University of Victoria
Department of Economics

Econometrics Working Paper EWP0402

ISSN 1485-6441

# TESTING FOR NORMALITY IN THE LINEAR REGRESSION MODEL: AN EMPIRICAL LIKELIHOOD RATIO TEST

**Lauren Bin Dong**
**&**
**David E. A. Giles**

*Department of Economics, University of Victoria*
*Victoria, B.C., Canada V8W 2Y2*

**April, 2004**

**Abstract**

The empirical likelihood ratio (ELR) test for the problem of testing for normality in a linear regression model is derived in this paper. The sampling properties of the ELR test and four other commonly used tests are explored and analyzed using Monte Carlo simulation. The ELR test has good power properties against various alternative hypotheses.

**Keywords:** Regression residual, empirical likelihood ratio, Monte Carlo simulation, normality

**JEL Classifications:** C12, C15, C16

---

**Author Contact:**
Lauren Dong, Statistics Canada; e-mail: Lauren.Dong@statcan.can; FAX: (613) 951-3292

David Giles*, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; FAX: (250) 721-6214

* *Corresponding co-author*

# 1 Introduction

The goal of this paper is to develop an empirical likelihood approach to the problem of testing for normality in a regression model. The maximum empirical likelihood (EL) method is a relatively new nonparametric technique (Owen, 1988, 2001; Qin and Lawless, 1994, 1995; and Mittelhammer *et al.*, 2000) for conducting estimation and hypothesis testing. It is a distribution-free method that still incorporates the notion of the likelihood function. It has several merits. First, it is able to avoid mis-specification problems that can be associated with parametric methods. Second, using the empirical likelihood method enables us to fully use the information available from the data in an asymptotically efficient way.

In this paper, as well as developing an empirical likelihood ratio (ELR) test for normality in a linear regression model, we analyze its sampling properties by undertaking a detailed power comparison of the ELR test and four other commonly used tests through Monte Carlo experiments. The framework of this paper parallels that of Dong and Giles (2004) with respect to the experimental design.

In the classical regression model:

$$y = X\beta + \varepsilon, \tag{1}$$

we usually assume that the error term is normally distributed, $\varepsilon \sim N(0,\ \sigma^2 I)$, and is not correlated with the regressors $x$. The regressor matrix $X$ is non-stochastic and of full rank. Under these assumptions, the OLS estimator is the best unbiased estimator. It is also the maximum likelihood estimator. Thus, we can apply the usual $t$ test and the $F$ test, for linear restrictions on the parameters and we can make other useful inferences. The importance of testing for the normality of the error term is well understood.

As the error term of the regression model, $\varepsilon$, is unobservable, we usually use the OLS residual vector $\hat{\varepsilon}$ from the linear regression model to replace the random error term for testing purposes. The main purpose of this paper is to construct a new test for the normality of the error terms in an ordinary multiple linear regression model. This new test is an ELR test.

The literature on testing for normality in this context is vast. For example, see D'Agostino (1971 and 1972), Bera and Jarque (1980 and 1981), White and MacDonald (1980), and Huang and Bolch (1974). These papers contain details of alternative tests, further references to the literature, and Monte Carlo evidence regarding the relative performances of the tests. We have chosen four of these existing tests to conduct a power comparison in this paper. These tests are the Jarque-Bera (JB) test, D'Agostino's $D$ test, Pearson's $\chi^2$ goodness of fit test, and the $\chi^{2*}$ test which is the adjusted $\chi^2$ test, as detailed below.

The outline of this paper is as follows. Section 2 sets up the model and derives the ELR test. The differences between the BLUS residuals and the OLS residuals are discussed. Section 3 provides the Monte Carlo simulations for the five tests. Random data sets for the error terms under the null hypothesis and under the alternative distributions are generated using the Gauss package (Aptec Systems, 2002). The empirical results are also presented and discussed in this section. Section 4 provides a summary and some conclusions.

## 2 The model

The classical linear regression model has the form:

$$y = X\beta + \varepsilon, \tag{2}$$

where $y$ is a $n \times 1$ vector of observed values of the random dependent variable, $X$ is a known $n \times k$ regressor matrix of rank $k$, $\beta$ is the $k \times 1$ vector of unknown parameters, and the error term $\varepsilon$ is a vector of unobservable stochastic disturbances, assumed to be normally distributed with mean zero and a scalar covariance matrix. That is, $\varepsilon \sim N(0,\ \sigma^2 I)$.

The independent variables in the linear regression model are assumed to be "fixed in repeated samples". In the Monte Carlo experiment described in section 3, we achieve this by drawing observations for these variables independently from a uniform distribution and holding the regressor matrix $X$ constant for each replication of the experiment based upon a given sample size.

## 2.1 OLS and BLUS residuals

The residual vector from the regression model (OLS) is a linear transformation of the error vector. It has the form:

$$\hat{\varepsilon} = M\varepsilon, \tag{3}$$

and has a distribution $\hat{\varepsilon} \sim N(0, \sigma^2 M)$, where $M = I - X(X'X)^{-1}X'$ is a $n \times n$ idempotent symmetric matrix with a rank of $n - k$. The coefficient vector $\beta$ does not appear in this expression. Therefore, we have no need to consider the true or estimated values of the coefficients in the context of testing for normality. The covariance matrix of the residual vector is $\sigma^2 M$, which is not a diagonal matrix and is singular. Therefore, the elements of the residual vector are not independently distributed.

