Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(2) 2010 157-170

Random Design in Regression Models

Moti L. Tiku[1] and Aysen D. Akkaya[2*]
[1]McMaster University, Hamilton, Ontario, Canada
[2]*Department of Statistics, Middle East Technical University, 06531 Ankara, Turkey*

Received 05 January 2010; Revised 06 February 2010; Accepted 10 March 2010

SUMMARY

In regression models the design variable has traditionally been assumed to be non-stochastic. In most real life situations, however, the design variable is stochastic having a non-normal distribution as the response error. Modified maximum likelihood method is utilized to estimate unknown parameters in such situations. The resulting estimators are shown to be efficient and robust. A real life example is given.

Keywords: Random design, Regression, Non-normality, Least squares, Maximum likelihood, Modified maximum likelihood, Efficiency, Outliers, Inliers, Robustness.

1. INTRODUCTION

In the regression model

$$y = \eta(x) + e \qquad (1.1)$$

the design variable X has traditionally been assumed to be non-stochastic and the random error e as normal $N(0, \sigma^2)$, $\eta(x)$ being a linear or non-linear function. In most real life situations, however, X is stochastic and both X and e have non-normal distributions (Hutchinson and Lai 1990, Vaughan and Tiku 2000, Sazak *et al*. 2006, Tiku *et al*. 2008). Two very general methods of estimation are available: (a) least squares, and (b) maximum likelihood. LSEs (least squares estimators), however, are neither efficient (for non-normal distributions) nor robust; see, for example, Islam and Tiku (2004), Sazak *et al*. (2006), Tiku *et al*. (2008, 2009) and Akkaya and Tiku (2008a). MLEs (maximum likelihood estimators) are elusive in most situations because the maximum likelihood equations involve nonlinear functions and are, consequently, very difficult to solve even iteratively. Moreover, iterative solutions can be problematic for reasons of (i) slow convergence, (ii) convergence to wrong values (e.g. to a local rather than the global maximum), and (iii) not converging at all (Puthenpura and Sinha 1986, Qumsiyeh (2007, pp. 8-14)). Therefore, we utilize the method of modified maximum likelihood estimation. The method was introduced and developed by Tiku (1967, 1968, 1989) and Tiku and Suresh (1992). The resulting MMLEs (modified maximum likelihood estimators) are explicit fucntions of sample observations and are, therefore, easy to compute. They are enormously more efficient (for non-normal distributions) than the LSEs, particularly for large n.

Realize that non-normal distributions occur frequently in practice (Pearson 1931, Spjϕtvoll and Aastveit 1980, Elveback *et al*. 1970). The method of modified maximum likelihood estimation is carried out in three steps: (i) the maximum likelihood equations are expressed in terms of the order statistics of a sample, (ii) the non-linear functions are replaced by linear approximations so that the differences between the two converge to zero as n becomes large, and (iii) the resulting equations are solved. The solutions are called MMLEs. Under some very general regularity conditions, MMLEs are known to be asymptotically

*Corresponding author : A.D. Akkaya
E-mail address : akkay@metu.edu.tr

fully efficient (unbiased and having minimum variances). A rigorous proof of this is given in Bhattacharyya (1985) for censored samples and Vaughan and Tiku (2000) for complete samples. For small sample sizes, MMLEs are known to be essentially as efficient as MLEs and the two are numerically very close to one another; see Tiku and Vaughan (1997, pp. 890-892), Schneider (1986, p. 104), Tiku *et al*. (1986, pp. 106-107), Vaughan (2002, p. 228), Tiku and Akkaya (2004, p.52) and Kantar and Senoglu (2008, Examples 1-2). In this paper we specifically take $\eta(x)$ to be a quadratic function while both X and e are stochastic and, as usual, mutually independent. We derive LSEs and MMLEs and show that the latter are considerably more efficient (for non-normal distributions) and robust to plausible deviations from the assumed distributions and to mild data anomalies (e.g. outliers). If the distributions of both X an e are normal, LSEs and MMLEs are identical. This paper should be read in conjunction with Vaughan and Tiku (2000), Sazak *et al*. (2006) and Tiku *et al*. (2008) who assume $\eta(x)$ to be a linear function; see also Islam and Tiku (2009). The purpose of this paper is to extend the results to quadratic functions while both X and e are stochastic. We show that non-normality of the design variable has devastating effect on the efficiencies of LSEs; assumption of a non-stochastic design obscures this fact. Compared to LSEs, MMLEs are shown to be considerably more efficient and robust although a little more difficult to compute.

2. STOCHASTIC MODEL

As an extension of a linear stochastic regression model (Vaughan and Tiku 2000, Sazak *et al*. 2006, Tiku *et al*. 2008, Islam and Tiku 2009), consider the quadratic stochastic model

$$y_i = \theta_0 + \theta_1 u_i + \theta_2 u_i^2 + e_i \ (1 \le i \le n),$$
$$u_i = (x_i - \mu_1)/\sigma_1 \qquad (2.1)$$

μ_1 and σ_1 are the location and scale parameters in the distribution of X, respectively. Assume that $E(e) = 0$ and $V(e) = \sigma_{2.1}^2 = \sigma^2$, X and e being mutually independent. We will show that the model (2.1) is advantageous because the MMLEs and the LSEs of θ_0, θ_1, θ_2 and σ are invariant to the location and scale of X. This invariance is very important in many applications (Dedieu and Ogorzalek 1997, Voss *et al*. 2004).

Long-tailed symmetric distributions: A broad range of long-tailed symmetric (LTS) distributions (kurtosis $\mu_4/\mu_2^2 > 3$) is given by ($k = 2p - 3$, $p \ge 2$)

$$f(z) = \frac{\Gamma(p)}{\sqrt{k}\,\Gamma(1/2)\Gamma(p-1/2)}\left(1+\frac{z^2}{k}\right)^{-p}, \ -\infty < z < \infty \qquad (2.2)$$

$E(Z) = 0$ and $V(Z) = 1$. The kurtosis of the distribution is $3(p\text{-}3/2)/(p\text{-}5/2)$. For $p = \infty$, (2.2) reduces to normal $N(0,1)$. Note that the distribution of $t = \sqrt{(\upsilon/k)}Z$ is Student's t with $\upsilon = 2p - 1$ degrees of freedom.

Assume that the distributions of $(x_i - \mu_1)/\sigma_1$ and e_i/σ are given by (2.2) with p equated to p_1 and p_2, respectively. It is easy to show that

$$E(Y) = \theta_0 + \theta_2 \text{ and} \qquad (2.3)$$
$$V(Y) = \sigma^2 + \theta_1^2 + 2\theta_2^2\,(p_1 - 1)/(p_1 - 5/2)$$

For the variances $\sigma_1^2 = V(X)$ and $\sigma^2 = V(e)$ to exist, both p_1 and p_2 have to be greater than 3/2. For the variance $V(Y)$ to exist, however, p_1 has to be greater than 5/2.

Comment: Sharp differences start appearing between the results when the design variable X is non-stochastic and when it is stochastic. In the former situation, for example, the variance of Y is the same as that of e (i.e., σ^2) which is considerably smaller than what it is in the latter situation even if X is normal ($p_1 = \infty$); see equation (2.3). Notice the very dominant role the quadratic term in the model (2.1) plays. If θ_2 was zero, the variance of Y would be much smaller, equal to $\sigma^2 + \theta_1^2$, and would not depend on the shape parameter p_1 in the distribution of X. The variance of Y heavily depends on p_1 if $\theta_2 \ne 0$.

