UVic
University of Victoria

Department of Economics

AN EMPIRICAL LIKELIHOOD RATIO TEST FOR NORMALITY

Lauren Bin Dong
&
David E. A. Giles

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

February, 2004

Abstract

The empirical likelihood ratio (ELR) test for the problem of testing for normality is derived in this paper. The sampling properties of the ELR test and four other commonly used tests are provided and analyzed using the Monte Carlo simulation technique. The power comparisons against a wide range of alternative distributions show that the ELR test is the most powerful of these tests in certain situations.

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

1 INTRODUCTION

The purpose of this paper is to develop an empirical likelihood approach to the problem of testing for normality in a population. The maximum empirical likelihood (EL) method is a relatively recently developed nonparametric technique (Owen, 1988) for conducting estimation and hypothesis testing. It is a distribution-free method that still incorporates the notions of the likelihood function and the likelihood ratio. 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 employ 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, we analyze its sampling properties by undertaking a detailed power comparison of the ELR test and four other commonly used tests. It is well known that a normal distribution has skewness coefficient $\alpha_3 = 0$ and kurtosis coefficient $\alpha_4 = 3$. The sample skewness and kurtosis statistics are excellent descriptive and inferential measures for evaluating normality. Any test based on skewness or kurtosis is usually called an omnibus test. An omnibus test is sensitive to various forms of departure from normality. Among the commonly used tests for normality, the Jarque-Bera (1980) test (JB), D'Agostino's (1971) D test, and Pearson's (1900) χ^2 goodness of fit test (χ^2 test) are selected. These are all omnibus tests. Using them separately gives us the opportunity of testing for departures from normality in different respects.

Random data sets are generated using the Gauss package (Aptech Systems, 2002). In each replication, the same data set is used for all of the tests that we consider. The five tests, the ELR, the JB, the D test, the χ^2, and the χ^{2*} (the adjusted χ^2 test to be defined in section 3.3) are all asymptotic tests. The properties of the tests in finite samples are unknown, although some of them have received some previous consideration in the literature. We simulate their actual sizes and calculate their size-adjusted critical values. These results allow us to undertake a power comparison of the tests at the same actual significance levels. One exception is the D test. The actual critical values of the D test are taken from D'Agostino (1971 and 1972). The reason for this is given Section 3.2. We find that the ELR test has good power properties and it is invariant with respect to the form of the information constraints. These results are robust with respect to various changes in the parameters and to the form of the alternative hypothesis. We recommend the use of the ELR test for normality.

1

The outline of this paper is as follows. Section 2 describes the approach of using the empirical likelihood method and the ELR test for the problem of testing for normality. Section 3 discusses the conventional tests that we consider. Section 4 outlines the Monte Carlo simulation experiment and provides the empirical results of the tests. Some of the computational issues associated with the ELR test are discussed in Section 5, and Section 6 provides a summary and our conclusions.

2 ELR TEST

The main focus of this section is to derive an ELR test. Consider a random data set of size n: y_1, y_2, ..., y_n which is i.i.d. and has a common distribution $F_0(\theta)$ that is unknown. θ is the parameter vector of the underlying distribution. In the context of testing for normality, the parameter vector becomes $\theta = (\mu, \sigma^2)'$. Our interest is to test for normality $H_0 : N(\mu, \sigma^2)$ using the information from the sample and the empirical likelihood approach.

2.1 EL Method

The EL method has many favorable features. First, the method utilizes the concept of likelihood functions, which is very important. The likelihood method is very flexible. It is able to incorporate the information from different data sources and knowledge arising from outside of the sample of data. The assumption of the underlying data distribution is important in constructing a parametric likelihood function. The usual parametric likelihood methods lead to asymptotically best estimators and asymptotically powerful tests of the parameters if the specification of the underlying distribution is correct. The term "best" means that the estimator has the minimum asymptotic variance. The likelihood ratio test and the Wald test can be constructed based on the estimates and distributional assumptions to make useful inferences. A problem with parametric likelihood inference is that we may not know the correct distributional family to use and there is usually not sufficient information to assume that a data set is from a specific parametric distribution family. Mis-specification can cause likelihood based estimates to be inefficient and inconsistent, and inferences based on the wrongly specified underlying distribution can be completely inappropriate. Using the

empirical likelihood method, we are able to avoid mis-specification problems that can be associated with parametric methods.

Second, the empirical likelihood method enables us to fully employ the information available from the data in an asymptotically efficient way. It is well known that the generalized method of moments (GMM) approach uses the estimating equations to provide asymptotically efficient estimates for parameters of interest using the information constraints. The empirical likelihood method is able to use the same set of estimating equations together with the empirical likelihood function approach to provide the empirical likelihood estimates for the parameters. The empirical likelihood estimator is obtained in an operationally optimal way and is asymptotically as efficient as the GMM estimator. The ability to incorporate both the likelihood approach and estimating equations should also benefit the ELR test from a power perspective.

2.2 ELR Test

The ELR test is based on the empirical likelihood function. First, we assign a probability parameter p_i to each data point y_i and then form the empirical likelihood function $L(F) = \prod_{i=1}^{n} p_i$. The p_i's are subject to the usual probability constraints: $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$. The maximum empirical likelihood method is to maximize the likelihood function subject to information constraints. These constraints arise from the data naturally: they are the moment equations and the probability constraints. We will match the sample and population moments. Let $h(y, \theta)$ be the moment function vector. Under the null hypothesis that the data are from a normal distribution with mean μ and variance σ^2, the first four unbiased empirical moment equations, $E_p(h(y, \theta)) = 0$, have the form:

$$\sum_{i=1}^{n} p_i y_i - \mu = 0 \tag{1}$$

$$\sum_{i=1}^{n} p_i y_i^2 - (\mu^2 + \sigma^2) = 0 \tag{2}$$

$$\sum_{i=1}^{n} p_i y_i^3 - (\mu^3 + 3\sigma^2\mu) = 0 \tag{3}$$

$$\sum_{i=1}^{n} p_i y_i^4 - (\mu^4 + 6\sigma^2\mu^2 + 3\sigma^4) = 0. \tag{4}$$

3

The first term on the left hand of each equation is the sample moment; the second term is the population moment under the null hypothesis H_0. We match the two terms to set up the moment equation. We denote this system of equations as $E_p(h(y, \theta)) = 0$. The probability constraints are the usual ones: $0 < p_i < 1$ and $\sum_{i=1}^{n} p_i = 1$.

The reasons that we have chosen to use the first four moment equations are as follows. First, we need at least three moment equations so that the number of moment equations, m, is greater than the number of parameters, p. If $m < p$, then the system is under-identified; there will be a set of solutions to the system. If $m = p$, then, the solution to the estimating-equation system is exactly the solution to the EL approach with $p_i = 1/n$. Only when $m > p$, the EL system will yield a unique and more efficient solution. Second, we would like to make the various tests that we have consider comparable. The JB test uses the standardized first four moments, therefore, the ELR test should use four moment equations.