Some researchers (*e.g.*,Huang and Bolch, 1974) prefer to use Theil's (1965, 1968) Best Linear Unbiased Scalar (BLUS) residuals to test for normality in the regression model. The BLUS residual vector is a linear transformation of the OLS residual vector. The BLUS residual vector, $\varepsilon^*$, is obtained from:

$$\varepsilon^* = A'\varepsilon, \tag{4}$$

where $A$ is a $n \times (n-k)$ matrix and it is in the null space of $X$, *i.e.* $X'A = 0$, and $A'A = I_{n-k}$. We note that:

$$E(\varepsilon^*) = 0 \ \text{ and } \ Var(\varepsilon^*) = \sigma^2 I_{n-k}. \tag{5}$$

The covariance matrix of the BLUS residual vector is diagonal and is of full rank. The BLUS residual vector is distributed $N(0,\ \sigma^2 I_{n-k})$ when the error term is from a normal distribution.

Huang and Bolch (1974) proved that *theoretically both the BLUS and the OLS residuals suffer from the common problem of lack of independence under the alternative hypothesis of non-normal disturbances*. The BLUS residuals are independent *if and only if* the error term is independent and normally distributed. Comparing the OLS and the BLUS residuals from the viewpoint of testing for normality, the OLS residuals are at least as good as the BLUS residuals when the underlying distribution is not normally distributed. This is relevant with regard to power considerations. Huang and Bolch (1974) report on Monte Carlo studies where the least squares residual vector $\hat{\varepsilon}$ led to a more powerful test than that obtained by using the BLUS residual vector $\varepsilon^*$. Thus, we focus on the least squares residuals in this study.

An additional problem in testing for normality in a regression model is that the probability distribution of the OLS residuals is always closer to the normal form than is the probability distribution of the disturbance, if the disturbances are not normal. White and McDonald (1980) show that the skewness (positive or negative) and the kurtosis of the OLS residuals will never exceed the skewness and kurtosis of the disturbance term. The residuals for small samples appear more normal than would the unobserved values of the error term, $\varepsilon$. This is called super-normality. Any test for normality using the residuals is more likely to fail in rejecting the null hypothesis when the null is false, than would be the case by using the error term itself (if this were in fact possible) in the construction of the test.

## 2.2 ELR test

Consider the least squares residual vector $\hat{\varepsilon}$ derived from the regression model $y = X\beta + \varepsilon$, where the disturbance term $\varepsilon$ has an unknown distribution with mean zero. Our interest is in testing for normality of the error term using the least squares residuals. The null hypothesis is that the error term is normally distributed, $H_0 : \varepsilon_i' s \sim iidN(0,\ \sigma^2)$, where $i = 1,\ 2,\ \ldots,\ n$. The corresponding residuals, $\hat{\varepsilon}_i$'s are used in the construction of the test.

For each $\hat{\varepsilon}_i$, we assign a probability parameter $p_i$. The empirical likelihood function is naturally formed as $\prod_{i=1}^{n} p_i$. The $p_i$'s are subject to the usual probability constraints: $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$. The EL method is to maximize the empirical likelihood function by choosing the $p_i$'s subject to certain unbiased moment conditions. These moment conditions are naturally derived from the problem in hand. We choose to use the first four unbiased moment equations to match the sample and population moments under the null hypothesis. In order to detect all of the possible departures from normality due to skewness and/or kurtosis, we have taken into account the third and the fourth moments of the residuals. The first four empirical unbiased moment equations have the following form:

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i = 0 \tag{6}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^2 - \sigma^2 = 0 \tag{7}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^3 = 0 \tag{8}$$

$$\sum_{i=1}^{n} p_i \hat{\varepsilon}_i^4 - 3\sigma^4 = 0. \tag{9}$$

As in Dong and Giles (2004), we denote these moment equations as $E_p(h(\hat{\varepsilon},\ \theta)) = 0$.

The Lagrangian function of the log empirical likelihood is formed as

$$\max_{\{p_i,\ \lambda,\ \theta\}} n^{-1} \sum_{i=1}^{n} \log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda' E_p h(\hat{\varepsilon}_i,\ \theta), \tag{10}$$

where $\theta = \sigma^2$, and $\lambda$ is the vector of the Lagrangian multipliers. The optimal value for $\eta$ is unity. The $p_i$ 's can be solved as functions of $\lambda$ and $\theta$: $p_i = n^{-1}(1 + \lambda' E_p(h(\hat{\varepsilon}_i,\ \theta)))^{-1}$. Substituting this information into the Lagrangian function, we get an optimization problem involving only the vector of the Lagrangian multipliers $\lambda$ and the parameter $\sigma^2$. The first order condition with respect to the parameter $\sigma^2$ is:

$$\sum_{i=1}^{n} p_i(\lambda_2 + 6\sigma^2\lambda_4) = 0. \tag{11}$$

With the four moment equations and the first order condition, we have a system of five equations to be solved for the five unknowns. The EL estimators are $\hat{\lambda}$ and $\hat{\sigma}^2$. Substituting these back into the formula for the $p_i$'s, we get the $\hat{p}_i$'s. The empirical likelihood ratio function is $R(p_i^c) = L(p_i^c)/L(p_i^u)$, where $p_i^c = \hat{p}_i$ is the constrained value and $p_i^u = 1/n$ is the unconstrained value. The ELR test statistic has the following form:

$$-2\log R(\sigma^2) = 2\log(1 + \hat{\lambda}' E_{\hat{p}} h(\hat{\varepsilon}_i, \hat{\sigma}^2)). \tag{12}$$

The limiting distribution of the test statistic is $\chi^2_{(d)}$ where the degrees of freedom, $d$, equals the number of the moment constraints less the number of parameters, which implies that $d = 3$. With these theoretical results, we are ready to conduct the Monte Carlo simulations for the ELR test. One point is worth noting: the ELR test has the advantage that it can be applied to problems of testing for *any* type of distribution provided that we adjust the moment equations accordingly. This distinguishes the ELR test from the widely used JB test for normality, for example.

