ADAPTIVE ESTIMATION AND HYPOTHESIS TESTING METHODS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYÇA DÖNMEZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPY
IN
STATISTICS

MARCH 2010

Approval of the thesis:

ADAPTIVE ESTIMATION AND HYPOTHESIS TESTING METHODS

submitted by **AYÇA DÖNMEZ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. H. Öztaş Ayhan _________________
Head of Department, **Statistics**

Prof. Dr. Moti Lal Tiku
Supervisor, **Statistics Dept., METU** _________________

Examining Committee Members:

Prof. Dr. Ayşen Akkaya _________________
Statistics Dept., METU

Prof. Dr. Moti Lal Tiku _________________
Statistics Dept., METU

Assoc. Prof. Dr. M. Qamarul Islam _________________
Economics Dept., Çankaya University

Assistant Prof. Dr. Hakan Savaş Sazak _________________
Statistics Dept., Ege University

Assistant Prof. Dr. Barış Sürücü _________________
Statistics Dept., METU

Date: 03.03.2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: **Ayça DÖNMEZ**

Signature:

ABSTRACT

ADAPTIVE ESTIMATION AND HYPOTHESIS TESTING METHODS

Dönmez, Ayça
Ph. D., Department of Statistics
Supervisor: Prof. Dr. Moti Lal Tiku

March 2010, 153 pages

For statistical estimation of population parameters, Fisher's maximum likelihood estimators (MLEs) are commonly used. They are consistent, unbiased and efficient, at any rate for large *n*. In most situations, however, MLEs are elusive because of computational difficulties. To alleviate these difficulties, Tiku's modified maximum likelihood estimators (MMLEs) are used. They are explicit functions of sample observations and easy to compute. They are asymptotically equivalent to MLEs and, for small *n*, are equally efficient. Moreover, MLEs and MMLEs are numerically very close to one another. For calculating MLEs and MMLEs, the functional form of the underlying distribution has to be known. For machine data processing, however, such is not the case. Instead, what is reasonable to assume for machine data processing is that the underlying distribution is a member of a broad class of distributions. Huber assumed that the underlying distribution is long-tailed symmetric and developed the so called M-estimators. It is very desirable for an estimator to be robust and have bounded influence function. M-estimators, however, implicitly censor certain sample observations which most practitioners do not appreciate. Tiku and Surucu suggested a modification to Tiku's MMLEs. The new MMLEs are robust and have bounded influence functions. In fact, these new estimators are overall more efficient than M-estimators for long-tailed symmetric distributions. In this thesis, we have proposed a new modification to MMLEs. The resulting estimators are robust and have bounded influence functions. We have also shown that they can be used not only for long-tailed symmetric distributions but for skew distributions as well. We have used the proposed modification in the context of experimental design and linear regression. We have shown that the resulting estimators and the hypothesis testing procedures based on them are indeed superior to earlier such estimators and tests.

Key words: robustness, modified maximum likelihood (MML) estimators, non-normality.

ÖZ

UYARLAMALI TAHMİN VE HİPOTEZ TESTİ YÖNTEMLERİ

Dönmez, Ayça
Doktora, İstatistik Bölümü
Tez Yöneticisi: Prof. Dr. Moti Lal Tiku

Mart 2010, 153 sayfa

Populasyon parametrelerinin istatistiksel tahmininde yaygın olarak Fisher en çok olabilirlik tahminleyicileri (MLEs) kullanılmaktadır. MLEs tutarlı, yansız ve etkinlerdir. Ancak birçok durumda hesaplamaya dayalı zorluklardan ötürü elde edilemezler. Bu zorlukları aşmak için Tiku uyarlanmış en çok olabilirlik tahminleyicileri (MMLEs) kullanılabilir. MMLEs gözlemlerin açık fonksiyonları olarak ifade edildiklerinden kolay hesaplanırlar. MMLEs asimptotik olarak MLEs'e eşit olmalarının yanında küçük örneklemlerde de eşdeğer etkinliktedirler. Ayrıca MLEs ve MMLEs sayısal olarak birbirlerine çok yakındırlar. Herikisinin de hesaplanabilmesi için dağılımın fonksiyonel formunun biliniyor olması gerekir. Ancak bu makine veri işlemesinde mümkün olmayabilir. Onun yerine esas dağılımın geniş bir dağılım ailesinin üyesi olduğunu varsaymak daha makuldür. Huber esas dağılımın uzun kuyruklu simetrik dağılım olduğunu varsaymış ve M-tahminleyicilerini geliştirmiştir. Bir tahminleyici için sağlam oluşu ve sınırlandırılmış bir etki fonksiyonuna sahip olması oldukça istenen özelliklerdir. Fakat M-tahminleyicilerinin örneklemdeki gözlemleri sansürlüyor oluşu uygulama yapanlar için sorun teşkil edebilir. Tiku ve Surucu MMLEs için bir değişiklik önermiştir. Yeni MMLEs sağlam olmalarının yanında sınırlandırılmış etki fonksiyonlarına da sahiplerdir. Bu yeni tahminleyicilerin uzun kuyruklu simetrik dağılımlar için M-tahminleyicilerine kıyasla toplamda daha etkin oldukları gözlenmiştir. Bu tez çalışmasında MMLEs için yeni bir değişiklik önerisinde bulunduk. Elde edilen tahminleyiciler sağlamdırlar ve sınırlandırılmış etki fonksiyonuna sahiplerdir. Bunun yanında yeni tahminleyicilerin yalnızca uzun kuyruklu simetrik dağılımlarda değil çarpık dağılımlarda da kullanılabileceğini gösterdik. Deneysel tasarım ve doğrusal regresyon alanlarında önerilen değişimi kullandık. Elde edilen tahminleyicilerin ve bunlar üzerine kurulmuş hipotez testi yöntemlerinin önceki benzerlerinden daha üstün olduğunu gördük.

Anahtar kelimeler: sağlamlık (robustness), uyarlanmış en çok olabilirlik (MML) tahmin edicileri, normal olmayan dağılımlar.

To the memory of Professor Moti Lal TIKU…

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

GENERAL METHODS OF ESTIMATION

Two methods of parameter estimation have numerous applications, namely, the method of least squares and the method of maximum likelihood. For estimating location and scale parameters, for example, they proceed as follows:

1.1 Least Squares

Let X be a random variable with mean $E(X)=\mu$ and variance $V(X)=\sigma^2$. A random sample

$$x_1, x_2, ..., x_n \tag{1.1.1}$$

is available. The objective is to estimate μ and σ (or σ^2). The least squares methodology postulates the model

$$x_i = \mu + e_i \quad (1 \le i \le n) \tag{1.1.2}$$

where e_i is a random error with mean $E(e_i)=0$ and variance $V(e_i)=\sigma^2$; e_i $(1 \le i \le n)$ are independent of one another. The least squares estimator of μ is obtained by minimizing the error sum of squares

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \ . \tag{1.1.3}$$

That gives

$$\tilde{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{(the sample mean).} \qquad (1.1.4)$$

The least squares estimator of σ^2 is defined as

$$\tilde{\sigma}^2 = \min \sum_{i=1}^{n} e_i^2 /(n-1) = \sum_{i=1}^{n} (x_i - \bar{x})^2 /(n-1) = s^2 \text{ (the sample variance);}$$

$(n-1)$ is called the *df* (degree of freedom) of s^2 since $(n-1)s^2 = \sigma^2$ constitutes an $(n-1)$ dimesional hyper sphere. It is easy to show that

$$E(\bar{x}) = \mu \ , \ V(\bar{x}) = \sigma^2 / n \text{ and } E(s^2) = \sigma^2 .$$

The LSE (least squares estimator) of σ is $\tilde{\sigma} = s$. For large n, $E(s) \cong \sigma$. However, $E(s^2) = \sigma^2$ for all n.

***Remark*:** The only assumption in using the method of least squares is that the mean and variance of e_i (i.e., the mean and variance of the underlying distribution) are finite. In that sense, the method is general.

If $E(e_i) = a\sigma$ (a being a constant), i.e. $E(X) = \mu + a\sigma$, then the LSE of μ is obtained by minimizing

$$\sum_{i=1}^{n} (x_i - \mu - a\sigma)^2 .$$

That gives,

$$\tilde{\mu} = \bar{x} - a\sigma \quad (\sigma \text{ to be replaced by } \tilde{\sigma}), \qquad (1.1.5)$$

and the LSE of the variance σ^2 is

$$\tilde{\sigma}^2 = \min \sum_{i=1}^{n} (x_i - \tilde{\mu} - a\sigma)^2 /(n-1) = s^2 .$$

Thus, $\tilde{\sigma} = s$ (as before).

Comment**:** While $V(\bar{x}) = \sigma^2 / n$ irrespective of the underlying distribution, $V(s)$ depends on the distribution. In fact Roy and Tiku (1962) showed that

$$V(s) \cong \frac{\sigma^2}{2n}\left(1 + \frac{1}{2}\lambda_4\right), \quad \lambda_4 = \beta_2 - 3\,, \tag{1.1.6}$$

$\beta_2 = \mu_4 / \mu_2^2$ being the kurtosis of the underlying distribution; see also Tan and Wong (1977). For a normal distribution, $\lambda_4 = 0$. Clearly, the variance of s increases with β_2. We will show later that even the sample mean $\bar{x}$ is inefficient when β_2 deviates from 3 by appreciable amounts. It may be noted that no distribution can have kurtosis $\beta_2 < 1$ (Pearson and Tiku, 1970).

1.1.1 Correlated Errors

In (1.1.2) we assume that the errors e_i $(1 \le i \le n)$ are independent of one another (hence, uncorrelated). That is not necessary. We now assume that

$$E(e_i) = 0 \text{ and } Cov(e_1, e_2, ..., e_n) = \Omega\sigma^2,$$

Ω being an nxn matrix with constant coefficients. In this situation, the LSE of μ is obtained by minimizing the generalized dispersion

$$e'\Omega^{-1}e,$$

$e' = (e_1, e_2, ..., e_n)$. That gives,

$$\tilde{\mu} = x'\Omega^{-1}x / 1'\Omega^{-1}1.$$

The method is very flexible indeed. If $Cov(e_i, e_j) = 0$,

$$\tilde{\mu} = \sum_{i=1}^{n}(1/\sigma_i^2)x_i \Big/ \sum_{i=1}^{n}(1/\sigma_i^2)$$

a weighted sum with weights inversely proportional to the variances. This result is well known.

1.2 Maximum Likelihood

Asssume that X has a location-scale distribution $(1/\sigma)f((x-\mu)/\sigma)$, i.e., the distribution of $Z=(X-\mu)/\sigma$ is free of μ and σ. The likelihood function (joint probability density function) of a random sample is

$$L=\left(\frac{1}{\sigma}\right)^n\prod_{i=1}^{n}f\left(\frac{x_i-\mu}{\sigma}\right). \tag{1.2.1}$$

The MLEs (maximum likelihood estimators) of μ and σ are those values of μ and σ which maximize L or $\ln L$, there being one-to-one correspondance between the two functions since L is always positive. To maximize $\ln L$ for μ and σ, we solve the equations

$$\partial\ln L/\partial\mu=0 \text{ and } \partial\ln L/\partial\sigma=0. \tag{1.2.2}$$

Under very general regularity conditions, essentially existence of first two derivatives of $\ln L$ and the third derivative being bounded, the variance-covariance matrix of the MLEs $\hat{\hat{\mu}}$ and $\hat{\hat{\sigma}}$ is for large n,

$$Cov(\hat{\hat{\mu}},\hat{\hat{\sigma}})=\begin{bmatrix}-E(\partial^2\ln L/\partial\mu^2) & -E(\partial^2\ln L/\partial\mu\partial\sigma)\\ -E(\partial^2\ln L/\partial\mu\partial\sigma) & -E(\partial^2\ln L/\partial\sigma^2)\end{bmatrix}^{-1}. \tag{1.2.3}$$

Given the functional form f of the underlying distribution, the MLEs have all the Fisherian optimal properties at least asymptotically, i.e., unbiasedness, consistency and efficiency. However, MLEs are often computationally intractable. To illustrate this assume that the underlying distribution is Logistic

$$f(x)=\frac{1}{\sigma}\exp\{-(x-\mu)/\sigma\}\Big/\left[1+\exp\{-(x-\mu)/\sigma\}\right]^2, \quad -\infty<x<\infty. \tag{1.2.4}$$

Here, $E(X)=\mu$ and $V(X)=3.2898\sigma^2$. The LSEs of μ and σ are

$\tilde{\mu}=\bar{x}$ and $\tilde{\sigma}=s/\sqrt{3.2898}$ (since s^2 estimates the population variance).

The maximum likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{2}{\sigma}\sum_{i=1}^{n} g(z_i) = 0 \quad \text{and} \tag{1.2.5}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_i - \frac{2}{\sigma}\sum_{i=1}^{n} z_i\, g(z_i) = 0 \tag{1.2.6}$$

where

$$z_i = (x_i - \mu)/\sigma \quad \text{and} \quad g(z_i) = \exp(-z_i)/\left[1 + \exp(-z_i)\right]. \tag{1.2.7}$$

The equations (1.2.5) and (1.2.6) have no explicit solutions. They have to be solved by iterations. Software is available to do that, e.g., Press, et al. (1992) and Venables and Ripley (2002).

In general, however, one can encounter difficulties in solving maximum likelihood equations by iterations: (i) the iterations may converge to wrong values, (ii) the iterations may not converge at all, and (iii) the iterations might locate local rather than global maxima due to multiple roots. This is illustrated adequately by Puthenpura and Sinha (1986) and Qumsiyeh (2007). Qumsiyeh (2007, pp. 6-10) had a random sample of size $n = 100$ from the logistic distribution (1.2.4). To calculate the MLEs of μ and σ, she used Powell hybrid algorithm. This algorithm is a variation of Newton's method and uses a finite-difference approximation to the Jacobian and takes precautions to avoid large steps (More et al., 1980). She started the iteration process with $\mu = -1.00$ and $\sigma = 5.00$. The process converged at 23rd iteration and gave the MLEs as

$$\hat{\hat{\mu}} = -0.14 \text{ and } \hat{\hat{\sigma}} = 1.05.$$

The true values being $\mu = 1$ and $\sigma = 1$, these estimates are quite reasonable. Qumsiyeh (2007, p.10) introduced 10% outlier in the sample. Started the iteration process again with the results in Table 1.1. The process never converged. Also, yielded a negative estimate of σ. That is disconcerting.

Table 1.1: The starting values of the iteration process

Iteration no.	$\partial \ln L / \partial \mu$	$\partial \ln L / \partial \sigma$	$\hat{\mu}$	$\hat{\sigma}$
1	10.16	-52.14	-1.00	5.00
2	10.16	-52.14	-1.00	5.00
3	10.16	-52.16	-1.00	5.00
4	$-\infty$	∞	0.88	-0.08

To alleviate the computational difficulties with maximum likelihood, Tiku (1967, 1968a,b, 1989) and Tiku and Suresh (1992) developed the methodology of modified maximum likelihood as follows.

1.3 Modified Maximum Likelihood

Assume that the pdf (probability density function) of X is $(1/\sigma) f((x-\mu)/\sigma)$. The likelihood function is

$$L = \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^{n} f\left(\frac{x_i - \mu}{\sigma}\right).$$

The maximum likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{\sigma} \sum_{i=1}^{n} g(z_i) = 0, \quad g(z) = f'(z)/f(z) \quad \text{and}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^{n} z_i \, g(z_i) = 0 .$$

The method of modified maximum likelihood estimation is implemented in three steps: (i) the equations are expressed in terms of ordered variates $z_{(i)} = (x_{(i)} - \mu)/\sigma \quad (1 \le i \le n)$, (ii) the functions $g\left(z_{(i)}\right)$ are replaced by linear approximations $g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)}$ so that the differences between the two tend to zero as n becomes large, and (iii) the resulting equations called maximum likelihood equations are solved. They are typically of the form

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{M}{\sigma}\left(K + D\sigma - \mu\right) = 0 \quad \text{and} \qquad (1.3.1)$$

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma^3}\left[M(K-\mu)(K+D\sigma-\mu)-\left(n\sigma^2 - B\sigma - C\right)\right] = 0\,. \quad (1.3.2)$$

Therefore, the MMLEs (modified maximum likelihood estimators) are of the form

$$\hat{\mu} = K + D\sigma \quad (\sigma \text{ to be replaced by } \hat{\sigma})$$

and (1.3.3)

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

the divisor $2n$ may be replaced by $2\sqrt{n(n-1)}$ as a bias-correction. If the distribution is symmetric, $D=0$ and $\hat{\mu} = K$ (is free of σ) which is a very interesting result.

Notice the form of $\partial \ln L^* / \partial \mu$ in (1.3.1). Since

$$\lim_{n \to 0} \frac{1}{n}\left\{\frac{\partial \ln L^*}{\partial \mu} - \frac{\partial \ln L}{\partial \mu}\right\} = 0\,,$$

it follows that the MMLE $\hat{\mu}$ is conditionally (σ known) the *MVB* (minimum variance bound) estimator of μ, for large n. Using a similar argument and the form of $\partial \ln L / \partial \mu = 0$ above, it follows that $\hat{\sigma}$ is conditionally (μ known) the *MVB* estimator of σ for large n.

EXAMPLE

Consider the situation when a random sample $x_1, x_2, ..., x_n$ comes from the Generalized Logistic distribution (b>0)

$$f(x) = \frac{b}{\sigma} \frac{\exp\{-(x-\mu)/\sigma\}}{\left[1+\exp\{-(x-\mu)/\sigma\}\right]^{b+1}}, \qquad -\infty < x < \infty\,. \qquad (1.3.4)$$

For $b<1$, (1.3.4) is negatively skewed. For $b=1$, it is the well known Logistic Distribution and is symmetric. For $b>1$, it is positively skewed. Here, the maximum likelihood equations expressed in terms of the ordered variates $z_{(i)} = (x_{(i)} - \mu)/\sigma$ (obtained simply by replacing $z_i = (x_i - \mu)/\sigma$ by $z_{(i)}$) are

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} g\left(z_{(i)}\right) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)}\, g\left(z_{(i)}\right) = 0$$

where $g(z) = \exp(-z)/\{1+\exp(-z)\}$. The equations clearly have no explicit solutions. In fact, for small or large values of b, the iterations have convergence problems.

To obtain modified likelihood equations, we consider the linear approximations (Tiku and Akkaya, 2004)

$$g\left(z_{(i)}\right) \cong \alpha_i - \beta_i z_{(i)} \quad (1 \le i \le n). \tag{1.3.5}$$

The coefficients α_i and β_i are obtained from Taylor series expansion of $g\left(z_{(i)}\right)$ about the population quantiles $t_{(i)}$ determined by

$$\int_{-\infty}^{t_{(i)}} f(z)dz = \frac{i}{n+1} \quad (1 \le i \le n). \tag{1.3.6}$$

That gives,

$$t_{(i)} = -\ln\left(q_i^{-1/b} - 1\right), \; q_i = i/(n+1), \tag{1.3.7}$$

$$\alpha_i = \frac{1+\exp(t)+t\exp(t)}{\{1+\exp(t)\}^2} \quad \text{and} \quad \beta_i = \frac{\exp(t)}{\{1+\exp(t)\}^2} > 0, \; t = t_{(i)}.$$

Realize that as n tends to infinity, $z_{(i)}$ converges to $t_{(i)}$. Hence, the differences $g\left(z_{(i)}\right) - \left(\alpha_i - \beta_i z_{(i)}\right)$ $(1 \le i \le n)$ converge to zero as n tends to infinity. Consequently, MLEs and MMLEs are asymptotically equivalent.

The modified likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n}\left(\alpha_i - \beta_i z_{(i)}\right) = \frac{M}{\sigma^2}(K + D\sigma - \mu) = 0 \quad \text{and} \qquad (1.3.8)$$

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)}\left(\alpha_i - \beta_i z_{(i)}\right) = \frac{1}{\sigma^3}\left[M(K-\mu)(K+D\sigma-\mu) - \left(n\sigma^2 - B\sigma - C\right)\right] = 0 \quad , \qquad (1.3.9)$$

where $M = (b+1)m$, $m = \sum_{i=1}^{n}\beta_i$, $K = \left(\sum_{i=1}^{n}\beta_i x_{(i)}\right)/m$,

$\Delta_i = (b+1)^{-1} - \alpha_i$, $D = \sum_{i=1}^{n}\Delta_i / m$, $B = (b+1)\sum_{i=1}^{n}\Delta_i (x_{(i)} - K)$ and

$C = (b+1)\left(\sum_{i=1}^{n}\beta_i x_{(i)}^2 - mK^2\right)$.

The solutions of (1.3.8) and (1.3.9) are the following MMLEs:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \left\{B + \sqrt{B^2 + 4nC}\right\} / 2n; \qquad (1.3.10)$$

n may be replaced by $\sqrt{n(n-1)}$ as a bias correction. For $b = 1$, $D = 0$.

For reasons given earlier, $\hat{\mu}$ and $\hat{\sigma}$ are conditionally *MVB* estimators for large n.

***Comment*:** MMLEs are known to be asymptotically equivalent to MLEs whenever the latter exist. Therefore, they are asymptotically unbiased and their variance-covariance matrix is given by I^{-1}, where I is the Fisher information matrix. A rigorous proof of this is given in Bhattacharyya (1985) for censored samples and in Vaughan and Tiku (2000) for complete samples (under very general regularity conditions). A huge literature is available and compares MMLEs with MLEs and concludes: (i) the two are numerically very close to one another, and (ii) MMLEs are as efficient as MLEs. See, for example, Schneider (1986), Tan and Tabatabai (1988), Tiku et al. (1986), Vaughan (2002), Tiku and Akkaya (2004), and Kantar and Senoglu (2008). In fact, the modified maximum likelihood method works very well when the methods of maximum likelihood or least squares fail. This is ilustrated in Puthenpura and Sinha (1986). Qumsiyeh (2007, p.14) had a sample of size $n = 100$ from Generalized Logistic distribution with $b = 0.5$. The true values are $\mu = 0$ and $\sigma = 1$. She calculated the MLEs, MMLEs and the LSEs of μ and σ with the following results:

Table 1.2: The MLEs, MMLEs and the LSEs of μ and σ of Generalized Logistic distribution with $b = 0.5$; $n = 100$.

	μ	σ
MLE	-2.891	-1.161
MMLE	-0.086	1.074
LSE	-0.200	0.995

She concluded that the iterations with maximum likelihood equations can converge to wrong values and the method of least squares can give highly biased results. She stated that MMLEs are fine in all respects. This agrees with the results of Puthenpura and Sinha (1986).

1.3.1 Robustness

LSEs are not distribution based but MMLEs are (i.e., in calculating them a particular distribution is assumed). However, the reality is that the underlying distribution is hardly ever known exactly. It is also naive to believe that nothing is known about the underlying distribution. There are graphical techniques and goodness-of-fit tests available to identify the underlying distribution (Surucu, 2008; Tiku and Akkaya 2004, Chapter 9). They may not locate the exact distribution but can ascertain distributions in close proximity. On the other hand, a sample might contain outliers perhaps due to some misadventure in experimentation. Strong outliers can readily be identified by using computer graphics and outlier detection procedures (Tiku and Akkaya 2004, Chapter 9). What is difficult indeed is to identify mild outliers in a sample. Situations which cannot be readily distinguished from an assumed distribution are called plausible alternatives (Tiku et al., 1986, Preface; Tiku and Akkaya, 2004, Preface). An estimator is called robust if it is fully efficient (at any rate for large sample size n) for an assumed distribution and maintains high efficiency for plausible alternatives. A fully efficient estimator is unbiased and has minimum variance. There is a huge literature investigating the efficiency and robustness properties of LSEs and MMLEs; see, for example, Tiku and Akkaya (2004), Islam and Tiku (2004), Senoglu (2005), Oral (2006), and Tiku et al. (2008). The conclusion is that LSEs are efficient only for normal and near-normal distributions. They are not, however, robust to deviations from an assumed distribution. The MMLEs have excellent efficiency and robustness properties although they are somewhat more difficult to compute than the LSEs. To repeat, MMLEs are model based, i.e., in computing them a particular distribution is assumed. However, they are remarkably efficient and robust to plausible deviations from an assumed distribution, and to mild data anomalies.

1.3.2 Machine Data Processing

It is argued (Hampel et al. 1986, Preface) that in machine data processing there is no opportunity to explore the nature of the underlying distribution but one can rightfully assume that it is, for example, long-tailed symmetric. They define a robust estimator to be one which has: (i) high efficiency (whatever the distribution is as long as it is long-tailed symmetric) or the sample has outliers (irrespective of whether they are mild or strong), and (ii) has high breakdown, i.e., if a number of observations are shifted to infinity in either direction, the estimator continues to assume finite values and, hence, finite mean (which should preferably be its population value, i.e. the estimator be unbiased), and finite variance (preferably not much bigger than *MVB*).

Consider, for example, the sample mean $\bar{x}$. It is efficient only for estimating the mean of a normal or near-normal distribution. If an observation is shifted to infinity, it will assume an infinite value. Thus, $\bar{x}$ is not robust. They show that the following M-estimators of the population mean are robust for long-tailed symmetric distributions.

1.4 M-Estimators

Let $x_1, x_2, ..., x_n$ be a random sample from a long-tailed symmetric distribution of the type $(1/\sigma)f((x-\mu)/\sigma)$. The log-likelihood function for estimating μ is

$$\ln L = \sum_{i=1}^{n} \ln f(z_i) = \sum_{i=1}^{n} \rho(z_i), \quad z_i = (x_i - \mu)/\sigma . \tag{1.4.1}$$

If the functional form of $f(z)$ is known, the MLE of μ (for given σ) is obtained from the equation

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{\sigma}\sum_{i=1}^{n} \rho'(z_i) = \frac{1}{\sigma}\sum_{i=1}^{n} \psi(z_i) = 0 . \tag{1.4.2}$$

For normal and double exponential

$$f(z) \propto e^{-z^2/2} \quad \text{and} \quad f(z) \propto e^{|z|} \qquad (-\infty < z < \infty),$$

for example, the $\rho(z)$ and $\psi(z)$ functions are given by

$$\rho(z) = \frac{1}{2}\ln(2\pi) + \frac{1}{2}z^2, \quad \psi(z) = z$$

and

$$\rho(z) = \ln 2 + |z|, \quad \psi(z) = \text{sgn}(z) \quad (z \neq 0),$$

respectively. Writing $w_i = w_i(z) = \psi_i(z_i)/z_i$, (1.4.2) can be expressed as

$$\sum_{i=1}^{n} w_i(x_i - \mu) = 0 \tag{1.4.3}$$

which gives

$$\mu = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i .$$

Given σ and $\psi(z)$, one may solve (1.4.3) iteratively.

