REGRESSION AND EDF TESTS OF FIT

by

John Spinelli

B.Sc. Simon Fraser University 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

August 1980

APPROVAL

Name: John Spinelli

Degree: Master of Science

Title of a thesis: Regression and EDF Tests of Fit

Examining Committee:

Chairman: D.M. Eaves

M.A. Stephens
Senior Supervisor

R. Lockhart

R. Koopman

Wm. Warren
External Examiner
Research Scientist
Western Forest Products Laboratory
Canadian Forestry Service

Date Approved: August 15, 1980

Title of Thesis/Dissertation:

REGRESSION AND EDF TESTS OF FIT

Author: ______________________

(signature)

JOHN SPINELLI

(name)

17/9/80

(date)

ABSTRACT

Many procedures have been proposed to test the hypothesis that a set of data comes from a certain distribution. These procedures are known as goodness-of-fit tests or tests of fit. Two important types of goodness-of-fit tests are tests based on regression theory and tests based on the empirical distribution function (EDF).

In this thesis several new tests of fit based on regression theory are proposed and some results on these new tests are given. A framework with which to discuss these new tests and already existing regression tests is presented. In addition, EDF tests are extended to the case of testing for the two-parameter exponential distribution.

A Monte Carlo power study is conducted in order to compare empirically the power of regression, EDF, and other tests of fit when testing for the normal, exponential, and extreme value distributions.

ACKNOWLEDGEMENTS

I would like to thank the members of my supervisory committee, Dr. R. Lockhart and Dr. R. Koopman, for their many helpful suggestions, and especially Dr. M.A. Stephens for his support, encouragement and guidance during the preparation of this thesis and throughout the course of my studies.

Thanks also go to fellow graduate students Al Byrd and Luc Janicot for their interest and assistance.

Most of all, I wish to thank my wife, Monica, for the great help she gave to me in producing this work, and for having the patience to bear with me through the many months that it took to finish this thesis.

TABLE OF CONTENTS

LIST OF TABLES

I. Definitions and Background Theory

1.1 Introduction

In order to satisfy the assumptions of a statistical technique, or to model a phenomenon, a researcher will often want to test that his or her data come from a certain distribution. Let y_1, y_2, ..., y_n be independent and identically distributed observations from a cumulative distribution function, $F(y)$. The researcher wants to test the hypothesis

$$F(y) = F_0(y) \quad , \qquad (1)$$

where $F_0(y)$ is some cumulative distribution function. Occasionally, this distribution will be completely specified, but more often it will have one or more unknown parameters. This problem is known as goodness-of-fit testing.

The classical procedure for solving this problem is the χ^2 test proposed by Karl Pearson (1900). The advantage of this test is that it is simple to calculate and easily understood. With continuous distributions, however, the test requires grouping, which causes a loss of information and thus a loss of power. Because of this, the test will not be discussed further. (For a complete discussion of the χ^2 test see Kendall and Stuart, 1973)

There have been many goodness-of-fit procedures suggested for continuous distributions. The oldest and most common methods use the empirical distribution function (EDF),

$$F_n(y) = \frac{\#\text{ of observed values} \le y}{n} . \qquad (2)$$

The fit is judged by the degree of closeness between the EDF and the distribution function, F(y). Tests of this type are known as EDF tests.

Another common procedure for testing fit is to plot $y_{(i)}$, the i-th largest observation in the sample, against some function of i. The fit is then judged by the strength of the linearity of the graphed values. This procedure leads to regression tests of fit.

In this thesis regression and EDF tests of fit are reviewed, new tests are introduced, and some results on these new tests are given. These tests are then compared with other goodness-of-fit tests. The comparisons will be made for three distributions; exponential, extreme value, and normal.

Regression and EDF tests are presented in chapters 2 and 3, respectively. In chapter 4 other tests of fit which are to be compared are introduced, and some examples of testing for the different distributions are given in chapter 5. The results of the power comparisons, and the conclusions drawn from them are presented in the final chapter. The remainder of this chapter is used to state the basis by which the various tests will be compared, define and present relevant facts about the three distributions, and present some important estimation theory.

1.2 Principles of Goodness-of-Fit

The different tests of fit will be compared on the basis of three principles of goodness-of-fit testing that were proposed by Stephens (1978).

Principle 1

The statistic should keep close to the original data, not make an elaborate transformation to new numbers which mean little to the researcher. Thus, when a test statistic is significant, the researcher will be able to interpret this in terms of some irregularity in the original data.

For example, when the χ^2 test statistic mentioned above is significant, the significance can be seen as a value in one or more of the groups that is much different than expected. Both EDF tests and regression tests follow principle 1, and this principle will not be discussed further in this thesis.

Principle 2

A test should be consistent and unbiased.

Let the null hypothesis be

$$F(y) = F_0(y;a,b) \ , \qquad (3)$$

where a and b are unknown location and scale parameters, and let t_α be the value of the statistic, t, such that

$$P(t > t_\alpha \mid F(y) = F_0(c+dy)) = \alpha . \qquad (4)$$

for any constants c and d. A test is consistent if

$$\lim_{n\to\infty} P(t > t_\alpha \mid F(y) \neq F_0(c+dy)) = 1 . \qquad (5)$$

This concept was introduced by Wald and Wolfowitz (1940) and is a minimum quality needed for a goodness-of-fit test, as the corresponding concept of consistency in estimation is a minimum requirement for an estimator.

A related but not as essential a quality is that a test be unbiased. A test, t, is unbiased if

$$P(t > t_\alpha \mid F(y) \neq F_0(c+dy)) > \alpha \qquad (6)$$

for all sample sizes. A consistent test will always be asymptotically unbiased, but an unbiased test will not necessarily be consistent.

<u>Principle 3</u>

A test should be powerful over a wide range of alternatives.

The power of a test is

$$P(\ t > t_\alpha \mid F(y) \neq F_0(c+dy)\ )\ . \tag{7}$$

Clearly no one test will be most powerful for all alternatives, so we desire a test that has good relative power for all or, at least, most alternatives. Extensive Monte Carlo work has been performed in an attempt to find tests which follow this principle. The powers of regression, EDF and other tests of fit has been compared for the normal, exponential and extreme value distributions. The power of a test against a certain alternative distribution is calculated by looking at several samples from that alternative and calculating the proportion of times the test rejects the null hypothesis.

1.3 The Distributions

The Exponential Distribution

The two parameter exponential distribution is defined as,

$$F(y;a,b) = 1 - \exp[-(y - a)/b]\ , \tag{8}$$

for $y \geq a$ and $b > 0$. The parameters, a and b, are location and scale parameters, and in addition, a is a threshold parameter, meaning that either a response cannot occur or cannot be measured below that value.

The distribution has had wide application in lifetime testing and other events that occur at random in time. Because time is defined only on the positive axis, the distribution is most often used with a known, and equal to 0. There are other cases, however, where it might be supposed that some variable declines exponentially after a certain value, a, which is unknown. In this thesis only this case, where both a and b are unknown, will be considered.

An important property of the exponential distribution is that it has a constant hazard rate. This means that an item that has an exponential failure rate is ageless. The hazard rate is defined as

$$h(y) = f(y)/[1 - F(y)] \quad , \tag{9}$$

where f(y) is the density function, f(y) = dF(y)/dy. This can be interpreted, for a lifetime distribution, as the probability that an item will fail in time dy, given that it has survived up to time y. For the exponential distribution, it is easily seen that h(y)=1/b. In other words the future survival of the item does not depend on the age of the item. Several goodness-of-fit tests have been derived based on this property.

Another characteristic of the exponential distribution is that it is a special case of two other distributions that are often used to model events in time; the Weibull distribution, (which will be defined in the next section), and the gamma distribution. Because of this, these distributions are important alternatives to be guarded against when testing that a random variable is exponentially distributed.

The Extreme Value Distribution

The maximum value of a random sample, x_n, has a nondegenerate limiting distribution if there exists a sequence of constants $\{a_n\}$ and $\{b_n\}$ with $a_n > 0$ such that

$$\lim_{n \to \infty} P[(x_n - b_n)/a_n < y] = F(y) \; . \tag{10}$$

If the distribution of $(x_n - b_n)/a_n$ is nondegenerate, then F(y) must be one of three types. The Type I distribution, which is commonly referred to as the extreme value distribution, is defined as

$$F(y;a,b) = \exp\{-\exp[(y - a)/b]\} \; , \tag{11}$$

where a and b are location and scale parameters, respectively, and b > 0. The type II distribution is defined as

$$F(y;a,b,k) = \exp\{-[(y - a)/b]^{-k}\} \; , \tag{12}$$

for $y \geq a$, and b,k > 0. The type III distribution is defined as

$$F(y;a,b,k) = \exp\{-[(a - y)/b]^{k}\} \; , \tag{13}$$

for $y \leq a$ and b,k > 0.

The type I distribution arises if the parent population has a right tail which declines at least as quickly as the exponential distribution, the second form arises if the right tail is thicker than the exponential distribution, and the third form arises if the distribution is bounded above. If we let $c = 1/k$, then the type I distribution can be seen as the crossover point between the two distributions, that is, where $c=0$.

The corresponding distributions of smallest values can be found by looking at the distributions of $(-Y)$. The type III distribution for smallest values is also known as the Weibull distribution.

The type II and III distributions can be transformed to the type I distribution by the transformations

$$X = \log (Y - a) \text{ and } X = -\log (a - Y) , \tag{14}$$

respectively. From this it is easily seen that if Y has a Weibull distribution, defined

$$F(y;c,d,k) = 1 - \exp\{-[(y - c)/d]^k\} , \tag{15}$$

for $y \geq c$, $d,k > 0$, and c known, then $X = -\log(Y - c)$ has the extreme value distribution, where $a = -\log(d)$ and $b = 1/k$. This fact enables us to test for the two parameter Weibull and the extreme value distribution with one procedure.

The extreme value distributions have applications in many problems involving the maximum or minimum values of a sample. Some examples are the weakest link in a chain, the maximum height of a river, the maximum rainfall, etc. The Weibull distribution has also been used quite often in reliability and lifetime models.

The Normal Distribution

A random variable is normally distributed if it has the probability density

$$f(y;a,b) = 1/[b\sqrt{(2\pi)}]\ \exp\{-(1/2)[(y - a)/b]^2\}\ , \qquad (16)$$

for $b > 0$. The parameters, a and b, are once again location and scale parameters, respectively. This distribution is the most widely used distribution in statistics, and has applications to almost every field of statistics.

1.4 Estimation Procedures

Ordinary Least Squares

Let y be a vector of observations and X be a matrix with columns $(1,x)$, where x is a vector of known values and $1' = (1,1,\ldots 1)$. Let y be linearly related to x; that is,

$$y = XB + e\ , \qquad (17)$$

where $B' = (a,b)$ is the vector of parameters, and e is a vector of random errors with

$$E(e) = 0 , \tag{18}$$

and

$$E(ee') = \sigma^2 I . \tag{19}$$

The ordinary least squares (OLS) estimator of B minimizes the sum of squares

$$SS = (y - XB)'(y - XB) . \tag{20}$$

Differentiating and setting equal to 0, gives

$$dSS/dB = 2X'(y - XB) = 0 , \tag{21}$$

which implies

$$B(OLS) = (X'X)^{-1}X'y . \tag{22}$$

Therefore, the OLS estimates of a and b are

$$b(OLS) = [x'y - (1'y)(1'x)/n]/[x'x - (1'x)^2/n] \tag{23}$$

and

$$a(OLS) = 1'y/n - b(OLS)(1'x)/n . \tag{24}$$

It is known by the Gauss-Markov theorem that a(OLS) and b(OLS) are the minimum variance linear unbiased estimators of a and b for this model. (e.g. Kendall and Stuart, 1973) If (19) does not hold, the OLS estimators are still unbiased estimators of a and b, but the are no longer the minimum variance (MV) linear estimators. The MV linear estimators are the generalized least squares estimators which will be defined in the next section.

Generalized Least Squares Theory

Once again, let y be a vector of observations and X be a matrix with columns (1,x), such that

$$y = XB + e \quad , \tag{25}$$

where $B' = (a,b)$ is the vector of parameters and e is the vector of random errors with expectation equal to 0. Now, let

$$E(ee') = b^2 V \quad . \tag{26}$$

The generalized least squares (GLS) estimate of B, given by Aitken (1935), (also, see Kendall and Stuart, 1973), is found by transforming y to i.i.d variables. That is, decompose V^{-1} into LL', and let

$$z = L'y \quad , \tag{27}$$

$$W = L'X \quad , \tag{28}$$

and

$$u = L'e \quad . \tag{29}$$

The new model is then

$$z = WB + u \ , \tag{30}$$

where

$$E(uu') = I \ . \tag{31}$$

Thus, using ordinary least squares we can find the estimate of B,

$$\begin{aligned} B(OLS) &= (W'W)^{-1} W'y \\ &= (X'LL'X)^{-1} X'LL'y \\ &= (X'V^{-1}X)^{-1} X'V^{-1}y \ . \end{aligned} \tag{32}$$

Therefore, the GLS estimate of B is

$$B(GLS) = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{33}$$

which implies that

$$a(GLS) = -x'Gy \tag{34}$$

and

$$b(GLS) = 1'Gy \ , \tag{35}$$

where

$$G = V^{-1}(1x' - x1')V^{-1}/D \ , \tag{36}$$

and

$$D = (1'V^{-1}1)(x'V^{-1}x) - (1'V^{-1}x)^2 \ . \qquad (37)$$

It can be shown that these are the best (MV) linear unbiased estimates of a and b. (Kendall and Stuart, 1973)

The GLS estimates are also the values which minimize the generalized error sum of squares,

$$R = (y - XB)'V^{-1}(y - XB) \ . \qquad (38)$$

Differentiating and setting equal to 0 leads to,

$$dR/dB = 2\ X'V^{-1}\ (y - XB) = 0 \qquad (39)$$

which implies

$$B(GLS) = (X'V^{-1}X)^{-1}\ X'V^{-1}y \ . \qquad (40)$$

The fact that the GLS estimates minimize this weighted sum of squares is useful in understanding the different goodness-of-fit tests to be discussed in chapter 2.

A particular situation where the generalized least squares estimators are of value is in the estimation of parameters from the order statistics of a sample. Let y be a continuous random variable with distribution F(y;a,b), where a and b are location and scale parameters. Also let $y_{(1)}, \le y_{(2)}, \le \cdots, \le y_{(n)}$ be the order statistics of a sample of size n from this distribution. Then

$$y = a + bt \quad , \tag{41}$$

where t is the reduced random variable with distribution $F(t;0,1)$.

Let m_i be the expected value of the i-th order statistics of a sample of size n drawn from $F(t;0,1)$, and m be the vector with components m_i. Then,

$$E[y_{(i)}] = a + bm_i \quad , \tag{42}$$

and

$$y = XB + e \quad , \tag{43}$$

where in this case y is the vector of order statistics, and X is the matrix with columns (1,m).

Since the $y_{(i)}$ are order statistics, they are not independent and identically distributed (i.i.d.) random variables. Let V be the covariance matrix of the order statistics from the distribution $F(t;0,1)$, and therefore of e; i.e.,

$$E(ee') = b^2 V \; . \tag{44}$$

Estimates of a and b can then be found by applying (34) and (35).

The Normal Distribution

For some distributions, the properties of m and V simplify the GLS theory. (see LLoyd, 1952; and Kendall and Stuart, 1973) For symmetric distributions,

$$Jm = -m , \tag{45}$$

and

$$V^{-1} = J\, V^{-1}\, J , \tag{46}$$

where

$$J = \begin{vmatrix} 0 & & & 1 \\ & & \cdot & \\ & \cdot & & \\ \cdot & & & \\ 1 & & & 0 \end{vmatrix} . \tag{47}$$

Applying this fact,

$$\begin{aligned} 1'V^{-1}m &= 1'(JV^{-1}J)m \\ &= 1'J\, V^{-1}\, Jm \\ &= 1'V^{-1}(-m) \\ &= -\, 1'V^{-1}m , \end{aligned} \tag{48}$$

which implies that

$$1'V^{-1}m = 0 . \tag{49}$$

Using this result, (34) and (35) reduce to

$$a(GLS) = 1'V^{-1}y/1'V^{-1}1 \tag{50}$$

and

$$b(GLS) = m'V^{-1}y/m'V^{-1}m \tag{51}$$

Furthermore, for the normal distribution it is known that

$$V^{-1}1 = 1 , \tag{52}$$

implying that (34) further reduces to

$$a(GLS) = 1'y/n = \bar{y} \quad . \tag{53}$$

The Exponential Distribution

Another special distribution for which simplifying properties of m and V^{-1} are known is the exponential distribution. In this case

$$m_i = \sum_{j=1}^{i} (n - j + 1)^{-1}, \tag{54}$$

and V^{-1} has all elements zero except

$$V^{-1}_{r,r+1} = (n - r)^2 + (n - r + 1)^2 \tag{55}$$

and

$$V^{-1}_{r,r+1} = V^{-1}_{r+1,r} = -(n - r)^2 \ . \tag{56}$$

These properties imply that

$$1'm = n , \tag{57}$$

$$V^{-1}m = 1 , \tag{58}$$

$$1'V^{-1} = (n^2, 0, \ldots, 0) , \tag{59}$$

$$1'V^{-1}m = n , \tag{60}$$

$$m'V^{-1}m = n , \tag{61}$$

$$1'V^{-1}1 = n^2 , \tag{62}$$

and

$$1'V^{-1}y = n^2 y_{(1)} . \tag{63}$$

Hence (34) and (35) become

$$a(GLS) = y_{(1)} - [\bar{y} - y_{(1)}]/(n - 1) , \tag{64}$$

and

$$b(GLS) = n[\bar{y} - y_{(1)}]/(n - 1) . \tag{65}$$

These special forms of the GLS estimators will become important in studying the regression tests for the exponential and normal distributions.