# 3 Monte Carlo simulation

## 3.1 Experimental design

The linear regression model used here, $y = X\beta + \varepsilon$, is the one we presented in Section 2 with four regressors, *i.e.*, $k = 4$. The regressor matrix $X$ is constructed by first obtaining three $2,000 \times 1$ vectors of uniformly distributed random variables. Then we transform the vectors to have zero mean and theoretical unit variance. By adding a vector of ones, the basic $2,000 \times 4$ regressor matrix is formed. For those sample sizes where $n$ is smaller than $2,000$, the regressor matrices are simply obtained by taking the first $n$ rows of the basic regressor $X$ matrix. In the Monte Carlo simulations, for each sample size, the regressor matrix is kept fixed. $10,000$ replications are made for the null and for each of the alternative distributions to provide the sampling properties of the tests, given the various sample sizes from 30 to $2,000$.

The vector of the random disturbance $\varepsilon$ is drawn from the null distribution $N(0,\ \sigma^2 I)$ with the true value $\sigma^2 = 1$. In computing the power of a test, the error vectors of the random disturbances are drawn from the following four alternative distributions : Lognormal, $\chi^2_{(2)}$, Students $t(5)$, and Double Exponential. In each case, the residual vector $\hat{\varepsilon}$ is standardized to have zero mean and unit variance. This particular transformation does not result in any loss of generality. All of the tests that we have considered are invariant with respect to the mean and the variance of the error term $\varepsilon$.

For each replication, five normality tests are applied to the OLS residuals. These five tests are: the ELR test, the Jarque-Bera (1980) (JB) test, D'Agostino's (1971) $D$ test, Pearson's (1900) $\chi^2$ goodness of fit ($\chi^2$) test, and $\chi^{2*}$ test which is the $\chi^2$ goodness of fit test after adjusting for the expected frequencies in each category to be no less than five. The ELR test and the two $\chi^2$ goodness of fit tests are appropriate and readily applicable to the OLS residuals directly. The asymptotic distributions of these test statistics are chi-squared. As for the $D$ test, there is theoretical evidence that it is applicable to the residuals as well. White and McDonald (1980) showed that the $D$ test, and the skewness coefficient, $\sqrt{\alpha_3}$, and the kurtosis coefficient, $\alpha_4$, can be applied to the residuals in testing for normality of the disturbances in regressions. The JB test was proposed specifically for the problem of testing for normality in a regression model.

There are two issues that we should keep in mind. One is that the distribution of the OLS residuals is always closer to normal than is the non-normal random error term itself. Thus, we would expect that any test for normality using regression residuals is more likely to fail in rejecting the null hypothesis when the null is false than would be the case if the test were able to be constructed using the true random error term itself. Second, the distribution of the OLS residual vector $\hat{\varepsilon}$ depends on the distribution of the error term $\varepsilon$, the number of regressors $k$ and the elements of the regressor matrix $X$, and the sample size $n$. Thus, the performance of any test for normality depends on these factors as well.

In the Monte Carlo experiments that are conducted here, we have simulated the size-adjusted critical values for the tests that we consider. These critical values are specific to the regressor matrices we have chosen; they are not applicable to other situations with a different regressor matrix $X$. For those researchers who may be interested in using the ELR test for normality in the context of regression, we will be providing, on the internet, a small library that contains two procedures. One procedure will intake a general regressor matrix $X$ and will calculate the size and the correct size-adjusted critical values. The second procedure will intake the $X$ matrix and the associated size-adjusted critical values and will calculate the actual powers of the ELR test.

## 3.2 Results

Tables 1 to 5 present the results of the Monte Carlo experiments. Table 1 presents the size and the size-adjusted critical values for the ELR test, the JB test, and the two $\chi^2$ tests. The size of a test is computed here as the empirical rejection rate when the null hypothesis is true, given the nominal significance level. The size distortion is the difference between the empirical size of the test and the nominal significance level. We choose to illustrate the size at four nominal significance levels, 10%, 5%, 2%, 1%, in order to provide a broad picture of the sampling properties of the tests. The null hypothesis is $N(0,\ \sigma^2)$ where the true value of $\sigma^2 = 1$. The size and the size-adjusted critical values for the $D$ test are not provided in this table because the percentile points of the $D$ test are taken from D'Agostino (1971 and 1972). For each experimental replication the same data set is used for the construction of all five tests to make the comparisons valid.

The size of the ELR test is larger than the nominal significance level, but it converges nicely to the nominal level as the sample size grows. For example, the size of the test changes from 28.11% to 6.02% when the sample size varies from 30 to 2000, at the nominal significance level of 5%.

