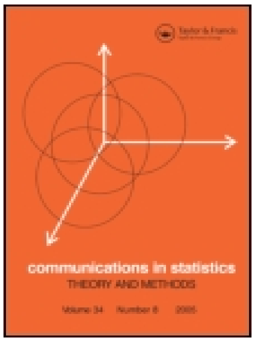


Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: www.tandfonline.com/journals/lsta20

# NONNORMAL REGRESSION. I. SKEW DISTRIBUTIONS

M. Qamarul Islam, Distinguished Visiting Professor Moti L. Tiku & F. Yildirim

COMMUN. STATIST.—THEORY METH., 30(6), 993–1020 (2001)

# NONNORMAL REGRESSION. I. SKEW DISTRIBUTIONS

**M. Qamarul Islam,[1] Moti L. Tiku,[2,*] and F. Yildirim[3]**

[1]Department of Economics, Cankaya University, Ankara 06530, Turkey
[2]Department of Mathematics and Statistics, McMaster University, Hamilton, ON L8S 4K1, Canada
[3]Department of Statistics, Middle East Technical University, Ankara 06531, Turkey

## ABSTRACT

In a linear regression model of the type $y = \theta X + e$, it is often assumed that the random error $e$ is normally distributed. In numerous situations, e.g., when $y$ measures life times or reaction times, $e$ typically has a skew distribution. We consider two important families of skew distributions, (a) Weibull with support IR: $(0, \infty)$ on the real line, and (b) generalised logistic with support IR: $(-\infty, \infty)$. Since the maximum likelihood estimators are intractable in these situations, we derive modified likelihood estimators which have explicit algebraic forms and are, therefore, easy to compute. We show that these estimators are remarkably efficient, and robust. We develop hypothesis testing procedures and give a real life example.

* Corresponding author is a Distinguished Visiting Professor at Middle East Technical University, Ankara. E-mail: tikumoti@metu.edu.tr

Symmetric families of distributions, both long and short tailed, will be considered in a future paper.

*Key Words*: Robustness; Maximum likelihood; Modified maximum likelihood; Least squares; Weibull; Generalised logistic

## 1. INTRODUCTION

In a linear regression model of the type

$$y = \theta X + e \tag{1.1}$$

it is usual practice to assume that the errors $e_i$, $1 \le i \le n$, are iid normal $N(0, \sigma^2)$. In practice, however, $e_i$ are often nonnormal. In this paper we assume that $e_i$ have a skew distribution. For illustration we consider two important skew distributions, (a) the Weibull with support IR: $(0, \infty)$, and (b) generalised logistic with support IR: $(-\infty, \infty)$. The likelihood equations are, however, intractable and solving them by iteration can be problematic (Barnett [1], Lee et al. [2], Tiku et al. [3], Vaughan [4]). If the data contains outliers, iterations with likelihood equations are often nonconverging; see, for example, Puthenpura and Sinha [5]. These difficulties are indeed debilitating. To alleviate these difficulties, we utilise the method of modified likelihood (Tiku [6], [7], [8]; Tiku and Suresh [9]). Tan [10], and Tan and Balakrishnan [11], give a Bayesian insight of this method. The method first expresses the likelihood equations in terms of order statistics and then linearizes the intractable terms. For estimating the location (mean) and the scale (standard deviation) parameters of location-scale distributions, the modified likelihood equations have explicit solutions called MML (modified maximum likelihood) estimators. These estimators are known to be asymptotically fully efficient under regularity conditions (Bhattacharyya [12], Vaughan and Tiku [13]) and almost as efficient as the ML (maximum likelihood) estimators for small sample sizes (Smith et al. [14], Lee et al. [2], Tan [10]); see also Tiku et al. [3], Tiku and Suresh [9], Vaughan [4], and Bian and Tiku [15], [16]. We extend this method to linear models (1.1) and show that the MML estimators of $\theta$ and $\sigma$, $\sigma^2 = V(e)/\tau^2$, are explicit and remarkably efficient, and robust. See also Tiku and Selçuk [17] and Tiku et al. [18], [19] who discuss applications of this method to time series data.

## 2. WEIBULL DISTRIBUTION

Consider in first place the linear model

$$y_i = \theta_0 + \theta_1 x_i + e_i, \quad 1 \le i \le n, \tag{2.1}$$

where $e_i$ are iid and have the Weibull distribution

$$W(p, \sigma) : (p/\sigma^p) e^{p-1} \exp\{-(e/\sigma)^p\}, \quad 0 < e < \infty. \tag{2.2}$$

Since $\theta_0 + \theta_1 x + E(e)$ will often be used as a predictor of the expected response $E(y)$, a model which can result in values of the probability

$$prob\{y \ge \theta_0 + \theta_1 x + E(e)\} = \exp\{-[\Gamma(1 + 1/p)]^p\}$$

substantially smaller or larger than 0.5 (say, $\le$ 0.4 or $\ge$ 0.6) is hardly of any practical interest. If $e$ has the Weibull distribution $W(p, \sigma)$, then the values of this probability are

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $p$ = | 0.5 | 1.0 | 1.1 | 1.2 | 1.3 | 1.5 | 2.0 | 3.0 | 6.0 |
| $prob$ = | 0.24 | 0.37 | 0.38 | 0.39 | 0.41 | 0.42 | 0.46 | 0.49 | 0.53 |

In this paper, therefore, we are primarily interested in values of $p \ge 1.3$. See also Cohen and Whitten [20] who argue that in most applications $p$ is greater than 1. Writing $z_i = e_i/\sigma = (y_i - \theta_0 - \theta_1 x_i)/\sigma$, $1 \le i \le n$, the likelihood equations $\partial \ln L/\partial\theta_0 = 0$, $\partial \ln L/\partial\theta_1 = 0$ and $\partial \ln L/\partial\sigma = 0$ are nonlinear functions and are expressions in terms of $z_i^{-1}$ and $z_i^{p-1}$. They have no explicit solutions and solving them by iteration is indeed problematic; see also Smith [21], and Yildirim and Korasli [22]. To derive modified likelihood equations which have explicit solutions, and are under regularity conditions asymptotically equivalent to the likelihood equations, we first order $w_i = y_i - \theta_1 x_i$ (for a given $\theta_1$) so that

$$w_{(1)} \le w_{(2)} \le \cdots \le w_{(n)}; \quad w_{(i)} = y_{[i]} - \theta_1 x_{[i]}. \tag{2.3}$$

We define the ordered variates $z_{(i)} = \{w_{(i)} - \theta_0\}/\sigma$, $1 \le i \le n$; $(y_{[i]}, x_{[i]})$ may be called concomitants of $z_{(i)}$ and is that pair of $(y_j, x_j)$ values which determines $w_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations can be written in terms of $z_{(i)}$:

$$\frac{\partial \ln L}{\partial\theta_0} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)}^{p-1} = 0 \tag{2.4}$$

$$\frac{\partial \ln L}{\partial\theta_1} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} x_{[i]} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} x_{[i]} z_{(i)}^{p-1} = 0 \tag{2.5}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{p-1} = 0. \qquad (2.6)$$

Realise the difficulties which can arise if $z_{(1)}$ tends to zero in which case (2.4)–(2.5) are not defined.

## 3. MODIFIED LIKELIHOOD

Write $t_{(i)} = E\{z_{(i)}\}$, $1 \le i \le n$, and note that

$$E\{z_{(i)}\} = \frac{n!}{(i-1)!(n-i)!}\sum_{j=0}^{i-1} \frac{(-1)^j \binom{i-1}{j} \Gamma(1+1/p)}{(n-i+j+1)^{(p+1)/p}}. \qquad (3.1)$$

The computation of (3.1) is rather cumbersome for large $n$ (say, $n \ge 10$). For $n \ge 10$, however, the approximate values of $t_{(i)}$ are used and obtained from the equations

$$\int_0^{t_{(i)}} p z^{p-1} \exp(-z^p)\, dz = \frac{i}{n+1}$$

which gives

$$t_{(i)} = [-\ln\{1 - i/(n+1)\}]^{1/p}, \quad 1 \le i \le n. \qquad (3.2)$$

Since $z_{(i)}^{p-1}$ is almost linear in small intervals around $z_{(i)}$, we linearize $z_{(i)}^{p-1}$ by using the first two terms of a Taylor series expansion (Tiku [6], [7]; Tiku and Suresh [9]):

$$z_{(i)}^{p-1} \cong \alpha_i + \beta_i z_{(i)}; \quad \alpha_i = (2-p)t_{(i)}^{p-1} \quad \text{and} \quad \beta_i = (p-1)t_{(i)}^{p-2} \quad (1 \le i \le n). \qquad (3.3)$$

Similarly,

$$z_{(i)}^{-1} \cong \alpha_{i0} - \beta_{i0} z_{(i)}; \quad \alpha_{i0} = 2t_{(i)}^{-1} \quad \text{and} \quad \beta_{i0} = t_{(i)}^{-2} \quad (1 \le i \le n). \qquad (3.4)$$

Incorporating (3.3)–(3.4) in (2.4)–(2.6), we get the modified likelihood equations $\partial \ln L^*/\partial\theta_0 = 0$, $\partial \ln L^*/\partial\theta_1 = 0$ and $\partial \ln L^*/\partial\sigma = 0$. The solutions of these equations are the MML estimators:

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]} - (\Delta/m)\hat{\sigma} \qquad (3.5)$$

$$\hat{\theta}_1 = K - D\hat{\sigma} \tag{3.6}$$

and

$$\hat{\sigma} = \left\{-B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{\{n(n-2)\}} \tag{3.7}$$

where

$$\begin{aligned}
&\delta_i = (p-1)\beta_{i0} + p\beta_i, \quad \Delta_i = (p-1)\alpha_{i0} - p\alpha_i;\\
&m = \sum_{i=1}^{n} \delta_i, \quad \Delta = \sum_{i=1}^{n} \Delta_i;\\
&\bar{y}_{[.]} = (1/m)\sum_{i=1}^{n} \delta_i y_{[i]}, \quad \bar{x}_{[.]} = (1/m)\sum_{i=1}^{n} \delta_i x_{[i]};\\
&K = \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} \Big/ \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2,\\
&D = \sum_{i=1}^{n} \Delta_i (x_{[i]} - \bar{x}_{[.]}) \Big/ \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2;\\
&B = \sum_{i=1}^{n} \Delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\},\\
&C = \sum_{i=1}^{n} \delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}^2\\
&\quad = \sum_{i=1}^{n} \delta_i (y_{[i]} - \bar{y}_{[.]})^2 - K\sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]}.
\end{aligned} \tag{3.8}$$

Note that $\delta_i > 0$ for all $p > 1$. The ML estimator of $\sigma$ can cease to be real or positive (see, for example, Lawless [23] (Chapter 6)), but the MML estimator $\hat{\sigma}$ is always real and positive.

**Remark:** It is not difficult to prove that (see, for example, Vaughan and Tiku [13]) asymptotically ($p > 2$)

$$\begin{aligned}
&\frac{1}{n}\left\{\frac{\partial \ln L}{\partial \theta_0} - \frac{\partial \ln L^*}{\partial \theta_0}\right\} = 0, \quad \frac{1}{n}\left\{\frac{\partial \ln L}{\partial \theta_1} - \frac{\partial \ln L^*}{\partial \theta_1}\right\} = 0,\\
&\text{and} \quad \frac{1}{n}\left\{\frac{\partial \ln L}{\partial \sigma} - \frac{\partial \ln L^*}{\partial \sigma}\right\} = 0;
\end{aligned} \tag{3.9}$$

see also the Appendix. Thus, the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are asymptotically equivalent to the ML (maximum likelihood) estimators for $p > 2$.

Note that $p > 2$ is a necessary regularity condition for the Fisher information matrix to exist.

**Computations:** The computations are carried out in two iterations. In the first iteration, $w_{(i)}$ are obtained by ordering $w_i = y_i - \tilde{\theta}_1 x_i$ $(1 \le i \le n)$ in ascending order, where $\tilde{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i / \sum_{i=1}^{n}(x_i - \bar{x})^2$ is the LS (least squares) estimator of $\theta_1$. Then, $\hat{\theta}_1$ is calculated from (3.6)–(3.7). In the second iteration, $w_{(i)}$ are obtained by ordering $w_i = y_i - \hat{\theta}_1 x_i$, $1 \le i \le n$. The resulting concomitants are used to compute the MML estimators from (3.5)–(3.7). In all our computations, only two iterations were needed for the estimates to stabilise sufficiently enough. The reason perhaps is that the MML estimators only depend on the concomitants $(y_{[i]}, x_{[i]})$ and the concomitant indices are determined by the relative magnitudes, not necessarily the true values, of $w_i$ $(1 \le i \le n)$.

## 4. ASYMPTOTIC EFFICIENCY

Again, it is not difficult to prove that (Vaughan and Tiku [13] (Appendix A)) asymptotically ($p > 2$)

$$\begin{aligned} &E\left[\frac{1}{n}\sum_{i=1}^{n}\left\{z_{(i)}^{p-1} - (\alpha_i + \beta_i z_{(i)})\right\}\eta_i\right] = 0 \quad \text{and} \\ &E\left[\frac{1}{n}\sum_{i=1}^{n}\left\{z_{(i)}^{-1} - (\alpha_{i0} + \beta_{i0} z_{(i)})\right\}\eta_i\right] = 0 \end{aligned} \tag{4.1}$$

for $\eta_i = 1$, $\eta_i = x_{[i]}$ or $\eta_i = z_{(i)}$. As a consequence of this, the expected values of all the first order and the second order derivatives of $\ln L^*$ are exactly the same (asymptotically) as the corresponding values for $\ln L$. As a consequence of this, the following result is true; see also the Appendix. See also Vaughan and Tiku [13] (Appendix A).

**Theorem 1:** For $p > 2$, the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are asymptotically fully efficient, i.e., they are asymptotically unbiased and their covariance matrix is $I^{-1}(\theta_0, \theta_1, \sigma)$, where

$$I = \frac{np^2}{\sigma^2}\begin{bmatrix} \left(1-\frac{1}{p}\right)^2\Gamma\left(1-\frac{2}{p}\right) & \left(1-\frac{1}{p}\right)^2\Gamma\left(1-\frac{2}{p}\right)\sum x_i/n & \Gamma\left(2-\frac{1}{p}\right) \\ & \left(1-\frac{1}{p}\right)^2\Gamma\left(1-\frac{2}{p}\right)\sum x_i^2/n & \Gamma\left(2-\frac{1}{p}\right)\sum x_i/n \\ & & 1 \end{bmatrix} \tag{4.2}$$

is the Fisher information matrix. Note that if $\sum_{i=1}^{n} x_i = 0$ (e.g., a symmetric design in the interval –1 to 1), $\hat{\theta}_1$ is uncorrelated with $\hat{\theta}_0$ and $\hat{\sigma}$.

**Comment:** The values calculated from (4.2) provide close approximations to the true values even for moderate sample sizes and may be used for all $n > 50$ ($p \geq 2.5$), at any rate for $\hat{\theta}_1$ and $\hat{\sigma}$. Given in Table 1 are the values calculated from the diagonal elements of $I^{-1}$, namely,

$$\begin{aligned}
V(\hat{\theta}_0) &= \frac{\sigma^2}{n(p-1)^2}\left\{\frac{1}{\Gamma(1-2/p)-\Gamma^2(1-1/p)} + \frac{n\bar{x}^2}{\Gamma(1-2/p)\sum(x_i-\bar{x})^2}\right\} \\
V(\hat{\theta}_1) &= \frac{\sigma^2}{(p-1)^2\Gamma(1-2/p)\sum(x_i-\bar{x})^2}, \\
V(\hat{\sigma}) &= \frac{\Gamma(1-2/p)\sigma^2}{np^2\{\Gamma(1-2/p)-\Gamma^2(1-1/p)\}},
\end{aligned} \tag{4.3}$$

and give the asymptotic variances. Also given are the corresponding simulated values based on $N = [10{,}0000/n]$ Monte Carlo runs. A set of design points $x_i$, $1 \leq i \leq n$, was randomly generated from a *Uniform* $(0, 1)$ distribution and was common to all the $N$ number of random samples $(y_1, y_2, \ldots, y_n)$ generated from the Weibull $W(p, \sigma)$. Without loss of generality $\theta_0$, $\theta_1$ and $\sigma$ are in the rest of the paper taken to be equal to 0, 1 and 1, respectively. For brevity, the values of the covariances are not reproduced but there is close agreement between their simulated values and the values calculated from $I^{-1}$.

The small difference between the two sets of values in Table 1 clearly indicate that the MML estimators are remarkably efficient for large $n$. Since MVB (minimum variance bound) estimators do not exist, all estimators will have their variances greater than the MVB; see also Smith et al. [14], Tan [10] and Vaughan [4].

***Table 1.*** Values of $n\times$(Variance) of the MML Estimators

| | | $\hat{\theta}_0$ | | $\hat{\theta}_1$ | | $\hat{\sigma}$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $p$ | Asymp | Simul | Asymp | Simul | Asymp | Simul |
| 50 | 2.5 | 0.565 | 0.832 | 1.201 | 1.607 | 0.310 | 0.406 |
| | 3.0 | 0.659 | 0.786 | 1.158 | 1.267 | 0.352 | 0.431 |
| | 4.0 | 0.655 | 0.701 | 0.778 | 0.780 | 0.409 | 0.435 |
| 100 | 2.5 | 0.494 | 0.651 | 1.179 | 1.425 | 0.310 | 0.394 |
| | 3.0 | 0.591 | 0.673 | 1.137 | 1.233 | 0.352 | 0.402 |
| | 4.0 | 0.609 | 0.657 | 0.764 | 0.787 | 0.409 | 0.434 |

## 5. RELATIVE EFFICIENCY

We have already established the fact that for $p > 2$ the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are asymptotically fully efficient. To have an idea about their efficiencies relative to some of the commonly used estimators, e.g., the LS (least squares) estimators which when corrected for bias are (see also Cohen and Whitten [20])

$$\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1 \bar{x} - \Gamma(1 + 1/p)\tilde{\sigma}, \quad \tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \bigg/ \sum_{i=1}^{n} (x_i - \bar{x})^2$$

and

$$\tilde{\sigma} = \left\{ \sum_{i=1}^{n} [y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x})]^2 / (n-2)[\Gamma(1 + 2/p) - \Gamma^2(1 + 1/p)] \right\}^{1/2}; \tag{5.1}$$

$\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$. It is easy to show that $E(\tilde{\theta}_1) = \theta_1$ and

$$V(\tilde{\theta}_1) = \left\{ \Gamma(1 + 2/p) - \Gamma^2(1 + 1/p) \right\} \sigma^2 \bigg/ \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{5.2}$$

The values of the asymptotic relative efficiency of $\tilde{\theta}_1$, i.e., $100 V(\hat{\theta}_1)/V(\tilde{\theta}_1)$, are 67, 88, 91 and 81 percent for $p = 2.5$, 3.0, 6.0 and 10.0, respectively. Clearly, the LS estimator $\tilde{\theta}_1$ is considerably less efficient than the MML estimator $\hat{\theta}_1$.