<u>Maximum Likelihood</u>

Let $y_1, y_2, \ldots, y_n$ be a random sample from a density $f(y;B)$, where B is a vector of unknown parameters. The

Likelihood Function is defined

$$L(y|B) = f(y_1|B)f(y_2|B)\ldots f(y_n|B) \; . \qquad (66)$$

The maximum likelihood estimators are the set of values of the parameters, B, which maximize the Likelihood Function. ML estimators, first introduced by Fisher (1921), have several nice statistical properties, including the fact that they are asymptotic MV unbiased estimators. For a complete discussion of ML estimators and their properties, see Kendall and Stuart (1973).

The Exponential Distribution

For the exponential distribution, the maximum likelihood estimators of a and b are

$$a(ML) = y_{(1)} \text{ and } b(ML) = \bar{y} - y_{(1)} \; , \qquad (67)$$

where

$$\bar{y} = \sum_{i=1}^{n} y_i/n \; , \qquad (68)$$

and $y_{(1)}$ is the smallest value in the sample. a(ML) and b(ML) are biased, and it is important to note that the variance of a(ML) is proportional to $1/n^2$ rather than $1/n$ This fact will be utilized in chapter 3.

The similarity, in this case, between the GLS estimators and the maximum likelihood (ML) estimators is easily seen. In fact, the GLS estimators can be expressed as functions of the ML estimators;

$$a(GLS) = a(ML) - b(ML)/n \ , \tag{69}$$

and

$$b(GLS) = nb(ML)/(n - 1) \ . \tag{70}$$

The Extreme Value Distribution

The maximum likelihood estimators of a and b for the extreme value distribution (Type I) are

$$a(ML) = -b(ML) \log \{ \sum_{i=1}^{n} \exp[-y_i/b(ML)]/n\} \ , \tag{71}$$

and

$$b(ML) = (1/n) \sum_{i=1}^{n} y_i - \{ \sum_{i=1}^{n} y_i \exp[y_i/b(ML)]\}/\{ \sum_{i=1}^{n} \exp[y_i/b(ML)]\} \ , \tag{72}$$

where b(ML) must be solved for iteratively. These are biased estimates of a and b.

The Normal Distribution

For the normal distribution, the maximum likelihood estimates of a and b are

$$a(ML) = \bar{y}, \text{ and } b(ML) = (1/n) \sum_{i=1}^{n} (y_i - \bar{y})^2 . \qquad (73)$$

It is interesting to note that a(GLS) = a(ML), in this case.

II. Regression Tests

2.1 Introduction

The forerunners of regression tests of fit were the graphical goodness-of-fit methods. In these methods the ordered sample is plotted against some function of i, and the goodness-of-fit is graphically estimated by the closeness of this plot to linearity. Regression tests can be seen as attempts to define the meaning of a good or bad fit, by the use of a measure of the linearity of the plot.

Several different methods of measuring linearity exist, and these methods are discussed in section 2.2. In section 2.3 the various regression tests are presented, and the procedure to calculate the test statistics is given in section 2.4. In the final section of this chapter results on some of these regression tests are given.

2.2 Measures of Linearity

Most regression statistics utilize the following model,

$$y = a1 + bx + e \qquad (74)$$

for some x, where

$$E(e) = 0 \,, \tag{75}$$

$$E(ee') = b^2 V \,, \tag{76}$$

and y is the vector of order statistics. A measure of how well the data fits this linear model is desired. This measure of the linearity of the data should not depend on the parameters, a and b, of the hypothesized distribution. That is, the measure should be location and scale invariant. This allows one test statistic to be used for all values of a and b.

Measures of Linearity using OLS Estimators

The most common measure of linearity is the ordinary r^2, the squared correlation between a vector of values y, and a vector of values x, defined

$$r^2 = \frac{[(x - \bar{x}1)'(y - \bar{y}1)]^2}{(x - \bar{x}1)'(x - \bar{x}1)\ (y - \bar{y}1)'(y - \bar{y}1)} \,. \tag{77}$$

Looking at $R = 1 - r^2$ is more intuitive to goodness-of-fit testing since the hypothesis of a certain model is rejected by large values of the statistic. A procedure which produces a statistic equal to R is to regress the ordered values on some function of i, estimate a and b by ordinary least squares (OLS), and calculate the error (or unexplained) sum of squares, ESS(OLS), divided by the total sum of squares, TSS. The statistic, R, is then defined

$$R = ESS(OLS)/TSS \ , \tag{78}$$

where

$$ESS(OLS) = (y - \hat{y})'(y - \hat{y}) \ , \tag{79}$$

$$TSS = (y - \bar{y}1)'(y - \bar{y}1) \ , \tag{80}$$

$$\hat{y} = a(OLS) + b(OLS)x \ , \tag{81}$$

and a(OLS), b(OLS), and y are defined in equations (23), (24), and (74), respectively.

Because of the identity, TSS = ESS(OLS) + RSS(OLS), where

$$RSS(OLS) = (\hat{y} - \bar{y}1)'(\hat{y} - \bar{y}1) \ , \tag{82}$$

R always lies between 0 and 1. Clearly, r^2 is equivalent to RSS(OLS)/TSS or 1 - ESS(OLS)/TSS.

Measures of Linearity using GLS Estimators

The main problem with the measures of linearity involving estimation by OLS is that since we are ignoring the covariance structure of the order statistics, we are not getting the best estimates of a and b. Thus, because we are not using this information about the distribution, we expect a drop in power from this type of statistic.

If we wish to utilize the information of the covariance structure in the estimation procedure, the parameters a and b should be estimated using generalized least squares (GLS). After this has been done, it seems reasonable, for a measure of the fit to the line to look at some function of the residuals, $y - \hat{y}$. The problem is that there is no function that has the relationship to r^2, the squared correlation coefficient, that the ordinary error sum of squares has. Therefore, several methods of measuring linearity, using GLS estimators, will be considered.

The first method of measuring linearity using GLS estimators, is to look at the sum of squares of the residuals using GLS estimators divided by the total sum of squares. The error sum of squares using GLS estimators is defined

$$ESS(GLS) = (y - \hat{y})'(y - \hat{y}) , \qquad (83)$$

where

$$\hat{y} = a(GLS) + b(GLS)m , \qquad (84)$$

and a(GLS) and b(GLS) are defined in equations (34) and (35), respectively. This measure has an intuitive appeal, and as will be seen later, it has some nice statistical properties.

Measures of Linearity using a Decomposition of the Covariance Matrix

The second method of utilizing the GLS estimates in a measure of linearity uses the way in which the GLS estimates are found. That is, we first decompose V^{-1}, the inverse covariance matrix of e, the random error term, into

$$V^{-1} = LL' \tag{85}$$

as was presented in section 1.4 above. Then, transform the y vector and the X matrix into

$$z = L'y \tag{86}$$

$$W = L'X \; . \tag{87}$$

The covariance matrix of z is now the identity matrix. We can then look at

$$ESS_z = (z - \hat{z})'(z - \hat{z}) \tag{88}$$

divided by

$$TSS_z = (z - \bar{z}1)'(z - \bar{z}1) \; , \tag{89}$$

where

$$\hat{z} = WB(OLS) \tag{90}$$

and

$$\bar{z} = 1'z/n \; . \tag{91}$$

A problem with this method is that the TSS is not independent of the transformation used. Since there is more than one way to decompose the matrix, V^{-1}, into a matrix times its transpose, this method is not desirable.

Measures of Linearity using the Generalized Error Sum of Squares

Another method of measuring linearity is based on the generalized error sum of squares, GESS. GESS is defined

$$GESS = (y - \hat{y})'V^{-1}(y - \hat{y}), \tag{92}$$

where

$$\hat{y} = a(GLS) + b(GLS)m . \tag{93}$$

Since this is the quantity that is minimized by the GLS estimation procedure, it is an appropriate measure to be used when the parameters are estimated by GLS. In order to make the measure of linearity scale invariant, the GESS must be divided by another sum of squares. What is wanted is a generalized TSS of the form

$$(y - c1)'V^{-1}(y - c1) , \tag{94}$$

where c is some suitable scalar quantity. Buse (1973) proposed the quantity $c=\bar{y}(GLS)$, where

$$\bar{y}(GLS) = 1'V^{-1}y/1'V^{-1}1 . \tag{95}$$

The generalized total sum of squares is then

$$GTSS = (y - \bar{y}(GLS)1)'V^{-1}(y - \bar{y}(GLS)1) . \tag{96}$$

This quantity has properties similar to the TSS in ordinary regression. Just as $\bar{y}$ is the scalar that minimizes the unweighted sum of squared deviations $(y - c1)'(y - c1)$, $\bar{y}(GLS)$ minimizes the weighted sum of squares, (94). Also, GTSS can be decomposed into a generalized regression sum of squares, and the GESS defined above. The decomposition is

$$GTSS = (\hat{y} - \bar{y}(GLS)1)'V^{-1}(\hat{y} - \bar{y}(GLS)1) + GESS \ . \qquad (97)$$

Thus, the measure of linearity consisting of GESS divided by GTSS, has the property of lying between 0 and 1.

Measures of Linearity using ML Estimators

A final possible class of measures of linearity utilize the maximum likelihood estimators. If the ML estimators of a and b, a(ML) and b(ML), exist, then these estimators can then be inserted in the regression equation and the fit measured by one of the already mentioned procedures. Since the covariance matrix, V, is not explicitly used in the estimation procedure, the most acceptable measure is the ordinary ESS(ML)/TSS, where ESS(ML) is the residual sum of squares using maximum likelihood estimators. Although this method has no least squares interpretation, it has the advantage that the ML estimators do not require the expected values or covariances of the order statistics in order to be calculated.

In the next section the tests of fit corresponding to these measures of linearity will be presented.

2.3 Definitions of Regression Tests

A number of different choices for the vector x have been proposed or implied for the regresssion model given in (74). Regression tests will be presented in the framework of the vector, x, and the measure of linearity that are utilized.

Regression Tests Using the Expected Values of Order Statistics

The most common set of values of x is the vector of expected values of the order statistics, m. If the OLS estimators are used, several statistics equivalent to the ordinary r^2 exist. For the normal distribution, Shapiro and Francia (1972) proposed the statistic,

$$W' = (m'y)^2/(n-1)s^2m'm \ , \qquad (98)$$

where

$$(n-1)s^2 = (y - \bar{y}1)'(y - \bar{y}1) \ . \qquad (99)$$

It is easily seen that this is equivalent to r^2, since $1'm = 0$ for the normal distribution. Sarkadi (1975) showed that the statistic r^2 is consistent for all location-scale distributions with finite variance, and introduced the corresponding statistic for the exponential distribution, although he gives no percentage points for the statistic. Gerlach (1979) showed that a modified r^2 statistic for censored samples is consistent for all general location-scale distributions, and he gave points for the extreme value (Weibull) case. (For the purpose of comparison, R will be looked at rather than r^2.)

Until now, no goodness-of-fit tests have been proposed using generalized least squares or maximum likelihood estimators and measuring the departure from linearity using one of the methods described in the preceding section. The test statistic which uses generalized least squares to estimate the parameters, and measures the departure from linearity with ESS(GLS)/TSS, will be referred to as G1. G1 is defined as,

$$G1 = (y - \hat{y})'(y - \hat{y})/(y - \bar{y}1)'(y - \bar{y}1) \ , \qquad (100)$$

where

$$\hat{y} = a(GLS) + b(GLS)m \ . \qquad (101)$$

The corresponding test statistic using maximum likelihood estimators and measuring linearity by ESS(ML)/TSS, will be called G2. The test statistic which utilizes the measure of linearity, GESS/GTSS defined in (92) and (96) respectively, will be referred to G3.

In this thesis the finite and asymptotic expectations of some regression tests have been found. Also, using Sarkadi's proof of the consistency of r^2, G1 is shown to be a consistent test. These results will be presented in the next section.

A major drawback with the GLS methods is that V has only been tabulated for small sample sizes (e.g. for the extreme value distribution for sample sizes up to n = 25, (Mann, 1968); and for the normal distribution for sample sizes up to n = 50, (Tietjen, et al, 1977)). There is no problem for the exponential distribution, since the inverse covariance matrix can be easily computed, but for other distributions it must be approximated using an algorithm such as Davis and Stephens' (1977, 1978) approximation of the covariance matrix for the normal distribution.

Regression Tests using the Inverse Cumulative Distribution Function

There have been many regression tests suggested using the inverse cumulative distribution function of some function of i, for the i-th element of x. These functions are easily calculated approximations of m, chosen either to avoid storing tables of the expected values of the order statistics, or for very large sample sizes for which m has not been tabulated. (The expected values of the order statistics have been tabulated for the normal distribution for n = 1(1),100(25),200 by Pearson and Hartley (1972), and for the extreme value distribution for n = 1(1),100 by White (1967).) Also, the asymptotic distribution of the statistics using these quantities is often easier to calculate than the distribution of statistics using the expected values of the order statistics.

Since there would be no reason to approximate m with some function to avoid storage, and then store V or V^{-1}, all of the tests involving approximations of this type propose estimating a and b by OLS and measuring the linearity by r^2. One could, of course, estimate V as well with an algorithm such as Davis and Stephens (1977,1978) for the normal distribution, and then proceed with one of the methods using the GLS estimates of a and b.

These approximations of m utilize the probability integral transformation which is a method of transforming any continuous distribution to a uniform(0,1) distribution. This is, if X has distribution F(x), then setting

$$u = F(x) \quad , \tag{102}$$

implies u has a uniform(0,1) distribution. Therefore, $E[u_{(i)}] = i/(n+1)$ is a reasonable first approximation to $F\{E[x_{(i)}]\}$. For this reason,

$$h = F^{-1}[E(u)] \tag{103}$$

where u is the vector of uniform order statistics, has been suggested. It should be noted that this approximation can be very bad for the extreme values of i, especially if the tails of F^{-1} are steep.

The obvious test statistic to use is the squared correlation between y and h. DeWet and Venter (1973) proposed this statistic for the normal distribution, and gave a table of the asymptotic distribution of the statistic. Smith and Bain (1976) give some critical values of the statistic, $Rh = 1 - r^2(y,h)$, for the normal, exponential, and extreme value case.

Since the maximum likelihood estimates require no storage, another possible statistic is G2, ESS(ML)/TSS, with m replaced by h. This statistic will be referred to as G2h.

Depending on the distribution wished to be tested for, several other functions have been suggested. For example, Weisberg and Bingham (1975), following Blom (1958), propose using

$$x_i = F^{-1}[(i - .375)/(n + .25)] \quad , \qquad (104)$$

as an approximation to m_i for the normal distribution. For other distributions there are, no doubt, other functions that approximate m well. However, if the approximation is good, it is expected that any test statistic using the approximation will have the same power properties as the corresponding statistic using m. Indeed, this is what Weisberg and Bingham found for their test. Because of this, no further comparisons will be made with tests utilizing approximations other than h.

The Shapiro Wilk Statistic

Shapiro and Wilk (1968, 1972) devised the W tests for normality and exponentiality, respectively. The procedure they propose is to regress the order statistics on m as above, estimate b by GLS, and to look at the normalized ratio of the square of b(GLS) to the usual estimate of b^2, s^2 which is defined in (99). Thus, in the case of normality,

$$W = R^4[b(GLS)]^2/(n-1)C^2s^2 \quad , \tag{105}$$

where $C^2 = m'V^{-1}V^{-1}m$ and $R^2 = m'V^{-1}m$. R^2 and C^2 are constants placed in the formula so that W lies between 0 and 1.

There is another way, however, to view this statistic. If one regresses the order statistics on the values

$$k = 1'G/d \tag{106}$$

where G is given in (36), and d is the normalization constant to make $k'k = 1$, estimates a and b by OLS, and looks at the ordinary r^2, one gets a statistic equivalent to W. In the case of normality this reduces to regressing the order statistics on $m'V^{-1}/\surd(C^2)$ where C^2 is given above, and in the case of exponentiality it reduces to regressing the order statistics on a set of values the first of which is at $1/n - 1$, and the rest at $1/n$.