We transform the objective function by taking the natural logarithm of the likelihood function. This is an affine transformation and it does not alter the location of the maximum of the objective function. The log empirical likelihood is of the form: $l(F) = \sum_{i=1}^{n} \log p_i$. The constrained optimization problem is then set up in the Lagrangian function form:

$$G = n^{-1} \sum_{i=1}^{n} \log p_i - \eta(\sum_{i=1}^{n} p_i - 1) - \lambda' E_p h(y, \theta). \tag{5}$$

Making use of the first order conditions of the Lagrangian function with respect to the probability parameter p_i's and the constraint of p_i's, we find that

$$p_i = n^{-1}(1 + \lambda' E_p(h(y_i, \ \theta)))^{-1}.$$

The optimal value for the Lagrangian multiplier η is unity. Substituting the p_i's and η into the Lagrangian function, the original maximization problem over the p_i's, λ, and θ is transformed into an optimization problem over a smaller number of parameters, namely the elements of the vector λ and θ.

The first order conditions of the Lagrangian function with respect to the parameter

vector $\theta = (\mu, \ \sigma^2)'$ have the form:

$$\sum_{i=1}^{n} p_i(\lambda_1 + 2\mu\lambda_2 + 3(\mu^2 + \sigma^2)\lambda_3 + 4(\mu^3 + 3\sigma^2\mu)\lambda_4) = 0 \tag{6}$$

$$\sum_{i=1}^{n} p_i(\lambda_2 + 3\mu\lambda_3 + 6(\mu^2 + \sigma^2)\lambda_4) = 0. \tag{7}$$

With the four moment equations and the two first order conditions, the solution $\hat{\theta}$ and $\hat{\lambda}$ can be obtained using a nonlinear equation solver procedure. In our study, we use Eqsolve, one of the nonlinear equation solver in the Gauss package. The log likelihood function here is log-concave, and the constraint functions are well behaved with positive coefficients associated with parameter terms. Therefore, the conditions for a unique solution are satisfied.

The EL estimator of the parameter vector is $\hat{\theta}$ and the estimated Lagrangian multiplier vector is $\hat{\lambda}$. Substituting these values into the formula for the p_i's, we get the $\hat{p}_i$'s as the estimated probability values for the y_i's. The estimated maximum value of the empirical likelihood function is $L(\hat{F}) = \prod_{i=1}^{n} \hat{p}_i$.

The null hypothesis and the alternative hypothesis for the ELR test are:

$$H_0 : y_i's \sim iidN(\mu, \sigma^2); \qquad H_a : \text{not } H_0.$$

The empirical likelihood ratio function has the form: $R(F) = \frac{L(F)}{L(F_n)}$, where F is the underlying distribution and $L(F_n) = n^{-n}$. Under the null hypothesis, minus two times the log empirical likelihood ratio has the limiting distribution (Owen, 1988):

$$-2 \log R(F) \xrightarrow{d} \chi^2_{(m-p)}$$

where m is the number of moment equations and p is the number of parameters of interest.

The value of the ELR test statistic based on the values of the restricted and unrestricted empirical likelihood functions is:

$$-2 \log R(\hat{\theta}) = -2 \log(L(\hat{F})/L(F_n)) \tag{8}$$

$$= 2 \sum_{i=1}^{n} \log(1 + \hat{\lambda}'h(y_i, \hat{\theta})). \tag{9}$$

The ELR test is an asymptotic test. The actual sizes of the ELR test for finite samples are unknown and are therefore computed using Monte Carlo simulations. We reject the null hypothesis when the value of the test statistic is greater than the critical value based on the asymptotic distribution of the test statistic. The total number of the rejections are counted and are divided by the number of replications, which gives us the actual rejection rate. This rejection rate is considered as the actual size of the test for this value of n, given that the number of the replication is large enough, say $10,000$. The values of the ELR test statistic are stored and sorted in ascending order so that the percentiles of their empirical distribution can be determined. In this way we can obtain, say, 10%, 5%, 2% and 1%, size-adjusted critical values. In another words the size-adjusted critical values are the values of the test statistic when the actual sizes of the test equal the nominal significance levels. These critical values can then be used to simulate the power of the test in finite samples, by considering various forms of the alternative hypothesis

3   CONVENTIONAL TESTS

We have chosen other four commonly used tests in testing for normality as the competitors. They are the Jarque-Bera test, the D'Agostino's test, Peason's χ^2 goodness of fit test and the χ^{2*} test which is the adjusted χ^2 goodness of fit test. The set-up of each test is given below.

3.1   Jarque-Bera Test (JB)

The JB test was proposed by Jarque and Bera (1980). This test is based on the difference between the skewness and kurtosis of the data set $\{y_1, \ y_2, \ \ldots, \ y_n\}$ and those of the assumed normal distribution.

The null hypothesis and the alternative for the JB test are:

$$H_0 : y_i's \sim iidN(\mu, \sigma^2); \qquad H_a : \text{not } H_0.$$

The JB test statistic is:

$$JB = n(\frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24}),\tag{10}$$

where

$$\alpha_3 \equiv \frac{n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^3}{s^3} \tag{11}$$

$$\alpha_4 \equiv \frac{n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^4}{s^4} \tag{12}$$

$$s^2 \equiv n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{13}$$

Here, $\bar{y}$ is the sample mean, and s^2, α_3 and α_4 are the second, third, and fourth sample moments about the mean, respectively. The JB statistic has an asymptotic distribution which is $\chi^2_{(2)}$ under the null hypothesis.

The JB test is known to have very good power properties in testing for normality; it is clearly easy to compute; and it is commonly used in the regression context in econometrics. One limitation of the test is that it is designed only for testing for normality, while the ELR test can be applied to test for any types of underlying distribution with some appropriate modification to the moment equations.

3.2 D'Agostino's Test (D)

The D test was originally proposed by D'Agostino (1971). It has been widely used for testing for normality. Suppose y_1, y_2, $\ldots, y_n$ is the data set. $y_{1,n}$, $y_{2,n}$, $\ldots$, $y_{n,n}$ are the ordered observations, where $y_{1,n} \leq y_{2,n} \leq \ldots \leq y_{n,n}$. The D test statistic has the form:

$$D = \frac{T}{n^2 s}, \tag{14}$$

where s is the sample standard deviation, which is the square root of s^2 as defined in the context of the JB test, and $T = \sum_{i=1}^{n} \{i - \frac{n+1}{2}\} y_{i,n}$. If the sample is drawn from a normal distribution, then

$$E(D) = \frac{(n-1)\Gamma(\frac{n}{2} - \frac{1}{2})}{2\sqrt{2n\pi}\Gamma(\frac{n}{2})} \approx (2\sqrt{\pi})^{-1} \approx 0.28209479. \tag{15}$$

The asymptotic standard deviation of the D test statistic is:

$$asd(D) = (\frac{12\sqrt{3} - 37 + 2\pi}{24n\pi})^{\frac{1}{2}} \approx 0.02998598/\sqrt{n}. \qquad (16)$$

The standardized D test statistic is:

$$D^* = \frac{D - E(D)}{asd(D)}, \qquad (17)$$

and the null hypothesis and the alternative for the D test are:

$$H_0 : y_i's \sim iidN(\mu, \ \sigma^2); \qquad H_a : \text{not } H_0.$$

Under the null hypothesis, D^* is asymptotically distributed as $N(0, 1)$. If the sample is drawn from a distribution other than normal, $E(D^*)$ tends to differ from zero. If the underlying distribution has greater than normal kurtosis, then, $E(D*) < 0$. If it has less than normal kurtosis, then, $E(D^*) > 0$. So to guard against both possibilities, the test is a two-sided test.