LSEs: The least squares estimators are obtained by minimizing the error sums of squares

$$\sum_{i=1}^{n}(x_i - \mu_1)^2 \text{ and } \sum_{i=1}^{n}(x_i - \theta_0 - \theta_1 u_i - \theta_2 u_i^2)^2 \qquad (2.4)$$

They are

$$\tilde{\mu}_1 = \bar{x},\ \tilde{\sigma}_1 = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)} \text{ (bias-corrected)}$$

$$\tilde{\theta} = (\tilde{\mathbf{U}}'\tilde{\mathbf{U}})^{-1}(\tilde{\mathbf{U}}'\mathbf{Y}) \text{ and } \tilde{\sigma} = s_e \quad (2.5)$$

where

$$s_e^2 = \sum_{i=1}^{n}(y_i - \tilde{\theta}_0 - \tilde{\theta}_1\tilde{u}_i - \tilde{\theta}_2 u_i^2)^2/(n-3)$$

$$\tilde{u}_i = (x_i - \tilde{\mu}_1)/\tilde{\sigma}_1$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad \tilde{\mathbf{U}} = \begin{bmatrix} 1 & \tilde{u}_1 & \tilde{u}_1^2 \\ 1 & \tilde{u}_2 & \tilde{u}_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \tilde{u}_n & \tilde{u}_n^2 \end{bmatrix} \text{ and } \tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\theta}_0 \\ \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{bmatrix}$$

The variance $V(\tilde{\mu}_1) = \sigma_1^2/n$. It is not easy to work out the variance of $\tilde{\sigma}_1$; $Cov(\tilde{\mu}_1, \tilde{\sigma}_1) = 0$ follows from the symmetry of the distribution (2.2). Asymptotically (Roy and Tiku 1962),

$$V(\tilde{\sigma}_1) \cong \frac{\sigma_1^2}{2n}(1 + \tfrac{1}{2}\lambda_4) \quad (2.6)$$

where $\lambda_4 = (\mu_4/\mu_2^2) - 3$ and μ_4/μ_2^2 is the kurtosis of the distribution. For a normal distribution, $\lambda_4 = 0$.

MMLEs: Realizing that $L_e = L_{y|x}$, the likelihood function L is

$$L = L_x L_e \propto \left(\frac{1}{\sigma_1}\right)^n \left[\prod_{i=1}^{n}\left(1 + \frac{u_i^2}{k_1}\right)^{-p_1}\right]$$

$$\left(\frac{1}{\sigma}\right)^n \left[\prod_{i=1}^{n}\left(1 + \frac{z_i^2}{k_2}\right)^{-p_2}\right], \; (z_i = e_i/\sigma) \quad (2.7)$$

$k_1 = 2p_1 - 3$ and $k_2 = 2p_2 - 3$. The maximum likelihood equations expressed in terms of the ordered (in increasing order of magnitude) variates $u_{(i)}$ and

$$z_{(i)} = (y_{[i]} - \theta_0 - \theta_1 u_{[i]} - \theta_2 u_{[i]}^2)/\sigma \quad (1 \le i \le n)$$

are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{2p_1}{k_1\sigma_1}\sum_{i=1}^{n} g_1(u_{(i)}) - \frac{2p_2}{\sigma_1\sigma\, k_2}\sum_{i=1}^{n}(\theta_1 + 2\theta_2 u_{[i]})\, g_2(z_{(i)}) = 0$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{2p_1}{\sigma_1 k_1}\sum_{i=1}^{n} u_{(i)}\, g_1(u_{(i)}) - \frac{2p_2}{\sigma_1\sigma\, k_2}\sum_{i=1}^{n}(\theta_1 u_{[i]} + 2\theta_2 u_{[i]}^2)\, g_2(z_{(i)}) = 0$$

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{2p_2}{k_2\sigma}\sum_{i=1}^{n} g_2(z_{(i)}) = 0$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{2p_2}{k_2\sigma}\sum_{i=1}^{n} u_{[i]}\, g_2(z_{(i)}) = 0 \quad (2.8)$$

$$\frac{\partial \ln L}{\partial \theta_2} = \frac{2p_2}{k_2\sigma}\sum_{i=1}^{n} u_{[i]}^2\, g_2(z_{(i)}) = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p_2}{\sigma\, k_2}\sum_{i=1}^{n} z_{(i)}\, g_2(z_{(i)}) = 0$$

where

$g_1(u) = u/\{1 + (1/k_1)u^2\}$ and $g_2(z) = z/\{1 + (1/k_2)z^2\}$; $(y_{[i]}, u_{[i]})$ are the concomitants of $z_{(i)}$, i.e., the pair (y_i, u_i) associated with the i^{th} ordered value $z_{(i)}$ obtained by arranging z_i $(1 \le i \le n)$ in increasing order of magnitude.

The solutions of the equations (2.8) are the MLEs. These equations, however, have no explicit solutions. Solving so many equations by iteration is very difficult indeed and there can be problems of convergence as said earlier.

To work out MMLEs we linearize the functions

$$g_1(u_{(i)}) \cong \alpha_{1i} + \beta_{1i}u_{(i)} \text{ and } g_2(z_{(i)}) \cong \alpha_{2i} + \beta_{2i}z_{(i)} \quad (1 \le i \le n) \quad (2.9)$$

For $p_1 = p_2$, $\alpha_{1i} = \alpha_{2i}$ and $\beta_{1i} = \beta_{2i}$. The values of $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are given in Appendix *A* with *p* equated to p_1 and p_2, respectively. Incorporating (2.9) in (2.8) gives the modified maximum likelihood equations. Their solutions are the following MMLEs:

$$\hat{\mu}_1 = (1/m_1)\sum_{i=1}^{n}\beta_{1i}x_{(i)} \quad \left(m_1 = \sum_{i=1}^{n}\beta_{1i}\right)$$

$$\hat{\sigma}_1 = \left(B_1 + \sqrt{B_1^2 + 4nC_1}\right)/2n$$

$$B_1 = \frac{2p_1}{k_1}\sum_{i=1}^{n}\alpha_{1i}(x_{(i)} - \hat{\mu}_1)$$

$$C_1 = \frac{2p_1}{k_1}\sum_{i=1}^{n}\beta_{1i}(x_{(i)} - \hat{\mu}_1)^2 \qquad (2.10)$$

$$\hat{\boldsymbol{\theta}} = \mathbf{K} + \mathbf{D}\hat{\sigma} \text{ and}$$

$$\hat{\sigma} = \left(B + \sqrt{B^2 + 4nC}\right)/2\sqrt{n(n-2)} \quad \text{(bias-corrected)} \qquad (2.11)$$

$$B = \frac{2p_2}{k_2}\sum_{i=1}^{n}\alpha_{2i}\{y_{[i]} - K_0 - K_1\hat{u}_{[i]} - K_2\hat{u}_{[i]}^2\}$$

$$C = \frac{2p_2}{k_2}\sum_{i=1}^{n}\beta_{2i}\{y_{[i]} - K_0 - K_1\hat{u}_{[i]} - K_2\hat{u}_{[i]}^2\}^2$$

$$\mathbf{Y} = \begin{bmatrix} y_{[1]} \\ y_{[2]} \\ \cdot \\ \cdot \\ \cdot \\ y_{[n]} \end{bmatrix}, \quad \hat{\mathbf{W}} = \begin{bmatrix} 1 & \hat{u}_{[1]} & \hat{u}_{[1]}^2 \\ 1 & \hat{u}_{[2]} & \hat{u}_{[2]}^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \hat{u}_{[n]} & \hat{u}_{[n]}^2 \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} K_0 \\ K_1 \\ K_2 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} D_0 \\ D_1 \\ D_2 \end{bmatrix} \qquad (2.12)$$

$$\mathbf{K} = (\hat{\mathbf{W}}'\beta\hat{\mathbf{W}})^{-1}(\hat{\mathbf{W}}'\beta\mathbf{Y})$$

and $\mathbf{D} = (\hat{\mathbf{W}}'\beta\hat{\mathbf{W}})^{-1}(\hat{\mathbf{W}}'\alpha\mathbf{1})$

Remark. The divisor $2\sqrt{n(n-2)}$ (replacing $2n$) in the expression for $\hat{\sigma}$ minimizes the bias in $\hat{\sigma}$ overall.