It is very difficult to find the expected values and variances of $\tilde{\theta}_0$ and $\tilde{\sigma}$ (and, of course, the covariances between $\tilde{\theta}_0$, $\tilde{\theta}_1$ and $\tilde{\sigma}$) even asymptotically. To compare the efficiencies of the LS and the MML estimators, therefore, we simulated their means and variances. The bias in all these estimators were found to be negligible, although the bias in $\hat{\theta}_0$ for small $n$ is a little larger than that of $\tilde{\theta}_0$. For $n = 20$ and 100, for example, we have the following values of the means:

Simulated values of the Means; $\theta_0 = 0$, $\theta_1 = 1$, $\sigma = 1$.

| | $p = 1.5$ | | | | | | $p = 3.0$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tilde{\theta}_0$ | $\hat{\theta}_0$ | $\tilde{\theta}_1$ | $\hat{\theta}_1$ | $\tilde{\sigma}$ | $\hat{\sigma}$ | $\tilde{\theta}_0$ | $\hat{\theta}_0$ | $\tilde{\theta}_1$ | $\hat{\theta}_1$ | $\tilde{\sigma}$ | $\hat{\sigma}$ |
| 20 | 0.02 | 0.15 | 1.00 | 0.99 | 0.98 | 0.94 | 0.01 | 0.04 | 1.00 | 1.00 | 0.98 | 0.96 |
| 100 | −0.00 | 0.04 | 1.01 | 1.00 | 0.99 | 0.98 | 0.00 | 0.02 | 1.00 | 1.00 | 1.00 | 0.99 |

The MML estimators are, however, considerably more efficient. Given in Table 2 are the simulated values of the variances of the MML estimators and the relative efficiencies

$$E_1 = 100\{V(\hat{\theta}_0)/V(\tilde{\theta}_0)\}, \quad E_2 = 100\{V(\hat{\theta}_1)/V(\tilde{\theta}_1)\}, \quad E_3 = 100\{V(\hat{\sigma})/V(\tilde{\sigma})\} \tag{5.3}$$

of the LS estimators. The LS estimators have a disconcerting feature, namely, their relative efficiencies decrease as $n$ increases and stabilise at values considerably less than 100%, especially for smaller values of $p$.

**Remark:** We also calculated means and variances of the estimators for other designs, e.g., $x_i$ $(1 \le i \le n)$ generated from normal $N(0, 1)$. The biases were

***Table 2.*** Variances of the MML Estimators and the Relative Efficiencies of the LS Estimators: $(1) = nV(\hat{\theta}_0)$, $(2) = nV(\hat{\theta}_1)$, $(3) = nV(\hat{\sigma})$

| $n$ | (1) | $E_1$ | (2) | $E_2$ | (3) | $E_3$ |
|---|---|---|---|---|---|---|
| | | | $p = 1.3$ | | | |
| 10 | 1.54 | 97 | 5.55 | 109 | 0.637 | 61 |
| 20 | 1.83 | 79 | 5.11 | 94 | 0.602 | 58 |
| 30 | 1.86 | 77 | 5.19 | 94 | 0.581 | 58 |
| 50 | 1.83 | 70 | 5.40 | 81 | 0.615 | 57 |
| 100 | 1.52 | 70 | 4.68 | 75 | 0.574 | 56 |
| | | | $p = 1.5$ | | | |
| 10 | 0.801 | 61 | 2.63 | 69 | 0.597 | 70 |
| 20 | 0.907 | 52 | 2.26 | 56 | 0.537 | 64 |
| 30 | 0.876 | 45 | 1.96 | 47 | 0.541 | 65 |
| 50 | 0.776 | 40 | 1.92 | 40 | 0.510 | 59 |
| 100 | 0.473 | 30 | 1.47 | 34 | 0.495 | 61 |
| | | | $p = 2.0$ | | | |
| 10 | 0.706 | 76 | 21.774 | 86 | 0.533 | 79 |
| 20 | 0.856 | 75 | 1.800 | 79 | 0.454 | 78 |
| 30 | 0.851 | 68 | 1.646 | 72 | 0.451 | 77 |
| 50 | 0.829 | 64 | 1.818 | 67 | 0.420 | 73 |
| 100 | 0.569 | 55 | 1.608 | 60 | 0.389 | 67 |
| | | | $p = 6.0$ | | | |
| 10 | 0.651 | 95 | 0.295 | 94 | 0.564 | 92 |
| 20 | 0.646 | 96 | 0.324 | 94 | 0.517 | 94 |
| 30 | 0.642 | 96 | 0.321 | 93 | 0.497 | 94 |
| 50 | 0.652 | 96 | 0.381 | 93 | 0.479 | 93 |
| 100 | 0.624 | 93 | 0.376 | 91 | 0.470 | 92 |

negligible, and the relative efficiencies $E_1$, $E_2$ and $E_3$ of the LS estimators turned out to be essentially the same as in Table 2.

## 6. ROBUSTNESS

In practice the shape parameter $p$ in $W(p, \sigma)$ (eq. 2.2) might be somewhat misspecified or the sample might contain outliers. From a practical point of view, therefore, it is very important for an estimator to have efficiency robustness; see, for example, Huber [24], Tiku et al. [3], and Tan and Tiku [25]. Such an estimator is fully efficient (or nearly so) for an assumed model but maintains high efficiency for plausible alternatives to the assumed model. We assume the model to be $W(p, \sigma)$ with $p=2$, the scale $\sigma$ being unknown. The value $p=2$ is chosen for illustration but, of course, any other value of $p$ can be chosen with similar results. The alternatives to this model are called sample models. Out of a large number of plausible sample models, we choose a representative few as follows. These models represent different types of distributions which, like the assumed Weibull $W(2, \sigma)$, have the longer tail on the right hand side.

(a) The Weibull $W(p, \sigma)$: (1) $p = 1.3$, (2) $p = 2.0$, (3) Exponential $(1/\sigma)\exp(-e/\sigma)$, $0 < e < \infty$, i.e., $W(1, \sigma)$.
(b) Dixon's single outlier model: (4) $e_1, e_2, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n$ are $(n-1)$ random deviates from $W(2, \sigma)$ but $e_i$ (for some $i$) is from $W(2, \lambda\sigma)$, $\lambda = 4$.
(c) Tiku's single outlier model (Hawkins [26], Tiku [27]): (5) $e_{(1)}, e_{(2)}, \ldots, e_{(n-1)}$ are the first $(n-1)$ order statistics of a random sample of size $n$ from $W(2, \sigma)$ and $e_{(n)}$ is the largest order statistic of this sample plus $\lambda\sigma$, $\lambda = 4$.
(d) Contamination model: (6) $0.90W(2, \sigma) + 0.10W(1.3, \sigma)$.

Both outlier models (b) and (c) extend to more than one outlier (Tiku [28], Mann [29]) which are not considered here for brevity. The models (b)–(d) are very important from a practical point of view. In fact, Huber [24] stated that the occurrence of five to ten percent outliers in a sample is a rule not an exception.

Realise that the sample model (2) is also the assumed population model. Therefore, the $t_{(i)}$ values (and $\delta_i$ and $\Delta_i$ coefficients) are calculated from (3.1) or (3.2) with $p=2$ and used for all the sample models (a)–(d). Note, however, that $\tilde{\theta}_0$ and $\hat{\theta}_0$ are not location and scale invariant and it is difficult to figure out the parameter they are estimating under a sample model other than model (2) in which case they both estimate $\theta_0$. For that reason we are not reporting their means and variances in Table 3, although