A possible justification for regressing the order statistics against these unusual values comes from the results of Chernoff and Lieberman (1956) and Barnett (1975). They found that if one constrains estimation of a and b to OLS, then the "best" values to use as k for the exponential and normal distributions are those used for W. Here "best" means that the OLS estimators are also the GLS estimators. This seems to suggest that, at least for normality and exponentiality, W will be a more powerful statistic than those based on r^2 from regressing the order statistics on m. Although no "best" values for k exist for the extreme value distribution, or to any other well known distribution, the W test was extended to the extreme value distribution in this thesis.

Using the fact due to Stephens (1975) that for the normal distribution,

$$V^{-1}m \to 2m \tag{107}$$

as n approaches infinity, it can be seen that the W test and the Shapiro-Francia test, W′, are asymptotically equivalent. (In fact W′ was initially derived as a large sample approximation to W. For a comparison of the two statistics, see Shapiro and Francia (1972), Weisberg (1974), and Weisberg and Bingham (1975).) Sarkadi conjectures that since Shapiro-Francia′s test is consistent, that this implies W for normality will also be consistent.

In the exponential case, however, there is no relationship between the Shapiro-Wilk and Shapiro-Francia tests. In fact for exponentiality

$$W = n[\bar{y} - y_{(1)}]^2/(n - 1)s^2 \tag{108}$$

which is similar to the coefficient of variation, $s/\bar{y}$. Sarkadi implies that because many distributions have the same coefficient of variation as the exponential distribution, that W is not consistent. It is not quite as simple as this, however. For example, if one is given a normal distribution with mean equal to 1, and variance equal to 1, (coefficient of variation equal to 1), the Shapiro-Wilk statistic will have asymptotic power equal to 1 since

$$\lim_{n \to \infty} E[y_{(1)}] = -\infty \ . \tag{109}$$

It is possible to construct a distribution for which the W test for exponentiality will not have asymptotic power equal to 1 if attention is restricted to distributions with a threshold parameter. (Since W is known to be location invariant, the threshold can be taken to equal 0 without loss of generality.) In these distributions

$$\lim_{n \to \infty} E[y_{(1)}] = 0 \ . \tag{110}$$

Since the distributions of $\bar{y}$ and s^2, and thus the W test for exponentiality, depend only on the first four moments of a distribution, one method to show that the W test for exponentiality is inconsistent is to find a distribution in this restricted class with the same first four moments of any translated exponential distribution. If this was done, the Shapiro-Wilk statistic would have the same asymptotic distribution for this new distribution as for the null distribution, and thus could not be consistent.

Another way to show the inconsistency of the W test is to find a distribution restricted to the positive axis, with the same coefficient of variation as the exponential distribution. Some examples of distributions restricted to the positive axis with coefficient of variation equal to 1 are; the Beta(a,b) distribution with $a < 1$ and $b=a(a+1)/(1-a)$, the F(a,b) distribution with $b > 6$ and $a=(2b-4)/(b-6)$, and the lognormal(0,b) distribution with $b=\ln(2)$.

The existence of these distributions alone does not prove the lack of consistency of W. It would also have to be shown that the variance of W for these distributions approaches 0 at least as quickly as for the exponential distribution. No results of this type have been proved, but the power of W was compared for several of the above distributions for sample sizes 10, 20, 50, and 100. The mean and standard deviation of W were also calculated for 1000 samples of each size and for each distribution. These values can be compared with those for the exponential distribution in table 1. (The power of another test, A2, which will be defined in the next chapter, was also included for the purpose of comparison.) The results strongly suggest that the W test is not a consistent test, and for some distributions is asymptotically biased.

Table 1

Expected value, standard deviation and power (in %) of the Shapiro-Wilk W-test for exponentiality against selected alternatives.

Alternative Distribution	Sample Size	Expected Value	Standard Deviation	Power W	Power A2
Exponential	10	0.135	0.058		
	20	0.060	0.022		
	50	0.022	0.006		
	100	0.010	0.002		
Beta (.25,.417)	10	0.130	0.073	15.8	61.9
	20	0.056	0.021	9.4	92.4
	50	0.021	0.005	7.2	100.0
	100	0.010	0.001	2.2	100.0
Beta (.5,1.5)	10	0.129	0.056	7.7	22.1
	20	0.057	0.019	5.8	38.2
	50	0.021	0.004	4.4	73.1
	100	0.010	0.001	1.5	96.4
Beta (.75,2.625)	10	0.134	0.063	8.4	12.3
	20	0.097	0.067	6.7	14.6
	50	0.021	0.004	5.5	22.1
	100	0.010	0.002	2.5	38.4
Beta (.99,197.01)	10	0.134	0.061	9.9	9.9
	20	0.060	0.021	8.4	9.1
	50	0.022	0.006	11.4	10.3
	100	0.010	0.002	6.4	10.6
F (6,8)	10	0.145	0.076	14.9	12.0
	20	0.064	0.031	21.1	19.8
	50	0.022	0.008	29.7	45.2
	100	0.011	0.003	29.6	78.9
F (4,10)	10	0.143	0.074	16.8	13.3
	20	0.061	0.026	16.1	15.6
	50	0.022	0.008	26.5	32.8
	100	0.011	0.003	24.0	60.3
Log-normal (0,ln(2))	10	0.134	0.069	14.4	12.5
	20	0.059	0.026	19.2	16.5
	50	0.021	0.008	25.7	34.5
	100	0.010	0.003	22.9	63.3

2.4 Procedure to Calculate the Regression Statistics

Let y_1, y_2, ..., y_n be the random sample of size n from the distribution F(y;a,b).

1. Sort the sample in ascending order, and let y be the vector of order statistics.
2. Find

$$y = \hat{a} + \hat{b}k \ , \qquad (111)$$

where $\hat{a}$ and $\hat{b}$ are OLS, GLS, or ML estimates, and k is the vector of expected values of order statistics or some other function, depending on the statistic that is being calculated.

3. Calculate the appropriate error sum of squares and total sum of squares;
4. Compute the regression statistic according to the formula given in table 2.

Table 2

Definitions of regression test statistics studied.

Notation:

$ESS(c,x) = [y - a(c)1 - b(c)x]'[y - a(c)1 - b(c)x]$

$TSS = (y - \bar{y})'(y - \bar{y})$

$GESS = [y - a(GLS)1 - b(GLS)m]'V^{-1}[y - a(GLS)1 - b(GLS)m]'$

$GTSS = (y - \bar{y})V^{-1}(y - \bar{y})$

Name	Formula
R	ESS(OLS,m)/TSS
Rh	ESS(OLS,h)/TSS
G1	ESS(GLS,m)/TSS
G2	ESS(ML,m)/TSS
G2h	ESS(ML,h)/TSS
G3	GESS/GTSS

2.5 Theory of Regression Tests

Expectations of G1 and R

Since in the case of order statistics the errors cannot be assumed to be normally distributed, very little distribution theory is known for the regression statistics. From least squares theory, however, the expectations of the statistics under the null hypothesis can often be found.

Before finding the expectations of R and G1 under the null hypothesis, we will prove four lemmas. For the remainder of this section, let y, X, B, and e be as defined in (25), and V be as defined in (26).

Lemma 1

$$E[ESS(OLS)] = tr(V) - tr[X'VX(X'X)^{-1}] \tag{112}$$

Proof: Let $\hat{y}$ be defined in (81). It is known that

$$\begin{aligned}[y - \hat{y}] &= XB + e - XB(OLS) \\ &= X[B - B(OLS)] + e \\ &= X[B - (X'X)^{-1}X'y] + e \\ &= X[B - (X'X)^{-1}X'(XB + e)] + e\end{aligned}$$

$$= e - X(X'X)^{-1}X'e$$

$$= [In - X(X'X)^{-1}X']e$$

$$= Qe , \tag{113}$$

where

$$Q = [I - X(X'X)^{-1}X'] . \tag{114}$$

This implies that

$$E[ESS(OLS)] = E[(y - \hat{y})'(y - \hat{y})]$$

$$= E\ (e'Q'Qe)$$

$$= b^2\ tr(Q'QV) , \tag{115}$$

where tr is the trace operator. (Kendall and Stuart, 1973) Furthermore, since the trace operator is invariant under cyclic permutation of matrices,

$$tr(Q'QV) = tr\{[I - X(X'X)^{-1}X']'[I - X(X'X)^{-1}X']V\}$$

$$= \{tr[V - 2X(X'X)^{-1}X'V +$$

$$X(X'X)^{-1}X'X(X'X)^{-1}X'V]\}$$

$$= \{tr(V) - tr[X'VX(X'X)^{-1}]\} . \tag{116}$$

Lemma 2

Let $\hat{y}$ be the vector of estimates of y defined in (84),

$$E[\hat{y} - E(\hat{y})][\hat{y} - E(\hat{y})]' = b^2\ XS^{-1}X' , \tag{117}$$

where

$$S^{-1} = (X'V^{-1}X)^{-1} \quad . \tag{118}$$

Proof: First, it can be shown that

$$\begin{aligned} \hat{y} &= XB(GLS) \\ &= XS^{-1}X'V^{-1}y \\ &= XS^{-1}X'V^{-1}(XB + e) \\ &= XB + XS^{-1}X'V^{-1}e \quad , \end{aligned} \tag{119}$$

and thus

$$E(\hat{y}) = XB \quad . \tag{120}$$

From (119) and (120),

$$\begin{aligned} E[\hat{y} - E(\hat{y})][\hat{y} - E(\hat{y})]' &= E(XS^{-1}X'V^{-1}ee'V^{-1}XS^{-1}X'] \\ &= b^2 \; XS^{-1}X'V^{-1}XS^{-1}X' \\ &= b^2 \; XS^{-1}X'. \end{aligned} \tag{121}$$

<u>Lemma 3</u>

$$E[ESS(GLS)] = tr(V) - tr(X'XS^{-1}) \tag{122}$$

Proof: Let $\hat{y}$ be defined in (84). Following the proof of Lemma 1, it can be seen that

$$\begin{aligned}(y - \hat{y}) &= XB + e - XB(GLS) \\ &= X[B - B(GLS)] + e \\ &= X[B - (X'V^{-1}X)^{-1}X'V^{-1}y] + e \\ &= X[B - (X'V^{-1}X)^{-1}X'V^{-1}(XB + e)] + e \\ &= e - X(X'V^{-1}X)^{-1}X'V^{-1}e \\ &= [In - XS^{-1}X'V^{-1}]e \\ &= Pe , \end{aligned} \tag{123}$$

where

$$P = [I - XS^{-1}X'V^{-1}] , \tag{124}$$

and S^{-1} is defined in (118). This implies that

$$\begin{aligned}E[ESS(GLS)] &= E[(y - \hat{y})'(y - \hat{y})] \\ &= E\ (e'P'Pe) \\ &= b^2\ tr(P'PV) , \end{aligned} \tag{125}$$

where

$$\begin{aligned}tr(P'PV) &= tr\{[I - XS^{-1}X'V^{-1}]'[I - XS^{-1}X'V^{-1}]V\} \\ &= tr\{[I - 2XS^{-1}X'V^{-1} + V^{-1}XS^{-1}X'XS^{-1}X']V\} \\ &= tr[V - 2X'XS^{-1} + V^{-1}X(X'V^{-1}X)X'XS^{-1}X'] \\ &= tr(V) - 2tr(X'XS^{-1}) + tr(X'XS^{-1}) \\ &= tr(V) - tr(X'XS^{-1}) . \end{aligned} \tag{126}$$

Lemma 4.

$$\lim_{n\to\infty} E[ESS(GLS)/n] = 0 \,. \tag{127}$$

for all distributions with finite second moment.

Proof: First, it can be shown that for any distribution where σ^2 and μ^2 are finite, that

$$tr(V) = n(\sigma^2 + \mu^2) - m'm \tag{128}$$

where σ^2 is the variance of the reduced variable t, and μ is the mean of t. Furthermore, Hoeffding (1953) proved that

$$\lim_{n\to\infty} m'm/n = \int y^2 \, dF(y)$$

$$= \sigma^2 + \mu^2 \tag{129}$$

whenever the right hand side is finite. Therefore,

$$\lim_{n\to\infty} [tr(V)/n)] = 0 \,. \tag{130}$$

By Lemma 2, we know that since $X'XS^{-1}$ is a covariance matrix,

$$tr(X'XS^{-1}) \geq 0 \,, \tag{131}$$

which implies that

$$tr(V) \geq tr(V) - tr(X'XS^{-1}) ,$$
$$= E[ESS(GLS)] \qquad (132)$$

Therefore, since E[ESS(GLS)] is positive and less than tr(V), the result follows from (130).

The significance of the above lemmas is that ESS(OLS) and ESS(GLS) are the quantities in the numerators of the statistics, R and G1. In both of these cases the quantity in the denominator is s^2 which is known to converge in probability to $b^2\sigma^2$. This implies that for any e > 0, there exists an N such that

$$E(ESS/TSS) = E\{[ESS/(N-1)]/s^2\}$$
$$\leq 1/(b^2\sigma^2 - e)\ E[ESS/(N-1)] , \qquad (133)$$

which implies that

$$\lim_{n\to\infty} E(G1) = \lim_{n\to\infty} E(ESS(GLS)/TSS)$$
$$= 0. \qquad (134)$$

Furthermore, by the Markov inequality which states that for any positive random variable X,

$$P(X \geq t) \leq E(X)/t , \qquad (135)$$

the critical values of G1 approach 0 in the limit.

The limiting expectation of R, and thus of the critical values of R, has been shown to be 0 by Sarkadi (1975), and can be obtained from Hoeffding's (1953) theorem. From this theorem it can be derived that there is perfect asymptotic correlation between the expected values of the order statistics and the order statistics. (Kendall and Stuart, 1973).

In the special case of the normal distribution, ESS/s^2 and s^2 are independent due to the sufficiency of s^2. (Hogg and Craig, 1956) This implies

$$E(ESS/s^2) = E(ESS)/E(s^2) \tag{136}$$

for any sample size. This allows the expectations of R and G1 to be calculated exactly for finite sample sizes.

No statement of independence can be made for the exponential or extreme value distributions, but Monte Carlo results for finite sample sizes suggest that (136) is approximately true for these two distributions.

Consistency of G1

In order to prove the consistency of r^2 (=1 - R) for any distribution, $F_0(x)$, with finite variance, Sarkadi (1975) effectively shows that there exists an e > 0, such that

$$\lim_{n \to \infty} P(R > e \mid H_1) = 1 , \tag{137}$$

and that for any $e_1 > 0$, in particular $e_1 = e$,

$$\lim_{n \to \infty} P(R > e_1 \mid H_0) = 0 , \tag{138}$$

where H_0 is the null hypothesis, $F(x) = F_0(x)$, and H_1 is the alternative hypothesis, $F(x) = F_1(x)$. Using these two results from Sarkadi and Lemma 4 of the preceding section, we prove the following theorem.

<u>Theorem 1</u>

G1, defined in (100), is a consistent test.

Proof: It is well known that a(OLS) and b(OLS) are the estimators which minimize the ESS. Therefore, if any other estimators are used, the ESS will be greater than or equal to the ESS using OLS estimators. In particular, it is easily seen that

$$ESS(OLS) \leq ESS(GLS) \qquad (139)$$

which implies that $R \leq G1$. This fact, with result (137), implies that there exists an $e > 0$ such that

$$\lim_{n \to \infty} P(\ G1 > e \mid H_1\ ) = 1\ . \qquad (140)$$

It was shown above that the limiting expectation of G1 is 0, and this, along with the Markov inequality, implies that for any $e > 0$,

$$\lim_{n \to \infty} P(\ G1 > e \mid H_0\ ) = 0\ . \qquad (141)$$

Therefore, G1 is a consistent test.

The Expectation of G3

The expectation of GESS, defined in (92), is

$$E(GESS) = b^2(n - 2) . \tag{142}$$

(Kendall and Stuart, 1973) The expectation of GTSS can be found as follows:

$$\begin{aligned} E(GTSS) &= E[y'V^{-1}y - (1'V^{-1}y)^2/1'V^{-1}1] \\ &= E\{(a1 + bm + e)'V^{-1}(a1 + bm + e) - \\ &\quad [1'V^{-1}(a1 + bm + e)]^2/1'V^{-1}1\} \\ &= E[b^2m'V^{-1}m - (e'V^{-1}1)(1'V^{-1}e)/1'V^{-1}1 + \\ &\quad e'V^{-1}e - b^2(1'V^{-1}m)^2/1'V^{-1}1] \\ &= b^2[m'V^{-1}m - (1'V^{-1}m)^2/1'V^{-1}1 + \\ &\quad n - tr(V^{-1}11'V^{-1}V)/1'V^{-1}1] \\ &= b^2[m'V^{-1}m - (1'V^{-1}m)^2/1'V^{-1}1 + \\ &\quad (n - 1)] . \end{aligned} \tag{143}$$

Using the fact that $1'V^{-1}m = 0$ and $V^{-1}m \to m$, in the case of the normal distribution,

$$E(GTSS) = b^2(n - 1 + m'V^{-1}m) , \tag{144}$$

which implies that

$$\lim_{n \to \infty} E(GTSS/n) = 3b^2 . \tag{145}$$

In the exponential case (143) reduces to

$$E(GTSS) = 2b^2(n - 1) , \qquad (146)$$

which implies that

$$\lim_{n\to\infty} E(GTSS/n) = 2b^2 . \qquad (147)$$

Thus, if it could be shown that GTSS and GESS/GTSS were independent or that GTSS converged in probability to its expected value, the asymptotic expectation of G3 would equal 1/2 in the exponential case and 1/3 in the normal case.

