

**Goodness of fit tests for the extreme value distribution  
based on regression, EDF  
and the stabilized probability plot**

**Mir Nabi Pirouzi Fard**

**and**

**Björn Holmquist**

**2007:1**



**LUND**  
UNIVERSITY

**DEPARTMENT OF  
STATISTICS**

S-220 07 LUND  
SWEDEN

**Goodness of fit tests for the extreme value distribution based on  
regression, EDF and the stabilized probability plot**

BY MIR NABI PIROUZI FARD AND BJÖRN HOLMQUIST

*Department of Statistics, Lund University, Box 743, S-220 07 Lund, Sweden*

mirnabi.pirouzifard@stat.lu.se, bjorn.holmquist@stat.lu.se

SUMMARY

Procedures for goodness-of-fit test of the extreme value distribution are investigated. The different procedures are all based on order statistics and take into account the dependence structure within the ordered sample. The power of the test statistics are examined, and shows that two tests that outperform the others can be found.

*Some key words:* Empirical Distribution Function; Extreme value distribution; Generalized Least Squares; Means, variances and covariances of order statistics; Regression test.

1. INTRODUCTION

The extreme value distribution is widely used in lifetime testing, in the study of size effects on material strengths, the reliability of systems made up of a large number of components, in assessing the level of air pollution and in the flood frequency analysis. This distribution has an important role in modelling lifetime data and hence considerable efforts have been dedicated to testing the hypothesis of extreme value distribution. For reviews of the subject the reader is referred to D'Agostino and Stephens (1986) and Balakrishnan and Rao (1998).

Let  $Y$  have an extreme value distribution with cumulative distribution function

$$F(y) = 1 - \exp(-\exp(\frac{y - \alpha}{\beta})) \quad -\infty < y < \infty \quad (1)$$

where the parameters  $\beta > 0$  and  $-\infty < \alpha < \infty$ . The mean and variance of this distribution, sometimes referred to as the Gumbel distribution, are respectively,

$$E(Y) = \alpha - \gamma\beta \quad \text{var}(Y) = \beta^2\pi^2/6$$

where  $\gamma \approx 0.57721$  is Euler's constant.

In this paper we consider tests of fit based either on regression, the Empirical Distribution Function (EDF) or the stabilized probability plot. A test of fit is a test of  $H_0$ : a random sample of  $Y$ -values comes from an extreme value distribution with unknown parameters  $\alpha$  and  $\beta$ . In section 2 we discuss the test statistics for random samples from an extreme value distribution. The results of the power comparisons and tables of significance points are given in section 3.

## 2. TEST STATISTICS

Goodness of fit tests mostly require estimation of location and scale parameters in the tested distribution  $F(y)$  which is the cumulative distribution function in (1).

Let  $x_1 \leq \dots \leq x_n$  denote an ordered random sample of size  $n$  from equation (1) with  $\alpha = 0$  and  $\beta = 1$ , and

$$m_i = E(x_i) \quad (i = 1, \dots, n) \quad \text{and} \quad \sigma_{ij} = \text{cov}(x_i, x_j) \quad (i, j = 1, \dots, n)$$

Define  $m$  to be the  $(n \times 1)$  vector of the expected values  $m_i$ ,  $\Sigma$  the  $(n \times n)$  matrix of variances and covariances  $\sigma_{ij}$ . If we let  $Z$  be a vector of ordered random

observations from equation (1) for general  $\alpha$  and  $\beta$ , then the elements  $z_i$  of  $Z$  may be expressed as

$$z_i = \mu + \theta x_i, \quad i = 1, \dots, n \quad (2)$$

where  $\mu = \alpha$  is a location parameter and  $\theta = \beta$  a scale parameter.

In a regression model

$$z_i = \mu + \theta m_i + \varepsilon_i, \quad i = 1, \dots, n \quad (3)$$

the points of the  $n$  pairs  $(m_i, z_i)$  should be approximately a straight line with intercept  $\mu$  on the vertical axis and slope  $\theta$ . The parameters  $\mu$  and  $\theta$ , in equation (3) can be estimated by a suitable method.