From Tables 1 to 5 we see that the size of the JB test is lower than the nominal significance level; it converges to the nominal level as the sample size grows. For instance, the size of the test increases from 2.78% to 4.88% when the sample size varies from 30 to 2000 at the nominal level of 5%. Comparing the actual sizes of the ELR test and the JB test, it is clearly that the ELR test is always over-rejecting while the JB test is under-rejecting. In the context of testing for normality in regression residuals, given the fact that the residuals are closer to the normal than the error term itself, the situation seems to lend a certain advantage to the ELR test. Suppose there is an alternative distribution for the error term that is very close to normal, then the ELR test would have a higher possibility of rejecting the null hypothesis than would the JB test.

The size of the basic $\chi^2$ test is closer to the nominal significance level than is the case for the ELR test for small samples. For example, the size is 5.09% at $n = 30$, and with a nominal significance level of 5%. However, it settles down to 9.1% when the sample size grows to 2000. The size distortion of the test does not vanish when $n$ grows. This is avoidable if we adjust the $\chi^2$ test to take into account the consideration that the expected frequencies in each category should be no less than five. The actual size of the $\chi^{2*}$ test is very close to the nominal significance level at all sample sizes. The size distortion vanishes when $n$ grows. Comparing the four tests, the size of the $\chi^{2*}$ test is the best in the sense it is the closest to the nominal significance level. It exhibits the least size distortion, overall.

The power comparisons of the five selected tests are presented in Tables 2 to 5. The simulated size-adjusted critical values are used in computing the power of each test. That is, each test is now applied at the same actual significance level by using the critical values that are appropriate for the particular value of $n$.

The power of the ELR test increases as the sample size grows, given any alternative distribution. For example, the power increases from 69.39% to 100% when $n$ grows from 30 to 100, given that the alternative distribution is the Lognormal, and at a true significance level of 5%. All of other tests are also consistent while the power of the $\chi^2$ test converges more slowly than others. The powers of the first three tests, the ELR test, the JB test, and the $D$ test, are close to each other for moderate samples. For instance, the powers of these tests are 97.91%, 99.14%, 93.37% at $n = 50$ and with a significance level of 5%. For the small sample size $n = 30$, the JB test is the most powerful test, and the ELR test is the second most powerful. The $\chi^2$ tests are not applicable for some of the smaller sample sizes that are considered. Overall, all of the five tests have good power against the lognormal alternative distribution.

When the alternative distribution is $\chi^2_{(2)}$, the power of the ELR test is 55.29% at $n = 30$ and reaches 100% at $n = 100$ and at the nominal level of 5%. The power of the JB test is slightly better than that of the ELR test. The power of the tests in descending order is: the JB test, the ELR test, the $D$, the $\chi^2$ and the $\chi^{2*}$ tests for all sample sizes.

Both the Lognormal and the $\chi^2_{(2)}$ are asymmetric alternative distributions. It is, relative to symmetric alternatives, easier for any of these tests to detect such a departure from the normality. All of these tests have high power for medium and large samples.

The Student $t_{(5)}$ and the Double Exponential (DE) alternative distributions are both symmetric with fatter tails than the normal distribution. It is difficult for any commonly used tests to detect for the departure from normality in such cases. All of the tests have lower power for small sample size $n = 30$ over these two alternatives than was the case with asymmetric alternative distributions. The power of the ELR test is only 5.68% for the $St_{(5)}$ and is 6.8% for the DE at $n = 30$. For the alternative $St_{(5)}$, the power of the D test overrides the JB test when $n > 150$. For the alternative distribution DE, the D test is the most powerful test among the five tests when $n > 30$. The power of the ELR test reaches 100% at $n = 1000$.

# 4 Summary and conclusions

A new empirical likelihood ratio test for normality in a regression model is derived in this paper. Monte Carlo experiments are employed to simulate the sampling properties of the ELR test and other four commonly used tests in the context of regression. These properties include the actual sizes and the size-adjusted critical values which are then used to conduct power comparisons for all of the tests. The experiment results indicate that the actual size of the ELR test is still large relative to the nominal significance level. However, using the actual significance level, the ELR test has good power properties relative to other tests, and is recommended for use in regression context. For future research, it would be worthwhile to explore the techniques that can reduce the size distortion and maintain the good power properties for the ELR test.

## Acknowledgements

This paper is based on one chapter of the first author's Ph.D. dissertation, completed in the Department of Economics, University of Victoria, in December 2003. Special thanks go to Don Ferguson, Ron Mittelhammer, Min Tsao, Graham Voss and Julie Zhou for their many helpful suggestions and contributions.