III. EDF Statistics

3.1 Background and Definitions

Let $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ be the order statistics of a sample of size n from the distribution F(y). The EDF statistics measure the discrepancy between the empirical distribution function, $F_n(y)$, defined in chapter 1, and the hypothesized distribution function, $F_0(y)$. These test statistics were developed for the case where $F_0(y)$ is continuous, and completely specified, that is, with no unknown parameters.

The most famous EDF statistic, D, was proposed by Kolmogorov (1933), and is defined by

$$D = \sup_y |F_n(y) - F_0(y)| \quad . \tag{148}$$

This is the largest vertical gap between the empirical distribution function and $F_0(y)$. Kolmogorov also found the asymptotic distribution of D, and gave a recurrence relation for the distribution for finite n.

Smirnov (1939) introduced the corresponding one-tailed test statistics, D+ and D−, defined by

$$D+ = \sup_y [F_n(y) - F_0(y)] \quad , \tag{149}$$

and

$$D- = \sup_y [F_0(y) - F_n(y)] \quad , \tag{150}$$

and found their asymptotic distribution. These three statistics, D, D+, and D-, are known as the Kolmogorov-Smirnov statistics.

Another test statistic of this general type was proposed by Kuiper (1960), and is defined

$$V = D+ + D- \quad . \tag{151}$$

This statistic was proposed in order to test for randomness on a circle. The Kolmogorov-Smirnov statistics can produce different values depending on where the origin of the circle is measured, whereas V is independent of the origin of the circle. Kuiper found the asymptotic distribution of V, and Stephens (1965) gave the exact distribution of V for finite n.

Because of the Glivenko-Cantelli theorem (Glivenko, 1933; Cantelli, 1933) which states that if $F(y) = F_0(y)$, then

$$P(\lim_{n \to \infty} D = 0) = 1 \ , \tag{152}$$

it is known that the Kolmogorov-Smirnov tests and the Kuiper test are consistent tests.

A second family of EDF statistics for continuous distributions was first proposed by Cramer(1928), and generalized by von Mises (1931) and Smirnov (1939). The general form is

$$K = n \int g[F_0(y)][F_n(y) - F_0(y)]^2 \, dF_0(y) \; . \qquad (153)$$

The statistic with g=1 is called the Cramer-von Mises statistic, W2. The asymptotic distribution of this statistic was investigated by Smirnov (1936), and found by Anderson and Darling (1952). Pearson and Stephens (1962) found the distribution of W2 for finite n.

Anderson and Darling (1952,1954) proposed and found the asymptotic distribution of the statistic with the weight function,

$$g[F_0(y)] = \{F_0(y)[1 - F_0(y)]\}^{-1} \; , \qquad (154)$$

which is known as A2. The corresponding statistic for goodness-of-fit on the circle, invariant with respect to the origin, was proposed by Watson (1961). The statistic, U2, is defined by

$$U2 = n \int (F_n(y) - F_0(y) - t)^2 \, dF_0(y) \qquad (155)$$

where

$$t = \int [F_n(y) - F_0(y)] \, dF_0(y) \; . \qquad (156)$$

Stephens (1963,1964) found the asymptotic and finite distribution of U.

Stephens (1970) found modifications for all of the above statistics, so that the asymptotic significance points can be used for finite sample sizes. This allows the use of these statistics with only one set of tables. For a detailed survey of the distribution theory of EDF tests, see Durbin (1973).

In practice, to calculate the EDF statistics, the probability integral transformation, referred to in the last chapter, is utilized. The order statistics from F(y) are transformed to uniform order statistics by this transformation, and the empirical distribution function of these new uniform order statistics is compared with the uniform distribution function, $F(z) = z$. Since this can be done for all continuous distributions, the distributions of the EDF statistics are independent of the distribution being tested for.

The formulas for the 6 EDF statistics defined above, are

$$D+ = \max_i \, [i/n - z_{(i)}] \, , \tag{157}$$

$$D- = \max_i \, (z_{(i)} - (i-1)/n) \, , \tag{158}$$

$$D = \max \, (D+, D-) \, , \tag{159}$$

$$V = D+ + D- \, , \tag{160}$$

$$W2 = \{ \sum_{i=1}^{n} [z_{(i)} - (2i - 1)/2n]^2 \} + 1/12n \, , \tag{161}$$

$$U2 = W2 - n(\bar{z} - 1/2)^2 \text{ where } \bar{z} = \sum_{i=1}^{n} z_i/n \, , \tag{162}$$

and

$$A2 = \{ \sum_{i=1}^{n} (2i - 1)[\ln(z_{(i)}) + \ln(1 - z_{(n+1-i)})]\}/n - n \quad ,(163)$$

where $z_{(i)}$ is the i-th transformed order statistic.

If the distribution being tested has one or more unknown parameters, then the EDF statistics are no longer distribution free. It is known, however, that if the unknown parameters are location and scale, then the distributions of the EDF statistics do not depend on the unknown parameters.

When testing for $F_0(y;a,b)$ with unknown a and b, location and scale parameters, respectively, the following procedure can be used. First, estimate the parameters, a and b, by maximum likelihood or some other efficient method, and calculate

$$w_{(i)} = (y_{(i)} - \hat{a})/\hat{b} \quad , \qquad (164)$$

where $\hat{a}$ and $\hat{b}$ are the estimates of a and b. Then, calculate

$$z_{(i)} = F(w_{(i)};0,1) \quad . \qquad (165)$$

In this case the $z_{(i)}$ are no longer uniform random variables and the distribution of the statistics will depend on the $F_0(y)$ being tested for.

Stephens (1974,1977,1979) found the moments and approximated the asymptotic distribution of the EDF statistics for the normal, extreme value, and logistic distributions with one or both parameters unknown, and the exponential distribution with scale parameter unknown. He also gave modifications for the test statistics for finite n in each of these cases. Durbin (1975) found the exact distributions for finite n in the case of the exponential distribution with scale parameter unknown for the statistics D, D+, and D-.

3.2 Application to the Two-parameter Exponential Distribution

Until now, very little work has been done on EDF statistics in the case of testing for the two-parameter exponential distribution. One reason for this is that there exists a simple technique to eliminate the unknown location parameter, a. If $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ are the order statistics from a sample of size n from the exponential(a,b) distribution, then

$$z_{(i)} = y_{(i+1)} - y_{(1)} \tag{166}$$

are the order statistics of a sample of size n-1 from the exponential(0,b) distribution. The z's can then be tested using the procedure above and Stephens' (1974) modifications.

As an alternative to the above method, one can proceed with the usual EDF method for distributions with unknown parameters. That is, estimate both parameters by some procedure and perform the probability integral transformation. If the maximum likelihood (or GLS) estimators are used the asymptotic distribution of the statistics is the same as for the case where a is known. This is because a(ML) has variance of the order $1/n^2$, and so is super efficient, as mentioned in chapter 1. (Darling, 1955) There will, however, be a difference in the distribution for finite sample sizes.

In most other cases involving estimation of parameters, the maximum likelihood estimators have been utilized. If the maximum likelihood estimators, $y_{(1)}$ and $\bar{y} - y_{(1)}$ are used in this case, then $w_{(1)} = y_{(1)} - y_{(1)}$ will always be equal to 0. This implies that $z_{(1)}$ is also 0, which produces an infinite value for A2. Since A2 has been found to be a very important EDF statistic (see Stephens, 1974), another procedure is desired.

A good solution is to use the unbiased generalized least squares estimates rather than the maximum likelihood estimates. It has already been noted that the GLS estimators can be expressed as functions of the ML estimators. In chapter 6, these two estimation procedures, as well as the transformation to one parameter exponential will be compared on the basis of their power against a wide range of alternatives, when testing for the two-parameter exponential distribution.

IV. Other Tests of Fit

4.1 Introduction

In addition to the EDF and regression tests, which arise out of a procedure applicable to testing for many different distributions, various special tests have frequently been proposed by different authors to test for specific distributions. Some power comparisons have been made for these tests either because they were believed to be serious rivals to the tests already mentioned, or because no other power comparisons have been made for them in the three cases under consideration in this thesis. Most of the tests were developed with the object of testing for one particular distribution, and they will be presented under the heading of that particular distribution.

4.2 Exponential Distribution

Other than the tests introduced in this thesis, only two other tests have been developed primarily for testing for the two-parameter exponential distribution. They are; the Shapiro-Wilk test which was presented in chapter 2, and the Tiku, Rai, and Mead (1974) test. The Tiku, Rai, and Mead test statistic, TRM, is the ratio of an estimator of b which is calculated on only a part of the sample, with the maximum likelihood estimator. TRM is defined,

$$TRM = (1 - 1/n)b_c/[1 - 1/n(1 - q)]b(ML) \qquad (167)$$

where

$$b_c = [(1/n)\sum_{i=1}^{n-r} y_{(i)} + qy_{(n-r)} - y_{(1)}]/(1 - q) , \qquad (168)$$

$q = r/n$, $r = [.5 + .5n]$, and $[x]$ is the greatest integer in x. It was also shown that TRM has a beta distribution, which for large sample sizes can be approximated by a normal distribution.

Several other tests were chosen to be compared on the basis of their good power properties for the one-parameter exponential distribution. (For a complete power survey for the exponential(0,b) distribution, see Stephens, (1978)) A test which is known to be uniformly most powerful against the gamma distribution alternative when testing for the exponential(0,b) distribution (Shorack, 1972), was suggested by Moran (1947,1951). The test, M, is defined

$$M = -2\sum_{i=1}^{n} \ln[y_{(i)}/\bar{y}] . \qquad (169)$$

It is also known that M has a χ^2 distribution with 2n degrees of freedom.

Another test in this category is L, introduced by Lewis (1965). He proposed the test

$$L = 2n - \sum_{i=1}^{n} iy_{(i)}/n\bar{y} \ . \qquad (170)$$

A third test was proposed by Jackson (1967). This is another test which is the ratio of two estimators of b. In this case it is the OLS estimator divided by the maximum likelihood estimator, for exponential(0,b). That is,

$$T = m'y/1'y \ , \qquad (171)$$

where y and m are the vectors of order statistics and expected order statistics, respectively.

It was necessary to modify these three tests to apply them to testing for the two-parameter exponential distribution. The obvious method is to apply the "docking-off" transformation described in the last chapter, to the order statistics, and calculate the test statistics on the z values. This was done for all three tests.

For T, however, the OLS and maximum likelihood estimators for the two-parameter exponential could also be used. In this case T becomes,

$$T2 = (m'y - \bar{y})/n(\bar{y} - y_{(1)}) \ . \qquad (172)$$

It is easily shown that

$$T1 = T2 + (n - 1)/n \ , \tag{173}$$

where T1 is the T test obtained from the "docking-off" procedure. Therefore, for Jackson's T, unlike for the EDF tests, the two procedures yield equivalent results.

Several EDF tests with special modifications for the one-parameter exponential distribution, used with the docking off technique, were also looked at. The first of these is a modified version of W2 suggest by Finkelstein and Schafer (1971). Their test statistic, FS, which was found to have good power for the one parameter case, is defined

$$FS = \sum_{i=1}^{n} \max(|F(y_{(i)};b) - (i - 1)/n|, |F(y_{(i)};b) - i/n|) \ . \tag{174}$$

The other EDF tests utilize two well known properties of the exponential distribution. The first, due to Pyke (1965), is that if $y_{(i)}$ are order statistics from an exponential distribution, that

$$d_i = (n + 1 - i)\ (y_{(i)} - y_{(i-1)}) \tag{175}$$

where y(0)=0, also have an exponential distribution with the same scale parameter. This fact is also utilized by the M, L, and TRM tests introduced above. The second property is that if

$$u_{(j)} = \sum_{i=1}^{j} y_{(i)}/n\bar{y}, \quad j = 1, \ldots, n-1 , \tag{176}$$

then the $u_{(j)}$ are ordered uniform random variables. After subtracting $y_{(1)}$, these two transformations can be made to get n - 2 uniform distributed variables which can be tested by the EDF procedure for a completely specified distribution. This procedure will be referred to as the K transformation. All six of the EDF tests were compared using this procedure.

The final two tests that were compared for the exponential distribution are based on the hazard rate defined in (9). It is known that if the hazard rate of a distribution increases (decreases) monotonically, then the normalized sample spacings, d_i, tend to increase (decrease) with i for fixed sample size. (Lin and Mudholkar, 1980) Thus, the constant hazard rate of the exponential distribution implies that the normalized spacings should be approximately the same size. Several tests have been proposed utilizing this fact. Fercho and Ringer (1972) found Gnedenko's F-test defined

$$Q(r) = (n-r)S_1/rS_u, \tag{177}$$

where

$$S_1 = \sum_{i=1}^{r} d_i \quad \text{and} \quad S_u = \sum_{i=r+1}^{n} d_i, \tag{178}$$

to have the best power among tests of this type. Q(r) is known to have an F distribution with 2r and 2(n-r) degrees of freedom. Fercho and Ringer suggest using r = [n/2].

In order to guard against alternatives with hazard rates that are not monotonic (e.g. the lognormal distribution), Lin and Mudholkar (1980) propose the test that rejects the null hypothesis if either of Fl or Fu fall in the rejection region. The two tests are defined

$$Fl = (n-2r)S_1/rS_m, \text{ and } Fu = (n-2r)S_u/rS_m, \qquad (179)$$

where

$$S_m = \sum_{i=r+1}^{n-r} d_i, \quad S_u = \sum_{i=n-r+1}^{n} d_i, \qquad (180)$$

and S_1 is defined above. The critical values are found using the approximation

$$P(a < Fl < b \;,\; a < Fu < b \mid H_0) < [P(a < F < b)]^2 \;, \qquad (181)$$

where F has an F distribution with 2r and 2(n-2r) degrees of freedom. Lin and Mudholkar found this approximation to be very good. They suggest using r = [n/10] since the quality of the approximation increases as the denominator degrees of freedom increase.

4.3 Extreme Value Distribution

For the extreme value distribution, five additional tests were studied. Two of these tests were especially designed to test for the extreme value distribution (type I) against type II and type III extreme value distribution alternatives.

The first of these was introduced by Bardsley (1977). It is the ratio between the sample standard deviation, s, and the standard deviation of the largest member of all possible pairs in the sample, s'. The test utilizes the fact that asymptotically, $s = s'$ for the type I distribution, for the type II distribution $s < s'$, and for the type III distribution $s > s'$. (Jenkinson, 1955) The test is defined,

$$B = (s^2/s'^2 - 1)/s_t, \tag{182}$$

where

$$s_t = \sqrt{\{[1/n(n - 1)] \sum_{i=1}^{n} (t_i - \bar{t})^2\}} , \tag{183}$$

$$\bar{t} = \sum_{i=1}^{n} t_i/n , \tag{184}$$

and

$$t_i = ns^2/s'^2 - (n - 1)x_i , \tag{185}$$

where x_i is the ratio of s^2/s'^2 calculated after excluding the i-th value of the sample. Bardsley presents a large sample approximation to the critical values of B, and recommends not using the approximation for sample sizes less than 45.

The second test statistic designed to test for an extreme value type I distribution against type II and type III alternatives was proposed by van Montfort and Otten (1978). The test is an approximation of the locally most powerful test against the type II and type III alternatives. Let the i-th normalized sample spacing be defined

$$d_i = [y_{(i)} - y_{(i-1)}]/(m_i - m_{i-1}) \ , \tag{186}$$

for i=2, ..., n , and g_i be the derivative of $E(d_i)$ which in the extreme value case is approximated by

$$g_i = \log\{-\log[(i-.5)/(n+1)]\} \ . \tag{187}$$

Van Montfort and Otten's test statistic, A, is defined

$$A = [(\sum_{i=2}^{n} l_i d_i / \sum_{i=2}^{n} l_i) - \bar{d}] / (v^2/n) \tag{188}$$

where l_i is an approximation of the i-th normalized sample spacing that uses h_i rather than m_i,

$$\bar{d} = \sum_{i=2}^{n} d_i/(n-1) \ , \tag{189}$$

and

$$v^2 = \sum_{i=2}^{n} (d_i - \bar{d})^2/(n - 1) \quad . \tag{190}$$

This test statistic was primarily designed as a one-sided test to be used to test against either the type II or type III distribution. The test, however, can be used as a two-sided test against all alternatives. van Montfort and Otten give approximations to the distribution of A, and also give critical points for some finite sample sizes.