The Ordinary Least Squares (OLS) is a method for obtaining estimates of parameters in equation (3). If the variances of the dependent variable in equation (3) are constant and the covariances are equal to 0 the estimated parameters by this method are the minimum variance linear unbiased estimators of  $\mu$  and  $\theta$ .

But since the observed values in equation (3) are order statistics with  $\text{var}(z_i) = \theta^2 \text{var}(x_i)$  and  $\text{cov}(z_i, z_j) = \theta^2 \text{cov}(x_i, x_j)$ , and  $\text{var}(x_i)$  depend on  $i$  and also  $\text{cov}(x_i, x_j)$  depends on  $i$  and  $j$ . Thus the OLS estimators will not be the minimum variance estimators.

The best linear unbiased estimates of  $\mu$  and  $\theta$  can be obtained from the generalized least-squares (GLS) regression of the order statistic (Aitken, 1935; Lloyd, 1952) and are given by

$$\hat{\mu} = \frac{m' \Sigma^{-1} (m \mathbf{1}' - \mathbf{1} m') \Sigma^{-1} Z}{\mathbf{1}' \Sigma^{-1} \mathbf{1} m' \Sigma^{-1} m - (\mathbf{1}' \Sigma^{-1} m)^2} \quad (4)$$

and

$$\hat{\theta} = \frac{\mathbf{1}' \Sigma^{-1} (m \mathbf{1}' - \mathbf{1} m') \Sigma^{-1} Z}{\mathbf{1}' \Sigma^{-1} \mathbf{1} m' \Sigma^{-1} m - (\mathbf{1}' \Sigma^{-1} m)^2} \quad (5)$$

where  $\mathbf{1}$  is a  $n$ -dimensional vector of ones.

Thus the  $i$ th fitted value  $\hat{z}_i$  is given by the equation

$$\hat{z}_i = \hat{\mu} + \hat{\theta}m_i \quad i = 1, \dots, n \quad (6)$$

The use of GLS requires information on the expected values and the variance-covariance matrix of the order statistics from the standard extreme value distribution.

Lieblein and Zelen (1956) presented the expected values, variances and covariances of the order statistics from the standard extreme value distribution for  $n = 1(1)6$ . Lieblein and Salzer (1957) presented a table of expected values of order statistics for  $n = 1(1)10(5)25$  and the first 26 largest values for  $n = 30(5)60(10)100$ . White (1967, 1969), tabulated means and variances of order statistics for  $n = 1(1)50(5)100$ . For sample sizes  $n = 1(1)15(5)30$ , Balakrishnan and Chan (1992) presented tables of means, variances and covariances of the order statistics.

To calculate  $\hat{\mu}$  and  $\hat{\theta}$  in equations (4) and (5) we use the approximate values of means of order statistics, suggested by Pirouzi Fard and Holmquist (2007)

$$m_i \approx \begin{cases} -\log n - \gamma & \text{for } i = 1, \\ \log(-\log(1 - ((i - 0.4866)/(n + 0.1840)))) & \text{for } i = 2, \dots, n, \end{cases} \quad (7)$$

We also apply the approximate variances and covariances of the order statistics, given by Pirouzi Fard and Holmquist (2006) in an unpublished statistical research report as

$$\sigma_{ij} \approx \begin{cases} \pi^2/6 & \text{for } i = j = 1 \\ \frac{(i-0.469)[(n+0.831-i)(n+0.073)]^{-1}}{\log(\frac{n+0.831-i}{n+0.356}) \log(\frac{n+0.779-i}{n+0.356})} & 1 \leq i \leq j \leq n \end{cases} \quad (8)$$

and  $\sigma_{ji} = \sigma_{ij}$  is the covariance of the  $i$ th and  $j$ th order statistics of standard extreme value distribution.