***Table 3.*** Simulated Means and Variances of the LS and the MML Estimators of $\theta_1$ and $\sigma$; $\theta_1 = 1$, $\sigma = 1$

| | Mean | | | | $n$ (Variance) | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | $\tilde{\theta}_1$ | $\hat{\theta}_1$ | $\tilde{\sigma}$ | $\hat{\sigma}$ | $\tilde{\theta}_1$ | $\hat{\theta}_1$ | $\tilde{\sigma}$ | $\hat{\sigma}$ |
| | | | | Model (1), $\tau = 1.547$ | | | | |
| 20 | 1.00 | 0.99 | 1.50 | 1.35 | 5.49 | 3.04 | 2.40 | 1.56 |
| 30 | 0.99 | 0.99 | 1.52 | 1.36 | 5.60 | 2.81 | 2.50 | 1.58 |
| 50 | 1.00 | 1.00 | 1.52 | 1.35 | 6.60 | 3.09 | 2.47 | 1.48 |
| 100 | 0.99 | 1.00 | 1.54 | 1.36 | 6.17 | 2.91 | 2.50 | 1.44 |
| | | | | Model (2), $\tau = 1$ | | | | |
| 20 | 1.00 | 0.99 | 0.98 | 0.94 | 2.28 | 1.80 | 0.58 | 0.45 |
| 30 | 1.00 | 0.99 | 0.99 | 0.96 | 2.28 | 1.65 | 0.58 | 0.45 |
| 50 | 1.00 | 1.00 | 0.99 | 0.97 | 2.71 | 1.82 | 0.57 | 0.42 |
| 100 | 1.00 | 1.00 | 1.00 | 0.98 | 2.70 | 1.61 | 0.58 | 0.49 |
| | | | | Model (3), $\tau = 2.159$ | | | | |
| 20 | 1.00 | 0.99 | 2.06 | 1.78 | 10.7 | 5.00 | 7.39 | 4.55 |
| 30 | 0.99 | 0.99 | 2.09 | 1.79 | 10.9 | 4.82 | 7.96 | 4.74 |
| 50 | 1.00 | 1.00 | 2.10 | 1.77 | 12.8 | 6.07 | 8.04 | 4.54 |
| 100 | 0.98 | 0.99 | 2.14 | 1.78 | 12.1 | 8.12 | 8.33 | 4.50 |
| | | | | Model (4), $\tau = *$ | | | | |
| 20 | 0.20 | 0.54 | 1.62 | 1.42 | 8.08 | 3.70 | 7.80 | 4.30 |
| 30 | 0.42 | 0.69 | 1.48 | 1.30 | 6.58 | 3.08 | 7.21 | 3.52 |
| 50 | 0.64 | 0.82 | 1.33 | 1.18 | 6.12 | 2.90 | 6.70 | 2.82 |
| 100 | 0.83 | 0.92 | 1.20 | 1.09 | 3.79 | 1.92 | 5.13 | 1.78 |
| | | | | Model (5), $\tau = **$ | | | | |
| 20 | 0.99 | 1.00 | 2.57 | 2.03 | 15.1 | 5.78 | 0.63 | 0.37 |
| 30 | 0.99 | 0.99 | 2.20 | 1.74 | 11.4 | 4.49 | 0.56 | 0.36 |
| 50 | 1.01 | 1.01 | 1.84 | 1.47 | 9.08 | 3.80 | 0.52 | 0.35 |
| 100 | 0.99 | 1.00 | 1.50 | 1.25 | 5.77 | 2.32 | 0.54 | 0.36 |
| | | | | Model (6), $\tau = 1.068$ | | | | |
| 20 | 1.00 | 0.99 | 1.32 | 1.18 | 4.34 | 2.45 | 2.88 | 1.72 |
| 30 | 0.99 | 0.99 | 1.34 | 1.19 | 4.45 | 2.34 | 3.02 | 1.74 |
| 50 | 1.00 | 1.00 | 1.35 | 1.20 | 5.26 | 2.56 | 2.94 | 1.60 |
| 100 | 0.99 | 1.00 | 1.37 | 1.21 | 4.90 | 2.18 | 3.04 | 1.56 |

*The values of $\tau$ are 1.32, 1.22, 1.14, and 1.07 for $n = 20$, 30, 50, and 100, respectively.
**The values of $\tau$ are 1.58, 1.36, 1.15, and 0.94 for $n = 20$, 30, 50, and 100, respectively.

the variances of $\hat{\theta}_0$ are considerably smaller than those of $\tilde{\theta}_0$. The estimators $\tilde{\sigma}$ and $\hat{\sigma}$ are both estimating $\tau\sigma(\tau > 0)$, where $\tau$ is the ratio of the standard deviation of the sample model to the standard deviation of the population model. The values of $\tau$ are given in Table 3. Note that $\tau$ has absolutely no role to play in the computation of the LS or the MML estimators. Its values are given only for mean square error calculations. The estimators $\tilde{\theta}_1$ and $\hat{\theta}_1$ are location invariant and both are estimating $\theta_1$ under all the models above. The simulated values of the mean and variance are given in Table 3.

Under Dixon's outlier model both $\tilde{\theta}_1$ and $\hat{\theta}_1$ develop bias for small $n$, but $\hat{\theta}_1$ has considerably smaller bias. Under Tiku's outlier model, however, they have hardly any bias and that is a very interesting finding for one of the authors of this paper. For all the models above, the MML estimators $\hat{\theta}_1$ and $\hat{\sigma}$ not only have smaller bias than the LS estimators $\tilde{\theta}_1$ and $\tilde{\sigma}$, but they also have much smaller variances (and mean square errors). Besides, $\hat{\theta}_1$ and $\hat{\sigma}$ are remarkably efficient (Table 1) for the assumed population model, i.e., the sample model (2) above. Thus, the MML estimators $\hat{\theta}_1$ and $\hat{\sigma}$ are robust. The reason for their robustness is that $D$ and $B/\sqrt{nC}$ in equations (3.8) are small and, as a consequence, the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$, and $\hat{\sigma}$ are essentially the solutions of the equations

$$\frac{\partial \ln L^*}{\partial \theta_0} \cong -\frac{1}{\sigma}\sum_{i=1}^{n} \delta_i z_{(i)} = 0, \quad \frac{\partial \ln L^*}{\partial \theta_1} \cong -\frac{1}{\sigma}\sum_{i=1}^{n} x_{[i]-1}(\delta_i z_{(i)}) = 0, \quad \text{and} \tag{6.1}$$

$$\frac{\partial \ln L^*}{\partial \sigma} \cong -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} \delta_i z_{(i)}^2 = 0. \tag{6.2}$$

It is clear from these equations that the ordered residuals $e_{(i)} = \sigma z_{(i)}$ (and their squares $e_{(i)}^2$) are assigned the weight $\delta_i$. But for the Weibull $W(2, \sigma)$, $\delta_i$ have half-umbrella ordering, that is, $\delta_i$ $(1 \le i \le n)$ is a decreasing sequence of positive numbers. For $n = 20$ $(p = 2)$, for example,

$\delta_i$ = 22.50, 11.99, 8.49, 6.73, 5.68, 4.97, 4.47, 4.08, 3.79, 3.55, 3.35, 3.18, 3.04, 2.91, 2.80, 2.70, 2.60, 2.51, 2.42, 2.33.

Consequently, the extreme ordered residuals $e_{(i)}$ on the right hand side (and their squares $e_{(i)}^2$) are assigned small weights. Thus, the influence of long tails or outliers (on the right hand side) is automatically depleted. This gives the MML estimators the feature of robustness to plausible deviations from the assumed model. On the other hand, the LS estimators are the solutions of (6.1)–(6.2) with $\delta_i$ $(1 \leq i \leq n)$ equal to 1. Thus, all ordered residuals $e_{(i)}$ (and their squares) are assigned the same weight. This exposes the LS estimators to the dominant influence of long tails or outliers resulting in their considerably larger variances (mean square errors), as illustrated in Table 3.

## 7. HYPOTHESIS TESTING

Testing the null hypothesis $H_0: \theta_1 = 0$ is of great practical interest. In that regard, we have the following result.

**Lemma 1:** Conditionally ($\sigma$ known), $\hat{\theta}_1(\sigma)$ is asymptotically the MVB estimator and is normally distributed with mean $\theta_1$ and variance $\sigma^2 / \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2$, $p > 2$.

**Proof:** The result follows from the fact that $\partial \ln L^* / \partial \theta_1 = 0$ can, in view of $\partial \ln L^* / \partial \theta_0 = 0$, be put in the form

$$\frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2}{\sigma^2} \left\{ (K - D\sigma) - \theta_1 \right\} = 0. \tag{7.1}$$

The result then follows from the fact that the modified likelihood equation $\partial \ln L^* / \partial \theta_1 = 0$ is asymptotically equivalent to the likelihood equation $\partial \ln L / \partial \theta_1 = 0$ $(p > 2)$, and the third and higher derivatives of $\partial \ln L^* / \partial \theta_1$ are zero (Bartlett [30]); see also Kendall and Stuart [31] (Chapter 18).

**Lemma 2:** Conditionally ($\theta_1$ known), $\hat{\sigma}(\theta_1)$ is asymptotically the MVB estimator of $\sigma$ and $(n-1)\hat{\sigma}^2(\theta_1)$ is distributed as a multiple of chi-square.

**Proof:** The modified likelihood equation $\partial \ln L^* / \partial \sigma = 0$ can, in view of $\partial \ln L^* / \partial \theta_0 = 0$, be put in the form

$$\begin{aligned} \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma^3} &\left( \left\{ -B_0 + \sqrt{(B_0^2 + 4nC_0} \right\} \Big/ 2n - \sigma \right) \\ &\times \left( \left\{ -B_0 - \sqrt{(B_0^2 + 4nC_0} \right\} \Big/ 2n - \sigma \right) \end{aligned} \tag{7.2}$$

where

$$B_0 = \sum_{i=1}^{n} \Delta_i \{y_{[i]} - \bar{y}_{[.]} - \theta_1 (x_{[i]} - \bar{x}_{[.]})\} \quad \text{and}$$

$$C_0 = \sum_{i=1}^{n} \delta_i \{y_{[i]} - \bar{y}_{[.]} - \theta_1 (x_{[i]} - \bar{x}_{[.]})\}^2.$$

Since the only admissible root of (7.2) is

$$\hat{\sigma}(\theta_1) = \left\{ -B_0 + \sqrt{(B_0^2 + 4nC_0)} \right\} \Big/ 2n \tag{7.3}$$

the result follows; see, for example, Bartlett [30] and Kendall and Stuart [31] (p. 52).