However, σ and $\psi(z)$ are not known in practice. Huber (1964) proposed $\psi(z)$ as

$$\psi(z)=\begin{cases} z & if \quad |z|\le c \\ c\,\mathrm{sgn}(z) & if \quad |z|>c \end{cases}, \tag{1.4.4}$$

which corresponds to a normal distribution in the middle and double exponential in the tails. The popular choice of the c values are 1.345, 1.5 and 2 which correspond to roughly 10%, 5% and 2.5% truncation of the tails of a normal distribution $N(0,1)$. Birch and Myers (1982) proposed that σ be replaced by

$$mad = median|\,x_i - median(x_i)\,| \,/\, 0.6745.$$

For a normal distribution, mad is an asymptotically unbiased estimator of σ. In the so called Princeton study (Andrews, et al. 1972), sixty five $\psi(z)$ functions were examined. The following three functions were found to be particularly useful. Incidentally, they are descending functions, i.e., they decrease with increasing $|z|$:

1. The wave function (Andrews et al., 1972, Andrews, 1974)

$$\psi(z)=\begin{cases} \sin(z) & if \quad |z|\le \pi \\ 0 & if \quad |z|>\pi. \end{cases} \tag{1.4.5}$$

2. The bisquare function (Beaton and Tukey, 1974)

$$\psi(z)=\begin{cases} z\,(1-z^2)^2 & if \quad |z|\le 1 \\ 0 & if \quad |z|>1\,. \end{cases} \tag{1.4.6}$$

3. The piecewise linear function (Hampel, 1974)

$$\psi(z)=\begin{cases} |z| & if \quad 0\le |z|<a \\ a & if \quad a\le |z|<b \\ \dfrac{c-|z|}{c-b} & if \quad b\le |z|<c \\ 0 & if \quad c\le |z|\,. \end{cases} \tag{1.4.7}$$

This results in different estimators for different values of a, b and c, in fact, a large number of them. In an extensive numerical study, Gross (1976) examined 25 representative estimators (out of 65 studied in Priceton study) of location parameter μ and scale parameter σ and concluded that the three descending functions $\psi(z)$ above with certain specified adjusting constants *h* (named W24, BS82 and H22) were generally the most efficient. We reproduce the estimators W24, BS82 and H22 from Gross (1976) as follows; see also Tiku (1980):

$$T_0 = median\{x_i\} \text{ and } S_0 = median\{|x_i - T_0|\} \quad (i = 1, 2, ..., n).$$

W24

$$\hat{\mu}_w = T_0 + (h S_0) \tan^{-1}\left[\frac{\sum_{i=1}^{n} \sin(z_i)}{\sum_{i=1}^{n} \cos(z_i)}\right] \text{ and } \hat{\sigma}_w = (h S_0)\left[n \frac{\sum_{i=1}^{n} \sin^2(z_i)}{\left(\sum_{i=1}^{n} \cos(z_i)\right)^2}\right]^{1/2}, \quad (1.4.8)$$

$h = 2.4$ and summations include only those i for which

$$|z_i| < \pi, \quad z_i = (x_i - T_0)/hS_0 .$$

BS82

$$\hat{\mu}_B = T_0 + (h S_0)\frac{\sum_{i=1}^{n} \psi(z_i)}{\sum_{i=1}^{n} \psi'(z_i)} \text{ and } \hat{\sigma}_B = (h S_0)\left[n \frac{\sum_{i=1}^{n} \psi^2(z_i)}{\left(\sum_{i=1}^{n} \psi'(z_i)\right)^2}\right]^{1/2}, \quad (1.4.9)$$

$h = 8.2$, $z_i = (x_i - T_0)/hS_0$ and $\psi(z)$ is the Beaton-Tukey function given in (1.4.6) and $\psi'(z)$ is its derivative.

H22

$$\hat{\mu}_H = T_0 + S_0 \frac{\sum_{i=1}^{n} \psi(z_i)}{\sum_{i=1}^{n} \psi'(z_i)} \text{ and } \hat{\sigma}_H = S_0 \left[n \frac{\sum_{i=1}^{n} \psi^2(z_i)}{\left(\sum_{i=1}^{n} \psi'(z_i)\right)^2}\right]^{1/2}, \quad (1.4.10)$$

where $\psi(z)$ is the piecewise linear function given in (1.4.7) and $\psi'(z)$ is its derivative; $a = 2.25$, $b = 3.75$ and $c = 15$.

Extensive simulations have been caried out to explore the efficiency properties of M-estimators (1.4.9); see, for example, the Princeton study, Tiku (1980) and Dunnett (1982).

The conclusions are that

(a) for long-tailed symmetric distributions, the M-estimators of μ are unbiased and have very good efficiency, but

(b) the M-estimators of σ can have substantial downward bias, even asymptotically.

1.4.1 Influence Function

Hampel (1974) introduced the concept of '*influence function*', equivalently '*breakdown*', to ascertain the robustness of an estimator. Observations in a sample are shifted in either direction to infinity and its effect on the estimator ascertained. If an estimator assumes infinite values (and, consequently, its mean and variance are infinite), the estimator is non-robust. Apparently, the sample mean is non-robust but the sample median is robust. However, the sample median is not efficient other than for extreme distributions like Cauchy. Nevertheless, one has to aim for high efficiency when only a small portion of observations are shifted to infinity to ascertain robustness. In that regard, the M-estimators (1.4.9) are robust. Incidentally, empirical influence function is a graphical plot of the values an estimator assumes when an observation(s) in a random sample is shifted (in either direction) to infinity (Hampel et al. 1986, p.93). A smooth (bounded) plot establishes robustness of an estimator. M-estimators have bounded influence functions and so have the following estimators based on censored samples.

1.5 Trimmed Mean and Variance

Let

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \tag{1.5.1}$$

be the order statistics of a random sample of size *n*. Deleting *r* smallest and *r* largest observations, Tukey defined the following estimators:

$$\hat{\mu}_T = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_{(i)}$$

and (1.5.2)

$$\hat{\sigma}_T = \left[\frac{1}{n-2r-1}\left\{\sum_{i=r+1}^{n-r}\left(x_{(i)} - \hat{\mu}_T\right)^2 + r\left[\left(x_{(r+1)} - \hat{\mu}_T\right)^2 + \left(x_{(n-r)} - \hat{\mu}_T\right)^2\right]\right\}\right]^{1/2}.$$

The estimators $\hat{\mu}_T$ and $\hat{\sigma}_T$ have bounded influence functions as long as not more than *r* observations are shifted in either direction to infinity.

Taking $r = [0.5 + 0.1n]$ (integer value), in comprehensive simulation studies, Tiku (1980) and Dunnett (1982) showed that $\hat{\mu}_T$ and $\hat{\sigma}_T$ are as efficient as M-estimators for long-tailed symmetric distributions with finite variances and in situations when the sample contains mild outliers. For extreme situations, e.g. the underlying distribution is Cauchy or the sample has a considerable number of strong outliers, taking $r = [0.5 + 0.3n]$ renders $\hat{\mu}_T$ and $\hat{\sigma}_T$ as efficient as M-estimators. The estimator $\hat{\sigma}_T$, like the M-estimators of σ, can have substantial downward bias.

1.6 Estimators Based on Censored Normal Samples

Tiku et al. (1986, p. 22-23) have an interesting result. They show that if the tails of a long-tailed symmetric distribution are truncated, the resulting truncated distribution has $\beta_2 = \mu_4 / \mu_2^2$ closer to 3 (kurtosis of a normal distribution) although its variance is understandably less than that of the untruncated distribution. Consider, for example, the family of long-tailed symmetric distributions

$$f(x) = \frac{1}{\sigma\sqrt{k}} \frac{\Gamma(p)}{\Gamma(1/2)\Gamma(p-1/2)} \left[1 + \frac{(x-\mu)^2}{k\sigma^2}\right]^{-p}, \qquad -\infty < x < \infty; \quad (1.6.1)$$

$k = 2p - 3$, $p \geq 2$. Note that $E(X) = \mu$ and $V(X) = \sigma^2$. Now, consider the truncated distribution

$$f_T(z) \propto \left(1 + \frac{z^2}{k}\right)^{-p}, \quad z = (x-\mu)/\sigma; \quad -z_0 < z < z_0. \qquad (1.6.2)$$

Tiku et al. (1986, p.23) give the following values of the standard deviation $\sqrt{\mu_2}$ and the kurtosis β_2^* of (1.6.2):

Table 1.3: The standard deviation and the kurtosis of the truncated distribution.

$p = 5/2$			$p = 7/2$			$p = \infty$ (normal)		
z_0	$\sqrt{\mu_2}$	β_2^*	z_0	$\sqrt{\mu_2}$	β_2^*	z_0	$\sqrt{\mu_2}$	β_2^*
∞	1	∞	∞	1	6	∞	1	3
2.650	0.837	3.32	2.566	0.886	3.04	2.326	0.935	2.54
1.508	0.657	2.46	1.586	0.716	2.36	1.645	0.789	2.19

They noticed that nearly 10% truncation of either tail brings the distribution close to normal so far as its kurtosis is concerned. Since truncation of tails is equivalent to censoring the extreme observations in a sample, they considered the censored sample

$$x_{(r+1)}, x_{(r+2)}, \ldots, x_{(n-r)}. \qquad (1.6.3)$$

Assuming that the underlying distribution is normal $N(\mu, \sigma^2)$, they determined the efficiency and robustness properties of the MMLEs based on (1.6.3). They are

$$\hat{\mu} = \left\{ \sum_{i=r+1}^{n-r} x_{(i)} + r\beta \left(x_{(r+1)} + x_{(n-r)} \right) \right\} \Big/ m, \quad m = (n-2r) + 2r\beta$$

and (1.6.4)

$$\hat{\sigma} = \left\{ B + \sqrt{B^2 + 4AC} \right\} \Big/ 2\sqrt{A(A-1)}, \quad A = n - 2r;$$

$$B = r\alpha \left(x_{(n-r)} - x_{(r+1)} \right) \text{ and } C = \sum_{i=r+1}^{n-r} \left(x_{(i)} - \hat{\mu} \right)^2 + r\beta \left\{ \left(x_{(r+1)} - \hat{\mu} \right)^2 + \left(x_{(n-r)} - \hat{\mu} \right)^2 \right\}.$$

The coefficients α and β in (1.6.4) are calculated from the following equations:

$$\beta = -\frac{f(t)}{q} \left[t - \frac{f(t)}{q} \right] \quad \text{and} \quad \alpha = \frac{f(t)}{q} - \beta t; \qquad (1.6.5)$$

$q = r/n$, $f(t) = (2\pi)^{-1/2} \exp(-t^2/2)$ and t is the value such that

$$(2\pi)^{-1/2} \int_{-\infty}^{t_{(i)}} \exp(-z^2/2) dz = q.$$

The estimators $\hat{\mu}$ and $\hat{\sigma}$ above were initially obtained by Tiku (1967). With $r = [0.5 + 0.1n]$, Tiku (1980) and Dunnett (1982) showed that for long-tailed symmetric distributions with a finite variance, $\hat{\mu}$ and $\hat{\sigma}$ above are as efficient as M-estimators. Like M-estimators, however, $\hat{\sigma}$ can have considerable downward bias. For extreme distributions with infinite variance (e.g. Cauchy) $\hat{\mu}$ and $\hat{\sigma}$ with $r = [0.5 + 0.3n]$ in (1.6.4) are competitive with M-estimators.

1.6.1 Random Censoring

It is clear from equations (1.4.9) that a 'random' number of extreme observations in a sample are censored to calculate the M-estimators. A similar mechanism can be implemented to calculate MMLEs (1.6.4) as follows (Tiku, 1980).

In (1.6.4) – (1.6.5), replace $q = r/n$ by $q^* = r^*/n$ where

$$r^* = \begin{cases} 0 & if \quad k^*/n = 0 \\ \left[0.5 + 0.1n\right] & if \quad 0 < k^*/n \le 0.1 \\ \left[0.5 + 0.3n\right] & if \quad k^*/n > 0.1 \end{cases} \qquad (1.6.1.1)$$

k^* is the number of values of

$$|z_i| = |x_i - median\{x_i\}| / \tilde{\sigma}_0 \quad (1 \le i \le n)$$

which exceed 3.0, $\tilde{\sigma}_0 = 1.483|x_i - median\{x_i\}|$. Denote the resulting MMLEs by $\hat{\mu}^*$ and $\hat{\sigma}^*$, i.e., the estimators (1.6.4) with $q = r/n$ replaced by $q^* = r^*/n$. It may be noted that $\tilde{\sigma}_0$ is asymptotically unbiased if the distribution is normal.

As for $(\hat{\mu}, \hat{\sigma})$, Tiku (1980) carried out extensive simulations to study the efficiencies of $(\hat{\mu}^*, \hat{\sigma}^*)$. He showed that even for situations of extreme type, $\hat{\mu}^*$ is overall more efficient than the M-estimators of μ, and $\hat{\sigma}^*$ has overall less bias than the M-estimators of σ. Consider, for example, the following models:

Outlier models: (1) $(n - r_1)N(0, \sigma^2)$ & $r_1 N(0, 9\sigma^2)$

(2) $(n - r_1)N(0, \sigma^2)$ & $r_1 N(0, 100\sigma^2)$, $r_1 = \left[0.5 + 0.2n\right]$

(3) Student's *t* with 2 *df*, (4) Cauchy, (5) Slash (Normal/Cauchy).

We reproduce his results in Table 1.4; $\sigma = 1$ without loss of generality, $n = 20$.

Table 1.4: The results of MML, W24 and H22 estimators under different distribution models.

		Model				
	Estimator	(1)	(2)	(3)	(4)	(5)
Mean	$\hat{\sigma}_W$	1.33	1.39	1.37	1.80	2.53
	$\hat{\sigma}_H$	1.32	1.38	1.36	1.79	2.51
	$\hat{\sigma}^*$	1.15	1.26	1.15	1.41	2.09
Variance	$\hat{\mu}_W$	0.098	0.105	0.106	0.185	0.362
	$\hat{\mu}_H$	0.097	0.104	0.105	0.183	0.359
	$\hat{\mu}^*$	0.092	0.113	0.095	0.167	0.329

Because of the randomness of q^*, however, it is difficult to derive the distribution of $\sqrt{n}\,\hat{\mu}^* / \hat{\sigma}^*$. Like the M-estimators, this restricts the use of $\hat{\mu}^*$ and $\hat{\sigma}^*$ primarily to estimation of location and scale parameters.

The question is whether the estimators (1.3.10) can be formulated such that they can be used for all sorts of say long-tailed symmetric distributions or when a sample has mild to strong outliers, and be at least as efficient as the M-estimators. The purpose of this thesis is to develop such estimators. They are particularly useful for machine data processing when a statistician has no opportunity to investigate the nature of the underlying distribution. Admitedly, such situations are very common in practice.

1.7 Hypothesis Testing

So far we have talked about parameter estimation, particularly of the location and scale parameters. Another important problem is that of hypothesis testing. Given a random sample $x_1, x_2, ..., x_n$ one wants to test, for example, the null hypothesis $H_0 : \mu = 0$. The statistic that is used most often is Student's t:

$$t = \sqrt{n}\bar{x} / s . \qquad (1.7.1)$$

If the underlying distribution is normal $N(\mu, \sigma^2)$, the null distribution of t is Student's t with $\nu = n-1$ degrees of freedom. To test H_0 against $H_1 : \mu > 0$, if the computed value of t is greater than $t_{1-\alpha}(\nu)$, H_0 is rejected at α percent significance level. The non-null distribution of t is noncentral t with $\nu = n-1$ degrees of freedom and non-centrality parameter $\lambda^2 = n(\mu/\sigma)^2$. The t-test is UMP (uniformly most powerful).

Several authors investigated the effect of non-normality on the Type I error and power of the t-test. Most prominent among them are Gayen (1949) and Srivastava (1958). Both these authors assumed that the underlying distribution is Edgeworth series, $z = (x - \mu)/\sigma$:

$$f(z) = \left\{1 + \frac{1}{6}\lambda_3 H_3(z) + \frac{1}{24}\lambda_4 H_4(z) + \frac{1}{72}\lambda_4^2 H_6(z)\right\}\phi(z) \qquad (1.7.2)$$

where $\lambda_3 = \mu_3 / \mu_2^{3/2}$ and $\lambda_4 = (\mu_4 / \mu_2^2) - 3$ are the standardized third and fourth cumulants, $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ $(-\infty < z < \infty)$, and $H_r(z)$ is the r^{th} Hermite polynomial. They obtained the exact null distribution of t and its power function. For various combinations of values of (λ_3, λ_4), they calculated the exact values of the Type I error and power. Although Gayen's and Srivastava's work had great deal of mathematical charm but it failed to be conclusive, the reason being the limitation of the Edgeworth series; Barton and Dennis (1952) showed that the Edgeworth series is a genuine probability density function only for a small range of values of (λ_3, λ_4). Therefore, Gayen's and Srivastava's work had validity only for near-normal distributions. For small departures from normality, both Type I error and power of the t-test are not affected in any substantial way.

Tiku (1964; 1971a,b) introduced a different approach which is not restricted like Gayen's and Srivastava's. He developed the sampling distributions of s^2 (sample variance), t^2 and ANOVA F statistics in terms of Laguerre polynomials and Gamma density functions. Thus, he calculated the Type I error and power for a much broader range of non-normal distributions than was possible with Gayen's and Srivastava's approach. He concluded that non-normality

(a) does not affect the Type I error to a remarkable degree, but

(b) has a substantial downward effect on the power.

What is, therefore, needed are test procedures that are robust (both in terms of Type I error and power) to departures from normality and to data anomalies, e.g., outliers. In this thesis, we develop such procedures by using modified maximum likelihood estimators and variants of them.

CHAPTER 2

REVISED MODIFIED MAXIMUM LIKELIHOOD ESTIMATION

In machine data processing, there is no opportunity to ascertain the nature of the underlying distribution but one may be justified in assuming that it is a long-tailed symmetric distribution (Hampel et al., 1986, Preface). The MMLEs (1.3.3) are model based, i.e., in calculating them, a particular distribution is assumed. One may, for example, assume that the underlying distribution is one of the long-tailed symmetric family

$$f(x) = \frac{1}{\sigma\sqrt{k}} \frac{1}{\beta(1/2, p-1/2)} \left[1 + \frac{(x-\mu)^2}{k\sigma^2}\right]^{-p}, \qquad -\infty < x < \infty; \tag{2.1}$$

$k = 2p - 3$, $p \geq 2$ and $\beta(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. It may be noted that $E(X) = \mu$ and $V(X) = \sigma^2$. For a given $p\ (\geq 2)$, the MMLEs of μ and σ are

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i\, x_{(i)} \Big/ m \quad \left(m = \sum_{i=1}^{n} \beta_i\right) \text{ and } \hat{\sigma} = \left\{B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{n(n-1)} \tag{2.2}$$

where

$$B = (2p/k)\sum_{i=1}^{n} \alpha_i (x_{(i)} - \hat{\mu}) \text{ and } C = (2p/k)\sum_{i=1}^{n} \beta_i \left(x_{(i)} - \hat{\mu}\right)^2 .$$

The coefficients α_i and β_i are given by (Islam and Tiku, 2004, Equation (4.5))

$$\alpha_i = (1/k)t_{(i)}^3 \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2 \text{ and } \beta_i = 1 \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2; \tag{2.3}$$

$t_{(i)} = E\left(z_{(i)}\right)$ $(1 \leq i \leq n)$. These coefficients are essentially obtained from Taylor series expansions. Tables of $t_{(i)}$ and the variances of $z_{(i)}$ and the covariances of

$\left(z_{(i)}, z_{(j)}\right)$ are given in Tiku and Kumra (1981) for $p = 2\,(.5)\,10$ and $n \le 20$. For $n \ge 10$, however, $t_{(i)}$ may be calculated from (as a close approximation to the true values)

$$\frac{1}{\sqrt{k}\,\beta(1/2, p-1/2)} \int_{-\infty}^{t_{(i)}} \left(1+\frac{z^2}{k}\right)^{-p} dz = \frac{i}{n+1} \quad (1 \le i \le n)\,. \qquad (2.4)$$

An IMSL subroutine is available to evaluate (2.4). For a given p, $\hat{\mu}$ and $\hat{\sigma}$ are known to be as efficient as the MLEs whenever the latter are authentic. However, the MLEs are not readily available since they are analytically and computationally too involved as said earlier.

With p chosen to be 3 or 3.5, $\hat{\mu}$ and $\hat{\sigma}$ are remarkably robust to long-tailed symmetric distributions having finite variances and to situations when a sample contains mild outliers or other mild data anomalies (Tiku and Akkaya, 2004; Oral, 2006; Tiku et al., 2008). For machine data processing, however, long-tailed symmetric distributions need to be inclusive of extreme distributions like Cauchy and also to situations when a sample contains strong outliers and other strong data anomalies (Hampel et al., 1986). In this thesis, we develop such estimators.

What we show first, following Tiku and Surucu (2009), is that when α_i and β_i in (2.3) are estimated from a given sample, the resulting estimators $\hat{\mu}_x$ and $\hat{\sigma}_x$ have very high breakdown and are overall more efficient than the M-estimators mentioned earlier. To estimate the coefficients α_i and β_i $(1 \le i \le n)$, as in Huber (1981), let

$$T_0 = median\{x_i\} \text{ and } S_0 = 1.483 median\{|\,x_i - T_0\,|\} \quad (1 \le i \le n)\,. \qquad (2.5)$$

Realize that T_0 is an unbiased estimator of μ (for symmetric distributions) and S_0 is asymptotically an unbiased estimator of σ (for a normal distribution). Obviously, $t_{(i)}$ in (2.3) can be estimated by $\tilde{t}_{(i)} = \left(x_{(i)} - T_0\right)/S_0$. We also write $\tilde{t}_i = \left(x_i - T_0\right)/S_0$. Since complete sums are invariant to ordering,

$$\hat{\mu}_x = \sum_{i=1}^{n} w_i\, x_i \Big/ w \quad \left(w = \sum_{i=1}^{n} w_i \right) \text{ and } \hat{\sigma}_x = \left\{ B + \sqrt{\left(B^2 + 4nC\right)} \right\} \Big/ 2\sqrt{n(n-1)} \,. \quad (2.6)$$

Here

$$B \cong (2p/k)\sum_{i=1}^{n} \nu_i (x_i - \hat{\mu}_x) \quad \left(\nu_i = (w_i / k)\tilde{t}_i \right),$$

and (2.7)

$$C = (2p/k)\sum_{i=1}^{n} w_i (x_i - \hat{\mu}_x)^2 \;;$$

$$w_i = 1 \Bigg/ \left\{ 1 + \frac{1}{k}\left(\frac{x_i - T_0}{S_0} \right)^2 \right\}^2 \quad (1 \le i \le n) \,. \qquad (2.8)$$

***Note*:** It may be noted that $\hat{\mu}_x$ is a nonlinear function and so is $\hat{\sigma}_x$.

***Remark*:** As with M-estimators, the only assumption for using $\hat{\mu}_x$ and $\hat{\sigma}_x$ is that the underlying distribution is long-tailed symmetric. Their asymptotic properties are given in Appendix A. Realize that the coefficient ν_i in the above expression for B has been obtained from α_i by equating $\tilde{t}_i^{\,2}$ to its expected value which is 1 (almost) for $p = 16.5$ as chosen in the next section. This is necessary to have a bounded influence function.

2.1 Choice of *k*

As pointed out by Tiku and Surucu (2009), if we choose k very large, w_i $(1 \le i \le n)$ essentially reduce to 1 and $\hat{\mu}_x$ reduces to the sample mean $\bar{x}$ which, although fully efficient for a normal distribution, has zero breakdown and is not efficient (and robust) for long-tailed symmetric distributions or even to moderate outliers in a sample. On the other hand, if we choose k small, $\hat{\mu}_x$ and $\hat{\sigma}_x$ are enormously inefficient for normal and near-normal distributions. The choice $k = 30\,(p = 16.5)$ turns out to be a good compromise. We denote the corresponding MMLEs by MML30. The emprical influence functions of $\hat{\mu}_x$ and

$\hat{\sigma}_x$ are given in Figure 2.1 and Figure 2.2, respectively. They illustrate high breakdown of MML30; see also Tiku and Surucu (2009). This was to be expected since the associated terms in the expressions for w, $\hat{\mu}_x$, B and C tend to 0, respectively, as the i^{th} observation x_i is shifted (in either direction) to infinity.

Tiku and Surucu (2009) estimator of μ is exactly the same as in (2.6) but their estimator of σ is

$$\sqrt{1.13\sum_{i=1}^{n} w_i (x_i - \hat{\mu}_x)^2 \Big/ w}\,. \tag{2.1.1}$$

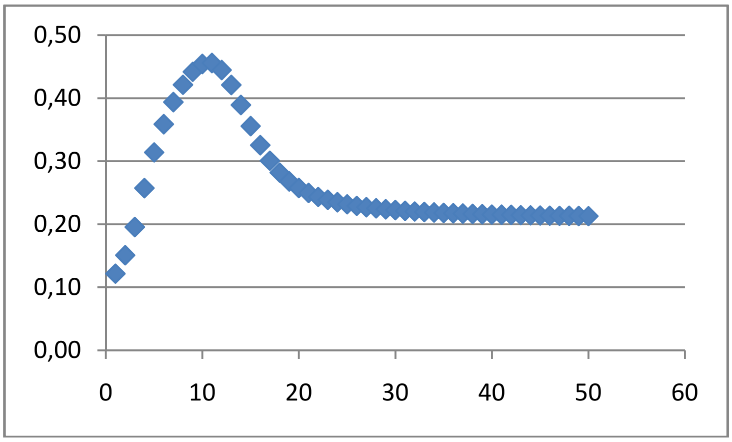

Figure 2.1: Empirical influence function of $\hat{\mu}_x$ for $p = 3.5$ ($n = 10$).

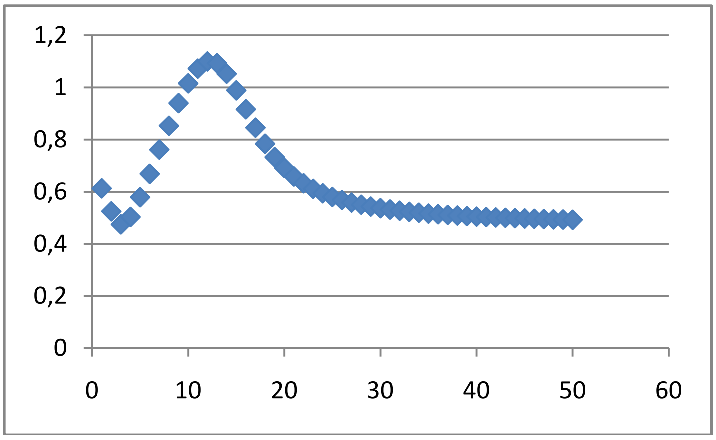

Figure 2.2: Empirical influence function of $\hat{\sigma}_x$ for $p = 3.5$ ($n = 10$).

2.2 Efficiency and Robustness

To evaluate the efficiency and robustness of MML30 given in (2.6)-(2.8), we consider a normal distribution and a very broad range of long-tailed symmetric distributions and samples containing data anomalies as follows, see also Tiku and Surucu (2009), μ taken to be zero without any loss of generality:

(1) Normal $N(0,\sigma^2)$

The family (2.1) with

(2) $p=5$, (3) $p=3.5$, (4) $p=2.5$, (5) $p=2$

Outlier models: $(n-r)$ x_i come from $N(0,\sigma^2)$ and r (we do not know which) come from

(6) $N(0,4\sigma^2)$, (7) $N(0,16\sigma^2)$; $r=\left[0.5+0.1n\right]$ (integer value).

Mixture models:

(8) $0.90N(0,\sigma^2)+0.10N(0,4\sigma^2)$, (9) $0.90N(0,\sigma^2)+0.10N(0,16\sigma^2)$

(10) Student's *t* distribution with 2 *df*,

(11) Cauchy distribution,

(12) Slash (Normal/Uniform) distribution

Models (1)-(9) have finite mean and variance, (10) has finite mean but non-existent variance, and (11)-(12) have non-existent mean and variance.

2.3 Simulations

We generated $N = [100{,}000/n]$ (integer value) samples (consisting of independently distributed observations) of size n from each of the models (1)-(12). The observations generated from models (6)-(9) were divided by suitable constants to make their variances equal to σ^2. From the resulting N values of MML30 and W24 (one of the most efficient estimators of μ and σ), we computed their means and variances. They are given in Table 2.1 and Table 2.2. For the normal distribution, $\hat{\mu}_x$ is a little less efficient than $\hat{\mu}_w$. For models (2)-(9), $\hat{\mu}_x$ is overall more efficient than $\hat{\mu}_w$. For models (10)-(12), $\hat{\mu}_x$ is considerably more efficient than $\hat{\mu}_w$. Realizing that $\hat{\mu}_x$ also has high breakdown, there does not seem to be any advantage in using the highly acclaimed M-estimators of μ.

Table 2.1: Simulated* values of $(n/\sigma^2)Var(\hat{\mu}_x)$ and $(n/\sigma^2)Var(\hat{\mu}_w)$.

	n = 10		n = 20		n = 50		n = 100	
Model	$\hat{\mu}_x$	$\hat{\mu}_w$	$\hat{\mu}_x$	$\hat{\mu}_w$	$\hat{\mu}_x$	$\hat{\mu}_w$	$\hat{\mu}_x$	$\hat{\mu}_w$
1	1.064	1.030	1.057	1.022	1.034	1.005	1.022	1.002
2	0.945	0.949	0.936	0.945	0.963	0.969	0.945	0.959
3	0.905	0.922	0.871	0.898	0.878	0.908	0.895	0.928
4	0.761	0.798	0.748	0.798	0.731	0.778	0.722	0.769
5	0.574	0.626	0.555	0.605	0.548	0.600	0.539	0.594
6	0.962	0.963	0.953	0.950	0.917	0.923	0.945	0.952
7	0.553	0.589	0.550	0.587	0.545	0.580	0.553	0.592
8	0.946	0.945	0.940	0.943	0.934	0.939	0.940	0.948
9	0.586	0.632	0.566	0.610	0.575	0.619	0.564	0.601
10	2.273	2.620	2.099	2.411	1.985	2.302	1.956	2.278
11	4.869	6.389	3.973	5.171	3.341	4.307	3.285	4.238
12	8.917	11.274	7.595	9.410	7.118	8.789	6.577	8.105
Sum/12	1.946	2.311	1.737	2.029	1.631	1.893	1.579	1.830
Tiku-Surucu	1.901	2.252	1.728	2.009	1.600	1.863	1.597	1.827

*Means are not given since both estimators are unbiased

Table 2.2: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance of $\hat{\sigma}_x$ and $\hat{\sigma}_w$.