Remark. It is easy to verify that the LSEs and the MMLEs above are invariant to location and scale of the design, i.e., if x_i are replaced by $a + bx_i$ $(1 \le i \le n)$, their values (hence, their variances and covariances) are unchanged. This invariance is very important particularly in engineering applications (Voss *et al*. (2004), Akkaya and Tiku (2008a)).

Computations. Computation of the MMLEs $\hat{\mu}_1$ and $\hat{\sigma}_1$ is straightforward. To compute the MMLEs of θ_0, θ_1, θ_2 and σ, we need the concomitants $(y_{[i]}, \hat{u}_{[i]})$, $\hat{u}_i = (x_i - \hat{\mu}_1)/\hat{\sigma}_1$ $(1 \le i \le n)$.

To identify them, we first use the LSEs and order $\tilde{e}_i = y_i - \tilde{\theta}_0 - \tilde{\theta}_1\hat{u}_i - \tilde{\theta}_2\hat{u}_i^2$ $(1 \le i \le n)$ in increasing order of magnitude. The pair $(y_i, \hat{u}_i)$ associated with $\tilde{e}_{(i)}$ are the concomitants $(y_{[i]}, \hat{u}_{[i]})$. Using these concomitants, we compute the MMLEs. We now order $\hat{e}_i = y_i - \hat{\theta}_0 - \hat{\theta}_1\hat{u}_i - \hat{\theta}_2\hat{u}_i^2$ $(1 \le i \le n)$ and identify the new concomitants. We repeat the process one more time. Thus, the MMLEs are computed in two iterations besides computing the LSEs initially. Not more than two iterations are needed for the estimates to stabilize sufficiently. The reason is that only the relative magnitudes (not necessarily the true values) of e_i $(1 \le i \le n)$ are needed to identify the concomitants. See also Islam and Tiku (2004) and Akkaya and Tiku (2008a).

Variances and covariances. The asymptotic variances and covariances of the MMLEs are given by the inverse of Fisher information matrix **I** because they are under regularity conditions equivalent (asymptotically) to the corresponding MLEs as said earlier. Realizing that

$$\int_{-\infty}^{\infty}\left(1 + \frac{1}{k}z^2\right)^{-j}dz = \sqrt{k}\ \Gamma(1/2)\Gamma(j - 1/2)/\Gamma(j) \qquad (2.13)$$

the non-zero elements of $\mathbf{I}^{-1}$ in the present situation are obtained.

The unconditional (all parameters unknown) asymptotic variances of $\hat{\mu}_1$ and $\hat{\sigma}_1$ are

$$V(\hat{\mu}_1) \cong \frac{(p_1+1)(p_1-3/2)}{p_1(p_1-1/2)}\frac{\sigma_1^2}{n}$$

and $$V(\hat{\sigma}_1) \cong \frac{(p_1+1)}{(p_1-1/2)} \frac{\sigma_1^2}{2n} \tag{2.14}$$

Cov($\hat{\mu}_1, \hat{\sigma}_1$) = 0 by symmetry of (2.2). The conditional (μ_1 and σ_1 known) asymptotic variances (and non-zero covariances) of $\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\sigma}$ are

$$V(\hat{\theta}_0) \cong \frac{3}{2}\frac{(p_1-3/2)}{(p_1-1)} \frac{(p_2+1)(p_2-3/2)}{p_2(p_2-1/2)} \frac{\sigma^2}{n} \tag{2.15}$$

$$V(\hat{\theta}_1) \cong \frac{(p_1-1)}{(p_1-2)} \frac{(p_2+1)(p_2-3/2)}{p_2(p_2-1/2)} \frac{\sigma^2}{n} \tag{2.16}$$

$$V(\hat{\theta}_2) = -\text{Cov}(\hat{\theta}_0, \hat{\theta}_2)$$
$$\cong \frac{(p_1-5/2)}{2(p_1-1)} \frac{(p_2+1)(p_2-3/2)}{p_2(p_2-1/2)} \frac{\sigma^2}{n} \tag{2.17}$$

and

$$V(\hat{\sigma}) \cong \frac{(p_2+1)}{(p_2-1/2)} \frac{\sigma^2}{2n} \tag{2.18}$$

the multiple $(p_1 - 1)/(p_1 - 2)$ in (2.16) is an adjustment factor which improves the accuracy of the results, $p_1 > 2$.

Comment. For the conditional variance of $\hat{\theta}_2$ to be positive as it should, p_1 has to be greater than 5/2, i.e., at least the first four moments of X should exist. This leads us to define a "tight design" in the context of a quadratic stochastic regression model as follows.

Tight design. The design ($x_1, x_2, \ldots, x_n$) is called tight if at least the first four moments of X exist.

Remark. For the Fisher information matrix to exist, it is necessary that the design be tight, e.g., $p_1 > 5/2$ in (2.2).

To verify the accuracy of (2.14)-(2.18), we carried out extensive simulations. It must be said that simulation techniques have become very reliable and give accurate results. What we found is that (2.14) give accurate values of the variances of $\hat{\mu}_1$ and $\hat{\sigma}_1$ for all n, and (2.18) gives accurate values of the variance of $\hat{\sigma}$ for $p_i \geq 5$ (i = 1, 2) and $n \geq 20$. For example, we have the values given in Table 1 calculated from the asymptotic equations (2.14) and (2.18). Also given are the corresponding simulated values based on [100,000/n] (integer value) Monte Carlo runs. Simulated means are not given because the bias in the estimators was found to be negligibly small; σ_1 and σ were taken to be equal to 1 without loss of generality.

Tiku and Suresh (1992) give the exact values of the variances of $\hat{\mu}_1$. The corresponding simulated values above are in agreement with their values. For the family (2.2), Tiku and Suresh concluded that $\hat{\mu}_1$ is essentially as efficient as BLUE (best linear unbiased estimator) of μ_1 and $\hat{\sigma}_1$ is more efficient than the BLUE of σ_1. The estimators $\hat{\mu}_1$ and $\hat{\sigma}_1$ attain minimum variance bounds very quickly as n increases (Tiku and Suresh 1992, Senoglu and Tiku 2001).

Equations (2.15)-(2.17) give the asymptotic conditional (μ_1 and σ_1 known) variances of $\hat{\theta}_0, \hat{\theta}_1$ and $\hat{\theta}_2$. They are given in Table 2. Also given are the corresponding simulated variances. The simulated means are not given because the bias in all the estimators turned out to be negligibly small.

Table 1. Unconditional (all parameters unknown) variances* of $\hat{\mu}_1$, $\hat{\sigma}_1$ and $\hat{\sigma}$.

	$\hat{\mu}_1$		$\hat{\sigma}_1$		$\hat{\sigma}$
	Asymp.	Simul.	Asymp .	Simul.	Simul.
n	$p_1 = p_2 = 3$				
20	0.040	0.041	0.040	0.046	0.054
50	0.016	0.017	0.016	0.020	0.027
	$p_1 = p_2 = 5$				
20	0.047	0.047	0.033	0.036	0.038
50	0.018	0.018	0.013	0.014	0.014
	$p_1 = p_2 = 10$				
20	0.049	0.051	0.029	0.029	0.032
50	0.020	0.020	0.012	0.012	0.012

*Asymptotic variances of $\hat{\sigma}$ are the same as those of $\hat{\sigma}_1$ because $p_1 = p_2$.

Table 2. Conditional (μ_1 and σ_1 known) variances of the MMLEs $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$; $\theta_0 = 0$, $\theta_1 = \theta_2 = 1$ and $\sigma = 1$.