Now, $B_0/\sqrt{nC_0} \cong 0$ for large $n$. Consequently,

$$\partial \ln L^*/\partial\sigma \cong (n/\sigma^3)\{(C_0/n) - \sigma^2\}. \tag{7.4}$$

Therefore, $\hat{\sigma}^2(\theta_1) \cong C_0/(n-1)$ is the MVB estimator of $\sigma^2$. It also follows from (7.4) that for large $n$, $(n-1)\hat{\sigma}^2(\theta_1)/\sigma^2$ is a chi-square with $n-1$ degrees of freedom; see, for example, Tiku [32] (p. 626). Consequently, $(n-2)\hat{\sigma}^2/\sigma^2$ is for large $n$ referred to a chi-square distribution with $n-2$ degrees of freedom.

**Testing** $\theta_1 = 0$**:** To test $H_0: \theta_1 = 0$ against the alternatives $H_0: \theta_1 > 0$, we define the statistic

$$\mathrm{T} = \hat{\theta}_1 \sqrt{\left\{ \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2 \right\}} \Big/ \hat{\sigma}. \tag{7.5}$$

Large values of $T$ lead to the rejection of $H_0$ in favour of $H_1$. Since $\hat{\sigma}$ converges to $\sigma$ as $n$ tends to infinity, in view of Lemma 1, the null distribution of $T$ is asymptotically normal $N(0, 1)$. For $n > 20$, in fact, the $N(0, 1)$ distribution provides close approximations to the percentage points (Table 4) for all $p \geq 1.4$, in spite of the fact that for the regularity conditions to hold $p$ has to be greater than 2. For $n \leq 20$, the null distribution of $T$ is referred to Student's $t$ with $n-2$ degrees of freedom; this essentially is a consequence of Lemmas 1 and 2.

**Power Function:** The asymptotic power function of the $T$-test is given by

$$P\left\{ Z \geq z_\alpha - (\theta_1/\sigma)\sqrt{\sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2} \right\}, \tag{7.6}$$

*Table 4.* Simulated Type I Errors, Presumed Value is 0.05

| | $p$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | 1.3 | 1.4 | 1.5 | 2.0 | 2.5 | 3.0 | 4.0 | 6.0 |
| 10 | 0.057 | 0.044 | 0.039 | 0.041 | 0.051 | 0.049 | 0.050 | 0.054 |
| 20 | 0.083 | 0.039 | 0.038 | 0.038 | 0.047 | 0.049 | 0.054 | 0.051 |
| 30 | 0.110 | 0.049 | 0.032 | 0.044 | 0.048 | 0.052 | 0.059 | 0.049 |
| 50 | 0.154 | 0.059 | 0.034 | 0.045 | 0.051 | 0.056 | 0.052 | 0.054 |
| 100 | 0.144 | 0.054 | 0.031 | 0.043 | 0.042 | 0.049 | 0.049 | 0.055 |

where $Z$ is normal $N(0, 1)$ and $z_\alpha$ is its $100(1-\alpha)\%$ point. Simulations reveal that (7.6) gives accurate approximations for all $n > 20$. For $n \leq 20$, the power function is adequately approximated by a noncentral $t$ distribution but we do not pursue it in any detail. The power of the $T$-test increases to 1 very rapidly. We do not give details for conciseness.

**Robustness:** A test is said to have criterion robustness if its Type I error for plausible alternatives is not substantially higher than that attained under an assumed model. The test is said to have efficiency robustness if its power is high; see, for example, Tiku et al. [3], Preface. Consider the above $T$-test and, for comparison, the analogous test based on the LS estimators $\tilde{\theta}_1$ and $\tilde{\sigma}$, namely,

$$G = \tilde{\theta}_1\sqrt{\textstyle\sum_{i=1}^{n}(x_i-\bar{x})^2}\Big/\left\{\left[\Gamma(1+2/p)-\Gamma^2(1+1/p)\right]\tilde{\sigma}^2\right\}^{1/2}. \quad (7.7)$$

Large values of G lead to the rejection of $H_0$ in favour of $H_1$. In absence of any readily available distributional results, the null distribution of $G$ is referred to normal $N(0, 1)$ for large $n$. For $n \leq 20$, the null distribution of $G$ is referred to Student's $t$ with $n-2$ degrees of freedom. The assumed model is the Weibull $W(2, \sigma)$, and the alternatives are the following for illustration ($\sigma$ unknown):

(I) $W(1,\sigma)$, (II) $W(1.3,\sigma)$, (III) $W(3,\sigma)$,

(IV) Contamination $0.90W(2,\sigma)+0.10W(1.3,\sigma)$.

The random deviates generated from the models (I) to (IV) were divided by $\tau = 1.356$, 1.198, 0.85, and 1.022, respectively, to equalise the variances. This is important for power comparison. The simulated values of the power are given in Table 5. It can be seen that the $T$-test is robust. This is essentially due to the half-umbrella ordering of the $\delta_i$ $(1 \leq i \leq n)$ coefficients as explained earlier. The $G$-test is clearly yielding poor results: This is partly

*Table 5.* Power of the $G$ and $T$ tests, $n = 20$. True Model $W(2, \sigma)$, $\sigma = 1$

| Model | I | | II | | $W(2, \sigma)$ | | III | | IV | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | $G$ | $T$ | $G$ | $T$ | $G$ | $T$ | $G$ | $T$ | $G$ | $T$ |
| 0.0 | 0.005 | 0.014 | 0.006 | 0.021 | 0.008 | 0.033 | 0.007 | 0.051 | 0.005 | 0.024 |
| 0.2 | 0.02 | 0.06 | 0.02 | 0.08 | 0.03 | 0.12 | 0.04 | 0.17 | 0.02 | 0.09 |
| 0.4 | 0.05 | 0.18 | 0.07 | 0.23 | 0.10 | 0.32 | 0.15 | 0.39 | 0.07 | 0.24 |
| 0.6 | 0.12 | 0.38 | 0.16 | 0.46 | 0.24 | 0.57 | 0.34 | 0.65 | 0.16 | 0.45 |
| 0.8 | 0.21 | 0.59 | 0.28 | 0.68 | 0.42 | 0.79 | 0.60 | 0.85 | 0.29 | 0.66 |
| 1.0 | 0.32 | 0.76 | 0.44 | 0.85 | 0.65 | 0.92 | 0.82 | 0.96 | 0.45 | 0.82 |
| 1.2 | 0.46 | 0.86 | 0.59 | 0.93 | 0.81 | 0.97 | 0.94 | 0.99 | 0.58 | 0.90 |
| 1.4 | 0.59 | 0.93 | 0.73 | 0.97 | 0.92 | 1.00 | 0.99 | 1.00 | 0.71 | 0.96 |

due to the inadequacy of the normal approximation (Tiku [33] (p. 146)) but mainly due to the fact that the LS estimator $\tilde{\theta}_1$ is much less efficient than the MML estimator $\hat{\theta}_1$. Since for a given model the relative efficiency $E = V(\hat{\theta}_1)/V(\tilde{\theta}_1)$ of the LS estimator is less than 1 and is the same under both $H_0$ as well as $H_1$, the $T$-test is bound to be more powerful; see Sundrum [34]. As can be seen from Table 5, the power of the $G$-test declines sharply when the sample model deviates from the population model. The $G$-test is clearly nonrobust.

**Example:** Consider the following data (Johnson and Johnson [35]) which represent the ordered survival times (the number of days/1000) of 43 patients suffering from granulocytic leukemia:

$e_{(i)}$: 0.007, 0.047, 0.058, 0.074, 0.177, 0.232, 0.273, 0.285, 0.317, 0.429, 0.440, 0.445, 0.455, 0.468, 0.495, 0.497, 0.532, 0.571, 0.579, 0.581, 0.650, 0.702, 0.715, 0.779, 0.881, 0.900, 0.930, 0.968, 1.077, 1.109, 1.314, 1.334, 1.367, 1.534, 1.712, 1.784, 1.877, 1.886, 2.045, 2.056, 2.260, 2.429, 2.509

To verify whether this data is genuinely from a Weibull distribution, we utilize the goodness-of-fit statistic $Z_W$ (Tiku [36], Tiku and Singh [37]) based on sample spacings. The statistic is location and scale invariant and, therefore, no parameter estimation is required in its computation.

For a given $p$, let

$$D_i = (n - i)\left\{e^p_{(i+1)} - e^p_{(i)}\right\}, \quad i = 1, 2, \ldots, n - 1, \tag{7.8}$$

be the sample spacings. Calculate the statistic

$$Z_W = 2\sum_{i=1}^{n-1}(n-1-i)D_i \Big/ (n-2)\sum_{i=1}^{n-1} D_i.$$

Small and large values of $Z_W$ lead to the rejection of Weibull (for a given $p$). The null distribution of $Z_W/2$ is the same as the distribution of the mean of $n-2$ iid *Uniform* $(0,1)$ variates (Tiku [36]). For $n \geq 7$, therefore, the null distribution of

$$Z = (Z_W - 1)\sqrt{3(n-2)} \tag{7.9}$$

is referred to $N(0,1)$. For testing exponentiality, the $Z_W$ test is known to be the most powerful test overall (Dyer and Harbin [38]).