	$n = 10$				$n = 50$				$n = 100$			
	Mean		Variance		Mean		Variance		Mean		Variance	
Model	$\hat{\sigma}_x$	$\hat{\sigma}_w$	$\hat{\sigma}_x$	$\hat{\sigma}_w$	$\hat{\sigma}_x$	$\hat{\sigma}_w$	$\hat{\sigma}_x$	$\hat{\sigma}_w$	$\hat{\sigma}_x$	$\hat{\sigma}_w$	$\hat{\sigma}_x$	$\hat{\sigma}_w$
1	0.92	0.93	0.579	0.545	0.97	0.99	0.531	0.521	0.97	1.00	0.531	0.525
2	0.90	0.91	0.649	0.637	0.93	0.96	0.631	0.656	0.94	0.97	0.591	0.615
3	0.87	0.88	0.678	0.687	0.91	0.94	0.661	0.696	0.91	0.94	0.636	0.680
4	0.81	0.82	0.682	0.701	0.84	0.87	0.660	0.714	0.84	0.88	0.650	0.707
5	0.71	0.72	0.652	0.688	0.73	0.76	0.582	0.640	0.74	0.77	0.587	0.646
6	0.89	0.89	0.594	0.587	0.93	0.96	0.544	0.554	0.93	0.96	0.538	0.552
7	0.72	0.71	0.464	0.466	0.75	0.76	0.435	0.463	0.75	0.76	0.448	0.478
8	0.90	0.90	0.652	0.651	0.93	0.96	0.624	0.646	0.93	0.96	0.592	0.622
9	0.72	0.72	0.707	0.771	0.75	0.76	0.646	0.705	0.75	0.76	0.629	0.684
10	1.42	1.43	3.269	3.522	1.44	1.49	2.906	3.231	1.44	1.50	2.885	3.243
11	2.06	2.08	14.003	15.840	1.94	2.03	9.091	10.809	1.93	2.03	8.939	10.610
12	2.85	2.85	24.422	27.047	2.75	2.84	15.821	18.464	2.73	2.83	14.704	17.215
Sum/12	1.15	1.15	3.946	4.345	1.16	1.19	2.761	3.175	1.16	1.20	2.644	3.048
Tiku-Surucu	1.15	1.15	3.853	4.034	1.20	1.19	3.059	3.097	1.21	1.20	3.111	3.127

MMLEs are clearly as good as M-estimators or better; see also Tiku and Surucu (2009). They are advantageous for two reasons: overall, (i) they have smaller bias, and (ii) they have smaller variance. MMLEs developed here are essentially as good as those of Tiku and Surucu (2009). In fact, the MMLE of σ developed here has a little less bias, and smaller mean square error. Like Tiku-Surucu estimators, our estimators have bounded influence functions. They are, therefore, as good as M-estimators or better and at least as good as Tiku-Surucu estimators.

2.4 Iterated MML30

In the expression for the weight w_i in (2.8), we used T_0 and S_0 as initial estimators of μ and σ, respectively. The question is whether replacing T_0 and S_0 by other estimators can result in improved efficiencies. In that regard, we first calculate $\hat{\mu}_x$ and $\hat{\sigma}_x$ iteratively as follows:

Initially, we use T_0 and S_0 and calculate $\hat{\mu}$ and $\hat{\sigma}$. We replace T_0 and S_0 by $\hat{\mu}$ and $\hat{\sigma}$, respectively, and calculate the new $\hat{\mu}$ and $\hat{\sigma}$. We repeat the process one more time and calculate $\hat{\mu}$ and $\hat{\sigma}$ and regard them as the desired MMLEs. Thus, the MMLEs are computed in two iterations besides computing them initally by using T_0 and S_0.

Given below are the simulated variances of $\hat{\mu}$ for the twelve models considered in Table 2.3.

Table 2.3: Simulated values of $(n/\sigma^2)Var(\hat{\mu})$ with two iterations.

Model	$n = 10$	$n = 20$	$n = 50$	$n = 100$
1	1.025	1.035	1.004	1.013
2	0.939	0.938	0.943	0.922
3	0.907	0.899	0.865	0.874
4	0.777	0.744	0.737	0.748
5	0.618	0.582	0.561	0.570
6	0.925	0.933	0.927	0.925
7	0.579	0.575	0.563	0.558
8	0.930	0.947	0.923	0.928
9	0.628	0.598	0.584	0.569
10	2.560	2.351	2.276	2.232
11	6.492	5.067	4.534	4.246
12	12.306	9.325	8.572	8.141
Sum/12	2.391	1.999	1.874	1.810

Increasing the iteration number does not have a significant effect on the results for models (1)-(9) all of which have finite moments. However, for models (10)-(12) which do not have finite moments, the variances are increased. Therefore, we conclude that iterations do not necessarily improve the results. This is in agreement with Huber's findings (Hampel et. al 1986, p.105).

2.5 Initial Estimators Based on Censored Samples

Obviously, estimators based on samples with extereme observations censored will not be subject to tail-effects (i.e., long-tails of a distribution or outliers in a sample) and can, therefore, make satisfactory initial estimators of μ and σ. Consider the censored sample

$$x_{(r+1)} \le x_{(r+2)} \le \ldots \le x_{(n-r)} \qquad (r = [0.5 + 0.3n]) \qquad (2.5.1)$$

and assume that the underlying distribution is normal $N(\mu, \sigma^2)$. The MMLEs of μ and σ are (Tiku, 1967)

$$\hat{\mu}_0 = \left\{ \sum_{i=r+1}^{n-r} x_{(i)} + r\beta\left(x_{(r+1)} + x_{(n-r)}\right) \right\} \Big/ m, \quad m = n - 2r + 2r\beta$$

and (2.5.2)

$$\hat{\sigma}_0 = \left\{ B + \sqrt{B^2 + 4AC} \right\} \Big/ 2\sqrt{A(A-1)}, \quad A = n - 2r;$$

$$B = r\alpha\left(x_{(n-r)} - x_{(r+1)}\right), \;\; C = \sum_{i=r+1}^{n-r}\left(x_{(i)} - \hat{\mu}_0\right)^2 + r\beta\left\{ \left(x_{(r+1)} - \hat{\mu}_0\right)^2 + \left(x_{(n-r)} - \hat{\mu}_0\right)^2 \right\},$$

$$\alpha = 0.7733 \text{ and } \beta = 0.7355 .$$

Replacing T_0 and S_0 by $\hat{\mu}_0$ and $\hat{\sigma}_0$, respectively, the means and variances of the resulting one-step MMLEs $\hat{\mu}$ and $\hat{\sigma}$ in (2.6) are given in Table 2.4 and Table 2.5. The means of $\hat{\mu}$ are not given because it is an unbiased estimator of μ for models (1)-(12) in section 2.2. It can be seen that the results are no better than those obtained by using T_0 and S_0 initially.

Table 2.4: Simulated values of $(n/\sigma^2)Var(\hat{\mu})$, using $\hat{\mu}_0$ and $\hat{\sigma}_0$ initially.

Model	n = 10	n = 20	n = 50	n = 100
1	1.059	1.028	1.031	0.985
2	0.963	0.950	0.943	0.983
3	0.917	0.877	0.911	0.947
4	0.764	0.784	0.816	0.870
5	0.594	0.572	0.603	0.679
6	0.964	0.944	0.948	1.002
7	0.574	0.598	0.653	0.723
8	0.974	0.959	0.953	0.974
9	0.603	0.615	0.653	0.725
10	2.251	2.297	2.510	2.832
11	5.870	4.752	5.080	5.998
12	10.294	9.151	9.861	11.821
Sum/12	2.152	1.961	2.080	2.378

Table 2.5: Simulated values of $(1/\sigma)\text{Mean}(\hat{\sigma})$ and $(n/\sigma^2)Var(\hat{\sigma})$.

	n = 10		n = 50		n = 100	
Model	Mean	Variance	Mean	Variance	Mean	Variance
1	0.91	0.663	1.01	0.534	1.03	0.534
2	0.87	0.706	0.99	0.702	1.02	0.731
3	0.86	0.788	0.98	0.795	1.01	0.879
4	0.79	0.785	0.93	0.905	0.97	1.057
5	0.69	0.707	0.83	0.849	0.88	1.095
6	0.87	0.677	1.00	0.622	1.02	0.684
7	0.70	0.574	0.87	0.729	0.92	0.937
8	0.88	0.732	1.00	0.721	1.02	0.795
9	0.71	0.784	0.87	1.139	0.92	1.531
10	1.37	3.596	1.71	4.725	1.86	6.628
11	1.98	15.261	2.52	16.813	2.91	23.387
12	2.75	25.234	3.55	30.798	4.10	43.080
Sum/12	1.12	4.209	1.35	4.944	1.47	6.778

When we compare Table 2.4 with Table 2.1 and Table 2.5 with Table 2.2, we observe an overall increase in the variances of the estimators. However, when the results of the mean of sigma are examined, it can be seen that the bias is reduced for models with finite variances. This might or might not be inconsequential for practical purpose.

2.6 Populations with Finite Mean and Variance

If the underlying distribution is known to be long-tailed symmetric with finite variance as in most situations, v_i in (2.7) may be taken to be equal to its original value, namely,

$$v_i = (w_i/k)\tilde{t}_i^{\,3}.$$

We give the simulated values of the mean and variance of the resulting MMLE of sigma in Table 2.6.

Table 2.6: Simulated values of $(1/\sigma)\text{Mean}(\hat{\sigma})$ and $(n/\sigma^2)\text{Var}(\hat{\sigma})$, where $v_i = (w_i/k)\tilde{t}_i^{\,3}$.

	n = 10		n = 50		n = 100	
Model	Mean	Variance	Mean	Variance	Mean	Variance
1	0.98	0.553	1.00	0.529	1.00	0.503
2	0.96	0.712	0.98	0.681	0.98	0.641
3	0.95	0.815	0.97	0.791	0.97	0.775
4	0.92	1.105	0.93	0.953	0.93	0.954
5	0.84	1.181	0.84	1.064	0.85	1.070
6	0.97	0.639	0.98	0.613	0.98	0.600
7	0.88	0.947	0.88	0.857	0.88	0.839
8	0.96	0.734	0.98	0.698	0.98	0.702
9	0.88	1.504	0.88	1.400	0.88	1.417
Sum/9	0.93	0.910	0.94	0.843	0.94	0.833
*	0.83	0.629	0.86	0.590	0.86	0.578

* Are the values for $\hat{\sigma}$ (developed in this chapter)

The MMLE of sigma has smaller bias for models with finite variances. However, the variance of the estimator is increased. This is a very common phenomenon, however.

From the point of view of having bounded influence functions, the MMLEs given in (2.6)-(2.8) are adventageous. We use them in further development of the subject matter.

CHAPTER 3

SKEW DISTRIBUTIONS

Consider the family of skew distributions represented by the Generalized Logistic (b>0)

$$f(x) = \frac{b}{\sigma} \frac{\exp\{-(x-\mu)/\sigma\}}{\left[1+\exp\{-(x-\mu)/\sigma\}\right]^{b+1}}, \qquad -\infty < x < \infty, \tag{3.1}$$

where μ is location, σ is scale and b is shape parameter. Note that different values of b characterize different types of distributions; b<1, b=1 (called Logistic Distribution) and b>1 respectively denotes negatively skewed, symmetric and positively skewed distributions.

Tiku and Akkaya (2004) give the mean and the variance of Generalized Logistic distribution in terms of *psi-function* (also called *digamma function*) $\Psi(.) = \Gamma'(.)/\Gamma(.)$, where $\Gamma(.)$ is the gamma function:

$$E(z) = \Psi(b) - \Psi(1) \text{ and } Var(z) = \Psi'(b) + \Psi'(1)$$

for $z = (x-\mu)/\sigma$. The values of $\Psi(b)$ and $\Psi'(b)$ are tabulated in Tiku and Akkaya (2004) as follows:

Table 3.1: Values of the psi-function $\Psi(b)$.

b	0.5	1	2	4	6	8
$\Psi(b)$	-1.9635	-0.5772	0.4228	1.2561	1.7061	2.0156
$\Psi'(b)$	4.9348	1.6449	0.6449	0.2838	0.1813	0.1331

Since μ is the location parameter and σ is the scale parameter, the mean of an observation *x* from Generalized Logistic distribution is

$$E(x) = \mu + \sigma\left[\Psi(b) - \Psi(1)\right] \tag{3.2}$$

while the variance of x is

$$Var(x) = \sigma^2\left[\Psi'(b) + \Psi'(1)\right]. \tag{3.3}$$

The values of its skewness and kurtosis are given below:

Table 3.2: The values of the skewness and kurtosis of generalized logistic distribution with shape parameter *b*.

$b =$	0.5	1	2	4	6
Skewness $\mu_3 / \mu_2^{3/2}$	-0.855	0.000	0.511	0.868	0.961
Kurtosis μ_4 / μ_2^2	5.400	4.200	4.333	4.758	4.951

Given a random sample $x_1, x_2, ..., x_n$, the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^{n} \frac{\exp\{-(x_i - \mu)/\sigma\}}{\left[1 + \exp\{-(x_i - \mu)/\sigma\}\right]^{b+1}} . \tag{3.4}$$

The maximum likelihood equations expressed in terms of the standardized ordered variates $z_{(i)} = (x_{(i)} - \mu)/\sigma$ are

$$\frac{d\ln L}{d\mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0$$

and (3.5)

$$\frac{d\ln L}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0;$$

$$g(z) = e^{-z}/(1 + e^{-z}) = 1/(1 + e^{z}) .$$

These equations have no explicit solutions. To obtain modified maximum likelihood equations, we linearize $g(z_{(i)})$:

$$g(z_{(i)}) \cong \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n; \tag{3.6}$$

α_i and β_i are obtained from the first two terms of a Taylor series expansion. That gives; $t = t_{(i)}$:

$$\alpha_i = \left(1 + \exp(t) + t\exp(t)\right)/\left(1 + \exp(t)\right)^2 \text{ and } \beta_i = \exp(t)/\left(1 + \exp(t)\right)^2 . \tag{3.7}$$

Here, we use the approximate values of $t_{(i)}$ for $n > 10$,

$$t = t_{(i)} = -\ln(q_i^{-1/b} - 1), \quad q_i = i/(n+1), \quad 1 \le i \le n. \qquad (3.8)$$

Balakrishnan and Leung (1988) tabulated the true values of $t_{(i)} = E\{z_{(i)}\}$ for $n \le 15$. However, using the approximate values does not alter the efficiencies of the resulting estimators in any substantial way (Tiku and Akkaya, 2004).

Incorporating (3.6) in (3.5) gives the modified maximum likelihood equations (Tiku and Akkaya, 2004):

$$\frac{d \ln L}{d\mu} \cong \frac{d \ln L^*}{d\mu} = \frac{(b+1)m}{\sigma^2}(K + D\sigma - \mu) = 0$$

and (3.9)

$$\frac{d \ln L}{d\sigma} \cong \frac{d \ln L^*}{d\sigma} = -\frac{1}{\sigma^3}\left[\left(n\sigma^2 - B\sigma - C\right) - (b+1)m(K-\mu)(K + D\sigma - \mu)\right] = 0$$

where

$$m = \sum_{i=1}^{n} \beta_i, \quad K = \left(\sum_{i=1}^{n} \beta_i x_{(i)}\right) / m, \quad D = \sum_{i=1}^{n} \Delta_i / m, \quad \Delta_i = (b+1)^{-1} - \alpha_i, \qquad (3.10)$$

$$B = (b+1)\sum_{i=1}^{n} \Delta_i (x_{(i)} - K) \text{ and} \qquad (3.11)$$

$$C = (b+1)\left(\sum_{i=1}^{n} \beta_i x_{(i)}^2 - mK^2\right) = (b+1)\sum_{i=1}^{n} \beta_i \left(x_{(i)} - K\right)^2. \qquad (3.12)$$

The solutions of (3.9) are the MMLEs:

$$\hat{\mu} = K + D\hat{\sigma} \text{ and } \hat{\sigma} = \left\{B + \sqrt{\left(B^2 + 4nC\right)}\right\} \Big/ 2\sqrt{n(n-1)}. \qquad (3.13)$$

***Remark*:** For a given *b*, the estimators $\hat{\mu}$ and $\hat{\sigma}$ have negligible bias and are highly efficient for all sample sizes. Asymptotically, they are unbiased and fully efficient. They are also robust to plausible deviations from an assumed distribution in the family (3.1) and to moderate data anomalies; see for example, Senoglu and Tiku (2001).

3.1 Unspecified Shape Parameter

In machine data processing, it might not be possible to specify *b*. Now, the only assumption we make is that the underlying distribution is one of (3.1). We proceed as follows:

As usual, we write

$$T_0 = median\{x_i\} \text{ and } S_0 = 1.483 median\{|x_i - T_0|\} \quad (1 \le i \le n).$$

For *b*=1 (logistic distribution which is symmetric), T_0 and S_0 are particularly good initial estimators of the median and the standard deviation $\sqrt{2\Psi'(1)}\sigma$.

3.2 Unknown *b*

Since we do not know the value of the shape parameter *b*, we estimate (*b*+1) from a given sample. We also estimate $t_{(i)}$'s and hence, α_i's and β_i's in (3.7).

The initial estimates of $t_{(i)}$ are

$$\tilde{t}_{(i)} = \left(x_{(i)} - T_0\right)/S_0 \quad (1 \le i \le n); \tag{3.2.1}$$

hence, the initial estimates of α_i and β_i, $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ respectively, are obtained by replacing $t_{(i)}$ by $\tilde{t}_{(i)}$ $(1 \le i \le n)$.

To estimate $(b+1)^{-1}$ and $(b+1)$ in (3.10)- (3.12), we note that

$$E\left(\frac{1}{1+e^{-z}}\right) = \int_{-\infty}^{\infty} \frac{e^{-z}}{\left(1+e^{-z}\right)^{b+2}} dz = \frac{b}{b+1} \quad (z = (x-\mu)/\sigma). \tag{3.2.2}$$

Writing

$$w_i = e^{z_i}/\left(1+e^{z_i}\right) = 1/\left(1+e^{-z_i}\right), \tag{3.2.3}$$

$E(\overline{w}) = E\left\{(1/n)\sum_{i=1}^{n} w_i\right\} = b/(b+1)$. An initial estimate of $b/(b+1)$ is

$$\tilde{\overline{w}} = (1/n)\sum_{i=1}^{n} \tilde{w}_i, \quad \tilde{w}_i = e^{\tilde{t}_{(i)}}/\left(1+e^{\tilde{t}_{(i)}}\right) = 1/\left(1+e^{-\tilde{t}_{(i)}}\right). \tag{3.2.4}$$

Thus,

$$1-\tilde{\tilde{w}} \text{ is an initial estimator of } (b+1)^{-1}$$

and (3.2.5)

$$1/\left(1-\tilde{\tilde{w}}\right) \text{ is an initial estimator of } (b+1).$$

***Remark*:** Since $\tilde{w}_i$ are bounded between 0 and 1, $1-\tilde{\tilde{w}}$ and $1/\left(1-\tilde{\tilde{w}}\right)$ converge to their expected values $(b+1)^{-1}$ and $(b+1)$, respectively, very quickly with increasing n.

The MMLEs are calculated by replacing α_i by $\tilde{\alpha}_i$, β_i by $\tilde{\beta}_i$, $(b+1)^{-1}$ by $(1-\tilde{\tilde{w}})$ and $(b+1)$ by $1/\left(1-\tilde{\tilde{w}}\right)$ in (3.10)-(3.12). Since complete sums are invariant to ordering, (3.10)-(3.12) can be written in terms of x_i and $\tilde{t}_i$ simply by droping the ordered symbol '( )' on them. The estimates are calculated from five iterations starting with T_0 and S_0. Calculation show that no more than five iterations are needed for the estimates to stabilize sufficiently. Here, more than two iterations are required for estimates to stabilize sufficiently because the underlying distributions are skewed. It may be noted that $\hat{\mu}$ is estimating the population median $\mu-\ln(2^{1/b}-1)\sigma$ and $\hat{\sigma}$ is estimating the population scale parameter σ.

3.3 Simulations

To study the properties of these new estimators, we carried out comprehensive simulation studies based on N=[100.000/n] Monte Carlo runs. Random samples were generated for a given b in (3.1), and N estimates of the median and scale computed. Random observations x_i $(1 \le i \le n)$ generated when $b \ne 1$ were multiplied by $\left[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}\right]^{1/2}$ so that the variances of x_i are always the same as when b = 1 (logistic distribution), i.e., $2\Psi'(1)\sigma^2 = 3.2898\sigma^2$. It may be noted that $\hat{\mu}$ is then estimating the scaled median

$$\textit{Scaled median} = \left\{\mu-\ln(2^{1/b}-1)\sigma\right\}\left[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}\right]^{1/2}; \qquad (3.3.1)$$

μ is taken to be zero without loss of generality, and $\hat{\sigma}$ is estimating the scale parameter σ which is taken to be 1 without loss of generality. If one wants to estimate the median, not the scaled median, the *x*-observations need not be multiplied by $[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}]^{1/2}$.

The means and variances of the *N* estimates of the scaled median and the scale parameter are given in Table 3.3. It is pleasing to notice that the new MMLEs have negligible bias in spite of the fact that $t_{(i)}$ $(1 \le i \le n)$ and $(b+1)^{-1}$ are estimated from a given sample.

Table 3.3: Simulated values of means and variances of the MMLEs $\hat{\mu}$ and $\hat{\sigma}$; $Scaled\, median = -\ln(2^{1/b}-1)[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}]^{1/2}\sigma,\ \mu=0$ without loss of generality.

	$b = 0.5$				$b = 1$			
	Scaled median = -0.777				Scaled median = 0			
n =	**10**	**20**	**50**	**100**	**10**	**20**	**50**	**100**
$(1/\sigma)\text{Mean}(\hat{\mu})$	-0.813	-0.808	-0.786	-0.780	-0.003	0.000	0.000	0.000
$(n/\sigma^2)\text{Var}(\hat{\mu})$	3.273	3.263	3.207	3.344	3.553	3.642	3.428	3.357
$(1/\sigma)\text{Mean}(\hat{\sigma})$	0.936	0.947	0.948	0.949	0.982	0.990	0.996	1.001
$(n/\sigma^2)\text{Var}(\hat{\sigma})$	0.751	0.717	0.704	0.739	0.755	0.733	0.759	0.785

	$b = 2$				$b = 4$			
	Scaled median = 1.056				Scaled median = 2.174			
n =	**10**	**20**	**50**	**100**	**10**	**20**	**50**	**100**
$(1/\sigma)\text{Mean}(\hat{\mu})$	1.083	1.069	1.071	1.066	2.211	2.194	2.185	2.179
$(n/\sigma^2)\text{Var}(\hat{\mu})$	3.573	3.678	3.935	3.876	3.717	3.708	3.874	3.913
$(1/\sigma)\text{Mean}(\hat{\sigma})$	0.994	1.003	1.008	1.015	0.993	1.001	1.008	1.014
$(n/\sigma^2)\text{Var}(\hat{\sigma})$	0.813	0.806	0.765	0.800	0.857	0.821	0.831	0.809

	$b = 6$				$b = 8$			
	Scaled median = 2.819				Scaled median = 3.268			
n =	**10**	**20**	**50**	**100**	**10**	**20**	**50**	**100**
$(1/\sigma)\text{Mean}(\hat{\mu})$	2.845	2.826	2.815	2.815	3.312	3.287	3.283	3.272
$(n/\sigma^2)\text{Var}(\hat{\mu})$	3.567	3.580	3.724	3.731	3.709	3.711	3.780	3.989
$(1/\sigma)\text{Mean}(\hat{\sigma})$	0.984	0.996	1.006	1.006	0.996	1.004	1.008	1.009
$(n/\sigma^2)\text{Var}(\hat{\sigma})$	0.874	0.804	0.827	0.845	0.923	0.868	0.878	0.840

The estimators $\hat{\mu}$ and $\hat{\sigma}$ work very well and are unbiased (almost) for each shape parameter (unknown to us) and sample size.

The results above are very promising indeed and extend Huber type work to skew distributions. Huber M-estimation is not applicable to skew distributions. It may be noted that the MMLEs above have bounded influence functions since for any k,

$$\lim_{|t|\to\infty}\left\{ t^k e^t \big/ \left(1+e^t\right)^2 = t^k e^{-t} \big/ \left(1+e^{-t}\right)^2 \right\} \to 0\,, \qquad (3.3.2)$$

k being 1 or 2 in our situation. See Appendix B for details.

3.4 Least Square Estimators

For $b = 1$, the LSE of the median is $\bar{x}$ with variance

$$nV(\bar{x}) = 2\Psi'(1)\sigma^2 = 3.290\sigma^2 .$$

For $n = 10$, the relative efficiency of the new MMLE (which does not assume any knowledge of b) is

100(Variance of $\bar{x}$ / Variance of MMLE) = 93%

which is indeed a promising result; for $n = 100$, it is 98%.

For $b = 1$, the LSE of σ is $s/\sqrt{2\Psi'(1)}$ with asymptotic variance (Roy and Tiku, 1962)

$$\frac{\sigma^2}{2n}\left(1+\frac{1}{2}\lambda_4\right),\ \lambda_4 = \beta_2 - 3 . \qquad (3.4.1)$$

For $b = 1$, $\beta_2 = 4.2$. For $n = 100$, the value of this variance is 0.80. The corresponding variance of the new MMLE $\hat{\sigma}$ is 0.785 (Table 3.3). Again, the result is very promising.

***Comment*:** The method may extend to other skew distributions. That needs further study.

***Remark*:** We give below the variances of the LSE $s/\sqrt{\Psi'(b)+\Psi'(1)}$ of σ for n = 100 calculated from (3.4.1); b assumed known:

$b =$	0.5	1	2	4	6
$(n/\sigma^2)Variance$	1.100	0.800	0.833	0.940	0.988

These may be compared with the corresponding values in Table 3.3. The MMLEs are not only more efficient than the LSEs but no knowledge of b is needed in calculating the former.

***Remark*:** The LSE of the population median is T_0. Given in Table 3.4 are the simulated means and variances of T_0 and $S_0^* = S_0/1.483$ (proposed by Huber as an initial estimator). It can be seen that T_0 and S_0^* have negligible bias. However, T_0 and S_0^* are jointly much less efficient than the LSEs we now propose. It can be seen that T_0 and S_0^* are good only as initial estimators.

Table 3.4: Simulated means and variances of T_0 and $S_0^* = S_0/1.483$.

	n = 10				n = 20			
	T_0		S_0^*		T_0		S_0^*	
b	Mean	nxVar	Mean	nxVar	Mean	nxVar	Mean	nxVar
0.5	-0.812	3.370	0.977	1.500	-0.795	3.504	1.015	1.622
1	-0.008	3.672	1.024	1.551	-0.002	3.772	1.069	1.672
2	1.088	3.783	1.035	1.569	1.066	3.882	1.079	1.711
4	2.215	3.785	1.027	1.542	2.193	3.966	1.064	1.678
6	2.838	3.809	1.014	1.518	2.827	3.987	1.054	1.697
8	3.305	3.919	1.016	1.600	3.287	4.116	1.062	1.737

	n = 50				n = 100			
	T_0		S_0^*		T_0		S_0^*	
b	Mean	nxVar	Mean	nxVar	Mean	nxVar	Mean	nxVar
0.5	-0.789	3.642	1.022	1.692	-0.780	3.798	1.031	1.756
1	-0.005	4.038	1.085	1.768	-0.011	4.285	1.096	1.760
2	1.062	4.202	1.097	1.779	1.057	4.295	1.113	1.722
4	2.165	4.373	1.089	1.823	2.177	4.403	1.105	1.844
6	2.806	4.019	1.082	1.787	2.812	4.215	1.086	1.807
8	3.276	4.468	1.079	1.831	3.273	4.249	1.090	1.664

Proposal: The proposed LSE of the scaled population median

$$\left\{\mu - \ln\left(2^{1/b} - 1\right)\sigma\right\}\left[2\Psi'(1)/\{\Psi'(b) + \Psi'(1)\}\right]^{1/2}$$

is

$$\tilde{\mu} = \left(\bar{x} - [\Psi(b) - \Psi(1)]\, s/\sqrt{\Psi'(b) + \Psi'(1)}\right)\left[2\Psi'(1)/\{\Psi'(b) + \Psi'(1)\}\right]^{1/2}; \quad (3.4.2)$$

μ may be taken to be zero without loss of generality. The LSE of σ is given on the previous page. Note that (3.4.2) is similar in form to $\hat{\mu}$ in (3.13). It is not possible to derive the variance of this estimator analytically because the $Cov(\bar{x}, s)$ is difficult to determine even asymptotically.

