Fisher Information Test of Normality

by

Yew-Haur Lee

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Approved by

Dr George R. Terrell, Chairman

Dr Clint W. Coakley

Dr Klaus Hinkelmann

Dr Eric P. Smith

Dr Keying Ye

September 3, 1998

Blacksburg, Virginia

Keywords: Normality testing, Fisher information for location, Power study, Residuals, Nonparametric density estimation, Calculus of variation

Fisher Information Test of Normality

by

Yew-Haur Lee

(ABSTRACT)

An extremal property of normal distributions is that they have the smallest Fisher Information for location among all distributions with the same variance. A new test of normality proposed by Terrell (1995) utilizes the above property by finding that density of maximum likelihood constrained on having the expected Fisher Information under normality based on the sample variance. The test statistic is then constructed as a ratio of the resulting likelihood against that of normality.

Since the asymptotic distribution of this test statistic is not available, the critical values for n = 3 to 200 have been obtained by simulation and smoothed using polynomials. An extensive power study shows that the test has superior power against distributions that are symmetric and leptokurtic (long-tailed). Another advantage of the test over existing ones is the direct depiction of any deviation from normality in the form of a density estimate. This is evident when the test is applied to several real data sets.

Testing of normality in residuals is also investigated. Various approaches in dealing with residuals being possibly heteroscedastic and correlated suffer from a loss of power. The approach with the fewest undesirable features is to use the Ordinary Least Squares (OLS) residuals in place of independent observations. From simulations, it is shown that one has to be careful about the levels of the normality tests and also in generalizing the results.

Acknowledgements

I would like to extend my sincere gratitude and appreciation to Dr George R. Terrell for his guidance throughout this period of research. He was always there when I needed his advice. I am indeed fortunate to have this opportunity to learn and work with him. I would also like to thank Dr Klaus Hinkelmann, Dr Eric P. Smith, Dr Clint W. Coakley, and Dr Keying Ye for their time in carefully reading my dissertation and for the many helpful suggestions and comments that greatly improve my dissertation. I would also like to extend my thanks to the Statistics Department for the invaluable learning experience that I received.

I would also like to thank Michele Marini and Bill Sydor for resolving the many computing obstacles that I encountered. I would also like to express my thanks to the many friends that I have come to know for their encouragement and companionship.

I would also like to express special thanks to my parents, my sisters and Poh Ling, for their love, encouragement and support all this while. This would not be possible without sacrifices they made so that I could pursue my goal.

Finally, I would like to thank God for his grace, mercy and strength that has sustained me throughout this time of my life.

Table of Contents

List of Tables

Table of Figures

Chapter 1 Introduction and Motivation

The problem of normality testing is well known and has generated plenty of attention from researchers, see Mardia (1980) and D'Agostino and Stephens (1986). This is because a lot of classical optimal procedures were developed based on the normality assumption. However, researchers soon realized that this assumption was not always satisfied.

Three approaches can be taken to deal with non-normality of data. The first approach is transforming the data to normality so that the classical procedures could still be used. The second approach is the use of nonparametric procedures. The third is to use robust procedures that are less sensitive to deviation from normality, especially tail behavior. Each of the three comes with strengths and weaknesses; there is no consensus on which is the best approach.

The role of normality testing is not just to see if the data are well approximated by the normal distribution; but also to provide information on the deviation from normality. This information would then guide the researchers to the best approach to dealing with the non-normality of their data.

1.1 Statement of the Problem

In this dissertation, it will be assumed that one is testing normality because the user wishes to fit a location model. Suppose the data collected $x_1, x_2, \ldots, x_n$ represent an independent and identically distributed (iid) random sample of size *n* from a population with probability density function f(x) and cumulative density function (cdf) F(x). Let Φ be the cdf of x that is normally distributed with unknown mean and variance. The null hypothesis in this problem of testing for normality is

$$H_0 : F(x) = \Phi(x)$$

and the alternative hypothesis simply states H_0 is false. Hence only *omnibus* tests will be considered in this dissertation. Here, *omnibus* refers to the ability of a test to detect any deviation from normality with an adequate sample size.

In this problem, the focus is on failing to reject H_0 so that the conclusion is that the data come from a normal distribution. As noted by D'Agostino and Stephens (1986), this distinguishes normality testing from most statistical tests. Also, with a vague alternative hypothesis, they commented that 'the appropriate statistical test will often be by no means clear and no general Neyman-Pearson type (test) appears applicable'. Hence, it will be unlikely to have a single test that will have power superior to their alternatives.

1.2 Direction of Research

There are literally hundreds of normality tests in the literature. Major power studies done by Shapiro et al. (1968) and Pearson et al. (1977) have not arrived at a definitive answer; but a general consensus has been reached about which tests are powerful.

Pearson's (1900) chi-squared test, which is possibly the oldest, is not very sensitive. Data are grouped and compared to the expected counts under normality. Since information is lost in the grouping and this test is not specially tailored for the normal distribution, the conclusion is not surprising.

Bowman and Shenton (1975) proposed the use of joint contour plots of the third and fourth moments for their test. It proves powerful among tests based on moments. For these tests, sample moments are compared to those which are expected from a normal distribution. However, these tests are not omnibus since a distribution could have skewness and kurtosis close to that of a normal but yet the distribution could be non-normal.

Another class of normality tests are based on the empirical cumulative distribution function (ECDF). Deviation from normality is measured as a function of the discrepancies between the empirical and hypothesized distribution functions. Stephen's (1974) version to the Anderson-Darling (1954) test has proved to be the most powerful among these tests. However, it is not clear if this measure of deviation is of primary importance in deciding what to do if the data are non-normal.

A new direction was established in normality testing when Shapiro and Wilk (1965) formalized the evaluation of normality in probability plotting. Probability plotting involves plotting of ordered observations against their expected values under normality. Normality is judged by the linearity of the plot. This test and its modifications have proved to be popular among researchers since it is as powerful, if not more so, against certain alternative distributions as the Anderson-Darling and the Bowman-Shenton tests. However, as with the ECDF tests, it is unclear if the measure of deviation from normality is what researchers are concerned about.

A well known property of the normal distribution is that it has the least Fisher information among all other distributions with the same variance. A new approach to normality testing suggested by Terrell (1995) essentially provides a nonparametric density estimate of the data constrained on the above property when finding the likelihood of the data. This constraint eliminates the need for a smoothing parameter. The test is then based on the ratio of that likelihood to that under normality. Excess information would be reflected in a poorer fit of the data to normality since the Fisher Information is underestimated. Hence, the test is omnibus and sensitive to departures from normality that are manifested in the excess information.

The goal of this dissertation is to gain an understanding into the workings of the Fisher Information Test for normality, provide a comprehensive table of critical values and evaluate the power performance against existing tests. The Fisher Information Test will also be modified to test for normality in residuals.

In the next chapter, we will give a brief review of the background and details of the existing tests of normality. Chapter 3 develops the theory and operational details behind the Fisher Information Test. Chapter 4 presents an evaluation of the sensitivity of the Fisher Information Test and a power comparison is done against existing tests. Chapter 5 investigates the testing of normality in residuals and another power comparison is done to see if the results differ from those using independent observations. Finally, Chapter 6 summarizes the findings and also discusses directions for future research.

Chapter 2 Existing Normality Tests

For a detailed survey of the literature, see Mardia (1980) and D'Agostino and Stephens (1986). In this chapter, the focus is on existing normality tests which are mentioned in Chapter 1 that are powerful. Each of these tests belongs to a different class of normality tests. A brief background to the general approach in each class is given before details are presented for each test.

2.1 Moments Tests - Bowman-Shenton K^2

Since the concepts of skewness and kurtosis can be used to differentiate between distributions, one of the earliest classes of normality tests is based on these moments. The standardized coefficients of skewness, $\sqrt{\beta_1}$, and kurtosis, β_2 are defined as

$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} \text{ and } \beta_2 = \frac{\mu_4}{\sigma^4}$$

where μ_i is the ith central moment.

Skewness refers to the symmetry of a distribution. For a symmetric distribution like the normal, $\sqrt{\beta_1} = 0$. A distribution that is skewed to the right has $\sqrt{\beta_1} > 0$ while one that is skewed to the left has $\sqrt{\beta_1} < 0$.

Kurtosis refers to the flatness or 'peakedness' of a distribution. The normal distribution has $\beta_2 = 3$ and is used as a reference for other distributions. A leptokurtic distribution is one that is more peaked and with heavier tails than the normal, resulting in $\beta_2 > 3$. A platykurtic distribution has a flatter distribution with shorter tails than the normal, hence $\beta_2 < 3$.

The sample skewness, $\sqrt{b_1}$, and kurtosis, b_2, are defined as

$$\sqrt{b_1} = \frac{m_3}{\left(s^2\right)^{\frac{3}{2}}} \text{ and } b_2 = \frac{m_4}{\left(s^2\right)^2}$$

where m_i is the *i*th sample moment. Since the moments of $\sqrt{b_1}$ and b_2 are known, their distributions have been approximated using Pearson curves. The critical values for the normality tests of skewness and kurtosis are tabulated in Pearson and Hartley (1972) for selected values of $n \geq 25$ at $\alpha = 0.02$ and 0.10. Normalizing transformations have been found for $\sqrt{b_1}$ and b_2 by D'Agostino (1970) and D'Agostino and Pearson (1973) respectively. $Z(\sqrt{b_1})$ and $Z(b_2)$ denote the resulting approximate standardized normal variables.

D'Agostino and Pearson suggested combining $\sqrt{b_1}$ and b_2 in the following way:

$$K^2 = Z^2\left(\sqrt{b_1}\right) + Z^2\left(b_2\right)$$

where K^2 is distributed as χ_2^2 since it is the sum of the squares of 2 standardized normal equivalent deviates. However, they assumed that the squared standardized normal equivalent deviates are independent which is not true especially for small sample sizes. Using simulation, Bowman and Shenton (1975) obtained 90%, 95% and 99% contours for K^2 for sample sizes between 20 and 1000. Carrying out this test would then only require calculating $\sqrt{b_1}$ and b_2, selecting the appropriate contour, and determining if ($\sqrt{b_1}$, b_2) falls within the contours. If it does not, then normality is rejected.

2.2 Distance/ECDF Tests - Anderson-Darling A^2

ECDF or distance tests are another broad class of normality tests that are based on a comparison between the ECDF, $F_n(x_{(i)}) = \frac{i}{n}$, and the hypothesized distribution under normality, Z_i, as defined by

$$Z_i = \Phi\left(\frac{x_{(i)} - \bar{x}}{s}\right)$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$. Stephens (1974) provided versions of the ECDF tests with unknown μ and σ^2.

ECDF tests can be further classified into those involving either the supremum or the square of the discrepancies, $F_n(x_{(i)}) - Z_i$. The most well known ECDF tests involving the supremum is the Kolmogorov-Smirnov statistic

$$K = \max(D^+, D^-)$$

where $D^+ = \max\left(\frac{i}{n} - Z_i\right)$ and $D^- = \max\left(Z_i - \frac{i-1}{n}\right)$

ECDF tests involving the square of the discrepancies are known as those from the Cramér-von Mises family with the general form

$$CvM = n\int [\frac{i}{n} - Z_i]^2 \psi(Z_i) dZ_i$$

where $\psi(Z_i)$ is the weighting function.

If $\psi(Z_i) = 1$, that is the Cramér-von Mises statistic itself, W^2. For the Anderson-Darling statistic, A^2, $\psi(Z_i) = \frac{1}{Z_i(1-Z_i)}$. This choice of $\psi(Z_i)$ gives emphasis to tail values and the computational form is given by

$$A^2 = -\frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\left[\ln(Z_i) + \ln(1 - Z_{n+1-i})\right]\right] - n$$

Stephens found that A^2 has the highest power among all ECDF tests. The asymptotic distribution is known and it was found that the critical values for finite samples quickly converge to their asymptotic values for n ≥ 5.

2.3 Regression/Correlation Tests - Shapiro-Wilk *W*

The main idea behind these tests is normal probability plotting. Normal probability plotting is a graphical technique to determine the normality of the data by looking for linearity in a plot of the ordered observations $x_{(i)}$ against the expected values of standard normal order statistics, m_i. Formal determination of the linearity uses regression or correlation techniques, hence the name of this group of tests.

If $x_{(i)}$ is indeed normal, then the slope would give the standard deviation of x_i, σ, and the intercept, the mean of the x_i's, μ. Since the ordered observations are not independent, let $V=(v_{ij})$ be the nxn covariance matrix, $x' = (x_1, x_2, \ldots, x_n)$ and $m' = (m_1, m_2, \ldots, m_n)$. The best linear unbiased estimators of the slope and intercept using generalized least squares are

$$\hat{\sigma} = \frac{m'V^{-1}x}{m'V^{-1}m} \text{ and } \hat{\mu} = \bar{x}$$

The usual symmetric estimate of the variance regardless of the distribution of x_i is given by s^2.

The Shapiro and Wilk (1965) *W* statistic is defined as

$$W = \frac{K\hat{\sigma}^2}{(n-1)s^2} = \frac{a'x}{(n-1)s^2} = \frac{\left[\sum_{i=1}^{n} a_i x_{(i)}\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where

$$a' = (a_1, a_2, \ldots, a_n) = m'V^{-1}\left[(m'V^{-1})(V^{-1}m)\right]^{-\frac{1}{2}}$$

$$K = \frac{m'V^{-1}m}{m'V^{-1}V^{-1}m}$$

W compares the ratio of two estimates of variance, $\hat{\sigma}^2$ and s^2, apart from a normalizing constant, *K*, and (*n*-1). If the distribution of x_i is normal, then *W* will be close to 1. Otherwise, *W* is less than 1.

The critical values of *W* are tabulated up to sample sizes of 50. However, values for $\{a_i\}$ are also needed to carry out this test. For larger sample sizes, Shapiro and Francia (1972) noted that the ordered observations, as *n* increases , may be treated as independent (i.e. $v_{ij} = 0$ for $i \neq j$) . Treating *V* as an identity matrix, *W* can be extended for n larger than 50 by

$$W' = \frac{\left[\sum_{i=1}^{n} m_i x_{(i)}\right]^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n} m_i^2}$$

Values of $\{m_i\}$ are available from Harter (1961) up to sample sizes of 400. However, two tables are still needed to carry out this test.

A further modification was suggested by Weisberg and Bingham (1975) that uses this approximation

$$m_i \approx \Phi^{-1}\left(\frac{i-\frac{3}{8}}{n+\frac{1}{4}}\right)$$

due to Blom (1958). This approximation was shown to be close even in small samples, and the null distribution of W was practically identical to W'. This simplifies the computation of the test statistics since separate values for m_i need not be kept.

Royston (1982) used another approximation suggested by Shapiro and Wilk (1965) for $\{a_i\}$ and applied the following normalizing transformation to W:

$$y = (1-W)^{\lambda} \text{ and } z = (y-\mu_y)/\sigma_y$$

where z is standard normal and λ, μ_y and σ_y are functions of n. λ is estimated by maximizing the correlation between certain empirical quantiles of W and the corresponding standard normal equivalent with weights given according to the variance of a normal quantile. The relation between μ_y and σ_y and n is then determined by applying λ to simulated values W. The normalizing transformation producing W^* does away with any special tables, besides the standard ones, needed to find the critical values of W.

Chapter 3 Fisher Information Test

In this chapter, the theory behind the Fisher Information Test as suggested by Terrell (1995) will be explained in detail. First, the Fisher Information Inequality will be derived. Next, the Fisher Information Test will be developed and an implementation algorithm will be presented. Lastly, details of the simulation done to generate the critical values will be given.

3.1 Normal Information Inequality (for location)

The *Fisher information number* which is denoted by I_F is given by

$$I_F = E[-(\log f)''] = E[\{(\log f)'\}^2] = \int \frac{(f')^2}{f}$$

where f is any density. It measures, on average, how fast the log-likelihood changes as the mean moves away from the center of the distribution. Another way of looking at it would be the ease of using a sample to locate the center of a distribution. A famous property of the normal distribution is that its I_F is the smallest among all distributions with the same variance. The implication is that it is hardest to tell where the mean is in a normal distribution. The proof given by Terrell (1995) is presented here since it is integral to the development of the Fisher Information Test.

The right hand side of the familiar property of a density given below

$$1 = \int f$$

is integrated by parts for any μ and f that goes to zero at the limits of its support that results in

$$1 = -\int (x - \mu) f'$$

Introducing f to the integral by replacing f' with $\frac{f'}{f} f$ results in

$$1 = -\int (x - \mu) \tfrac{f'}{f} f$$

Applying the Cauchy-Schwartz Inequality yields

$$1 = \int [-(x-\mu)]\frac{f'}{f} f \le \int [-(x-\mu)]^2 f \int \left(\frac{f'}{f}\right)^2 f = \int (x-\mu)^2 f \int \left(\frac{f'}{f}\right)^2 f$$

Choosing μ to minimize the first integral makes $\int (x-\mu)^2 f = \mathrm{var}(X)$. Hence, the inequality after rearranging becomes

$$I_F = \int \frac{(f')^2}{f} \ge \frac{1}{\mathrm{var}(X)}$$

where equality is achieved when $(x-\mu)$ is proportional to $\frac{f'}{f} = (\log f)'$ almost everywhere. For a normal distribution,

$$(\log f)' = \frac{d}{d\mu}[\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(x-\mu)^2] = \frac{1}{\sigma^2}(x-\mu) \propto (x-\mu)$$

Therefore, for the normal distribution, the equality is achieved and $I_F = \frac{1}{\sigma^2}$.

This Fisher Information Inequality is sort of a dual to the Cramér-Rao Inequality. The Fisher Information Inequality gives the lower bound for the Fisher Information for any distribution in terms of its variance while the Cramér-Rao Inequality gives the lower bound for the variance of any location estimator in terms of its Fisher Information.

3.2 Normal Information Statistic F

For any other distribution besides the normal, the Fisher Information number would be in excess of the inverse of its variance. This excess would thus be a natural measure of non-normality or deviation from normality. A natural test statistic for normality would be to get a direct estimate of I_F. However, that would require an estimate of the density which relies on nonparametric methods that are asymptotically inefficient. Moreover, there is also the need to specify a smoothing parameter. Terrell circumvented these two problems by formulating the problem using maximum likelihood in the following way

$$\max_f \sum_{i=1}^{n} \log f(x_i) \quad subject \quad to \quad I_F = \frac{1}{s^2} \quad and \quad \int f = 1 \tag{3-1}$$

where the first constraint estimates I_F using the asymptotically efficient statistic s^2 under normality. The second constraint gives the familiar property of a density.

The Normal Information statistic, F, is then a log-ratio of this likelihood to that under normality. If the data are normal, then F is close to 1 since I_F has been correctly estimated. Otherwise, I_F is underestimated and F reflects a poorer fit of the data.

Rewriting (3.1) with Lagrange multipliers gives

$$\max_f \sum_{i=1}^{n} \log f(x_i) - \lambda \int \frac{(f')^2}{f} - \gamma \int f \tag{3-2}$$

where the values of λ and γ are chosen so that the constraints are met. If λ is treated as a parameter, the above form is similar to the first penalized maximum likelihood density estimation problem of Good and Gaskins (1971). This problem is a formal dual to theirs, and the form of the solutions to both are similar. A solution in terms of exponential splines has been found by deMontricher et al. (1975). Hence, the Fisher Information Test of Normality gives a graphical tool in the form of a nonparametric density estimate with the smoothing parameter specified by the constraint on I_F.