### 2.1 Regression tests based on residuals

We are interested in using residuals to test how well the data fit  $\{\hat{z}_i\}$ . The residuals can be expressed as  $z_i - \hat{z}_i$  i.e. the differences between the observed values and the values given by the model. We will examine two methods to measure linearity of data. The first method of measure linearity is the error (or unexplained) sum of squares (ESS) divided by the total sum of squares (TSS) given by

$$T_1 = \frac{\text{ESS}}{\text{TSS}} = \frac{(Z - \hat{Z})'(Z - \hat{Z})}{(Z - \bar{Z}1)'(Z - \bar{Z}1)}$$

where  $\hat{Z}$  is the vector of estimated values (by using the GLS regression) with elements given in equation (6) and  $\bar{Z} = (1'Z)/(1'1)$ .

The other method to measure linearity of data is

$$T_2 = \frac{\text{GESS}}{\text{GTSS}} = \frac{(Z - \hat{Z})'\Sigma^{-1}(Z - \hat{Z})}{(Z - \bar{Z}_*1)'\Sigma^{-1}(Z - \bar{Z}_*1)}$$

where GESS is the generalized error (or unexplained) sum of squares, GTSS is the generalized total sum of squares and  $\bar{Z}_*$  is a constant which Buse (1973) has defined as  $(1'\Sigma^{-1}Z)/(1'\Sigma^{-1}1)$ . Both measures of linearity are location and scale invariant.

## 2.2 Test of fit based on EDF

Let  $x_1, \dots, x_n$  be the order statistics from a continuous distribution function  $F(y)$  of size  $n$ . The empirical distribution function for the sample, given by

$$F_n(y) = \frac{\text{number of observations } \leq y}{n}; \quad -\infty < y < \infty$$

is a step function with a step size  $1/n$  at the order statistics  $x_1 < \dots < x_n$ . The distance between the EDF and the hypothesized distribution,  $F(y) = \hat{F}(y)$  can be considered as a way of testing for  $H_0$ . Large values of the test statistics indicate that  $H_0$  should be rejected. In our case,  $\hat{F}(y)$  is given by

$$\hat{F}(y) = 1 - \exp(-e^{(y-\hat{\mu})/\hat{\theta}})$$

i.e. the estimated cumulative distribution function of the extreme value distribution, where the parameters are estimated by GLS according to equations (4) and (5).

In this study we discuss three such tests. The first of the EDF-based test we consider is the  $D_n$  statistic:

$$D_n = \sup_y |F_n(y) - \hat{F}(y)| = \max(D_n^+, D_n^-)$$

introduced by Kolmogorov-Smirnov, where  $D_n^+$  and  $D_n^-$  is obtained by

$$D_n^+ = \max_{1 \leq i \leq n} \left( \frac{i}{n} - \hat{F}(x_i) \right), \quad D_n^- = \max_{1 \leq i \leq n} \left( \hat{F}(x_i) - \frac{i-1}{n} \right)$$

The second EDF test statistic is the Cramer-von Mises statistic,  $W_n^2$ :

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(y) - \hat{F}(y)]^2 d\hat{F}(y)$$

which can also be expressed by the formula

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left( \widehat{F}(x_i) - \frac{i - 0.5}{n} \right)^2$$

The third EDF test is the Anderson-Darling statistic,  $A_n^2$ :

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(y) - \widehat{F}(y))^2}{\widehat{F}(y)(1 - \widehat{F}(y))} d\widehat{F}(y)$$

This test statistic can be calculated by the formula

$$A_n^2 = -n - \sum_{i=1}^n \frac{2i - 1}{n} [\log(\widehat{F}(x_i)) + \log(1 - \widehat{F}(x_{(n+1-i)}))]$$

