.9497	------
$Chi^2(2)$	0.9400	0.0078	0.0126	0.0050	0.9472	0.0025
Gamma(0.05,1)	0.7820	0.1658	0.8505	0.0984	0.9040	0.0457
Gamma(0.1,1)	0.8721	0.0757	0.8977	0.0512	0.9290	0.0207
Beta(2,0.05)	0.9035	0.0443	0.9212	0.0277	0.9370	0.0127
Beta(5,0.05)	0.8643	0.0835	0.8963	0.0526	0.9237	0.0260
Logn(1,1.1)	0.7989	0.1489	0.8541	0.0948	0.9051	0.0446
Logn(1,1.3)	0.5728	0.3750	0.7100	0.2389	0.8374	0.1123
Exp(2)	0.9396	0.0082	0.9437	0.0052	0.9473	0.0024
Weibull(0.5,0.5)	0.8373	0.1105	0.8787	0.0702	0.9168	0.0329
NCt(5,5)	0.9221	0.0257	0.9364	0.0125	0.9386	0.0092

Note: Ten out of forty distributions are listed. Rests of the thirty distributions have not shown significant deviations.

In this study, 40 distributions have been analyzed which cover the majority of the distributions used in the major power studies done so far in the literature. Among these, the most damaging ones appear to be the lognormal distributions, as shown in Table 1. The tests we have chosen are the most representative of their respective class of tests.

Test	**Class of Test**
Anderson-Darling (A^2) & Lilliefors (L)	ECDF
Jarque-Bera (JB) & D'Agostino & Pearson (K^2)	Moment
Shapiro-Francia (SF)	Correlation/Regression

We set $X_{i1} = 1(i = 1,2,\ldots,N)$ and generated X_2& $X_3$from a standard normal distribution. The regressors were fixed throughout the study. Note that the specific values of the means and variances of these regressors have no effect on the simulation results. This invariance property follows from the fact that, for a linear model with regressor matrix X the ordinary least-squares residuals are the same as those of a linear model with regressor matrix XR, where R is any $k \times k$ nonsingular matrix of constants (Weisberg, 1980, p.20)[*]. Our study shows that our concept is valid. We are able to pick out a unique best test from among the numerous alternatives, by finding the one which works best for the 'least favorable' or most damaging distribution.

[*] *See* Jarque & Bera, 1987

3. Simulation Study

In the first part of the simulation study, we have calculated the finite sample critical values for all five tests in our study for sample size n=30, 50 & 100 and for nominal level α = 0.01, 0.05 & 0.1 by using 100, 000 Monte Carlo replications.

In second part, we have performed the normality tests on the most damaging distributions; Lognormal (1, 1.3) & Weibull (0.5, 0.5). Power calculations are based on 10,000 Monte Carlo replications. Table 2 summarizes the empirical powers of the tests for sample size n = 30, 50 & 100 at α = 0.01, 0.05 & 0.1.

Table: 2 Power results against the most damaging distributions

Distribution	Test	N	α =0.01	α =0.05	α =0.1
Logn(1,1.3)	A^2	30	0.9931	0.9988	0.9996
		50	0.9999	1	1
		100	1	1	1
	SF	30	0.9926	0.9989	0.9993
		50	1	1	1
		100	1	1	1
	K^2	30	0.9194	0.9757	0.9914
		50	0.9944	0.9996	1
		100	1	1	1
	JB	30	0.9198	0.9937	0.9994
		50	0.9939	1	1
		100	1	1	1
	L	30	0.9527	0.9894	0.9961
		50	0.9989	0.9999	0.9999
		100	1	1	1
Weibull(0.5,0.5)	A^2	30	1	1	1
		50	1	1	1
		100	1	1	1
	SF	30	0.9996	1	1
		50	1	1	1
		100	1	1	1
	K^2	30	0.9558	0.9914	0.9992
		50	0.9991	1	1
		100	1	1	1
	JB	30	0.9548	0.9991	1
		50	0.9993	1	1
		100	1	1	1
	L	30	0.9957	0.9995	0.9999
		50	1	1	1
		100	1	1	1