3.3 Solution to the Problem

Since there is a need for the resulting density estimate to be non-negative, the technique from Good-Gaskins of getting the solution in terms of a function $h^2 = f$ will be used. This will eliminate the need for a non-negativity constraint. With this modification, $f' = 2hh'$. The expression for Fisher information in terms of h is then

$$I_F = \int \frac{(f')^2}{f} = \int \frac{(2hh')^2}{h^2} = 4\int (h')^2$$

The optimization can be modified by maximizing the average log-likelihood and (3-2) becomes

$$\max_h \frac{2}{n} \sum_{i=1}^{n} \log h(x_i) - 4\lambda \int (h')^2 - \gamma \int h^2 \tag{3-3}$$

where λ and g are chosen so that $4\int (h')^2 = \frac{1}{s^2}$ and $\int h^2 = 1$. (3-3) is a particular case of the so-called isoperimetric problem in the calculus of variation.

Calculus of variation is a technique using classical calculus methods to solve maximization and minimization problems where the solution is a function instead of a point while isoperimetric problems are those that involve derivatives in the constraints. The main idea is to write the objective function and constraints as a Lagrangian function with *h* being replaced by $g + \varepsilon p$. Here, *g* is assumed to be the solution to the problem and εp is the perturbing function with *p* being an arbitrary function that vanishes at -∞ and ∞ (same as g) and e is an arbitrary constant. The Lagrangian function (*V)* is then a function of e. Note that as $\varepsilon \to 0$, $h \to g$. Therefore, the first-order necessary condition for the problem is then given by $\left.\frac{dV}{d\varepsilon}\right|_{\varepsilon=0} = 0$ and the second-order sufficient condition for maximization problems is $\frac{d^2V}{d\varepsilon^2} < 0$. For a reference to calculus of variation, see Chiang (1992).

To get the Euler-Lagrange variational condition for (3-3) from first principles, (3-3) is re-written as a function of e as follows:

$$V(\varepsilon) = \frac{2}{n}\sum_{i=1}^{n}\log\{\, g(x_i) + \varepsilon p(x_i)\} - 4\lambda\int (g' + \varepsilon p')^2 - \gamma\int\{\, g + \varepsilon p\}^2 \tag{3-4}$$

Expanding (3-4) results in

$$V(\varepsilon) = \frac{2}{n}\sum_{i=1}^{n}\log\{\, g(x_i) + \varepsilon p(x_i)\} - 4\lambda\int\left\{ (g')^2 + 2\varepsilon g'p' + \varepsilon^2 (p')^2 \right\} - \gamma\int\left\{ g^2 + 2\varepsilon gp + \varepsilon^2 p^2 \right\}$$

Differentiating *V(*e*)* with respect to e,

$$\frac{dV(\varepsilon)}{d\varepsilon} = \frac{2}{n}\sum_{i=1}^{n}\frac{p(x_i)}{g(x_i) + \varepsilon p(x_i)} - 4\lambda\int\left\{ 2g'p' + 2\varepsilon(p')^2 \right\} - \gamma\int\left\{ 2gp + 2\varepsilon p^2 \right\}$$

and using the first-order necessary condition gives

$$\left.\frac{dV(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \frac{1}{n}\sum_{i=1}^{n}\frac{p(x_i)}{g(x_i)} - 4\lambda\int g'p' - \gamma\int gp = 0$$

Using the shifting property of the Dirac delta function, $p(x_i)$ can be written as

$$p(x_i) = \int p(x)\delta(x - x_i)$$

and integrating the second term by parts,

$$\int g'p' = g'p\Big|_{-\infty}^{\infty} - \int g''p = -\int g''p$$

where use is made of the fact that p vanishes at -∞ and ∞. Substituting the above into the first-order condition gives

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\int p(x)\delta(x - x_i)}{g(x_i)} + 4\lambda\int g''p - \gamma\int gp = 0$$

Factoring p and collecting terms gives

$$\int p\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta(x - x_i)}{g(x_i)} + 4\lambda g'' - \gamma g\right\} = 0$$

Since p is arbitrary, the above reduces to

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta(x - x_i)}{g(x_i)} + 4\lambda g'' - \gamma g = 0$$

As the Dirac delta functional is zero except at zero, the final form of the Euler-Lagrange variational condition is given by

$$\frac{\frac{1}{n}\sum_{n=1}^{n}\delta\left(x - x_i\right)}{g} = \gamma g - 4\lambda g'' \qquad (3\text{-}4)$$

For the second-order necessary condition,

$$\frac{d^2V}{d\varepsilon^2} = -\frac{2}{n}\sum_{i=1}^{n}\frac{p(x_i)^2}{\{g(x_i) + \varepsilon p(x_i)\}^2} - 8\lambda\int (p')^2 - 2\gamma\int p^2 < 0$$

which ensures that the solution g gives a maximum solution to the problem.

Klonias (1982) found that

$$g_h^*(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h\left(x - x_i\right)}{g_h^*\left(x_i\right)} \qquad (3\text{-}5)$$

characterized the solutions to (3-4) where $K_h = \frac{1}{h}K\left(\frac{x}{h}\right)$ and $K = \frac{1}{2}e^{-|x|}$. Here, the principle of supposition is employed where (3-4) is solved for each n and then the solution is added together for the final solution to (3-4). Kernel functions, K, are used to solve the equation $K - K'' = \delta$ and the scaled version, K_h, is a solution to the equation

$$K_h - h^2 K_h'' = \delta \tag{3-6}$$

which looks like an $n = 1$ version of (3-4) . Dividing (3-6) by $n g_h^*(x_i)$ and summing from 1 to n gives

$$\frac{1}{n}\sum_{i=1}^{n}\frac{K_h(x-x_i)}{g_h^*(x_i)} - h^2\,\frac{1}{n}\sum_{i=1}^{n}\frac{K_h''(x-x_i)}{g_h^*(x_i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta(x-x_i)}{g_h^*(x_i)}$$

$$g_h^*(x) - h^2 g_h^{*\prime\prime}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta(x-x_i)}{g_h^*(x_i)}$$

where $g_h^{*\prime\prime}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h''(x-x_i)}{g_h^*(x_i)}$ and (3-5) are used for the substitution. Since the delta functional is zero except at zero

$$g_h^*(x) - h^2 g_h^{*\prime\prime}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}\delta(x-x_i)}{g_h^*(x)} \tag{3-7}$$

Instead of solving for λ and γ, the solution has been reparameterized to only one parameter, h. To tackle the second constraint that $g^*{}_h$ is the square root of a density where the square will integrate to one, let $a = \sqrt{\int g_h^*(x)^2\,dx}$. Then $g_h(x) = \frac{g_h^*(x)}{a}$ where integrating the square of g_h gives one, which verifies that $g^*{}_h$ indeed results in a square root of a density. Replacing (3-7) with g_h results in

$$a^2 g_h(x) - a^2 h^2 g_h''(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}\delta(x-x_i)}{g_h(x)} \tag{3-8}$$

where $g_h^{*\prime\prime}(x) = a g_h''(x)$.

As for the first constraint on the Fisher Information, an expression for Fisher Information is obtained by multiplying (3-8) by $g_h(x)$ and integrating. The first term equals a^2 since $\int g_h^2(x)dx = 1$ by definition. Integrating the second by parts

$$-\int h^2 a^2 g_h''(x) g_h(x)dx = -h^2 a^2 [g_h(x) g_h'(x)\Big|_{-\infty}^{\infty} - \int \left(g_h'(x)\right)^2 dx] = h^2 a^2 \int \left(g_h'(x)\right)^2 dx$$

where use is made of the assumption that the density vanishes at the limits of its support. The third term equals 1 since

$$\int \frac{1}{n}\sum_{i=1}^{n} \delta\left(x - x_i\right)dx = \frac{1}{n}\sum_{i=1}^{n} \int \delta\left(x - x_i\right)dx = \frac{1}{n}\sum_{i=1}^{n} 1 = 1$$

After the above simplification, (3-8) becomes

$$a^2 + a^2 h^2 \int \left(g_h'\right)^2 = 1$$

Hence, $I_F = 4\int \left(g_h'\right)^2 = 4\frac{1-a^2}{a^2 h^2}$ where the Fisher Information is a continuously decreasing function of h. If the data have been standardized, solving h for $I_F = 1$ would produce the required density estimate.

The Normal Information Test statistic, F, is twice the log likelihood ratio that compares the estimated density, g^2, to that of normality, f_0:

$$F = 2\log\frac{l\left(g^2\right)}{l(f_0)} = 2\left[\sum_{i=1}^{n} \log g^2 - \sum_{i=1}^{n} \log f_0\right] \tag{3-9}$$

Using the maximum likelihood estimators, $\hat{\sigma}^2$ and $\bar{x}$, in f_0 yields

$$F = 2\left[\sum_{i=1}^{n} \log g^2 - \left\{\log\left(2\pi\right)^{-\frac{n}{2}} + \log\left(\hat{\sigma}^2\right)^{-\frac{n}{2}} - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right\}\right]$$

The third term vanishes since $\hat{\sigma}^2 = 1$ as the data have been standardized. In addition, the last term simplifies to $-\frac{n}{2}$ using the definition of $\hat{\sigma}^2$. Finally, F simplifies to

$$F = 4\sum_{i=1}^{n} \log g + n\log\left(2\pi\right) + n \tag{3-10}$$

3.4 Computational Algorithm

The algorithm to get F is as follows :

1. Standardize the data to get variance one.
2. Solve the fixed point equation, $g_h^*(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{K_h\left(x - x_i\right)}{g_h^*\left(x_i\right)}$, by

a. using an initial estimate $g^{(0)}$ by taking the square root of a Laplace kernel density estimate.

b. computing a second estimate by $g^{(1)}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(x - x_i)}{g^{(0)}(x_i)}$

c. suppressing oscillations by $g^{(2)}(x) = \frac{1}{2}\left[g^{(0)}(x) + g^{(1)}(x)\right]$ and iterating to convergence.

3. Normalize the density so that its square integrates to one.
4. Compute Fisher Information in terms of h.
5. Find h that gives Fisher Information of one using the secant method, and then calculate F.

The details of implementing this algorithm in FORTRAN are given in Appendices B.1 and B.3.

3.5 Generation of Critical Values Using Simulation

Since the distribution of F is unknown, critical values have been generated via simulation. Sets of normal deviates are obtained using the subroutine ran1 from Press et al. (1992). Ten thousand values of F were generated for each sample size, n = 3(1)100(5)200. Different sets of pseudo-random numbers were used for each simulation to avoid dependence between results. The critical values obtained were then smoothed using fifth degree polynomials. The resulting smoothed critical values are tabulated in Appendix A for a at 0.50, 0.25, 0.20, 0.15, 0.10, 0.05, 0.025, 0.02 and 0.01 where bigger a values are available for those who are more inclined to accepting non-normality in their data.

Chapter 4 Evaluations and Applications to Real Data

From the definition of F in (3-9), it can be expected that the power of F is driven by the discrepancies between g^2 and f_0. The exact relationship is given in (3-10) which shows that F depends on the sample size, n, and the resulting square root of the density, g. To evaluate the sensitivity of F to non-normality, features of the data that affect g^2 will be examined. Then, based on those features, conjectures will be formed to see what aspects of non-normality F will be sensitive to. These conjectures could then be confirmed by a power comparison of F against existing tests. Finally, F is applied to some real data sets.

4.1 Features of Data that Affect Density Estimate, g^2

If the data are normal, the theory behind F would indicate that g^2 would give a good estimate of the density. To get a rough bell-shaped density, one would expect clustering of data points in the center and tail behavior to greatly affect the shape of g^2. Figure 4-1 shows the estimated density plot for {-1, 0, 1}. There are spikes at each data point with the one in the center receiving more weight than the other two. For {-1, 0, 0, 1} in Figure 4-2, the middle spike has even more weight with the additional data point in the center.

With sparse data, the resulting density estimate has to fill the spaces between and around data points to have a density with area that sums to one. Hence, one would not expect F to be powerful. As sample size increases, the density estimate is increasingly driven by the location of data points and how they cluster together. As a result, the ability of F to detect non-normality increases.

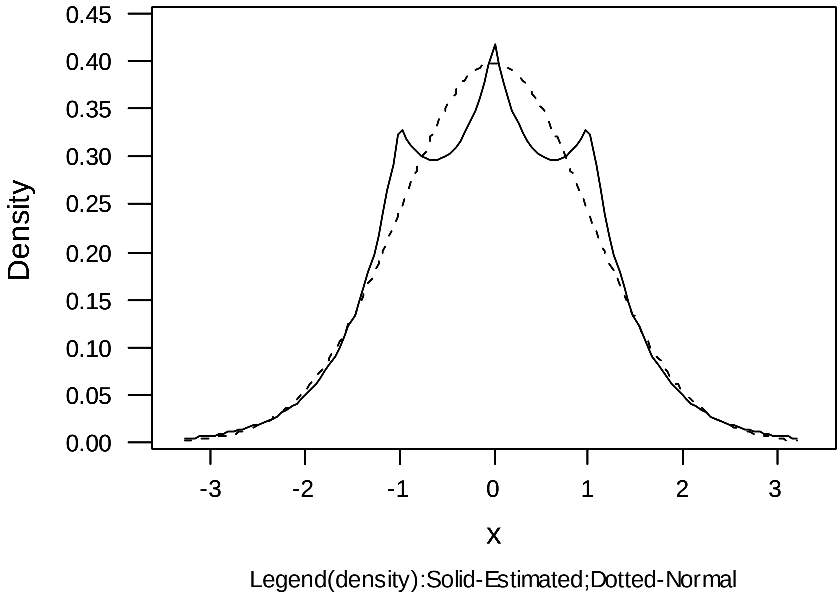


Figure 4-1 Estimated density for {-1, 0, 1}

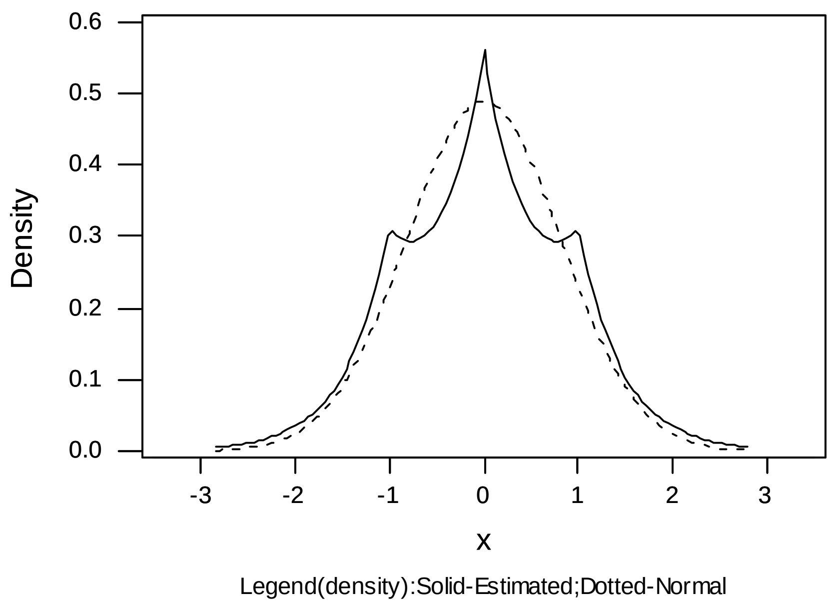


Figure 4-2 Estimated density for {-1, 0, 0, 1}

Note that for Figure 4-1 and Figure 4-2, the resulting tails are both tapering gently down at both ends. With no data in the tails, there is little discrepancy between g^2 and f_0. For {-1, 0, 0, 1, 5} with a prominent outlier, the spike in the right tail in Figure 4-3 testifies to the sensitivity of g^2 to tail behavior. Hence, tail behavior is another feature in the data that affects g^2.

Although g^2 is not a consistent density estimate of the underlying density with non-normal data, the discrepancies between g^2 and f_0 will have the potential to inflate F since the data might exhibit asymmetry and/or significant tail misbehavior. Since g^2 is affected by clustering of data and tail behavior, one would conjecture that F is most sensitive to leptokurtic, symmetric distributions since the ability to inflate F exists in both tails. Next would be leptokurtic, asymmetric where the ability is now confined to only one tail. With short tails in platykurtic distributions, F should be less powerful.

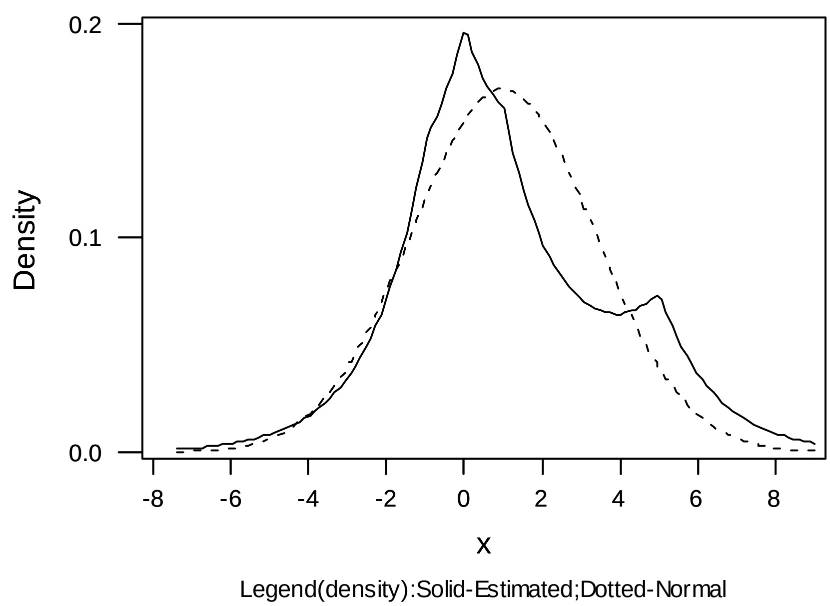


Figure 4-3 Estimated density for {-1, 0, 0, 1, 5}

4.2 Power Comparison

The power of *F* against existing normality tests is compared through a simulation study. For a review of other major power studies, see Shapiro et al. (1968) and Pearson et al. (1977). Refer to Section 1.2 for some general conclusions that have been reached from these major power studies.

4.2.1 Simulation Set-up

The simulation study was carried out with n = 10, 20, 50, 70 and 100 with 1000 samples drawn from 31 non-normal distributions specified in Table 4-1 and Table 4-2 for symmetric and asymmetric distributions, respectively. The distributions considered are classified according to the following groups:

I. symmetric, leptokurtic

II. symmetric, platykurtic

III. asymmetric, leptokurtic

IV. asymmetric, platykurtic

The distributions within each group are arranged in order of increasing departure from normality as measured by the standardized Fisher Information, $\mathrm{var}(X)I_F$. This measure is chosen so as to account for differing variances in distributions. Where $\mathrm{var}(X)I_F$ does not exist, the distributions are ordered on the basis of their standardized coefficient of kurtosis, β_2. In group I, the distributions include $SC(\varepsilon,\sigma_\varepsilon^2)$ which is the scale-contaminated normal with 100ε% of $N(0,\sigma_\varepsilon^2)$ being the contaminant. Similarly, $LC(\varepsilon,\mu_\varepsilon)$ is the location-contaminated normal with 100ε% of $N(\mu_\varepsilon,1)$ being the contaminant in group III.