### 2.3 Test of fit based on stabilized probability plot

Let  $x_1, \dots, x_n$  be the ordered observations in a random sample of size  $n$  from the distribution of the form  $F(x) = F_0((x - \mu)/\theta)$ , where  $\mu$  is a location parameter and  $\theta$  is a scale parameter. A probability plot is a plot of the  $x_i$  versus a corresponding theoretical quantities  $u_i = F_0^{-1}(d_i)$ , where  $d_i$  is an estimate of  $F_0((x_i - \mu)/\theta)$ . In such a plot the points should lie fairly near the line  $x_i = \mu + \theta u_i$ , and it indicates that the hypothesized distribution is a reasonable model for the data. The interpretation of the plot can be complicated due to the existence of outliers and the unequal variances of the plotted points. The stabilized probability plot is introduced by Michael (1983) to handle the problem. The plot is formed by plotting

$$s_i = (2/\pi) \sin^{-1} [F_0^{0.5}\{(x_i - \mu)/\theta\}] \quad (9)$$

against

$$r_i = (2/\pi) \sin^{-1} [\{(i - 0.5)/n\}^{0.5}] \quad (10)$$



where according to Michael (1983),  $s$  follows the sine distribution and all its order statistics have the same asymptotic variance. Hence by this transformation, the variance of the plotted points are approximately equal over the range of probability values.

A goodness-of-fit statistic based on stabilized probability plot is also suggested by Michael (1983) as

$$D_{sp} = \max |r_i - s_i| \quad (11)$$

Kimber (1985) used the statistic  $D_{sp}$  for testing of the extreme value distribution of maxima. He applied Downton's estimates (1966) of  $\mu$  and  $\theta$ , to obtain the critical values. Coles (1989) investigated the statistic  $D_{sp}$  for testing the extreme value distribution of minima and denoted it by  $D_{sp}^*$ . He estimated the parameters  $\alpha$  and  $\beta$  in equation (1), by using Blom's procedure (1958) and showed that due to the improved estimation procedure, the test statistic  $D_{sp}^*$  had higher power than Kimber's proposed test statistic.

The best linear unbiased estimates of the parameters whose distribution function is of a location - scale form have been considered by Lloyd (1952). Therefore we estimate the parameters in equation (1) by using equations (4) and (5) to get the test statistic

$$D_{sp}^+ = \max |r_i - \hat{s}_i| \quad (12)$$

where  $\hat{s}_i$  is obtained from equation (9) by replacing  $\mu$  and  $\theta$  by the estimates from (4) and (5).

A power comparison of the  $D_{sp}^*$ ,  $D_{sp}^+$  and some other test statistics are given in section 3.

### 3 THE RESULTS OF THE MONTE CARLO STUDY

#### 3.1 Power of the tests

The power of test statistics have been examined for samples from equation (1) against a range of alternative distributions based on 40000 replicates. As mentioned, the purpose of this study is to apply the expected values, variances and covariances of order statistics of the standard extreme value distribution in order to estimate the unknown parameters in equation (1) by GLS regression.

In Tables 1 and 2 the power of the test statistics by using approximations (equations (7) and (8)) are compared with the power when exact values are used for sample size 20. These tables show that the differences are negligible, most of the differences being in the second decimal place. The power also reveals that the test statistic  $T_1$  has higher power in comparison with the other test statistics for many of the alternatives.

In the case of the extreme value distribution, Stephens (1977) determined approximate critical values of the Cramer-von Mises statistic when using the maximum likelihood estimates of  $\mu$  and  $\theta$ . The  $W_n^2$  and  $A_n^2$  tests have also been studied by Littel *et al.* (1979), in which case again the maximum likelihood estimates were used in substitution for  $\mu$  and  $\theta$ .

Spinelli (1980), in an unpublished M.Sc. Thesis at Simon Fraser University, studied the test statistics based on regression and EDF for the extreme value distribution, of maxima. He pointed out problems arising when using the GLS method for the extreme value distribution due to the unavailability of the variances and covariances of the order statistics for large sample sizes.