We calculated $Z$ for various values of $p$. One of the authors of this paper was very pleased to notice that $Z$ monotonically decreases from positive to negative values as $p$ increases. It attains the value zero when $p = 1.314$. The Weibull $W(1.314, \sigma)$ is, therefore, the most plausible model for the data above. A Q-Q plot of $e_{(i)}$ against $t_{(i)}$, calculated from (3.2) with $p = 1.314$ yields "close to a straight line" pattern and, therefore, supports the Weibull model.

We introduced a design variable $x_i$ by taking $y_i = x_i + e_i$. The values of $x_i$ (generated from a Uniform distribution) are given below:

0.00, 0.08, 0.60, 0.89, 0.97, 0.19, 0.52, 0.40, 0.26, 0.74, 0.09, 0.56, 0.58, 0.81, 0.59, 0.51, 0.88, 0.99, 0.73, 0.97, 0.30, 0.43, 0.90, 0.65, 0.90, 0.96, 0.16, 0.86, 0.91, 0.29, 0.94, 0.42, 0.31, 0.52, 0.40, 0.79, 0.69, 0.54, 0.59, 0.09, 0.61, 0.43, 0.60.

The model (2.1) is now applicable with $p$ in $W(p, \sigma)$ estimated by 1.314. We assume, of course, that $\theta_0, \theta_1$ and $\sigma$ are not known. The LS and the MML estimates of $\theta_1$ and their standard errors (calculated from the equations above) are

| | Estimate | Standard Error |
|---|---|---|
| LS | 0.99 | $\pm 0.39$ |
| MML | 0.97 | $\pm 0.22$ |

Both the estimates are close to the population value ($\theta_1 = 1$) but the MML estimate has considerably smaller standard error, as expected.

**Remark:** For the model (2.1), it will be $y_i$ (not $e_i$) that will be known. The spacings $D_i$, therefore, are calculated by replacing $e_{(i)}$ by $\hat{e}_{(i)}$, $(1 \leq i \leq n)$, where $\hat{e}_{(i)}$ are the order statistics of $y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i$. The $Z$-test above is then applied. A Q-Q plot obtained by plotting $\hat{e}_{(i)}$ against $t_{(i)}$, $(1 \le i \le n)$, provides corroborative evidence; see also Tiku and Vaughan [39] and Tiku et al. [17] who show that such a procedure yields good results since the MML estimators are robust.

## 8. GENERALISED LOGISTIC

In model (2.1), suppose $e_i$ have the generalised logistic distribution

$$GL(b,\sigma) = \frac{b}{\sigma}\frac{\exp(-e/\sigma)}{\left\{1+\exp(-e/\sigma)\right\}^{b+1}}, \quad -\infty < e < \infty. \tag{8.1}$$

The values of $b$ that are of interest are $0.4 \le b \le 8$, for the probability

$$prob\left\{y \ge \theta_0 + \theta_1 x + E(e)\right\} = \left\{1 + e^{-c}\right\}^b \tag{8.2}$$

to have values between 0.4 and 0.6.

The likelihood equations for estimating $\theta_0$, $\theta_1$ and $\sigma$ can be written in terms of the ordered variates $z_{(i)}$, $1 \le i \le n$, as in (2.4)–(2.6), and are expression in terms of the awkward functions

$$\begin{aligned} &g(z) = e^{-z}/(1+e^{-z}); \quad z = z_{(i)} = (w_{(i)} - \theta_0)/\sigma, \\ &w_{(i)} = y_{[i]} - \theta_1 x_{[i]} \quad (1 \le i \le n). \end{aligned} \tag{8.3}$$

Since $g\{z_{(i)}\}$ is linear (almost) in the vicinity of $z_{(i)}$, we have as in (3.3),

$$g\{z_{(i)}\} \cong \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n, \tag{8.4}$$

where $(t = t_{(i)})$

$$\begin{aligned} &\alpha_i = (1 + e^t + te^t)/(1+e^t)^2 \quad \text{and} \quad \beta_i = e^t/(1+e^t)^2; \\ &t_{(i)} = -\ln\left(q_i^{-1/b} - 1\right), \quad q_i = i/(n+1). \end{aligned}$$

**MML Estimators:** Incorporating (8.4) in the likelihood equations, we obtain the modified likelihood equations. The solutions of these equations are the MML estimators:

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]} - (\Delta/m)\hat{\sigma} \tag{8.5}$$

$$\hat{\theta}_1 = K - D\hat{\sigma}, \quad \text{and} \quad \hat{\sigma} = \left\{-B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{\{n(n-2)\}}, \tag{8.6}$$

where

$$
\begin{aligned}
m &= \sum_{i=1}^{n} \beta_i; \quad \Delta_i = \alpha_i - (b+1)^{-1}; \quad \Delta = \sum_{i=1}^{n} \Delta_i; \\
\bar{y}_{[.]} &= (1/m) \sum_{i=1}^{n} \beta_i y_{[i]}, \quad \bar{x}_{[.]} = (1/m) \sum_{i=1}^{n} \beta_i x_{[i]}; \\
K &= \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} \Bigg/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2, \\
D &= \sum_{i=1}^{n} \Delta_i (x_{[i]} - \bar{x}_{[.]}) \Bigg/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2; \qquad (8.7) \\
B &= (b+1) \sum_{i=1}^{n} \Delta_i \{ y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}) \}, \\
C &= (b+1) \sum_{i=1}^{n} \beta_i \{ y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}) \}^2 \\
&= (b+1) \left\{ \sum_{i=1}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]})^2 - K \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} \right\}.
\end{aligned}
$$

Realise that $\beta_i > 0$ and, hence, $m > 0$.

The estimators are computed in two iterations, exactly the same way as before. Note that for $b = 1$ (logistic distribution), $\Delta = 0$ for all $n$.

## 9. RELATIVE EFFICIENCY

The Fisher information matrix is given in the Appendix. The variances and covariances of the MML estimators for large $n$ ($n > 50$) may be obtained from this matrix. The LS estimators corrected for bias are

$$
\begin{aligned}
&\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1 \bar{x} - \{\varphi(b) - \varphi(1)\}\tilde{\sigma}, \quad \tilde{\theta} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \Bigg/ \sum_{i=1}^{n} (x_i - \bar{x})^2, \\
&\text{and} \quad \tilde{\sigma} = \sqrt{\textstyle\sum_{i=1}^{n} \{y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x})\}^2 / (n-2) \{\varphi'(b) + \varphi'(1)\}}; \qquad (9.1)
\end{aligned}
$$

$\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$. The expression of the $\varphi$-function $\varphi(u)$ and its derivatives $\varphi'(u)$ are given in Abramowitz and Stegun

[40]. Their values are given in Tiku et al. [18] (Appendix) for several values of $b$.

To compare the efficiencies of the MML estimators with the LS estimators, we simulated their means and variances for several values of $b$. The biases in both the MML and the LS estimators were found to be negligible and are not, therefore, reported. Given in Table 6 are the simulated values of the variances of the MML estimators $\hat{\theta}_0, \hat{\theta}_1$ and $\hat{\sigma}$ and the relative efficiencies $E_1$, $E_2$ and $E_3$ (defined earlier) of the LS estimators $\tilde{\theta}_0, \tilde{\theta}_1$ and $\tilde{\sigma}$, respectively. Without loss of generality, $\theta_0, \theta_1$ and $\sigma$ were taken to be equal to 0, 1 and 1, respectively. The design points $x_i$, $1 \le i \le n$, were

***Table 6.*** Variances of the MML Estimators and the Relative Efficiencies of the LS Estimators: (1) $nV(\hat{\theta}_0)$, (2) $nV(\hat{\theta}_1)$, (3) $nV(\hat{\sigma})$

| $n$ | (1) | $E_1$ | (2) | $E_2$ | (3) | $E_3$ |
|---|---|---|---|---|---|---|
| | | | $b = 0.5$ | | | |
| 10 | 17.04 | 88 | 53.86 | 85 | 0.970 | 98 |
| 20 | 23.85 | 85 | 56.80 | 82 | 0.856 | 85 |
| 30 | 26.14 | 81 | 57.09 | 79 | 0.842 | 80 |
| 50 | 25.29 | 82 | 63.67 | 79 | 0.844 | 78 |
| 100 | 20.37 | 79 | 57.65 | 75 | 0.711 | 65 |
| | | | $b = 2.0$ | | | |
| 10 | 6.72 | 95 | 20.48 | 93 | 0.841 | 100 |
| 20 | 9.26 | 91 | 21.56 | 89 | 0.729 | 92 |
| 30 | 9.98 | 93 | 22.16 | 91 | 0.709 | 89 |
| 50 | 10.09 | 90 | 25.30 | 89 | 0.707 | 87 |
| 100 | 8.48 | 90 | 24.85 | 88 | 0.678 | 83 |
| | | | $b = 4.0$ | | | |
| 10 | 6.00 | 87 | 16.27 | 86 | 0.804 | 91 |
| 20 | 7.74 | 82 | 16.60 | 81 | 0.724 | 81 |
| 30 | 7.86 | 78 | 16.16 | 77 | 0.662 | 73 |
| 50 | 8.37 | 79 | 19.26 | 79 | 0.675 | 75 |
| 100 | 7.53 | 78 | 20.30 | 79 | 0.663 | 70 |
| | | | $b = 8.0$ | | | |
| 10 | 6.65 | 80 | 13.55 | 78 | 0.799 | 88 |
| 20 | 7.95 | 73 | 13.82 | 74 | 0.667 | 75 |
| 30 | 8.34 | 70 | 14.31 | 74 | 0.673 | 67 |
| 50 | 7.76 | 68 | 15.07 | 72 | 0.634 | 68 |
| 100 | 7.34 | 70 | 14.86 | 70 | 0.653 | 68 |

generated from a Uniform (0, 1) as before and were common to all the $N = [100{,}000\,/\,n]$ random samples $(y_1, y_2, \ldots, y_n)$ generated from the generalised logistic. It can be seen that the MML estimators are considerably more efficient. Note again the disconcerting feature of the LS estimators, i.e., their relative efficiencies decrease as the sample size $n$ increases.