Table 4-1 Properties of symmetric distributions used in simulation study

Distributions	Var(X)	β_2	I_F	Var(X)I_F
I Symmetric, leptokurtic				
Normal	1	3	1	1
t_{10}	1.25	4	0.85	1.06
Logistic	3.29	4.2	0.33	1.10
SC(0.05, 9)	1.4	7.65	0.88*	1.24
SC(0.10, 9)	1.8	8.33	0.79*	1.42
t_4	2	-	0.71	1.43
SC(0.05, 25)	2.2	19.96	0.89*	1.95
Laplace	2	6	1	2
SC(0.10, 25)	3.4	16.45	0.80*	2.71
t_2	-	-	0.6	-
Cauchy	-	-	0.5	-
II Symmetric, platykurtic				
U(0,1)	0.08	1.8	-	-
Beta(1.5, 1.5)	0.06	2	-	-
Beta(2,2)	0.05	2.14	-	-

*using numerical integration

Table 4-2 Properties of asymmetric distributions used in simulation study

Distributions	Var(X)	$\sqrt{\beta_1}$	β_2	I_F	Var(X)I_F
III Asymmetric, leptokurtic					
Weibull(2)	0.21	0.63	3.25	-	-
LC(0.05,3)	1.43	0.67	4.35	0.85*	1.22
LC(0.10,3)	1.81	0.80	4.02	0.77*	1.40
LC(0.20,3)	2.44	0.68	3.09	0.67*	1.63
Chi-squared(10)	20	0.89	4.2	0.08	1.67
LC(0.05,5)	2.19	1.65	7.44	0.95*	2.09
LC(0.10,5)	3.25	1.54	5.45	0.93*	3.03
LC(0.05,7)	3.33	2.42	10.37	1.00*	3.32
LC(0.20,5)	5	1.07	3.16	0.91*	4.54
LC(0.10,7)	5.41	1.96	6.40	0.99*	5.38
LC(0.20,7)	8.84	1.25	3.20	0.99*	8.78
Chi-squared(4)	8	1.41	6	-	-
Chi-squared(2)	4	2	9	-	-
Chi-squared(1)	2	2.83	15	-	-
Weibull(0.5)	20	6.62	87.72	-	-
Lognormal(0, 1)	4.67	6.18	113.94	14.78	69.03
IV Asymmetric, platykurtic					
Beta(3,2)	0.04	-0.29	2.36	-	-
Beta(2,1)	0.06	-0.57	2.4	-	-

*using numerical integration

Table 4-3 Discrete distribution with normal moments

X	P(X=x)
-4	0.115827
-1	0.384173
1	0.384173
4	0.115827

Table 4-4 Power estimates of discrete distribution with normal moments

Sample size, n	K^2	W	W^*	A^2	F
α = 0.05					
10	-	0.600	0.600	0.750	0.750
20	0.047	1.000	1.000	1.000	1.000
50	0.012	1.000	1.000	1.000	1.000
100	0.004	1.000	1.000	1.000	1.000
α = 0.10					
10	-	0.650	0.685	0.832	0.832
20	0.097	1.000	1.000	1.000	1.000
50	0.012	1.000	1.000	1.000	1.000
100	0.013	1.000	1.000	1.000	1.000

The existing normality tests considered in this study include *W(W')*, W^* and A^2. Recall that W is the Shapiro-Wilk (1965) test and A^2 is Stephen's (1974) version to the Anderson-Darling (1954) test. Where the sample size exceeds 50, Shapiro-Francia (1972) W' will be used in place of W since it extends the range of W from 50 and below to 400. W^*, which is Royston's (1982) approximation to *W(W')*, will be considered a separate test as it will be informative to compare its power to *W(W')*.

K^2 is left out of the power study since it is not an omnibus test. To illustrate this point, a discrete distribution with normal moments is added to this simulation study. Table 4-3 gives the details of such a discrete distribution that has the same first to fourth moments as the normal. The power of the normality tests with this distribution is given in Table 4-4. Results are not obtained for K^2 at n = 10 since the exact contours are not available. All the tests except K^2 had estimated power above 0.60 even for n as low as 10. For sample sizes 20 or larger, these tests had estimated power of 1.00. The power for K^2 is even lower than the nominal α value especially for higher sample sizes. Hence, K^2, in particular, and moments tests, in general, are only able to detect distributions with non-normal moments and are not omnibus tests.

To differentiate between the tests to see if one test is superior to another, the practice in the literature has been to determine which test has the highest power based on the same set of pseudo-random numbers for each distribution. To generalize the results across different distributions, the averaged rank calculated for each test is sometimes used. The fact that a different set of pseudo-random numbers might give rise to a different ordering of the power is usually ignored.

To account for this variability, a formal statistical test on the equality of the power of the tests is conducted in this power study. As all the tests are subjected to the same set of pseudo-random numbers, the powers of the individual tests are correlated. Hence, Cochran's Q is used to account for this correlation. In cases where the equal power hypothesis is rejected, McNemar's test with correction for continuity is used for pairwise comparisons to determine whether the test with the highest power is significantly different from the rest. To maintain the overall type I error rate at 0.05 in the presence of multiple testings, the idea from Fisher's Least Significance Difference is used here. This means that multiple comparisons are carried out only if the hypothesis of equal power using Cochran's Q is rejected. In addition, the same type I error rate is used for both Cochran's Q and McNemar's tests. For details of both tests, refer to Siegel and Castellan (1988).

The results from using Cochran's Q and McNemar's tests will be reflected as superscripts to the test with the highest power in this power study. The superscripts will denote the number of tests, including the one with the highest power, that are significantly better than the rest. Hence, a '1' would reflect that the test with the highest power has significantly higher power than the rest while a '4' would mean that all the tests have the same power.

The empirical level of each test is also given based on a normal sample of 10 000. 95% confidence intervals on the empirical level of each test will be used to assess if they contain the relevant nominal levels. This information is useful since it acts as a check on possible inflation/deflation of the power estimates.

For programming details involved in this simulation study, please refer to Appendix B.2.

4.2.2 Results

$\underline{n=10}$

Table C-1(a) shows the results for α = 0.05. The empirical level for each test is given by the power estimates for the normal distribution. Here, all the confidence intervals contain the nominal value of 0.05.

For group I, *F* is the most sensitive for most of the distributions, having significantly higher power than the other tests for SC(0.10, 9), SC(0.10, 25) and t_2. For t_{10} and SC(0.05, 25) where *F* did not have the highest power, all four tests have power that are not significantly different from one another. *W* is the least sensitive in distributions where not all tests have the same power.

For group II, A^2 has the highest power in all three distributions. However, its power is not significantly higher than *W** and *W* while *F* proves to be the least sensitive.

For asymmetric and leptokurtic distributions in group III, there is no clear dominance of any one test. For location contaminated normals (LCs), A^2, *W** and *F* have the highest power for different LCs, with A^2 having significantly higher power for LC(0.20,5). As for non-LCs, *W** clearly is the most sensitive with significantly higher power for all distributions except Weibull(2), Chi-squared(10) and Weibull(0.5); *F* is the least sensitive especially for those with higher var(X)I_F.

As for distributions in group IV, *W** has the highest power but it is not significantly different from *W* and A^2 while *F* again proves to be the least sensitive.

The results for α = 0.10 are given in Table C-6(a). Here, *F* has the highest power in most of the distributions in group I with the power being significantly higher for the Laplace distribution. Again, *W* is the least sensitive for distributions that are symmetric and leptokurtic. As for group II, all tests except *F* are equally good at detecting non-normality.

For non-LCs in group III, *W** is the most sensitive, with the power for Chi-squared(4), Chi-squared(2) and Lognormal(0,1) being significantly higher. As for LCs, A^2 and *F* are more sensitive than *W* and *W** in detecting non-normality, with A^2 having significantly higher power in LC(0.20,7). For group IV, both A^2 and *W* have the highest power but none of them are significantly higher than the rest. Once again, *F* is the least sensitive.

n=20

The results for a = 0.05 are given in Table C-2(a). *F* is the most sensitive in detecting non-normality in group I with significantly higher power in all distributions except t_{10} and SC(0.05, 9). For groups II, IV and non-LCs, *W* is the most sensitive in most distributions, with significantly higher power in Chi-squared(10). *W** proves to be equally good in most cases while *F* is the least sensitive.

As for LCs, *F* is the most sensitive in six of the distributions with those for LC(0.05,3), LC(0.10,3) and LC(0.05,7) being significantly higher. *W*, *W** and A^2 are equally sensitive in detecting non-normality for the remaining LCs but are not as dominant as *F*.

The results for a = 0.10 are given in Table C-7(a). For group I, *F* has significantly higher power in all distributions except in t_{10} where all four tests are equally sensitive. On the whole, both *W* and *W** are most sensitive in detecting non-normality in groups II and IV as well as non-LCs. As for LCs, *F* has significantly higher power in LC(0.05,3), LC(0.10,3) and LC(0.10,5). For the remaining LCs, *F*, *W* and *W** are equally sensitive.

n=50

The results for a = 0.05 are given in Table C-3(a) with the empirical level for *W* being much lower than the nominal value. Hence, the power for *W* is underestimated and it is not surprising that *W** emerged with significantly higher power in groups II and IV as well as in Weibull(2) and Chi-squared(10) for the non-LCs. Further, *F*'s position is unchallenged in group I with significantly higher power in all distributions except for the Cauchy. *F* is also most sensitive for most LCs with significantly higher power for LC(0.05,3), LC(0.10,3), LC(0.05,5) and LC(0.10,5).

The other thing to note is that certain distributions in group III with higher $var(X)I_F$ are beginning to be so extreme that all tests are equally adept at detecting them.

These distributions include LC(0.20,5), Chi-squared(1), Weibull(0.5) and Lognormal(0, 1).

Table C-8(a) contains the results for α =0.10. Here, the power for *W* is not underestimated. A fairer comparison can then be made of the sensitivity of *W** and *F*. The results are similar for *F* in group I and in LCs. In groups II and III, *W** still has significantly higher power in Beta(1,1), Beta(2,2) and Beta(3,2) but are equally sensitive for the remaining distributions as *W*. The same applies to non-LCs.

As for LCs, the only anomaly is that A^2 has significantly higher power for LC(0.20,3). Again, distributions with high var(X)I_F in group III are all detected by all of the normality tests.

<u>*n*=70</u>

Table C-4(a) displays the results for α = 0.05. The empirical level of *W'* of 0.067 is much higher than the nominal value of 0.05. This confirms the hesitance of Pearson et al. (1977) in recommending the use of *W'* since they pointed out that the empirical critical values were overstated as a result of being based only on 1 000 samples. They warned that this 'unfairly enhances' the power of *W'*. With this in mind, it is not surprising that for distributions in group I and some LCs, *W'* has significantly higher power than the other tests. For these distributions, *F* consistently has the second highest power for these distributions.

In spite of the inflated power for *W'*, both *W** and A^2 managed to have significantly higher power: *W** in groups II and IV as well as Weibull(2) in group III and A^2 in LC(0.20,3). In addition, the inflation of power in *W'* did not affect those distributions in group III that are detected by all the normality tests 100% of the time.

The results for α = 0.10 in Table C-9(a) are very similar.

<u>*n*=100</u>

Since the power of *W'* is inflated for *n*=70, the critical value used for *W'* for *n*=100 was the average empirical critical value of *W'* obtained by Pearson et al. (1977) to adjust for the inflation to get a fair comparison. This is reflected in Table C-5(a) and

Table C-10(a) where the nominal levels are contained in the 95% confidence intervals for the empirical levels.

In Table C-5(a) for $\alpha = 0.05$, *F* and *W* are sensitive in detecting non-normality in group I with *F* having significantly higher power for t_4 and the Laplace distribution. As for groups II and IV as well as non-LCs distribution like Weibull(2) and Chi-squared(10), *W** has significantly higher power. As for LCs, *W* has significantly higher power in LC(0.05,3) and LC(0.10,3) while A^2 excels in detecting LC(0.20,3). For the most of the remaining distributions in group III, all the tests are able to detect non-normality 100% of the time.

A look at Table C-10(a) for $\alpha = 0.10$ reveals similar findings. *W** has significantly higher power in group II as well as for Weibull(2) in group III and Beta(3,2) in group IV. Again, *W* and A^2 have significantly higher power in the same LCs and slightly more than half of the distributions in group III are detected 100% of the time by all the normality tests. However, in group I, *F* and *W* are now equally sensitive in detecting non-normality in group I.

4.2.3 Summary

As expected, no one test has significantly higher power than all other tests for all the distributions. However, some broad patterns have emerged regarding the sensitivity of each test to the different types of distribution. The following summarizes the results from the power study:

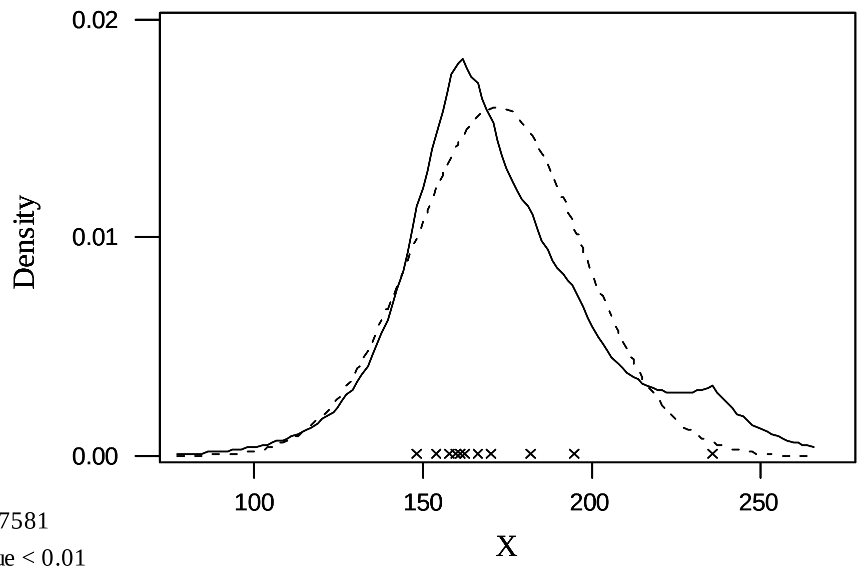
1. For distributions that are symmetric and leptokurtic, *F* is superior to the other tests for detecting non-normality.
2. For distributions that are platykurtic or asymmetric excluding LCs, *W** is superior for larger sample sizes ($n \geq 50$) while both *W* and *W** are equally sensitive for smaller ones.
3. LCs behaves like a continuum between leptokurtic distributions that are symmetric to those that are asymmetric as n, p and μ_ε increases. Hence, no one test is superior. With p and μ_ε small, *F* is more sensitive for smaller *n* (≤ 50) while *W* is better at larger *n*. As p and μ_ε increase, A^2 is more sensitive. However, there comes a point

when p and μ_ε become so big that all the normality tests easily detect non-normality 100% of the time.

4. It is not surprising that when sample sizes are small (n<50), *W** has power that is equal to *W*. However, at larger sample sizes, *W** is preferred, since its power is neither inflated/deflated. In some cases, *W** has significantly higher power than *W* even when *W's* power estimates are inflated. Hence *W** is preferred over *W*.
5. An examination of the power of *F* shows that besides increasing with *n*, it also varies directly with $\mathrm{Var(X)I}_F$, albeit the relationship is not a deterministic one.

4.3 Applications to Real Data Sets

In this section, *F* is applied to several real data sets. Here, the estimated density is plotted against the normal density with the same sample mean and variance as the data. This graphic best illustrates any deviation from normality and provides a ready explanation when normality is rejected.



F = 4.7581
P-value < 0.01

Legend(density):solid-estimated;dotted-normal

Figure 4-4 Density estimate of Male Weights Data (n=11)

4.3.1 Male Weights Data

Shapiro and Wilk (1965) used their test on a data set of 11 adult male weights taken from Snedecor (1946). These are, in pounds, 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, and 236. The resulting statistic F is 4.7581, which is beyond the 99^{th} percentile. This is consistent with the result given by W. The resulting density estimate in Figure 4-4 shows a prominent outlier at 236 with a peak in the right tail that accounts for

the rejection of normality for this data set.

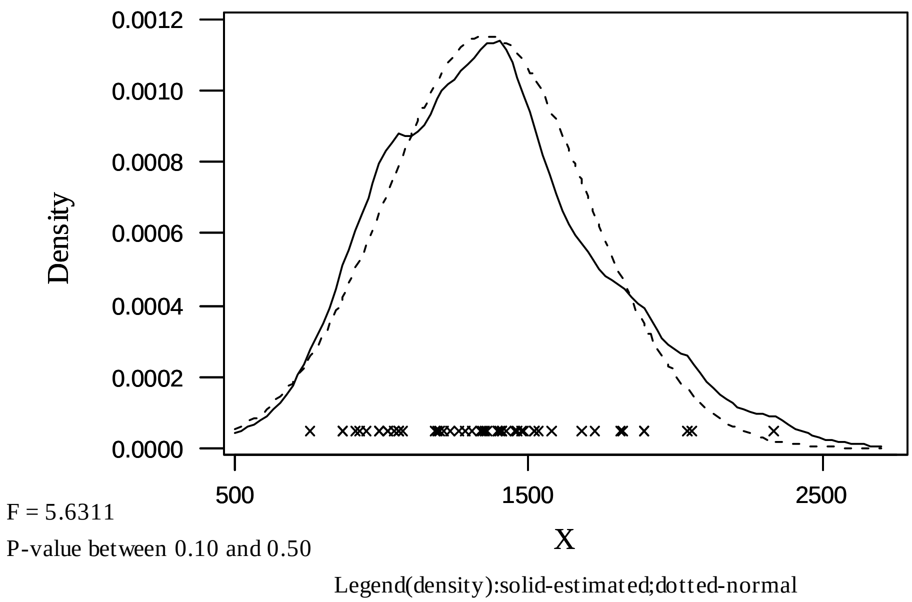
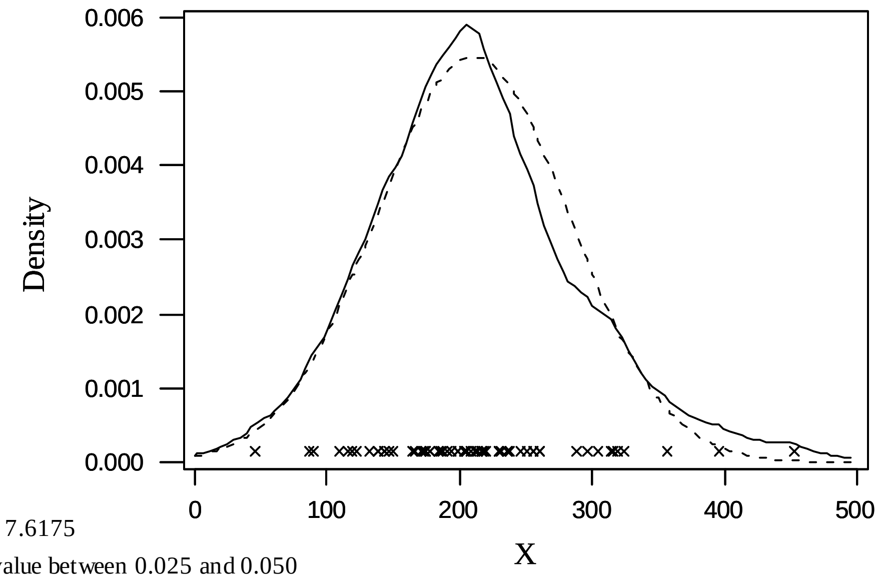


Figure 4-5 Density estimate of Mississippi River Data (n=49)

4.3.2 Mississippi River Data

Another example is taken from Gumbel (1943) which gives the maximum daily rates of discharge from the Mississippi River at Vicksburg in cubic feet per second for 49 years starting from 1890. Assuming that the data are independent, the resulting statistic F being 5.6311 is between the 50^{th} and 90^{th} percentile. From Figure 4-5, it can be seen that the data do not deviate much from normality and hence supports the contention that the data can be approximated by the normal distribution.