**Table 1**: Size and Size-adjusted Critical Values for the Tests in the OLS Residuals

| $m$ | 10,000 | | $H_0 : N(0, \sigma^2)$ where $\sigma^2 = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 30 | 50 | 70 | 100 | 150 | 200 | 250 | 500 | 1,000 | 2,000 |
| *ELR test at nominal levels:* | | | | | | | | | | |
| 10% | 0.2638 | 0.2276 | 0.2063 | 0.1771 | 0.1615 | 0.1408 | 0.1344 | 0.1076 | 0.0776 | 0.0648 |
| 5% | 0.1878 | 0.1584 | 0.1423 | 0.1241 | 0.1042 | 0.0924 | 0.0838 | 0.0640 | 0.0431 | 0.0345 |
| 2% | 0.1185 | 0.0974 | 0.0885 | 0.0771 | 0.0633 | 0.0550 | 0.0492 | 0.0358 | 0.0202 | 0.0155 |
| 1% | 0.0801 | 0.0674 | 0.0626 | 0.0520 | 0.0419 | 0.0368 | 0.0325 | 0.0238 | 0.0114 | 0.0089 |
| *Size-adjusted Critical Values:* | | | | | | | | | | |
| 10% | 10.43 | 9.70 | 9.25 | 8.68 | 7.98 | 7.49 | 7.27 | 6.47 | 5.61 | 5.19 |
| 5% | 12.95 | 12.59 | 12.21 | 11.45 | 10.65 | 10.17 | 9.74 | 8.61 | 7.47 | 6.83 |
| 2% | 16.64 | 16.27 | 16.11 | 15.47 | 14.42 | 13.87 | 13.28 | 11.80 | 9.85 | 9.20 |
| 1% | 19.29 | 18.81 | 18.58 | 18.05 | 17.32 | 16.60 | 16.07 | 14.36 | 11.77 | 10.95 |
| *JB test:* | | | | | | | | | | |
| 10% | 0.0446 | 0.0552 | 0.0613 | 0.0671 | 0.0737 | 0.0748 | 0.0805 | 0.0881 | 0.0927 | 0.0963 |
| 5% | 0.0284 | 0.0361 | 0.0382 | 0.0423 | 0.0438 | 0.0415 | 0.0463 | 0.0477 | 0.0515 | 0.0506 |
| 2% | 0.0181 | 0.0240 | 0.0228 | 0.0252 | 0.0256 | 0.0240 | 0.0248 | 0.0223 | 0.0246 | 0.0231 |
| 1% | 0.0126 | 0.0187 | 0.0161 | 0.0173 | 0.0179 | 0.0172 | 0.0174 | 0.0127 | 0.0146 | 0.0136 |
| *Size-adjusted Critical Values:* | | | | | | | | | | |
| 10% | 2.72 | 3.12 | 3.46 | 3.65 | 3.89 | 4.06 | 4.12 | 4.32 | 4.43 | 4.53 |
| 5% | 4.29 | 4.94 | 5.14 | 5.42 | 5.62 | 5.52 | 5.78 | 5.91 | 6.05 | 6.02 |
| 2% | 7.28 | 8.81 | 8.23 | 8.83 | 8.81 | 8.49 | 8.58 | 8.00 | 8.29 | 8.17 |
| 1% | 10.32 | 12.73 | 11.20 | 12.04 | 11.54 | 12.33 | 11.76 | 10.13 | 10.31 | 9.99 |
| $\chi^2$ *goodness of fit test:* | | | | | | | | | | |
| 10% | 0.0593 | 0.0662 | 0.0743 | 0.0880 | 0.1006 | 0.1063 | 0.1134 | 0.1193 | 0.1214 | 0.1270 |
| 5% | 0.0262 | 0.0351 | 0.0404 | 0.0528 | 0.0628 | 0.0675 | 0.0720 | 0.0758 | 0.0798 | 0.0782 |
| 2% | 0.0110 | 0.0170 | 0.0220 | 0.0296 | 0.0379 | 0.0441 | 0.0463 | 0.0485 | 0.0490 | 0.0482 |
| 1% | 0.0052 | 0.0110 | 0.0147 | 0.0216 | 0.0284 | 0.0329 | 0.0350 | 0.0356 | 0.0341 | 0.0363 |
| *Size-adjusted Critical Values:* | | | | | | | | | | |
| 10% | 6.50 | 10.51 | 14.43 | 20.59 | 30.52 | 38.78 | 46.92 | 57.19 | 58.86 | 60.14 |
| 5% | 8.11 | 12.73 | 16.99 | 23.96 | 34.98 | 44.23 | 52.68 | 63.90 | 65.47 | 66.54 |
| 2% | 9.98 | 15.75 | 20.89 | 29.16 | 42.34 | 53.12 | 62.82 | 73.67 | 76.18 | 80.09 |
| 1% | 11.79 | 18.46 | 24.99 | 36.86 | 51.78 | 67.35 | 79.34 | 85.79 | 90.22 | 97.87 |
| $\chi^{2*}$ *goodness of fit test:* | | | | | | | | | | |
| 10% | – | 0.0812 | 0.0800 | 0.0818 | 0.0820 | 0.0835 | 0.0859 | 0.0826 | 0.0799 | 0.0830 |
| 5% | – | 0.0393 | 0.0400 | 0.0414 | 0.0415 | 0.0407 | 0.0406 | 0.0413 | 0.0395 | 0.0398 |
| 2% | – | 0.0157 | 0.0151 | 0.0154 | 0.0161 | 0.0162 | 0.0168 | 0.0155 | 0.0177 | 0.0155 |
| 1% | – | 0.0073 | 0.0079 | 0.0076 | 0.0082 | 0.0089 | 0.0095 | 0.0085 | 0.0087 | 0.0086 |
| *Size-adjusted Critical Values:* | | | | | | | | | | |
| 10% | – | 9.14 | 11.13 | 14.37 | 19.72 | 24.58 | 29.41 | 39.72 | 44.07 | 47.48 |
| 5% | – | 11.08 | 13.37 | 16.87 | 22.41 | 27.58 | 32.42 | 43.42 | 47.92 | 51.34 |
| 2% | – | 13.42 | 16.13 | 19.98 | 25.79 | 31.33 | 36.58 | 47.53 | 52.93 | 56.06 |
| 1% | – | 15.34 | 18.19 | 21.68 | 28.48 | 33.75 | 39.74 | 50.94 | 56.17 | 59.50 |

Note: $m$ is the number of replications; $n$ is the sample size. The ELR test uses the first four raw moment equations with one parameter. The $\chi^2$ and $\chi^{2*}$ tests may not be applicable with small samples.