3.5 Comparison of MMLEs and Proposed LSEs

In this section, we let shape parameter b be unknown for the calculation of both the MMLEs and LSEs. We will estimate the unknown shape parameter b from (3.2.5) and incorporate it in our computations. The results of our simulations are given in Table 3.5 and Table 3.6 with sample sizes $n = 10$ and $n = 20$, respectively. True values of the scaled population median are given in Table 3.3. It may be noted that

$$\Psi'(d) = \sum_{i=1}^{\infty}(i + d - 1)^{-2} \quad (3.5.1)$$

while $\Psi(d)$ is computed by using FORTRAN subroutine 'psi' in IMSL/LIBRARY Special Functions.

It can be seen that the LSEs have larger bias than the MMLEs. Overall, the MMLEs are considerably more efficient (jointly) than the LSEs. This is very interesting indeed. Moreover, unlike the MMLEs, the LSEs do not have bounded influence functions; see Appendix B. That is a rerious drawback in the context of robustness.

Table 3.5: Simulated values of means and variances of the MMLEs and LSEs; $\mu = 0$ and $\sigma = 1$ without loss of generality, $n = 10$.

***n* = 10**	Median MMLE	Median LSE	Scale MMLE	Scale LSE	Median MMLE	Median LSE	Scale MMLE	Scale LSE
	$b = 0.5$				$b = 1$			
Mean	-0.812	-0.910	0.936	0.927	0.001	-0.003	0.983	0.961
nxVar	3.271	3.092	0.748	0.809	3.462	3.338*	0.793	0.803
	$b = 2$				$b = 4$			
Mean	1.077	1.109	0.997	0.977	2.207	2.246	0.987	0.972
nxVar	3.654	3.641	0.822	0.881	3.710	3.845	0.824	0.928
	$b = 6$				$b = 8$			
Mean	2.845	2.883	0.986	0.973	3.311	3.347	0.989	0.977
nxVar	3.577	3.757	0.861	0.984	3.645	3.844	0.857	0.989

* For b known, the variance of the LSE is 3.290.

Table 3.6: Simulated values of means and variances of the MMLEs and LSEs; $\mu = 0$ and $\sigma = 1$ without loss of generality, $n = 20$.

***n* = 20**	Median MMLE	Median LSE	Scale MMLE	Scale LSE	Median MMLE	Median LSE	Scale MMLE	Scale LSE
	$b = 0.5$				$b = 1$			
Mean	-0.797	-0.913	0.944	0.952	0.001	-0.002	0.992	0.981
nxVar	3.263	3.118	0.744	0.903	3.567	3.469	0.740	0.783
	$b = 2$				$b = 4$			
Mean	1.063	1.107	1.006	0.995	2.191	2.246	1.005	1.001
nxVar	3.650	3.706	0.778	0.878	3.730	3.975	0.846	1.032
	$b = 6$				$b = 8$			
Mean	2.826	2.884	1.000	1.000	3.292	3.349	0.998	1.000
nxVar	3.694	4.000	0.840	1.049	3.615	3.969	0.868	1.105

***Comment*:** Another way of estimating the shape parameter b would be to calculate $\hat{\mu}$ and $\hat{\sigma}$ for a series of values of b and choose that value (of b) which maximizes (Tiku and Akkaya, 2010)

$$(1/n)\{\ln L\}_{\mu=\hat{\mu},\sigma=\hat{\sigma}} . \tag{3.5.2}$$

This procedure is computationally more involved and will be considered in future research. Apparently, it might not yield substantially better results than those obtained by using $\tilde{\tilde{w}}$ in the estimation of b. This is because $\tilde{\tilde{w}}$ is bounded between 0 and 1 and converges to its expected value very quickly as the sample size n increases.

CHAPTER 4

ANALYSIS OF VARIANCE IN EXPERIMENTAL DESIGN

Experimental design is a very important area not only for applied but also for theoretical studies in statistics. The traditional assumption of normality of course leads to the development of an enormous amount of theory related to experimental design. The normality assumption makes it possible to test treatment effects by defining Fisher F-statistics. However, non-normal distributions occur more frequently in practice. In statistical literature, one can find many studies dealing with non-normal data in experimental design and the effects of non-normality on the F-statistics (Geary, 1947; Gayen, 1950; Srivastava, 1959; Tiku, 1964; Donaldson, 1968; Tiku 1971b; Spjøtvoll and Aastveit, 1980; Tan and Tiku, 1999, Senoglu and Tiku, 2001). In this chapter specifically we extend analysis of variance procedures given in Senoglu and Tiku (2001) to non-normal data in a single factor experimental design. In later chapters we extend the methodology to more complex data structures. In particular, our method makes it possible to extend Senoglu and Tiku (2002) results to situations where the shape parameters (of the assumed Generalized Logistic) in blocks are different and unknown. This is a very important advance because Senoglu and Tiku (2002) assume that shape parameters are different but known. It may be noted that different shape parameters create non-identical blocks, perceived to be a very difficult problem for statistical analyses.

4.1 One-Way Classification

Consider the one-way classification model

$$y_{ij} = \mu + \gamma_i + e_{ij} \quad (i = 1,2,\ldots,a;\ j = 1,2,\ldots,n), \tag{4.1.1}$$

where μ is a constant and γ_i is the effect due to i^{th} treatment (or block). This is a balanced design since the number of observations in each block, n, is the same. Without loss of generality, we assume that it is a fixed effects model and $\sum_{i=1}^{a} \gamma_i = 0$. Different types of distribution families for the errors e_{ij} are studied in the following subsections. In each of them, the errors e_{ij} are assumed to be iid. It may be noted that our method readily extends to situations where the number of observations in blocks are unequal (unbalanced design). In this thesis, we will confine ourselves to balanced designs.

4.1.1 Normal Distribution

We first assume that e_{ij} are iid normal $N(0,\sigma^2)$. The likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^{an} \prod_{i=1}^{a}\prod_{j=1}^{n} \exp\left\{-\left(y_{ij} - \mu - \gamma_i\right)^2 / 2\sigma^2\right\}.$$

The MLEs are solutions of the equations $\partial \ln L / \partial \mu = 0, \partial \ln L / \partial \gamma_i = 0$ $(i = 1,2,\ldots,a)$ and $\partial \ln L / \partial \sigma = 0$. They are

$$\hat{\hat{\mu}} = \bar{y}_{..},\ \hat{\hat{\gamma}}_i = \bar{y}_{i.} - \bar{y}_{..} \ \ (1 \le i \le a) \text{ and}$$

$$\hat{\hat{\sigma}}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{i.}\right)^2 / a(n-1) = \sum_{i=1}^{a} s_i^2 / a$$

where $\bar{y}_{i.} = (1/n)\sum_{j=1}^{n} y_{ij}$ and $\bar{y}_{..} = (1/an)\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$. All these estimators are unbiased; $\bar{y}_{i.}$ and $\bar{y}_{..}$ are also the *MVB* estimators.

Fisher decomposition of the total sum of squares is

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2 = n\sum_{i=1}^{a}\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{c}\left(y_{ij} - \bar{y}_{i.}\right)^2$$

or

$$S_T^2 = S_{block}^2 + S_{error}^2 .$$

The sums of squares S_{block}^2 and S_{error}^2 on the right hand side are called 'block' and 'error' sums of squares, respectively. Under the normality assumption, S_{block}^2 / σ^2 is distributed as chi-sqaure with $(a-1)$ degrees of freedom if the null hypothesis

$$H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_a = 0$$

is true; $S_{error}^2 / a(n-1)$ is independently distributed as chi-sqaure with $a(n-1)$ degrees of freedom. These results lead to Fisher *F* statistic

$$F = s_{block}^2 / s_{error}^2$$

where

$$s_{block}^2 = S_{block}^2 / (a-1) \text{ and } s_{error}^2 = S_{error}^2 / a(n-1)$$

are called 'block' and 'error' mean sums of squares, respectively. Large values of *F* lead to the rejection of H_0 in favor of H_1,

$$H_1 : \text{At least one } \gamma_i \neq 0 \ \ (1 \le i \le a) .$$

The null distribution of *F* is central F with $\nu_1 = a-1$ and $\nu_2 = a(n-1)$ degrees of freedom. Under H_1, the distribution of *F* is non-central F with (ν_1, ν_2) degrees of freedom and non-centrality parameter

$$\lambda^2 = n \sum_{i=1}^{a} \gamma_i^2 / \sigma^2 .$$

A Laguerre series expansion which is computationally straightforward is developed in Tiku (1965). See also Tiku (1985a,b).

Writing $\mu_i = \mu + \gamma_i$ in the linear function

$$\sum_{i=1}^{a} \ell_i \gamma_i = \sum_{i=1}^{a} \ell_i \mu_i ,$$

μ_i is estimated by $\bar{y}_{i.} = (1/n) \sum_{j=1}^{n} y_{ij}$; $\bar{y}_{i.}$ is unbiased and $Var(\bar{y}_{i.}) = \sigma^2 / n$. The estimators $\bar{y}_{i.}$ $(1 \le i \le a)$ are mutually independent.

Under H_0, every linear contrast is zero. To test that a particular linear contrast

$$\sum_{i=1}^{a} \ell_i \, \mu_i, \quad \sum_{i=1}^{a} \ell_i = 0,$$

is zero, the test statistic is

$$t = \sum_{i=1}^{a} \ell_i \, \bar{y}_{i.} \Big/ \hat{\hat{\sigma}} \sqrt{\sum_{i=1}^{a} \left(\ell_i^2 / n\right)}.$$

The null distribution of t is Student's t with $\nu = a(n-1)$ degrees of freedom.

Remark: If the distribution is a known location-scale distribution, Senoglu and Tiku (2001) worked out MMLEs of μ_i $(1 \le i \le a)$ and σ. They showed that the corresponding variance-ratio statistic is similar to the F statistic above. Our aim here is to develop methodology which can be used in machine data processing in the context of experimental design. In such a situation, the only information is that the underlying distribution is of certain types, e.g., long-tailed symmetric.

4.1.2 Long-Tailed Symmetric Family

Suppose that e_{ij} are iid and distrubuted as one of the distributions in the family

$$f(e) = \frac{1}{\sigma\sqrt{k}} \frac{1}{\beta(1/2, p-1/2)} \left[1 + \frac{e^2}{k\sigma^2}\right]^{-p}, \qquad -\infty < e < \infty \; ; \quad (4.1.2.1)$$

$k = 2p - 3$, $p \ge 2$ and $\beta(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Note that $E(e) = \mu$ and $V(e) = \sigma^2$. For this family, the likelihood function L of the observations y_{ij} $(1 \le i \le a, 1 \le j \le n)$ is

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{a} \prod_{j=1}^{n} \left\{1 + \frac{z_{ij}^2}{k}\right\}^{-p} \; ;$$

$N = an$, $z_{ij} = e_{ij}/\sigma = (y_{ij} - \mu - \gamma_i)/\sigma \quad (i = 1,2,\ldots,a;\; j = 1,2,\ldots,n)$.

The likelihood equations for estimating μ, γ_i $(1 \le i \le a)$ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{\sigma k} \sum_{i=1}^{a} \sum_{j=1}^{n} g\left(z_{ij}\right) = 0 \tag{4.1.2.2}$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{2p}{\sigma k} \sum_{j=1}^{n} g\left(z_{ij}\right) = 0 \tag{4.1.2.3}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{\sigma k} \sum_{i=1}^{a} \sum_{j=1}^{n} z_{ij}\, g\left(z_{ij}\right) = 0 \tag{4.1.2.4}$$

where the function $g(z)$ is given by

$$g(z) = z \big/ \left(1 + z^2/k\right). \tag{4.1.2.5}$$

In order to solve the equations (4.1.2.2)-(4.1.2.5) to obtain the MLEs, a+1 number of equations have to be iterated simultaneously. This is a difficult and time consuming task and there can be problems of convergence as stated before. Therefore, we will utilize the MMLEs in our analysis.

If we let $y_{i(1)} \le y_{i(2)} \le \ldots \le y_{i(n)}$ $(1 \le i \le a)$ be the order statistics of the n observations y_{ij} $(1 \le j \le n)$ in the i^{th} block, then

$$z_{i(j)} = \left(y_{i(j)} - \mu - \gamma_i\right)/\sigma \quad (i = 1,2,\ldots,a;\ j = 1,2,\ldots,n)$$

are ordered z_{ij} variates. After replacing z_{ij} by $z_{i(j)}$ and using the linear approximation

$$g\left(z_{i(j)}\right) \cong \alpha_j + \beta_j z_{i(j)} \quad (1 \le j \le n)$$

where (replacing $t_{(j)}^3$ by $t_{(j)}$ in α_j for reasons given earlier)

$$\alpha_j = \frac{(1/k)t_{(j)}}{\left\{1 + (1/k)t_{(j)}^2\right\}^2} \text{ and } \beta_j = \frac{1}{\left\{1 + (1/k)t_{(j)}^2\right\}^2},\ t_{(j)} = E\left\{z_{i(j)}\right\}$$

we obtain the modified likelihood equations $\partial \ln L^* / \partial \mu = 0, \partial \ln L^* / \partial \gamma_i = 0$, $\partial \ln L^* / \partial \sigma = 0$. One may like to drop the ordering on $t_{(j)}$ and $y_{i(j)}$ since the ordering of $z_{i(j)}$ can be disregarded as complete sums are invariant to ordering. This has also been explained in Chapter 2.

Remember that we proposed $\widetilde{t}_i = \left(x_i - T_0\right)/S_0, \left(i = 1,\ldots,n\right)$ as the initial estimate of $t_{(i)}$ in Chapter 2, where there exists only one block. When the block number is more than one, however, we need a modification since we have $T_{0i} = median\{y_{ij}\}$ and $S_{0i} = 1.483 median\{|y_{ij} - T_{0i}|\}$ $(1 \le i \le a)$ for each block. In the present situation:

$$\tilde{t}_{ij} = (y_{ij} - T_{0i})/S_{0i} .$$

The initial estimates of α_j and β_j, respectively, are obtained by replacing t_j by $\tilde{t}_{ij}$ $(1 \le i \le n)$. The resulting α_j and β_j coefficients will be denoted by $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_{ij}$, respectively.

The explicit solutions of the modified likelihood equations are the MMLEs of μ, γ_i $(1 \le i \le a)$ and σ:

$$\hat{\mu} = \sum_{i=1}^{a} \hat{\mu}_{i.}/a, \quad \hat{\gamma}_i = \hat{\mu}_{i.} - \hat{\mu} \quad \text{and } \hat{\sigma} = \left\{B + \sqrt{(B^2 + 4NC)}\right\}/(2N) \quad (4.1.2.6)$$

where

$$B = \sum_{i=1}^{a} B_i, \; C = \sum_{i=1}^{a} C_i,$$

$$B_i = \frac{2p}{k}\sum_{j=1}^{n} \tilde{\alpha}_{ij}\left(y_{ij} - \hat{\mu}_{i.}\right), \; C_i = \frac{2p}{k}\sum_{j=1}^{n} \tilde{\beta}_{ij}\left(y_{ij} - \hat{\mu}_{i.}\right)^2, \quad (4.1.2.7)$$

$$\hat{\mu}_{i.} = (1/m_i)\sum_{j=1}^{n} \tilde{\beta}_{ij} y_{ij} \quad \text{and} \quad m_i = \sum_{j=1}^{n} \tilde{\beta}_{ij}.$$

Note that, as in Chapter 2, $k = 30\,(p = 16.5)$.

A more conveniant form of the MMLE of σ is

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{a} \hat{\sigma}_i^2 \Big/ a}, \quad (4.1.2.8)$$

where $\hat{\sigma}_i = \left\{B_i + \sqrt{(B_i^2 + 4nC_i)}\right\}/ 2\sqrt{n(n-1)}$, B_i and C_i $(1 \le i \le a)$ are given in (4.1.2.7). Note that $\hat{\sigma}^2$ given in (4.1.2.8) is advantageous because it has the same form as the corresponding LSE, namely,

$$s^2 = \left(s_1^2 + s_2^2 + \ldots + s_a^2\right)/a; \quad s_i^2 = \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 \Big/ (n-1) \; (1 \le i \le a).$$

We will use it in rest of this chapter.

Note that the LSEs of μ and γ_i $(1 \le i \le a)$ are, respectively,

$$\tilde{\mu} = (1/N)\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij} \quad \text{and} \quad \tilde{\gamma}_i = \tilde{\mu}_{i.} - \tilde{\mu} \tag{4.1.2.9}$$

where $N = an$ and $\tilde{\mu}_{i.} = \bar{y}_{i.} = (1/n)\sum_{j=1}^{n} y_{ij}$.

4.1.2.1 Efficiency and Robustness

To evaluate the efficiency and robustness of the MMLEs given in (4.1.2.6), we consider a normal distribution and a very broad range of long-tailed symmetric distributions and samples containing data anomalies as follows:

(1) Normal $N(0,\sigma^2)$

The family (4.1.2.1) with

(2) $p = 5$, (3) $p = 3.5$, (4) $p = 2.5$, (5) $p = 2$

Outlier models: $(n-r)$ x_i come from $N(0,\sigma^2)$ and r (we do not know which) come from

(6) $N(0,4\sigma^2)$, (7) $N(0,16\sigma^2)$; $r = [0.5 + 0.1n]$ (integer value).

Mixture models:

(8) $0.90N(0,\sigma^2) + 0.10N(0,4\sigma^2)$, (9) $0.90N(0,\sigma^2) + 0.10N(0,16\sigma^2)$

(10) Student's *t* distribution with 2 *df* (degrees of freedom),

(11) Cauchy distribution,

(12) Slash (Normal/Uniform) distribution

Note that, models (1)-(9) have finite mean and variance, (10) has finite mean but non-existent variance, and (11)-(12) have non-existent mean and variance, as said earlier.

From the resulting N values of the MMLEs and LSEs, we computed their means and variances. The results of MMLEs are given in Table 4.1 while those of LSEs are given in Table 4.2.

Table 4.1: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\mu}$, $\hat{\sigma}$, $\hat{\gamma}_i$ $(1 \le i \le a)$ and the summation of $\hat{\gamma}_i$'s $(1 \le i \le a)$ for long tail symmetric family; $n = 10$.

Model	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\sum_{i=1}^{a} \hat{\gamma}_i$
1	-0.001	0.949	0.006	-0.004	0.002	-0.004	0.000
	[0.270]	[0.146]	[0.799]	[0.805]	[0.768]	[0.811]	
2	-0.002	0.922	0.001	0.000	-0.005	0.004	0.000
	[0.245]	[0.173]	[0.697]	[0.730]	[0.724]	[0.728]	
3	-0.001	0.894	-0.003	0.001	-0.005	0.007	0.000
	[0.227]	[0.185]	[0.678]	[0.666]	[0.658]	[0.655]	
4	0.000	0.839	-0.002	0.002	-0.001	0.000	0.000
	[0.195]	[0.197]	[0.555]	[0.557]	[0.561]	[0.557]	
5	0.001	0.740	0.001	0.001	-0.004	0.002	0.000
	[0.143]	[0.185]	[0.428]	[0.437]	[0.430]	[0.425]	
6	0.001	0.913	0.000	-0.003	-0.001	0.003	0.000
	[0.240]	[0.151]	[0.705]	[0.721]	[0.728]	[0.702]	
7	0.000	0.741	-0.001	0.003	-0.002	0.001	0.000
	[0.139]	[0.123]	[0.421]	[0.408]	[0.430]	[0.422]	
8	0.001	0.917	0.004	-0.001	-0.002	0.000	0.000
	[0.243]	[0.171]	[0.751]	[0.722]	[0.724]	[0.713]	
9	0.001	0.757	-0.001	0.002	-0.002	0.001	0.000
	[0.152]	[0.208]	[0.449]	[0.446]	[0.451]	[0.454]	
10	0.001	1.495	0.001	-0.003	0.000	0.003	0.000
	[0.558]	[1.028]	[1.697]	[1.643]	[1.685]	[1.682]	
11	-0.002	2.248	0.004	-0.006	0.005	-0.003	0.000
	[1.232]	[5.340]	[3.548]	[3.546]	[3.728]	[3.549]	
12	-0.002	3.076	0.011	-0.016	0.020	-0.014	0.000
	[2.211]	[9.003]	[6.387]	[6.619]	[6.664]	[6.712]	

* *Variances are given in brackets*

Table 4.2: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of LSEs $\tilde{\mu}$, $\tilde{\sigma}$, $\tilde{\gamma}_i$ $(1 \le i \le a)$ and the summation of $\tilde{\gamma}_i$'s $(1 \le i \le a)$ for long tail symmetric family; $n = 10$.

Model	$\tilde{\mu}$	$\tilde{\sigma}$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{\gamma}_3$	$\tilde{\gamma}_4$	$\sum_{i=1}^{a} \tilde{\gamma}_i$
1	0.000	0.994	-0.007	0.001	0.003	0.002	0.002
	[0.245]	[0.140]	[0.764]	[0.755]	[0.752]	[0.746]	
2	-0.001	0.989	0.000	0.000	0.002	-0.002	-0.002
	[0.251]	[0.197]	[0.754]	[0.739]	[0.756]	[0.742]	
3	0.000	0.984	-0.004	-0.005	0.007	0.002	0.002
	[0.248]	[0.271]	[0.757]	[0.765]	[0.759]	[0.737]	
4	0.000	0.973	-0.001	-0.001	0.002	0.000	0.000
	[0.253]	[0.518]	[0.764]	[0.731]	[0.747]	[0.762]	
5	-0.001	0.930	0.004	0.000	-0.004	0.000	0.000
	[0.245]	[1.212]	[0.750]	[0.738]	[0.725]	[0.706]	
6	-0.003	0.991	-0.001	0.001	-0.003	0.002	0.002
	[0.248]	0.192	0.734	0.750	0.748	0.763	
7	0.000	0.973	-0.004	0.003	0.000	0.002	0.002
	[0.251]	[0.476]	[0.743]	[0.762]	[0.735]	[0.747]	
8	0.000	0.989	-0.003	0.000	0.003	0.000	0.000
	[0.248]	[0.220]	[0.732]	[0.774]	[0.752]	[0.754]	
9	0.000	0.968	-0.004	0.001	0.004	0.000	0.000
	[0.252]	[0.657]	[0.739]	[0.771]	[0.755]	[0.759]	
10	-0.004	2.534	-0.004	0.011	0.004	-0.011	-0.011
	[3.261]	[66.172]	[10.506]	[9.164]	[8.409]	[10.627]	
11	-1.758	100.241	18.039	-1.767	-17.001	0.729	0.729
	[3.37E+06]	[1.35E+08]	[1.92E+07]	[3.97E+06]	[138E+07]	[3.56E+06]	
12	-1.629	47.405	1.922	-0.770	2.119	-3.271	-3.271
	[1.92E+05]	[7.66E+06]	[2.24E+05]	[2.77E+05]	[2.28E+05]	[1.57E+06]	

** Variances are given in brackets*

When the efficiencies of LSEs and MMLEs are compared, the estimators obtained by the method of MML are observed to be on the whole enormously more efficient than the LSEs and give less baised results as well. For the models (10)-(12) with non-existance variance, the differences between MMLEs and LSEs become very striking. The summation of $\hat{\gamma}_i$'s$(1 \le i \le a)$ is zero for each model, however, the summation of $\tilde{\gamma}_i$'s fails to be zero for models (10)-(12). Furthermore, the variances of the LSEs explode for distributions (10)-(12) because their influence functions, unlike the MMLEs, are not bounded. This is disastrous for machine data processing.

4.1.2.2 Linear Contrasts

Besides the overall block differences being examined by the variance ratio F statistics, it is advisable to construct linear contrasts to capture comparison of different combinations of block means. In this chapter we consider the linear contrast

$$\eta = \sum_{i=1}^{a} \ell_i \gamma_i = \sum_{i=1}^{a} \ell_i \mu_i \quad (\mu_i = \mu + \gamma_i), \quad \sum_{i=1}^{a} \ell_i = 0, \qquad (4.1.2.2.1)$$

where we assume without loss of generality that η is a standardized linear contrast, i.e., $\sum_{i=1}^{a} \ell_i^2 = 1$; ℓ_i $(1 \le i \le a)$ are constant coefficients. In order to construct all the possible $(a-1)$ standardized orthogonal linear constrasts, we use Helmert transformation:

$$\begin{aligned} \eta_1 &= (\mu_1 - \mu_2)/\sqrt{2} \\ \eta_2 &= (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6} \qquad (4.1.2.2.2) \\ &\vdots \\ \eta_{a-1} &= (\mu_1 + \mu_2 + \ldots + \mu_{a-1} - (a-1)\mu_a) \,/\, \sqrt{a(a-1)}. \end{aligned}$$

Note that, two contrasts $\eta_1 = \sum_{i=1}^{a} \ell_{1i} \mu_i$ and $\eta_2 = \sum_{i=1}^{a} \ell_{2i} \mu_i$ are orthogonal if $\sum_{i=1}^{a} \ell_{1i} \ell_{2i} = 0$. The constrasts in (4.1.2.2.2) are all orthogonal to one another and to the mean vector

$$(\mu_1 + \mu_2 + \ldots + \mu_a) \,/ a .$$

The LSE of the linear contrast η is obtained by replacing $\mu_i = \mu + \gamma_i$ in (4.1.2.2.1) by $\bar{y}_i$ $(1 \le i \le a)$:

$$\tilde{\eta} = \sum_{i=1}^{a} \ell_i \bar{y}_i .$$

Here, the variance of $\bar{y}_i$'s $(1 \le i \le a)$ is σ^2/n which is estimated by s^2/n and s^2 is the pooled sample variance:

$$s^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_i\right)^2 \Big/ \left[a\left(n-1\right)\right] = \sum_{i=1}^{a} s_i^2/a .$$

For independent (or uncorrelated) $\bar{y}_i$'s $(1 \le i \le a)$, the variance of $\tilde{\eta}$ is

$$Var\left(\tilde{\eta}\right) = \sum_{i=1}^{a} \ell_i^{\,2}\, Var\left(\bar{y}_i\right) = \frac{s^2}{n}\sum_{i=1}^{a} \ell_i^{\,2} = \frac{s^2}{n},$$

since η is a standardized linear contrast.

The MMLE of η is

$$\hat{\eta} = \sum_{i=1}^{a} \ell_i\, \hat{\mu}_i ,$$

and the variance of $\hat{\eta}$ is

$$Var\left(\hat{\eta}\right) = \sum_{i=1}^{a} \ell_i^{\,2}\, Var\left(\hat{\mu}_i\right) = \sigma^2 \sum_{i=1}^{a} \ell_i^{\,2}\, V_i , \quad Var\left(\hat{\mu}_i\right) = V_i\, \sigma^2$$

for independent (or uncorrelated) $\hat{\mu}_i$'s $(1 \le i \le a)$.

4.1.2.3 Hypothesis Testing

In order to study the differences between block means, our hypothesis to test is $H_0 : \eta = 0$ against $H_a : \eta = d \ne 0$.

The distribution of the test statistic based on the LSEs, $t = \sqrt{n}\left(\tilde{\eta}/s\right)$, is Student's t with *df* $v = a(n-1)$ under the null hypothesis if the distribution of $e_{ij}\ (1 \le i \le a, 1 \le j \le n)$ is normal.

We define the test statistic $T = \hat{\eta}\big/\sqrt{Var\left(\hat{\eta}\right)}$ by using the MMLEs. While the null-distribution of T is $N(0,1)$ for large *n*, it is referred to Student's t with $v = a(n-1)$ *df* for small *n*.

In general, the non-null distribution of both test statistics, t and T, are referred to noncentral Student's t with $v = a(n-1)$ *df* and noncentrality parameter $\lambda^2 = n\left(\eta/\sigma\right)^2$.