	$\hat{\theta}_0$		$\hat{\theta}_1$		$\hat{\theta}_2$	
	Asymp.	Simul.	Asymp.	Simul.	Asymp.	Simul.
n	p_1=5, p_2=3					
50	0.021	0.025	0.021	0.021	0.005	0.011
100	0.011	0.012	0.010	0.010	0.002	0.004
	$p_1 = p_2$=5					
50	0.024	0.029	0.025	0.025	0.006	0.015
100	0.012	0.016	0.012	0.012	0.003	0.012
	p_1= 7, p_2=3					
50	0.022	0.025	0.019	0.020	0.006	0.011
100	0.011	0.011	0.008	0.009	0.003	0.004
	$p_1 = p_2$=10					
50	0.028	0.030	0.022	0.024	0.008	0.014
100	0.014	0.015	0.011	0.011	0.004	0.005

It is seen that (2.15) and (2.17) provide fairly accurate approximations to the conditional variances of $\hat{\theta}_0$ and $\hat{\theta}_1$, for $p_1 \geq 5$ and $n \geq 50$. However, the asymptotic equation (2.17) fails to do so for $\hat{\theta}_2$ unless $p_1 \geq 7$ and $n \geq 100$. The reason for it is that the conditional variance (2.17) involves the population kurtosis μ_4/μ_2^2 of the distribution (2.2). The corresponding simulated variance implicitly involves the sample estimate of the population kurtosis. It is well known that a very large sample size n is required for the sample kurtosis to converge to the population kurtosis (Pearson 1963, Tiku and Akkaya 2004, p.231) and that too if at least the first eight population moments exist and are finite (Kendall and Stuart 1968). We recommend that the variances, of $\hat{\theta}_2$ in particular, be obtained by simulation.

Comment. For unknown location and scale parameters in the distribution of X the derivation of the Fisher infrmation matrix I also involves expected values of certain nonlinear functions of $\hat{u}_i = (x_i - \hat{\mu}_1)/\hat{\sigma}_1$ and is, therefore, complicated. Alternatively, the sample information matrix may be used; see Appendix C. See also Islam and Tiku (2009, Section 5.3).

3. RELATIVE EFFICIENCIES OF THE LSEs

To evaluate the relative efficiencies of the LSEs

RE = 100 (variance of MMLE)/(variance of LSE) (3.1)

Three situations need to be considered:

(a) The distribution of X is long-tailed symmetric and that of e is normal $N(\mu_1, \sigma_1^2)$.

(b) The distribution of X is normal and that of e is long-tailed symmetric.

(c) The distributions of both X and e are long-tailed symmetric.

There are, of course, many other situations, e.g., the distribution of X is short-tailed symmetric (kurtosis less than 3) and that of e is long-tailed symmetric; see the real life example in section 5 which justifies this scenario. All the situations can not be covered in a single paper. Further studies are inevitably needed.

Table 3 covers situation (a) Given are the simulated values of the unconditional (all parameters unknown) variances of the MMLEs. Also given are the relative efficiencies of the corresponding LSEs, σ_1 and σ taken to be 1 without loss of generality. Simulated means are not given because the bias in all the estimators turned out to be negligibly small. Note that for $p = 5$, (2.2) is indistinguishable from Logistic when scale adjusted to have the same variances; both the distributions have kurtosis 4.2.

From the values of Table 3, it can be seen that the MMLEs of μ_1, σ_1, θ_0, θ_1 and θ_2 are more efficient than the corresponding LSEs; notice that the MLEs $\hat{\sigma}_1$ and $\hat{\theta}_2$ are considerably more efficient than the corresponding LSEs. The MMLE $\hat{\sigma}$ is only marginally more efficient than the LSE $\tilde{\sigma}$ because the distribution of e is normal.

Table 4 covers situation (b) Given are the simulated values (similar to those in Table 3). We give values only for $n = 50$, for conciseness.

The effect of non-normality of the error e is now particularly pronounced on the LSE of σ. The LSEs of μ_1 and σ_1 are as efficient as the corresponding MMLEs because the distribution of X is normal.

Table 3. Variances of the MMLEs and relative efficiencies of the LSEs; *X* has LTS distribution (2.2) with $p = p_1$ and *e* is normal N(0, 1).

n		μ_1	σ_1	θ_0	θ_1	θ_2	σ
		$p_1 = 2.8$					
20	Var	0.040	0.049	0.122	0.327	0.306	0.028
	RE	80	72	89	81	67	95
50	Var	0.016	0.021	0.044	0.120	0.118	0.011
	RE	80	71	90	87	65	98
100	Var	0.008	0.010	0.020	0.054	0.050	0.006
	RE	80	51	90	73	36	99
		$p_1 = 3$					
20	Var	0.040	0.046	0.122	0.308	0.267	0.028
	RE	82	81	90	87	81	95
50	Var	0.016	0.022	0.042	0.117	0.132	0.011
	RE	83	71	91	84	56	98
100	Var	0.008	0.010	0.021	0.065	0.054	0.005
	RE	82	49	90	85	27	99
		$p_1 = 5$					
20	Var	0.046	0.036	0.131	0.303	0.196	0.027
	RE	95	95	97	97	98	95
50	Var	0.019	0.015	0.047	0.115	0.074	0.011
	RE	94	88	96	95	89	98
100	Var	0.009	0.007	0.022	0.052	0.033	0.005
	RE	93	86	97	93	88	99

Table 5 covers situation (c). For conciseness, we give values only for $n = 50$. It is seen that the LSEs are much less efficient than for situations (a) and (b).

We also considered the situation when the distribution of *X* is skewed, Generalized Logistic with shape parameter *b* (Tiku and Akkaya 2004, p. 31) and the distribution of $z = e/\sigma$ is (2.2) with shape parameter *p*. For $b = 4$, $p = 3.5$ and $n = 50$, for example, the relative efficiencies of the LSEs of μ_1, σ_1, θ_0, θ_1, θ_2 and σ are 80, 69, 88, 85, 67 and 87 per cent respectively, the bias in the MMLEs and LSEs both being negligible.

4. ROBUSTNESS

In practice, deviations from an assumed distribution are very common. One can not, therefore, feel comfortable with assuming a particular distribution

Table 4. Variances* of the MMLEs and relative efficiencies of the LSEs; *X* is Normal N(0, 1) and *e* has LTS distribution (2.2) with $p = p_2$; $n = 50$

	μ_1	σ_1	θ_0	θ_1	θ_2	σ
	$p_2 = 2$					
Var	0.020	0.010	0.037	0.0104	0.049	0.046
RE	100	98	70	91	88	34
	$p_2=2.5$					
Var	0.020	0.010	0.043	0.0102	0.051	0.026
RE	100	98	85	93	93	66
	$p_2=3.0$					
Var	0.020	0.010	0.049	0.0106	0.050	0.021
RE	100	98	91	95	93	79
	$p_2=5.0$					
Var	0.020	0.010	0.051	0.0106	0.049	0.015
RE	100	98	96	98	95	88

*When *X* has a normal distribution, $\hat{\mu}_1 = \tilde{\mu}_1$ and $\hat{\sigma}_1 = \sqrt{(1-1/n)}\,\tilde{\sigma}_1$.

Table 5. Variances of the MMLEs and relative efficiencies of the LSEs, *X* and *e* both having LTS distributions (2.2); $n = 50$.

	μ_1	σ_1	θ_0	θ_1	θ_2	σ
	$p_1= 2.8$, $p_2=2$					
Var	0.016	0.024	0.034	0.117	0.147	0.049
RE	79	61	66	72	37	35
	$p_1= 2.8$, $p_2= 5$					
Var	0.016	0.021	0.043	0.118	0.110	0.014
RE	80	71	88	86	67	90
	$p_1= 3$, $p_2=2$					
Var	0.017	0.024	0.033	0.116	0.100	0.046
RE	83	76	67	82	70	38
	$p_1=3$, $p_2=5$					
Var	0.017	0.020	0.045	0.126	0.100	0.014
RE	84	77	92	90	77	88
	$p_1= 5$, $p_2=2$					
Var	0.020	0.014	0.037	0.109	0.066	0.033
RE	94	92	74	89	90	46
	$p_1= 5$, $p_2=5$					
Var	0.019	0.013	0.049	0.117	0.064	0.016
RE	94	89	93	96	90	90

and believing it to be exactly correct. That brings the robustness issue in focus. An estimator is called robust if it is fully efficient (or nearly so) for an assumed distribution and maintains high efficiency for plausible alternatives or when a sample contains mild data anomalies (e.g., outliers); see for example Tiku *et al*. (1986, Preface), Tiku and Akkaya (2004, Preface), Senoglu (2005) and Oral (2006). Plausible alternatives are those situations which are difficult to distinguish from an assumed distribution. Strong deviations or strong data anomalies can easily be detected by graph plotting techniques (Hamilton 1992, p. 16) or goodness-of-fit tests (Surucu 2006, Tiku and Akkaya 2004, Chapter 9) and remedial action taken. For example, strong outliers in a sample can be separated from the bulk of observations and studied on their own (Islam and Tiku 2004, p. 2455, Akkaya and Tiku 2008a, Example 2).