**Table 2**: Power of The Five Tests in the OLS Residuals

| $m$ | 10,000 | | $H_a$ : Lognormal | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | 30 | 50 | 70 | 100 | 150 | 200 | 250 |
| *ELR test at actual levels:* | | | | | | | |
| 10% | 0.8637 | 0.9968 | 1 | 1 | 1 | 1 | 1 |
| 5% | 0.7785 | 0.9874 | 0.9996 | 1 | 1 | 1 | 1 |
| 2% | 0.6297 | 0.9647 | 0.9984 | 1 | 1 | 1 | 1 |
| 1% | 0.5174 | 0.9355 | 0.9961 | 1 | 1 | 1 | 1 |
| *JB test:* | | | | | | | |
| 10% | 0.9339 | 0.9978 | 1 | 1 | 1 | 1 | 1 |
| 5% | 0.8715 | 0.9895 | 0.9999 | 1 | 1 | 1 | 1 |
| 2% | 0.7670 | 0.9627 | 0.9981 | 1 | 1 | 1 | 1 |
| 1% | 0.6837 | 0.9255 | 0.9956 | 0.9997 | 1 | 1 | 1 |
| *D test:* | | | | | | | |
| 10% | 0.8262 | 0.9735 | 0.9970 | 0.9997 | 1 | 1 | 1 |
| 5% | 0.7648 | 0.9585 | 0.9957 | 0.9994 | 1 | 1 | 1 |
| 2% | 0.6898 | 0.9327 | 0.9915 | 0.9988 | 1 | 1 | 1 |
| 1% | 0.6272 | 0.9095 | 0.9869 | 0.9986 | 1 | 1 | 1 |
| $\chi^2$ *goodness of fit test:* | | | | | | | |
| 10% | – | – | – | 0.9819 | 0.9984 | 0.9991 | 0.9992 |
| 5% | – | – | – | 0.9696 | 0.997 | 0.9988 | 0.9991 |
| 2% | – | – | – | 0.9486 | 0.9932 | 0.9982 | 0.9989 |
| 1% | – | – | – | 0.9166 | 0.9879 | 0.9969 | 0.9984 |
| $\chi^{2*}$ *goodness of fit test:* | | | | | | | |
| 10% | – | – | – | – | – | 0.9999 | 1 |
| 5% | – | – | – | – | – | 0.9998 | 1 |
| 2% | – | – | – | – | – | 0.9992 | 1 |
| 1% | – | – | – | – | – | 0.9985 | 1 |

Note: $m$ is the number of replications; $n$ is the sample size. The ELR test uses the first four raw moment equations with one parameter. The $\chi^2$ and $\chi^{2*}$ tests may not be applicable with small samples.

**Table 3**: Power of The Five Tests in the OLS Residuals

| $m$ | 10,000 | | $H_a : \chi^2_{(2)}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 30 | 50 | 70 | 100 | 150 | 200 | 250 | 500 | 1,000 | 2,000 |
| *ELR test at actual levels:* | | | | | | | | | | |
| 10% | 0.7548 | 0.9803 | 0.9988 | 0.9999 | 1 | 1 | 1 | | | |
| 5% | 0.6362 | 0.953 | 0.9963 | 0.9999 | 1 | 1 | 1 | | | |
| 2% | 0.4774 | 0.8938 | 0.9881 | 0.9997 | 1 | 1 | 1 | | | |
| 1% | 0.3674 | 0.8362 | 0.9782 | 0.9993 | 1 | 1 | 1 | | | |
| *JB test:* | | | | | | | | | | |
| 10% | 0.8217 | 0.9841 | 0.9991 | 0.9999 | 1 | 1 | 1 | | | |
| 5% | 0.6855 | 0.9328 | 0.9929 | 0.9998 | 1 | 1 | 1 | | | |
| 2% | 0.5103 | 0.8101 | 0.9616 | 0.9976 | 1 | 1 | 1 | | | |
| 1% | 0.4025 | 0.6974 | 0.9165 | 0.9901 | 1 | 1 | 1 | | | |
| *D test:* | | | | | | | | | | |
| 10% | 0.5986 | 0.8434 | 0.9403 | 0.9891 | 0.9996 | 0.9999 | 1 | | | |
| 5% | 0.5000 | 0.7854 | 0.9116 | 0.9801 | 0.9988 | 0.9999 | 1 | | | |
| 2% | 0.3911 | 0.7025 | 0.8653 | 0.9656 | 0.9973 | 0.9998 | 1 | | | |
| 1% | 0.3172 | 0.6352 | 0.8250 | 0.9498 | 0.9949 | 0.9997 | 1 | | | |
| $\chi^2$ *goodness of fit test:* | | | | | | | | | | |
| 10% | – | 0.6409 | 0.8204 | 0.9546 | 0.9977 | 0.9997 | 1 | | | |
| 5% | – | 0.5411 | 0.743 | 0.9256 | 0.9945 | 0.9993 | 1 | | | |
| 2% | – | 0.4307 | 0.6343 | 0.865 | 0.9863 | 0.9983 | 0.9999 | | | |
| 1% | – | 0.3519 | 0.5283 | 0.7523 | 0.9595 | 0.9935 | 0.9993 | | | |
| $\chi^{2*}$ *goodness of fit test:* | | | | | | | | | | |
| 10% | – | – | – | 0.9237 | 0.9960 | 0.9996 | 1 | | | |
| 5% | – | – | – | 0.8729 | 0.9919 | 0.999 | 1 | | | |
| 2% | – | – | – | 0.7976 | 0.9828 | 0.998 | 1 | | | |
| 1% | – | – | – | 0.7481 | 0.9710 | 0.9967 | 0.9995 | | | |