For illustration, we do hypothesis testing with standardized orthogonal linear constrasts for four blocks, i.e., $a=4$. Linear constrasts, obtained by Helmert transformation, are

$$\eta_1 = (\mu_1 - \mu_2)/\sqrt{2} \qquad (4.1.2.3.1)$$

$$\eta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6} \qquad (4.1.2.3.2)$$

and

$$\eta_3 = (\mu_1 + \mu_2 + \mu_3 - 3\mu_4)/(2\sqrt{3}). \qquad (4.1.2.3.3)$$

The test statistic for testing $H_0 : \eta_c = 0$ is

$$T_c = \frac{\hat{\eta}_c}{\sqrt{Var(\hat{\eta}_c)}} \text{ , } c = 1,2,3 \text{ .} \qquad (4.1.2.3.4)$$

Large values of T_c lead to the rejection of H_0 in favor of $H_1 : \eta_c > 0$, $c = 1,2,3$.

Since $\partial \ln L^*/\partial\mu = 0$ is asymptotically equivalent to $\partial \ln L/\partial\mu = 0$ and has the form

$$\frac{\partial \ln L^*}{\partial \mu_i} = \frac{2pm_i}{k\sigma^2}\{\hat{\mu}_i(\sigma) - \hat{\mu}_i\} = 0 \text{ , } m_i = \sum_{j=1}^{n} \tilde{\beta}_{ij} \quad (1 \le i \le a) \text{ ,}$$

the estimator $\hat{\mu}_i(\sigma) = \sum_{j=1}^{n} \tilde{\beta}_{ij} y_{i(j)} \Big/ m_i$ is conditionally (σ known) the *MVB* estimator of $\hat{\mu}_i = \mu + \gamma_i$ and is normally distributed with variance

$$Var(\hat{\mu}_i(\sigma)) = \frac{k}{2p(m_i/n)}\frac{\sigma^2}{n} \text{ , } 1 \le i \le a \text{ .}$$

Therefore, the variance of $\hat{\eta}_c$ is

$$Var(\hat{\eta}_c) = \sum_{i=1}^{a} \ell_{ci}^{\ 2}\, Var(\hat{\mu}_i) = \frac{\sigma^2}{n}\frac{k}{2p}\sum_{i=1}^{a}\frac{\ell_{ci}^{\ 2}}{(m_i/n)} \quad \text{ , } c = 1,2,3 \text{ ,}$$

since the blocks are independent. Of course, the above results apply when *n* is large in which case (m_i/n) is a constant in the limit when *n* tends to infinity.

The test statistics for testing $H_0 : \eta_c = 0$ versus $H_a : \eta_c > 0$ becomes

$$T_c = \frac{\sqrt{n}\,\hat{\eta}}{\sigma} \sqrt{\frac{2p}{k\sum_{i=1}^{a} \frac{\ell_{ci}^{\,2}}{(m_i/n)}}} \;,\; c = 1,2,3\,; \qquad (4.1.2.3.5)$$

σ to be raplaced by $\hat{\sigma}$. The *MVB* estimator of μ_i $(1 \le i \le a)$ is

$$MVB(\mu_i) = \frac{(p-3/2)(p+1)}{np(p-1/2)}\sigma^2 . \qquad (4.1.2.3.6)$$

Note that we take $p = 16.5$ and $k = 30$, as in Chapter 2.

Rejection probabilities of the test statistics for testing

$$H_0 : \eta_3 = \mu_1 + \mu_2 + \mu_3 - 3\mu_4 = 0 \quad \text{against}$$
$$H_a : \eta_3 > 0$$

for different μ_i 's $(1 \le i \le 4)$ with samples from the distributions (1)-(12) in section 4.1.2.1 are obtained by simulation. We generated $nn = [100{,}000/n]$ (integer value) samples (consisting of independently distributed observations) of size n=10 from each of the models (1)-(12). The observations generated from models (6)-(9) were divided by suitable constants to make their variances equal to σ^2. Different number of iterations (e.g. 2, 3 and 5 iterations) were carried out. We observed that three iterations are enough to give stable results. The rejection probabilites using the test statistics with estimated *MVB*

$$M\hat{V}B(\mu_i) = \frac{(p-3/2)(p+1)}{np(p-1/2)}\hat{\sigma}^2$$

are given in Table 4.3 and Table 4.4, respectively. Without loss of generality, σ is taken to be 1 and the Type I error is assumed to be 0.05.

Table 4.3: Values of the power for long-tailed symmetric family estimators in which $\hat{\sigma}$ is directly used to calculate the test statistics under different values of μ; $n = 10$.

	μ_i											
	0.0		0.2		0.4		0.6		0.8		1.0	
Model	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE
1*	0.045	0.052	0.17	0.19	0.41	0.44	0.69	0.71	0.89	0.90	0.98	0.98
2	0.046	0.052	0.18	0.20	0.42	0.44	0.69	0.72	0.89	0.91	0.97	0.98
3	0.047	0.054	0.18	0.20	0.42	0.45	0.70	0.73	0.89	0.91	0.97	0.98
4	0.045	0.050	0.18	0.21	0.43	0.49	0.71	0.77	0.90	0.93	0.97	0.99
5	0.041	0.049	0.18	0.23	0.46	0.56	0.75	0.84	0.91	0.96	0.97	0.99
6	0.047	0.051	0.17	0.18	0.40	0.44	0.67	0.71	0.87	0.90	0.96	0.97
7	0.046	0.042	0.14	0.16	0.33	0.38	0.57	0.65	0.78	0.85	0.90	0.95
8	0.046	0.049	0.17	0.19	0.41	0.45	0.67	0.71	0.87	0.90	0.96	0.97
9	0.040	0.042	0.13	0.16	0.32	0.38	0.55	0.63	0.76	0.82	0.89	0.93
10	0.133	0.041	0.23	0.10	0.34	0.21	0.47	0.36	0.59	0.52	0.72	0.69
11	0.292	0.032	0.33	0.06	0.37	0.11	0.42	0.16	0.46	0.25	0.50	0.34
12	0.320	0.036	0.34	0.05	0.38	0.08	0.41	0.12	0.44	0.16	0.48	0.23

* The power values of the normal theory test are little bit smaller only because it s Type I error is smaller than the T_c-test. It should, in fact, be little bit larger for a common Type I error.

Table 4.4: Values of the power for long-tailed symmetric family estimators in which $M\hat{V}B(\mu_i)$ is used to calculate the test statistics under different values of μ; $n = 10$.

	μ_i											
	0.0		0.2		0.4		0.6		0.8		1.0	
Model	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE
1	0.049	0.056	0.17	0.19	0.41	0.43	0.69	0.69	0.89	0.89	0.98	0.98
2	0.048	0.052	0.17	0.19	0.42	0.44	0.70	0.72	0.89	0.91	0.97	0.98
3	0.048	0.054	0.17	0.19	0.41	0.45	0.71	0.74	0.89	0.91	0.97	0.98
4	0.044	0.048	0.18	0.21	0.43	0.50	0.71	0.78	0.90	0.94	0.97	0.99
5	0.041	0.048	0.17	0.23	0.45	0.57	0.74	0.85	0.91	0.97	0.97	0.99
6	0.046	0.053	0.16	0.17	0.41	0.45	0.68	0.72	0.88	0.90	0.96	0.98
7	0.045	0.043	0.14	0.15	0.33	0.39	0.57	0.65	0.78	0.86	0.91	0.96
8	0.047	0.050	0.17	0.18	0.41	0.45	0.67	0.71	0.87	0.90	0.96	0.97
9	0.044	0.043	0.14	0.17	0.31	0.37	0.56	0.66	0.77	0.85	0.89	0.95
10	0.134	0.042	0.22	0.11	0.34	0.22	0.47	0.38	0.60	0.56	0.72	0.72
11	0.289	0.040	0.33	0.07	0.37	0.11	0.42	0.21	0.45	0.31	0.50	0.42
12	0.312	0.033	0.35	0.06	0.38	0.09	0.41	0.14	0.45	0.20	0.47	0.27

Since distributions (10)-(12) have infinite variance, μ has to be very large for the non-centrality parameter $n(\eta/\sigma)^2$ to be appreciably greater than zero to yield a value of the power greater than the Type I error. Therefore, for distributions (10)-(12), we took $\mu = 0.0, 0.4, 0.8, 1,2, 1.6, 2.0$. The test based on LSEs had enormously large Type I error. For sake of comparison, we obtained their 95% points by simulation. The critical value of LSEs is 2.5 for model (10), 6.0 for model (11), and 7.0 for model (12). The results obtained with these critical values are tabulated in Table 4.5.

Table 4.5: Values of the power with simulated critical values using LSEs; $n=10$.

	μ_i											
	0.0		0.4		0.8		1.2		1.6		2.0	
Model	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE
10	0.052	0.042	0.17	0.21	0.39	0.52	0.65	0.81	0.83	0.94	0.91	0.98
11	0.052	0.034	0.06	0.10	0.08	0.25	0.09	0.46	0.12	0.65	0.16	0.78
12	0.053	0.031	0.06	0.09	0.07	0.17	0.08	0.29	0.08	0.46	0.10	0.60

Although the Type I error of the test using the LSEs is now 0.05, the power values are considerably less than the values for the MMLEs. Using LSEs is in vain unless the distributions are normal or near-normal.

The test statistics for testing

$$\begin{aligned} &H_0 : \eta_1 = \eta_2 = \eta_3 = 0 \quad \text{against} \\ &H_a : at\ least\ one\ \eta_c \neq 0 \quad (c = 1,2,3) \end{aligned} \tag{4.1.2.3.7}$$

can be obtained with MMLEs and LSEs, respectively, as

$$T^2 = T_1^2 + T_2^2 + T_3^2 \text{ and}$$

$$t^2 = t_1^2 + t_2^2 + t_3^2$$

where

$$T_c^2 = \left\{ \frac{\hat{\eta}_c}{\widehat{MVB}(\mu_i)} \right\}^2 \frac{2pn}{k\sum_{i=1}^{a} \frac{\ell_{ci}^2}{(m_i/n)}} \quad \text{and} \tag{4.1.2.3.8}$$

$$t_c^2 = n\left(\frac{\tilde{\eta}}{s}\right)^2, \quad c = 1,2,3\,. \tag{4.1.2.3.9}$$

The statistics (4.1.2.3.8) and (4.1.2.3.9) are distributed as chisquare with 1 *df*. Thus, T^2 is distributed as χ_3^2.

Table 4.6: Values of the power for testing $H_0 : \eta_1 = \eta_2 = \eta_3 = 0$ with long-tailed symmetric family; $n = 10$.

	μ_i											
	0.0		0.2		0.4		0.6		0.8		1.0	
Model	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE	LSE	MMLE
1	0.042	0.054	0.18	0.20	0.59	0.60	0.92	0.91	0.99	0.99	1.00	1.00
2	0.042	0.053	0.18	0.20	0.59	0.62	0.91	0.92	0.99	1.00	1.00	1.00
3	0.037	0.047	0.18	0.21	0.59	0.62	0.93	0.94	0.99	1.00	1.00	1.00
4	0.037	0.041	0.18	0.22	0.61	0.67	0.92	0.94	0.99	1.00	1.00	1.00
5	0.035	0.048	0.19	0.26	0.65	0.76	0.93	0.97	0.99	1.00	1.00	1.00
6	0.041	0.047	0.17	0.19	0.56	0.60	0.90	0.92	0.99	0.99	1.00	1.00
7	0.027	0.032	0.12	0.15	0.45	0.51	0.81	0.87	0.96	0.98	0.99	1.00
8	0.039	0.045	0.17	0.19	0.57	0.61	0.90	0.92	0.99	0.99	1.00	1.00
9	0.033	0.033	0.13	0.15	0.45	0.52	0.80	0.86	0.96	0.97	0.99	1.00
10	0.274	0.036	0.33	0.08	0.48	0.24	0.67	0.47	0.83	0.72	0.91	0.87
11	0.726	0.028	0.75	0.04	0.76	0.10	0.79	0.20	0.81	0.33	0.85	0.49
12	0.791	0.024	0.79	0.04	0.80	0.07	0.82	0.12	0.84	0.19	0.86	0.29

Results obtained by using MMLEs are enormously superior and no knowledge of p in (4.1.2.1) as such is assumed. This is indeed very advantageous for machine data processing, and also theoretically. It is interesting to note that the emprical values of the Type I error are smaller than the presumed value for extreme non-normal symmetric distributions. Therefore, the corresponding values of the power will be larger than those in the tables above when the Type I errors are 0.05.

***Remark*:** The above methods are readily applicable to situations when the shape parameter p does not have the same value from block to block. This is due to the fact that we are estimating the coefficients α_i and β_i $(1 \le i \le n)$ from each block.

4.1.3 Generalized Logistic Distribution

We now assume that the errors e_{ij} in (4.1.1) are iid and its distributions is a member of the family of Generalized Logistic ($b > 0$)

$$f(e) = \frac{b}{\sigma} \frac{\exp(-e/\sigma)}{\{1+\exp(-e/\sigma)\}^{b+1}}, \qquad -\infty < e < \infty, \qquad (4.1.3.1)$$

where σ is scale and b is a shape parameter.

The likelihood function L for the one-way classification fixed effects model is

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{a} \prod_{j=1}^{n} \frac{\exp(-z_{ij})}{\{1+\exp(-z_{ij})\}^{b+1}} \; ;$$

$$z_{ij} = (y_{ij} - \mu - \gamma_i)/\sigma .$$

To estimate μ, γ_i $(1 \le i \le a)$ and σ, the following likelihood equations are obtained:

$$\frac{\partial \ln L}{\partial \mu} = \frac{N}{\sigma} - \frac{(b+1)}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} g(z_{ij}) = 0, \qquad (4.1.3.2)$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma} \sum_{j=1}^{n} g(z_{ij}) = 0, \qquad (4.1.3.3)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} z_{ij} - \frac{(b+1)}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} z_{ij}\, g(z_{ij}) = 0 \qquad (4.1.3.4)$$

and the function $g(z)$ is given by

$$g(z) = e^{-z}/(1+e^{-z}).$$

In order to derive MLEs, one should solve the likelihood equations (4.1.3.2) to (4.1.3.4). However, the involvement of the function $g(z)$ makes it difficult to work out solutions.

By linearizing $g(z)$ as before, i.e.,

$$g(z_{i(j)}) \cong \alpha_j + \beta_j z_{i(j)} \quad (1 \le j \le n); \qquad (4.1.3.5)$$

$$\alpha_j = (1+e^t+te^t)/(1+e^t)^2, \; \beta_j = e^t/(1+e^t)^2, \; t = t_{(j)} = E\{z_{i(j)}\}, \; (4.1.3.6)$$

and incorporating (4.1.3.5) and (4.1.3.6) in likelihood equations (4.1.3.2)-(4.1.3.4), we obtain the modified likelihood equations $\partial \ln L^* / \partial \mu = 0$, $\partial \ln L^* / \partial \gamma_i = 0$ and $\partial \ln L^* / \partial \sigma = 0$ as in Senoglu and Tiku (2001). Here,

$$z_{i(j)} = \left(y_{i(j)} - \mu - \gamma_i \right) / \sigma \quad (i = 1,2,\ldots,a;\, j = 1,2,\ldots,n),$$

are the ordered variates and $y_{i(1)} \le y_{i(2)} \le \ldots \le y_{i(n)}$ $(1 \le i \le a)$ are the order stastistics of the n observations.

The values of $t_{(j)}$ can be estimated as in Chapter 3. In this case, however, we replace $t_{(j)}$ by $t_{i(j)}$ since we have $T_{0i} = median\{ y_{ij} \}$ and $S_{0i} = 1.483 median\{ | y_{ij} - T_{0i} | \}$ $(1 \le i \le a)$ for each block:

$$\tilde{t}_{i(j)} = \left(y_{i(j)} - T_{0i} \right) / S_{0i} \,.$$

Therefore, the initial estimates of α_j and β_j, respectively, are obtained by replacing $t_{(j)}$ by $\tilde{t}_{i(j)}$ $(1 \le i \le n)$ and denoted by $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_{ij}$, respectively.

The modified maximum likelihood equations can be written as

$$\frac{d \ln L}{d\mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{N}{\sigma} - \frac{(b+1)}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} \left(\tilde{\alpha}_{ij} - \tilde{\beta}_{ij}\, z_{i(j)} \right) = 0 \,, \tag{4.1.3.7}$$

$$\frac{\partial \ln L}{\partial \gamma_i} \cong \frac{\partial \ln L^*}{\partial \gamma_i} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma} \sum_{j=1}^{n} \left(\tilde{\alpha}_{ij} - \tilde{\beta}_{ij}\, z_{i(j)} \right) = 0 \text{ and} \tag{4.1.3.8}$$

$$\frac{d \ln L}{d\sigma} \cong \frac{d \ln L^*}{d\sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} z_{i(j)} - \frac{(b+1)}{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{n} z_{i(j)} \left(\tilde{\alpha}_{ij} - \tilde{\beta}_{ij}\, z_{i(j)} \right) = 0 \,. \tag{4.1.3.9}$$

The equations (4.1.3.7)-(4.1.3.9) have explicit solutions while (4.1.3.2)-(4.1.3.4) do not, and the solutions are the MMLEs,

$$\hat{\mu} = (1/m) \sum_{i=1}^{a} m_i \hat{\mu}_{i.} \,,\; \hat{\gamma}_i = \hat{\mu}_{i.} - \hat{\mu} \text{ and } \hat{\sigma} = \left\{ -B + \sqrt{\left(B^2 + 4NC \right)} \right\} / (2N) \,, \tag{4.1.3.10}$$

$$\Delta = \sum_{i=1}^{a} \sum_{j=1}^{n} \Delta_{ij} \,,\; \Delta_{ij} = \tilde{\alpha}_{ij} - (b+1)^{-1} \,,\; m_i = \sum_{j=1}^{n} \tilde{\beta}_{ij} \,,\; m = \sum_{i=1}^{a} m_i \,,$$

$$B=\sum_{i=1}^{a} B_i\ ,\ C=\sum_{i=1}^{a} C_i\ ,\ B_i=(b+1)\sum_{j=1}^{n}\Delta_{ij}\left(y_{i(j)}-K_i\right), \tag{4.1.3.11}$$

$$C_i=(b+1)\sum_{j=1}^{n}\tilde{\beta}_{ij}\left(y_{i(j)}-K_i\right)^2=(b+1)\left\{\sum_{j=1}^{n}\tilde{\beta}_{ij}y_{i(j)}{}^2-m_iK_i{}^2\right\}, \tag{4.1.3.12}$$

$$K_i=(1/m_i)\sum_{j=1}^{n}\tilde{\beta}_{ij}y_{i(j)}\ \text{ and }\ \hat{\mu}_{i.}=(1/m_i)\left\{\sum_{j=1}^{n}\tilde{\beta}_{ij}y_{i(j)}-\hat{\sigma}\sum_{j=1}^{n}\Delta_{ij}\right\}.$$

As explained earlier, it is more convenient to take

$$\hat{\sigma}=\sqrt{\sum_{i=1}^{a}\hat{\sigma}_i^2\Big/a}\ , \tag{4.1.3.13}$$

where $\hat{\sigma}_i=\left\{-B_i+\sqrt{\left(B_i{}^2+4nC_i\right)}\right\}\Big/2\sqrt{n(n-1)}$, B_i and C_i $(1\le i\le a)$ are as in (4.1.3.11) and (4.1.3.12), respectively.

Note that for each block, $(b+1)$ is initially estimated by $1/\left(1-\tilde{\tilde{w}}_i\right)$ $(1\le i\le a)$:

$$\tilde{\tilde{w}}_i=(1/n)\sum_{j=1}^{n}\tilde{w}_{ij}\ ,\ \ \tilde{w}_{ij}=e^{\tilde{t}_{i(j)}}\Big/\left(1+e^{\tilde{t}_{i(j)}}\right)=1\Big/\left(1+e^{-\tilde{t}_{i(j)}}\right)\quad 1\le j\le n,\quad 1\le i\le a\ .$$

See section 3.2 for details.

We generated $nn=[100{,}000/n]$ (integer value) random samples of size $n=10$ from generalized logistic family with various shape parameter values in order to study the MMLEs in (4.1.3.10)-(4.1.3.13). From the resulting nn values of MMLEs, we computed their means and variances which are given in Table 4.8. For $b\neq 1$, the random observations were multiplied by $[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}]^{1/2}$ in order to make the variances the same as the variance of logistic distribution (b = 1), which is equal to $2\Psi'(1)\sigma^2=3.2898\sigma^2$. As explained in Chapter 3, $\hat{\mu}$ is estimating the scaled median while $\hat{\sigma}$ is estimating the scale parameter σ which is taken to be 1 without loss of generality. Note that the scaled median is

$$-\ln(2^{1/b}-1)[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}]^{1/2} \tag{4.1.3.14}$$

where μ is taken to be zero without loss of generality. The values of the scaled median in (4.1.3.14) are given in Table 4.7

Table 4.7: The values of scaled median for different shape parameters.

Model	Scaled Median
$b = 0.5$	-0.777
$b = 1$	0.000
$b = 2$	1.056
$b = 4$	2.174
$b = 6$	2.819
$b = 8$	3.268

Table 4.8: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\mu}$, $\hat{\sigma}$, $\hat{\gamma}_i$ $(1 \le i \le a)$ and the summation of $\hat{\gamma}_i$'s $(1 \le i \le a)$ for generalized logistic family; $n = 10$.

Model	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\sum_{i=1}^{a}\hat{\gamma}_i$
$b = 0.5$	-0.815	0.965	0.005	-0.007	-0.001	0.005	0.002
	[0.793]	[0.200]	[2.457]	[2.459]	[2.454]	[2.413]	
$b = 1$	-0.005	1.012	0.008	0.003	-0.002	0.004	0.013
	[0.861]	[0.211]	[2.653]	[2.576]	[2.616]	[2.594]	
$b = 2$	1.076	1.024	0.004	-0.004	0.018	0.004	0.022
	[0.900]	[0.225]	[2.682]	[2.692]	[2.773]	[2.715]	
$b = 4$	2.208	1.028	0.004	0.009	0.010	0.004	0.027
	[0.908]	[0.235]	[2.733]	[2.801]	[2.791]	[2.780]	
$b = 6$	2.857	1.022	0.018	-0.002	0.007	0.005	0.029
	[0.915]	[0.235]	[2.747]	[2.700]	[2.770]	[2.766]	
$b = 8$	3.308	1.024	0.005	0.011	0.000	0.014	0.029
	[0.882]	[0.245]	[2.738]	[2.728]	[2.790]	[2.766]	

** Variances are given in brackets*

Although we do not assume a given value for the shape parameter *b*, the MMLEs are unbiased (almost) in estimating the scaled median and shape parameter. Remember that we are using a fixed effects model where $\sum_{i=1}^{a} \gamma_i = 0$ without loss of generality . The MMLEs $\hat{\gamma}_i$ $(1 \le i \le a)$ satisfy this condition (almost). Note that the *x*-observations need not be multiplied by $\sqrt{2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}}$ if one wants to estimate the median rather than the scaled-median.

4.1.3.1 Linear Contrasts and Hypothesis Testing

To study linear contrasts, we assume four blocks ($a = 4$) and use Helmert transformation in order to construct standardized orthogonal linear constrasts:

$$\eta_1 = (\mu_1 - \mu_2)/\sqrt{2}$$

$$\eta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6} \text{ and}$$

$$\eta_3 = (\mu_1 + \mu_2 + \mu_3 - 3\mu_4)/2\sqrt{3}\,.$$

The test statistics for testing

$$H_0 : \eta_c = 0 \quad \text{against} \quad H_a : \eta_c > 0,\; c = 1,2,3 \qquad (4.1.3.1.1)$$

is

$$T_c = \frac{\hat{\eta}_c}{\sqrt{Var(\hat{\eta}_c)}}\;,\; c = 1,2,3\,. \qquad (4.1.3.1.2)$$

Since $\partial \ln L^* / \partial \mu = 0$ is asymptotically equivalent to $\partial \ln L / \partial \mu = 0$ and has the form

$$\frac{\partial \ln L^*}{\partial \mu_i} = \frac{m_i (b+1)}{\sigma^2}\{\hat{\mu}_i(\sigma) - \hat{\mu}_i\} = 0,\; m_i = \sum_{j=1}^{n} \widetilde{\beta}_{ij}\,,$$

the estimator $\hat{\mu}_i(\sigma) = (1/m_i)\left\{\sum_{j=1}^{n} \widetilde{\beta}_{ij} y_{i(j)} - \sigma \sum_{j=1}^{n} \Delta_{ij}\right\}$ is conditionally (σ known) the *MVB* estimator of $\hat{\mu}_i = \mu + \gamma_i$ and is normally distributed with variance

$$Var(\hat{\mu}_i(\sigma)) = \frac{1}{(m_i/n)(b+1)} \frac{\sigma^2}{n}.$$

As before, the above results are true asymptotically because $\lim_{n\to\infty} m_i/n$ is a constant in the limit.

Since the blocks are independent, the variance of $\hat{\eta}_c$ ($c = 1,2,3$) is

$$Var(\hat{\eta}_c) = \sum_{i=1}^{a} \ell_{ci}^{\,2}\, Var(\hat{\mu}_i) = \frac{\sigma^2}{n} \frac{1}{(b+1)} \sum_{i=1}^{a} \frac{\ell_{ci}^{\,2}}{(m_i/n)}$$

where $\ell_1 = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \end{pmatrix}$,

$$\ell_2 = \left(1/\sqrt{6} \quad 1/\sqrt{6} \quad -2/\sqrt{6} \quad 0\right) \text{ and}$$

$$\ell_2 = \left(1/(2\sqrt{3}) \quad 1/(2\sqrt{3}) \quad 1/(2\sqrt{3}) \quad -3/(2\sqrt{3})\right).$$

However, since we are estimating $(b+1)$ for each block by $1/\left(1-\tilde{\tilde{w}}_i\right)$ $(1 \le i \le a)$, we have

$$Var(\hat{\eta}_c) = \frac{\sigma^2}{n} \sum_{i=1}^{a} \frac{(1-\tilde{\tilde{w}}_i)\,\ell_{ci}^{\;2}}{(m_i/n)}. \quad (4.1.3.1.3)$$

Therefore, the test statistics for testing $H_0 : \eta_c = 0$ versus $H_a : \eta_c > 0$ becomes

$$T_c = \frac{\sqrt{n}\,\hat{\eta}_c}{\sigma} \sqrt{\frac{1}{\sum_{i=1}^{a} \frac{(1-\tilde{\tilde{w}}_i)\,\ell_{ci}^{\;2}}{(m_i/n)}}}\ , \ c = 1,2,3\,; \quad (4.1.3.1.4)$$

σ is replaced by $\hat{\sigma}$.

The power results under different alternatives for testing $H_0 : \eta_c = 0$ against $H_a : \eta_c > 0$, $(c = 1,2,3)$ with 0.05 Type I error are tabulated in Tables 4.9-4.11.

Table 4.9: Values of the power for testing $H_0 : \eta_1 = 0$ under different alternatives for distributions with different shape parameters *b*.

n = 10	μ_i					
Model	0.0	0.2	0.4	0.6	0.8	1.0
b = 0.5	0.056	0.11	0.19	0.30	0.43	0.57
b = 1	0.057	0.14	0.27	0.44	0.63	0.79
b = 2	0.060	0.16	0.34	0.55	0.75	0.89
b = 4	0.063	0.18	0.39	0.61	0.82	0.93
b = 6	0.064	0.19	0.39	0.64	0.83	0.94
b = 8	0.063	0.20	0.40	0.64	0.84	0.94

Table 4.10: Values of the power for testing $H_0 : \eta_2 = 0$ under different alternatives for distributions with different shape parameters *b*.

$n = 10$	μ_i					
Model	0.0	0.2	0.4	0.6	0.8	1.0
$b = 0.5$	0.059	0.12	0.23	0.36	0.52	0.66
$b = 1$	0.063	0.16	0.31	0.53	0.73	0.88
$b = 2$	0.061	0.2	0.4	0.66	0.85	0.95
$b = 4$	0.059	0.21	0.47	0.72	0.89	0.97
$b = 6$	0.061	0.22	0.47	0.73	0.91	0.97
$b = 8$	0.058	0.22	0.48	0.76	0.91	0.98

Table 4.11: Values of the power for testing $H_0 : \eta_3 = 0$ under different alternatives for distributions with different shape parameters *b*.