The third test statistic looked at for the extreme value distribution is also based on the sample spacings. The test, S, proposed by Mann, Scheuer, and Fertig (1973), is defined

$$S = \left(\sum_{i=1}^{r} d_i \right) / \left(\sum_{i=1}^{n-1} d_i \right) \quad , \tag{191}$$

d_i is defined in (186), r = [m/2], and [x] is the greatest integer in x. The test statistic was developed to test for the type I extreme value distribution for smallest values, and utilizes the fact that the right tail of this distribution is steeper than than alternatives of the normal or distribution. It was also designed to test for the two-parameter Weibull (extreme value type III distribution for smallest values) against the three-parameter Weibull distribution (location parameter not equal to 0.) In this case when logarithms are taken, the right tail of the alternative is longer than the right tail of the extreme value distribution type I. (No reason was given for the choice of these alternatives.)

The test statistic was modified to test for the extreme value distribution of largest values by making the transformation X = -Y, and looking at the lower tail. It was also changed to a two-tailed test to test for all alternatives. Mann, Scheuer, and Fertig show that asymptotically S has a Beta distribution. Monte Carlo points were also given for sample sizes up to 25, and show that the distribution holds even for small sample sizes.

An approximation to Mann, Scheuer, and Fertig's test, replacing d_i by l_i where l_i is the same as for the test statistic, A, above, and r = [m/2], was also examined. This test is referred to as Sh.

The final two tests statistics for the extreme value distribution considered were ratio type tests of the Shapiro-Wilk type. In the normal case, W is proportional to the ratio of the generalized least squares estimate of b with the maximum likelihood estimate of b. The ratio of these two estimates, called W′, is not equivalent to W in the extreme value case, and was considered.

The last test was designed as a large sample approximation of W for the extreme value case. Since the correlation between the generalized least squares estimator and the maximum likelihood estimator is quite high even for small sample sizes, the ratio of the maximum likelihood estimator and s^2, which will be referred to as W′′, was considered.

4.4 Normal Distribution

Since goodness-of-fit tests for the normal distribution have been extensively analyzed with respect to power, (see Stephens, 1973; and Shapiro, Wilk, and Chen, 1968), only one other test was considered. This is van Montfort and Otten's test, A. In the normal distribution g_i, the derivative of $E(d_i)$, is approximated by

$$g_i = 1/4 + 3/4(F^{-1}(p_i))^2 \ , \qquad (192)$$

where $p_i = (i-7/8)/(n+1/4)$. The test statistic was designed as a one-tail test against alternatives of the Student-t distribution. However, Monte Carlo results showed that the test could be significant in either tail depending on the alternative, so in order to utilize this test statistic as a test against all alternatives it was used as a two-tailed test.

V. Monte Carlo Results and Examples

5.1 Finite Significance Points

New Statistics

The distributions for finite sample sizes of all of the statistics introduced for the first time in this thesis were approximated by Monte Carlo methods. These statistics are; G1, G2, G2h, G3, and the statistics W′, W′′, and Sh for the extreme value distribution. For each of several sample sizes, 10,000 samples from the hypothesized distribution were generated using the IMSL library (version 7), and the statistics were calculated for each sample. The finite distributions were calculated for sample sizes 5(5), 25 for all distributions. The significance points were also calculated for samples of size 50 from the normal distribution and if possible for the extreme value distribution, and samples of size 50 and 100 for the exponential distribution.

In order to avoid having a overwhelmingly large number of tables, significance points are presented only for those statistics which proved their value in the power comparisons. These statistics are; G1, G2, and G2h. Tables of the Monte Carlo significant points can be found in the Appendix.

New Cases for Existing Statistics

Several statistics that had been previously introduced were extended to new cases, and the distributions for finite n of these statistics were found. The statistics are; W for the extreme value distribution, R for the exponential distribution, and W2, U2, and A2 case 5 for the exponential distribution. Significance points for these test statistics based on 10,000 samples can be found in the Appendix.

Other Statistics

The finite distributions of some already existing statistics were also calculated. In some cases the significance points were unavailable for some of the sample sizes wanted for the power comparisons, but in most cases these results were used to check the adequacy of the Monte Carlo procedure. Either 10,000 or 5,000 samples were taken from the hypothesized distribution for these statistics.

5.2 Illustrations

Illustration 1

Thirty-two values of modulus of rupture measured on Douglas Fir and Larch 2x4s were kindly provided by Dr. W. Warren. These values were tested for exponentiality. The data and values of the regression and EDF statistics can be found in table 3. By looking at the modified values of W2, U2, and A2, and interpolating from the Monte Carlo tables for the other statistics it is seen that all statistics strongly reject the hypothesis that the variable is exponentially distributed.

Table 3

Thirty-two ordered values of modulus of rupture for Douglas Fir and Larch two-by-fours.

43.19	49.44	51.55	55.37	56.63	67.27	78.47	86.59
90.63	92.45	94.24	94.35	94.38	98.21	98.39	99.74
100.22	103.48	105.54	105.54	107.13	108.14	108.64	108.94
109.62	110.81	112.75	113.64	116.39	119.46	120.33	131.57

Values and significance levels of the regression and EDF test statistics for exponentiality calculated on the above data.

Name of Statistic	Value	Significance Level
R	.39	<.01
Rh	.36	<.01
G1	2.52	<.01
G2	2.32	<.01
G2h	1.71	<.01
G3	.66	<.01
W	.17	<.01
D+	1.11	<.025
D-	1.99	<.01
D	1.99	<.01
V	3.10	<.01
W2	1.07	<.01
U2	.78	<.01
A2	5.10	<.01

Illustration 2

van Montfort (1973) presents 47 values in cu. feet/sec. of the annual maxima of the discharges of the North Saskatchewan River at Edmonton. A random sample of twenty-five of these values were tested for the extreme value distribution. The sample and values of the statistics can be found in table 4.

Table 4

Twenty-five values of the annual maxima of the discharges of the North Saskatchewan River at Edmonton in cubic feet per second.

19.885,	20.94	21.82	24.888	27.5
28.1	28.6	30.38	31.5	38.1
39.02	40.0	40.0	40.4	44.7
50.33	51.442	58.8	61.2	65.597
66.0	84.1	106.6	121.97	185.56

Values and significance levels of the regression and EDF statistics calculated on the above data.

Name of Statistic	Value	Significance Level
R	.102	<.10
Rh	.131	>.10
G1	.176	<.05
G2	.176	<.05
G2h	.237	<.025
G3	.365	>.10
W	.310	<.02
W2	.103	<.10
U2	.079	>.10
A2	.745	<.10

Illustration 3

Modulus of elasticity was measured for sixty-four 2x4s of Douglas Fir or Larch. (Data provided by Dr. W. Warren.) A sample of 50 of these values was tested for normality. These values and the values of the regression and EDF statistics can be found in table 5. All the statistics fail to reject the hypothesis that the data is normally distributed.

Table 5

Fifty values of modulus of elasticity for Douglas Fir and Larch two-by-fours.

43.19	45.84	49.44	51.55	54.14
55.37	56.93	59.63	60.04	61.07
65.74	67.09	72.24	72.34	73.46
76.52	77.35	78.36	78.47	78.79
82.00	83.57	84.95	86.59	87.96
90.19	91.57	91.74	92.45	94.24
94.54	95.00	98.39	99.74	100.22
103.48	105.54	107.13	108.14	108.64
108.94	109.62	110.81	112.75	116.39
116.79	119.46	120.33	121.16	131.57

Values and significance levels of the regression and EDF statistics calculated on the above data.

Name of Statistic	Value	Significance Level
R	.019	>.10
Rh	.015	>.10
G1	.019	>.10
G2	.019	>.10
G2h	.018	>.10
G3	.263	>.10
W	.963	>.10
W2	.049	>.10
U2	.047	>.10
A2	.362	>.10

VI. Results and Conclusions

6.1 Introduction

In order to compare the power properties of the various test procedures for each distribution, a wide range of alternative distributions and sample sizes was investigated. The procedure was as follows:

1. A sample of size 10, 20, or 50 from some distribution other than the null distribution was generated. The random samples were generated using IMSL version 7 random generators.
2. All statistics to be compared were calculated, and it was recorded whether the statistic was significant.
3. This procedure was repeated several times and the percentage of significant samples was recorded. (Not all the statistics were compared using the same set of samples.)

The distributions generated were the appropriate alternatives for each null distribution. For each alternative distribution, 2500 (or in some cases, 1000) samples of each of size 10, 20 and 50, were generated. The maximum standard error of the power results is equal to $.5/\sqrt{n}$ which, for the cases where 2500 samples were generated, is 1%, and, for the cases where 1000 samples were generated, is equal to 1.6%.

Since the covariance matrix of the order statistics from an extreme value distribution have only been tabled for sample sizes up to n=25, any test for the extreme value distribution which utilized the GLS estimate was not calculated for samples of size 50.

A listing of all of the alternative distributions employed in the power study is given in table 6. The results of the power comparison for the extreme value distribution can be found in tables 7 to 9, for the exponential distribution in tables 10 to 16, and for the normal distribution in table 17. All results presented are for 10% level tests. The powers for a 5% level test are also available.

Table 6

Alternative distributions used in the power study.

Name	Definition of Standardized Distribution	Code
Beta(a,b)	$f(x) = \text{cnst.}\ x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1$	B(a,b)
Chi-squared(a)	$f(x) = \text{cnst.}\ x^{(a-2)/2}\exp(-x/2)\ ,\ x > 0$	C(a)
Exponential	$f(x) = \text{cnst.}\ \exp(-x)$	Exp
Extreme value type I	$F(x) = \exp[-\exp(-x)]$	EVI
Extreme value type II(a)	$F(x) = \exp[-(x)^{1/a}]\ ,\ x > 0$	EVII(a)
Extreme value type III(a)	$F(x) = \exp[-(-x)^{1/a}],\ x < 0$	EVIII(a)
F(a,b)	$f(x) = \text{cnst.}\ x^{(a-2)/2}[1+(a/b)x]^{-(a+b)/2},$ $x > 0$	F(a,b)
Half-Cauchy	y is T(1), x is $\|y\|$	HC
Half-normal	y is N, x is $\|y\|$	HN
Laplace	$f(x) = \text{cnst.}\ \exp(-\|x\|)$	L
Log-normal(a)	$f(x) = \text{cnst.}\ \exp[-(\ln x)^2/2a],\ x > 0$	LN(a)
Normal	$f(x) = \text{cnst.}\ \exp(-x^2/2)$	N
Student-t(a)	$f(x) = \text{cnst.}\ (1+x^2/a)^{-(a+1)/2}$	T(a)
Uniform	$f(x) = 1,\ 0 < x < 1$	U
Weibull(a)	$F(x) = 1 - \exp[(-x)^a]\ ,\ x > 0$	W(a)

Table 7

Power, (in %), of tests for the Extreme Value distribution against selected alternatives.

Alt. dist.	Sample size	Regression tests: Rh	R	G1	G2	G2h	G3	W	EDF tests: W2	U2	A2
B	10	9.4	7.6	6.6	8.2	14.7	8.1	11.0	16.8	16.3	18.8
(1,4)	20	7.0	7.2	6.9	7.3	14.0	8.9	11.0	25.4	26.0	30.4
	50	6.2	9.7		10.9	15.2			56.6	55.6	66.3
EVII	10	14.8	12.2	9.7	13.6	17.3	8.0	13.0	11.8	11.1	13.7
(.1)	20	18.7	16.3	15.0	16.6	22.6	6.5	16.2	13.9	12.9	16.3
	50	27.2	23.8		25.4	31.7			17.8	14.9	19.7
EVII	10	19.5	14.3	14.7	17.8	25.0	7.4	18.6	15.0	13.0	17.8
(.2)	20	29.2	25.2	26.4	27.2	35.8	4.8	30.0	22.9	20.2	27.8
	50	49.3	44.8		50.2	58.8			38.8	31.9	44.9
EVII	10	33.2	25.6	27.6	31.4	40.9	5.6	35.0	28.4	24.4	32.7
(.4)	20	53.6	49.6	51.2	53.7	62.1	3.6	57.4	48.8	43.9	56.2
	50	83.1	80.0		85.2	89.2			82.2	74.3	86.8
EVII	10	59.7	53.4	56.0	59.0	68.6	4.6	66.3	60.1	55.7	66.0
(.8)	20	85.2	83.4	85.7	86.7	90.4	5.6	90.0	87.7	84.4	91.8
	50	99.3	99.2		99.6	99.8			99.6	99.4	99.9
EVIII	10	9.5	10.7	12.0	10.4	6.2	14.6	10.6	10.2	10.2	9.4
(.1)	20	6.4	10.8	13.4	11.8	4.5	15.4	12.7	12.2	12.7	12.6
	50	5.4	12.5		18.6	4.6			17.0	15.9	17.1
EVIII	10	10.2	14.2	18.2	13.7	4.4	19.6	14.2	13.9	13.4	12.1
(.2)	20	9.6	18.4	25.8	20.7	6.2	24.1	25.0	18.1	18.4	18.4
	50	18.9	38.9		52.1	21.5			36.4	33.4	39.9
EVIII	10	24.3	32.2	34.9	32.1	11.0	31.8	28.9	25.6	24.8	24.4
(.4)	20	40.0	57.7	66.4	62.5	29.0	50.1	62.3	46.6	46.0	50.5
	50	83.2	95.9		97.9	83.5			82.9	80.4	88.7
EVIII	10	61.3	71.3	76.0	68.7	39.9	65.4	65.6	58.0	56.5	57.7
(.8)	20	93.0	97.8	95.7	97.2	85.8	78.5	91.2	90.2	89.4	92.6
	50	*	*		*	*			*	99.9	*

* = >99.9

Table 7 (continued)

Power, (in %), of tests for the Extreme Value distribution against selected alternatives.

		Regression tests							EDF tests		
Alt. dis.	Sample size	Rh	R	G1	G2	G2h	G3	W	W2	U2	A2
L	10	30.1	32.2	35.2	33.0	19.1	35.9	33.1	32.0	31.3	30.2
	20	44.2	48.7	53.8	54.0	38.0	51.9	51.6	56.4	57.0	58.1
	50	71.9	76.5		84.9	75.0			90.8	90.5	90.8
N	10	15.2	20.4	25.1	19.8	6.5	25.4	21.2	18.8	18.3	16.6
	20	22.8	35.1	42.5	38.2	16.1	36.4	41.8	31.0	30.6	33.4
	50	44.5	64.8		77.1	49.4			61.8	58.7	66.5
T (1)	10	69.9	67.7	67.3	68.8	62.8	48.8	62.5	67.0	66.0	67.3
	20	89.9	89.3	87.0	88.6	86.0	65.5	72.7	90.3	90.0	90.6
	50	99.7	99.7		99.7	99.6			99.9	99.9	99.9
T (6)	10	22.1	24.4	32.7	25.2	13.3	31.4	29.0	23.5	23.0	22.3
	20	32.9	41.5	48.0	46.4	27.2	44.1	47.4	41.6	41.9	43.5
	50	58.2	67.4		79.6	63.4			76.4	75.4	79.2
U	10	15.3	24.1	23.5	17.7	4.9	28.1	13.0	21.4	21.7	20.3
	20	27.8	52.2	46.3	40.2	10.9	39.2	20.8	37.8	38.8	42.2
	50	83.1	97.0		93.9	63.3			78.4	76.8	86.6

Table 8

Power, (in %), of tests for the Extreme Value distribution against selected alternatives.

Alt. dist.	Sample size	Att	Aut	Stt	Sut	S2tt	B	W′	W′′
B	10	11.2	2.4	17.2	4.5	16.4	12.0	15.7	5.7
(1,4)	20	13.1	1.0	24.0	2.2	24.3	13.1	30.0	8.7
	50	23.4	0.2	43.1	0.8	43.4	14.2		12.8
EVII	10	14.4	6.2	11.6	6.1	12.7	13.9	11.2	14.0
(.1)	20	17.6	4.1	14.4	4.6	15.0	15.8	11.2	15.8
	50	27.2	2.0	20.6	3.0	21.8	22.7		26.5
EVII	10	19.2	3.8	15.5	4.0	17.0	18.8	13.6	17.0
(.2)	20	29.7	2.0	24.8	2.9	26.5	25.8	16.9	29.3
	50	56.0	0.4	47.0	0.5	49.3	46.5		54.0
EVII	10	34.6	1.7	28.4	2.3	31.0	32.2	24.1	33.7
(.4)	20	58.4	0.4	53.5	0.6	55.5	51.8	40.3	57.8
	50	91.5	0.0	87.7	0.0	89.2	85.1		90.2
EVII	10	64.8	0.4	61.7	0.4	63.8	60.8	57.6	59.4
(.8)	20	90.9	0.0	89.5	0.0	90.8	86.0	86.4	88.8
	50	99.8	0.0	99.8	0.0	99.8	99.0		99.2
EVIII	10	10.0	15.1	10.9	15.3	11.0	8.8	8.6	10.8
(.1)	20	14.5	22.3	13.4	20.5	12.3	12.1	9.7	14.3
	50	22.8	35.6	16.9	26.6	17.5	17.8		22.9
EVIII	10	14.0	23.5	13.4	22.0	13.5	11.0	10.0	13.3
(.2)	20	24.1	36.7	18.3	29.5	18.1	17.0	12.4	26.4
	50	59.8	75.6	39.5	55.6	41.4	45.2		56.2
EVIII	10	32.0	46.2	26.4	39.4	26.6	16.1	12.7	30.1
(.4)	20	65.7	78.5	48.2	61.7	49.2	42.2	26.2	59.5
	50	98.6	99.7	83.8	91.6	86.5	94.1		97.2
EVIII	10	68.3	79.5	56.2	69.2	58.5	27.5	31.8	64.1
(.8)	20	97.0	98.4	88.7	93.3	90.4	65.1	60.5	89.4
	50	*	*	99.7	99.9	99.8	94.4		99.9

Table 8 (continued)

Power, (in %), of tests for the Extreme Value distribution against selected alternatives.