Assume, for illustration, that the distributions of X and e are (2.2) with $p_1 = 3.5$ and $p_2 = 3$, respectively. A large number of alternatives were considered, but for conciseness, we report the results only for the following three outlier models:

(1) n-r number of x_i come from (2.2) with $p = p_1 = 3.5$ (variance σ_1^2) and r (we do not know which) come from the same distribution with variance $9\sigma_1^2$; $r = [0.5+0.1n]$ (integer value). (4.1)

(2) n-r number of e_i come from (2.2) with $p = p_1 = 3$ (variance σ^2) and r (we do not know which) come from the same distribution with variance $9\sigma^2$. (4.2)

(3) In the two samples, $n - r$ x_i and e_i come from normal distributions N(0, σ_1^2) and N(0, σ^2) and r come from N(0, $9\sigma_1^2$) and N(0, $9\sigma^2$), respectively. (4.3)

The random numbers generated were divided by appropriate constants to make their variances the same as the assumed distributions. The variances and the relative efficiencies are given in Table 6, bias in all the estimators being negligible is not reported.

It is seen that data anomalies have devastating effect on the LSEs. Clearly, the MMLEs are robust. The results are the same for numerous other alternatives, e.g., mixture and contamination models or when the values of p_1 and p_2 are misspecified; see also Tiku *et al*. (2008, Section 4.1) and Islam and Tiku (2009, Section 5.4). The reason for the inherent robustness of the MMLEs is given in Appendix A.

Table 6. Variances of the MMLEs and the relative efficiencies of the LSEs for outlier models; $\sigma_1 = \sigma = 1$, $n = 50$.

	μ_1	σ_1	θ_0	θ_1	θ_2	σ
	Model (4.1)					
Var	0.013	0.025	0.036	0.099	0.109	0.022
RE	69	58	79	61	48	73
	Model (4.2)					
Var	0.017	0.018	0.036	0.115	0.090	0.033
RE	88	82	72	87	80	50
	Model (4.3)					
Var	0.014	0.025	0.031	0.115	0.114	0.027
RE	69	59	68	58	52	56

5. STS DISTRIBUTIONS FOR THE DESIGN VARIABLE

In some real life situations, the distribution of $U = (X - \mu_1)/\sigma_1$ is short-tailed symmetric (STS). Assume that the distribution of U is one in the family (Akkaya and Tiku 2008b)

$$f(u) = \frac{A}{\sqrt{2\pi}}\{1+\frac{1}{2h}u^2\}^2 e^{-u^2/2}, \ -\infty < u < \infty \qquad (5.1)$$

$h = 2 - d$ and $d < 2$ is a constant. Since

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u^{2j} e^{-u^2/2} du = (2j)!/(2^j\, j\,!) \qquad (5.2)$$

it is easy to evaluate the moments of U. Specifically,

$$A = 1/\{\sum_{j=0}^{2}\binom{2}{j}\left(\frac{1}{2h}\right)^j \frac{(2j)!}{2^j (j)!}\}$$

$$\mu_2 = V(U) = A\sum_{j=0}^{2}\binom{2}{j}\left(\frac{1}{2h}\right)^j \frac{\{2(j+1)\}!}{2^{j+1}(j+1)!} \text{ and}$$

$$\mu_4 = E(U^4) = A\sum_{j=0}^{2}\binom{2}{j}\left(\frac{1}{2h}\right)^j \frac{\{2(j+2)\}!}{2^{j+2}(j+2)!} \qquad (5.3)$$

The values of the variance μ_2 and kurtosis μ_4/μ_2^2 of the distribution of U are given below:

d =	$-\infty$	-0.5	0.0	0.5	1.0	1.5
Variance	1	1.84	2.04	2.33	2.82	3.67
Kurtosis	3	2.56	2.44	2.26	2.03	1.71

It may be noted that no distribution can have kurtosis less than 1 (Pearson and Tiku 1970).

The maximum likelihood equations $\partial \ln L/\partial\mu_1 = 0$ and $\partial \ln L/\partial\sigma_1 = 0$ based on a random sample $x_1, x_2, ..., x_n$ are expressions in terms of the nonlinear functions

$$g(u_i) = u_i/\{1+(1/2h)u_i^2\},\ u_i = (x_i - \mu_1)/\sigma_1 \qquad (5.4)$$

and have no explicit solutions. Solving them by iteration is frought with difficulties. Proceeding along the same lines as in Section 2, the MMLEs are obtained:

$$\hat{\mu}_1 = (1/m)\sum_{i=1}^{n}\beta_i x_{(i)} \quad (m = \sum_{i=1}^{n}\beta_i)$$

and
$$\hat{\sigma}_1 = (-B+\sqrt{B^2+4nC})/2n \qquad (5.5)$$

where

$$B = (2/h)\sum_{i=1}^{n}\alpha_i(x_{(i)} - \hat{\mu}_1) \text{ and } C = \sum_{i=1}^{n}\beta_i(x_{(i)} - \hat{\mu}_1)^2;$$

the values of α_i and β_i are given in Appendix B.

The LSEs of μ_1 and σ_1 are respectively,

$$\tilde{\mu}_1 = \bar{x} \text{ and } \tilde{\sigma}_1 = s_1/\sqrt{\mu_2} \qquad (5.6)$$

The MMLEs are highly efficient. Specifically, $\hat{\mu}_1$ is considerably more efficient than $\bar{x}$ (Akkaya and Tiku 2008b, Table 1).

The estimators $\hat{\mu}_1$ and $\hat{\sigma}_1$ are incorporated in (2.11)-(2.12) and the MMLEs $\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\sigma}$ calculated. Similarly, $\tilde{\mu}_1$ and $\tilde{\sigma}_1$ are incorporated in (2.5) and the LSEs $\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2$ and $\tilde{\sigma}$ calculated.

Remark: Here, the design $(x_1, x_2, ..., x_n)$ is 'tight' because all the moments of X exist and are finite.

Fisher information matrix: The elements of the Fisher information matrix can be obtained exactly along the same lines as in Section 2. The elements of $\mathbf{I}^{-1}$ give the unconditional or conditional variances (and covariance) as in (2.14)-(2.18). In particular,

$$V(\hat{\mu}_1) \cong \sigma_1^2/nD \text{ and } V(\hat{\sigma}_1) \cong \sigma_1^2/nD^* \qquad (5.7)$$

$$D = 1 - (2/h)[(1-a)/(1+2a+3a^2)]$$

and $$D^* = -1 + 3[(1+2a+11a^2)/(1+2a+3a^2)] \qquad (a = 1/2h)$$

The asymptotic equations (5.7) give accurate values for large n; see also Akkaya and Tiku (2008b, p.287). We recommend that the variances of the MMLEs and LSEs be obtained by simulation, at any rate for sample sizes less than $n = 100$.

To have an idea about the relative efficiencies, we give in Table 7 the variances of the MMLEs and the relative efficiencies of the LSEs. For illustration, the distribution of $U = (X - \mu_1)/\sigma_1$ is taken to be (5.1) with $d = 1$ and that of $Z = e/\sigma$ is taken to be (2.2) with $p = p_2 = 3.5$. Simulated means are not given because the bias in all the estimators was found to be negligible.