The percentage points for sample sizes, $n = 30, \ 50, \ 70, \ 100$ are given by D'Agostino (1972). They were constructed using Pearson curves fitted by moments and extensive simulations. The percentile points for larger sample sizes, $n = 150, \ 200, \ 500, \ 1000$, are provided by D'Agostino (1971) and they are based on Cornish-Fisher expansions. These percentile points were calculated and verified by D'Agostino (1972). In our study, instead of simulating critical values, we use these published values. The D test is an omnibus test in the sense of being able to appropriately detect deviations from normality due either to skewness or to kurtosis.

The Shapiro-Wilks (1965) W test for normality is also known to be a relatively powerful test. The W test is based on the ratio of the best linear unbiased estimator of the population standard deviation to the sample variance. Appropriate weights for the ordered sample observations are needed in computing the numerator of the W test statistic and in computing the percentile points of the null distribution of W for small samples. Each sample size requires a new set of appropriate weights. The W test is also an omnibus test. It has power properties that are superior to those of the chi-squared goodness of fit test in many situations. However, the more recent D test has power properties that compare favorably

with the W test (D'Agostino, 1971). Shapiro and Wilks did not extend their test beyond samples of size 50. D'Agostino (1971) commented on the W test that there are a number of indications that it is best not to make such an extension, although subsequently Royston (1982) did extend the W test for normality to large samples. We have chosen not to include the W test in our study as it is known to have power similar to the D test, while being more difficult to implement computationally.

3.3 Pearson's χ^2 Goodness of Fit Test (χ^2)

Pearson's (1900) χ^2 goodness of fit test was the first constructive test in the statistics literature and is a commonly used nonparametric test. It is based on the discrepancies between the observed and expected data frequencies. Consider a sample of independent observations of size n, y_1, y_2, ..., y_n, with a common distribution $F(y, \theta)$ unknown, where θ is the parameter vector. The null hypothesis is:

$$H_0 : F(y, \theta) = F_0(y, \theta),$$

where F_0 is the distribution function of a particular specified distribution.

In our study, we first transform the y_i data to be $x_i = \frac{y_i - \bar{\mu}}{\bar{\sigma}}$, where $\bar{\mu}$ and $\bar{\sigma}$ are the sample mean and sample deviation. Then, we specify F_0 as $N(0, 1)$. The sample of data is classified into k mutually exclusive categories. The number of categories, k, and the boundaries of the categories are determined in advance, independently of the data. Let p_{0i} denote the expected probability of an observation falling in the ith category, np_{0i} denote the expected frequencies, and n_i denote the observed frequencies, where $i = 1, 2, \ldots, k$.

The χ^2 test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_{0i})^2}{np_{0i}} \tag{18}$$

$$= \frac{1}{n}\sum_{i=1}^{k} \frac{n_i^2}{p_{0i}} - n \tag{19}$$

and it has a limiting distribution $\chi^2_{(k-3)}$ if the null hypothesis is true.

The number of mutually exclusive categories k is supposed to be arbitrary and inde-

pendent of the observed data. The asymptotic theory of the χ^2 test is valid no matter how the k categories are determined provided that they are determined without reference to the observations. There are some basic criteria that k should meet, for example $k < n$. Often, an additional restriction is imposed in practice on the choice of k. The resulting intervals should be such that $np_i \geq 5$, for all i. In this case we will denote the test statistic as χ^{2*}. The χ^2 and the χ^{2*} tests may not be applicable when the sample size is very small. Both the ELR test and the χ^2 tests are nonparametric and are applicable when testing for any type of underlying distributions.

4 MONTE CARLO SIMULATIONS AND RESULTS

This section discusses the Monte Carlo simulations applying the empirical likelihood ratio test and the four conventional tests that we have considered for testing for normality. The sampling properties of the tests are provided. In particular, power comparisons of the ELR test and the four other tests are conducted.

4.1 Data Generating Process

The null hypothesis is that the underlying population has a distribution that is $N(\mu, \sigma^2)$. The four alternative distributions that we consider are: Lognormal (LN); $\chi^2_{(2)}$; Student $t_{(5)}$; and the Double Exponential distribution (DE). These distributions cover a range of situations from symmetric, fat-tailed to skewed distributions. The Log-normal and the $\chi^2_{(2)}$ distributions are skewed to the right. The student $t_{(5)}$ is symmetric and fat-tailed. The Double Exponential is a symmetric and long-tailed distribution.

Without loss of generality, all of these distributions are standardized to have mean zero and variance unity. This serves only to fix the true values of the location and scale, possibly both unknown, and does not preclude inferences about those values. This approach is also taken by White and MacDonald (1980), for example.

1. Data for the standardized Lognormal distribution are generated by transforming the standard normal variable $z \sim N(0,1)$ to $y \sim LN(0,1)$:

$$y = \exp{(z)}/2.161197416 - 0.762873978. \tag{20}$$

2. Data for the standardized $\chi^2_{(2)}$ distribution are obtained from two independent standard normal variates: z_1 and z_2:

$$\chi^2_{(2)} = ((z_1^2 + z_2^2) - 2)/2. \tag{21}$$

3. Data for the standardized Student $t_{(5)}$ distribution are obtained by:

$$t_{(5)} = \frac{z\sqrt{3/5}}{\sqrt{\chi^2_{(5)}/5}}, \tag{22}$$

where $z \sim N(0, 1)$, z and $\chi^2_{(5)}$ are independent of each other.

4. Data for the standardized Double Exponential distribution are obtained by:

$$y = \frac{x_1 - x_2}{2\sqrt{2}}, \tag{23}$$

where $x_i \sim \chi^2_{(2)}$, $i = 1, \ 2$ are independent of each other.

Two particular questions are of interest. First, how do the five tests differ in terms of size distortion in finite samples? Second, how do the powers of the tests compare with each other across all of the alternatives, once the size distortion has been taken into account?

4.2 Size Distortion

All of the five tests are asymptotic tests and their sizes in finite samples are unknown. The actual sizes of all of the tests, except the D test, for finite samples are simulated and illustrated in Table 1 in the appendix. The size distortion is the difference between the actual size of the test and the nominal significance level. The size-adjusted critical values are the values that ensure that the actual sizes of a test equal the nominal significance levels based on the asymptotic distribution of the test statistic. Table 1 also provides the size-adjusted critical values for the tests (excluding the D test). The percentile points, *i.e.* the size-adjusted critical values, for the D test, do not appear in the table but are taken from D'Agostino (1971 and 1972), as the accuracy of these values have been verified by other authors.

The true size of the ELR test is quite large for small samples. For example, the actual size is 34.94% when the nominal significance level is 10%, at $n = 30$. The sizes come down quickly and converge to the correct nominal levels as n increases, as would be expected. The size of the JB test is much lower than the respective nominal level for small sample sizes. For example, the actual size is about 4.38% when the nominal significance level is 10% for $n = 30$. The sizes converge to the correct nominal levels when n grows. The size distortion of the χ^2 test is smaller than that of the ELR test for small samples. However, it is worse than that of the ELR test when the sample size grows. In particular, the size distortion does not vanish as the sample size $n \to \infty$. This problem is avoidable if the adjusted chi-square goodness of fit test, the χ^{2*} test, is used. The χ^{2*} test is the χ^2 goodness of fit test adjusted so that the expected frequencies in each category is greater than or equal to five.