We have also performed a power study for the extreme value distribution, of maxima. Comparison of the power of the tests for some different distributions with the results presented by Spinelli show that our results seems to be of the same order as Spinelli's results except for  $T_2$  statistic. Our result mostly give lower power for the  $T_2$  statistic than is indicated in Spinelli's study.

The test based on the  $T_1$  statistic is in terms of power, superior to most others test statistics in this study. The test based on  $D_{sp}^+$  is almost of the same power as the  $T_1$  test.

Table 1 also shows that due to the improved parameter estimation, the test statistic  $D_{sp}^+$  is generally more powerful than  $D_{sp}^*$ . The procedure for parameter estimation used in  $D_{sp}^+$  is simple to implement which is another advantage of this test statistic.

### 3.2 Percentage Points

The critical values of the  $T_1$  statistic for  $n = 10(10)100$  are given in Table 3. In this table for  $n = 10(10)30$ , the percentage points of the statistics based on both the exact and approximate values of means and covariances. Table 3 reveals that the approximation yield an error less than 6 % for all sample sizes. This encourages us to trust the use of approximations of means and variances in calculating the test statistics.

## REFERENCES

AITKEN, A.C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinb.*, **55**, 42.

- BLOM, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. Almqvist and Wiksell, Uppsala.
- BALAKRISHNAN, N. & CHAN, P.S. (1992). Order statistics from extreme value distribution I: Tables of means, variances and covariances, *Commun. Stat. Sim. Comp.* **21**, 1199-1217.
- BALAKRISHNAN, N. & Rao, C.R. (1998). *Handbook of Statistics 17*, Elsevier, Amsterdam.
- BUSE, A. (1973). Goodness of fit in Generalized Least Squares Estimation. *Amer. Statistician* **27**, 106-108.
- COLES, S.G (1989). On Goodness-of-Fit Tests for the Two-Parameter Weibull Distribution Derived from Stabilized Probability Plot. *Biometrika* **76**, 593-598.
- D'AGOSTINO, R.B. & STEPHENS, M.A. (1986). *Goodness-of-fit techniques*, Marcel Dekker, New York.
- DOWNTON, F. (1966). Linear Estimations of Parameters in the Extreme Value Distribution. *Technometrics* **8** 3-17.
- KIMBER, A.C. (1985). Tests for Exponential, Weibull and Gumbel Distributions based on the Stabilized Probability Plot. *Biometrika* **72**, 661-663.
- LIEBLEIN, J. & SALZAER, H.E. (1957). Table of the first moment of ranked extremes. *J. Res. Nat. Bur. Stand.* **59**, 203-206.
- LIEBLEIN, J. & ZELEN, M. (1956). Statistical investigation of the fatigue life of deep groove ball bearing. *J. Res. Nat. Bur. Stand.* **57**, 273-316.
- LITTEL, R.C., McCLAVE, J.T. & OFFEN, W.W. (1979). Goodness of fit tests for two-parameter Weibull distribution. *Commun. Stat.* **B8**, 257-269.
- LLOYD, E.H. (1952). Least-Squares Estimation of Location and Scale Parameters Using Order Statistics, *Biometrika* **39**, 88-95.
- MICHAEL, J.R. (1983). The stabilized probability plot. *Biometrika* **70**, 11-17.

- PIROUZI FARD, M.N. & HOLMQUIST, B. (2007). First moment approximations for order statistics from the extreme value distribution. *Statistical Methodology*, **4**, 196-203.
- PIROUZI FARD, M.N. & HOLMQUIST, B. (2006). Approximations of variances and covariances for order statistics from the standard extreme value distribution, *Statistical research report*, No. 2006:1, Lund University, Sweden.
- SPINELLI, J. (1980). Contributions to goodness-of-fit. *M.Sc. Thesis*, Department of Mathematics, Simon Fraser University.
- STEPHENS M.A. (1977). Goodness of Fit for the Extreme Value Distribution. *Biometrika*, **64** 583-588.
- WHITE, J.S. (1967). The moments of log-Weibull order statistics, *Research Publication GMR-717*, Research Laboratories, General Motors Corp., Warren, Mich.
- WHITE, J.S. (1969). Moments of Log-Weibull Order Statistics, *Technometrics* **11**, 373-386.