Note: $m$ is the number of replications; $n$ is the sample size. The ELR test uses the first four raw moment equations with one parameter. The $\chi^2$ and $\chi^{2*}$ tests may not be applicable with small samples.

**Table 4**: Power of the Five Tests in the OLS Residuals

| $m$ | 10,000 | | $H_a$ : Student $t_{(5)}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | 30 | 50 | 70 | 100 | 150 | 200 | 250 | 500 | 1,000 |
| *ELR test at actual levels:* | | | | | | | | | |
| 10% | 0.1008 | 0.1290 | 0.1987 | 0.3414 | 0.5541 | 0.7255 | 0.8291 | 0.9886 | 1 |
| 5% | 0.0568 | 0.0681 | 0.1063 | 0.2174 | 0.4175 | 0.5931 | 0.7330 | 0.9777 | 1 |
| 2% | 0.0239 | 0.0262 | 0.0433 | 0.1054 | 0.2505 | 0.4216 | 0.5837 | 0.9497 | 1 |
| 1% | 0.0130 | 0.0153 | 0.0248 | 0.0651 | 0.1636 | 0.3159 | 0.4693 | 0.9161 | 0.9998 |
| *JB test:* | | | | | | | | | |
| 10% | 0.3152 | 0.4632 | 0.5732 | 0.6962 | 0.8226 | 0.8975 | 0.9365 | 0.9963 | 1 |
| 5% | 0.2332 | 0.3687 | 0.4933 | 0.6162 | 0.7654 | 0.8644 | 0.9121 | 0.9939 | 1 |
| 2% | 0.1567 | 0.2693 | 0.3988 | 0.5120 | 0.6805 | 0.7975 | 0.8697 | 0.9896 | 1 |
| 1% | 0.1171 | 0.2121 | 0.3375 | 0.4494 | 0.6221 | 0.7268 | 0.8224 | 0.9838 | 1 |
| *D test:* | | | | | | | | | |
| 10% | 0.2533 | 0.4036 | 0.5453 | 0.6771 | 0.8226 | 0.9090 | 0.9479 | 0.9987 | 1 |
| 5% | 0.1700 | 0.3168 | 0.4526 | 0.5902 | 0.7625 | 0.8671 | 0.9192 | 0.9964 | 1 |
| 2% | 0.1080 | 0.2337 | 0.3546 | 0.4941 | 0.6764 | 0.8045 | 0.8798 | 0.9932 | 1 |
| 1% | 0.0791 | 0.1821 | 0.2892 | 0.4302 | 0.6149 | 0.7556 | 0.8448 | 0.9882 | 1 |
| $\chi^2$ *goodness of fit test:* | | | | | | | | | |
| 10% | – | 0.2982 | 0.3660 | 0.4393 | 0.5349 | 0.6233 | 0.6810 | 0.9153 | 0.9979 |
| 5% | – | 0.2184 | 0.2841 | 0.3620 | 0.4610 | 0.5504 | 0.6201 | 0.8767 | 0.9956 |
| 2% | – | 0.1566 | 0.2146 | 0.2928 | 0.3911 | 0.4809 | 0.5522 | 0.8279 | 0.9884 |
| 1% | – | 0.1271 | 0.1776 | 0.2414 | 0.3439 | 0.4216 | 0.4873 | 0.7765 | 0.9729 |
| $\chi^{2*}$ *goodness of fit test:* | | | | | | | | | |
| 10% | – | – | – | 0.2221 | 0.2335 | 0.2404 | 0.2404 | 0.3603 | 0.8587 |
| 5% | – | – | – | 0.1323 | 0.1476 | 0.1534 | 0.1561 | 0.2500 | 0.7758 |
| 2% | – | – | – | 0.0675 | 0.0782 | 0.0822 | 0.0805 | 0.1540 | 0.6439 |
| 1% | – | – | – | 0.0439 | 0.0463 | 0.0545 | 0.0472 | 0.0998 | 0.5473 |

Note: $m$ is the number of replications; $n$ is the sample size. The ELR test uses the first four raw moment equations with one parameter. The $\chi^2$ and $\chi^{2*}$ tests may not be applicable with small samples.