$n = 10$	μ_i					
Model	0.0	0.2	0.4	0.6	0.8	1.0
$b = 0.5$	0.063	0.10	0.15	0.23	0.33	0.44
$b = 1$	0.061	0.13	0.22	0.34	0.49	0.64
$b = 2$	0.059	0.14	0.28	0.44	0.62	0.76
$b = 4$	0.062	0.16	0.31	0.50	0.67	0.82
$b = 6$	0.065	0.16	0.32	0.52	0.70	0.84
$b = 8$	0.061	0.16	0.33	0.53	0.72	0.85

It is observed that under $H_0 : \eta_k = 0$ ($k = 1, 2, 3$), we obtained power values slightly greater than the Type I error. Therefore, it is decided to obtain the 95% point for $H_0 : \eta_3 = 0$ by simulation and the results are given in Table 4.12.

Table 4.12: Values of the power for testing $H_0 : \eta_3 = 0$ with simulated critical values.

$n = 10$	μ_i					
Model	0.0	0.2	0.4	0.6	0.8	1.0
$b = 0.5$	0.051	0.09	0.14	0.20	0.29	0.41
$b = 1$	0.051	0.11	0.20	0.31	0.45	0.61
$b = 2$	0.054	0.12	0.24	0.41	0.58	0.74
$b = 4$	0.051	0.14	0.28	0.46	0.64	0.80
$b = 6$	0.050	0.14	0.29	0.47	0.67	0.81
$b = 8$	0.054	0.15	0.29	0.48	0.68	0.83

When Table 4.11 and 4.12 are compared, a decrease in power values under $H_a : \eta_3 > 0$ is observed which is expected. However, it should be noted that the decrease is very small. In addition, note that the critical value obtained by simulation is 1.80, which is very close to the original 95% point.

The test statistic for testing

$$\begin{aligned} &H_0 : \eta_1 = \eta_2 = \eta_3 = 0 \quad \text{against} \\ &H_a : at\ least\ one\ \eta_c \neq 0 \quad c = 1,2,3. \end{aligned} \tag{4.1.3.1.5}$$

can be formulated by using MMLEs:

$$T^2 = T_1^2 + T_2^2 + T_3^2 . \tag{4.1.3.1.6}$$

where

$$T_c^2 = \frac{n\hat{\eta}_c^2}{\hat{\sigma}^2}\left\{\sum_{i=1}^{a} \frac{(1-\tilde{\tilde{w}}_i)\,\ell_{ci}^{\;2}}{m_i/n}\right\}^{-1} , \quad c = 1,2,3 . \tag{4.1.3.1.7}$$

Since T_c^2's ($c = 1,2,3$) are distributed as chisquare with 1 *df*, the test statistic T^2 is distributed as χ_3^2. Table 4.13 shows the power values of the test (4.1.3.1.5) by using T^2 given in (4.1.3.1.6).

Table 4.13: The table of power values for the test $H_0 : \eta_1 = \eta_2 = \eta_3 = 0$ with generalized logistic family estimators.

n = 10	μ_i					
Model	0.0	0.2	0.4	0.6	0.8	1.0
b = 0.5	0.049	0.07	0.12	0.20	0.34	0.49
b = 1	0.055	0.09	0.19	0.36	0.59	0.79
b = 2	0.061	0.10	0.25	0.50	0.76	0.91
b = 4	0.059	0.13	0.29	0.58	0.83	0.95
b = 6	0.063	0.13	0.31	0.61	0.85	0.97
b = 8	0.064	0.13	0.32	0.62	0.85	0.97

It can be seen that T^2 provides a powerful test and is succesful in attaining the presumed Type I error (almost). It is important to note that the generalized logistic has considerable amount of skewness for $b \geq 4$ (Tiku an Akkaya, 2004, Appendix 2D). Therefore, increasing the sample size will result in more accurate approximations, especially for $b \geq 4$.

4.1.3.2 Non-Identical Blocks

In some real life situatitions, the errors $e_{ij}\ (j=1,2,\ldots,n)$ in the i^{th} block might come from Generalized Logistic with shape parameter b_i and scale parameter σ. The shape parameters $b_i\ (1\le i\le a)$ are not necessarily equal. Senoglu and Tiku (2002) assumed that all b_i are known and gave a solution to estimate and test the block effects. Our method extends to the situation when $b_i\ (1\le i\le a)$ are not known because it uses the estimators of the shape parameters and not their true values. The test is exactly the same as (4.1.3.1.6), T_c^2 being the statistic (4.1.3.1.7) . The only restriction is that all $b_i\ (1\le i\le a)$ are either ≥ 1 or ≤ 1.

4.2 Two-Way Classification and Interaction

Consider the two-way classification model

$$y_{ijl}=\mu+\gamma_i+\delta_j+\tau_{ij}+e_{ijl} \quad (1\le i\le a;1\le j\le c,1\le l\le n), \qquad (4.2.1)$$

where μ is a constant, γ_i is the effect due to i^{th} block, δ_j is the effect due to j^{th} column, and τ_{ij} is the interaction between the i^{th} block and j^{th} column . The random errors e_{ijl} are iid. Without loss of generality, we assume that it is a fixed effects model where

$$\sum_{i=1}^{a}\gamma_i=\sum_{j=1}^{c}\delta_i=\sum_{i=1}^{a}\tau_{ij}=\sum_{j=1}^{c}\tau_{ij}=0\,. \qquad (4.2.2)$$

In the following subsections, various types of distribution families are assumed for the random errors e_{ijl} in order to study the estimators.

4.2.1 Normal Distribution

Traditionally, e_{ijl} have been assumed to be iid normal $N(0,\sigma^2)$, where the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^{anl} \prod_{i=1}^{a}\prod_{j=1}^{c}\prod_{l=1}^{n} \exp\left\{-\left(y_{ijl}-\mu-\gamma_i-\delta_j-\tau_{ij}\right)^2 \big/ 2\sigma^2\right\} .$$

Solving the maximum likelihood equations $\partial \ln L / \partial\mu = 0$, $\partial \ln L / \partial\gamma_i = 0$ $(i = 1,2,\ldots,a)$, $\partial \ln L / \partial\delta_j = 0$ $(j = 1,2,\ldots,c)$ and $\partial \ln L / \partial\sigma = 0$ leads to the following MLEs:

$$\hat{\hat{\mu}} = \bar{y}_{...},\ \hat{\hat{\gamma}}_i = \bar{y}_{i..} - \bar{y}_{...}\ (1 \le i \le a),\ \hat{\hat{\delta}}_j = \bar{y}_{.j.} - \bar{y}_{...}\ (1 \le j \le c),$$

$$\hat{\hat{\tau}}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \text{ and}$$

$$\hat{\hat{\sigma}}^2 = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n}\left(y_{ijl} - \bar{y}_{ij.}\right)^2 \Big/ ac\,(n-1) = \sum_{i=1}^{a}\sum_{j=1}^{c} s_{ij}^2 \Big/ ac$$

where $\bar{y}_{i..} = (1/cn)\sum_{j=1}^{c}\sum_{l=1}^{n} y_{ijl}$ is the mean of the i^{th} block, $\bar{y}_{.j.} = (1/an)\sum_{i=1}^{a}\sum_{l=1}^{n} y_{ijl}$ is the mean of the j^{th} column, $\bar{y}_{ij.} = (1/n)\sum_{l=1}^{n} y_{ijl}$ is the mean of the $(i,j)^{th}$ cell and $\bar{y}_{...} = (1/acn)\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n} y_{ijl}$ is the overall mean.

The main interest in two-way classification with interaction is to test the null hypotheses

$$H_{01} : \gamma_i = 0, \forall i,\ H_{02} : \delta_j = 0, \forall j \text{ and } H_{03} : \tau_{ij} = 0, \forall i, j .$$

Fisher decomposition of the total sum of squares is

$$\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n}\left(y_{ijl} - \bar{y}_{..}\right)^2 = cn\sum_{i=1}^{a}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 + an\sum_{j=1}^{c}\left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2$$
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{c}\left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n}\left(y_{ijl} - \bar{y}_{ij.}\right)^2$$

or

$$S_T^2 = S_{block}^2 + S_{column}^2 + S_{interaction}^2 + S_{error}^2 .$$

They are instrumental in constructing the test statistics. Note that S_{block}^2 , S_{column}^2, $S_{interaction}^2$ and S_{error}^2 on the right hand side are the sums of squares due to 'blocks', 'columns', 'interactions' and 'error', respectively. If e_{ijl} are iid normal $N(0,\sigma^2)$, then S_{block}^2 / σ^2 , S_{column}^2 / σ^2 and $S_{interaction}^2 / \sigma^2$ are distributed as chi-sqaure with $(a-1)$, $(c-1)$ and $(a-1)(c-1)$ degrees of freedoms under the null hypotheses

$$H_{01} : \gamma_1 = \gamma_2 = \cdots = \gamma_a = 0 ,$$

$$H_{02} : \delta_1 = \delta_2 = \cdots = \delta_c = 0 ,$$

$$H_{03} : \tau_{11} = \tau_{12} = \cdots = \tau_{ac} = 0 ,$$

respectively. Since $S_{error}^2 / ac(n-1)$ is independently distributed as chi-sqaure with $ac(n-1)$ degrees of freedom when errors are normal, the test statistics to test $H_{0k}, k = 1,2,3$ are respectively

$$F_1 = s_{block}^2 / s_{error}^2 ,$$

$$F_2 = s_{column}^2 / s_{error}^2 ,$$

$$F_3 = s_{interaction}^2 / s_{error}^2$$

where

$$s_{block}^2 = S_{block}^2 / (a-1) ,\ s_{column}^2 = S_{column}^2 / (c-1) \text{ and } s_{error}^2 = S_{error}^2 / ac(n-1)$$

are called 'block', 'column' and 'error' mean sums of squares, respectively. Large values of F lead to the rejection of $H_{0k}, k = 1,2,3$, respectively, in favor of

$$H_{11} : At\ least\ one\ \gamma_i \neq 0 \ \ (1 \le i \le a) ,$$

$$H_{12} : At\ least\ one\ \delta_j \neq 0 \ \ (1 \le j \le c) ,$$

$$H_{13} : At\ least\ one\ \tau_{ij} \neq 0 .$$

The null distributions of the test statistics is central F with the denominator degrees of freedom $\nu_2 = ac(n-1)$. The numerator degrees of freedom ν_1 of F_1 is $(a-1)$, of F_2 is $(c-1)$ and of F_3 is $(a-1)(c-1)$. Under the alternatives, the distributions of the test statistics $F_k, (k = 1,2,3)$ are non-central F with (ν_1, ν_2) degrees of freedom and non-centrality parameter

$$\lambda_1^2 = cn \sum_{i=1}^{a} \gamma_i^2 / \sigma^2 \quad \text{for } F_1,$$

$$\lambda_2^2 = an \sum_{j=1}^{c} \delta_j^2 / \sigma^2 \quad \text{for } F_2 \text{ and}$$

$$\lambda_3^2 = n \sum_{i=1}^{a} \sum_{j=1}^{c} \tau_{ij}^2 / \sigma^2 \quad \text{for } F_3 .$$

4.2.2 Long-Tailed Symmetric Family

Assuming that the distribution of error terms in (4.2.1) belongs to long-tailed symmetric family, we obtain the following likelihood function:

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{a} \prod_{j=1}^{c} \prod_{l=''}^{n} \left\{1 + \frac{z_{ij}^2}{k}\right\}^{-p} \quad ; \qquad (4.2.2.1)$$

$z_{ij} = e_{ij} / \sigma = (y_{ij} - \mu - \gamma_i - \delta_j - \tau_{ij}) / \sigma \quad (i = 1,2,\ldots,a;\ j = 1,2,\ldots,c; l = 1,2,\ldots,n)$, $N = acn$.

The likelihood equations for estimating μ, $\gamma_i \quad (1 \le i \le a)$, $\delta_j \quad (1 \le j \le c)$, τ_{ij} and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{\sigma k} \sum_{i=1}^{a} \sum_{j=1}^{c} \sum_{l=1}^{n} g(z_{ijl}) = 0 , \qquad (4.2.2.2)$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{2p}{\sigma k} \sum_{j=1}^{c} \sum_{l=1}^{n} g(z_{ijl}) = 0 , \qquad (4.2.2.3)$$

$$\frac{\partial \ln L}{\partial \delta_j} = \frac{2p}{\sigma k} \sum_{i=1}^{a} \sum_{l=1}^{n} g(z_{ijl}) = 0 , \qquad (4.2.2.4)$$

$$\frac{\partial \ln L}{\partial \tau_{ij}} = \frac{2p}{\sigma k} \sum_{l=1}^{n} g(z_{ijl}) = 0 , \qquad (4.2.2.5)$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{\sigma k}\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n} z_{ij}\, g\left(z_{ij}\right) = 0 \qquad (4.2.2.6)$$

where the function $g(z)$ is given by

$$g(z) = z\big/\left(1 + z^2/k\right). \qquad (4.2.2.7)$$

Due to the intractability of the equations (4.2.2.2) to (4.2.2.6) which have no explicit solutions, we utilize the MMLEs obtained as follows:

The linear approximations we consider are

$$g\left(z_{ij(l)}\right) \cong \alpha_{ijl} + \beta_{ijl} z_{ij(l)} \quad (1 \le i \le a; 1 \le j \le c; 1 \le l \le n) \qquad (4.2.2.8)$$

where

$$\alpha_{ijl} = \frac{(1/k)t_{ij(l)}}{\left\{1 + (1/k)t_{ij(l)}^2\right\}^2} \text{ and } \beta_{ijl} = \frac{1}{\left\{1 + (1/k)t_{ij(l)}^2\right\}^2},\ t_{ij(l)} = E\left\{z_{ij(l)}\right\}\ (k = 30). \qquad (4.2.2.9)$$

Incorporating (4.2.2.9) in (4.2.2.2)-(4.2.2.7) gives the modified likelihood equations $\partial \ln L^* / \partial \mu = 0, \partial \ln L^* / \partial \gamma_i = 0$ and $\partial \ln L^* / \partial \sigma = 0$.

We disregard the ordering of z_{ijl} as before and take $\tilde{t}_{ijl} = (y_{ijl} - T_{0ij}) / S_{0ij}$ $(j = 1, \ldots, n)$ as the initial estimate of t_{ijl}, where $T_{0ij} = median\{y_{ijl}\}$ and $S_{0ij} = 1.483\, median\{\left| y_{ijl} - T_{0ij} \right|\}$ $(1 \le i \le a; 1 \le j \le c)$ for the $(i, j)^{th}$ cell.

The initial estimates of α_{ijl} and β_{ijl} are obtained by replacing t_{ijl} by $\tilde{t}_{ijl}$, and the resulting coefficients are denoted by $\tilde{\alpha}_{ijl}$ and $\tilde{\beta}_{ijl}$, respectively.

The resulting MMLEs of μ, γ_i $(1 \le i \le a)$, δ_j $(1 \le j \le c)$, τ_{ij} and σ are

$$\hat{\mu} = \sum_{i=1}^{a}\sum_{j=1}^{c} m_{ij}\, \hat{\mu}_{ij.} \Big/ \sum_{i=1}^{a}\sum_{j=1}^{c} m_{ij},\ \hat{\gamma}_i = \hat{\mu}_{i..} - \hat{\mu},\ \hat{\delta}_j = \hat{\mu}_{.j.} - \hat{\mu}, \qquad (4.2.2.10)$$

$$\hat{\tau}_{ij} = \hat{\mu}_{ij.} - \hat{\mu}_{i..} - \hat{\mu}_{.j.} + \hat{\mu} \text{ and } \hat{\sigma} = \left\{B + \sqrt{\left(B^2 + 4NC\right)}\right\} \Big/ (2N) \qquad (4.2.2.11)$$

where

$$B = \sum_{i=1}^{a}\sum_{j=1}^{c} B_{ij},\ C = \sum_{i=1}^{a}\sum_{j=1}^{c} C_{ij},$$

$$B_{ij} = \frac{2p}{k}\sum_{l=1}^{n} \tilde{\alpha}_{ijl}\left(y_{ijl} - \hat{\mu}_{ij.}\right),\ C_{ij} = \frac{2p}{k}\sum_{l=1}^{n} \tilde{\beta}_{ijl}\left(y_{ijl} - \hat{\mu}_{ij.}\right)^2, \qquad (4.2.2.12)$$

$$\hat{\mu}_{i..} = \sum_{j=1}^{c} m_{ij}\,\hat{\mu}_{ij.} \Big/ \sum_{j=1}^{c} m_{ij}\ ,\ \hat{\mu}_{.j.} = \sum_{i=1}^{a} m_{ij}\,\hat{\mu}_{ij.} \Big/ \sum_{i=1}^{a} m_{ij}\ ,$$

$$\hat{\mu}_{ij.} = (1/m_{ij})\sum_{l=1}^{n} \tilde{\beta}_{ijl}\, y_{ijl} \quad \text{and} \quad m_{ij} = \sum_{l=1}^{n} \tilde{\beta}_{ijl}\ .$$

Note that, as in previous chapters, $k = 30\,(p = 16.5)$.

A more convenient form of the MMLE of σ is

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{a}\sum_{j=1}^{c} \hat{\sigma}_{ij}^{2} \Big/ ac}\ , \qquad (4.2.2.13)$$

where $\hat{\sigma}_{ij} = \left\{B_{ij} + \sqrt{\left(B_{ij}^{\ 2} + 4nC_{ij}\right)}\right\} \Big/ 2\sqrt{n(n-1)}$, B_{ij} and C_{ij} $(1 \le i \le a; 1 \le j \le c)$ are given in (4.2.2.12). Notice that $\hat{\sigma}$ given in (4.2.2.13) has the same form as the corresponding LSE, namely,

$$s = \sqrt{\sum_{i=1}^{a}\sum_{j=1}^{c} s_{ij}^{2} \Big/ ac}\ ; \quad s_{ij}^{2} = \sum_{l=1}^{n}\left(y_{ijl} - \bar{y}_{ij.}\right)^{2} \Big/ (n-1) \ \ (1 \le i \le a; 1 \le j \le c)\,.$$

The LSEs of μ , γ_i $(1 \le i \le a)$ δ_j $(1 \le j \le c)$ and τ_{ij} are

$$\tilde{\mu} = (1/N)\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n} y_{ijl}\ , \qquad \tilde{\gamma}_i = \tilde{\mu}_{i..} - \tilde{\mu}\,, \qquad \tilde{\delta}_j = \tilde{\mu}_{.j.} - \tilde{\mu}\,,$$

(4.2.2.14)

$$\tilde{\tau}_{ij} = \tilde{\mu}_{ij.} - \tilde{\mu}_{i..} - \tilde{\mu}_{.j.} + \tilde{\mu}$$

where

$$N = acn,\ \tilde{\mu}_{i..} = (1/cn)\sum_{j=1}^{c}\sum_{l=1}^{n} y_{ijl}\ ,\ \tilde{\mu}_{.j.} = (1/an)\sum_{i=1}^{a}\sum_{l=1}^{n} y_{ijl} \ \text{ and } \ \tilde{\mu}_{ij.} = (1/n)\sum_{l=1}^{n} y_{ijl}\ .$$

4.2.2.1 Efficiency and Robustness

In order to study the efficiency and robustness of MMLEs given in (4.2.2.10), (4.2.2.11) and (4.2.2.13), we use the distributions (1)-(12) given in section 4.1.2.1. The MMLEs and LSEs are computed and their means and variances are given in Table 4.14 and Table 4.15. Note that no iteration is done for calculating the MMLEs. We have 3 blocks and 3 columns in our design. The sample size n in each cell is taken to be 4 in Table 4.14 while it is 8 in Table 4.15.

Table 4.14: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\mu}$, $\hat{\sigma}$, $\hat{\tau}_{11}$ and LSEs $\tilde{\mu}$, $\tilde{\sigma}$, $\tilde{\tau}_{11}$ for long tail symmetric family; $n=4$.

Model	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\sigma}$	$\tilde{\sigma}$	$\hat{\tau}_{11}$	$\tilde{\tau}_{11}$
1	-0.001	-0.001	0.927	0.989	0.003	0.004
	[0.121]	[0.112]	[0.077]	[0.071]	[0.487]	[0.443]
2	0.001	0.001	0.910	0.987	0.002	0.004
	[0.110]	[0.110]	[0.094]	[0.101]	[0.443]	[0.440]
3	-0.001	0.000	0.897	0.985	0.003	0.004
	[0.106]	[0.110]	[0.108]	[0.134]	[0.421]	[0.442]
4	0.000	0.000	0.853	0.971	-0.004	-0.003
	[0.092]	[0.112]	[0.125]	[0.260]	[0.376]	[0.462]
5	-0.001	0.001	0.772	0.926	0.002	0.003
	[0.073]	[0.110]	[0.133]	[0.562]	[0.289]	[0.434]
6	0.000	0.001	0.926	0.990	-0.001	0.000
	[0.123]	[0.113]	[0.078]	[0.073]	[0.490]	[0.445]
7	-0.001	-0.001	0.928	0.991	0.001	0.001
	[0.121]	[0.113]	[0.079]	[0.074]	[0.485]	[0.443]
8	-0.002	-0.001	0.907	0.985	-0.002	-0.002
	[0.107]	[0.107]	[0.095]	[0.107]	[0.452]	[0.444]
9	-0.002	-0.002	0.796	0.965	-0.003	-0.004
	[0.077]	[0.111]	[0.160]	[0.293]	[0.307]	[0.451]
10	0.001	0.013	1.666	2.552	0.007	0.008
	[0.329]	[2.631]	[1.167]	[68.397]	[1.292]	[11.964]
11	0.004	-0.814	3.709	35.668	0.026	-0.001
	[0.27E+01]	[0.17E+05]	[0.11E+03]	[0.59E+06]	[0.10E+02]	[0.59E+05]
12	0.002	-9.064	4.638	112.461	0.004	19.968
	[0.32E+01]	[0.39E+07]	[0.99E+02]	[0.14E+09]	[0.14E+02]	[0.16E+08]

** Variances are given in brackets*

Table 4.15: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\mu}$, $\hat{\sigma}$, $\hat{\tau}_{11}$ and LSEs $\tilde{\mu}$, $\tilde{\sigma}$, $\tilde{\tau}_{11}$ for long tail symmetric family; $n=8$.

Model	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\sigma}$	$\tilde{\sigma}$	$\hat{\tau}_{11}$	$\tilde{\tau}_{11}$
1	0.001	0.001	0.950	0.997	0.002	0.001
	[0.117]	[0.109]	[0.068]	[0.063]	[0.484]	[0.454]
2	-0.001	0.000	0.922	0.993	0.000	0.000
	[0.108]	[0.112]	[0.083]	[0.095]	[0.442]	[0.461]
3	0.000	0.001	0.900	0.992	0.001	0.003
	[0.101]	[0.110]	[0.087]	[0.125]	[0.408]	[0.448]
4	0.001	0.003	0.844	0.981	0.000	-0.001
	[0.084]	[0.112]	[0.094]	[0.281]	[0.344]	[0.442]
5	0.000	0.001	0.749	0.946	0.000	0.000
	[0.065]	[0.106]	[0.092]	[0.615]	[0.262]	[0.434]
6	-0.002	-0.002	0.912	0.996	0.001	0.001
	[0.105]	[0.112]	[0.070]	[0.090]	[0.422]	[0.450]
7	-0.001	0.000	0.725	0.982	0.003	0.001
	[0.058]	[0.110]	[0.057]	[0.211]	[0.235]	[0.441]
8	0.001	0.001	0.919	0.994	-0.001	0.000
	[0.107]	[0.112]	[0.080]	[0.100]	[0.427]	[0.434]
9	0.000	0.000	0.764	0.979	0.002	0.000
	[0.068]	[0.112]	[0.105]	[0.315]	[0.274]	[0.444]
10	0.001	0.002	1.534	2.686	-0.001	-0.004
	[0.270]	[1.537]	[0.564]	[49.181]	[1.031]	[6.068]
11	0.001	-0.022	2.452	50.489	0.001	1.065
	[0.62E+00]	[0.58E+05]	[0.45E+01]	[0.42E+07]	[0.24E+01]	[0.77E+05]
12	0.003	0.839	3.286	103.073	0.000	-4.894
	[0.11E+01]	[0.55E+06]	[0.65E+01]	[0.39E+08]	[0.43E+01]	[0.29E+07]

* *Variances are given in brackets*

As expected, the variance decreases by about half when we increase the sample size from 4 to 8.

After comparing LSEs and MMLEs, it can be seen that the estimators obtained by the method of MML are on the whole considerably more efficient than the LSEs. While the variances and the means of the LSEs explode for distributions (10)-(12) with non-existance variance, MMLEs give much more stable results for the same scenarios because of their bounded influence functions.

4.2.3 Generalized Logistic Distribution

Suppose that the distribution of error terms in (4.2.1) is a member of generalized logistic family. The likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{a}\prod_{j=1}^{c}\prod_{l=1}^{n} \frac{\exp(-z_{ij})}{\{1+\exp(-z_{ij})\}^{b+1}} \;; \qquad (4.2.3.1)$$

where $z_{ij} = e_{ij}/\sigma = (y_{ij} - \mu - \gamma_i - \delta_j - \tau_{ij})/\sigma$ $(1 \le i \le a; 1 \le j \le c; 1 \le l \le n)$ and $N = acn$.

The likelihood equations to estimate μ , γ_i $(1 \le i \le a)$, δ_j $(1 \le j \le c)$, τ_{ij} and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{N}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n} g(z_{ijl}) = 0\,, \qquad (4.2.3.2)$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{nc}{\sigma} - \frac{(b+1)}{\sigma}\sum_{j=1}^{c}\sum_{l=1}^{n} g(z_{ijl}) = 0\,, \qquad (4.2.3.3)$$

$$\frac{\partial \ln L}{\partial \delta_j} = \frac{na}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{a}\sum_{l=1}^{n} g(z_{ijl}) = 0\,, \qquad (4.2.3.4)$$

$$\frac{\partial \ln L}{\partial \tau_{ij}} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{l=1}^{n} g(z_{ijl}) = 0\,, \qquad (4.2.3.5)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{l=1}^{n} z_{ijl} - \frac{(b+1)}{\sigma}\sum_{i=1}^{a}\sum_{j=1}^{n}\sum_{l=1}^{n} z_{ijl}\, g(z_{ijl}) = 0 \qquad (4.2.3.6)$$

where the function $g(z)$ is

$$g(z) = e^{-z}/\left(1+e^{-z}\right). \qquad (4.2.3.7)$$

The solutions of the equations (4.2.3.2)-(4.2.3.6) are the MLEs. However, these equations have no explicit solutions because of the non-linear function $g(z)$. Although solving them by iteration is possible, Barnett (1966), Lee et al. (1980) and Vaughan (1992) show that iteration may lead to multiple roots, nonconvergence (especially when data contains outliers; see Puthenpura and Sinha (1986) for details), or convergence to wrong values.

As in Senoglu and Tiku (2001), we propose modified likelihood equations $\partial \ln L^* / \partial \mu = 0$, $\partial \ln L^* / \partial \gamma_i = 0$, $\partial \ln L^* / \partial \delta_j = 0$, $\partial \ln L^* / \partial \tau_{ij} = 0$ and $\partial \ln L^* / \partial \sigma = 0$ which are obtained after linearizing $g(z)$:

$$g\left(z_{ij(l)}\right) \cong \alpha_{ijl} - \beta_{ijl} z_{ij(l)}; \tag{4.2.3.8}$$

$$\alpha_{ijl} = \left(1 + e^t + te^t\right) / \left(1 + e^t\right)^2, \ \beta_{ijl} = e^t / \left(1 + e^t\right)^2, \ t = t_{ij(l)} = E\{z_{ij(l)}\}, \tag{4.2.3.9}$$

$z_{ij(l)} = \left(y_{ij(l)} - \mu - \gamma_i - \delta_j - \tau_{ij}\right) / \sigma \ (1 \le i \le a; 1 \le j \le c; 1 \le l \le n)$.

Here, $z_{ij(l)}$ are the ordered variates and $y_{ij(1)} \le y_{ij(2)} \le \ldots \le y_{ij(n)}$ $(1 \le i \le a; 1 \le j \le c)$ are the order stastistics of the *n* observations in each cell.

The values of $t_{ij(l)}$ are estimated as in section 4.1.3, however, in this case we replace $\widetilde{t}_{i(j)}$ by $\widetilde{t}_{ij(l)}$ since we have $T_{0ij} = median\{y_{ijl}\}$ and $S_{0ij} = 1.483 median\{|\, y_{ijl} - T_{0ij}\,|\}$ $(1 \le i \le a; 1 \le j \le c)$ for each cell:

$$\widetilde{t}_{ij(l)} = \left(y_{ji(l)} - T_{0ij}\right) / S_{0ij}. \tag{4.2.3.10}$$

Thus, replacing t_{ijl} by $\widetilde{t}_{ijl}$, the initial estimates of α_{ijl} and β_{ijl} are obtained and they are denoted by $\widetilde{\alpha}_{ijl}$ and $\widetilde{\beta}_{ijl}$, respectively.