The ELR test is an asymptotic test with a limiting distribution of χ^2. The purpose of the Monte Carlo simulation study is to provide the actual distribution for the test statistic in finite samples. The fact that the size distortion of the ELR test is relative large indicates that the approximation of the finite sample distribution in small samples using the asymptotic χ^2 is relatively poor.

Owen (1990) suggested that, for small sample size n, we should replace $\chi^2_{(d)}$ with $\frac{(n-1)d}{(n-d)}$ times $F(d, n-d)$ for a better approximation. This would be very effective to reduce the size of the ELR test. For example, at $n = 30$, the following are the critical values of $\chi^2_{(2)}$ and $(n-1)d/(n-d)F(d, n-d)$, where $d = 2$ and $n = 30$:

α	$\chi^2_{(2)}$	$\frac{29 \times 2}{28} F(2, 28)$
10%	4.6052	5.1786
5%	5.9915	6.9354
2%	7.8240	9.3484
1%	9.2100	11.3141

From the table above, we can see that the critical values under the adjusted F distribution are larger than that of $\chi^2_{(2)}$. Using the critical values of F distribution, the size distortion of the ELR test will be smaller. However, we did not explore this point further in this study.

4.3 Power Comparisons

Tables 2 to 5 give the power comparisons of the five tests for normality across certain alternatives. It is recalled that the null distribution for all the tests is $N(\mu, \sigma^2)$, where the true values for the parameters are $\mu = 0$ and $\sigma^2 = 1$. In order to conduct the power comparisons, we need to use the same standards for the different tests. The size-adjusted critical values are used for this purpose. We have seen that the five tests have different actual sizes in finite samples. By using the size-adjusted critical values, we are able to compare the power of the five tests at the same actual significance levels, 10%, 5%, 2%, and 1%.

Table 2 gives the results when the alternative distribution is Lognormal. The ELR test has the highest power among the tests for significance levels of 5%, 2%, 1% . The power of the JB is in the same range as that of the ELR test, especially for small sample sizes. For example, the power of the ELR test is 93.76% for $n = 30$ and an actual level of 5%, while the JB test has a power of 92.77%. Both the ELR and the JB tests are very powerful for this skewed alternative distribution. The powers of the two tests converge to 100% at $n = 100$. The power of the D test is inferior to that of the ELR and the JB tests for small sample sizes. The χ^2 and the χ^{2*} tests are not applicable for some of the smaller sample sizes. The powers of all of the tests converge to 100% as n grows, though more slowly for the χ^{2*} test than for the other ones.

Table 3 gives the results when the alternative distribution is $\chi^2_{(2)}$. The ELR test is the most powerful one among all of the five tests considered for the various sample sizes. The power is 93.07% when $n = 30$, compared with 92.33% for the JB test, when the significance level is 10%. It's power converges to 100% faster than for any of the other tests and it reaches 100% at $n = 50$. Again, the power of the D test and χ^{2*} tests are lower than those of the ELR and the JB tests.

When the alternative distributions are symmetric, as for the Student $t(5)$ and the Double Exponential distributions, all of the tests have quite low power. It is difficult for any test to detect this forms of departure from the normality. Tables 4 and 5 illustrate these case. The JB test in this situation is the most powerful test among those considered. With a true significance level of 10%, its power is 37.7% and 46.5% against the $St(5)$ and the DE distributions, respectively, when $n = 30$; while the ELR test has a power of 10.6% and 13.2%, respectively, at a true significance level of 10%. The power of the ELR is even lower than

that of the D test for small sample sizes. All of the three tests, the ELR, the JB, and the D, have higher power when the sample size reaches $n = 200$. The powers of the three tests are about 100% at $n = 500$. This indicates that the power of the ELR test improves quickly as the sample size increases from $n = 30$ to $n = 500$, even though it starts at a low value for the small samples for the symmetric alternative distributions.

The relatively good power properties of the ELR test result from the ability of the EL method to incorporate the most information available. For instance, in the context of testing for normality, using the first four moment equations, the EL method is able to take into account the information of the sample mean, the variance, the skewness, and the kurtosis. The JB test has the same advantage as the ELR test with four moment equations since the design of the JB test incorporates the standardized third and fourth sample moments. Moreover, the EL method naturally utilizes the likelihood function which may lead to some efficiency gain. Therefore, the ELR test exhibits some attractive features in the application of testing for normality.

To provide some guidance for practitioners in taking the advantage of the good power properties of the ELR test in finite samples, we would suggest that one could use the size-adjusted critical values that we have provided in this study when the values of one's parameters match the values that we have considered. In addition, it would be worthwhile to devote some future effort to the provision of the size-adjusted critical values for a more extensive range of sample sizes.

4.4   Invariance of the ELR Test

In this section, we show that the ELR test is robust to changes in the functional form of the unbiased moment equations. Instead of using the first four *raw* moment equations, we consider the first four *standardized central* moment equations. The data are distributed i.i.d. $N(\mu,\ \sigma^2)$. We standardize the data so that, theoretically, the data will be i.i.d. $N(0,1)$.

The transformation from the raw moments to the standardized central moments is a smooth and nonlinear transformation in the parameter space. The raw and the standardized central moment equations have the form as follows.

Raw moment conditions	Standardized central moment conditions

$$E(y - \mu) = 0 \qquad\qquad E(\frac{y-\mu}{\sigma}) = 0$$

$$E(y^2 - (\mu^2 + \sigma^2)) = 0 \qquad\qquad E(\frac{(y-\mu)^2}{\sigma^2} - 1) = 0$$

$$E(y^3 - (\mu^3 + 3\sigma^2\mu)) = 0 \qquad\qquad E(\frac{(y-\mu)^3}{\sigma^3}) = 0$$

$$E(y^4 - (\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4)) = 0 \qquad E(\frac{(y-\mu)^4}{\sigma^4} - 3) = 0$$

Using the standardized central moment conditions places the ELR test on the same basis as the JB test. The JB test uses the skewness and the kurtosis coefficients of the data which are in the form of standardized central moments.

Table 6 gives the actual sizes and size-adjusted critical values of the ELR test, and Table 7 gives the power of the ELR test using the first four standardized moment conditions about the mean with two unknown parameters. The null distribution is again $N(\mu, \sigma^2)$. From Table 6, it is easy to see that the range and the pattern of the size distortion of the ELR test using standardized moment conditions about the mean are the same as the ones in the nonstandardized case. The size is approximately 33% at $n = 30$ and converges to the nominal level of 10% as n increases. In Table 7, the power of the ELR test is also in the same relative range as it is in the nonstandardized case. It is slightly higher at the lower actual significance levels for asymmetric alternatives and lower for symmetric ones, relatively.