F = 7.6175

P-value between 0.025 and 0.050

Legend(density):solid-estimated;dotted-normal

Figure 4-6 Density estimate of PCB Data (n=65)

4.3.3 PCB Data

A third example is taken from Risebrough (1972) who was studying concentrations of polychlorinated biphenyl (PCB), an industrial pollutant, in the yolk lipids of pelican eggs. He had a sample size of 65 and the resulting F is 7.6175, which is between the 95th and 97.5th percentile. Figure 4-6 shows the resulting density plot which is close to normal except for two outlying points in the right tail. Rejection of normality using α of 0.05 is therefore not surprising.

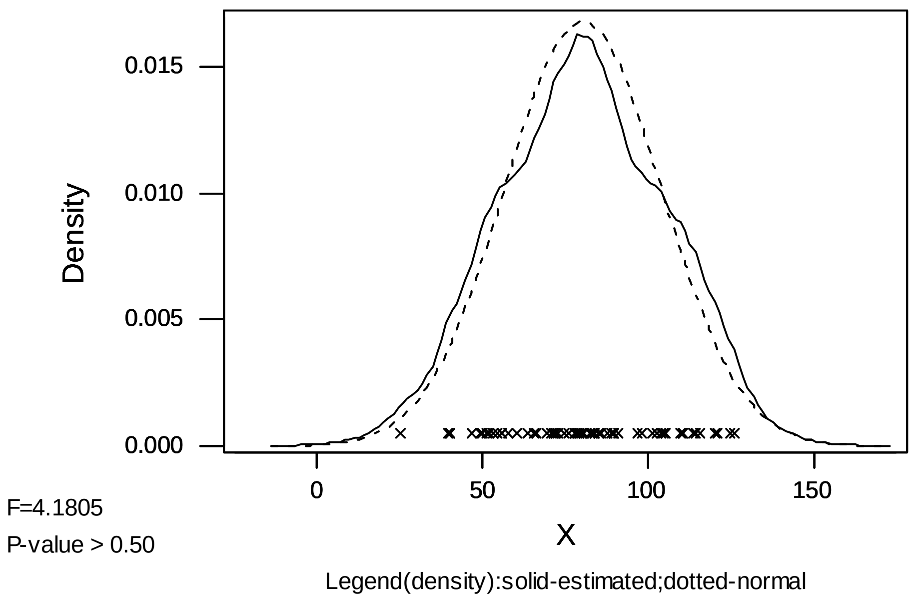


Figure 4-7 Density estimate of Buffalo Snowfall Data (n=63)

4.3.4 Buffalo Snowfall Data

This is a popular example used in the density estimation literature; see Silverman (1986). The data are a record of the amount of winter snowfall (in inches) at Buffalo, New York, for 63 winters from 1910/11 to 1972/73. Silverman showed that the data could either be unimodal or trimodal, depending on the smoothing parameter used. Subjecting the data to a normality test results in getting a density estimate as shown in Figure 4-7. The data fit a normal distribution quite well. F is below the 50^{th} percentile and normality is not rejected. In the framework of F, the data are unimodal.

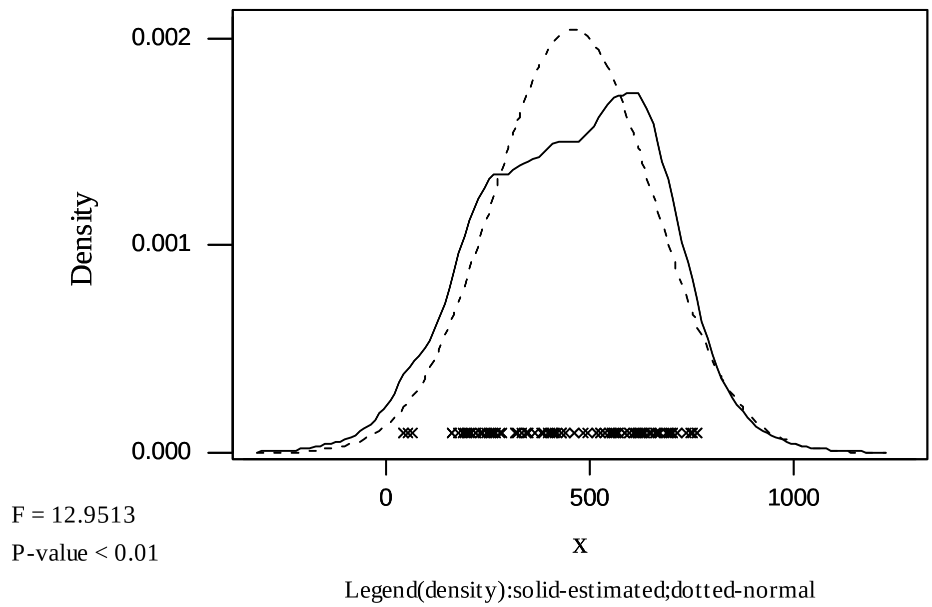


Figure 4-8 Density estimate of Mice Data (n=99)

4.3.5 Mice Data

This example is taken from Hoel (1972) who was looking at survival times of mice that were exposed to radiation. He had a sample size of 99 and the resulting F is 12.9513, which is beyond the 99^{th} percentile. Figure 4-8 shows the resulting density plot which is skewed to the left. Rejection of normality using α of 0.05 is therefore not surprising.

Chapter 5 Testing Normality of Residuals

In this chapter, the testing of normality in residuals will be investigated. Firstly, the background needed for testing residuals will be given. This includes a review of the existing methods in the literature. Next, *F* will then be modified to take into account the special nature of residuals as opposed to independent observations. Lastly, a power comparison is done to see if the results of testing normality in residuals differ from those using independent observations.

5.1 Background

Consider the following general linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

where $\underline{y}$ is a *n* x 1 vector of the observed responses, X is a known non-stochastic *n* x *k* matrix of rank *k*, $\underline{\beta}$ is a *k* x 1 vector of unknown parameters and $\underline{\varepsilon}$ is an *n* x 1 vector of unobserved error terms which are assumed to be independent and normally distributed with mean zero and constant variance, σ^2.

After the above model is fitted using least squares, the resulting ordinary least squares (OLS) residuals

$$\hat{\underline{\varepsilon}}_{OLS} = \left(I - X\left(X'X\right)^{-1}X'\right)\underline{\varepsilon} \tag{5-1}$$

are distributed as

$$\hat{\underline{\varepsilon}}_{OLS} \sim N\left[\underline{0}, \left(I - X\left(X'X\right)^{-1}X'\right)\sigma^2\right]$$

with $\left(I - X\left(X'X\right)^{-1}X'\right)$ being a *n* x *n* symmetric idempotent matrix with rank *n* – *k*. Thus $\hat{\underline{\varepsilon}}_{OLS}$ has a singular normal distribution that may be heteroscedastic and correlated.

Since the normality tests considered in Chapter 4 are developed based on independent observations, they cannot be used to test residuals without a study on the effect on the levels and powers of the normality tests. In the literature, one of the ways to deal with this problem is to transform the residuals by reducing the dimensionality so that they become homoscedastic and uncorrelated. See Cook and Weisberg (1982) for details on two types of transformations. One of them is Theil's (1965) best linear unbiased scalar (BLUS) residuals.

Huang and Bolch (1974) compared the power of normality tests using OLS and BLUS residuals in their simulations and found that using OLS gives superior power over BLUS especially for *n* greater than 30. Moreover, the one-to-one link between residuals and cases or data points no longer exists for BLUS residuals, making interpretation difficult in situations where there are outliers. In addition, there will be no advantage in using BLUS over OLS residuals since both suffer from the lack of independence when the underlying error distribution is not normal.

In view of the above two disadvantages in using BLUS residuals, the direction in the literature turned to finding conditions under which OLS residuals could be used to test normality of the true but unobserved errors. White and Macdonald (1980) found that under certain conditions, the distributions of certain test statistics used in normality tests remain asymptotically valid when OLS residuals are used instead of the true errors. To look at how the approximation works in finite samples, they carried out simulations to compare the power of normality tests using OLS residuals against their true unobserved error in addition to a comparison among normality tests. Using their terminology, test statistics that use the true errors are known as 'true' statistics while those using the OLS residuals are known as the 'modified' statistics.

In their simulation, they looked at n = 20, 35, 50 and 100 with a = 0.1 and the *k-1* regressors in X were generated from a uniform distribution with the first regressor being a column of ones. They only considered the case where k = 4 and used the following five non-normal distributions (arranged in order of departure from normality): Teichroew's (1956) heteroscedastic normal, t_5, Laplace (double exponential), χ^2_2 (exponential) and lognormal.

They found that the powers of the modified statistics are lower than the true ones. At n = 20, the powers of the modified statistics are not less than 60 percent of those of the true statistics. This percentage rises to 80 percent at n = 35. It gets beyond 90% for n = 100. At this sample size, the powers of the true and modified statistics were considered to be practically the same. Weisberg (1980) provided an explanation for the lower powers in modified statistics by writing (5-1) in terms of the *i*th OLS residual as follows:

$$\hat{\varepsilon}_i = (1 - h_{ii})\,\varepsilon_i - \sum_{j \neq i} h_{ij}\varepsilon_i \qquad (5\text{-}2)$$

where h_{ij} is the (i, j)th element of $H = X(X'X)^{-1}X'$. Since the second term in (5-2) is a sum, $\hat{\varepsilon}_i$ will tend to be normally distributed even if ε_i itself is non-normal. This phenomenon is known as supernormality, as first coined by Gnanadesikan (1977).

Not surprisingly, the discrepancies between the powers of the modified and true statistics also differ across different distributions – the greater the error distribution's departure from normality, the higher are the discrepancies. In addition, it was also found that no one test is dominant for all the error distributions considered.

Besides looking at power, White and Macdonald(1980) used measures such as r (correlation between true and modified statistics) and *m* (maximum absolute deviation between the true and modified statistics divided by the standard error of true statistic) to quantify the correspondence between the true and modified statistics. Their results showed that as *n* increases, *m* decreases while r increases to unity. In addition, the increase in these measures are also affected by the test itself as well as by the underlying error distribution.

In their conclusion, they warned that their results might not generalize well to all situations since they only looked at a specific way of generating the regressors in the X matrix.

Weisberg (1980) pointed out that any conclusion drawn from simulation studies involving the use of modified statistics must take into account variations in *n*, *k* and *H* (through how the regressors in *X* are generated). He supplemented the results from White and Macdonald by looking at two other ways of generating the regressors in the X matrix and extending *k* to 6, 8 and 10. However, he only considered *W'* at $n = 20$ and showed that the results do indeed depend on *k* and *H* besides *n*.

5.2 Power Comparisons

Arising from Section 5.1, it would be interesting to extend the work done by White and Macdonald (1980) and Weisberg (1980) to see if *F* still has superior power in cases where the error distributions are leptokurtic and symmetric. This is of special interest to those in the field of robust statistics where it is well known that leptokurtic error distributions make OLS estimators very inefficient.

Since the exact distribution of F is unknown at this point, it would not be possible to prove theoretically that the distribution of the modified test statistic for F is asymptotically valid. However, the case for finite samples could be examined by using the independent observations generated in the power study in Section 4.2 as true errors and comparing their power to those of the resulting OLS residuals.

5.2.1 Simulation Set-up

For this power study to be comparable to previous work done, the regressors in the X matrix will first be chosen from a uniform distribution with mean zero and variance 25. This mimics what White and Macdonald (1980) used in their study. Weisberg (1980) considered this formulation of X as the ideal case for normality tests since the first term in (5-2) is of order 1 while the second is of order n^{-1}. This implies that the ith OLS residual is primarily determined by the ith true residual as n increases. Hence, the effect of supernormality is mitigated with increasing n.

The relative effects on the power of modified statistics compared to their true counterparts will be evaluated at n = 10, 20, 35, 50 and 100 with a = 0.1 and the same five error distributions used in White and Macdonald (1980). To further generalize the results across a wider array of distributions, the entire range of distributions considered in the power study of Section 4.2 will be used here.

The second choice of X will be taken from Set 1 in Weisberg (1980). This choice of X seems to consist of generating the different columns from different distributions. Here, the relative effects on the power between true and modified statistics will be considered across k = 4, 6, 8 and 10 with the same five error distributions used in White and Macdonald (1980).

For both parts of this study, r and m will be used to measure the level of correspondence between the true and modified statistics. In addition, Cochran’s Q and McNemar’s tests will be used to evaluate the performance of the individual modified statistics to see if any test has superior power in detecting non-normality for each of the error distributions considered.

For programming details, please refer to Appendix B.2.

5.2.2 Results

<u>Comparison across *n*</u>

Table 5-1 contains the results for comparing the power between the true and modified statistics across different *n* using the regressors in X being generated from a uniform distribution.

For the true statistics, the 95% confidence interval of the levels of each test all contain the nominal value of 0.10 except for *W* at *n* = 35. Hence, the power estimates for *W* at *n* = 35 are underestimated although the effects are not that serious since *W* has already attained the highest power for Chi-squared(2) and the Lognormal(0,1) distribution.

In general, the levels of the normality tests do not seem to be affected by the use of OLS residuals except for *n* = 10 where the levels are smaller than the nominal value of 0.1. This is consistent with the idea of supernormality. Moreover, the practical effect of this is to have conservative normality tests at small sample sizes when modified statistics are used.

At *n* = 35, only the levels of *W* and *F* seem to be affected. Their power estimates are underestimated since the 95% confidence intervals for their empirical levels are below the nominal value. However, this does not change the results for *F* since it already attains significantly higher power for t_5 and the Laplace distribution. Similarly, the general conclusion for W is unchanged for the Lognormal(0,1) distribution, since it has the highest power, although the effects on the other distributions are harder to predict.

Table 5-1 Power comparisons of normality tests on iid observations and OLS residuals across different values of n based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1

Distributions	iid observations(true statistics)				OLS residuals(modified statistics)			
	W(W')	*W**	A^2	*F*	*W(W')*	*W**	A^2	*F*
n=10								
Normal[a]	0.095	0.100	0.102	0.100	0.078[b]	0.083[b]	0.087[b]	0.087[b]
Hetero. Normal	0.133	0.139	0.146[4]	0.139	0.071	0.077	0.083	0.088[2]
t_5	0.165	0.171	0.162	0.173[4]	0.100	0.117	0.121	0.129[3]
Laplace	0.204	0.213	0.216	0.242[1]	0.101	0.110	0.118	0.123[2]
Chi-squared(2)	0.542	0.550[1]	0.519	0.513	0.153	0.164	0.160	0.173[2]
Lognormal(0,1)	0.669	0.678[1]	0.648	0.641	0.233	0.247	0.230	0.261[2]
n=20								
Normal[a]	0.099	0.097	0.104	0.102	0.098	0.095	0.102	0.098
Hetero. Normal	0.105	0.100	0.116	0.123[2]	0.102	0.100	0.113	0.114[4]
t_5	0.265	0.261	0.246	0.287[1]	0.187	0.180	0.193	0.220[1]
Laplace	0.335	0.327	0.352	0.405[1]	0.250	0.248	0.244	0.286[1]
Chi-squared(2)	0.906[2]	0.905	0.863	0.851	0.573[2]	0.569	0.543	0.535
Lognormal(0,1)	0.967[2]	0.965	0.941	0.938	0.750[2]	0.749	0.724	0.729
n=35								
Normal[a]	0.092[b]	0.097	0.103	0.095	0.089[b]	0.095	0.099	0.092[b]
Hetero. Normal	0.113	0.118	0.122	0.137[2]	0.111	0.116	0.125	0.137[2]
t_5	0.307	0.312	0.328	0.401[1]	0.262	0.269	0.277	0.354[1]
Laplace	0.428	0.437	0.523	0.579[1]	0.349	0.354	0.416	0.485[1]
Chi-squared(2)	0.997[2]	0.997[2]	0.988	0.986	0.927	0.928[2]	0.897	0.904
Lognormal(0,1)	1.000[4]	1.000[4]	0.998	0.998	0.972[4]	0.972[4]	0.969	0.967
n=50								
Normal[a]	0.097	0.100	0.102	0.102	0.100	0.102	0.103	0.099
Hetero. Normal	0.104	0.108	0.129	0.151[1]	0.114	0.117	0.130	0.159[1]
t_5	0.342	0.343	0.405	0.519[1]	0.304	0.307	0.346	0.447[1]
Laplace	0.528	0.532	0.669	0.717[1]	0.444	0.449	0.568	0.642[1]
Chi-squared(2)	1.000[4]	1.000[4]	0.999	1.000[4]	0.987	0.988[2]	0.978	0.980
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
n=100								
Normal[a]	0.096	0.098	0.098	0.099	0.095	0.100	0.096	0.096
Hetero. Normal	0.159	0.092	0.135	0.165[2]	0.157	0.086	0.125	0.166[2]
t_5	0.702	0.391	0.575	0.709[2]	0.685[2]	0.357	0.550	0.679
Laplace	0.888	0.603	0.881	0.898[2]	0.858	0.540	0.834	0.874[2]
Chi-squared(2)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

In general, the percent decrease in power for modified statistics as observed by White and Macdonald (1980) is consistent with the results here. The superiority of the modified statistics for *F*, in symmetric, leptokurtic distributions, and *W*, *W**, in the remaining distributions, mimics the results to those using the true statistics. However, *F* seems to perform better at $n = 10$ with the highest power for all the non-normal distributions although none of them are significantly higher than the other tests. This might be due mainly from a smaller dip in its empirical level.

Appendices C-1(b) to C-10(b) contain the results for a wider array of error distributions used in Section 4.2 at both a = 0.05 and 0.1. They generally support the conclusions made above. The observation that *F* is superior at $n = 10$ holds. Appendix C-6(b) contains the results for a = 0.1 where the same empirical levels of the tests as those in Table 5-1 apply. Here, *F* has significantly higher power in LC(0.05,5), LC(0.10,5), LC(0.05,7), LC(0.10,7) and Chi-squared(4) – distributions for which *F* never had the highest power before – in addition to those in group I. In Appendix C-1(b) where a = 0.05, *F* has the smallest empirical level and yet it still has significantly higher power in LC(0.05,7), LC(0.10,7), Weibull(0.5) and Lognormal(0,1).

Table 5-2 contains the results for the measures of correspondence between the true and modified statistics. In general, *F* and *W* have the highest r for smaller and larger *n* respectively. However, *F* seems to have the smallest *m* in most cases except for $n = 100$ where *W* dominates.