Among all distributions studied, the least favorable distribution was the Lognormal (1, 1.3). We found that the Anderson-Darling (AD) test was the best among the five tests studied for this case. Our concept was that the test which was best for the least favorable case would also work well in the other cases. This is borne out by the fact that the AD test is also the best for next worse distribution, which is the Weibull (0.5, 0.5). For other distributions which do not cause deviations in the probabilities associated with the t-statistic, it does not matter how well a normality test does at picking up the deviation from normality.

Jarque-Bera (JB-test) is the most popular and widely use test in the field of economics but our results suggests the overall superiority of Anderson-Darling (AD-test) to Jarque-Bera (JB-test). So, AD-test is recommended for use if the goal is to protect the t-statistic.

4. Conclusion

We have explored 40 distributions and calculated how much they can be damaging for t-statistic. Lognormal (1, 1.3) is the worst distribution for t-statistic among the 40 distributions in our study with 37.5% deviation. Among the tests studied, Anderson-Darling test is the best choice not only against this distribution but also for all other distributions in question to ensure the validity of inferences based on t-statistic. This study has been confined to tests and alternative distributions appearing in the literature, but the approach can easily be generalized.

References

Bartolucci, F., & Scaccia, L. (2005). The use of mixtures for dealing with non-normal regression errors. *Computational Statistics & Data Analysis , 48*, 821-834.

Bonett, D. G., & Seier, E. (2002). A test of normality with high uniform power. *Computational Statistics & Data Analysis , 40*, 435 – 445.

Brys, G., Hubert, M., & Struyf, A. (2004). A Robustification of the Jarque-Bera test of Normality. *Physica-Verlag / Springer* .

Cho, D. W., & Im, K. S. (2002). A Test of Normality Using Geary's Skewness and Kurtosis Statistics. *Working Papers, College of Business & Administration, University of Central Florida*.

D'Agostino, R. B., & Pearson, E.S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\surd b_1$. *Biometrika , 60*, 613-622.

Dufour, J.-M. F. (1998). Simulation-based finite sample normality tests in linear regressions. *Economic Journal* , 1, 154-173.

Geary, R.C. (1947). Testing for normality. *Biometrika*, 34, 209-242.

Gel, Y. R., Miao, W., & Gastwirth, J. L. (2007). Robust direct tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis , 51*, 2734-2746.

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

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review , 55* (2), 163-172.

Onder, A. O., & Zaman, A. (2003). A test for normality based on robust regression residuals. *In: Dutter et. al (Eds.), Development in Robust Statistics. Physica-Verlag, Heidelberg* , 296-306.

Pearson, E. S., D'Agostino, R. B., & Bowman, K. O. (1977). Tests for Departure from Normality: Comparison of Powers. *Biometrika , 64* (2), 231-246.

Seier, E. (2002). Comparison of Tests for Univariate Normality. *Working Papar, East Tennessee State University, USA* , 1-17.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A Comparative Study of Various Tests of Normality. *Journal of American Statist. Assoc. , 63*, 1343-72.

Thadewald & Büning. (2007). Jarque-Bera Test and its Competitors for Testing Normality – A Power Comparison. *Journal of Applied Statistics,* 34 (1), 87-105.

Urzua, C., 1996. On the correct use of onimbus tests for normality. *Economics Letters 53, 247-251.*

Yanagihara, H. (2003). Asymptotic expansion of the null distribution of test statistic for linear hypothesis in non-normal linear models. *Journal of Multivariate Analysis , 84*, 222-246.

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation , 77* (2), 175-183.

Zaman, A., & Onder, O. (2005). Robust tests for normality of errors in regression models. *Economics Letters , 86, 63-68.*

Zaman, A., Rousseuw, P. J., & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters , 71*, 1-8.