**Testing $\theta_1 = 0$:** As in Section 7, it is easy to show that $\hat{\theta}_1(\sigma)$ is conditionally the MVB estimator (asymptotically) and is normally distributed with variance

$$\sigma^2/\{(b+1)\sum_{i=1}^{n}\beta_i(x_{[i]} - \bar{x}_{[.]})^2\}. \tag{9.2}$$

To test $H_0$, therefore, the statistic

$$T = \hat{\theta}_1\sqrt{\left\{(b+1)\sum_{i=1}^{n}\beta_i(x_{[i]} - \bar{x}_{[.]})^2\right\}}\bigg/\hat{\sigma} \tag{9.3}$$

can be used. Large values of $T$ lead to the rejection of $H_0$ in favour of $H_1 : \theta_1 > 0$. The null distribution of $T$ is referred to Student's $t$ with $n-2$ degrees of freedom for $n \le 20$, and to normal $N(0, 1)$ for $n > 20$. These distributions give accurate approximations for the probabilities and the percentage points of $T$. The simulated values of the Type I errors, for example, are given in Table 7.

The $T$-test is considerably more powerful than the analogous $G$-test based on the LS estimators, as expected (Sundrum [34]). We omit details for conciseness.

**Robustness**: The $T$ test is robust to reasonable deviations in the values of $b$ in (8.1), and to outlier and mixture models. This is due to the fact that the coefficients $\beta_i$ which correspond to largest ordered residuals $|\hat{e}_{(i)}|$ in the

***Table 7.*** Simulated Values of the Type I Error of the $T$ Test, Presumed Value is 0.05

| | $b$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | 0.2 | 0.5 | 1.0 | 2.0 | 4.0 | 6.0 | 8.0 |
| 10 | 0.062 | 0.051 | 0.052 | 0.051 | 0.049 | 0.055 | 0.051 |
| 20 | 0.066 | 0.050 | 0.051 | 0.046 | 0.046 | 0.046 | 0.045 |
| 30 | 0.064 | 0.062 | 0.053 | 0.049 | 0.053 | 0.051 | 0.051 |
| 50 | 0.060 | 0.058 | 0.060 | 0.051 | 0.049 | 0.045 | 0.052 |
| 100 | 0.045 | 0.041 | 0.055 | 0.056 | 0.061 | 0.056 | 0.048 |

direction of the long tail are small. Thus, the effect of long tails and outliers is automatically depleted.

**Comment:** One might argue that the shape parameters $p$ and $b$ in (2.2) and (8.1), respectively, should also be estimated rather than their plausible values located through Q-Q plots and goodness-of-fit tests as in the numerical example above. It may be noted, however, that shape parameters are very difficult to estimate and it takes a very large sample size to reduce their bias and variance to desirable limits; see also Pearson and Hartley [41] (p. 87–9). Using a strategically chosen value of a shape parameter leads to more robust and efficient estimators of the parameters in models like (2.1) than those based on their estimates obtained, for example, from the likelihood equations. This has been amply demonstrated in Tiku et al. [19] (Section 8). See also Tiku and Vaughan [39].

## 10. GENERALISATION

The results above readily generalise to multiple linear regression models

$$y_i = \theta_0 + \theta_1 x_{1i} + \cdots + \theta_k x_{ki} + e_i, \quad 1 \le i \le n. \tag{10.1}$$

Suppose $e_i$ are iid and have the Weibull distribution $W(p, \sigma)$. Let $(y_{[i]}, x_{1[i]}, \ldots, x_{k[i]})$ be the concomitants of $e_{(i)}$, $1 \le i \le n$. The MML estimator of $\theta_0$ is

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{1[.]} - \cdots - \hat{\theta}_k \bar{x}_{k[.]} - (\Delta/m)\hat{\sigma}. \tag{10.2}$$

Writing $Y_{[i]} = y_{[i]} - \bar{y}_{[.]}$, $X_{1[i]} = x_{1[i]} - \bar{x}_{1[.]}, \ldots, X_{k[i]} = x_{k[i]} - \bar{x}_{k[.]}$, the MML estimators of $\theta_i$ $(1 \le i \le k)$ and $\sigma$ are

$$\hat{\theta} = (X'\Gamma_\delta X)^{-1}\{X'\Gamma_\delta Y - \hat{\sigma} X'\Delta\} \tag{10.3}$$

and

$$\hat{\sigma} = \left\{-B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{\{n(n-k-1)\}}. \tag{10.4}$$

Here,

$$Y = \begin{bmatrix} Y_{[1]} \\ Y_{[2]} \\ \vdots \\ Y_{[n]} \end{bmatrix}, \quad X = \begin{bmatrix} X_{1[1]} & X_{2[1]} & \cdots & X_{k[1]} \\ X_{1[2]} & X_{2[2]} & \cdots & X_{k[2]} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1[n]} & X_{2[n]} & \cdots & X_{k[n]} \end{bmatrix};$$

$$\Gamma_\delta = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & \delta_n \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & \Delta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & \Delta_n \end{bmatrix}; \tag{10.5}$$

$$B = \sum_{i=1}^{n} \Delta_i \{ Y_{[i]} - K_1 X_{1[i]} - \cdots - K_k X_{k[i]}$$

and

$$C = \sum_{i=1}^{n} \{\delta_i Y_{[i]}^2 - K_1 Q_1 - \cdots - K_k Q_k\},$$

$$Q_j = \sum_{i=1}^{n} \delta_i (X_{j[i]} - \bar{X}_{j[.]}) y_{[i]};$$

$$K_i = \sum_{i=1}^{n} \delta_i (x_{j[i]} - \bar{x}_{j[.]}) y_{[i]} \Bigg/ \sum_{i=1}^{n} \delta_i (x_{j[i]} - \bar{x}_{j[.]})^2, \quad (1 \le j \le k).$$

The coefficients $\delta_i$ and $\Delta_i$ are defined in (3.8). The estimators in (10.2)–(10.4) have essentially the same efficiency and robustness properties as for the case $k = 1$. We omit details for conciseness.

## 11. CONCLUDING REMARKS

It is widely recognised that nonnormal distributions, particularly asymmetric, occur so frequently in practice. It is also recognised that samples often contain outliers. In such situations, the maximum likelihood estimation can be problematic, rather debilitating (Puthenpura and Sinha [5]). In this paper, we have used the method of modified likelihood for estimating parameters in a linear model with asymmetric error distributions, Weibull and generalised logistic for illustration. The resulting estimators, called MML estimators, are explicit functions of sample observations and are, therefore, easy to compute. Being explicit functions, they are also amenable to analytic studies. We have shown that the MML estimators are remarkably efficient and robust. In fact, they are asymptotically the MVB estimators under general regularity conditions. We have also developed hypothesis testing procedures and shown them to be robust and powerful. We believe that the method developed in this paper can be successfully adopted when the error distributions in linear models are symmetric, both long and short tailed, e.g.,

$$f(e) \propto (1/\sigma)\{1 + e^2/k\sigma^2\}^{-p}, \quad -\infty < e < \infty, \tag{11.1}$$

with $k = 2p - 3$, $p \geq 2$, and

$$f(e) \propto (1/\sigma)[1 + e^2/2(r-a)\sigma^2]^r\{1 + e^2/2k\sigma^2\}^{-p}, \quad -\infty < e < \infty, \tag{11.2}$$

with $k = p - 3/2$, $p > r + 3/2$, $r > a$ and $r$ is an integer. The family (11.1) represents long tailed symmetric distributions with kurtosis greater than 3. The family (11.2) was recently introduced by Tiku and Vaughan [42] and represents short tailed symmetric distributions with kurtosis less than 3.