**Table 5**: Power of the Five Tests in the OLS Residuals

| $m$ | 10,000 | $H_a$ : Double Exponential | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | 30 | 50 | 70 | 100 | 150 | 200 | 250 | 500 | 1,000 |
| *ELR test at actual levels:* | | | | | | | | | |
| 10% | 0.1057 | 0.1695 | 0.3066 | 0.5374 | 0.8045 | 0.9299 | 0.9751 | 1 | 1 |
| 5% | 0.0618 | 0.0893 | 0.1752 | 0.3750 | 0.6824 | 0.8606 | 0.9458 | 1 | 1 |
| 2% | 0.0296 | 0.0397 | 0.0738 | 0.1978 | 0.4949 | 0.7389 | 0.8792 | 0.9995 | 1 |
| 1% | 0.0170 | 0.0228 | 0.0442 | 0.1277 | 0.3648 | 0.6264 | 0.8084 | 0.9977 | 1 |
| *JB test:* | | | | | | | | | |
| 10% | 0.3919 | 0.5738 | 0.7044 | 0.8340 | 0.9380 | 0.9763 | 0.9916 | 1 | 1 |
| 5% | 0.2925 | 0.4692 | 0.6161 | 0.7645 | 0.9003 | 0.9604 | 0.9828 | 1 | 1 |
| 2% | 0.1951 | 0.3383 | 0.5073 | 0.6523 | 0.8244 | 0.9248 | 0.9676 | 0.9999 | 1 |
| 1% | 0.1419 | 0.2640 | 0.4300 | 0.5682 | 0.7680 | 0.8712 | 0.9413 | 0.9999 | 1 |
| *D test:* | | | | | | | | | |
| 10% | 0.3453 | 0.5813 | 0.7514 | 0.8914 | 0.9713 | 0.9936 | 0.9988 | 1 | 1 |
| 5% | 0.2409 | 0.4714 | 0.6611 | 0.8332 | 0.9489 | 0.9871 | 0.9963 | 1 | 1 |
| 2% | 0.1530 | 0.3471 | 0.5396 | 0.7415 | 0.9115 | 0.9721 | 0.9902 | 1 | 1 |
| 1% | 0.1083 | 0.2708 | 0.4577 | 0.6726 | 0.8725 | 0.9559 | 0.9844 | 1 | 1 |
| $\chi^2$ *goodness of fit test:* | | | | | | | | | |
| 10% | – | 0.4301 | 0.5381 | 0.6545 | 0.7771 | 0.8667 | 0.9152 | 0.9980 | 1 |
| 5% | – | 0.3221 | 0.4310 | 0.5472 | 0.6848 | 0.7870 | 0.8588 | 0.9956 | 1 |
| 2% | – | 0.2238 | 0.3187 | 0.4230 | 0.5489 | 0.6611 | 0.7510 | 0.9855 | 1 |
| 1% | – | 0.1727 | 0.2483 | 0.3215 | 0.4381 | 0.5237 | 0.6165 | 0.9567 | 1 |
| $\chi^2$ *goodness of fit test:* | | | | | | | | | |
| 10% | – | 0.2733 | 0.3389 | 0.4272 | 0.5649 | 0.6623 | 0.7296 | 0.9744 | 1 |
| 5% | – | 0.1827 | 0.2258 | 0.3013 | 0.4365 | 0.5387 | 0.6238 | 0.9489 | 1 |
| 2% | – | 0.1101 | 0.1319 | 0.1809 | 0.3029 | 0.3947 | 0.4840 | 0.9036 | 1 |
| 1% | – | 0.0707 | 0.0873 | 0.1351 | 0.2198 | 0.3124 | 0.3716 | 0.8508 | 1 |

Note: $m$ is the number of replications; $n$ is the sample size. The ELR test uses the first four raw moment equations with one parameter. The $\chi^2$ and $\chi^{2*}$ tests may not be applicable with small samples.

# References

Aptech Systems 2002, Gauss 5.0 for Windows NT, Aptech Systems, Inc., Maple Valley WA.

Bera, A., Jarque, C., 1980. Efficient tests for normality, heteroscedasticity, and serial independence of regression residuals, Economics Letters 6, 255-259.

Bera, A., Jarque, C., 1981. Efficient tests for normality, heteroscedasticity, and serial independence of regression residuals: monte carlo evidence, Economics Letters 7, 313-318.

D'Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples, Biometrika 58, 341 - 348.

D'Agostino, R.B., 1972. Small sample probability points for the d test of normality, Biometrika 59, 219-221.

Dong, L.B., Giles, D.E.A., 2004. An empirical likelihood ratio test for normality, Working paper EWP0401, Department of Economics University of Victoria.

Huang, C.J., Bolch, B.W., 1974. On the testing of regression disturbances for normality, Journal of the American Statistical Association 69, 330-335.

Mittelhammer, R., Judge, G., Miller, D., 2000. Econometric foundations (Cambridge University Press, Cambridge).

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75, 237 - 249.

Owen, A.B., 2001. Empirical likelihood (Chapman & Hall/CRC, New York).

Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philosophical Magazine Series 50, 157 - 175.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations, The Annals of Statistics, 22, 300 - 325.

Qin, J., Lawless, J., 1995. Estimating equations, empirical likelihood and constraints on parameters, Canadian Journal of Statistics 23, 145 - 159.

Theil, H., 1965. The analysis of disturbances in regression analysis, Journal of the American Statistical Association 60, 1067 - 1079.

Theil, H., 1968. A simplification of the BLUS procedure for analyzing regression disturbances, Journal of the American Statistical Association 63, 242 - 251.

White, H., MacDonald, G., 1980. Some large-sample tests for nonnormality in the linear regression model, Journal of the American Statistical Association 75, 16 - 31.