The modified likelihood equations obtained by linearizing $g(z)$ have explicit solutions and the solutions are the MMLEs,

$$\hat{\mu} = \mu_{...} - \hat{\sigma} \sum_{i=1}^{a} \sum_{j=1}^{c} \Delta_{ij} \Big/ \sum_{i=1}^{a} \sum_{j=1}^{c} m_{ij}, \ \hat{\gamma}_i = \hat{\mu}_{i..} - \hat{\mu}_{...}, \ \hat{\delta}_j = \hat{\mu}_{.j.} - \hat{\mu}_{...}, \tag{4.2.3.11}$$

$$\hat{\tau}_{ij} = \hat{\mu}_{ij.} - \hat{\mu}_{i..} - \hat{\mu}_{.j.} + \hat{\mu}_{...} \text{ and } \hat{\sigma} = \left\{-B + \sqrt{\left(B^2 + 4NC\right)}\right\} / (2N) \tag{4.2.3.12}$$

where

$$B = \sum_{i=1}^{a} \sum_{j=1}^{c} B_{ij}, \ C = \sum_{i=1}^{a} \sum_{j=1}^{c} C_{ij},$$

$$B_{ij} = (b+1) \sum_{l=1}^{n} \Delta_{ijl} \left(y_{ij(l)} - \hat{\mu}_{ij}\right), \ C_{ij} = (b+1) \sum_{l=1}^{n} \widetilde{\beta}_{ijl} \left(y_{ij(l)} - \hat{\mu}_{ij}\right)^2, \tag{4.2.3.13}$$

$$\hat{\mu}_{i..} = \sum_{j=1}^{c} m_{ij} \hat{\mu}_{ij.} \Big/ \sum_{j=1}^{c} m_{ij}, \ \hat{\mu}_{.j.} = \sum_{i=1}^{a} m_{ij} \hat{\mu}_{ij.} \Big/ \sum_{i=1}^{a} m_{ij},$$

$$\hat{\mu}_{...} = \sum_{i=1}^{a} \sum_{j=1}^{c} m_{ij} \hat{\mu}_{ij.} \Big/ \sum_{i=1}^{a} \sum_{j=1}^{c} m_{ij}, \ \hat{\mu}_{ij.} = \left(1/m_{ij}\right) \sum_{l=1}^{n} \widetilde{\beta}_{ijl} y_{ij(l)},$$

$\Delta_{ijl} = \tilde{\alpha}_{ijl} - (b+1)^{-1}$, $\Delta_{ij} = \sum_{l=1}^{n} \Delta_{ijl}$ and $m_{ij} = \sum_{l=1}^{n} \tilde{\beta}_{ijl}$.

Writing $\hat{\sigma}_{ij} = \left\{ B_{ij} + \sqrt{\left({B_{ij}}^2 + 4nC_{ij} \right)} \right\} \Big/ 2\sqrt{n(n-1)}$, B_{ij} and C_{ij} $(1 \le i \le a;$ $1 \le j \le c)$ in (4.2.3.13), a more convenient form of the MMLE of σ can be obtained:

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{a} \sum_{j=1}^{c} \hat{\sigma}_{ij}^2 \Big/ ac} \, , \tag{4.2.3.14}$$

For each $(i, j)^{th}$ cell, $(b+1)$ is initially estimated by $1 \big/ \left(1 - \tilde{\tilde{w}}_{ij} \right)$:

$$\tilde{\tilde{w}}_{ij} = (1/n) \sum_{l=1}^{n} \tilde{w}_{ijl} \, , \quad \tilde{w}_{ijl} = e^{\tilde{t}_{ij(l)}} \Big/ \left(1 + e^{\tilde{t}_{ij(l)}} \right) = 1 \Big/ \left(1 + e^{-\tilde{t}_{ij(l)}} \right) \quad (1 \le i \le a; 1 \le j \le c) \, .$$

See section 3.2 for details.

Note that since complete sums are invariant to ordering, we can ignore the ordering.

4.2.3.1 Efficiency and Robustness

We study the efficiency and robustness of the MMLEs of a two-way classification model with interaction whose error comes from generalized logistic family with various shape parameters. In order to do this, we carried out simulation studies based on *N*=[100.000/*n*] Monte Carlo runs. The means and variances of the MMLEs are given in Table 4.16 where the sample size for each cell n is taken to be 4 and in Table 4.17 where n is 8. We do this with 3 blocks and 3 columns. The number of iterations is 5. The results for block and column effects being essentially the same as the block effects in one-way classification, the variances of the corresponding estimators are not reproduced.

Random errors e_{ijl} $(1 \le i \le a; 1 \le j \le c; 1 \le l \le n)$ generated when $b \ne 1$ were multiplied by $\left[2\Psi'(1) / \{ \Psi'(b) + \Psi'(1) \} \right]^{1/2}$ so that the variances of e_{ijl} are always the same as when *b* = 1 (logistic distribution), i.e., $2\Psi'(1)\sigma^2 = 3.2898\sigma^2$ where σ is taken to be one without loss of generality.

Table 4.16: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\sigma}$ and $\hat{\tau}_{11}$ for generalized logistic family; $n = 4$.

Model	$\hat{\sigma}$	$\hat{\tau}_{11}$
$b = 0.5$	1.012	-0.001
	[0.119]	[1.460]
$b = 1$	1.046	0.002
	[0.115]	[1.563]
$b = 2$	1.061	0.001
	[0.127]	[1.591]
$b = 4$	1.064	0.007
	[0.139]	[1.601]
$b = 6$	1.063	0.006
	[0.142]	[1.566]
$b = 8$	1.066	0.007
	[0.147]	[1.583]

** Variances are given in brackets*

Table 4.17: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance* of MMLEs $\hat{\sigma}$ and $\hat{\tau}_{11}$ for generalized logistic family; $n = 8$.

Model	$\hat{\sigma}$	$\hat{\tau}_{11}$
$b = 0.5$	0.976	-0.007
	[0.096]	[1.439]
$b = 1$	1.020	0.007
	[0.098]	[1.557]
$b = 2$	1.034	0.014
	[0.102]	[1.596]
$b = 4$	1.034	0.004
	[0.109]	[1.632]
$b = 6$	1.030	0.010
	[0.112]	[1.612]
$b = 8$	1.035	0.006
	[0.117]	[1.607]

** Variances are given in brackets*

The estimators $\hat{\sigma}$ and $\hat{\tau}_{11}$ are unbiased (almost) for each shape parameter (unknown to us). When we compare the results with different sizes, it is observed that increasing the sample size from $n = 4$ to $n = 8$ reduces the variances by about one-half, especially of $\hat{\tau}_{ij}$.

CHAPTER 5

ROBUST LINEAR REGRESSION

In many practical situations, a variable *Y* depends on another variable *X*. However, *Y* cannot be measured exactly and is subject to a measurement error but *X* can supposedly be measured without error. For example, *Y* is the blood pressure of an elderly person and *X* is his/her age. A very important statistical problem is to model the dependence of *Y* on *X*. Usually, a functional relationship

$$y = \eta(x) + e \tag{5.1}$$

is assumed; $\eta(x)$ is a mathematical function involving certain unknown parameters and e is a random error having a particular distribution. Given a random sample (y_i, x_i) $(1 \le i \le n)$, the problem is to estimate the unknown parameters in (5.1). The situation which occurs most often in practice is that $\eta(x)$ is linear. The equation

$$y = \theta_0 + \theta_1 x + e \tag{5.2}$$

is called a linear regression model. Experimental data (y_i, x_i) $(1 \le i \le n)$ is available which supposedly follows this model. Thus,

$$y_i = \theta_0 + \theta_1 x_i + e_i , \quad 1 \le i \le n ; \tag{5.3}$$

e_i $(1 \le i \le n)$ are assumed to be iid with mean zero and unknown variance σ^2. In certain situations, e_i will have nonzero mean; we will consider those situations later.

To estimate θ_0, θ_1 and σ^2 (or σ), a very popular method is least squares estimation. The error sum of squares

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$$

is minimized with respect to θ_0 and θ_1. This, in particular, gives the LSE (least square estimator) of θ_1 as

$$\tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{5.4}$$

It is easy to show that

$$Var(\tilde{\theta}_1) = \sigma^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{5.5}$$

Akkaya and Tiku (2008) point out that $\tilde{\theta}_1$ is very vulnerable to the design $(x_1, x_2, \ldots, x_n)$. If, for example, an outlier occurs in the design, $\sum_{i=1}^{n} (x_i - \bar{x})^2$ will become very large in which case $\tilde{\theta}_1$ will appear to be very efficient. That is nonsensical. To rectify the situation, they proposed the reparametrized model

$$y_i = \theta_0 + \theta_1 u_i + e_i , \quad 1 \le i \le n , \tag{5.6}$$

where $u_i = (x_i - \bar{x})/s$, $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ and $n s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

We will show that the LSEs, and other estimators we develop in this chapter, are invariant to location and scale of the design. That is, if x_i are replaced by $a + bx_i$ $(1 \le i \le n)$, the estimators and their variances (and covariances) remain unchanged, *a* and *b* being constants.

The LSEs of θ_0 and θ_1 are now obtained by minimizing

$$\sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 u_i)^2 .$$

That gives

$$\tilde{\theta}_0 = \bar{y} \text{ and } \tilde{\theta}_1 = \sum_{i=1}^{n} u_i y_i \Big/ \sum_{i=1}^{n} u_i^2 \quad \left(\sum_{i=1}^{n} u_i^2 = n \right). \tag{5.7}$$

The LSE of σ^2 is obtained by minimizing $\sum_{i=1}^{n} e_i^2$,

$$\tilde{\sigma}^2 = \sum_{i=1}^{n} \left\{ y_i - \bar{y} - \tilde{\theta}_1 u_i \right\}^2 \Big/ (n-2) \text{ (bias corrected).} \tag{5.8}$$

It is very easy to verify that the LSEs (5.7) - (5.8) are invariant to location and scale of the design. It may be noted that

$$Var\left(\tilde{\theta}_0\right) = \sigma^2 / n \text{ and } Var\left(\tilde{\theta}_1\right) = \sigma^2 / n .$$

Both variances do not depend on the design which is a very useful result.

As said earlier, $\tilde{\sigma}$ is unbiased (asymptotically) and

$$Var(\tilde{\sigma}) \cong \frac{\sigma^2}{2n}\left(1 + \frac{1}{2}\lambda_4\right), \ \lambda_4 = \mu_4 / \mu_2^2 - 3 \ ; \tag{5.9}$$

μ_4 / μ_2^2 is the kurtosis of the underlying distribution. Note that (5.9) is also invariant to the design.

The only assumption for deriving LSEs is that the random errors e_i $(1 \le i \le n)$ are iid with mean zero and variance σ^2. Suppose that the common distribution of e_i is known. We can then try to obtain the MLEs (maximum likelihood estimators) of θ_0 , θ_1 and σ.

Assuming that e_i $(1 \le i \le n)$ are iid normal $N\left(0, \sigma^2\right)$, the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n \exp\left\{ - \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 u_i)^2 / 2\sigma^2 \right\}. \tag{5.10}$$

Solving the maximum likelihood equations $\partial \ln L / \partial \theta_0 = 0, \partial \ln L / \partial \theta_1 = 0$

and $\partial \ln L/\partial\sigma = 0$ gives the MLEs $\hat{\hat{\theta}}_0$, $\hat{\hat{\theta}}_1$ and $\hat{\hat{\sigma}}$. In this case, i.e. when e_i are normal $N(0,\sigma^2)$, the MLEs and LSEs are identical. The Fisher information matrix is

$$I = \begin{bmatrix} \frac{n}{\sigma^2} & 0 & 0 \\ 0 & \frac{n}{\sigma^2} & 0 \\ 0 & 0 & \frac{2n}{\sigma^2} \end{bmatrix}; \tag{5.11}$$

I^{-1} gives the asymptotic variances as

$$Var\left(\hat{\hat{\theta}}_0\right) \cong \sigma^2/n,\ Var\left(\hat{\hat{\theta}}_1\right) \cong \sigma^2/n \text{ and } Var\left(\hat{\hat{\sigma}}\right) \cong \sigma^2/2n. \tag{5.12}$$

In the present situation, however, the first two variances are exact for all *n* as said earlier. In fact, $\hat{\hat{\theta}}_0$ and $\hat{\hat{\theta}}_1$ are the *MVB* estimators. This follows from the fact that

$$\frac{\partial \ln L}{\partial\theta_0} = \frac{n}{\sigma^2}\left(\hat{\hat{\theta}}_0 - \theta_0\right) \text{ and } \frac{\partial \ln L}{\partial\theta_1} = \frac{n}{\sigma^2}\left(\hat{\hat{\theta}}_1 - \theta_1\right).$$

Remark: When the distribution of the random error *e* is non-normal, MLEs are generally elusive. Therefore, we utulize modified maximum likelihood estimation as follows.

5. 1 Long-Tailed Symmetric Distributions

Consider the simple linear regression model (5.6) with random error *e* having one the distributions in the long-tailed symmetric family

$$f(e) = \frac{1}{\sigma\sqrt{k}}\,\frac{1}{\beta(1/2, p-1/2)}\left[1 + \frac{e^2}{k\sigma^2}\right]^{-p}, \qquad -\infty < e < \infty;$$

$k = 2p - 3$, $p \geq 2$. It may be noted that $E(e) = 0$ and $V(e) = \sigma^2$. For $1 \leq p < 2$, $V(e)$ does not exist in which case σ is a scale parameter. Writing $z_i = (y_i - \theta_1 u_i - \theta_0)/\sigma$ $(1 \leq i \leq n)$ and $g(z) = z/\left(1 + z^2/k\right)$, we have the following maximum likelihood equations; $u_i = (x_i - \bar{x})/s$:

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{2p}{\sigma k} \sum_{i=1}^{n} g(z_i) = 0 \tag{5.1.1}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{2p}{\sigma k} \sum_{i=1}^{n} u_i \, g(z_i) = 0 \tag{5.1.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{\sigma k} \sum_{i=1}^{n} z_i \, g(z_i) = 0 . \tag{5.1.3}$$

The likelihood equations (5.1.1) - (5.1.3) do not have explicit solutions since $g(z)$ is a nonlinear function. Vaughan (1992) and Tiku and Suresh (1992) showed that (5.1.1) has multiple roots for all $p < \infty$, hence, calculation of MLEs is problematic. Tiku et al. (2001) proposed modified maximum likelihood estimation as an alternative and showed that MMLEs have all the desirable properties; see also Islam and Tiku (2004) and Akkaya and Tiku (2008) who derive MMLEs for parameters in a multiple linear regression model. Their estimators, however, do not have bounded influence functions.

Here, we derive MMLEs which have bounded influence functions. We obtain such MMLEs by using the linear approximations

$$g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)} \quad (1 \leq i \leq n) \tag{5.1.4}$$

where

$$\alpha_i = \frac{(1/k)t_{(i)}}{\left\{1 + (1/k)t_{(i)}^2\right\}^2} \text{ and } \beta_i = \frac{1}{\left\{1 + (1/k)t_{(i)}^2\right\}^2}, \tag{5.1.5}$$

$$t_{(i)} = E\left\{z_{(i)}\right\} \; (k = 30).$$

Here, we define $z_{(i)} = \left(w_{(i)} - \theta_0\right)/\sigma$ and $w_{(i)} = y_{[i]} - \theta_1 u_{[i]}$ $(1 \leq i \leq n)$ where $z_{(i)}$ are the ordered variates and $\left(y_{[i]}, u_{[i]}\right)$ are the concomitants of $z_{(i)}$.

In order to estimate $t_{(i)}$, we propose its initial estimator as

$$\tilde{t}_{(i)} = \frac{y_{[i]} - T_0 - T_1 u_{[i]}}{S_0} \tag{5.1.6}$$

where

$$T_1 = median\{r_\ell\};\ r_\ell = \frac{y_{\ell+1} - y_\ell}{u_{\ell+1} - u_\ell} \quad (1 \le \ell \le n-1) \tag{5.1.7}$$

is the initial estimator of the regression coefficient θ_1,

$$T_0 = median\{\tilde{w}_i\},\ \tilde{w}_i = y_i - T_1 u_i \ (1 \le i \le n), \tag{5.1.8}$$

is the initial estimator of the intercept θ_0 and

$$S_0 = 1.483\, median\{|\tilde{w}_i - T_0|\} \tag{5.1.9}$$

is the initial estimator of σ. These initial estimators were obtained by noticing that

$$\theta_1 = E(y_{i+1} - y_i)/(u_{i+1} - u_i)\ (1 \le i \le n-1).$$

Since complete sums are invariant to ordering, we use $\tilde{t}_i$ rather than $\tilde{t}_{(i)}$. Replacing t_i by $\tilde{t}_i$ leads to the initial estimates of α_i and β_i. They are $\tilde{\alpha}_i$ and $\tilde{\beta}_i$, respectively. Also, (5.1.4) can be written as

$$g(z_i) \cong \tilde{\alpha}_i + \tilde{\beta}_i z_i \quad (1 \le i \le n). \tag{5.1.10}$$

The solutions of the resulting modified likelihood equations are the following MMLEs:

$$\hat{\theta}_0 = \hat{\bar{y}} - \hat{\theta}_1 \hat{\bar{u}},\ \hat{\theta}_1 = K + D\hat{\sigma} \text{ and } \hat{\sigma} = \left\{B + \sqrt{(B^2 + 4nC)}\right\} / \left(2\sqrt{n(n-2)}\right) \tag{5.1.11}$$

where $K = \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}}) y_i \Big/ \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}})^2$, $D = \sum_{i=1}^{n} \alpha_i (u_i - \hat{\bar{u}}) \Big/ \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}})^2$,

$$B = \frac{2p}{k} \sum_{i=1}^{n} \alpha_i \{y_i - \hat{\bar{y}} - K(u_i - \hat{\bar{u}})\},\ C = \frac{2p}{k} \sum_{i=1}^{n} \beta_i \{y_i - \hat{\bar{y}} - K(u_i - \hat{\bar{u}})\}^2,$$

$$\hat{\bar{y}} = \sum_{i=1}^{n} \beta_i y_i \Big/ \sum_{i=1}^{n} \beta_i \text{ and } \hat{\bar{u}} = \sum_{i=1}^{n} \beta_i u_i \Big/ \sum_{i=1}^{n} \beta_i .$$

The calculation of the MMLEs proceeds in two steps. In the first step, we obtain the initial estimates (5.1.6) - (5.1.10) and use them to calculate the initial MMLEs. In the second step, we use these initial MMLEs to calculate (5.1.6) and then we use (5.1.5) to obtain the final MMLEs.

5.1.1 Simulations

To evaluate the efficiency and robustness of the MMLEs given in (5.1.11), we use the models (1)-(12) given in section 2.2 with $\mu = 0$. In addition, the LSEs given in (5.7) and (5.8) are computed for the same models. Without loss of generality, we assume that $\theta_0 = 0, \theta_1 = 1$, $\sigma = 1$. We generated $N = [100{,}000/n]$ (integer value) samples of independently distributed random errors of size n from each of the models (1)-(12). Note that models (1)-(9) have finite variance, (10) has finite mean but non-existent variance, and (11)-(12) have non-existent mean and variance. The random errors generated from models (6)-(9) were divided by suitable constants to make their variances equal to σ^2. The nonstochastic independent variables x_i's $(1 \le i \le n)$ were generated from a uniform distribution. They were standardized by replacing x_i with $u_i = (x_i - \bar{x})/s$ $(1 \le i \le n)$. From the resulting N values of the MMLEs and LSEs, we computed their means and variances. They are given in Table 5.1, Table 5.2 and Table 5.3 with sample sizes $n = 10$, $n = 20$ and $n = 50$, respectively.

Table 5.1: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model; $n = 10$.

Model	$\hat{\theta}_0$	$\tilde{\theta}_0$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
1	0.000	0.000	0.986	1.000	0.920	0.971
	[1.079]	[1.011]	[1.073]	[1.014]	[0.671]	[0.611]
2	0.000	0.001	0.985	1.003	0.890	0.963
	[1.000]	[1.012]	[1.001]	[1.000]	[0.746]	[0.813]
3	0.001	0.000	0.982	0.996	0.871	0.951
	[0.935]	[0.988]	[0.987]	[1.013]	[0.803]	[0.970]
4	-0.005	-0.004	0.986	1.004	0.818	0.930
	[0.829]	[1.023]	[0.883]	[1.019]	[0.849]	[1.528]
5	0.002	0.002	0.980	0.995	0.723	0.870
	[0.647]	[0.977]	[0.743]	[1.035]	[0.816]	[2.482]
6	-0.004	-0.003	0.992	1.005	0.878	0.961
	[0.940]	[0.978]	[0.901]	[0.954]	[0.654]	[0.754]
7	-0.002	0.002	0.993	1.004	0.708	0.931
	[0.561]	[1.010]	[0.525]	[0.910]	[0.547]	[1.593]
8	0.002	0.002	0.988	1.003	0.888	0.958
	[0.975]	[0.984]	[1.006]	[0.994]	[0.756]	[0.827]
9	-0.001	-0.001	0.981	0.999	0.738	0.897
	[0.687]	[0.998]	[0.784]	[1.004]	[0.926]	[1.947]
10	-0.005	-0.001	0.975	1.003	1.475	2.157
	[2.670]	[0.112E+02]	[3.050]	[0.929E+01]	[4.502]	[0.680E+02]
11	0.010	1.621	0.962	1.124	2.367	18.793
	[9.358]	[0.663E+05]	[14.110]	[0.576E+05]	[37.321]	[0.671E+06]
12	-0.001	-1.559	0.965	-0.308	3.187	31.428
	[15.019]	[0.298E+06]	[22.223]	[0.687E+05]	[56.555]	[0.326E+07]

Table 5.2: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model; $n=20$.

Model	$\hat{\theta}_0$	$\tilde{\theta}_0$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
1	0.000	0.001	0.991	0.999	0.950	0.985
	[1.070]	[1.022]	[1.035]	[0.987]	[0.589]	[0.548]
2	0.001	0.002	0.989	0.998	0.922	0.983
	[0.989]	[1.027]	[1.003]	[1.020]	[0.678]	[0.773]
3	0.001	0.000	0.992	1.001	0.893	0.972
	[0.886]	[0.980]	[0.942]	[1.032]	[0.724]	[1.002]
4	-0.002	-0.002	0.995	1.002	0.835	0.951
	[0.781]	[0.994]	[0.795]	[0.996]	[0.758]	[1.709]
5	-0.002	0.002	0.991	0.996	0.736	0.911
	[0.594]	[1.076]	[0.611]	[0.979]	[0.694]	[4.434]
6	-0.001	-0.001	0.990	0.997	0.917	0.979
	[0.929]	[0.983]	[0.956]	[0.999]	[0.611]	[0.724]
7	-0.001	0.001	0.992	1.000	0.751	0.955
	[0.563]	[0.989]	[0.586]	[1.027]	[0.536]	[1.729]
8	0.001	0.000	0.992	1.000	0.918	0.978
	[0.955]	[0.995]	[0.966]	[1.005]	[0.686]	[0.835]
9	0.000	0.002	0.993	1.002	0.751	0.943
	[0.616]	[1.015]	[0.627]	[1.000]	[0.769]	[2.303]
10	0.006	0.012	0.989	1.001	1.475	2.330
	[2.273]	[0.114E+02]	[2.489]	[0.104E+02]	[3.780]	[0.116E+03]
11	0.006	0.004	0.980	2.621	2.172	31.342
	[4.793]	[0.321E+06]	[5.404]	[0.280E+06]	[15.907]	[0.645E+07]
12	-0.008	-10.142	0.999	15.256	3.010	67.401
	[9.266]	[0.122E+08]	[11.171]	[0.426E+08]	[29.130]	[0.211E+09]

Table 5.3: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model; $n=50$.

Model	$\hat{\theta}_0$	$\tilde{\theta}_0$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
1	-0.001	-0.001	0.997	1.000	0.969	0.996
	[1.054]	[1.015]	[1.042]	[1.006]	[0.540]	[0.514]
2	0.000	0.000	0.997	1.000	0.935	0.993
	[0.963]	[1.026]	[0.934]	[0.971]	[0.638]	[0.801]
3	0.002	0.002	0.996	0.999	0.906	0.987
	[0.893]	[1.025]	[0.891]	[1.001]	[0.678]	[1.091]
4	-0.002	-0.002	0.997	1.000	0.843	0.977
	[0.738]	[0.990]	[0.759]	[0.981]	[0.681]	[2.143]
5	0.000	-0.002	0.996	0.998	0.740	0.943
	[0.548]	[0.990]	[0.579]	[0.975]	[0.618]	[5.304]
6	-0.001	-0.001	0.998	1.000	0.931	0.993
	[0.950]	[1.004]	[0.850]	[0.866]	[0.576]	[0.745]
7	0.002	0.002	0.999	1.000	0.751	0.983
	[0.552]	[0.994]	[0.476]	[0.678]	[0.459]	[1.934]
8	0.000	0.000	0.998	1.000	0.932	0.992
	[0.947]	[1.000]	[0.947]	[0.998]	[0.625]	[0.835]
9	-0.002	-0.003	0.999	1.002	0.754	0.974
	[0.558]	[0.983]	[0.584]	[1.015]	[0.691]	[2.661]
10	0.001	0.008	0.991	1.001	1.465	2.589
	[2.082]	[0.153E+02]	[2.172]	[0.163E+02]	[3.185]	[0.430E+03]
11	-0.002	6.445	0.988	5.921	2.032	88.917
	[3.792]	[0.154E+08]	[4.208]	[0.349E+08]	[11.834]	[0.752E+09]
12	0.004	-4.074	0.990	3.365	2.855	78.031
	[7.465]	[0.258E+07]	[8.044]	[0.168E+07]	[19.197]	[0.130E+09]

The results show that MMLEs are enormously more efficient than the LSEs other than for the normal distribution in which case they are a little less efficient. Both MMLEs and LSEs of θ_0 and θ_1 are unbiased for models (1)-(10); see also Appendix C. For models (11) and (12), however, the LSEs are not even unbiased. For model (10), the variances of the LSEs $\tilde{\theta}_0$ and $\tilde{\theta}_1$ are much larger than the corresponding MMLEs although they are unbiased.

The above results are very promising because the only assumption we make is that the underlying distribution is long-tailed symmetric including distributions as extreme as Cauchy.

5.2 Generalized Logistic

Let the random error in the simple linear regression model (5.6) come from generalized logistic distribution

$$f(e) = \frac{b}{\sigma} \frac{\exp(-e/\sigma)}{\{1+\exp\,(-e/\sigma)\}^{b+1}}, \qquad -\infty < e < \infty, \qquad (5.2.1)$$

where σ is scale and b is shape parameter.

The likelihood function L in terms of $z_i = (y_i - \theta_1 u_i - \theta_0)/\sigma$ $(1 \le i \le n)$ is

$$L = \left(\frac{b}{\sigma}\right)^n \prod_{i=1}^{n} \frac{\exp(-z_i)}{\{1+\exp\,(-z_i)\}^{b+1}} .$$

Writing $g(z) = e^{-z}/(1+e^{-z})$, we estimate θ_0, θ_1 and σ by using the following maximum likelihood equations:

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} g(z_i) = 0, \qquad (5.2.2)$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{1}{\sigma}\sum_{i=1}^{n} u_i - \frac{(b+1)}{\sigma}\sum_{j=1}^{n} u_i g(z_i) = 0 \text{ and} \qquad (5.2.3)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_i - \frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_i g(z_i) = 0 . \qquad (5.2.4)$$

We utilize the method of modified maximum likelihood estimation to solve these intractable equations. Islam et al (2001) derived the MMLEs for a generalized logistic distribution. We extend their study by assuming an unknown shape parameter; Islam et al (2001) assume that the shape parameter is known.

We initially estimate $(b+1)$ by $1/(1-\tilde{\tilde{w}}_i)$ $(1 \le i \le n)$ where

$$\tilde{\tilde{w}} = (1/n)\sum_{i=1}^{n} \tilde{w}_i\,, \quad \tilde{w}_i = e^{\tilde{t}_i}/(1+e^{\tilde{t}_i}) = 1/(1+e^{-\tilde{t}_i}) \quad 1 \le i \le n .$$

Refer to section 3.2 for details.