Table 5-2 Measures of correspondence between true and modified test statistics across different values of n based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1

Distributions	r (correlation)				m(max. abs. dev. over std. error of true statistics)			
	W(W')	*W**	A^2	*F*	*W(W')*	*W**	A^2	*F*
n=10								
Normal[a]	0.171	0.179	0.167	0.193	7.553	7.605	7.316	6.519
Hetero. Normal	0.200	0.221	0.202	0.236	5.969	6.036	6.042	5.593
t_5	0.237	0.255	0.251	0.300	5.188	5.137	5.135	4.624
Laplace	0.298	0.313	0.296	0.341	6.560	6.492	6.718	5.620
Chi-squared(2)	0.382	0.402	0.405	0.438	4.970	1.913	4.962	4.776
Lognormal(0,1)	0.357	0.378	0.403	0.450	4.557	4.513	4.790	4.344
n=20								
Normal[a]	0.485	0.482	0.453	0.475	6.468	6.455	7.364	5.070
Hetero. Normal	0.525	0.522	0.487	0.529	4.357	4.317	4.385	4.108
t_5	0.743	0.739	0.654	0.747	4.245	4.286	5.310	3.236
Laplace	0.726	0.721	0.647	0.723	3.321	3.340	4.121	2.946
Chi-squared(2)	0.655	0.650	0.559	0.690	3.862	3.883	5.133	3.631
Lognormal(0,1)	0.778	0.775	0.701	0.800	3.712	3.732	4.391	3.512
n=35								
Normal[a]	0.642	0.642	0.611	0.641	5.622	5.621	5.653	4.590
Hetero. Normal	0.685	0.685	0.654	0.696	3.330	3.332	3.729	2.940
t_5	0.906	0.906	0.863	0.899	2.879	2.880	3.865	2.588
Laplace	0.881	0.881	0.830	0.878	2.853	2.853	3.186	2.311
Chi-squared(2)	0.830	0.830	0.758	0.858	3.049	3.050	4.008	2.836
Lognormal(0,1)	0.885	0.885	0.796	0.901	2.333	2.333	4.187	2.150
n=50								
Normal[a]	0.731	0.731	0.704	0.718	4.068	4.069	5.302	4.233
Hetero. Normal	0.736	0.736	0.709	0.767	3.173	3.172	3.808	2.920
t_5	0.954	0.954	0.921	0.950	2.023	2.023	3.376	1.556
Laplace	0.931	0.931	0.877	0.929	1.907	1.907	2.845	1.916
Chi-squared(2)	0.855	0.855	0.807	0.890	3.240	3.240	3.235	2.547
Lognormal(0,1)	0.919	0.919	0.852	0.935	2.178	2.178	3.227	2.081
n=100								
Normal[a]	0.869	0.857	0.824	0.817	3.748	2.979	4.737	3.660
Hetero. Normal	0.901	0.852	0.819	0.857	2.021	2.666	2.866	2.357
t_5	0.990	0.987	0.969	0.986	1.019	1.264	1.998	0.841
Laplace	0.972	0.960	0.928	0.960	1.349	1.566	2.146	1.573
Chi-squared(2)	0.936	0.912	0.883	0.937	1.990	2.275	2.609	2.138
Lognormal(0,1)	0.975	0.967	0.925	0.976	1.185	1.350	2.080	1.160

Notes : a. based on 10 000 samples

Comparison across *k*

The results using set 1 in Weisberg (1980) which compares the power results for k = 4, 6, 8 and 10 are given in Table 5-3. For the true statistics, the nominal value of 0.1 is within the 95% confidence interval of the levels of each test. However, the effect on the modified statistics depends on the normality test itself. For *W* and *W'*, the empirical levels are below the nominal value for $k = 4$. As for A^2, the empirical levels are higher than the nominal for $k = 6$, 8 and 10. The empirical levels for *F* are below the nominal for $k = 4$ but they rise above the nominal as *k* increases.

Since the levels of modified statistics are affected by this particular X matrix, the power estimates are also not reflective of their true power. However, examining the pattern of deflation/inflation might be instructive.

For $k = 4$, A^2 has the highest empirical level that still captures the nominal value in its 95% confidence interval. However, it fails to have significant power for any of the distributions. This implies that A^2 has inferior power to the other tests in testing non-normality. No obvious conclusion can be drawn about which of the remaining tests are superior. *W** has the lowest empirical level which implies that its power estimates are slightly deflated compared to the rest. However, its power is still on par with *W*, in Chi-squared(2) and Lognormal, and with the rest in the heteroscedastic normal distribution. As for t_5 and Laplace, *F* seems superior since it is inconceivable that *W** could have significantly higher power even if the power estimates could be adjusted to account for the deflation.

As for the remaining values of *k*, both A^2 and *F* have inflated power estimates. However, both *W* and *W** have significantly higher power than the rest with the Lognormal(0,1) distribution at $k = 6$ and Chi-squared distribution at $k = 8$. This suggests that both these tests retain their superiority in testing distributions that are asymmetric and leptokurtic. The conclusion is less clear for *F* since its empirical levels get highly inflated especially at $k = 10$. However, looking at the difference in the power estimates between *F* and the other tests for t_5 and Laplace distributions, it is conceivable that *F* still retains its superiority in detecting non-normality for distributions that are symmetric and leptokurtic.

Table 5-3 Power comparisons of normality tests on iid observations and OLS residuals across different values of k based on 1000 samples using X=data set 1 from Weisberg (1980) at α=0.1 and n=20

Distributions	iid observations(true statistics)				OLS residuals(modified statistics)			
	W	W^*	A^2	F	W	W^*	A^2	F
<u>k=4</u>								
Normal[a]	0.099	0.097	0.101	0.100	0.092[b]	0.090[b]	0.096	0.092[b]
Hetero. Normal	0.108	0.105	0.106	0.118[4]	0.100	0.099	0.111	0.111[4]
t_5	0.231	0.227	0.252	0.274[1]	0.196	0.196	0.185	0.216[1]
Laplace	0.359	0.353	0.374	0.436[1]	0.245	0.239	0.238	0.287[1]
Chi-squared(2)	0.893[2]	0.893[2]	0.847	0.846	0.578[2]	0.574	0.544	0.541
Lognormal(0,1)	0.971[2]	0.971	0.947	0.946	0.776[3]	0.774	0.745	0.768
<u>k=6</u>								
Normal[a]	0.098	0.095	0.098	0.102	0.102	0.101	0.109[b]	0.113[b]
Hetero. Normal	0.099	0.098	0.104	0.115[4]	0.114	0.111	0.120	0.125[4]
t_5	0.274	0.272	0.266	0.322[1]	0.206	0.203	0.212	0.236[1]
Laplace	0.360	0.350	0.373	0.414[1]	0.248	0.241	0.248	0.293[1]
Chi-squared(2)	0.896[2]	0.896[2]	0.856	0.841	0.477	0.473	0.436	0.478[3]
Lognormal(0,1)	0.961[2]	0.961[2]	0.939	0.932	0.676[2]	0.669	0.631	0.668
<u>k=8</u>								
Normal[a]	0.103	0.102	0.105	0.103	0.105	0.103	0.109[b]	0.113[b]
Hetero. Normal	0.098	0.092	0.102	0.112[2]	0.101	0.098	0.111	0.123[2]
t_5	0.239	0.237	0.241	0.286[1]	0.150	0.145	0.152	0.182[1]
Laplace	0.390	0.381	0.398	0.443[1]	0.198	0.196	0.206	0.234[1]
Chi-squared(2)	0.899	0.900[2]	0.848	0.844	0.336[3]	0.335	0.315	0.328
Lognormal(0,1)	0.974[2]	0.972	0.945	0.949	0.478	0.474	0.463	0.504[1]
<u>k=10</u>								
Normal[a]	0.103	0.102	0.103	0.104	0.105	0.103	0.111[b]	0.133[b]
Hetero. Normal	0.133[4]	0.127	0.121	0.126	0.100	0.099	0.121	0.127[2]
t_5	0.247	0.244	0.239	0.288[1]	0.153	0.149	0.163	0.211[1]
Laplace	0.340	0.338	0.366	0.417[1]	0.194	0.190	0.188	0.246[1]
Chi-squared(2)	0.897[2]	0.896	0.859	0.857	0.258	0.254	0.236	0.289[1]
Lognormal(0,1)	0.942[2]	0.942[2]	0.909	0.908	0.376	0.365	0.365	0.430[1]

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

Table 5-4 gives the results for the measures of correspondence between the true and modified statistics. Here, *F* has the highest r and smallest *m* for the almost all of the error distributions and values of *k*. This shows that using the modified statistic for *F* is the closest to the true statistic among all normality tests considered for this configuration of X. Another interesting result is that as *k* increases, the correspondence between the true and modified statistics decreases. This shows that *k* also must be considered in the use of modified statistics to test normality of residuals.

5.2.3 Conclusion

From the results above, it seems that the modified statistics can indeed be used to detect non-normality of residuals albeit with some loss of power due to the presence of supernormality. In particular, *F* has, in general, the highest measures of correspondence between the true and modified statistic across the combinations of *n, k* and *X* considered in this chapter. However, one has to be careful about the levels of the normality tests and also generalizing any results since they do vary according to the specific combination of *n*, *k* and *X*.

From the two specific formulation of the X matrix, the conclusions regarding the superiority of *F*, *W* and *W** seem unchanged from those where the observations are independent. In the case where X is generated from a uniform distribution for *n* as small as 10, the pattern of superiority seems to take on a different characterization. *F* is now superior across a wider range of distributions besides those that are symmetric and leptokurtic.

Table 5-4 Measures of correspondence between true and modified test statistics across different values of k based on 1000 samples using X=data set 1 from Weisberg (1980) at α=0.1 and n=20

Distributions	r (correlation)				*m*(max. abs. dev. over std. error of true statistics)			
	W	*W**	A^2	*F*	*W*	*W**	A^2	*F*
<u>k=4</u>								
Normal[a]	0.439	0.436	0.415	0.441	6.463	6.471	7.942	5.550
Hetero. Normal	0.446	0.442	0.400	0.467	4.416	4.381	4.720	3.968
t_5	0.771	0.766	0.708	0.778	3.697	3.727	5.230	3.442
Laplace	0.694	0.690	0.609	0.701	4.626	4.659	4.057	3.432
Chi-squared(2)	0.687	0.683	0.606	0.721	3.953	3.976	4.245	3.528
Lognormal(0,1)	0.786	0.783	0.683	0.815	2.866	2.884	3.568	2.690
<u>k=6</u>								
Normal[a]	0.300	0.300	0.258	0.317	7.833	7.786	7.794	7.203
Hetero. Normal	0.337	0.334	0.267	0.348	5.581	5.675	5.766	5.277
t_5	0.640	0.635	0.518	0.636	3.753	3.771	5.609	3.407
Laplace	0.545	0.541	0.475	0.564	4.480	4.504	4.513	4.105
Chi-squared(2)	0.603	0.600	0.514	0.630	4.802	4.795	4.697	4.698
Lognormal(0,1)	0.664	0.661	0.583	0.695	4.925	4.930	5.391	4.876
<u>k=8</u>								
Normal[a]	0.167	0.163	0.144	0.191	8.892	8.813	8.500	8.202
Hetero. Normal	0.236	0.232	0.212	0.254	6.551	6.539	5.636	5.503
t_5	0.387	0.382	0.329	0.400	5.510	5.525	5.151	5.092
Laplace	0.416	0.411	0.357	0.440	5.619	5.635	5.097	5.002
Chi-squared(2)	0.440	0.435	0.370	0.479	4.880	4.871	4.462	4.764
Lognormal(0,1)	0.533	0.529	0.466	0.562	4.409	4.427	4.974	4.109
<u>k=10</u>								
Normal[a]	0.135	0.131	0.112	0.160	9.543	9.522	9.662	8.214
Hetero. Normal	0.166	0.162	0.142	0.196	6.012	5.968	5.434	5.729
t_5	0.334	0.331	0.283	0.341	6.300	6.336	6.406	4.865
Laplace	0.320	0.316	0.291	0.353	6.052	6.061	5.530	5.362
Chi-squared(2)	0.325	0.321	0.288	0.356	5.195	5.189	4.941	5.045
Lognormal(0,1)	0.447	0.443	0.433	0.493	4.595	4.609	5.135	4.311

Notes : a. based on 10 000 samples

Chapter 6 Summary and Discussion

From the power studies, it is clear that both W^* and F are needed in an arsenal of normality tests to detect non-normality over a wide range of distributions. However, F possesses certain advantages over W^*, in particular, and existing normality tests, in general.

Firstly, it provides a direct, graphical depiction of the deviation from non-normality through the estimated density plots. Aspects of non-normality like asymmetry and significant tail behavior show up more clearly in the estimated density plot than in a normal probability plot.

Secondly, the power of F is superior when the underlying distribution is symmetric and leptokurtic. This is precisely the situation where a formal test of normality is most needed to distinguish one's data from normality.

The above two advantages carry over when OLS residuals are used in place of the true errors in testing normality of residuals. Moreover, the superiority in power has added significance since it is well known that the OLS estimates obtained are very inefficient when the underlying distribution is leptokurtic or long-tailed. However, as the results have shown, the levels of the test might differ from the nominal depending on the particular combination of n, k and V. As there could be a myriad of combinations between these factors, more research needs to be done to see how to make the results more generalizable.

As for other directions for further research, more work needs to be done to either get an asymptotic or approximate distribution for F so as to do away with the need to consult tables for the critical values. Further, the ideas behind F could be generalized to the multivariate case for a multivariate normality test.

On a more general level, it seems the two advantages might even translate to tests for other distributions based on the main ideas used in deriving F. Terrell (1985) has outlined the basic arguments in creating tests for membership in the Pearson family by creating pseudo-information functionals that are extremal for certain distributions. This mimics the extremal property of I_F in normal distributions. Although Terrell cautioned that the derived tests will not be based on maximum likelihood estimates of the corresponding parameters, the strong advantage in this approach is the resulting density estimator that is far superior to existing graphical methods.

Appendix A Critical Values of the Normal Information Statistic, F

n\α	0.500	0.250	0.200	0.150	0.100	0.050	0.025	0.020	0.010
3	1.5901	1.8664	1.9512	2.0470	2.1727	2.3666	2.5381	2.5894	2.7494
4	1.6925	1.9937	2.0853	2.1916	2.3318	2.5519	2.7517	2.8123	3.0003
5	1.7919	2.1172	2.2155	2.3317	2.4859	2.7312	2.9582	3.0277	3.2424
6	1.8883	2.2369	2.3416	2.4675	2.6350	2.9045	3.1576	3.2356	3.4761
7	1.9820	2.3530	2.4639	2.5990	2.7795	3.0721	3.3502	3.4364	3.7015
8	2.0729	2.4655	2.5825	2.7264	2.9193	3.2342	3.5362	3.6302	3.9189
9	2.1611	2.5746	2.6974	2.8498	3.0546	3.3909	3.7159	3.8173	4.1286
10	2.2467	2.6804	2.8089	2.9693	3.1856	3.5425	3.8894	3.9980	4.3308
11	2.3298	2.7829	2.9169	3.0851	3.3124	3.6889	4.0569	4.1723	4.5258
12	2.4104	2.8824	3.0216	3.1973	3.4351	3.8305	4.2186	4.3405	4.7138
13	2.4888	2.9788	3.1232	3.3060	3.5540	3.9674	4.3747	4.5029	4.8951
14	2.5648	3.0722	3.2216	3.4112	3.6689	4.0997	4.5254	4.6596	5.0698
15	2.6386	3.1629	3.3170	3.5132	3.7803	4.2276	4.6709	4.8108	5.2382
16	2.7102	3.2508	3.4095	3.6120	3.8880	4.3512	4.8113	4.9566	5.4005
17	2.7798	3.3360	3.4993	3.7077	3.9923	4.4707	4.9469	5.0974	5.5569
18	2.8474	3.4187	3.5863	3.8005	4.0933	4.5862	5.0778	5.2332	5.7077
19	2.9131	3.4988	3.6706	3.8903	4.1911	4.6979	5.2041	5.3642	5.8530
20	2.9769	3.5766	3.7525	3.9774	4.2857	4.8059	5.3260	5.4906	5.9930
21	3.0388	3.6521	3.8319	4.0618	4.3774	4.9103	5.4437	5.6126	6.1279
22	3.0990	3.7253	3.9089	4.1437	4.4662	5.0112	5.5574	5.7303	6.2580
23	3.1576	3.7963	3.9836	4.2230	4.5522	5.1088	5.6671	5.8440	6.3834
24	3.2145	3.8653	4.0561	4.2999	4.6355	5.2033	5.7731	5.9536	6.5042
25	3.2698	3.9323	4.1265	4.3745	4.7162	5.2946	5.8754	6.0595	6.6207
26	3.3236	3.9973	4.1948	4.4469	4.7944	5.3830	5.9743	6.1617	6.7331
27	3.3760	4.0605	4.2612	4.5172	4.8703	5.4686	6.0698	6.2604	6.8414
28	3.4269	4.1218	4.3256	4.5853	4.9438	5.5514	6.1621	6.3557	6.9459
29	3.4765	4.1815	4.3883	4.6515	5.0151	5.6315	6.2514	6.4478	7.0467
30	3.5248	4.2394	4.4491	4.7158	5.0843	5.7092	6.3376	6.5368	7.1440
31	3.5718	4.2958	4.5083	4.7782	5.1515	5.7844	6.4211	6.6228	7.2380
32	3.6177	4.3506	4.5658	4.8388	5.2167	5.8573	6.5018	6.7060	7.3287
33	3.6623	4.4040	4.6218	4.8978	5.2800	5.9280	6.5800	6.7865	7.4163
34	3.7059	4.4559	4.6763	4.9552	5.3415	5.9966	6.6556	6.8644	7.5009
35	3.7484	4.5065	4.7293	5.0109	5.4013	6.0631	6.7289	6.9398	7.5828
36	3.7899	4.5558	4.7809	5.0652	5.4594	6.1277	6.7999	7.0128	7.6620
37	3.8304	4.6038	4.8313	5.1181	5.5160	6.1904	6.8688	7.0836	7.7386
38	3.8699	4.6506	4.8803	5.1696	5.5710	6.2514	6.9355	7.1522	7.8128
39	3.9086	4.6963	4.9281	5.2198	5.6246	6.3107	7.0004	7.2188	7.8846
40	3.9463	4.7409	4.9748	5.2688	5.6769	6.3683	7.0633	7.2834	7.9543
41	3.9833	4.7845	5.0204	5.3165	5.7278	6.4244	7.1245	7.3461	8.0219
42	4.0195	4.8270	5.0649	5.3632	5.7774	6.4790	7.1840	7.4072	8.0876
43	4.0549	4.8686	5.1083	5.4087	5.8259	6.5323	7.2419	7.4665	8.1513
44	4.0896	4.9093	5.1509	5.4532	5.8732	6.5842	7.2983	7.5242	8.2133
45	4.1236	4.9491	5.1925	5.4968	5.9195	6.6349	7.3532	7.5805	8.2736
46	4.1570	4.9881	5.2332	5.5394	5.9647	6.6844	7.4068	7.6354	8.3324
47	4.1897	5.0263	5.2731	5.5811	6.0089	6.7328	7.4591	7.6889	8.3896