Owen (2001, p. 50) discusses the transformation invariance of EL. This relates to the fact that empirical likelihood confidence regions are invariant under one to one parameter transformations and are also invariant under one to one invertible transformations of data. The empirical evidence of the invariance of the EL method that we have found in this study is that the distribution of the ELR test statistic in finite samples is invariant with respect to the functional form of the moment equations. There is an implicit connection between these two types of invariance of the EL method. The confidence regions and the power of the ELR test are two sides of the same coin. It would be worthwhile to explore the theoretical underpinnings of these findings in more detail. Indeed, a deeper understanding of this issue may also assist in preparing practical guidance for practitioners regarding size adjustment when applying the various EL-based tests.

4.5 The ELR Test with Increased Number of Moment Equations

Table 8 gives the sizes and the size-adjusted critical values for the ELR test using five moment equations, rather than four. Table 9 contains the power comparisons of the ELR tests, the ELR_4 and ELR_5 with the first four and the first five moment equations, and the JB test for small and medium sample sizes. In Tables 2 to 5, we have seen that the ELR test has very good power properties for large sample sizes over various types of alternative distributions and it is the most powerful test (among those considered) for small sample sizes against skewed alternative distributions. However, it is inferior to the JB test with respect to symmetric alternative distributions such as the $St(5)$ and the Double Exponential distributions. The purpose of this section is to see if the power of the ELR test can be improved by using an increased number of functionally independent moment equations.

Mittelhamer *et al.* (2000) conjectured that the power of the ELR test increases with the number of moment conditions. Hopefully the power of the ELR test can be improved in the case of symmetric alternative distributions. However, we should be aware of the following three issues. First, for the $St(5)$ distribution, the integer-order moments exist only up to four at most. Second, there is a potential problem of infeasibility in finite samples in computational practice of the EL method. Given the constraints on the p_i's, a set of over-identified moment equations may not provide a valid solution for θ. The probability of this infeasibility is small. However, when we increase the number of correctly specified moment equations, this potential may increase. That is, the probability of the potential problem may increase as the number of the moment equations increases. Third, the increased degree of over-identification may cause an increase in the computing time for the method.

The null distribution is still the same, namely $H_0 : N(\mu, \sigma^2)$. We illustrate that for small samples, *i.e.* $n = 30$, the power of the ELR test does increase significantly, especially at small significance levels such as 1% and 2%. Table 3.9 shows that the power of the ELR_5 test has increased up to 17% in small sample sizes against skewed distributions, with especially large increases over the low significance levels for each alternative distribution. For example, for the alternative $\chi^2_{(2)}$ distribution, the increment is approximately 17% at the actual significance level of 1% and at $n = 30$, which is quite significant. For the symmetric alternative distribution, the Double Exponential, at $n = 30$ and $\alpha = 1\%$, the increment is about 3%. The results overall are consistent with the conjecture in Mittelhammer *et al.* (2000). Unfortunately, the power of the ELR_5 test is still lower than that of the JB test in

small samples against the alternative distribution of the Double Exponential.

5   COMPUTING ISSUES

The nonlinear equation solver, the Eqsolve algorithm, works very well in the application considered in this study. Each draw from the underlying population is valid in the sense that the sample of data is able to work well for all of the tests: the ELR test, the JB test, the D test, the χ^2 test, and the χ^{2*} test. If a sample draw from the underlying distribution could not provide a valid numerical solution for the ELR test either because the estimated $\hat{p}_i$ is not in the $(0,\ 1)$ range or because the iteration could not converge to provide a valid solution, then, the sample would be thrown away. There are very few data sets being thrown away in small samples and no data sets being thrown away when the sample size is greater than fifty. That is, there is no selection bias when using the EL approach in this application.

The computing time in testing for normality is very reasonable. For example, it takes approximately one minute of processing time on a Pentium 4 2.0 GHZ personal computer to conduct a simulation experiment with $10,000$ replications to determine just the empirical size of the ELR test when $n = 30$. It takes about three minutes for $10,000$ replications and all of the five tests when $n = 30$.

6   SUMMARY AND CONCLUSIONS

In this paper, we have developed an empirical likelihood ratio test for the problem of testing for normality. Monte Carlo simulations are used to provide the actual sizes and the size-adjusted critical values for the ELR test and for four other tests. These critical values are used in computing the power of each test and conducting power comparisons between the tests. The empirical results provide evidence that the ELR test is a relatively powerful test. It is the most powerful test over asymmetric alternative distributions among all of the five tests considered here. For the symmetric alternative distributions, the power of the ELR test is slightly inferior to that of the JB test. The power of the ELR test can be improved by increasing the number of moment equations we use. The ELR test is invariant to the form

17

of the moment equations. Overall, the ELR test for normality has good power properties, and it is quite easily implemented.

ACKNOWLEDGEMENTS

APPENDIX: TABLES IN NORMALITY TEST

Table 1: Size and Size-adjusted Critical Values for the Four Tests: ELR, JB, χ^2, and χ^{2*}

m :	10,000					$H_0 : N(\mu,\ \sigma^2)$					
n :	30	50	70	100	200	250	500	1,000	2,000	5,000	10,000
ELR test at nominal levels:											
10%	0.3494	0.3215	0.2910	0.2668	0.2142	0.2043	0.1650	0.1423	0.1235	0.1129	0.1051
5%	0.2490	0.2203	0.2044	0.1825	0.1438	0.1325	0.1030	0.0819	0.0695	0.0562	0.0554
2%	0.1534	0.1402	0.1251	0.1152	0.0843	0.0738	0.0572	0.0411	0.0297	0.0267	0.0209
1%	0.1049	0.0988	0.0866	0.0792	0.0562	0.0494	0.0360	0.0255	0.0149	0.0141	0.0111
Size-adjusted Critical Values:											
10%	9.3778	9.1638	8.7144	8.3184	7.2353	6.8555	6.0854	5.4348	5.1198	4.8842	4.7309
5%	12.1792	11.6566	11.1896	10.9248	9.5447	9.1718	8.2100	7.2881	6.6690	6.2568	6.1587
2%	15.4973	14.8743	14.4297	14.2330	12.7728	12.3614	11.2863	9.8560	8.6991	8.5000	7.9667
1%	17.7115	17.4968	17.1119	16.8595	15.5300	15.7825	13.6090	11.7415	10.4135	9.8666	9.5787
JB test at nominal levels:											
10%	0.0438	0.0543	0.0569	0.0633	0.0785	0.083	0.0865	0.0881	0.0937	0.0976	0.0984
5%	0.0294	0.0353	0.0366	0.0390	0.0472	0.0467	0.0475	0.0455	0.0455	0.0484	0.0491
2%	0.0192	0.0229	0.0229	0.0246	0.0268	0.0278	0.0221	0.0231	0.0196	0.0189	0.0186
1%	0.0147	0.0165	0.0176	0.0183	0.0191	0.0194	0.0147	0.0149	0.0110	0.0102	0.0106
Size-adjusted Critical Values:											
10%	2.7415	3.1072	3.3437	3.5628	4.0822	4.164	4.3016	4.3573	4.5162	4.5533	4.5620
5%	4.2229	4.8305	4.9965	5.2076	5.7969	5.8276	5.8854	5.7599	5.7988	5.9441	5.9562
2%	7.6184	8.3630	8.7360	8.8227	8.9930	9.1034	8.0600	8.2455	7.7146	7.7384	7.6884
1%	11.2731	12.4052	12.4829	12.6074	11.8813	12.5295	10.7754	10.4652	9.3900	9.2140	9.3332
χ^2 **goodness of fit test at nominal levels:**											
10%	0.1320	0.1115	0.1022	0.1071	0.1239	0.1324	0.1357	0.1366	0.1398	0.1361	0.1267
5%	0.0605	0.0596	0.0537	0.0654	0.0768	0.0837	0.0851	0.0833	0.0863	0.0866	0.0788
2%	0.0214	0.0262	0.0267	0.0355	0.0461	0.0514	0.0479	0.0487	0.0510	0.0504	0.0454
1%	0.0105	0.0149	0.0162	0.0241	0.0344	0.0371	0.0350	0.0352	0.0359	0.0361	0.0322
Size-adjusted Critical Values:											
10%	5.0784	9.3466	12.9873	19.0312	37.3835	45.7889	55.8576	57.7062	58.8775	59.5980	59.4830
5%	6.2818	11.3592	15.2381	22.3365	42.1121	51.1546	61.9765	63.2299	64.3628	65.5994	65.5305
2%	7.8279	13.6176	18.4865	26.2892	50.2357	59.7278	71.3387	73.1178	73.5503	75.5200	74.6156
1%	9.1078	15.4547	20.8135	30.9140	63.5271	74.2101	85.9508	88.8786	90.9052	95.0253	87.8157
χ^{2*} **goodness of fit test at nominal levels:**											
10%	–	0.1693	0.1316	0.1187	0.1079	0.1104	0.1082	0.1045	0.1064	0.1012	0.0991
5%	–	0.0888	0.0671	0.0616	0.0568	0.0552	0.0541	0.0538	0.0547	0.0519	0.0501
2%	–	0.0410	0.0279	0.0267	0.0230	0.0231	0.0209	0.0213	0.0210	0.0224	0.0207
1%	–	0.0212	0.0137	0.0132	0.0117	0.0115	0.0096	0.0115	0.0105	0.0113	0.0111
Size-adjusted Critical Values:											
10%	–	7.4047	9.1804	12.4743	22.8229	27.7303	38.1887	43.0001	46.4041	49.2212	50.7836
5%	–	9.3701	11.1657	14.6779	25.5606	30.8296	41.6029	46.7956	50.3262	53.4114	55.1265
2%	–	11.6226	13.6931	17.5926	29.0460	34.3290	45.7909	51.4873	54.6466	58.5462	60.1098
1%	–	13.2294	15.4926	19.5909	31.2562	37.3967	48.6595	55.1539	58.1896	62.3658	63.3077