Alt. dist.	Sample size	Att	Aut	Stt	Sut	S2tt	B	W′	W′′
L	10	32.2	40.4	27.0	38.1	26.9	13.8	23.0	35.1
	20	51.8	58.9	43.8	56.1	43.4	16.6	48.0	55.4
	50	80.6	86.8	77.1	85.9	78.2	19.2		83.9
N	10	20.1	31.4	18.6	28.4	18.6	12.0	12.1	22.2
	20	42.4	55.5	31.9	45.8	31.9	25.8	20.2	42.2
	50	82.7	90.6	64.0	78.0	66.6	59.8		80.3
T	10	62.0	47.2	51.0	46.7	51.4	21.8	53.2	63.4
(1)	20	73.4	53.7	63.2	55.9	63.2	18.4	77.7	81.6
	50	85.6	62.8	79.7	70.9	79.4	18.6		91.9
T	10	25.3	35.0	22.9	32.3	23.0	12.0	17.8	27.8
(6)	20	48.0	57.7	37.8	51.6	37.4	19.0	33.6	49.1
	50	79.7	86.0	70.0	81.6	71.2	30.6		81.4
U	10	14.4	26.2	19.0	23.8	18.9	12.6	9.4	13.1
	20	31.4	45.9	26.7	35.3	26.4	35.0	11.0	20.2
	50	70.7	85.3	42.6	54.0	45.0	89.5		51.2

Table 9

Power, (in %), of tests for the Weibull distribution against selected alternatives.

		Regression tests							EDF tests		
Alt. dist.	Sample size	Rh	R	G1	G2	G2h	G3	W	W2	U2	A2
B	10	11.2	9.2	9.5	9.9	13.9	9.4	10.4	11.2	10.8	11.9
(1,4)	20	11.2	10.5	10.8	10.5	15.6	6.8	11.7	10.7	11.1	13.4
	50	16.3	13.1		14.4	19.9			15.2	14.2	17.0
C	10	12.1	9.8	9.6	11.2	15.1	7.6	11.7	10.9	10.2	12.0
(1)	20	15.6	13.2	11.2	14.1	20.2	6.3	13.0	12.4	12.0	14.7
	50	19.0	14.7		17.0	24.7			19.0	17.0	20.3
C	10	9.6	9.8	11.7	9.9	8.2	12.8	11.1	10.3	10.3	9.4
(3)	20	8.2	11.1	13.2	11.0	7.3	13.0	12.0	11.8	12.6	12.2
	50	7.0	10.3		12.5	5.8			13.2	12.8	12.5
C	10	9.3	10.2	12.1	10.3	7.2	14.1	11.0	11.6	11.2	10.5
(4)	20	7.4	12.2	14.0	12.0	5.7	13.0	12.5	11.8	12.3	11.6
	50	7.3	12.7		15.6	5.4			15.0	14.4	15.0
U	10	20.2	16.0	16.6	19.1	28.8	6.6	26.0	24.5	22.4	28.9
	20	31.7	27.6	32.2	32.1	43.2	6.1	42.7	41.7	39.4	51.8
	50	54.4	52.6		64.8	73.9			81.9	76.5	89.3

Alt. dist.	Sample size	Att	Aut	Stt	Sut	S2tt	B	W′	W′′
B	10	10.8	6.4	10.6	6.7	11.2	11.7	11.2	12.1
(1,4)	20	13.0	4.7	12.4	5.3	12.7	13.5	10.8	12.5
	50	19.2	2.0	17.6	2.6	18.2	15.8		14.6
C	10	12.4	5.6	10.9	6.7	11.4	12.1	10.6	11.8
(1)	20	15.5	4.4	13.9	5.2	14.4	15.5	9.8	12.7
	50	20.0	1.4	20.1	2.1	20.5	20.8		16.6
C	10	10.0	12.1	10.8	13.0	10.2	9.4	9.9	8.3
(3)	20	11.4	16.0	11.2	14.6	10.4	11.0	9.2	10.6
	50	13.0	20.2	12.3	16.4	12.0	10.5		13.3
C	10	10.5	14.1	11.4	13.4	11.0	10.0	9.7	10.4
(4)	20	12.2	18.7	12.0	17.2	11.2	11.7	9.9	12.9
	50	20.1	31.3	16.2	25.6	16.5	15.1		17.7
U	10	23.9	0.8	26.0	2.0	26.8	21.9	25.9	20.8
	20	44.0	0.2	46.8	0.6	48.6	34.3	51.5	37.5
	50	79.4	0.0	78.9	0.0	80.6	62.8		68.1

Table 10

Power, (in %), of the regression tests for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	Rh	R	G1	G2	G2h	G3	W
B	10	8.6	13.4	17.7	13.4	8.4	11.5	11.4
(1,4)	20	5.6	13.2	22.3	17.9	4.8	13.2	17.3
	50	6.7	22.2	42.8	38.6	10.0	14.6	43.5
C	10	18.4	13.4	11.0	19.4	39.0	25.4	25.0
(1)	20	23.0	17.8	20.9	28.5	41.5	31.7	39.0
	50	36.5	29.4	54.2	60.3	69.7	48.7	78.6
C	10	9.8	13.5	20.9	16.2	9.4	12.2	16.0
(4)	20	7.5	13.2	31.7	26.9	9.2	15.1	29.2
	50	9.4	19.4	62.8	58.3	29.2	31.0	72.7
HC	10	44.1	36.8	35.0	44.8	59.1	40.7	48.3
	20	68.1	63.0	64.2	68.5	75.3	61.7	71.6
	50	93.8	92.0	94.4	95.0	96.0	86.7	96.0
HN	10	10.2	15.8	24.7	18.7	8.5	13.4	17.0
	20	9.3	19.3	35.6	29.4	9.8	17.3	29.6
	50	12.6	32.2	68.2	64.3	28.6	22.9	72.4
LN	10	19.0	14.4	12.9	17.5	29.9	15.2	19.1
(1)	20	29.5	24.4	21.4	25.0	32.9	19.2	26.5
	50	49.4	42.8	42.5	45.8	52.1	26.6	47.4
LN	10	64.4	54.2	56.8	71.4	83.7	70.8	76.4
(2.4)	20	86.6	81.5	89.0	93.0	96.0	90.8	96.0
	50	99.1	98.8	99.8	99.9	*	99.6	*
U	10	40.2	56.0	66.0	58.8	33.4	26.8	48.0
	20	71.3	88.7	94.6	92.5	69.0	40.0	82.6
	50	99.5	*	*	*	99.8	66.9	99.8
W	10	40.6	30.3	30.6	46.3	67.7	48.6	54.4
(.5)	20	58.4	49.8	61.8	70.7	82.4	67.6	82.0
	50	85.8	82.5	97.5	98.2	98.9	92.9	99.4
W	10	18.8	27.4	46.8	37.5	18.4	20.8	37.8
(2)	20	22.0	38.8	76.1	71.0	39.2	34.5	73.3
	50	48.0	71.6	99.2	99.0	92.4	73.8	99.7

Table 11

Power, (in %), of tests for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	FS	J	L	MOR	TRM	BF	Q
B	10	14.3	12.7	12.5	13.0	13.7	8.9	12.7
	20	18.5	19.3	15.5	14.3	14.3	13.2	18.3
	50	30.5	42.9	37.4	25.9	30.3	20.9	27.3
C	10	28.8	28.5	33.6	38.8	31.4	22.3	31.1
(1)	20	54.2	46.1	57.2	67.7	49.9	43.6	54.9
	50	90.9	79.2	91.4	97.7	82.7	86.3	83.3
C	10	21.3	17.7	18.7	19.9	14.9	13.7	15.9
(4)	20	36.4	33.8	33.7	35.8	25.7	22.3	27.0
	50	74.5	67.7	78.7	78.5	58.2	50.0	59.0
HC	10	42.4	46.5	46.8	39.5	37.5	35.3	41.7
	20	70.1	74.8	72.8	62.0	63.0	59.1	63.4
	50	93.7	95.9	94.9	87.7	90.2	90.4	92.7
HN	10	18.3	16.0	16.1	16.1	16.8	11.3	16.7
	20	33.0	33.9	30.5	26.5	25.8	14.2	26.3
	50	61.6	70.8	69.8	50.3	53.5	30.8	55.7
LN	10	15.3	20.0	19.1	14.2	13.2	14.8	16.1
(1)	20	24.6	28.3	25.2	14.5	19.1	21.8	16.9
	50	37.2	46.2	36.4	16.6	27.1	38.5	26.8
L	10	78.6	74.7	79.8	83.2	79.2	61.5	81.2
(2.4)	20	98.3	97.4	98.7	98.6	98.1	91.6	97.2
	50	*	*	*	*	*	*	*
U	10	50.7	53.4	47.4	35.2	42.3	22.7	40.7
	20	78.2	89.7	77.1	54.3	68.5	61.6	71.8
	50	99.6	*	99.8	87.9	96.7	99.5	97.6
W	10	57.9	55.4	63.5	69.3	61.1	43.1	60.5
(.5)	20	89.2	85.0	91.5	94.0	84.4	75.3	86.6
	50	*	99.1	99.9	99.9	99.6	99.8	99.3
W	10	40.5	37.4	37.6	35.6	32.6	18.9	33.1
(2)	20	75.1	76.4	75.1	65.8	62.6	41.2	63.1
	50	99.5	99.6	99.7	97.0	97.8	86.0	97.6

Table 12

Power, (in %), of the EDF tests for the Exponential distribution, case 4 (a known), against selected alternatives.

Alt. dist.	Sample size	D+	D-	D	V	W2	U2	A2
B	10	3.6	16.8	11.7	11.4	13.0	13.5	10.6
(1,4)	20	2.5	22.7	15.4	14.4	17.7	15.7	14.4
	50	1.1	39.5	27.4	23.6	31.2	24.4	27.8
C	10	54.0	1.7	33.6	30.9	38.1	33.6	56.2
(1)	20	72.2	0.7	57.2	50.4	62.6	53.8	71.7
	50	96.5	1.0	91.4	87.3	94.5	89.0	98.3
C	10	1.2	43.3	32.5	29.7	37.9	34.4	31.7
(4)	20	1.3	68.7	56.1	50.8	63.7	55.5	61.6
	50	4.5	95.8	90.5	86.7	95.0	90.6	96.3
HC	10	53.9	2.8	42.7	36.7	45.1	37.8	46.0
	20	75.8	0.8	65.6	57.7	68.1	59.2	68.5
	50	95.4	0.0	92.8	88.5	94.2	89.1	93.9
HN	10	1.6	26.2	19.7	19.0	21.2	20.4	17.0
	20	0.8	38.6	27.0	24.8	31.7	27.2	27.4
	50	1.0	66.5	54.4	48.7	62.1	51.6	58.8
LN	10	16.0	14.4	16.6	16.3	17.4	17.5	14.9
(1)	20	22.5	15.6	22.5	23.2	25.6	26.5	24.8
	50	36.0	23.8	36.7	48.4	43.3	51.2	50.5
LN	10	91.2	0.2	83.2	78.3	86.0	79.8	89.8
(2.4)	20	99.6	0.0	98.4	97.0	99.1	97.0	99.4
	50	100.0	0.1	100.0	100.0	100.0	100.0	100.0
U	10	3.4	53.1	42.2	47.9	51.5	48.0	44.7
	20	17.6	80.0	69.0	78.5	81.2	74.7	77.5
	50	84.9	98.8	97.0	99.7	99.4	98.6	99.8
W	10	79.4	0.3	63.8	57.1	68.9	59.1	81.1
(.5)	20	95.2	0.0	91.0	84.1	92.2	86.2	96.8
	50	100.0	0.5	100.0	99.9	100.0	99.9	100.0
W	10	2.3	75.7	65.8	64.1	76.2	69.3	69.6
(2)	20	9.8	96.4	91.8	91.8	97.0	93.8	96.6
	50	49.8	100.0	100.0	100.0	100.0	100.0	100.0

Table 13

Power, (in %), of the EDF tests, case 4a, for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	D+	D-	D	V	W2	U2	A2
B	10	4.4	16.8	12.5	12.0	13.8	13.1	11.6
(1,4)	20	2.9	22.4	14.8	14.6	17.4	15.0	14.8
	50	1.2	38.7	27.0	23.0	31.0	24.0	27.0
C	10	40.3	2.5	23.0	21.6	25.3	22.2	36.6
(1)	20	63.2	1.0	46.9	40.7	51.2	44.4	64.1
	50	94.4	1.0	87.3	81.6	91.3	84.9	96.0
C	10	2.6	24.8	18.2	16.4	20.4	18.4	15.6
(4)	20	1.4	42.1	31.0	28.0	35.0	30.6	31.7
	50	2.2	78.0	68.0	63.5	74.5	67.3	74.6
HC	10	53.8	2.6	41.8	35.7	44.4	37.5	45.4
	20	75.6	0.6	65.5	57.3	68.4	58.8	68.7
	50	95.3	0.0	92.5	88.4	94.0	88.9	93.9
HN	10	1.8	25.0	17.7	17.0	19.6	18.8	15.3
	20	1.2	34.8	24.6	22.5	28.3	24.5	23.8
	50	0.9	64.9	51.8	47.1	60.1	49.8	56.6
LN	10	21.8	7.2	14.2	13.0	15.5	13.8	14.9
(1)	20	29.1	6.0	20.7	16.3	22.6	18.9	22.2
	50	42.4	5.4	32.1	29.5	36.4	32.3	35.3
LN	10	86.9	0.4	76.3	71.6	79.4	73.4	84.6
(2.4)	20	99.0	0.0	97.4	95.4	98.3	95.8	99.1
	50	100.0	0.1	100.0	100.0	100.0	100.0	100.0
U	10	3.2	48.5	37.9	43.0	45.5	43.6	39.0
	20	15.4	78.0	66.7	75.8	78.0	72.8	74.9
	50	83.0	98.9	96.8	99.7	99.6	98.3	99.6
W	10	68.4	0.5	52.5	45.2	55.4	48.2	66.4
(.5)	20	92.0	0.0	85.3	77.2	88.0	79.8	92.8
	50	100.0	0.4	100.0	99.6	100.0	99.8	100.0
W	10	1.1	42.6	33.2	31.4	38.9	34.8	31.8
(2)	20	2.9	76.0	65.4	63.4	74.2	66.4	70.1
	50	23.2	99.5	98.3	98.0	99.4	98.7	99.3

Table 14

Power, (in %), of the EDF tests, using the k transformation, for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	D+	D-	D	V	W2	U2	A2
B	10	3.4	17.0	10.6	8.9	10.8	8.8	11.6
(1,4)	20	1.9	23.9	14.2	10.0	15.8	10.6	16.5
	50	0.5	44.4	29.5	17.1	32.8	18.2	34.8
C	10	39.6	2.2	28.8	20.2	30.2	19.5	35.9
(1)	20	63.9	0.6	51.0	34.5	54.0	33.7	61.7
	50	94.0	0.0	88.6	72.6	91.2	70.4	94.8
C	10	2.0	26.5	15.2	12.1	16.5	12.1	16.1
(4)	20	0.5	43.7	30.0	18.4	32.5	19.0	32.2
	50	0.0	79.8	68.3	47.9	74.3	47.5	74.0
HC	10	56.1	2.8	50.0	42.2	50.0	43.2	48.9
	20	78.6	0.6	72.1	65.9	73.0	66.5	72.0
	50	96.6	0.0	95.1	93.0	95.3	93.0	95.3
HN	10	1.4	25.2	15.6	11.8	16.0	11.8	15.4
	20	0.5	37.6	24.4	14.0	26.4	14.9	27.2
	50	0.1	71.0	56.5	35.1	63.5	35.2	65.0
LN	10	22.0	7.0	17.9	17.4	17.8	17.6	16.4
(1)	20	32.5	5.5	26.1	24.5	27.0	25.8	25.4
	50	46.9	4.2	38.7	43.9	38.7	45.6	38.2
LN	10	87.4	0.3	81.2	68.7	82.3	68.1	85.0
(2.4)	20	99.2	0.0	98.5	95.2	98.6	94.2	99.2
	50	100.0	0.0	100.0	100.0	100.0	100.0	100.0
U	10	0.4	53.0	37.3	20.7	40.8	21.1	48.3
	20	0.0	84.9	73.0	47.4	79.4	45.4	87.1
	50	0.0	99.9	99.5	94.7	99.8	92.0	100.0
W	10	69.1	0.4	59.1	42.9	61.2	42.2	66.4
(.5)	20	92.4	0.0	87.3	74.1	89.1	72.2	92.7
	50	100.0	0.0	99.8	99.5	99.9	98.9	100.0
W	10	0.4	44.7	30.4	17.3	33.9	17.0	34.1
(2)	20	0.0	78.6	67.0	43.0	73.4	42.2	73.0
	50	0.0	99.7	99.2	93.2	99.6	90.6	99.6