n\α	0.500	0.250	0.200	0.150	0.100	0.050	0.025	0.020	0.010
48	4.2218	5.0638	5.3121	5.6220	6.0523	6.7801	7.5101	7.7411	8.4455
49	4.2534	5.1006	5.3505	5.6621	6.0947	6.8264	7.5601	7.7922	8.5001
50	4.2844	5.1366	5.3881	5.7014	6.1364	6.8717	7.6089	7.8421	8.5534
51	4.3150	5.1721	5.4250	5.7400	6.1772	6.9161	7.6568	7.8910	8.6056
52	4.3450	5.2069	5.4613	5.7779	6.2173	6.9597	7.7037	7.9390	8.6567
53	4.3746	5.2412	5.4970	5.8152	6.2568	7.0026	7.7497	7.9860	8.7068
54	4.4038	5.2749	5.5321	5.8519	6.2955	7.0446	7.7948	8.0321	8.7559
55	4.4325	5.3082	5.5667	5.8880	6.3337	7.0860	7.8392	8.0775	8.8042
56	4.4608	5.3409	5.6007	5.9236	6.3712	7.1267	7.8829	8.1221	8.8517
57	4.4888	5.3732	5.6343	5.9587	6.4082	7.1668	7.9258	8.1660	8.8985
58	4.5165	5.4051	5.6674	5.9933	6.4448	7.2063	7.9682	8.2093	8.9445
59	4.5438	5.4365	5.7000	6.0275	6.4808	7.2453	8.0100	8.2519	8.9900
60	4.5707	5.4676	5.7323	6.0612	6.5164	7.2838	8.0512	8.2941	9.0348
61	4.5974	5.4983	5.7642	6.0945	6.5515	7.3218	8.0919	8.3357	9.0792
62	4.6239	5.5287	5.7957	6.1275	6.5863	7.3594	8.1322	8.3769	9.1230
63	4.6500	5.5588	5.8269	6.1602	6.6207	7.3967	8.1720	8.4176	9.1665
64	4.6759	5.5886	5.8578	6.1925	6.6548	7.4335	8.2115	8.4579	9.2095
65	4.7016	5.6181	5.8884	6.2245	6.6886	7.4700	8.2506	8.4979	9.2522
66	4.7271	5.6473	5.9187	6.2563	6.7220	7.5062	8.2894	8.5376	9.2945
67	4.7524	5.6763	5.9487	6.2878	6.7552	7.5421	8.3279	8.5770	9.3366
68	4.7774	5.7051	5.9785	6.3190	6.7882	7.5778	8.3661	8.6161	9.3785
69	4.8023	5.7337	6.0081	6.3501	6.8209	7.6132	8.4041	8.6550	9.4201
70	4.8271	5.7621	6.0375	6.3809	6.8534	7.6483	8.4419	8.6937	9.4615
71	4.8516	5.7903	6.0667	6.4115	6.8857	7.6833	8.4795	8.7322	9.5028
72	4.8761	5.8184	6.0957	6.4420	6.9178	7.7181	8.5169	8.7705	9.5439
73	4.9003	5.8463	6.1245	6.4723	6.9498	7.7528	8.5541	8.8087	9.5849
74	4.9245	5.8740	6.1532	6.5024	6.9816	7.7872	8.5912	8.8468	9.6258
75	4.9485	5.9017	6.1817	6.5325	7.0132	7.8216	8.6282	8.8847	9.6667
76	4.9724	5.9292	6.2101	6.5623	7.0448	7.8558	8.6651	8.9226	9.7075
77	4.9962	5.9565	6.2384	6.5921	7.0762	7.8899	8.7019	8.9603	9.7482
78	5.0199	5.9838	6.2665	6.6217	7.1075	7.9239	8.7386	8.9980	9.7889
79	5.0435	6.0110	6.2946	6.6513	7.1387	7.9578	8.7752	9.0357	9.8296
80	5.0670	6.0380	6.3225	6.6807	7.1698	7.9916	8.8117	9.0732	9.8702
81	5.0904	6.0650	6.3503	6.7100	7.2008	8.0254	8.8482	9.1108	9.9109
82	5.1137	6.0919	6.3781	6.7393	7.2317	8.0590	8.8847	9.1483	9.9515
83	5.1369	6.1187	6.4057	6.7685	7.2625	8.0926	8.9211	9.1857	9.9922
84	5.1601	6.1454	6.4333	6.7975	7.2933	8.1262	8.9574	9.2232	10.0329
85	5.1832	6.1721	6.4608	6.8266	7.3240	8.1596	8.9937	9.2606	10.0736
86	5.2062	6.1987	6.4882	6.8555	7.3546	8.1930	9.0300	9.2979	10.1143
87	5.2291	6.2252	6.5155	6.8844	7.3851	8.2264	9.0662	9.3353	10.1550
88	5.2519	6.2516	6.5428	6.9131	7.4156	8.2596	9.1024	9.3726	10.1957

n\α	0.500	0.250	0.200	0.150	0.100	0.050	0.025	0.020	0.010
89	5.2747	6.2780	6.5700	6.9419	7.4460	8.2929	9.1385	9.4099	10.2364
90	5.2974	6.3043	6.5971	6.9705	7.4764	8.3260	9.1746	9.4471	10.2771
91	5.3201	6.3305	6.6241	6.9991	7.5066	8.3591	9.2106	9.4843	10.3177
92	5.3426	6.3567	6.6511	7.0276	7.5368	8.3921	9.2466	9.5214	10.3584
93	5.3651	6.3828	6.6780	7.0560	7.5669	8.4250	9.2825	9.5585	10.3991
94	5.3875	6.4088	6.7048	7.0843	7.5969	8.4579	9.3183	9.5955	10.4397
95	5.4098	6.4348	6.7315	7.1126	7.6269	8.4907	9.3541	9.6325	10.4803
96	5.4320	6.4607	6.7582	7.1408	7.6567	8.5233	9.3898	9.6694	10.5208
97	5.4542	6.4865	6.7848	7.1689	7.6865	8.5559	9.4254	9.7062	10.5612
98	5.4763	6.5122	6.8112	7.1969	7.7161	8.5884	9.4608	9.7429	10.6016
99	5.4982	6.5378	6.8376	7.2247	7.7457	8.6208	9.4962	9.7795	10.6419
100	5.5201	6.5633	6.8639	7.2525	7.7751	8.6530	9.5315	9.8159	10.6820
105	5.6281	6.6895	6.9939	7.3897	7.9205	8.8122	9.7054	9.9960	10.8808
110	5.7333	6.8127	7.1207	7.5234	8.0620	8.9670	9.8745	10.1712	11.0744
115	5.8351	6.9321	7.2438	7.6527	8.1987	9.1161	10.0370	10.3394	11.2607
120	5.9331	7.0471	7.3622	7.7768	8.3295	9.2582	10.1913	10.4990	11.4373
125	6.0265	7.1568	7.4753	7.8945	8.4534	9.3919	10.3355	10.6481	11.6022
130	6.1151	7.2606	7.5822	8.0053	8.5694	9.5162	10.4685	10.7850	11.7532
135	6.1983	7.3578	7.6825	8.1084	8.6769	9.6301	10.5890	10.9087	11.8889
140	6.2761	7.4481	7.7758	8.2035	8.7755	9.7333	10.6964	11.0183	12.0083
145	6.3483	7.5314	7.8619	8.2906	8.8652	9.8256	10.7907	11.1137	12.1112
150	6.4154	7.6078	7.9411	8.3698	8.9464	9.9077	10.8726	11.1955	12.1981
155	6.4778	7.6781	8.0141	8.4421	9.0200	9.9809	10.9437	11.2654	12.2707
160	6.5365	7.7430	8.0818	8.5088	9.0879	10.0473	11.0064	11.3258	12.3319
165	6.5928	7.8043	8.1460	8.5719	9.1522	10.1098	11.0644	11.3804	12.3859
170	6.6486	7.8638	8.2088	8.6340	9.2161	10.1723	11.1226	11.4342	12.4383
175	6.7062	7.9244	8.2731	8.6985	9.2836	10.2400	11.1870	11.4938	12.4965
180	6.7683	7.9894	8.3425	8.7697	9.3598	10.3192	11.2654	11.5670	12.5696
185	6.8384	8.0630	8.4212	8.8528	9.4507	10.4175	11.3671	11.6635	12.6688
190	6.9207	8.1499	8.5145	8.9538	9.5634	10.5438	11.5031	11.7948	12.8073
195	7.0200	8.2561	8.6286	9.0801	9.7065	10.7089	11.6863	11.9745	13.0007
200	7.1417	8.3883	8.7704	9.2399	9.8896	10.9248	11.9316	12.2180	13.2668

Appendix B Computational Details

All programs are written in FORTRAN and double precision representation is used.

B.1 Details of programming for F

The program to get F and the resulting density estimate is listed in Appendix B.3. The subroutine rtsec from Press et al. (1992) was used to implement the secant method in searching for the value of h and gives $I_F = 1$.

B.2 Details of power study

Critical values used were obtained from Shapiro (1980) for A^2, W and W'. The only exception is for W' at n=100 where the values are taken as an average of the values available in Pearson et al. (1977). Critical values for F were those that were obtained from the original simulation (unsmoothed). A full list of the values used is available in Table B-1.

Table B-1 List of critical values used in power study

n	α=0.05 (A^2=0.752[a])		α=0.10 (A^2=0.631[a])	
	$W(W')^a$	F^b	$W(W')^a$	F^b
10	0.842	3.6747	0.869	3.2759
20	0.905	4.8394	0.920	4.3292
35	-	-	0.944	5.3879
50	0.947	6.8451	0.955	6.1177
70	0.968	7.6560	0.973	6.8537
100	0.9743	8.6528	0.9787	7.7811

Sources: a. Shapiro (1980) except for n=100 which is taken from Pearson et al. (1977)

b. Using simulated critical values (unsmoothed)

The pseudo-random number subroutines used were those available from IMSL (STAT/LIBRARY FORTRAN Subroutines for Statistical Analysis Version 1.0, April 1987). For distributions that do not have a subroutine of their own like the Laplace, the inverse cumulative distribution function (cdf) method is used where a uniform deviate is generated first before using the inverse cdf to get the required pseudo-random number.

The FORTRAN code for obtaining the test statistics for K^2, W, W' and A^2 were adapted from a collection of FORTRAN subroutines entitled ‘Tests of Composite Distributional Hypotheses for the Analysis of Biological & Environmental Data’ available at Statlib (http://lib.stat.cmu.edu) that was written by Paul Johnson. The subroutine for W^* is available in IMSL.

B.3 Program Listing for *F*

```
c
c       FisherInfoTest.f
c
c       date : Aug 20, 98
c
c       aim  : Calculate test statistic F and provide density estimate in
c               original units (optional)
c
c       input : data in ASCII file
c
c       output :
c       a.  test statistic F
c        b.  ASCII file containing x (1st column) and density estimate(2nd
c           column) in original units (optional)
c
c

c
c     declaration of variables
c
      integer ndata, count
      double precision samave, samstdev, kernel, FisherInfo, h
      double precision mmm(1:200), sdata(1:200), root(1:200)
      double precision sum, sumsq, log2pi
      double precision h1, h2, fl1, f2, root11(1:200)
      double precision xl, swap, dx, factor
      double precision min, max, increment, x, density, rplot(1:200)
      logical Fisher1
      character PrintPlot
      data log2pi, big, Maxplot /1.83787706640934548356, 1.0D20, 128/
      data top, bottom /4, -4/
      character DataFile*20
      character PlotFileName*20

c
c     get data and determine minimum and maximum


      write(*,*)'Enter name of data file '
      read(*,*)DataFile
      open(unit=7, file=DataFile, status='old')
      ndata=0
      min=big
      max=-big
5     read(7,*, END=7)mmm(ndata+1)
      if (mmm(ndata+1).lt.min) min=mmm(ndata+1)
      if (mmm(ndata+1).gt.min) max=mmm(ndata+1)
          ndata=ndata+1
          go to 5

c
```

```
c     get mean and standard deviation using subroutine avestdev
c

7     call avestdev(mmm,ndata,samave,samstdev)

c
c     standardize data
c

      do 10 i = 1, ndata
         sdata(i)=(mmm(i)-samave)/samstdev
10    continue


c
c     get h where Fisher Information is 1
c
      h1=0.5
      h2= 2.0

20    call GetFisherInfo(ndata,h1,sdata,root11,fl1)
      call GetFisherInfo(ndata,h2,sdata,root11,f2)

      if (abs(fl1).lt.abs(f2)) then
         h=h1
         xl=h2
         swap=fl1
         fl1=f2
         f2=swap
      else
         xl=h1
         h=h2
      endif
      count=1
      Fisher1=.false.
30    if (.not.Fisher1) then
         count=count+1
         dx=(xl-h)*f2/(f2-fl1)
         xl=h
         fl1=f2
         h=h+dx
         if (h.lt.0) h=exp(h)
         call GetFisherInfo(ndata,h,sdata,root,f2)
         if (abs(dx).lt.0.0000000000001.or.f2.eq.0.) Fisher1=.true.
         go to 30
      endif
      FisherInfo=f2+1.0
      sum=0
      do 50 i=1, ndata
         sum = sum + log(root(i))
50    continue
      sumsq=4/(FisherInfo*h*h+4)
      sum = 4*sum - ndata*(2*log(sumsq)-1-log2pi)
      if (sum.le.0.) then
        h1=0.5
         h2=2.1
         go to 20
      endif
      write(*,*)'test statistic, F =', sum
c
c     print density plot
c
      write(*,*)'Plot density?(y/n)'
      read(*,*)PrintPlot
      if (PrintPlot.eq.'y') then
         write(*,*)'Enter name of file to store density'
         read(*,*)PlotFileName
         open(unit=8, file=PlotFileName, status='new')
         increment=(top-bottom)/Maxplot
         x=bottom
         do 200 i=1, Maxplot
```

```
            density=0.0
            do 100 j =1, ndata
               density=density+kernel(x,sdata(j),h)/root(j)
100         continue
            rplot(i)=density/ndata
            x=x+increment
200      continue
         factor=1/samstdev/sumsq
         x=bottom
         do 300 i=1, Maxplot
            write(8,250)samave+samstdev*x, rplot(i)*rplot(i)*factor
250         format(f10.4,5X,e15.10)
            x=x+increment
300      continue
         close(unit=8)
      endif
      close(unit=7)
      end

c
c     subroutine avestdev to calculate average and standard deviation
c
      SUBROUTINE AVESTDEV(DATA,N,AVE,STDEV)
      DOUBLE PRECISION DATA(N), AVE, STDEV, S
      AVE=0.0
      STDEV=0.0
      DO 11 J=1,N
        AVE=AVE+DATA(J)
11    CONTINUE
      AVE=AVE/N
      DO 12 J=1,N
        S=DATA(J)-AVE
        STDEV=STDEV+S*S
12    CONTINUE
      STDEV=STDEV/N
      STDEV=SQRT(STDEV)
      RETURN
      END

c
c     function kernel to get kernel values
c
      double precision function kernel(x,y,z)
      double precision x, y, z
      kernel=exp(-abs(x-y)/z)/z/2.0
      return
      end

c
c     subroutine to get Fisher Information
c

      subroutine GetFisherInfo(gndata,gh,gsdata,groot,gFisherInfo)
      logical converge, convergerest
      integer gndata, i, j, c, count
      double precision gh, gsdata(gndata), groot(gndata), gFisherInfo
      double precision kernel, density, datum, groot2(1:200)
      double precision oldroot(1:200), gsumsq, x

      c=1
      count=1
      converge = .false.

      do 1 i=1, gndata
         oldroot(i) = 0.0
1     continue

         do 10 i=1, gndata
            density=0.0
            do 5 j=1, gndata
               density = density + kernel(gsdata(i),gsdata(j),gh)
```

```
5           continue
            groot(i)=sqrt(density/gndata)
10       continue
15     if (.not.converge) then
         count = count + 1
         do 30 i=1, gndata
            density = 1.0/2.0/gh/groot(i)
            datum = gsdata(i)
            do 20 j=1, gndata
            if (i.ne.j) then
               density = density + kernel(datum,gsdata(j),gh)/groot(j)
            endif
20          continue
            groot2(i)=density/gndata
30       continue
         do 40 i=1, gndata
            groot(i) = (groot(i)+groot2(i))/2.0
40       continue
         if (abs(oldroot(c)-groot(c)).lt.0.00000000000001) then
            c = c + 1
            convergerest=.true.
            do 50 i=c, gndata
             if (abs(oldroot(c)-groot(c)).gt.0.00000000000001) then
                  convergerest=.false.
             endif
50          continue
            if (convergerest) converge=.true.
         endif
         do 60 i=1, gndata
            oldroot(i) = groot(i)
60       continue
         go to 15
      endif
      gsumsq=0
      do 80 i=1, gndata-1
         gsumsq = gsumsq + 1/groot(i)/groot(i)
         do 70 j=i+1, gndata
            x = abs(gsdata(i)-gsdata(j))/gh
            gsumsq = gsumsq + 2*exp(-x)*(1+x)/groot(i)/groot(j)
70    continue
80    continue
      gsumsq = gsumsq + 1/groot(gndata)/groot(gndata)
      gsumsq = gsumsq/4/gndata/gndata/gh
      gFisherInfo = 4*(1/gsumsq -1)/gh/gh
      gFisherInfo = gFisherInfo - 1.0
      return
      end
```

Appendix C Results from Power Study

This appendix contains the results on the power comparisons of normality tests on iid observations and the resulting OLS residuals. The power estimates are obtained based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution. The combination of *n* and α for each table is given below:

n	α	Table
10	0.05	C-1
20	0.05	C-2
50	0.05	C-3
70	0.05	C-4
100	0.05	C-5
10	0.10	C-6
20	0.10	C-7
50	0.10	C-8
70	0.10	C-9
100	0.10	C-10

Note: To make it easier to assess the results of the power study for each table, the following system of notation is adopted:

a. A superscript appears on the estimate of the test with the highest power.

b. The number in the superscript denote the number of tests, including the one with the highest power, that are significantly better than the rest. Hence, a ‘1’ in the superscript would reflect that the test, with the highest power, has significantly higher power than the rest while a ‘4’ in the superscript would mean that all the tests have the same power. Cochran’s Q is used to determine if the power estimates of all the tests are the same while McNemar’s test is used for multiple comparisons when the hypothesis of equal power for all tests is rejected. Refer to Section 4.2.1 for more details.

c. The superscript notation will not be used for cases where all the estimates reflect the same numerical value.