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu,\ \sigma^2)' = (0,\ 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 2: Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			$H_a : Lognormal(0,1)$			
n	30	50	70	100	150	200	250
ELR test:							
10%	0.9811	0.9999	1	1	1	1	1
5%	0.9376	0.9991	1	1	1	1	1
2%	0.8368	0.9966	0.9999	1	1	1	1
1%	0.7212	0.9897	0.9997	1	1	1	1
JB test:							
10%	0.9854	0.9995	1	1	1	1	1
5%	0.9277	0.9970	1	1	1	1	1
2%	0.8081	0.9642	0.9976	1	1	1	1
1%	0.7037	0.9088	0.9878	0.9998	1	1	1
D test:							
10%	0.8905	0.9761	0.9963	0.9999	1	1	1
5%	0.8374	0.9592	0.9931	0.9996	1	1	1
2%	0.7579	0.9300	0.9852	0.9990	1	1	1
1%	0.6982	0.9016	0.9771	0.9979	1	1	1
χ^2 *test:*							
10%	–	–	0.8887	0.9662	0.9984	0.9997	1
5%	–	–	0.8434	0.9498	0.9973	0.9997	1
2%	–	–	0.7809	0.9252	0.9938	0.9994	0.9998
1%	–	–	0.7072	0.8957	0.9894	0.9980	0.9995
χ^{2*} *test:*							
10%	–	–	–	0.8991	0.9962	0.9997	1
5%	–	–	–	0.8236	0.9909	0.9997	1
2%	–	–	–	0.7041	0.9784	0.9990	1
1%	–	–	–	0.6178	0.9642	0.9984	1

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 3: Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			$H_a : \chi^2_{(2)}(0,1)$			
n	30	50	70	100	150	200	250
ELR test:							
10%	0.9307	0.9975	1	1	1	1	1
5%	0.8295	0.9929	0.9998	1	1	1	1
2%	0.6411	0.9734	0.9986	1	1	1	1
1%	0.4767	0.9383	0.9970	1	1	1	1
JB test:							
10%	0.9233	0.9974	0.9999	1	1	1	1
5%	0.7475	0.9661	0.9975	1	1	1	1
2%	0.5535	0.8240	0.9599	0.9990	1	1	1
1%	0.4331	0.6909	0.8824	0.9867	1	1	1
D test:							
10%	0.6536	0.8621	0.9467	0.9888	0.9995	0.9999	1
5%	0.5533	0.7941	0.9150	0.9775	0.9984	0.9997	1
2%	0.4480	0.6997	0.8561	0.9579	0.9957	0.9995	0.9999
1%	0.3819	0.6265	0.8021	0.9352	0.9929	0.9991	0.9999
χ^2 *test:*							
10%	–	0.6111	0.7521	0.8922	0.9830	0.9974	0.9998
5%	–	0.4812	0.6569	0.8364	0.9748	0.9960	0.9996
2%	–	0.3615	0.5310	0.7573	0.9497	0.9914	0.9988
1%	–	0.2949	0.4580	0.6649	0.9183	0.9771	0.9959
χ^{2*} *test:*							
10%	–	0.2669	0.4057	0.6569	0.9411	0.9927	0.9989
5%	–	0.1349	0.2537	0.5204	0.8874	0.9831	0.9975
2%	–	0.0637	0.1390	0.3640	0.8088	0.9659	0.9959
1%	–	0.0390	0.0927	0.2743	0.7345	0.9450	0.9925

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 4: Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000				H_a : Student $t_{(5)}(0,1)$				
n	30	50	70	100	150	200	250	500	1,000
ELR test:									
10%	0.1060	0.1635	0.2394	0.3746	0.5818	0.7506	0.8502	0.9924	1
5%	0.0483	0.0874	0.1375	0.2464	0.4307	0.6346	0.7675	0.9829	1
2%	0.0208	0.0393	0.0659	0.1347	0.2752	0.4831	0.6330	0.9580	1
1%	0.0097	0.0196	0.0364	0.0786	0.1876	0.3704	0.4858	0.9303	0.9997
JB test:									
10%	0.3767	0.5033	0.6050	0.7182	0.8262	0.9037	0.9419	0.9972	1
5%	0.2913	0.4167	0.5212	0.6501	0.7801	0.8678	0.9183	0.9947	1
2%	0.2071	0.3222	0.4198	0.5454	0.6963	0.8017	0.8655	0.9901	1
1%	0.1575	0.2605	0.3501	0.4686	0.6257	0.7532	0.8191	0.9822	1
D test:									
10%	0.3157	0.4594	0.5745	0.7099	0.8369	0.9121	0.9563	0.9981	1
5%	0.2342	0.3663	0.4868	0.6314	0.7762	0.8710	0.929	0.9963	1
2%	0.1615	0.2753	0.3852	0.5330	0.6928	0.8080	0.8894	0.9906	1
1%	0.1251	0.2267	0.3224	0.4656	0.6334	0.7574	0.8541	0.9853	1
χ^2 *test:*									
10%	0.1885	0.2863	0.3758	0.4695	0.5853	0.6889	0.7463	0.9494	0.9990
5%	0.1220	0.2133	0.3040	0.3932	0.5230	0.6270	0.6879	0.9268	0.9983
2%	0.0774	0.1621	0.2396	0.3317	0.4465	0.5474	0.6225	0.8904	0.9956
1%	0.0604	0.1393	0.2010	0.2854	0.4033	0.4711	0.5461	0.8365	0.9880
χ^{2*} *test:*									
10%	–	0.1407	0.1671	0.1864	0.2148	0.2282	0.2308	0.3559	0.8340
5%	–	0.0735	0.0985	0.1089	0.1271	0.1460	0.1462	0.2532	0.7535
2%	–	0.0343	0.0477	0.0562	0.0697	0.0785	0.0844	0.1577	0.6365
1%	–	0.0229	0.0265	0.0367	0.0441	0.0531	0.0509	0.1121	0.5374