Table 15

Power, (in %), of the EDF tests, case 5, for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	D+	D-	D	V	W2	U2	A2
B	10	4.3	16.0	11.8	11.6	13.7	11.9	12.8
(1,4)	20	3.1	22.0	15.1	15.6	17.6	14.4	16.2
	50	1.0	38.6	26.9	22.8	31.4	24.7	28.6
C	10	39.8	2.4	23.5	22.1	26.8	25.4	31.5
(1)	20	64.3	0.9	48.1	42.8	52.1	46.1	59.0
	50	93.8	1.0	87.3	81.5	91.3	85.7	95.0
C	10	2.5	23.6	16.4	16.1	20.0	16.7	18.0
(4)	20	1.6	41.6	31.1	29.1	35.1	29.7	33.6
	50	1.7	77.9	67.7	63.3	75.0	67.8	75.6
HC	10	53.4	2.4	42.0	36.0	45.9	40.6	47.9
	20	76.3	0.6	66.1	58.8	69.0	61.2	70.2
	50	95.1	0.0	92.5	88.4	94.2	89.6	94.1
HN	10	1.5	23.9	16.7	16.5	19.4	17.0	17.6
	20	1.4	34.2	24.8	23.9	28.3	23.2	25.8
	50	0.7	64.8	51.5	46.7	60.6	50.3	58.6
LN	10	20.3	6.7	14.4	13.0	16.1	14.5	17.0
(1)	20	29.8	5.9	21.3	17.9	23.1	19.4	24.0
	50	40.6	5.4	32.1	29.3	36.8	33.1	37.0
LN	10	86.6	0.3	76.9	71.9	80.4	77.5	82.9
(2.4)	20	99.0	0.0	97.5	95.8	98.4	96.5	99.0
	50	100.0	0.1	100.0	100.0	100.0	100.0	100.0
U	10	2.5	47.2	35.6	42.6	45.0	40.4	44.4
	20	16.7	77.6	66.8	77.4	78.1	71.2	79.0
	50	80.9	98.9	96.8	99.7	99.6	98.3	99.7
W	10	68.0	0.5	53.2	45.8	57.3	53.2	62.0
(.5)	20	92.5	0.0	86.0	78.6	88.4	82.4	91.0
	50	100.0	0.4	100.0	99.5	100.0	99.8	100.0
W	10	0.8	41.6	31.4	30.8	38.4	31.8	36.1
(2)	20	3.4	75.8	65.5	65.0	74.2	65.2	72.8
	50	20.7	99.5	98.3	98.0	99.4	98.8	99.4

Table 16

Power, (in %), of the EDF tests, case 5a, for the Exponential distribution against selected alternatives.

Alt. dist.	Sample size	D+	D-	D	V	W2	U2
B	10	4.7	16.6	7.9	10.0	7.5	9.4
(1,4)	20	3.3	22.2	11.5	12.2	12.1	13.5
	50	1.1	40.2	21.6	20.2	24.1	21.8
C	10	43.1	3.2	35.6	34.3	39.7	32.8
(1)	20	67.7	1.4	59.4	52.2	64.7	54.0
	50	95.0	1.4	91.4	87.2	94.6	89.1
C	10	2.3	22.4	9.2	10.2	8.0	10.3
(4)	20	0.8	38.9	21.7	19.2	22.6	21.3
	50	0.5	76.2	59.6	55.0	65.5	59.5
HC	10	51.9	2.4	47.8	40.1	50.4	38.0
	20	74.7	0.6	69.6	60.6	72.7	59.5
	50	94.8	0.1	92.9	88.6	94.5	88.8
HN	10	2.2	23.8	10.0	12.2	9.0	12.6
	20	1.1	34.5	18.0	17.1	18.2	18.3
	50	0.5	65.1	44.7	41.1	51.6	44.9
LN	10	19.2	6.1	16.7	14.0	17.3	12.6
(1)	20	27.5	5.1	23.0	16.2	25.4	15.4
	50	37.9	4.4	32.2	26.1	35.6	27.5
LN	10	87.6	0.7	84.6	80.1	87.7	79.1
(2.4)	20	99.3	0.2	98.5	97.3	99.3	97.2
	50	100.0	0.3	100.0	100.0	100.0	100.0
U	10	0.7	49.8	23.8	30.4	25.2	31.8
	20	3.5	79.6	57.5	68.6	68.4	66.0
	50	65.1	99.3	95.9	99.4	99.0	98.0
W	10	71.8	1.2	65.6	60.0	69.8	57.5
(.5)	20	93.6	0.3	91.0	85.2	92.8	86.2
	50	100.0	1.0	100.0	99.8	100.0	99.9
W	10	0.7	40.0	18.2	18.4	17.7	20.8
(2)	20	0.6	74.9	54.9	51.7	60.5	55.1
	50	9.8	99.5	97.2	96.9	98.9	98.0

Table 17

Power, (in %), of tests for the Normal distribution against selected alternatives.

		Regression tests								EDF tests		
Alt. dist.	Sample size	Rh	R	G1	G2	G2h	G3	W	A	W2	U2	A2
C (1)	10	75.8	78.1	77.2	78.0	76.0	61.2	83.7	32.6	74.8	74.7	79.0
	20	98.6	98.9	98.9	98.9	98.6	85.4	99.6	47.2	97.9	97.3	99.0
	50	*	*	*	*	*	99.0	*	68.1	*	*	*
C (3)	10	39.6	41.6	41.6	41.5	39.8	28.2	47.2	20.2	37.2	36.6	40.2
	20	67.7	70.6	70.6	70.7	67.6	42.6	76.9	26.9	64.1	60.6	73.2
	50	98.3	99.0	99.0	99.0	98.3	71.4	99.5	35.7	96.2	94.4	98.3
C (4)	10	31.9	33.6	33.7	33.6	32.2	23.2	38.0	19.7	31.1	30.8	33.6
	20	57.6	59.4	59.4	59.8	57.6	34.5	64.5	23.9	53.5	50.5	61.4
	50	94.7	96.6	96.6	96.6	94.8	57.6	97.9	33.0	90.8	87.1	94.5
Exp.	10	53.3	55.2	55.2	55.2	53.5	39.6	60.9	25.6	51.4	50.9	55.0
	20	83.1	85.6	85.6	85.8	83.1	57.7	90.3	32.1	81.1	78.5	87.4
	50	99.9	*	*	*	99.9	88.1	*	45.5	99.8	99.4	99.9
EV	10	21.3	21.9	21.9	21.9	21.5	16.7	23.5	15.8	21.0	20.8	23.0
	20	41.4	41.2	41.3	41.5	41.4	24.4	43.4	22.5	36.5	34.2	42.4
	50	72.9	75.4	75.4	75.4	73.3	36.7	76.1	30.5	65.5	59.6	70.7
L	10	27.1	26.5	26.6	26.5	27.3	17.4	22.4	18.6	23.3	23.5	24.4
	20	43.2	38.9	39.0	39.4	43.1	22.8	33.4	33.0	35.0	35.6	38.6
	50	73.0	69.4	69.4	69.4	73.3	36.6	49.8	61.8	64.7	66.8	66.1
LN (1)	10	67.7	68.9	69.0	68.9	67.8	52.2	72.8	35.9	65.7	64.8	68.6
	20	92.7	93.6	93.6	93.7	92.7	78.0	95.3	54.9	92.2	90.6	94.4
	50	*	*	*	*	*	98.1	*	80.4	*	*	*

Table 17 (continued)

Power, (in %), of tests for the Normal distribution against selected alternatives.

		Regression tests								EDF tests		
Alt. dist.	Sample size	Rh	R	G1	G2	G2h	G3	W	A	W2	U2	A2
T (1)	10	70.2	68.9	69.0	68.9	70.4	58.9	64.6	56.1	67.7	67.6	67.8
	20	92.9	91.8	91.8	91.8	92.8	83.8	90.1	87.1	91.1	90.8	92.0
	50	*	*	*	*	*	98.1	99.6	99.6	*	99.9	*
T (4)	10	22.6	21.9	21.9	21.9	22.7	15.3	20.5	16.8	19.4	19.3	20.4
	20	36.8	34.2	34.2	34.3	36.8	20.8	29.7	28.8	28.3	28.1	31.7
	50	63.4	60.1	60.1	60.1	63.9	33.4	44.0	53.6	47.1	47.1	50.4
T (6)	10	18.6	17.8	17.8	17.8	18.6	12.6	16.5	13.5	15.6	16.0	16.5
	20	26.0	24.0	24.0	24.1	26.0	16.2	22.2	19.4	19.0	19.3	23.7
	50	45.6	42.5	42.5	42.6	45.8	20.5	27.5	34.0	28.2	28.5	32.1
U	10	7.8	10.4	10.4	10.4	8.0	16.8	21.6	20.6	13.7	15.4	15.7
	20	8.6	15.6	15.7	15.8	8.6	23.7	35.8	49.2	24.5	27.3	32.4
	50	43.0	68.2	68.2	68.4	43.6	40.2	94.2	97.2	59.6	64.0	73.6
W (.5)	10	91.0	92.4	92.4	92.4	91.2	81.3	94.4	49.8	90.2	90.4	92.5
	20	*	*	*	*	*	97.9	*	72.1	99.9	99.8	*
	50	*	*	*	*	*	*	*	94.4	*	*	*
W (2)	10	13.7	14.4	14.4	14.4	14.1	12.4	18.8	11.6	14.1	13.9	15.0
	20	21.3	22.2	22.2	22.3	21.2	15.9	26.4	15.4	20.6	19.4	26.0
	50	44.8	50.1	50.1	50.2	45.3	21.8	58.8	15.4	38.7	33.5	45.0
W (3.3)	10	8.2	8.2	8.2	8.2	8.3	9.4	13.9	10.0	9.2	10.2	9.7
	20	6.8	6.9	7.0	7.1	6.8	8.6	8.6	9.6	9.9	10.2	11.0
	50	5.3	6.4	6.4	6.4	5.3	10.2	10.5	10.7	8.4	8.9	8.9
W (4)	10	9.0	9.3	9.4	9.3	9.1	9.2	14.9	8.6	9.8	10.6	10.4
	20	8.0	7.7	7.7	7.7	7.9	9.3	9.0	9.2	10.0	10.5	11.8
	50	6.1	7.1	7.1	7.1	6.2	10.8	10.4	10.2	10.1	10.4	10.7

6.2 Results of Power Comparisons

The Extreme Value (Weibull) Distribution

1. As expected there is no one statistic with best power for all alternatives and sample sizes.
2. Several statistics have poor power relative to the other statistics, and can be eliminated from further discussion. They are: Bardsley's statistic, B; G3; and the ratio of the GLS and ML estimators, W'.
3. Statistics which utilize h, the vector with elements $F^{-1}[i/(n+1)]$, have slightly better power (< 5% improvement) than the corresponding statistic using the vector of expected values of the order statistics against some alternatives, but in several cases have much less power than the statistics using m. It is clear that, whenever possible, statistics using m such as R and G2, should be used rather than the corresponding statistics, Rh and G2h. It should be pointed out that this drop in power does not occur when the spacings are used rather than the order statistics. The statistics S and Sh have very similar power properties. In fact, it seems that, overall, Sh has slightly better power

than the actual statistic proposed by Mann, Scheuer, and Fertig, S.

4. S and Sh have very similar power properties with von Montfort and Otten's statistic, A. With the exception of the alternative Beta(1,4) distribution, S and Sh are beaten by von Montfort and Otten's statistic, A, in all cases, both as a one-tailed and as a two-tailed test.
5. All three of the EDF statistics compared have good overall power properties. A2 is clearly the most powerful of the three, with W2 marginally better than U2.
6. Of the regression statistics, R, G1 and G2, R has slightly lower overall power. G1 has the best overall power of the three for samples of size 10 and 20, but cannot be calculated for samples of size greater than 25.
7. The Shapiro-Wilk statistic, W, has fairly good overall power for samples of size 10 and 20. W'', the ratio of the squared ML estimator and s^2, has power very similar to that of W, and requires no stored data. Therefore, it should be used rather than W even for small sample sizes.
8. All of the following statistics have good power, with high relative power for some alternatives and low relative power for other alternatives. These statistics are A, A2, G2, and W''. Of these four, A is overall marginally better than the others.

Comparison of Power Results with Results of Other Power Studies

1. These results for the Weibull and Extreme Value distributions agree with other results in this field.
2. Mann, Fertig and Scheuer (1971) compared the one-tailed version of S with other statistics, including W2 and A2, for sample sizes 5 and 10, and found S to be better, when testing for the Weibull distribution. The only alternative distributions compared were the 3-parameter Weibull and log-normal (normal for testing for e.v.) distributions.
3. Littell, McClave, and Offen (1979) compared statistics including A2, W2, Rh and S for a limited number of alternatives and found A2 and W2 overall best. Their results differ a bit from those in this study in that their results give Rh slightly better power than is indicated here.
4. van Montfort and Otten (1978) compare one-sided versions of A and B when testing against extreme value type II or III alternatives, and find that A is uniformly superior to B. These results strongly support this conclusion.

<u>The Exponential Distribution</u>

1. The EDF statistics are overall better than the other statistics compared with A2 and U2 having the best power of the EDF statistics.

a. The EDF statistics using ML estimators (case 5a) have no better power than the statistics using GLS estimators (case 5). Since A2, which has been found to be a very powerful statistic when testing for other distributions, cannot be calculated using ML estimators, estimation by GLS is the preferred procedure.

b. The EDF statistics using the k transformation (mormalized spacings) and the docking off procedure are marginally more powerful than the EDF statistics using the docking off procedure only.

c. A2 using GLS estimators, and A2 using the procedure of subtracting $y_{(1)}$ and performing the K transformation have the best overall power. They are marginally more powerful than W2 in either of those two situations as well as W2 using ML estimators and W2 using the docking off procedure alone. Either A2 or W2 with any of these procedures is a powerful test statistic.

d. Because of the simplicity, and superior power of the case 5 EDF statistics, this procedure is recommended over the other three.

2. The regression statistics do very badly in this case with G3, Rh, G2h, and R having the worse power overall among all statistics compared. G1 and G2 have the best power among the regression statistics, but much worse power than the EDF statistics especially against alternatives chi-square with 1

degree of freedom and the Weibull(.5) distribution.

3. The procedures utilizing h have far less power than the corresponding statistics using m. Since for the exponential distribution, m is easily calculated, the expected values of the order statistics should not be approximated by h.
4. Lin and Mudholkar's statistic, BF, and Gnedenko's statistic, Q, do very badly relative to the other statistics.
5. The statistics which are a ratio of two estimators of the scale parameter; Jackson's J statistic, the L statistic proposed by Lewis, and Shapiro-Wilk's W, all do very well overall. Of the three procedures, Shapiro-Wilk is marginally best, but it was earlier shown not to be consistent and should not be used for this reason.
6. Moran's statistic, M, which is uniformly most powerful against gamma alternatives, does very well against chi-square (as expected), and also Weibull alternatives. It has very poor power, however, against some alternatives including the lognormal(0,1) distribution.

Comparison of Power Results with Results of Other Power Studies

1. Tiku, Rai, and Mead (1974) compared their statistic, TRM, with the Shapiro-Wilk statistic, W, against a wide range of alternative distributions. The power results agree, but they conclude that TRM has power of "comparable magnitude" with

W, whereas in this thesis it is concluded that, overall, TRM is much worse than W.

2. In general the power results are comparable to Stephens' (1978) study of the one-parameter exponential distribution. That is, modified versions of the statistics that were shown to be powerful for the one-parameter exponential distribution, were shown to be powerful in this study.
3. It has been pointed out by Dyer (1974) and Stephens (1974) that it is better not to know the true mean and variance of the alternative distribution, but rather to estimate them from the data when testing for the normal distribution using the EDF statistics. It can be seen, using the case 4 (location known) and case 5 power tables, that when testing for exponentiality there is substantially less power for some alternatives when the location parameter is estimated from the data than when it is known. Therefore, in this case it is better to know the true value of the location parameter rather than estimate it from the data.

The Normal Distribution

1. With the exception of the G3 statistic and van Montfort and Otten's statistic, A, all the statistics compared are very close in power.
2. Of the EDF statistics, A2 had at least as good, or better

power than W2 or U2 against all alternatives. W2 and U2 had very similar power properties with W2 having a slight edge.