Table C-1 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.05 and n=10

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	A^2	*F*	*W*	*W**	A^2	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.046	0.051	0.051	0.051	0.036[b]	0.040[b]	0.039[b]	0.029[b]
t_{10}	0.068	0.072^{4}	0.070	0.071	0.043	0.046^{4}	0.046^{4}	0.045
Logistic	0.062	0.070	0.068	0.073^{4}	0.046	0.053^{4}	0.046	0.045
SC(0.05, 9)	0.113	0.117	0.118	0.120^{4}	0.046	0.052	0.051	0.060^{3}
SC(0.10, 9)	0.165	0.168	0.164	0.183^{1}	0.063	0.073	0.074	0.074^{3}
t_4	0.125	0.142	0.129	0.150^{2}	0.052	0.058^{4}	0.056	0.056
SC(0.05, 25)	0.201	0.210^{4}	0.206	0.208	0.083	0.093	0.093	0.107^{1}
Laplace	0.149	0.153	0.156	0.166^{2}	0.065	0.074	0.065	0.078^{2}
SC(0.10, 25)	0.280	0.290	0.288	0.303^{1}	0.091	0.099	0.103	0.110^{2}
t_2	0.281	0.293	0.288	0.311^{1}	0.105	0.114	0.108	0.124^{2}
Cauchy	0.561	0.576	0.587	0.595^{2}	0.238	0.254	0.247	0.279^{1}
<u>II. Sym., platy</u>								
Beta(1,1)	0.088	0.086	0.093^{3}	0.067	0.033	0.035	0.044^{1}	0.031
Beta(1.5, 1.5)	0.036	0.038	0.044^{3}	0.033	0.031	0.033^{4}	0.030	0.032
Beta(2,2)	0.032	0.031	0.034^{4}	0.028	0.027	0.026	0.031^{4}	0.030
<u>III. Asym., lep</u>.								
Weibull(2)	0.066	0.074^{4}	0.071	0.068	0.028	0.032^{4}	0.030	0.032
LC(0.05, 3)	0.082	0.090^{2}	0.079	0.088	0.047	0.057	0.051	0.060^{3}
LC(0.10, 3)	0.104	0.117	0.113	0.118^{3}	0.041	0.044	0.046	0.053^{3}
LC(0.20, 3)	0.143	0.154^{4}	0.154^{4}	0.146	0.044	0.051	0.053^{4}	0.052
Chi-squared(10)	0.131	0.139	0.144^{3}	0.132	0.062	0.070^{3}	0.061	0.068
LC(0.05, 5)	0.295	0.305	0.281	0.307^{3}	0.091	0.103	0.095	0.104^{4}
LC(0.10, 5)	0.384	0.397^{4}	0.384	0.396	0.128	0.138	0.127	0.144^{2}
LC(0.05, 7)	0.420	0.424	0.412	0.427^{3}	0.135	0.149	0.141	0.163^{1}
LC(0.20, 5)	0.457	0.468	0.486^{1}	0.464	0.088	0.100	0.092	0.103^{3}
LC(0.10, 7)	0.609	0.612	0.620^{4}	0.616	0.168	0.176	0.169	0.194^{1}
LC(0.20, 7)	0.776	0.781	0.792^{2}	0.768	0.133	0.141	0.154	0.155^{2}
Chi-squared(4)	0.225	0.236^{1}	0.223	0.210	0.063	0.066^{4}	0.066^{4}	0.066^{4}
Chi-squared(2)	0.427	0.438^{1}	0.413	0.387	0.098	0.106^{4}	0.103	0.104
Chi-squared(1)	0.709	0.715^{1}	0.681	0.649	0.163	0.180^{2}	0.166	0.177
Weibull(0.5)	0.892	0.895^{2}	0.878	0.857	0.227	0.148	0.248	0.275^{1}
Lognormal(0,1)	0.620	0.633^{1}	0.604	0.580	0.166	0.174	0.180	0.193^{1}
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.045	0.048^{4}	0.046	0.040	0.043	0.044	0.045^{4}	0.044
Beta(2,1)	0.110	0.112^{3}	0.108	0.088	0.044	0.041^{4}	0.041^{4}	0.038

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

Table C-2 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.05 and n=20

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	A^2	*F*	*W*	*W**	A^2	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.052	0.049	0.049	0.052	0.046	0.044	0.045	0.047
t_{10}	0.102	0.097	0.090	0.110[3]	0.073	0.071	0.063	0.080[3]
Logistic	0.110	0.108	0.100	0.138[1]	0.085	0.084	0.088	0.096[3]
SC(0.05, 9)	0.191	0.185	0.166	0.203[2]	0.154	0.146	0.125	0.165[2]
SC(0.10, 9)	0.294	0.286	0.243	0.320[1]	0.207	0.198	0.170	0.232[1]
t_4	0.250	0.246	0.232	0.286[1]	0.160	0.157	0.158	0.187[1]
SC(0.05, 25)	0.374	0.372	0.354	0.386[1]	0.301	0.293	0.274	0.321[1]
Laplace	0.289	0.283	0.286	0.343[1]	0.191	0.182	0.174	0.219[1]
SC(0.10, 25)	0.541	0.536	0.508	0.561[1]	0.402	0.397	0.369	0.433[1]
t_2	0.555	0.541	0.553	0.615[1]	0.413	0.402	0.381	0.458[1]
Cauchy	0.886	0.883	0.899	0.914[1]	0.745	0.738	0.726	0.789[1]
<u>II. Sym., platy</u>								
Beta(1,1)	0.184[2]	0.182	0.162	0.096	0.073[3]	0.073[3]	0.072	0.044
Beta(1.5, 1.5)	0.094[3]	0.091	0.093	0.042	0.061	0.059	0.066[3]	0.033
Beta(2,2)	0.053	0.052	0.057[3]	0.034	0.054	0.054	0.059[3]	0.032
<u>III. Asym., lep</u>.								
Weibull(2)	0.146[2]	0.145	0.123	0.120	0.107	0.1404	0.113[4]	0.099
LC(0.05, 3)	0.192	0.188	0.167	0.209[1]	0.135	0.132	0.111	0.140[3]
LC(0.10, 3)	0.265	0.259	0.234	0.281[1]	0.165	0.154	0.139	0.169[3]
LC(0.20, 3)	0.266[2]	0.258	0.265	0.243	0.135	0.132	0.153[1]	0.126
Chi-squared(10)	0.240[1]	0.235	0.202	0.209	0.144[2]	0.142	0.114	0.128
LC(0.05, 5)	0.554	0.547	0.483	0.562[2]	0.358	0.353	0.293	0.376[1]
LC(0.10, 5)	0.779	0.773	0.728	0.786[2]	0.499	0.490	0.459	0.509[2]
LC(0.05, 7)	0.651	0.650	0.645	0.657[1]	0.558	0.553	0.505	0.575[1]
LC(0.20, 5)	0.852	0.846	0.856[3]	0.822	0.475	0.471	0.486[3]	0.462
LC(0.10, 7)	0.880[4]	0.880[4]	0.877	0.878	0.697	0.693	0.663	0.703[3]
LC(0.20, 7)	0.986[4]	0.986[4]	0.984	0.986[4]	0.656[3]	0.651	0.647	0.636
Chi-squared(4)	0.541[2]	0.537	0.472	0.459	0.309[2]	0.307	0.271	0.277
Chi-squared(2)	0.835[2]	0.834	0.764	0.758	0.482[1]	0.475	0.417	0.431
Chi-squared(1)	0.988[2]	0.986	0.974	0.972	0.704[1]	0.698	0.679	0.689
Weibull(0.5)	0.997[4]	0.997[4]	0.995	0.993	0.858[3]	0.851	0.829	0.842
Lognormal(0,1)	0.930[2]	0.927	0.901	0.897	0.686[1]	0.680	0.636	0.660
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.075[3]	0.073	0.069	0.044	0.043	0.043	0.046[3]	0.036
Beta(2,1)	0.295[2]	0.290	0.255	0.199	0.135[3]	0.131	0.122	0.104

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

Table C-3 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.05 and n=50

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	A^2	*F*	*W*	*W**	A^2	*F*
I.Sym., lep								
Normal[a]	0.041[b]	0.048	0.049	0.049	0.040[b]	0.047	0.048	0.046
t_{10}	0.097	0.102	0.113	0.170[1]	0.082	0.098	0.100	0.155[1]
Logistic	0.127	0.138	0.157	0.245[1]	0.114	0.124	0.145	0.220[1]
SC(0.05, 9)	0.309	0.320	0.286	0.406[1]	0.276	0.285	0.251	0.374[1]
SC(0.10, 9)	0.468	0.484	0.443	0.593[1]	0.416	0.427	0.397	0.542[1]
t_4	0.356	0.372	0.414	0.516[1]	0.301	0.316	0.350	0.466[1]
SC(0.05, 25)	0.621	0.627	0.610	0.686[1]	0.598	0.608	0.577	0.666[1]
Laplace	0.382	0.404	0.537	0.603[1]	0.304	0.325	0.440	0.504[1]
SC(0.10, 25)	0.831	0.833	0.833	0.896[1]	0.806	0.812	0.804	0.879[1]
t_2	0.808	0.820	0.860	0.905[1]	0.738	0.748	0.804	0.860[1]
Cauchy	0.996	0.996	0.997	0.999[4]	0.988	0.989	0.991	0.998[1]
II. Sym., platy								
Beta(1,1)	0.853	0.871[1]	0.582	0.392	0.552	0.591[1]	0.379	0.206
Beta(1.5, 1.5)	0.471	0.512[1]	0.271	0.115	0.284	0.309[1]	0.190	0.071
Beta(2,2)	0.235	0.262[1]	0.147	0.066	0.161	0.186[1]	0.088	0.029
III. Asym., lep.								
Weibull(2)	0.420	0.452[1]	0.335	0.307	0.326	0.357[1]	0.278	0.262
LC(0.05, 3)	0.304	0.329	0.302	0.400[1]	0.250	0.267	0.241	0.338[1]
LC(0.10, 3)	0.496	0.510	0.502	0.542[1]	0.415	0.438	0.422	0.457[1]
LC(0.20, 3)	0.582	0.605	0.625[2]	0.531	0.449	0.483[1]	0.511	0.410
Chi-squared(10)	0.558	0.582[1]	0.484	0.490	0.449	0.471[1]	0.377	0.399
LC(0.05, 5)	0.846	0.851	0.800	0.879[1]	0.811	0.814	0.767	0.852[1]
LC(0.10, 5)	0.975	0.976	0.970	0.984[1]	0.958	0.960	0.945	0.969[1]
LC(0.05, 7)	0.908	0.909	0.892	0.914[3]	0.904	0.905	0.886	0.913[1]
LC(0.20, 5)	1.000	1.000	1.000	1.000	0.984	0.987[3]	0.986	0.971
LC(0.10, 7)	0.993[4]	0.993[4]	0.991	0.993[4]	0.992	0.992	0.990	0.993[4]
LC(0.20, 7)	1.000[4]	1.000[4]	0.999	1.000[4]	0.998[4]	0.998[4]	0.998[4]	0.994
Chi-squared(4)	0.952	0.956[2]	0.877	0.879	0.835	0.845[1]	0.765	0.761
Chi-squared(2)	1.000[4]	1.000[4]	0.998	0.998	0.973[2]	0.973[2]	0.960	0.958
Chi-squared(1)	1.000	1.000	1.000	1.000	0.999[4]	0.999[4]	0.999[4]	0.998
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	0.999[4]	0.999[4]	0.998	0.994
IV.Asym.,platy								
Beta(3,2)	0.268	0.293[1]	0.183	0.086	0.176	0.194[1]	0.145	0.075
Beta(2,1)	0.881	0.897[1]	0.709	0.593	0.619	0.653[1]	0.510	0.372

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

Table C-4 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.05 and n=70

Distributions	(a) iid observations				(b) OLS residuals			
	W'	*W**	A^2	*F*	*W'*	*W**	A^2	*F*
I.Sym., lep								
Normal[a]	0.067[b]	0.046	0.047	0.051	0.064[b]	0.046	0.048	0.048
t_{10}	0.257[1]	0.109	0.141	0.231	0.234[1]	0.096	0.116	0.202
Logistic	0.348[1]	0.149	0.219	0.323	0.320[1]	0.141	0.178	0.279
SC(0.05, 9)	0.558[1]	0.386	0.366	0.532	0.541[1]	0.360	0.338	0.511
SC(0.10, 9)	0.738[1]	0.542	0.555	0.718	0.727[1]	0.498	0.515	0.693
t_4	0.696[1]	0.429	0.545	0.662	0.650[1]	0.396	0.495	0.617
SC(0.05, 25)	0.816[1]	0.735	0.724	0.801	0.811[1]	0.711	0.695	0.803
Laplace	0.757[1]	0.476	0.693	0.743	0.688[1]	0.400	0.595	0.669
SC(0.10, 25)	0.959[2]	0.914	0.916	0.954	0.958[2]	0.903	0.904	0.954
t_2	0.968[1]	0.889	0.942	0.960	0.953[1]	0.868	0.917	0.947
Cauchy	1.000[4]	0.999	1.000[4]	1.000[4]	0.999	0.998	1.000[4]	1.000[4]
II. Sym., platy								
Beta(1,1)	0.822	0.991[1]	0.781	0.690	0.574	0.864[1]	0.625	0.439
Beta(1.5, 1.5)	0.352	0.796[1]	0.403	0.231	0.241	0.609[1]	0.304	0.151
Beta(2,2)	0.131	0.469[1]	0.183	0.083	0.089	0.342[1]	0.139	0.064
III. Asym., lep.								
Weibull(2)	0.590	0.645[1]	0.457	0.461	0.511	0.538[1]	0.399	0.380
LC(0.05, 3)	0.582[1]	0.394	0.381	0.511	0.533[1]	0.348	0.337	0.477
LC(0.10, 3)	0.745[1]	0.634	0.629	0.655	0.691[1]	0.571	0.594	0.629
LC(0.20, 3)	0.763	0.763	0.795[1]	0.679	0.662	0.667	0.704[1]	0.590
Chi-squared(10)	0.773[2]	0.761	0.635	0.655	0.704[1]	0.679	0.567	0.592
LC(0.05, 5)	0.961[1]	0.925	0.885	0.952	0.956[1]	0.913	0.869	0.943
LC(0.10, 5)	0.998[4]	0.997	0.994	0.998[4]	0.997[4]	0.994	0.995	0.997[4]
LC(0.05, 7)	0.980[3]	0.978	0.973	0.980[3]	0.979[3]	0.977	0.968	0.979[3]
LC(0.20, 5)	1.000	1.000	1.000	1.000	0.996	0.997[4]	0.997[4]	0.995
LC(0.10, 7)	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]
LC(0.20, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(4)	0.990	0.993[2]	0.967	0.974	0.968[2]	0.964	0.929	0.931
Chi-squared(2)	1.000	1.000	1.000	1.000	0.999	1.000[4]	0.997	0.999
Chi-squared(1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
IV.Asym.,platy								
Beta(3,2)	0.248	0.475[1]	0.280	0.164	0.175	0.368[1]	0.221	0.111
Beta(2,1)	0.936	0.990[1]	0.891	0.845	0.794	0.908[1]	0.779	0.646

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

Table C-5 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.05 and n=100

Distributions	(a) iid observations				(b) OLS residuals			
	W'	*W**	A^2	*F*	*W'*	*W**	A^2	*F*
I.Sym., lep								
Normal[a]	0.050	0.050	0.048	0.049	0.050	0.050	0.048	0.050
t_{10}	0.292	0.097	0.173	0.303[2]	0.269	0.093	0.147	0.277[2]
Logistic	0.372	0.131	0.238	0.384[2]	0.345	0.120	0.225	0.350[2]
SC(0.05, 9)	0.628[2]	0.397	0.408	0.627	0.614[2]	0.374	0.405	0.612
SC(0.10, 9)	0.826[2]	0.611	0.676	0.822	0.806[2]	0.575	0.647	0.804
t_4	0.749	0.469	0.653	0.760[1]	0.727	0.444	0.618	0.745[1]
SC(0.05, 25)	0.904[2]	0.830	0.828	0.903	0.900[2]	0.815	0.808	0.896
Laplace	0.826	0.540	0.820	0.849[1]	0.785	0.464	0.754	0.812[1]
SC(0.10, 25)	0.983[2]	0.950	0.961	0.982	0.987[2]	0.948	0.958	0.985
t_2	0.982	0.935	0.974	0.983[2]	0.974	0.912	0.967	0.975[2]
Cauchy	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
II. Sym., platy								
Beta(1,1)	0.961	1.000[1]	0.953	0.929	0.820	0.984[1]	0.870	0.771
Beta(1.5, 1.5)	0.513	0.954[1]	0.563	0.428	0.348	0.853[1]	0.458	0.295
Beta(2,2)	0.204	0.763[1]	0.310	0.154	0.155	0.647[1]	0.266	0.127
III. Asym., lep.								
Weibull(2)	0.700	0.809[1]	0.617	0.610	0.624	0.716[1]	0.544	0.526
LC(0.05, 3)	0.668[1]	0.494	0.508	0.641	0.631[1]	0.453	0.469	0.611
LC(0.10, 3)	0.847[1]	0.755	0.798	0.820	0.811[1]	0.711	0.765	0.793
LC(0.20, 3)	0.884	0.892	0.930[1]	0.864	0.838	0.854	0.895[1]	0.812
Chi-squared(10)	0.860	0.888[1]	0.801	0.800	0.820	0.845[1]	0.752	0.748
LC(0.05, 5)	0.989[2]	0.962	0.954	0.987	0.987[2]	0.957	0.952	0.986
LC(0.10, 5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.05, 7)	0.996[4]	0.996[4]	0.992	0.996[4]	0.996[4]	0.995	0.994	0.996[4]
LC(0.20, 5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.10, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.20, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(4)	0.998[4]	0.998[4]	0.996	0.997	0.996[4]	0.994	0.991	0.993
Chi-squared(2)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
IV.Asym.,platy								
Beta(3,2)	0.303	0.720[1]	0.399	0.230	0.241	0.614[1]	0.335	0.185
Beta(2,1)	0.991	1.000[1]	0.984	0.975	0.934	0.988[1]	0.946	0.883

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

Table C-6 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1 and n=10

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	*A^2*	*F*	*W*	*W**	*A^2*	*F*
I.Sym., lep								
Normal[a]	0.095	0.100	0.102	0.100	0.078[b]	0.083[b]	0.087[b]	0.087[b]
t_{10}	0.119	0.128	0.124	0.135[3]	0.090	0.094[4]	0.093	0.091
Logistic	0.147	0.159[4]	0.148	0.152	0.092	0.098	0.090	0.099[4]
SC(0.05, 9)	0.160	0.167	0.177[3]	0.177[3]	0.107	0.114	0.116	0.132[1]
SC(0.10, 9)	0.194	0.204	0.207	0.218[2]	0.112	0.125	0.121	0.136[2]
t_4	0.206	0.214	0.223	0.229[2]	0.111	0.120	0.110	0.130[2]
SC(0.05, 25)	0.232	0.245	0.237	0.248[3]	0.129	0.135	0.134	0.142[3]
Laplace	0.204	0.213	0.216	0.242[1]	0.101	0.110	0.118	0.123[2]
SC(0.10, 25)	0.346	0.362	0.366	0.373[2]	0.173	0.186	0.171	0.211[1]
t_2	0.383	0.397	0.406	0.412[2]	0.168	0.179	0.180	0.207[1]
Cauchy	0.659	0.674	0.690	0.700[2]	0.302	0.314	0.327	0.351[1]
II. Sym., platy								
Beta(1,1)	0.154[3]	0.150	0.152	0.123	0.080	0.078	0.082[4]	0.073
Beta(1.5, 1.5)	0.103	0.106[3]	0.104	0.088	0.092	0.093[4]	0.093[4]	0.083
Beta(2,2)	0.080	0.082	0.093[2]	0.062	0.078	0.076	0.079	0.080[4]
III. Asym., lep.								
Weibull(2)	0.155	0.165[2]	0.154	0.138	0.083	0.084	0.086	0.087[4]
LC(0.05, 3)	0.161	0.176[3]	0.165	0.176[3]	0.079	0.089	0.085	0.095[3]
LC(0.10, 3)	0.185	0.195	0.195	0.205[3]	0.101	0.108	0.098	0.117[2]
LC(0.20, 3)	0.176	0.187	0.202[2]	0.189	0.090	0.095	0.098	0.108[2]
Chi-squared(10)	0.176	0.180[4]	0.173	0.171	0.099	0.102	0.108[4]	0.100
LC(0.05, 5)	0.348	0.355	0.363	0.368[3]	0.169	0.181	0.182	0.201[1]
LC(0.10, 5)	0.489	0.504	0.498	0.505[3]	0.168	0.181	0.173	0.197[1]
LC(0.05, 7)	0.447	0.450	0.457[4]	0.453	0.185	0.194	0.197	0.219[1]
LC(0.20, 5)	0.580	0.594	0.606[3]	0.570	0.158	0.172	0.171	0.179[3]
LC(0.10, 7)	0.670	0.670	0.670	0.675[4]	0.252	0.266	0.270	0.290[1]
LC(0.20, 7)	0.861	0.859	0.872[1]	0.855	0.226	0.230	0.234[4]	0.232
Chi-squared(4)	0.313	0.326[1]	0.305	0.297	0.109	0.117	0.114	0.130[1]
Chi-squared(2)	0.542	0.550[1]	0.519	0.513	0.153	0.164	0.160	0.173[2]
Chi-squared(1)	0.816	0.818[2]	0.787	0.779	0.235	0.253	0.244	0.260[2]
Weibull(0.5)	0.935[2]	0.935[2]	0.922	0.908	0.305	0.317	0.331	0.335[2]
Lognormal(0,1)	0.669	0.678[1]	0.648	0.641	0.233	0.247	0.230	0.261[2]
IV.Asym.,platy								
Beta(3,2)	0.107	0.108	0.113[4]	0.102	0.089[4]	0.086	0.089[4]	0.085
Beta(2,1)	0.230[2]	0.227	0.210	0.191	0.086	0.095[4]	0.092	0.093