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 5: Power Comparison of the ELR Test with JB, D, χ^2, χ^{2*} Tests

m	10,000			$H_a : DoubleExponential(0,1)$					
n	30	50	70	100	150	200	250	500	1,000
ELR test:									
10%	0.1318	0.2163	0.3484	0.5824	0.8286	0.9486	0.9843	1	1
5%	0.0720	0.1277	0.2222	0.4207	0.7102	0.8974	0.9650	0.9996	1
2%	0.0337	0.0665	0.1139	0.2636	0.5260	0.7986	0.9157	0.9992	1
1%	0.0188	0.0397	0.0635	0.1670	0.4147	0.6953	0.8371	0.9986	1
JB test:									
10%	0.4648	0.6266	0.7359	0.8541	0.9433	0.9805	0.9938	0.9998	1
5%	0.3620	0.5281	0.6541	0.7888	0.9126	0.9660	0.9895	0.9997	1
2%	0.2562	0.4059	0.5162	0.6744	0.8470	0.9326	0.9719	0.9994	1
1%	0.1866	0.3196	0.4279	0.5800	0.7764	0.8928	0.9458	0.9993	1
D test:									
10%	0.4540	0.6648	0.7972	0.9079	0.9773	0.9948	0.9990	1	1
5%	0.3418	0.5551	0.7128	0.8603	0.9612	0.9897	0.9977	1	1
2%	0.2367	0.4329	0.5968	0.7796	0.9258	0.9772	0.9945	1	1
1%	0.1784	0.3546	0.5125	0.7176	0.8923	0.9631	0.9906	1	1
χ^2 *test:*									
10%	0.2851	0.4421	0.5694	0.6980	0.8390	0.9124	0.9478	0.9993	1
5%	0.1956	0.3373	0.4747	0.5995	0.7688	0.8596	0.9154	0.9980	1
2%	0.1260	0.2504	0.3628	0.5002	0.6564	0.7621	0.8493	0.9925	1
1%	0.0950	0.2039	0.3040	0.4080	0.5739	0.6155	0.7289	0.9766	1
χ^{2*} *test:*									
10%	–	0.2588	0.3601	0.4755	0.6180	0.7253	0.7929	0.9816	1
5%	–	0.1483	0.2341	0.3432	0.4813	0.6094	0.6872	0.9610	1
2%	–	0.0758	0.1311	0.2070	0.3408	0.4614	0.5588	0.9218	0.9999
1%	–	0.0465	0.0832	0.1442	0.2464	0.3689	0.4488	0.8856	0.9999

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$. The χ^2 tests may not be applicable with some small sample sizes.

Table 6: Size and Size-adjusted Critical Values of the ELR Test Using Standardized Central Moment Equations

m	10,000			$H_0 : N(\mu, \sigma^2)$							
n	30	50	70	100	150	200	500	1,000	2,000	5,000	10,000
ELR test at nominal levels:											
10%	0.3321	0.2852	0.2579	0.2399	0.2388	0.1904	0.1497	0.1273	0.1153	0.1113	0.1057
5%	0.2338	0.2004	0.1792	0.1625	0.1574	0.1162	0.0880	0.0729	0.0616	0.0563	0.0526
2%	0.1666	0.1393	0.1239	0.1162	0.1092	0.0789	0.0528	0.0411	0.0347	0.0284	0.0260
1%	0.1054	0.0855	0.0767	0.0711	0.0657	0.0462	0.0297	0.0205	0.0156	0.0111	0.0104
Size and Size-adjusted critical values:											
10%	9.43	8.55	8.17	7.95	7.71	6.48	5.64	5.15	4.87	4.82	4.72
5%	12.02	11.21	11.09	10.51	10.20	8.95	7.56	6.88	6.53	6.19	6.09
2%	15.71	14.39	14.74	14.15	13.49	12.19	10.27	9.28	8.67	8.04	7.87
1%	18.44	16.72	17.46	16.95	16.62	14.45	12.46	11.31	10.05	9.46	9.27

Notes to table: The data is standardized to be $x_i = (y_i - \mu)/\sigma$, for $i = 1, 2, \ldots, n$. The true value of $\theta = (\mu, \sigma^2)'$ is $(0, 1)'$. m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

Table 7: Power of The ELR Test Using Standardized Central Moment Equations

m	10,000			$H_0 : N(\mu, \sigma^2)$				
n	30	50	70	100	150	200	250	500
$H_a : Lognormal$								
10%	0.9787	1	1					
5%	0.9429	0.9991	1					
2%	0.8454	0.9958	1					
1%	0.7308	0.9916	0.9998					
$H_a : \chi^2_{(2)}$								
10%	0.9440	0.9984	1					
5%	0.8689	0.9945	1					
2%	0.7061	0.9802	0.9992					
1%	0.5634	0.9600	0.9975					
$H_a : Student\ t_{(5)}$								
10%	0.0759	0.1163	0.1850	0.3147	0.5781	0.7596	0.8917	0.9953
5%	0.0398	0.0593	0.0835	0.1862	0.4336	0.6283	0.8254	0.9874
2%	0.0168	0.0269	0.0342	0.0782	0.2814	0.4646	0.7203	0.9679
1%	0.0090	0.0157	0.0194	0.0409	0.1793	0.3628	0.6241	0.9442
$H_a : Double\ Exponential$								
10%	0.0876	0.1939	0.3501	0.5758	0.8399	0.9525	0.9898	0.9999
5%	0.0437	0.0996	0.1870	0.4116	0.7295	0.8999	0.9776	0.9999
2%	0.0188	0.0431	0.0843	0.2252	0.5703	0.8026	0.9450	0.9997
1%	0.0116	0.0255	0.0474	0.1273	0.4194	0.7147	0.9072	0.9993

Notes to table: The data is standardized to be $x_i = (y_i - \mu)/\sigma$, for $i = 1, 2, \ldots, n$. m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