## APPENDIX

A rigorous proof of the asymptotic equality, under regularity conditions, of the expected values $-E(d\ln L^*/d\theta)$ and $-E(d^2\ln L^*/d\theta^2)$ of a modified likelihood equation $d\ln L^*/d\theta = 0$ and the expected values $-E(d\ln L/d\theta)$ and $-E(d^2\ln L/d\theta^2)$ of the corresponding likelihood equation $d\ln L/d\theta = 0$, is given by Vaughan and Tiku [13] (Appendix A); see also Bhattacharyya [12]. The proof is based on a result due to Hoeffding [43], namely, if a function $g$ is such that $|g(y)| \leq h(y)$ for some non-negative convex function $h$ with finite expectation $\int_0^\infty h(y)f(y)\,dy$, $f(y)$ being the pdf of $y$, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(t_{i:n}) = \int_0^\infty g(z)f(z)\,dz; \tag{A.1}$$

$t_{i:n}$ denotes the expected value of the $i$th order statistic in a random sample of size $n$ from $f(z)$. A heuristic proof follows along the following lines.

Consider for example, equation (8.4). Since $g(z)$ is bounded and the variance $V\{z_{(i)}\}$ tends to zero as $n$ tends to infinity, (8.4) is asymptotically an equality since the differences $g\{z_{(i)}\} - (\alpha_i - \beta_i z_{(i)})$ tend to zero. Thus the differential coefficients with respect to $\theta_0$, $\theta_1$ and $\sigma$ on both sides of (8.4) are asymptotically equal. This immediately gives the result that the MML estimators $\hat{\theta}_0, \hat{\theta}_1$ and $\hat{\sigma}$ are asymptotically unbiased and their covariance matrix (asymptotic) is $I^{-1}$, where $I$ is the Fisher information matrix. For the generalised logistic, the elements of $I$ are

$$\begin{aligned}
&I_{11} = nb/(b+2)\sigma^2, \quad I_{12} = b\sum x_i/(b+2)\sigma^2, \\
&I_{13} = nb\{\varphi(b+1) - \varphi(2)\}/(b+2)\sigma^2 \\
&I_{22} = b\sum x_i^2/(b+2)\sigma^2, \quad I_{23} = b\{\varphi(b+1) - \varphi(2)\}\sum x_i/(b+2)\sigma^2 \\
&I_{33} = nb\{(b+2)/b + [\varphi'(b+1) + \varphi'(2)] + [\varphi(b+1) - \varphi(2)]^2\}/(b+2)\sigma^2.
\end{aligned} \tag{A.2}$$

The numerical values of $\varphi$ and $\varphi'$ functions are given in Tiku et al. [18] (Appendix) for various values of b.

## ACKNOWLEDGMENT

We are grateful to the referee and the Associate Editor for very useful comments. The second author would like to thank METU (Middle East Technical University) for a Visiting Professorship and to the NSERC (Canada) for a research grant.

## REFERENCES

1. Barnett, V. D. (1966). "Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots". *Biometrika*, 53, 151–165.
2. Lee, K. R., Kapadia, C. H. and Dwight, B. B. (1980). "On estimating the scale parameter of Rayleigh distribution from censored samples". *Statist. Hefte*, 21, 14–20.
3. Tiku, M. L., Tan, W. Y. and Balakrishnan, N. (1986). *Robust Inference*. New York: Marcel Dekker.
4. Vaughan, D. C. (1992). "On the Tiku-Suresh method of estimation". Commun. Stat.—Theory Meth., 21, 451–469.
5. Puthenpura, S. and Sinha, N. K. (1986). "Modified maximum likelihood method for the robust estimation of system parameters from very noisy data". *Automatica*, 22, 231–235.

6. Tiku, M. L. (1967). "Estimating the mean and standard deviation from censored normal samples". *Biometrika*, 54, 155–165.
7. Tiku, M. L. (1968). "Estimating the parameters of normal and logistic distributions from censored samples". *Australian J. Stat.*, 10, 64–74.
8. Tiku, M. L. (1973). "Testing group effects from Type II censored normal samples in experimental design". *Biometrics*, 29, 25–33.
9. Tiku, M. L. and Suresh, R. P. (1992). "A new method of estimation for location and scale parameters". *J. Stat. Plann. Inf.*, 30, 281–292.
10. Tan, W. Y. (1985). "On Tiku's robust procedure – a Bayesian insight". *J. Stat. Plann. Inf.*, 11, 329–340.
11. Tan, W. Y. and Balakrishnan, N. (1986). "Bayesian insight into Tiku's robust procedure based on asymmetric censored samples". *J. Stat. Comput. Simul.*, 24, 17–31.
12. Bhattacharyya, G. K. (1985). "The asymptotics of maximum likelihood and related estimators based on Type II censored data". *J. Amer. Stat. Assoc.*, 80, 398–404.
13. Vaughan, D. C. and Tiku, M. L. (2000). "Estimation and hypothesis testing for a non-normal bivariate distribution and applications". *J. Mathematical and Computer Modelling*, 32, 53–67.
14. Smith, W. B., Zeis, C. D. and Syler, G. W. (1973). "Three parameter lognormal estimation from censored data". *J. Indian Stat. Assoc.*, 11, 15–31.
15. Bian, G. and Tiku, M. L. (1974a). "Bayesian inference based on robust priors and MML estimators: Part I, symmetric location-scale distributions". *Statistics*, 29, 317–345.
16. Bian, G. and Tiku, M. L. (1974b). "Bayesian inference based on robust priors and MML estimators: Part II, skew location-scale distributions". *Statistics*, 29, 81–99.
17. Tiku, M. L. and Selçuk, A. S. (2000). "Robust time series: some engineering applications". *IEEE*. (Invited paper): Proceedings of Goa Conference (ICIT-2000), 466–473.
18. Tiku, M. L., Wong, W. K. and Bian, G. (1999). "Time Series models with asymmetric innovations". *Commun. Stat.*—Theory Meth., 28, 1131–1160.
19. Tiku, M. L., Wong, W. K. and Vaughan, D. C. (2000). "Time series models in non-normal situations: symmetric innovations". *J. Time Series Analysis*, 21, 571–596.
20. Cohen, A. C. and Whitten, B. J. (1988). *Prameter Estimation in Reliability and Life Span Models*. New York: Marcel Dekker.
21. Smith, R. L. (1985). "Maximum likelihood estimation in a class of nonregular cases". *Biometrika*, 72, 67–90.

22. Yildirim, F. and Korasli, C. (1992). "Statistical approach for determining break down voltage of Gas-insulated cables. *IEEE Transactions on Electrical Insulation*, 27, 1186–1192.
23. Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley.
24. Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley.
25. Tan, W. Y. and Tiku, M. L. (1999). *Sampling Distributions In Terms Of Laguerre Polynomials And Applications*. New Delhi, India: New Age International (Formerly, Wiley Eastern).
26. Hawkins, D. M. (1977). "Comment on 'A new statistic for testing suspected outliers' ". *Commun. Stat.*, A6, 435–438.
27. Tiku, M. L. (1977). Rejoinder: "Comment on 'A new statistic for testing suspected outliers' ". *Commun. Stat.*, A6, 1417–1422.
28. Tiku, M. L. (1975). "A new statistic for testing suspected outliers". Commun. Stat., 4, 737–752.
29. Mann, N. R. (1982). "Optimal outlier tests for a Weibull model to identify process changes or to predict failure times". *TIMS/ Studies in the Management Sciences*, 19, 261–279.
30. Bartlett, M. S. (1953). "Approximate confidence intervals". *Biometrika*, 40, 12–19.
31. Kendall, M. G. and Stuart, A. (1979). *The Advanced Theory of Statistics*. London: Charles Griffin.
32. Tiku, M. L. (1982). "Testing linear contrasts of means in experimental design without assuming normality and homogeneity of variances". *Biometrical J.*, 24, 613–627. (Invited paper).
33. Tiku, M. L. (1971). "Student's t distribution under non-normal situations". *Australian J. Stat.*, 13, 142–148.
34. Sundrum, R. M. (1954). "On the relation between estimating efficiency and the power of tests". *Biometrika*, 41, 542–548.
35. Johnson, R. C. and Johnson, N. L. (1979). *Survival Models and Data Analysis*. New York: John Wiley.
36. Tiku, M. L. (1980). "Goodness of fit statistics based on the spacings of complete or censored samples". *Australian J. Stat.*, 22, 260–275.
37. Tiku, M. L. and Singh, M. (1981). "Testing the two parameter Weibull distribution". *Commun. Stat.—Theory Meth.*, 10, 907–918.
38. Dyer, D. and Harbin, M. S. (1981). "An empirical power study of multi-sample tests for exponentiality". *J. Stat. Comp. Simul.*, 12, 277–291.
39. Tiku, M. L. and Vaughan, D. C. (1997). "Logistic and nonlogistic density functions in binary regression with nonstochastic covariates". *Biometrical J.*, 39, 883–898.

40. Abramowitz, M. and Stegun, I. A. (1985). *Handbook of Mathematical Functions*. New York: Dover.
41. Pearson, E. S. and Hartley, H. O. (1972). *Biometrika Tables For Statisticians, Vol. II*. Cambridge: The University Press.
42. Tiku, M. L. and Vaughan, D. C. (1999). ''A family of short-tailed symmetric distributions''. *Technical Report*: McMaster University, Canada.
43. Hoeffding, W. (1953). ''On the distribution of the expected values of the order statistics''. *Ann. Math. Stat*., 24, 93–100.

Received December 1999
Revised February 2001