3. The three regression statistics R, G1, and G2 had almost identical distributions under the null hypothesis, and likewise, nearly identical power properties.
4. Substitution of h for m results in changes in power of up to 5%. The statistics using h, Rh and G2h, have better power than the statistics using m, R and G2, for wide tailed alternatives, and worse power for alternatives that are bounded. The drop in power from using Rh and G2h rather than R and G2 for the bounded alternatives is, in general, larger than the increase in power for wide tailed alternatives.
5. The Shapiro-Wilk statistic, W, is the statistic with the best power agianst alternatives that are bounded, but has worse power than the EDF or other regression statistics against alternatives with wider tails than the normal distribution.
6. van Montfort and Otten's two-tailed A statistic has very bad power relative to the other tests. It was the most powerful test compared agafnst a uniform distribution alternative. All significant values of A calculated on samples from the uniform distribution were significant in the lower tail. For most other distributions compared, including the t alternatives for which A was designed, the significant values of A were primarily in the upper tail.

Comparison of Power Results with Results of Other Power Studies

1. Shapiro, Wilk, and Chen (1968) did a comprehensive study of the major tests for normality, including, D, W2, A2, χ^2, and W. They found W to be far superior among all tests compared. The results for W in this thesis match Shapiro, Wilk, and Chen's results very closely. It has been pointed out by Stephens (1974), that the powers of W2 and A2 in Shapiro Wilk and Chen's study were incorrect.
2. The powers of the EDF statistics and W agree with Stephens' (1974) results. He concluded that W had marginally better power than W2 or A2.
3. Warren (1980) compared D, A2 and Weisberg and Bingham's approximation of R against a wide range of Weibull alternatives. The results in this thesis agree very closely with his results.

6.3 Conclusions and Future Research

Some overall conclusions on goodness-of-fit testing can be made from the results presented in this thesis.

1. Regression tests do relatively well for the extreme value and normal distributions, but the poor power properties of the tests for the exponential distribution indicate that these tests are not desirable for general goodness-of-fit

testing. The statistics, G1 and G2, proposed in this thesis tend to be more powerful than the most common regression statistic, R, but it is doubtful whether this increased power is worth the extra computation needed to calculate these statistics.

2. It was shown that the Shapiro-Wilk test, W, for exponentiality is not a consistent test. For the normal distribution, arguments for the consistency of W can be presented, but there is no reason to believe that this test will be consistent for any other distribution. Since ratio goodness-of-fit tests do not utilize any specific property of the distribution being tested for, there is no reason to believe that any other ratio type test of fit will be a consistent test. In fact, the consistency has not been shown for any other ratio test presented in this thesis. Thus, another possible principle of goodness-of-fit is that tests of fit should utilize a characterization or unique property of the distribution being tested. Several statistics that do utilize specific characterictics of the hypothesized distribution have been presented in this thesis, and any of these tests are preferable to the ratio tests. Since the distribution function and expected order statistics are unique properties of a distribution, both the regression and EDF type tests are tests which utilize characterizations of the distribution being tested.

3. Several tests which have been presented as tests for specific distributions were found to have no better power than the EDF and regression statistics, and in some cases much worse power. The results indicate that there is little to be gained by using these specialized tests of fit.
4. There is much more work to be done on both the asymptotic and finite distributions of goodness-of-fit tests. Also, with the exception of the EDF tests, and now the regression tests, few results have been found on the consistency of tests of fit, and hence, much more work is needed in this area.
5. There are several distributions including the gamma distribution and the three-parameter Weibull distribution with all parameters unknown which cannot as yet be tested for. Application of regression and EDF tests to other important distributions such as these would allow these tests to be used as indicators of the distribution which best fits the data, possibly by comparing the significance levels of the particular statistic when testing for the different distributions.

Appendix

Monte Carlo significance points for tests of the Exponential distribution.

Statistic	n	Significance level (%) 15	10	5	2.5	1
R	5	.178	.200	.260	.328	.392
	10	.132	.156	.192	.232	.272
	15	.110	.128	.162	.194	.232
	20	.096	.112	.140	.168	.210
	25	.082	.096	.122	.152	.188
	50	.056	.066	.084	.108	.142
	100	.036	.042	.056	.074	.105

Statistic	n	Significance level (%) 15	10	5	2.5	1
G1	5	.303	.401	.604	.848	1.239
	10	.229	.283	.401	.555	.789
	15	.183	.227	.307	.400	.550
	20	.157	.191	.261	.339	.446
	25	.133	.163	.217	.273	.359
	50	.083	.101	.133	.167	.223
	100	.051	.061	.081	.101	.139

Statistic	n	Significance level (%) 15	10	5	2.5	1
G2	5	.272	.322	.391	.481	.727
	10	.208	.251	.330	.423	.560
	15	.170	.206	.271	.344	.450
	20	.148	.180	.238	.300	.382
	25	.126	.154	.202	.250	.330
	50	.082	.098	.128	.164	.212
	100	.050	.060	.080	.100	.140

Statistic	n	Significance level (%) 15	10	5	2.5	1
G2h	5	.370	.406	.466	.510	.556
	10	.256	.294	.354	.408	.474
	15	.204	.234	.290	.344	.413
	20	.170	.202	.252	.298	.370
	25	.114	.172	.222	.270	.326
	50	.094	.116	.148	.184	.230
	100	.058	.070	.096	.120	.158

Statistic	n	Significance level (%) 15	10	5	2.5	1
W2	5	.102	.117	.141	.166	.197
	10	.121	.140	.176	.209	.265
	15	.129	.151	.188	.230	.281
	20	.133	.156	.198	.241	.292
	25	.138	.162	.202	.249	.313
	50	.143	.166	.206	.248	.310
	100	.144	.166	.210	.258	.320

Statistic	n	Significance level (%) 15	10	5	2.5	1
U2	5	.083	.093	.113	.131	.153
	10	.095	.109	.131	.156	.189
	15	.010	.114	.140	.166	.196
	20	.102	.119	.147	.172	.209
	25	.105	.122	.148	.176	.214
	50	.109	.123	.149	.176	.213
	100	.110	.124	.152	.184	.224

Statistic	n	Significance level (%) 15	10	5	2.5	1
A2	5	.555	.621	.735	.848	.989
	10	.660	.752	.926	1.096	1.338
	15	.715	.815	1.012	1.208	1.496
	20	.747	.862	1.070	1.290	1.555
	25	.780	.891	1.102	1.351	1.720
	50	.837	.950	1.161	1.377	1.740
	100	.854	.978	1.218	1.462	1.794

Monte Carlo significance points for tests of the Extreme Value distribution.

Statistic	n	15	10	5	2.5	1
		Significance level (%)				
G1	5	.232	.280	.360	.456	.580
	10	.170	.206	.264	.319	.400
	15	.134	.158	.206	.252	.317
	20	.112	.134	.172	.212	.268
	25	.096	.116	.152	.184	.228

Statistic	n	15	10	5	2.5	1
		Significance level (%)				
G2	5	.228	.264	.326	.377	.451
	10	.166	.196	.250	.300	.372
	15	.130	.154	.198	.248	.308
	20	.110	.130	.168	.210	.266
	25	.094	.114	.150	.184	.230
	50	.059	.071	.092	.118	.146

Statistic	n	15	10	5	2.5	1
		Significance level (%)				
G2h	5	.294	.334	.396	.442	.482
	10	.198	.230	.286	.342	.397
	15	.156	.184	.230	.273	.334
	20	.128	.154	.200	.243	.304
	25	.112	.136	.172	.214	.270
	50	.067	.083	.109	.137	.179

Statistic	n	Upper-tail Significance level (%)				
		15	10	5	2.5	1
W	5	1.090	1.159	1.242	1.295	1.323
	10	.949	1.004	1.084	1.152	1.224
	15	.884	.931	1.006	1.064	1.133
	20	.843	.886	.949	1.005	1.060
	25	.820	.858	.917	.962	1.022

Statistic	n	Lower-tail Significance level (%)				
		15	10	5	2.5	1
W	5	.599	.460	.388	.335	.282
	10	.583	.447	.387	.337	.294
	15	.559	.455	.402	.358	.314
	20	.556	.461	.408	.362	.319
	25	.559	.472	.426	.387	.337

References

Aitken, A.C. (1935). On least squares and linear combinations of observations. Proc. Roy. Soc. Edin., 55, 42-8.

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Ann. Math. Stat., 23, 193-212

Anderson, T.W. and Darling, D.A. (1954). A test for goodness-of-fit. J. Amer. Stat. Assoc., 49, 765-9.

Bardsley, W.E. (1977). A test for distinguishing between extreme value distributions. J. Hydrol., 34, 377-381.

Barnett, V. (1975). Probability plotting methods and order statistics. Appl. Stat., 24, 95-108.

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

Buse, A. (1973). Goodness of fit in Generalized Least Squares Estimation. Amer. Statistician, 27, 106-8.

Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilita, Giorn. Ist. Ital. Attuari., 4, 421-4.

Chernoff, H. and Lieberman, G. (1956). The use of generalized probability paper for continuous distributions. Ann. Math. Stat., 27, 806-18.

Cramer, H. (1928). On the composition of elementary errors. Second paper: Statistical applications. Skand. Aktuartidskr., 11, 141-180.

Darling, D.A. (1955). The Cramer-Smirnov test in the parametric case. Ann. Math. Stat., 26, 189-211.

Davis, C.S. and Stephens, M.A. (1977). The covariance matrix of normal order statistics. Comm. Stat.-Simul. Comput., B, 6, 75-81.

Davis, C.S. and Stephens, M.A. (1978). Approximating the covariance matrix of normal order statistics. Appl. Stat., 27, 206-212.

de Wet, T. and Venter, J.H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. Ann. Stat., 1, 380-387.

Durbin, J. (1973). Distribution Theory for Tests Based on the Sample Distribution Function. Philadelphia: Society for Industrial and Applied Mathematics.

Durbin, J. (1975). Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika, 62, 5-22.

Dyer, A.R. (1974). Comparisons of tests for normality with a cautionary note. Biometrika, 61, 185-189.

Fercho, W.W. and Ringer, L.J. (1972). Small sample power of some tests of the constant failure rate. Technometrics, 14, 713-724.

Finkelstein, J.M. and Schafer, R.E. (1971). Improved goodness-of-fit tests. Biometrika, 58, 641-645.

Fisher, R.A. (1921). On the mathematical foundations of theoretical statistics. Phil. Trans., A, 222, 309.

Gerlach, B. (1979). A consistent correlation-type goodness-of-fit test; with application to the two-parameter Weibull distribution. Math. Operationsforsch. Statist., Ser. Stat., 10, 427-452.

Glivenko, V. Sulla determinazione empirica della legge di probabilita, Giorn. Ist. Ital. Attuari., 4, 92-99.

Hoeffding, W. (1953). On the distribution of the expected values of the order statistics. Ann. Math. Stat., 24, 93-100.

Hogg, R.V. and Craig, A.J. (1956). Sufficient statistics in elementary distribution theory. Sankya, 17, 209-216.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quar. J. Roy. Meteor. Soc., 81, 158-171.

Kendall, M.G. and Stuart, A. (1973). The Advanced Theory of Statistics, Vol. 2, third edition. London: Griffin.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Attuari., 4, 83-91.

Kuiper, N.H. (1960). Tests concerning random points on a circle. Proc. Koninkl. Nederl. Akad. Van Wettensch., Ser. A, 63, 38-47.

LaBrecque, John (1977). Goodness-of-fit tests based on nonlinearity in probability plots. Technometrics, 19, 292-306.

Lewis, P.A.W. (1973). Some results on tests for Poisson processes. Biometrika, 52, 67-77.

Lin, C. and Mudholkar, G.S. (1980). A test of exponentiality based on the bivariate F distribution. Technometrics, 22, 79-82.

Littell, R.C., McClave, J.T., and Offen, W.W. (1979). Goodness-of-fit tests for the two parameter Weibull distribution. Comm. Stat.-Simul. Comput., B, 8, 257-269.

Lloyd, E.H. (1952). Least-squares estimation of location and scale parameters using order statistics. Biometrika, 39, 88-95.

Mann, N.R. (1968). Results on statistical estimation and hypothesis testing with application to the Weibull and extreme-value distributions. Aerospace Research Laboratories Report, ARL 68-0068, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio.

Mann, N.R., Fertig, K.W., and Scheuer, E.M. (1971). Confidence and tolerance bounds and a new goodness-of-fit test for the two-parameter Weibull or extreme-value distributions (with tables for censored samples of size 3(1)25). Aerospace Research Laboratories Report, ARL 71-0077, Air Force Systems Command, United States Air Force, Wright-Patterson Air Force Base, Ohio.

Mann, N.R., Scheuer, E.M., and Fertig, K.W. (1973). A new goodness-of-fit test for the two-parameter Weibull or extreme-value distribution. Comm. Stat. 2, 383-400.

Moran, P.A.P. (1947). The random division of an interval - part 1. J. Roy. Stat. Soc., B, 9, 92-98.

Moran, P.A.P. (1951). The random division of an interval - part 2. J. Roy. Stat. Soc., B, 13, 147-150.

Otten, A. and van Montfort, M.A.J. (1978) The power of two tests on the type of distributions of extremes. J. of Hydrology, 37, 195-199.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. Phil. Mag., (5), 50, 157.

Pearson, E.S. and Hartley, H.O. (1972). Biometrika Tables for Statisticians, Volume 2. Cambridge: University Press.

Pearson, E.S. and Stephens, M.A. (1962). The goodness-of-fit tests based on W and U . Biometrika, 49, 397-402.

Pyke, R. (1965). Spacings. J. Roy. Stat. Soc., B, 27, 421-427.

Sarkadi, K. (1975). The consistency of the Shapiro-Francia test. Biometrika, 62, 445-450.

Shapiro, S.S. and Francia, R.S. (1972). An approximate analysis of variance test for normality. J. Amer. Stat. Assoc., 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. and Wilk, M.B. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics*, 14, 355-370.

Shapiro, S.S., Wilk, M.B., and Chen, H.J. (1968). A comparative study of various tests for normality, *J. Amer. Stat. Assoc.*, 63, 1343-1372.

Shorack, G.R. (1972). The best test of exponentiality against gamma alternatives. *J. Amer. Stat. Assoc.*, 67, 213-214.

Smirnov, N.V. (1936). Sur la distribution de (Criterium de M. v. Mises) *C. R. Acad. Sci. Paris*, 202, 449-452.

Smirnov, N.V. (1939). Sur les escartes de la courbe de distribution empirique. *Mat. Sbornik (N.S.)*, 6, (48), 3-26.

Smith, R.M. and Bain, L.J. (1976). Correlation type goodness-of-fit statistics with censored sampling. *Comm. Stat.-Theor. Meth.*, *A*, 5, 119-132.

Stephens M.A. (1963). The distribution of the goodness-of-fit statistic U . I. *Biometrika*, 50, 303-313.

Stephens M.A. (1964). The distribution of the goodness-of-fit statistic U . II. *Biometrika*, 51, 393-398.

Stephens, M.A. (1965). The goodness-of-fit statistic V : distribution and significance points. *Biometrika*, 52, 309-321.

Stephens, M.A. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *J. Roy. Stat. Soc.*, *B*, 32, 115-122.

Stephens, M.A. (1974). EDF statistics for goodness-of-fit and some comparisons. J. Amer. Stat. Assoc., 69, 730-737.

Stephens, M.A. (1975). Asymptotic properties for covariance matrices of order statistics. Biometrika, 62, 23-27.

Stephens, M.A. (1977). Goodness of fit for the extreme value distribution. Biometrika, 64, 583-588.

Stephens, M.A. (1978). Goodness of fit tests with special reference to tests for exponentiality. Technical Report No. 262, Stanford University, Stanford, California.

Stephens, M.A. (1979). Tests of fit for the logistic distribution based on the empirical distribution function. Biometrika, 66, 591-595.

Tietjen, G.L., Kahaner, D.K., and Beckman, R.J. (1977). Variances and covariances of the normal order statistics for sample sizes 2 to 50. in Selected Tables in Mathematical Statistics, Vol V, eds. Owen, D.B. and Odeh, R.E., Providence, R.I.: American Mathematical Society.

Tiku, M.L., Rai, K., and Mead, E. (1974). A new statistic for testing exponentiality. Comm. Stat., 3, 485-493.

van Montfort, M.A.J. (1973). An asymmetric test on the type of the distribution of extremes. Meded. Landbouwhogeschool Wageningen, 73-18.

van Montfort, M.A.J. and Otten, A. (1978). On testing a shape parameter in the prescence of a location and a scale parameter. Math. Operationsforsch. Stat. Ser., 9, 91-104.

von Mises, R. (1931). Wahrscheinlichkeitsrechnung. Leipzig: Wein.

Warren, W.G. (1980). The power of some tests of normality against Weibull alternatives. Unpublished Manuscript.

Watson, G.S. (1961). Goodness-of-fot tests on a circle. Biometrika, 48, 109-114.

Wiesberg, S. (1974). An empirical comparison of the percentage points of W and W′. Biometrika, 61, 644-646.

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

White, J.S. (1967). The moments of log-Weibull order statistics. General Motors Research Publication, GMR-717, General Motors Research Laboratories, Warren, Michigan.

Wilk, M.B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.