Table 1: Empirical power for tests on level 0.10, for selected alternative distributions based on 40000 replicates by using the exact values of means, variances and covariances of order statistics for  $n=20$ . <sup>1</sup>Density  $4(1-x)^3, 0 < x < 1$ . <sup>2</sup>Density  $\frac{1}{16}xe^{-x/4}, x \geq 0$ . <sup>3</sup>Density  $10x^4e^{-2x^5}, x \geq 0$ .

Distribution	$T_1$	$T_2$	$W_n^2$	$A_n^2$	$D_n$	$D_{sp}^+$	$D_{sp}^*$
$U(0, 1)$	0.46	0.22	0.34	0.38	0.27	0.47	0.40
Beta(1,4) <sup>1</sup>	0.98	0.41	0.88	0.92	0.81	0.98	0.96
$N(0, 1)$	0.43	0.13	0.32	0.33	0.26	0.38	0.33
Student $t(6)$	0.48	0.21	0.42	0.44	0.35	0.45	0.42
Gamma(2,4) <sup>2</sup>	0.97	0.46	0.87	0.90	0.78	0.96	0.93
$\chi^2(5)$	0.95	0.42	0.83	0.86	0.73	0.93	0.90
Weibull(2,5) <sup>3</sup>	0.26	0.10	0.19	0.20	0.17	0.24	0.20
Empirical Critical							
Values, level= 0.10	0.14	0.47	0.10	0.61	0.17	0.11	0.11

Table 2: Empirical power for tests on level 0.10, for selected alternative distributions based on 40000 replicates by using the estimated values of means, variances and covariances of order statistics for  $n=20$ .

Distribution	$T_1$	$T_2$	$W_n^2$	$A_n^2$	$D_n$	$D_{sp}^+$	$D_{sp}^*$
$U(0, 1)$	0.46	0.23	0.35	0.39	0.28	0.47	0.40
Beta(1,4)	0.98	0.42	0.88	0.92	0.81	0.98	0.97
$N(0, 1)$	0.44	0.14	0.32	0.33	0.26	0.38	0.33
Student $t(6)$	0.48	0.21	0.42	0.44	0.35	0.45	0.41
Gamma(2,4)	0.97	0.47	0.87	0.90	0.78	0.96	0.94
$\chi^2(5)$	0.95	0.42	0.83	0.87	0.73	0.94	0.90
Weibull(2,5)	0.27	0.10	0.19	0.20	0.17	0.24	0.20
Empirical Critical							
Values, level= 0.10	0.13	0.46	0.10	0.61	0.17	0.11	0.11

Table 3: *Quantiles of the test statistic  $T_1$  based on 40000 replicates.*

$n$	0.80	0.85	0.90	0.95	0.975	0.99	
10	0.15	0.17	0.21	0.27	0.33	0.41	exact
	0.15	0.17	0.20	0.26	0.32	0.41	approx.
20	0.10	0.11	0.14	0.17	0.22	0.28	exact
	0.10	0.11	0.14	0.18	0.22	0.28	approx.
30	0.07	0.09	0.10	0.13	0.17	0.21	exact
	0.07	0.09	0.10	0.13	0.17	0.21	approx.
40	0.06	0.07	0.09	0.11	0.14	0.18	
50	0.05	0.06	0.07	0.10	0.12	0.15	
60	0.05	0.05	0.06	0.08	0.10	0.13	
70	0.04	0.05	0.06	0.07	0.09	0.12	
80	0.04	0.04	0.05	0.07	0.08	0.11	
90	0.03	0.04	0.05	0.06	0.08	0.10	
100	0.03	0.04	0.04	0.06	0.07	0.09	