The MMLEs are jointly enormously more efficient than the LSEs.

Remark: The MMLEs $\hat{\mu}_1$ and $\hat{\sigma}_1$ are robust to plausible deviations from an assumed STS distribution and to inliers in a sample. The reason for that is given in Appendix B. Two inlier models are introduced in Tiku *et al*. (2001, p. 1031) and Akkaya and Tiku (2008b, p. 288). There is room for ideas to define and model inliers in a sample.

Table 7. Variances of the MMLEs and the relative efficiencies of the LSEs; X has STS distribution (5.1) with $d = 1$ and e has LTS distribution (2.2) with $p = p_2 = 3.5$; $\theta_0 = 0$, $\theta_1 = \theta_2 = 1$, $\sigma_1 = \sigma = 1$.

		μ_1	σ_1	θ_0	θ_1	θ_2	σ
n = 20	Var	0.1094	0.0124	0.242	0.453	0.058	0.047
	RE	79	92	83	81	93	90
50	Var	0.0432	0.0045	0.081	0.182	0.021	0.019
	RE	77	90	80	78	91	84
100	Var	0.0216	0.0022	0.041	0.089	0.010	0.009
	RE	76	90	81	77	91	76

Example. Williams (1959) has the following data on Y (Janka hardness of Australian timber) and X (density of the timber). The data is also reproduced in Hand *et al*. (1994, p. 274). Since density is easier to measure, it is desirable to find a model so that hardness can be predicted from the density. The data is reproduced here for ready reference; n = 36.

Density x	Hardness y	Density x	Hardness y	Density x	Hardness y
24.7	484	39.4	1210	53.4	1880
24.8	427	39.9	989	56.0	1980
27.3	413	40.3	1160	56.5	1820
28.4	517	40.6	1010	57.3	2020
28.4	549	40.7	1100	57.6	1980
29.0	648	40.7	1130	59.2	2310
30.3	587	42.9	1270	59.8	1940
32.7	704	45.8	1180	66.0	3260
35.6	979	46.9	1400	67.4	2700
38.5	914	48.2	1760	68.8	2890
38.8	1070	51.5	1710	69.1	2740
39.3	1020	51.5	2010	69.1	3140

Since X and Y are subject to measurement errors, both ought to be treated as random variables. We run the data through EXCEL and plot the ordered estimated residuals $\tilde{e}_{(i)}$ (using LSEs to calculate them) against the quantiles $t_{(i)}$ of a standard normal distribution:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t_{(i)}} e^{-u^2/2}\mathrm{d}u = \frac{i}{n+1} \quad (1 \le i \le 36)$$

The plot (called Q-Q plot) is given in Fig. 1. It is clear that one data point (66.0, 3260) is grossly anomalous. Since such data have undue influence on any statistical analysis, it is common practice to set aside such data points; see for example, Akkaya and Tiku (2008a, p. 414). We now run the remaining n = 35 data points through EXCEL to have Q-Q plots of $x_{(i)}$ and $\tilde{e}_{(i)}$ values, and plot of (y_i, x_i) $(1 \le i \le 35)$ values. They are given in Figs. 2, 3 and 4.

It is clear that $\eta(x)$ in (1.1) can appropriately be modeled by a quadratic, the distribution of x_i can be well represented by a member of the STS family (5.1), and the distribution of e_i can be well represented by a member of the LTS family (2.2); see Hamilton (1992, p. 16) for various aspects of Q-Q plots and how to interpret them. The appropriateness of these distributions can, of course, be verified by employing goodness-of-fit tests (Tiku 1988, Sürücü 2008) but we do not pursue that here. Other such real life examples are given in Hand *et al*. (1994). See also Tiku and Akkaya (2004, Chapter 11).

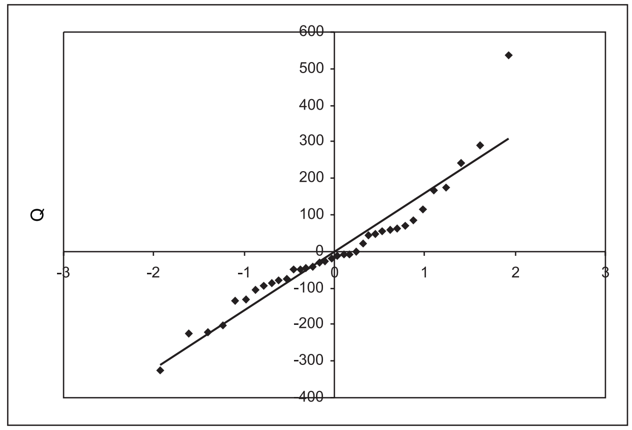

Fig. 1. Q-Q plot of the errors based on 36 observations.

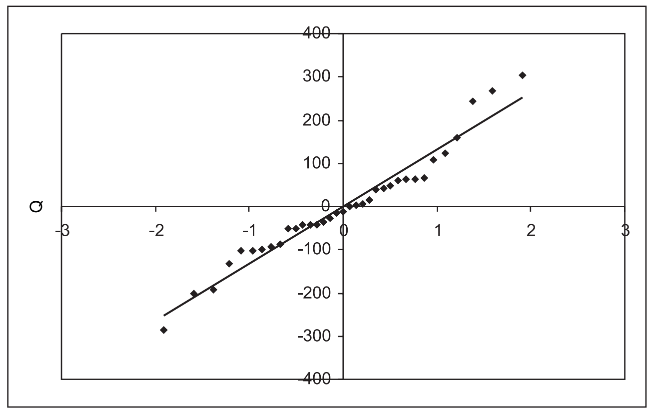

Fig. 2. Q-Q plot of the errors based on 35 observations.

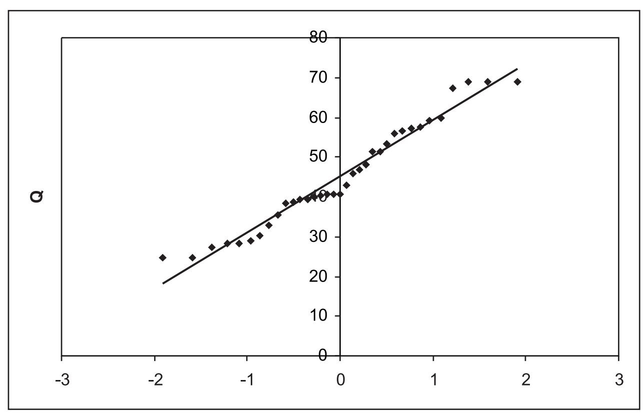

Fig. 3. Q-Q plot of density x_i $(1 \le i \le 35)$

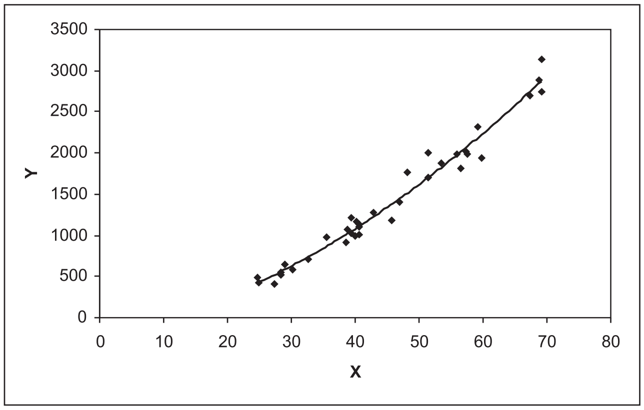

Fig. 4. Q-Q plot of (x_i, y_i) $(1 \le i \le 35)$

To calculate the MMLEs, we need values of d and $p = p_2$ in (5.1) and (2.2), respectively. First, consider determination of d for the distribution of X. This is done by calculating $\hat{\mu}_1$ and $\hat{\sigma}_1$ from (5.5) for several values of d. The chosen value is that value of d which maximizes $(1/n)\ln \hat{L}_x$, where (ignoring $1/\sqrt{2\pi}$ which does not contain d)

$$(1/n)\ln \hat{L}_x = \ln A + \frac{2}{n}\sum_{i=1}^{n} \ln\{1+(1/2h)\hat{u}_i^2\} - \frac{1}{2n}\sum_{i=1}^{n}\hat{u}_i^2$$

$$\hat{u}_i = (x_i - \hat{\mu}_1)/\hat{\sigma}_1; \; h = 2 - d$$

For the x_i $(1 \le i \le 35)$ observations above, we have the following:

$d =$	–0.5	0	0.5	1.0	1.5
$(1/n)\ln \hat{L}_x$	–3.968	–3.961	–3.954	–3.958	–4.032

The chosen value of d is, therefore, 0.5. The corresponding estimates $\hat{\mu}_1$ and $\hat{\sigma}_1$ are the desired MMLEs of μ_1 and σ_1, respectively. They are now incorporated in (2.2) to find the value of $p = p_2$ which maximizes $(1/n)\ln \hat{L}_{y|x} = (1/n)\ \ln \hat{L}_e$; θ_0, θ_1, θ_2 and σ are replaced by their MMLEs. Thus, we have the following values:

$p = p_2 =$	2.5	3	3.5	4	4.5	5
$(1/n)\ln \hat{L}_e$	–6.225	–6.225	–6.223	–6.224	–6.226	–6.227

The chosen value is $p = p_2 = 3.5$. The corresponding estimates are the desired MMLEs. Given below are the estimates and their parametric bootstrap standard errors:

	μ_1	σ_1	θ_0	θ_1	θ_2	σ
MMLE	45.91	8.69	1384.48	471.52	30.54	131.04
SE	2.04	0.79	114.90	46.65	10.31	20.76
LSE	45.15	8.72	1347.50	469.96	31.24	129.30
SE	2.25	0.82	125.51	49.01	11.00	22.03

The estimates are in league but the MMLEs have smaller standard errors and, therefore, greater precision. See also Tiku *et al*. (2008, pp. 1734-1740).

ACKNOWLEDGEMENT

We are very grateful to the Editors for inviting the senior author to contribute an article to this Special Issue. Our thanks are due to the referee for valuable comments.

REFERENCES

Akkaya, A.D. and Tiku, M.L. (2008a). Robust estimation in multiple linear regression model with Non-Gaussian noise. *Automatica,* **44**, 407-417.

Akkaya, A.D. and Tiku, M.L. (2008b). Short-tailed distributions and inliers. *Test*, **17**, 282-296.

Bhattacharyya, G.K. (1985). The asymptotics of maximum likelihood and related estimators based on type II censored data. *J. Amer. Statist. Assoc*., **80**, 398-404.

Dedieu, H. and Ogorzalek, M.J. (1997). Identifiability and identification of chaotic systems based on adaptive synchronization. IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, **44**, 948-962.

Elveback, L.R., Guillier, C.L. and Keating, F.R. (1970). Health, normality and the ghost of Gauss. *J. Amer. Medical Assoc*., **211**, 69-75.

Hamilton, L.C. (1992). *Regression with Graphics*. Brooks/Cole Publishing Company, California.

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E. (1994). *Small Data Sets*. Chapman and Hall, New York.

Hutchinson, T.P. and Lai, C.D. (1990*). Continuous Bivariate Distributions Emphasising Applications*. Rumsby Scientific, Adelaide.

Islam, M.Q. and Tiku, M.L. (2004). Multiple linear regression model under non- normality. *Comm. Statist.-Theory Methods*, **33**, 2443-2467.

Islam, M.Q. and Tiku, M.L. (2009). Multiple linear regression model with stochastic design variables. *J. Appl. Statist.*, 1-21. DOI: 10.1080/02664760902939612.

Kantar, Y.M. and Senoglu, B. (2008). A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter. *Comput. Geosci.*, **34**, 1900-1909.

Kendall, M.G. and Stuart, A. (1968). *The Advanced Theory of Statistics*, Vol. 3. Charles Griffin, London.

Oral, E. (2006). Binary regression with stochastic covariates. *Comm. Statist.- Theory Methods*, **35**, 1429-1447.

Pearson, E.S. (1931). The analysis of variance in case of non-normal variation. *Biometrika*, **23**, 114-133.

Pearson, E.S. (1963). Some problems arising in approximating to probability distributions using moments. *Biometrika*, **50**, 95-112.

Pearson, E.S. and Tiku, M.L. (1970). Some notes on the relationship between the distributions of central and non-central F. *Biometrika*, **57**, 175-179.

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.

Qumsiyeh, S.B. (2007). Non-normal bivariate distributions: Estimation and hypothesis testing. Ph. D thesis, Middle East Technical University, Ankara.

Roy, J. and Tiku, M.L. (1962). A Laguerre series approximation to the sampling distribution of the variance. *Sankhya*, **24**, 181-184.

Sazak, H.S., Tiku, M.L. and Islam, M.Q. (2006). Regression analysis with a stochastic design variable. *Intern. Statist. Rev*., **74**, 77-88.

Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations*. Marcel Dekker, New York.

Senoglu, B. (2005). Robust 2^k factorial design with Weibull error distributions. *J. Appl. Statist*., **32**, 1051-1066.

Senoglu, B. and Tiku, M.L. (2001). Analysis of variance in experimental design with non- normal error distributions. *Comm. Statist.- Theory Methods*, **30**, 1335-1352.

Spjϕtvoll, E. and Aastveit, A.H. (1980). Comparison of robust estimators on some data from field experiments. *Scand. J. Statist.*, **7**, 1-13.

Surucu, B. (2006). Goodness-of-fit tests for multivariate distributions. *Comm. Statist.-Theory Methods*, **35**, 1319-1331.

Surucu, B. (2008). A power comparison and simulation study of goodness-of-fit tests. *Compu. Math. Appl.*, **56**, 1617-1625.

Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, **54**, 155-165.

Tiku, M.L. (1968). Estimating the parameters of normal and logistic distributions from censored samples. *Austr. J. Statist.*, **10**, 64-74.

Tiku, M.L. (1988). Order statistics in goodness-of-fit tests. *Comm. Statist.-Theory Methods*, **17**, 2369-2387.

Tiku, M.L. (1989). Modified likelihood estimation. In: Kotz,S. and Johnson, N.L.(Eds), *Encyclo. Statist. Sci.*, Supplement Volume, 24.

Tiku, M.L. and Suresh, R.P. (1992). A new method of estimation for location and scale parameters. *J. Statist. Plann. Inf.*, **30**, 281-292.

Tiku, M.L. and Vaughan, D.C. (1997). Logistic and nonlogistic density functions in binary regression with nonstochastic covariates. *Biom. J.*, **39**, 883-898.

Tiku, M.L. and Akkaya, A.D. (2004). *Robust Estimation and Hypothesis Testing*. New Age International Publishers, New Delhli. Also, Oscar Publications.

Tiku, M.L., Tan, W.Y. and Balakrishnan, N. (1986). *Robust Inference*. Marcel Dekker, New York.

Tiku, M.L., Islam, M.Q. and Selcuk, S. (2001). Nonnormal regression, II. Symmetric distributions. *Comm. Statist.-Theory Methods*, **30**, 1021-1045.

Tiku, M.L., Islam, M.Q. and Sazak, H.S. (2008). Estimation is bivariate nonnormal distributions with stochastic variance functions. *Compu. Statist. Data Analy.*, **52**, 1728-1745.

Tiku, M.L., Islam, M.Q. and Qumsiyeh, S.B. (2009). Mahalanobis distance under non-normality. *Statistics*., 1-16, DOI:10.1080/02331880903043223.

Tiku, M.L. and Akkaya, A.D. (2010). Estimation in multifactor polynomial regression under non-normality. *Pak. J. Statist*., **26**(1), 49-68 (Silver Jubilee invited paper).