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

Table C-7 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1 and n=20

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	*A*^{2}	*F*	*W*	*W**	*A*^{2}	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.099	0.097	0.104	0.102	0.098	0.095	0.102	0.098
t_{10}	0.156	0.152	0.146	0.157[4]	0.121	0.118	0.122	0.138[2]
Logistic	0.160	0.155	0.163	0.189[1]	0.144	0.141	0.130	0.163[1]
SC(0.05, 9)	0.259	0.255	0.247	0.284[1]	0.227	0.223	0.195	0.242[1]
SC(0.10, 9)	0.330	0.326	0.306	0.364[1]	0.257	0.256	0.242	0.283[1]
t_4	0.306	0.297	0.294	0.343[1]	0.235	0.230	0.218	0.267[1]
SC(0.05, 25)	0.400	0.398	0.390	0.421[1]	0.314	0.308	0.301	0.348[1]
Laplace	0.335	0.327	0.352	0.405[1]	0.250	0.248	0.244	0.286[1]
SC(0.10, 25)	0.614	0.609	0.597	0.632[1]	0.497	0.491	0.460	0.540[1]
t_2	0.619	0.616	0.622	0.666[1]	0.497	0.489	0.479	0.540[1]
Cauchy	0.899	0.896	0.905	0.921[1]	0.781	0.777	0.777	0.814[1]
<u>II. Sym., platy</u>								
Beta(1,1)	0.381	0.382[2]	0.317	0.208	0.164	0.165[3]	0.153	0.103
Beta(1.5, 1.5)	0.168	0.169[3]	0.155	0.104	0.121[3]	0.121[3]	0.115	0.080
Beta(2,2)	0.120[3]	0.119	0.110	0.070	0.098	0.096	0.105[3]	0.069
<u>III. Asym., lep</u>.								
Weibull(2)	0.266[2]	0.263	0.241	0.226	0.175[2]	0.174	0.158	0.148
LC(0.05, 3)	0.284	0.279	0.264	0.305[1]	0.210	0.205	0.196	0.235[1]
LC(0.10, 3)	0.368	0.363	0.352	0.388[1]	0.243	0.240	0.244	0.266[1]
LC(0.20, 3)	0.405[3]	0.403	0.392	0.361	0.228[3]	0.228[3]	0.211	0.206
Chi-squared(10)	0.320[2]	0.314	0.285	0.287	0.225[3]	0.223	0.205	0.218
LC(0.05, 5)	0.624	0.619	0.581	0.629[3]	0.453	0.447	0.413	0.488[1]
LC(0.10, 5)	0.800	0.798	0.784	0.814[1]	0.569	0.565	0.538	0.574[3]
LC(0.05, 7)	0.640[4]	0.637	0.635	0.638	0.572	0.567	0.532	0.578[3]
LC(0.20, 5)	0.927	0.927	0.931[3]	0.915	0.584	0.582	0.593[3]	0.554
LC(0.10, 7)	0.885	0.884	0.882	0.888[4]	0.751	0.751	0.739	0.765[1]
LC(0.20, 7)	0.993[4]	0.993[4]	0.993[4]	0.993[4]	0.757[3]	0.756	0.754	0.733
Chi-squared(4)	0.643[2]	0.640	0.566	0.562	0.408[2]	0.406	0.378	0.384
Chi-squared(2)	0.906[2]	0.905	0.863	0.851	0.573[2]	0.569	0.543	0.535
Chi-squared(1)	0.996[2]	0.996[2]	0.988	0.988	0.769[2]	0.769[2]	0.748	0.752
Weibull(0.5)	1.000	1.000	1.000	1.000	0.913[3]	0.910	0.887	0.907
Lognormal(0,1)	0.967[2]	0.965	0.941	0.938	0.750[2]	0.749	0.724	0.729
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.152[3]	0.152[3]	0.151	0.111	0.111	0.108	0.117[3]	0.091
Beta(2,1)	0.456[2]	0.454	0.377	0.310	0.220[3]	0.219	0.215	0.167

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

Table C-8 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1 and n=50

Distributions	(a) iid observations				(b) OLS residuals			
	W	*W**	A^2	*F*	*W*	*W**	A^2	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.097	0.100	0.102	0.102	0.100	0.102	0.103	0.099
t_{10}	0.169	0.172	0.202	0.285[1]	0.151	0.155	0.174	0.248[1]
Logistic	0.202	0.208	0.249	0.332[1]	0.186	0.191	0.226	0.306[1]
SC(0.05, 9)	0.364	0.368	0.354	0.451[1]	0.344	0.347	0.326	0.428[1]
SC(0.10, 9)	0.524	0.526	0.531	0.662[1]	0.490	0.493	0.492	0.634[1]
t_4	0.471	0.473	0.535	0.619[1]	0.426	0.429	0.472	0.579[1]
SC(0.05, 25)	0.656	0.657	0.639	0.710[1]	0.636	0.639	0.610	0.699[1]
Laplace	0.528	0.532	0.669	0.717[1]	0.444	0.449	0.568	0.642[1]
SC(0.10, 25)	0.870	0.872	0.871	0.922[1]	0.847	0.850	0.851	0.896[1]
t_2	0.850	0.852	0.895	0.926[1]	0.800	0.801	0.870	0.910[1]
Cauchy	0.998[4]	0.998[4]	0.998[4]	0.998[4]	0.991	0.991	0.994	0.996[4]
<u>II. Sym., platy</u>								
Beta(1,1)	0.954	0.961[1]	0.739	0.612	0.733	0.738[2]	0.538	0.362
Beta(1.5, 1.5)	0.656	0.661[2]	0.380	0.222	0.451	0.460[1]	0.283	0.147
Beta(2,2)	0.426	0.432[1]	0.234	0.116	0.320	0.324[2]	0.191	0.097
<u>III. Asym., lep</u>.								
Weibull(2)	0.577	0.585[1]	0.446	0.429	0.481	0.485[2]	0.382	0.359
LC(0.05, 3)	0.408	0.412	0.397	0.493[1]	0.350	0.354	0.359	0.440[1]
LC(0.10, 3)	0.613	0.618	0.617	0.657[1]	0.538	0.545	0.544	0.573[1]
LC(0.20, 3)	0.730	0.733	0.754[1]	0.668	0.606	0.614	0.639[1]	0.571
Chi-squared(10)	0.697	0.702[2]	0.610	0.626	0.615	0.621[1]	0.540	0.537
LC(0.05, 5)	0.874	0.875	0.851	0.899[1]	0.850	0.853	0.820	0.888[1]
LC(0.10, 5)	0.986	0.986	0.985	0.993[1]	0.976	0.976	0.968	0.989[1]
LC(0.05, 7)	0.934	0.934	0.929	0.935[4]	0.931	0.931	0.923	0.934[3]
LC(0.20, 5)	1.000	1.000	1.000	1.000	0.987	0.987	0.995[2]	0.989
LC(0.10, 7)	0.994	0.994	0.994	0.995[4]	0.995[4]	0.995[4]	0.991	0.995[4]
LC(0.20, 7)	1.000	1.000	1.000	1.000	0.999[4]	0.999[4]	0.999[4]	0.998
Chi-squared(4)	0.977	0.978[2]	0.942	0.936	0.917	0.918[2]	0.878	0.870
Chi-squared(2)	1.000[4]	1.000[4]	0.999	1.000[4]	0.987	0.988[2]	0.978	0.980
Chi-squared(1)	1.000	1.000	1.000	1.000	0.999[4]	0.999[4]	0.998	0.997
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	0.999[4]	0.999[4]	0.998	0.998
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.434	0.440[1]	0.284	0.188	0.322	0.328[1]	0.240	0.151
Beta(2,1)	0.972[2]	0.972[2]	0.858	0.789	0.811	0.819[1]	0.713	0.588

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

Table C-9 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1 and n=70

Distributions	(a) iid observations				(b) OLS residuals			
	W'	*W**	*A^2*	*F*	*W'*	*W**	*A^2*	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.118[b]	0.101	0.098	0.098	0.117[b]	0.094	0.094	0.098
t_{10}	0.359[1]	0.175	0.234	0.333	0.319[1]	0.159	0.209	0.299
Logistic	0.436[1]	0.207	0.288	0.410	0.421[1]	0.190	0.290	0.396
SC(0.05, 9)	0.601[1]	0.432	0.447	0.584	0.588[1]	0.411	0.407	0.565
SC(0.10, 9)	0.798[1]	0.587	0.635	0.765	0.755[1]	0.553	0.601	0.736
t_4	0.747[1]	0.501	0.625	0.729	0.721[1]	0.486	0.582	0.699
SC(0.05, 25)	0.848[1]	0.761	0.755	0.837	0.825[1]	0.733	0.735	0.813
Laplace	0.830[2]	0.533	0.786	0.828	0.769[1]	0.462	0.681	0.759
SC(0.10, 25)	0.965[2]	0.922	0.938	0.963	0.962[2]	0.908	0.924	0.958
t_2	0.971[3]	0.922	0.963	0.968	0.964[1]	0.898	0.940	0.956
Cauchy	1.000[4]	0.998	1.000[4]	1.000[4]	0.998[4]	0.997	0.998[4]	0.998[4]
<u>II. Sym., platy</u>								
Beta(1,1)	0.907	0.996[1]	0.883	0.846	0.718	0.927[1]	0.756	0.624
Beta(1.5, 1.5)	0.503	0.890[1]	0.530	0.386	0.349	0.692[1]	0.395	0.263
Beta(2,2)	0.265	0.673[1]	0.346	0.197	0.211	0.535[1]	0.284	0.146
<u>III. Asym., lep</u>.								
Weibull(2)	0.675	0.767[1]	0.564	0.566	0.579	0.630[1]	0.477	0.477
LC(0.05, 3)	0.654[1]	0.476	0.474	0.607	0.617[1]	0.431	0.446	0.567
LC(0.10, 3)	0.827[1]	0.739	0.751	0.776	0.784[1]	0.696	0.702	0.740
LC(0.20, 3)	0.861	0.852	0.878[1]	0.808	0.796	0.795	0.820[1]	0.747
Chi-squared(10)	0.827	0.841[2]	0.733	0.747	0.762[2]	0.752	0.675	0.673
LC(0.05, 5)	0.962[2]	0.925	0.911	0.960	0.960[2]	0.914	0.905	0.956
LC(0.10, 5)	0.999[4]	0.998	0.998	0.999[4]	1.000[4]	0.995	0.997	1.000[4]
LC(0.05, 7)	0.968[4]	0.966	0.964	0.968[4]	0.969[3]	0.967	0.959	0.968
LC(0.20, 5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.10, 7)	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.999[4]	0.998	0.999[4]
LC(0.20, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(4)	0.998	0.999[3]	0.988	0.995	0.990[2]	0.990[2]	0.970	0.971
Chi-squared(2)	1.000	1.000	1.000	1.000	1.000[4]	1.000[4]	1.000[4]	0.999
Chi-squared(1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.397	0.690[1]	0.428	0.283	0.321	0.542[1]	0.346	0.241
Beta(2,1)	0.972	0.995[1]	0.954	0.933	0.886	0.956[1]	0.882	0.798

Notes :

a. based on 10 000 samples

b. 95% confidence interval for the level does not contain the nominal value. Power estimates obtained do not reflect their true values.

- Refer to page 56 on the system of notation used for the superscript.

Table C-10 Power comparisons of normality tests on iid observations and OLS residuals based on 1000 samples using k=4 with X_1=1 and X_i, i=2,3,4 drawn from the uniform distribution at α=0.1 and n=100

Distributions	(a) iid observations				(b) OLS residuals			
	W’	*W**	A^2	*F*	*W’*	*W**	A^2	*F*
<u>I.Sym., lep</u>								
Normal[a]	0.096	0.098	0.098	0.099	0.095	0.100	0.096	0.096
t_{10}	0.377[2]	0.142	0.244	0.366	0.362[2]	0.141	0.239	0.356
Logistic	0.476	0.187	0.358	0.477[2]	0.447	0.157	0.326	0.456[2]
SC(0.05, 9)	0.684	0.456	0.502	0.686[2]	0.670[2]	0.436	0.485	0.668
SC(0.10, 9)	0.875[2]	0.662	0.763	0.872	0.860[2]	0.641	0.734	0.859
t_4	0.818	0.541	0.745	0.824[2]	0.797	0.516	0.709	0.801[2]
SC(0.05, 25)	0.894	0.819	0.836	0.895[2]	0.893	0.808	0.824	0.895[2]
Laplace	0.888	0.603	0.881	0.898[2]	0.858	0.540	0.834	0.874[1]
SC(0.10, 25)	0.990	0.969	0.981	0.992[2]	0.988	0.965	0.978	0.988[2]
t_2	0.994[2]	0.959	0.985	0.993	0.993	0.950	0.985	0.995[2]
Cauchy	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<u>II. Sym., platy</u>								
Beta(1,1)	0.989	1.000[1]	0.973	0.971	0.917	0.996[1]	0.934	0.872
Beta(1.5, 1.5)	0.714	0.992[1]	0.723	0.620	0.555	0.939[1]	0.630	0.479
Beta(2,2)	0.395	0.876[1]	0.487	0.324	0.317	0.788[1]	0.436	0.262
<u>III. Asym., lep</u>.								
Weibull(2)	0.833	0.900[1]	0.736	0.735	0.757	0.826[1]	0.696	0.674
LC(0.05, 3)	0.781[1]	0.588	0.634	0.761	0.755[1]	0.567	0.617	0.731
LC(0.10, 3)	0.886[1]	0.818	0.859	0.872	0.855[1]	0.787	0.822	0.831
LC(0.20, 3)	0.942	0.944	0.964[1]	0.926	0.884	0.900	0.920[1]	0.865
Chi-squared(10)	0.935	0.936[2]	0.887	0.902	0.905[2]	0.905[2]	0.862	0.874
LC(0.05, 5)	0.990[2]	0.972	0.972	0.989	0.988[2]	0.971	0.969	0.988[2]
LC(0.10, 5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.05, 7)	0.992[4]	0.991	0.990	0.992[4]	0.992[4]	0.991	0.992[4]	0.992[4]
LC(0.20, 5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.10, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
LC(0.20, 7)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(4)	1.000	1.000	1.000	1.000	1.000[4]	1.000[4]	0.998	0.999
Chi-squared(2)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chi-squared(1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Weibull(0.5)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lognormal(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<u>IV.Asym.,platy</u>								
Beta(3,2)	0.463	0.871[1]	0.511	0.364	0.403	0.758[1]	0.453	0.314
Beta(2,1)	0.997	1.000[3]	0.996	0.993	0.969	0.997[1]	0.974	0.951

Notes :

a. based on 10 000 samples

- Refer to page 56 on the system of notation used for the superscript.

References

Anderson, T.W., and Darling, D.A. (1954) "A test of goodness-of-fit" *Journal of the American Statistical Association*, **49**, 765-769.

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. Wiley, New York.

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

Chiang, A.C. (1992). *Elements of Dynamic Optimization.* McGraw-Hill, New York.

Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, New York.

D'Agostino, R.B. (1970). "Transformation to normality of the null distribution of g_1" *Biometrika*, **57**, 679-681

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

D'Agostino, R.B., and Stephens, M.A. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

deMontricher, G.F., Tapia, R.A., and Thompson, J.R. (1975). "Nonparametric maximum likelihood estimation of probability densities by penalty function methods" *Annals of Statistics*, **3**, 1329-1348

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley & Sons.

Good, I.J., and Gaskins, R.A. (1971) "Nonparametric roughness penalties for probability densities" *Biometrika*, **58**, 255-277

Gumbel, E.J. (1943). "On the reliability of the classical chi-square test" *Annals of Mathematical Statistics*, **14** , 253-263.

Harter, H.L. (1961). "Expected values of normal order statistics" *Biometrika*, **48,** 151-165.

Hoel, D.G. (1972). "A representation of mortality by competing risks" *Biometrics*, **28**, 475-488.

Huang, C.J., and Bolch, B.W. (1974) "On the testing of Regression Disturbances for normality" *Journal of the American Statistical Association*, **69**, 330-335.

Klonias, V.K. (1982) "Consistency of two nonparametric maximum penalized likelihood estimators of the probability density function" *Annals of Statistics* **10**, 811-824

Mardia, K.V. (1980) "Tests of univariate and multivariate normality" In *Handbook of Statistics*, Vol 1, Ed. P.R. Krishnaiah. pp. 279-320. North-Holland, Amsterdam.

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 reasonably be supposed to have risen from random sampling" *Philosophical Magazine* (5), **50**, 157-175.

Pearson, E.S., D'Agostino, R.B., and Bowman, K.O. (1977) "Tests for departure from normality : Comparison of powers" *Biometrika*, **64**, 231-246

Pearson, E.S., and Hartley, H.O. (1972) *Biometrika Tables for Statisticians*, Vol 1. Cambridge University Press.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992) *Numerical Recipes in FORTRAN : The Art of Scientific Computing,* 2nd ed. Cambridge University Press.

Risebrough, R.W. (1972) "Effects of environmental pollutants upon animals other than man" *Proceedings of the 6th Berkeley Symposium on Mathematics and Statistics VI*, Berkeley, University of California Press. 443-463.

Royston, J.P. (1982) "An extension of Shapiro and Wilk's W test for normality to large samples" *Applied Statistics*, **31**, No.2, 115-124.

Shapiro, S.S. (1980). *How to Test Normality and other Distributional Assumptions* American Society for Quality Control, Wisconsin.

Shapiro, S.S., and Francia, R.S. (1972). "An approximate analysis of variance test for normality" *Journal of the American Statistical Association*, **67**, 215-216.

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

Shapiro, S.S., Wilk, M.B., and Chen, H.J. (1968). "A comparative study of various tests of normality" *Journal of the American Statistical Association*, **63**, 1343-1372.

Siegel, S., and Castellan, N.J. (1988) *Nonparametric Statistics for the Behavioral Sciences*, 2ed. McGraw-Hill.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Snedecor, G.W. (1946). *Statistical Methods*, 4th ed. Ames:Iowa State College Press.

Stephens, M.A. (1974). "EDF statistics for goodness of fit and some comparisons" *Journal of the American Statistical Association*, **69**, 730-737.

Teichroew, D. (1956). "The mixture of normal distributions with different variances" *Annals of Mathematical Statistics*, **28**, 510-512.

Terrell, G.R. (1995). "A Fisher information test for Pearson-family membership" *Proceedings of the Statistical Computing Section*, Joint Statistical Meetings, Orlando, Florida. 230-234.

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

Weisberg, S. (1980) Comment on "Some large-sample tests for nonnormality in the linear regression model" *Journal of the American Statistical Association*, **75**, 28-31.

Weisberg, S., and Bingham, C. (1975) "An approximate analysis of variance test for non-normality suitable for machine calculation" *Technometrics*, **17**, 133-134.

White, H., and Macdonald, G. M. (1980) "Some large-sample tests for nonnormality in the linear regression model" *Journal of the American Statistical Association*, **75**, 16-28.

Vita

Yew-Haur Lee was born on August 12, 1966 to Willard Lee and Cheong Moo in Singapore. He graduated from Queenstown Secondary Technical School, Singapore, in 1982 and was awarded the Most Outstanding Student Award. He then graduated from Raffles Junior College, Singapore, in 1984.

In 1991, he graduated with Second Class (Upper) Honors from the National University of Singapore, Singapore, with a Bachelor of Social Science in Economics (Statistics Stream).

On January 13, 1995 and September 3, 1998, he received his Master of Science and Doctor of Philosophy degrees in Statistics, respectively, from Virginia Polytechnic Institute and State University, Blacksburg VA.

The author is a member of the National Statistics Honor Society (Mu Sigma Rho) and the American Statistical Association.