Table 8: Size of the ELR Test with Five Moment Equations

m	10,000			$H_0 : N(\mu, \sigma^2)$				
n	30	50	70	100	150	200	250	500
ELR test at nominal levels:								
10%	0.4301	0.4216	0.3993	0.3817	0.3478	0.3283	0.2999	0.2546
5%	0.3349	0.3259	0.3099	0.2877	0.2655	0.2437	0.2156	0.1797
2%	0.2325	0.2345	0.2172	0.1976	0.1820	0.1631	0.1455	0.1139
1%	0.1786	0.1750	0.1641	0.1485	0.1383	0.1246	0.1076	0.0817
Size-adjusted Critical Values:								
10%	14.3394	14.2768	14.1680	13.7818	13.2607	12.6498	11.7367	10.4588
5%	18.3187	17.8126	17.6145	17.3194	17.0302	16.5965	15.4773	13.5068
2%	24.2292	22.5592	22.6917	22.1746	21.8368	21.8703	20.2833	18.1422
1%	30.3333	26.8072	25.9431	25.7542	25.2727	25.6406	24.2938	22.2712

Notes to table: m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

Table 9: Power of the ELR Test Using Five Moment Equations

m	10,000		$H_0 : N(\mu, \sigma^2)$					
n	30	50			30	50		

	H_a : *Lognormal*				$H_a : \chi^2_{(2)}$			
ELR_4 *test:*								
10%	0.9811	0.9999			0.9307	0.9975		
5%	0.9376	0.9991			0.8295	0.9929		
2%	0.8368	0.9966			0.6411	0.9734		
1%	0.7212	0.9897			0.4767	0.9383		
ELR_5 *test:*								
10%	0.9874	0.9965			0.9751	0.9997		
5%	0.9642	0.9944			0.9352	0.9984		
2%	0.9006	0.9868			0.8114	0.9950		
1%	0.7810	0.9770			0.6465	0.9874		
JB test:								
10%	0.9854	0.9995			0.9233	0.9974		
5%	0.9277	0.9970			0.7475	0.9661		
2%	0.8081	0.9642			0.5535	0.8240		
1%	0.7037	0.9088			0.4331	0.6909		

n	30	50	70	100	150	200	250	500
			H_a : *Double Exponential*					
ELR_4 *test:*								
10%	0.1318	0.2163	0.3484	0.5824	0.8286	0.9486	0.9843	1
5%	0.0720	0.1277	0.2222	0.4207	0.7102	0.8974	0.9650	0.9996
2%	0.0337	0.0665	0.1139	0.2636	0.5260	0.7986	0.9157	0.9992
1%	0.0188	0.0397	0.0635	0.1670	0.4147	0.6953	0.8371	0.9986
ELR_5 *test:*								
10%	0.1692	0.1659	0.2132	0.3582	0.6445	0.8398	0.9389	0.9998
5%	0.1025	0.0992	0.1219	0.2130	0.4672	0.7030	0.8599	0.9996
2%	0.0525	0.0513	0.0564	0.0956	0.2655	0.4887	0.7169	0.9959
1%	0.0303	0.0317	0.0352	0.0533	0.1640	0.3475	0.5824	0.9886
JB test:								
10%	0.4648	0.6266	0.7359	0.8541	0.9433	0.9805	0.9938	0.9998
5%	0.3620	0.5281	0.6541	0.7888	0.9126	0.9660	0.9895	0.9997
2%	0.2562	0.4059	0.5162	0.6744	0.8470	0.9326	0.9719	0.9994
1%	0.1866	0.3196	0.4279	0.5800	0.7764	0.8928	0.9458	0.9993

Notes to table: ELR_4 and ELR_5 are the ELR test with four and five moment equations. The degrees of freedom of the ELR_5 test is 3. The alternative of student $t_{(}5)$ is not applicable. m and n are the number of replications and the sample size. The true values of the parameters $(\mu, \sigma^2)' = (0, 1)'$.

REFERENCES

Bera, A. and Y. Bilias (2002), "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis", *Journal of Econometrics*, 107, 51 - 86.

Bera, A. and C. Jarque (1980), "Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals" *Economics Letters*, 6, 255-259.

Bera, A. and C. Jarque (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

DiCiccio, T., P. Hall, and J. Romano (1989), "Comparison of Parametric and Empirical Likelihood Functions", *Biometrika*, 76, 465 - 476.

DiCiccio, T. J. and J. P. Romana (1989), "On Adjustments Based on the Signed of the Empirical Likelihood Ratio Statistic", *Biometrica*, 76, 447-456.

DiCiccio, T., P. Hall, and J. Romano (1991), "Empirical Likelihood is Bartlett-Correctable", *Annals of Statistics*, 19, 1053 - 1061.

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

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029 - 1054.

Imben, G. W., R. H. Spady, and P. Johnson (1998), "Information Theoretic Approaches To Inference in Moment Condition Models", *Econometrica*, 66, 333 - 357.

Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes", *Annals of Statistics*, 25, 2084 - 2102.

Kendall, M. and A. Stuart (1963, 1969, and 1977), *The Advanced Theory of Statistics*,

Charles Griffin & Company Limited, London.

Mittelhammer, R., G. Judge, and D. Miller (2000), *Econometric Foundations*, Cambridge University Press, Cambridge.

Mittelhammer, R., G. Judge, and R. Schoenberg (2003), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods", *Working Paper*, Washington State University, University of California, Berkeley, and Aptech Systems, Inc.

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional", *Biometrika*, 75, 237 - 249.

Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Region", *The Annals of Statistics*, 18, 90 - 120.

Owen, A. B. (1991), "Empirical Likelihood for Linear Models", *The Annals of Statistics*, 19, 1725 - 1747.

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

Pearson, K. (1894), "Contribution to the Mathematical Theory of Evolution", *Philosophical Transactions of the Royal Society of London*, Series A 186, 343 - 414.

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.

Pearson, K. (1902), "On the Systematic Fitting of Curves to Observations and Measurements, Parts I and II", *Biometrika*, 1, 265 - 303; 2, 1 - 23.

Pearson, K. (1936), "Method of Moments and Method of Maximum Likelihood", *Biometrika*, 28, 34 - 59.

Qin, J. (1991), "Likelihood and Empirical Likelihood Ratio Confidence Intervals in Two Sample Semi-parametric Models", *Technical Report Series University of Waterloo*, Stat-91-6.

Qin, J. (1993), "Empirical Likelihood in Biased Sample Problems", *Annals of Statistics*, 21,

1182 - 1196.

Qin, J. and J. Lawless (1991), "Empirical Likelihood and General Estimating Equations I", *Technical Report Series University of Waterloo*, Stat-91-10.

Qin, J. and J. Lawless (1994), "Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300 - 325.

Qin, J. and J. Lawless (1995), "Estimating Equations, Empirical Likelihood and Constraints on Parameters", *Canadian Journal of Statistics*, 23, 145 - 159.

Royston, J. P. (1982), "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples", *Applied Statistics*, 31, 115 - 124.

Shapiro, S. S. and M. B. Wilk (1965), "An Analysis of Variance Test for Normality (Complete Samples)", *Biometrika*, 52, 591 - 611.

Welch, B. L. (1947), "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved", *Biometrika*, 34, 28 - 35.

White, H. and G. MacDonald (1980), "Some Large-Sample Tests for Nonnormality in the Linear Regression Model", *Journal of the American Statistical Association*, 75, 16 - 31.

Zaman, A. (1996), *Statistical Foundations for Econometric Techniques*, Academic Press, New York.