Vaughan, D.C. (2002). The generalized secant hyperbolic distribution and its properties. *Comm. Statist.-Theory Methods*, **31**, 219-238.

Vaughan, D.C. and Tiku, M.L. (2000). Estimation and hypothesis testing for a non-normal bivariate distribution. *J. Math. Comp. Modelling*, **32**, 53-67.

Voss, H.U., Timmer, J. and Kurths, J. (2004). Nonlinear dynamic system identification from uncertain and indirect measurements. *Intern. J. Birfurcation Choas*, **14**, 1905-1933.

Williams, E.J. (1959). *Regression Analysis*. Wiley, New York.

APPENDIX A

Let (x_1, x_2, ..., x_n) be a random sample of size n from the scaled Student's t distribution (2.2), $z_i = (x_i - \mu)/\sigma$ $(1 \le i \le n)$. The maximum likelihod equations $\partial \ln L/\partial\mu = 0$ and $\partial \ln L/\partial\sigma = 0$ are expressions in terms of the intractable functions

$$g(z_i) = z_i/\{1+(1/k)z_i^2\} \quad (1 \le i \le n) \qquad \text{(A.1)}$$

and have, therefore, no explicit solutions. Modified maximum likelihood equations are obtained by first expressing the maximum likelihood equations in terms of the order statistics of the sample. This is accomplished simply by replacing z_i by $z_{(i)} = (x_{(i)} - \mu)/\sigma$. Then, $g(z_{(i)})$ are replaced by linear approximations,

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)} \quad (1 \le i \le n) \qquad \text{(A.2)}$$

α_i and β_i are usually obtained from the first two terms of a Taylor series expansion of $g(z_{(i)})$ about the population quantiles $t_{(i)}$. The values of $t_{(i)}$ are obtained from

$$\frac{\Gamma(p)}{\sqrt{k}\,\Gamma(1/2)\Gamma(p-1/2)}\int_{-\infty}^{t_{(i)}} (1+\frac{z^2}{k})^{-p}\,dz = \frac{i}{n+1}$$

$$(1 \le i \le n) \qquad \text{(A.3)}$$

An IMSL subroutine in FORTRAN is available to determine $t_{(i)}$ from (A.3) . The coefficients α_i and β_i are given by

$$\alpha_i = (2/k)t_{(i)}^3/\{1+(1/k)t_{(i)}^2\}^2$$

and $$\beta_i = \{1-(1/k)t_{(i)}^2\}/\{1+(1/k)t_{(i)}^2\}^2 \qquad \text{(A.4)}$$

The linear approximations (A.2) are incorporated in the maximum likelihood equations. The solutions of the resulting equations are the MMLEs (modified maximum likelihood estimators).

For $\hat{\sigma}$ to be real and positive, β_i $(1 \le i \le n)$ have to be positive. These coefficients have umbrella ordering, that is, they increase until the middle value and then decrease in a symmetric fashion. Therefore, if $\beta_1 > 0$ then all the β_i are positive. For small p and large n, however, β_1 can be negative. To rectify this situation if $\beta_1 < 0$, α_i is replaced by α_i^* and β_i is replaced by β_i^* (Islam and Tiku 2004, p.2451)

$$\alpha_i^* = (1/k)t_{(i)}^3/\{1+(1/k)t_{(i)}^2\}^2$$

and $$\beta_i^* = 1/\{1+(1/k)t_{(i)}^2\}^2 \qquad \text{(A.5)}$$

This operation does not alter the asymptotic properties of the MMLEs because (asymptotically)

$$g(z_{(i)}) \cong \alpha_i^* + \beta_i^* z_{(i)} \quad (1 \le i \le n) \qquad \text{(A.6)}$$

Note that $g(z_{(i)})$ and α_i and β_i, and α_i^* and β_i^*, are all bounded.

Remark: Because of the umbrella ordering of the β_i $(1 \le i \le n)$ coefficients, the extreme x_i-observations and the extreme errors $\hat{e}_i$ automatically receive small weights in calculating the MMLEs. That depletes the influence of long-tails and data anomalies, e.g., outliers. As a result, MMLEs are robust to deviations from an assumed long-tailed distribution and to outliers in a sample.

APPENDIX B

For the MMLEs in (5.5), the coefficients are the following; see also Tiku and Akkaya (2010) who consider multifactor polynomial regression with nonstochastic design variables.

For $d \le 0$,

$$\alpha_i = (1/h)t_{(i)}^3/\{1+(1/2h)t_{(i)}^2\}^2 \text{ and } \beta_i = 1-(2/h)\gamma_i \qquad \text{(B.1)}$$

$$\gamma_i = \{1-(1/2h)t_{(i)}^2\}/\{1+(1/2h)t_{(i)}^2\}^2$$

For $d > 0$, α_i and β_i are replaced by α_i^* and γ_i^*, respectively:

$$\alpha_i^* = \{(1/h)t_{(i)}^3 + (1-h/2)t_{(i)}\}/\{1+(1/2h)t_{(i)}^2\}^2$$

and $$\beta_i^* = 1-(2/h)\gamma_i^* \qquad \text{(B.2)}$$

$$\gamma_i^* = \{(h/2)-(1/2h)t_{(i)}^2\}/\{1+(1/2h)t_{(i)}^2\}^2$$

It may be noted that the coefficients β_i and β_i^* are all positive, the former for $d \le 0$. The values of $t_{(i)}$ in (B.1) and (B.2) are obtained from the equation

$$\frac{A}{\sqrt{2\pi}}\int_{-\infty}^{t_{(i)}} \{1+\frac{1}{2h}u^2\}^2 e^{-u^2/2}du = \frac{i}{n+1} (1 \le i \le n)$$

$$\text{(B.3)}$$

A simple algorithm written by our colleague Dr. M.Q. Islam to calculate $t_{(i)}$ from (B.3) is available with the authors.

Remark: The coefficients β_i and β_i^* in (B.1) and (B.2) have inverted-umbrella ordering, i.e., they decrease until the middle value and then increase in a symmetric fashion. Thus, the middle x_i-observations receive small weights. Thus, their influence is automatically depleted. As a result, MMLEs are robust to inliers. See also Tiku and Akkaya (2010).

APPENDIX C

The sample information matrix is –1 times the second derivatives of lnL evaluated at $\mu_1 = \hat{\mu}_1$, $\sigma_1 = \hat{\sigma}_1$, $\theta_0 = \hat{\theta}_0$, etc. Inverse of this matrix gives asymptotic variances and covariances. They provide accurate approximations to finite sample size variances and covariances for $p_1 + p_2 > 5$ and $n \geq 50$. For example, we have the following values based on [100,000/n] Monte Carlo runs. They are (1) values calculated from the sample information matrix and (2) values obtained by simulation, σ_1 and σ taken to be 1 without loss of generality.

Values of the variances; $\theta_1 = 0$, $\theta_1 = \theta_2 = 1$.

$n = 50$	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}$
	p_1=3, p_2=2					
(1)	0.017	0.018	0.034	0.118	0.093	0.032
(2)	0.017	0.021	0.033	0.116	0.100	0.046
	p_1=5, p_2=2					
(1)	0.019	0.013	0.037	0.110	0.064	0.028
(2)	0.020	0.014	0.037	0.109	0.066	0.033
	p_1=2.8, p_2=5					
(1)	0.017	0.019	0.044	0.129	0.102	0.015
(2)	0.016	0.021	0.043	0.118	0.110	0.014
	p_1=3, p_2=5					
(1)	0.017	0.018	0.045	0.128	0.099	0.015
(2)	0.017	0.020	0.045	0.126	0.100	0.014
	p_1=5, p_2=5					
(1)	0.019	0.013	0.048	0.117	0.068	0.015
(2)	0.019	0.013	0.049	0.117	0.064	0.016

See also Islam and Tiku (2009, Section 5.3) who have similar results for a linear stochastic model. The results also imply that MMLEs are highly efficient, at any rate for large *n*.