Note that Taylor series expansion of $g(z)$ is used to work out MMLEs as before:

$$g(z_{(i)}) \cong \alpha_i + \beta_j z_{(i)} \quad (1 \le i \le n) \qquad (5.2.5)$$

where

$$\alpha_i = (1+e^t+te^t)/(1+e^t)^2\,,\ \beta_i = e^t/(1+e^t)^2\,,\ t = t_{(i)} = E\{z_{(i)}\},\ 1 \le i \le n .$$

To estimate $t_{(i)}$, we use $\tilde{t}_{(i)}$ as an initial estimator,

$$\tilde{t}_{(i)} = \frac{y_{[i]} - T_0 - T_1 u_{[i]}}{S_0} \tag{5.2.6}$$

where

$$T_1 = median\{r_\ell\};\ r_\ell = \frac{y_{\ell+1} - y_\ell}{u_{\ell+1} - u_\ell} \quad (1 \le \ell \le n-1) \tag{5.2.7}$$

is the initial estimator of the regression coefficient θ_1 ,

$$T_0 = median\{\tilde{\vartheta}_i\};\ \tilde{\vartheta}_i = y_i - T_1 u_i \ \ (1 \le i \le n) \tag{5.2.8}$$

is the initial estimator of the intercept θ_0 and

$$S_0 = 1.483 median\{|\tilde{\vartheta}_i - T_0|\} \tag{5.2.9}$$

is the initial estimator of σ and $(y_{[i]}, u_{[i]})$ are the concomitants of the ordered variates $z_{(i)} = (y_{[i]} - \theta_1 u_{[i]} - \theta_0)/\sigma \ \ (1 \le i \le n)$.

Since complete sums are invariant to ordering, we use $\tilde{t}_i$ rather than $\tilde{t}_{(i)}$. We replace t_i by $\tilde{t}_i$ and get the initial estimates of α_i and β_i denoted by $\tilde{\alpha}_i$ and $\tilde{\beta}_i$, respectively.

The resulting MMLEs are

$$\hat{\theta}_0 = \hat{\bar{y}} - \hat{\theta}_1 \hat{\bar{u}} + (\Delta/m)\hat{\sigma},\ \hat{\theta}_1 = K + D\hat{\sigma} \tag{5.2.10}$$

$$\text{and } \hat{\sigma} = \left\{B + \sqrt{(B^2 + 4nC)}\right\} \Big/ \left(2\sqrt{n(n-2)}\right) \tag{5.2.11}$$

where $K = \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}}) y_i \Big/ \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}})^2$, $D = \sum_{i=1}^{n} \Delta_i (u_i - \hat{\bar{u}}) \Big/ \sum_{i=1}^{n} \beta_i (u_i - \hat{\bar{u}})^2$,

$B = (b+1)\sum_{j=1}^{n} \Delta_i \{y_i - \hat{\bar{y}} - K(u_i - \hat{\bar{u}})\}$, $C = (b+1)\sum_{j=1}^{n} \beta_i \{y_i - \hat{\bar{y}} - K(u_i - \hat{\bar{u}})\}^2$,

$\Delta_i = (b+1)^{-1} - \alpha_i$, $\Delta = \sum_{i=1}^{n} \Delta_i$, $\hat{\bar{y}} = \sum_{i=1}^{n} \beta_i y_i \Big/ \sum_{i=1}^{n} \beta_i$ and $\hat{\bar{u}} = \sum_{i=1}^{n} \beta_i u_i \Big/ \sum_{i=1}^{n} \beta_i$.

It may be noted that $\hat{\theta}_0$ is estimating

$\tau = \theta_0 +$ scaled median of the error distribution.

The calculation of the MMLEs proceed in three steps. In the first step, we calculate the MMLEs by using the initial estimates (5.2.6) - (5.2.9). In the second step, we use these MMLEs to calculate (5.2.6). Then, we calculate the MMLEs (5.2.10) and (5.2.11). Third, we use these new MMLEs to do the fourth iteration and obtain the final MMLEs.

5.2.1 Simulations

In this section, we study the efficiency and robustness of the MMLEs given in (5.2.10) and (5.2.11) by generating $N = \left[100{,}000/n\right]$ (integer value) samples of independently distributed random errors e_i of size n from generalized logistic distribution with different values of the shape parameter b; e_i were multiplied by $\left[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}\right]^{1/2}$. Without any loss of generality, we assume that $\theta_0 = 0$, $\theta_1 = 1$ and $\sigma = 1$. The nonstochastic values x_i's $(1 \le i \le n)$ are generated from uniform distribution and $u_i = (x_i - \bar{x})/s$. The means and variances of the resulting N values of MMLEs are given in Table 5.4, Table 5.5 and Table 5.6 with sample sizes $n = 10$, $n = 20$ and $n = 50$, respectively.

It may be noted that $\hat{\tau}$ and $\tilde{\tau}$ are estimating -0.777, 0, 1.056, 2.174, 2.819, and 3.268 for b = 0.5, 1, 2, 4, 6 and 8, respectively.

Table 5.4: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model for generalized logistic distribution with shape parameter *b*; $n = 10$.

b	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
0.5	-0.851	-0.906	0.989	0.996	0.941	1.003
	[3.403]	[3.481]	[2.974]	[3.299]	[0.853]	[0.862]
1	0.001	-0.001	0.998	1.003	0.978	1.030
	[3.579]	[3.576]	[3.186]	[3.326]	[0.862]	[0.816]
2	1.118	1.150	0.991	0.996	0.994	1.046
	[3.823]	[3.792]	[3.144]	[3.312]	[0.895]	[0.888]
4	2.245	2.296	0.996	1.001	0.985	1.042
	[3.719]	[3.717]	[3.063]	[3.263]	[0.915]	[0.964]
6	2.897	2.953	0.989	0.994	0.984	1.041
	[3.712]	[3.709]	[2.995]	[3.211]	[0.947]	[0.998]
8	3.359	3.419	0.999	1.004	0.983	1.044
	[3.741]	[3.745]	[3.016]	[3.234]	[0.951]	[1.013]

Table 5.5: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model for generalized logistic distribution with shape parameter *b*; $n = 20$.

b	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
0.5	-0.826	-0.904	0.997	0.997	0.950	1.000
	[3.389]	[3.519]	[2.877]	[3.294]	[0.808]	[0.904]
1	0.004	0.001	0.996	0.996	0.990	1.023
	[3.749]	[3.764]	[3.034]	[3.247]	[0.807]	[0.819]
2	1.088	1.141	0.998	0.999	1.000	1.035
	[3.857]	[3.862]	[3.165]	[3.359]	[0.818]	[0.867]
4	2.227	2.307	1.007	1.007	0.999	1.041
	[3.839]	[3.841]	[3.103]	[3.347]	[0.852]	[0.986]
6	2.863	2.951	0.995	0.998	0.990	1.035
	[3.780]	[3.835]	[2.986]	[3.281]	[0.872]	[1.029]
8	3.328	3.422	1.000	1.002	1.000	1.046
	[3.764]	[3.812]	[2.998]	[3.327]	[0.873]	[1.041]

Table 5.6: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of parameters in a simple linear regression model for generalized logistic distribution with shape parameter b; $n = 50$.

b	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
0.5	-0.801	-0.891	0.999	1.000	0.950	1.001
	[3.253]	[3.335]	[2.785]	[3.333]	[0.751]	[0.927]
1	-0.004	-0.003	0.999	0.999	0.998	1.028
	[3.590]	[3.561]	[3.004]	[3.265]	[0.761]	[0.814]
2	1.072	1.135	0.998	0.998	1.008	1.038
	[3.785]	[3.739]	[3.083]	[3.324]	[0.798]	[0.892]
4	2.196	2.293	1.001	1.001	1.008	1.047
	[3.806]	[3.784]	[2.969]	[3.298]	[0.831]	[1.033]
6	2.829	2.934	0.999	0.999	1.001	1.044
	[3.801]	[3.751]	[2.863]	[3.193]	[0.825]	[1.050]
8	3.294	3.405	1.000	0.999	1.005	1.049
	[3.878]	[3.860]	[2.925]	[3.307]	[0.863]	[1.119]

The MMLEs are seen to be enormously more efficient (jointly) than the LSEs besides having negligible bias. The results are very interesting from theoretical and practical considerations. It may be noted that the MMLEs have bounded influence functions. For illustration, the emprical influence function of $\hat{\theta}_1$ is given below.

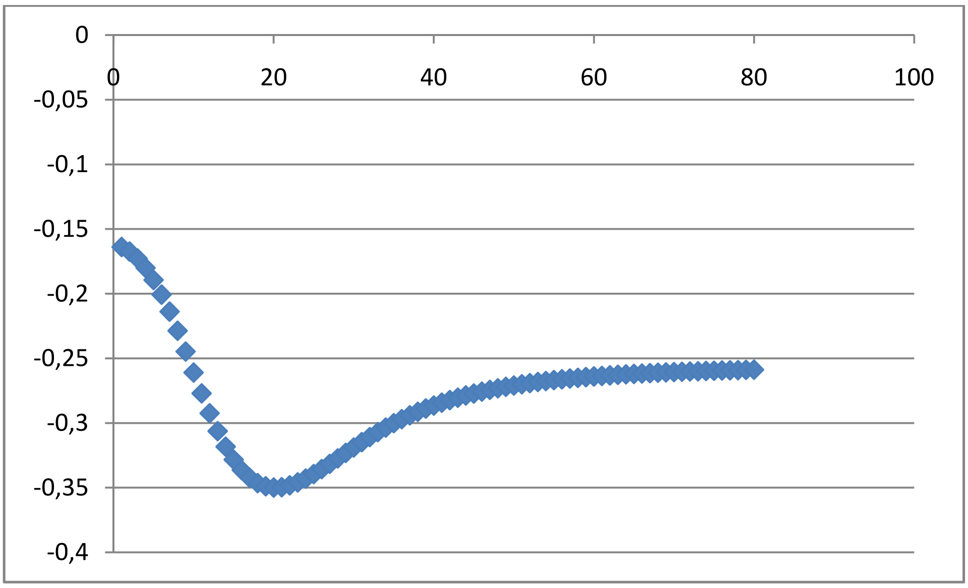

Figure 5.1: Emprical influence function of $\hat{\theta}_1$ for long-tailed symmetric distribution, $p = 3.5$.

CHAPTER 6

ROBUST MULTIPLE LINEAR REGRESSION

This chapter expands our findings on simple linear regression model to multiple linear regression model

$$Y = X\theta^* + e \tag{6.1}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{nx1}, \; X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix}_{nx(q+1)}, \; \theta^* = \begin{bmatrix} \theta_0^* \\ \theta_1^* \\ \vdots \\ \theta_q^* \end{bmatrix}_{(q+1)x1} \text{ and } e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{nx1};$$

$e_i \; (1 \le i \le n)$ are iid errors with mean zero and unknown variance σ^2, X is a nonstochastic design matrix consisting of $q+1$ explanatory variables which have no linear relationship with each other, Y is a vector of observed responses and θ is a vector of parameters to be estimated.

The commonly used least squares estimation method minimizes the error sum of squares $e'e$ and this results in closed-form least square estimators (LSEs) of θ and σ, namely,

$$\tilde{\theta}^* = (X'X)^{-1} X'Y \text{ and} \tag{6.2}$$

$$\tilde{\sigma}^* = \sqrt{\sum_{i=1}^{n} \left\{ y_i - \bar{y} - \sum_{j=1}^{q} \tilde{\theta}_j^* \left(x_{ij} - \bar{x}_j \right) \right\}^2 \Big/ (n-q-1)};$$

$\bar{x}_j = (1/n)\sum_{i=1}^{n} x_{ij}$ and $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$.

However, Puthenpura and Sinha (1986) show that these LSEs are not efficient if the data is very noisy. Additionally, Islam and Tiku (2004) show that the efficiencies of these LSEs are low for non-normal error distributions and generally decrease as n increases.

Note that the variance-covariance matrix of the estimator $\tilde{\theta}^*$ is

$$Cov(\tilde{\theta}^*) = (X'X)^{-1}\sigma^2 .$$

Akkaya and Tiku (2008) point out that the variances (and covariances) of LSEs are too vulnerable to design anomalies. To rectify the situation, they propose a reparametrized multiple linear regression model

$$Y = 1\theta_0 + U\theta + e \tag{6.3}$$

where

$$1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{nx1} , \quad U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1q} \\ u_{21} & u_{22} & \cdots & u_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nq} \end{bmatrix}_{nxq} , \quad u_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad , \quad \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad \text{and}$$

$$s_j^2 = (1/n)\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \quad (1 \le i \le n, 1 \le j \le q).$$

The LSEs of the model (6.3) become

$$\tilde{\theta}_0 = \bar{y}, \; \tilde{\theta} = (U'U)^{-1}U'Y \text{ and} \tag{6.4}$$

$$\tilde{\sigma} = s_e ; \; s_e^2 = \sum_{i=1}^{n} \left\{ y_i - \bar{y} - \sum_{j=1}^{q} \tilde{\theta}_j u_{ij} \right\}^2 \Big/ (n - q - 1) \tag{6.5}$$

since $\bar{u}_j = (1/n)\sum_{i=1}^{n} u_{ij} = 0$. The variance-covariance matrix of the estimator $\tilde{\theta}$ is

$$Cov(\tilde{\theta}) = (U'U)^{-1}\sigma^2 ,$$

while the variances of $\tilde{\theta}_0$ and $\tilde{\sigma}$ are as follows:

$$Var(\tilde{\theta}_0) = \sigma^2/n$$

$$Var(\tilde{\sigma}) \cong \frac{\sigma^2}{2n}\left(1+\frac{1}{2}\lambda_4\right), \ \lambda_4 = \mu_4/\mu_2^2 - 3,$$

as said earlier. The estimator $\tilde{\theta}_0$ is uncorrelated with $\tilde{\sigma}$ when the distribution of random errors is symmetric.

Akkaya and Tiku (2008) show that the parameters $\tilde{\theta}_j$ $(1 \le j \le q)$ are invariant to location and scale of x_{ij} $(1 \le i \le n)$ and so are $\tilde{\theta}_0$ and $\tilde{\sigma}$. That is, if x_{ij} are replaced by $a_j + b_j x_{ij}$ $(1 \le i \le n)$, a_j and b_j being constants, the values of $\tilde{\theta}_0$, $\tilde{\theta}_j$ and $\tilde{\sigma}$ do not change neither do their covariances and variances.

When e_i $(1 \le i \le n)$ are iid normal $N(0, \sigma^2)$, LSEs are identical to the MLEs and are fully efficient. Here, we are interested in developing estimators which have high efficiencies and have bounded influence functions. To do that we proceed as follows.

6. 1 Long-Tailed Symmetric Distributions

Assume that the errors e_i $(1 \le i \le n)$ have one of the distributions in the family

$$f(e) = \frac{1}{\sigma\sqrt{k}}\,\frac{1}{\beta(1/2, p-1/2)}\left[1+\frac{e^2}{k\sigma^2}\right]^{-p}, \qquad -\infty < e < \infty;$$

$k = 2p-3, \ p \ge 2.$

The maximum likelihood equations are

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{2p}{\sigma k}\sum_{i=1}^{n} g(z_i) = 0 \tag{6.1.1}$$

$$\frac{\partial \ln L}{\partial \theta_j} = \frac{2p}{\sigma k}\sum_{i=1}^{n} u_{ij}\, g(z_i) = 0 \quad (1 \le j \le q) \tag{6.1.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{\sigma k}\sum_{i=1}^{n} z_i\, g\left(z_i\right) = 0 \tag{6.1.3}$$

where

$$z_i = \left(y_i - \theta_0 - \sum_{j=1}^{n}\theta_j u_{ij}\right)\Big/\sigma, \qquad g(z) = z\big/\left(1+z^2/k\right), \qquad u_{ij} = \left(x_{ij} - \bar{x}_j\right)\big/s_j,$$

$$\bar{x}_j = (1/n)\sum_{i=1}^{n} x_{ij} \text{ and } s_j = \sqrt{(1/n)\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2} \quad \left(1 \le i \le n, 1 \le j \le q\right).$$

Since the likelihood equations (6.1.1)-(6.1.3) are intractable, we utilize the modified maximum likelihood method by using the linear approximations

$$g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)} \quad (1 \le i \le n); \tag{6.1.4}$$

$$\alpha_i = \frac{(1/k)t_{(i)}}{\left\{1+(1/k)t_{(i)}^2\right\}^2} \text{ and } \beta_i = \frac{1}{\left\{1+(1/k)t_{(i)}^2\right\}^2}, \tag{6.1.5}$$

$$t_{(i)} = E\left\{z_{(i)}\right\} \ (k = 30).$$

In order to find appropriate initial estimate of $t_{(i)}$, we take θ_j $\left(1 \le j \le q\right)$ all equal; say $\dot{\theta}$. This is motivated by the fact that *a priori* there is no reason to believe that one design variable is more important than others. Thus, (6.3) can *a priori* be expressed as

$$y_i = \theta_0 + \dot{\theta} v_i + e_i\,;\ v_i = \sum_{j=1}^{q} u_{ij} \quad \left(1 \le i \le n\right).$$

An initial estimate of $t_{(i)}$ is, therefore,

$$\tilde{t}_{(i)} = \frac{y_{[i]} - T_0 - T_1 v_{[i]}}{S_0} \tag{6.1.6}$$

where

$$T_1 = median\left\{r_\ell\right\},\ r_\ell = \frac{y_{\ell+1} - y_\ell}{v_{\ell+1} - v_\ell} \quad (1 \le \ell \le n-1) \tag{6.1.7}$$

is an initial estimator of $\dot{\theta}$,

$$T_0 = median\left\{\tilde{w}_i\right\},\ \tilde{w}_i = y_i - T_1 v_i \ (1 \le i \le n), \tag{6.1.8}$$

is an initial estimator of the intercept θ_0 and

$$S_0 = 1.483\, median\left\{\left|\, \tilde{w}_i - T_0 \,\right|\right\} \tag{6.1.9}$$

is an initial estimator of σ; $\left(y_{[i]}, v_{[i]}\right)$ are the concomitants of the ordered variates $z_{(i)} = \left(y_{[i]} - \dot{\theta}\, v_{[i]} - \theta_0\right)/\sigma$ $(1 \le i \le n)$.

The motivation for the initial estimate T_1 is that $y_{\ell+1} - y_{\ell} = \dot{\theta}\left(v_{\ell+1} - v_{\ell}\right)$ $(1 \le \ell \le n-1)$. Since complete sums are invariant to ordering, we can drop the ordering and use $\widetilde{t}_i$ rather than $\widetilde{t}_{(i)}$. After using the initial estimate of t_i,

$$\widetilde{t}_i = \frac{y_i - T_0 - T_1 v_i}{S_0} \tag{6.1.10}$$

we obtain the initial estimates of α_i and β_i; $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$, respectively.

The modified likelihood estimators obtained as in the previous chapter are

$$\hat{\theta}_0 = \hat{\bar{y}} - \sum_{j=1}^{q} \hat{\theta}_j \hat{\bar{u}}_j \,,\ \hat{\theta} = K + D\hat{\sigma} \text{ and} \tag{6.1.11}$$

$$\hat{\sigma} = \left\{B + \sqrt{\left(B^2 + 4nC\right)}\right\} \Big/ \left(2\sqrt{n\left(n-q-1\right)}\right) \tag{6.1.12}$$

where

$$B = \frac{2p}{k}\sum_{i=1}^{n} \alpha_i \left\{y_i - \hat{\bar{y}} - \sum_{j=1}^{q} K_j \mathrm{M}_{ij}\right\},\ C = \frac{2p}{k}\sum_{i=1}^{n} \beta_i \left\{y_i - \hat{\bar{y}} - \sum_{j=1}^{q} K_j \mathrm{M}_{ij}\right\}^2,$$

$$K = \left(\mathrm{M}'\theta\mathrm{M}\right)^{-1}\left(\mathrm{M}'\theta Y\right) = \left(K_j\right)_{qx1},\ D = \left(\mathrm{M}'\theta\mathrm{M}\right)^{-1}\left(\mathrm{M}'\alpha Y\right) = \left(D_j\right)_{qx1},$$

$$\mathrm{M}_{ij} = u_{ij} - \hat{\bar{u}}_j,\ \mathrm{M} = \left(\mathrm{M}_{ij}\right)_{nxq}, \beta = diag\left(\beta_i\right)_{nxn},\ \alpha = diag\left(\alpha_i\right)_{nxn},$$

$$\hat{\bar{y}} = \sum_{i=1}^{n} \beta_i y_i \Big/ \sum_{i=1}^{n} \beta_i \text{ and } \hat{\bar{u}}_j = \sum_{i=1}^{n} \beta_i u_{ij} \Big/ \sum_{i=1}^{n} \beta_i \left(1 \le i \le n, 1 \le j \le q\right).$$

We calculate the MMLEs in two steps: First, we assume that all θ_j $\left(1 \le j \le q\right)$ are equal and calculate the MMLEs $\hat{\theta}_{00}$, $\hat{\theta}_{j0}$ $\left(1 \le j \le q\right)$ and $\hat{\sigma}_0$ from (6.1.10). Second, we use these MMLEs to calculate

$$\widetilde{t}_i = \frac{y_i - \hat{\theta}_{00} - \sum_{j=1}^{q} \hat{\theta}_{j0} u_{ij}}{\hat{\sigma}_0}$$

which lead us to the final MMLEs.

6.1.1 Simulations

To study the efficiency and robustness of the MMLEs (6.1.11)-(6.1.12), we consider the twelve distributions in section 2.2 with $\mu = 0$. The LSEs (6.4)-(6.5) are also computed for the same models. Models (1)-(9) have finite mean and variance, while model (10) has finite mean but non-existent variance, and models (11)-(12) have non-existent mean and variance. The random errors generated from models (6)-(9) were divided by suitable constants to make their variances equal to σ^2. We generated the nonstochastic design variables x_{ij}'s $(1 \le i \le n, 1 \le j \le q)$ from a uniform distribution and, for illustration, we consider $q = 4$ explanatory variables. We generated $N = [100{,}000/n]$ (integer value) samples of independently distributed random errors of size n from each of the models (1)-(12). Without loss of generality, we take $\theta_0 = 0$, $\theta_j = 1$ $(1 \le j \le q)$, $\sigma = 1$. From the resulting N values of the MMLEs and LSEs, we computed their means and variances. The results are given in Table 6.1-Table 6.2 with $n = 20$ and Table 6.3-Table 6.4 with $n = 50$.

Table 6.1: Simulated values of $(1/\sigma)$Meanand (n/σ^2)Variance(in brackets) of MMLEs and LSEs of θ_0, θ_1 and σ in multiple linear regression model where random errors are assumed to come from a long tailed symmetric family; $n = 20$.

Model	$\hat{\theta}_0$	$\tilde{\theta}_0$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
1	0.001	0.002	0.997	0.998	0.977	0.986
	[1.053]	[1.040]	[1.076]	[1.084]	[0.654]	[0.657]
2	-0.003	-0.002	0.999	0.999	0.951	0.976
	[0.966]	[1.010]	[0.977]	[1.026]	[0.756]	[0.858]
3	0.003	0.002	1.001	1.002	0.936	0.972
	[0.918]	[1.003]	[1.018]	[1.115]	[0.858]	[1.137]
4	-0.001	-0.001	1.001	1.002	0.888	0.956
	[0.843]	[1.029]	[0.815]	[1.001]	[0.941]	[1.851]
5	-0.002	-0.003	1.003	1.003	0.796	0.910
	[0.636]	[0.981]	[0.671]	[1.087]	[0.906]	[3.365]
6	0.000	0.000	1.000	1.000	0.956	0.984
	[0.931]	[0.985]	[1.156]	[1.306]	[0.704]	[0.847]
7	-0.003	-0.002	0.998	0.999	0.815	0.958
	[0.584]	[0.987]	[0.908]	[1.723]	[0.712]	[1.914]
8	-0.002	-0.002	1.004	1.004	0.950	0.974
	[0.964]	[1.003]	[1.006]	[1.067]	[0.776]	[0.899]
9	-0.002	-0.003	0.995	0.995	0.811	0.932
	[0.661]	[0.992]	[0.703]	[1.034]	[1.098]	[2.359]
10	-0.005	-0.003	1.004	1.019	1.640	2.340
	[2.718]	[0.108E+02]	[3.035]	[0.119E+02]	[5.323]	[0.105E+03]
11	-0.001	-6.954	0.998	-5.305	2.598	55.561
	[6.962]	[0.460E+07]	[7.658]	[0.488E+07]	[27.628]	[0.792E+08]
12	0.007	-7.376	1.006	-1.841	3.529	60.486
	[12.488]	[0.350E+07]	[14.649]	[0.145E+07]	[46.957]	[0.783E+08]

Table 6.2: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of θ_2, θ_3 and θ_4 in multiple linear regression model where random errors are assumed to come from a long tailed symmetric family; $n = 20$.

Model	$\hat{\theta}_2$	$\tilde{\theta}_2$	$\hat{\theta}_3$	$\tilde{\theta}_3$	$\hat{\theta}_4$	$\tilde{\theta}_4$
1	1.004	1.003	1.002	1.002	1.000	1.000
	[1.031]	[1.042]	[1.027]	[1.037]	[1.037]	[1.041]
2	0.995	0.995	0.995	0.995	1.000	0.999
	[1.006]	[1.060]	[0.982]	[1.038]	[0.983]	[1.034]
3	1.004	1.004	0.998	0.997	1.000	1.000
	[0.988]	[1.101]	[0.957]	[1.059]	[0.933]	[1.030]
4	0.999	0.997	1.003	1.002	0.997	0.997
	[0.849]	[1.067]	[0.828]	[1.031]	[0.840]	[1.056]
5	1.005	1.005	0.998	0.998	1.000	1.000
	[0.682]	[1.066]	[0.640]	[1.054]	[0.642]	[1.018]
6	0.998	0.999	0.998	0.999	0.995	0.996
	[0.980]	[1.063]	[0.799]	[0.814]	[0.957]	[1.042]
7	1.001	1.001	1.000	1.000	0.996	0.996
	[0.648]	[1.089]	[0.435]	[0.476]	[0.655]	[1.108]
8	1.000	1.000	0.997	0.997	1.004	1.004
	[0.983]	[1.047]	[1.004]	[1.063]	[0.979]	[1.040]
9	0.999	0.999	1.000	1.000	0.998	0.997
	[0.675]	[1.034]	[0.685]	[1.032]	[0.683]	[1.010]
10	0.999	0.996	0.998	0.987	1.011	1.003
	[3.029]	[0.109E+02]	[2.862]	[0.998E+01]	[2.749]	[0.116E+02]
11	1.010	9.419	1.001	-8.504	0.992	-3.960
	[7.966]	[0.627E+07]	[7.223]	[0.920E+07]	[7.524]	[0.312E+07]
12	1.009	-3.692	1.019	6.233	0.996	5.328
	[14.804]	[0.212E+07]	[12.739]	[0.330E+07]	[13.103]	[0.112E+07]

Table 6.3: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of θ_0, θ_1 and σ in multiple linear regression model where random errors are assumed to come from a long tailed symmetric family; $n = 50$.

Model	$\hat{\theta}_0$	$\tilde{\theta}_0$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
1	0.004	0.004	0.998	0.998	0.972	0.993
	[0.997]	[0.986]	[1.080]	[1.078]	[0.562]	[0.565]
2	0.002	0.002	1.000	0.999	0.951	0.992
	[0.922]	[0.980]	[1.006]	[1.060]	[0.638]	[0.762]
3	-0.006	-0.004	1.004	1.004	0.933	0.993
	[0.942]	[1.067]	[1.014]	[1.173]	[0.699]	[1.139]
4	-0.004	-0.006	0.998	0.999	0.877	0.978
	[0.767]	[0.988]	[0.848]	[1.084]	[0.794]	[2.064]
5	0.003	0.003	1.001	1.001	0.777	0.935
	[0.596]	[0.955]	[0.624]	[1.090]	[0.722]	[4.803]
6	-0.005	-0.005	0.998	0.997	0.947	0.991
	[0.911]	[0.983]	[1.090]	[1.189]	[0.565]	[0.747]
7	-0.003	-0.004	0.998	0.997	0.787	0.969
	[0.575]	[0.995]	[0.833]	[1.457]	[0.527]	[1.866]
8	0.005	0.006	1.000	1.000	0.947	0.993
	[0.957]	[1.012]	[1.045]	[1.125]	[0.655]	[0.885]
9	-0.002	0.000	1.002	1.005	0.800	0.983
	[0.595]	[1.037]	[0.676]	[1.143]	[0.930]	[2.797]
10	-0.004	0.007	0.997	1.021	1.585	2.757
	[2.297]	[0.478E+02]	[2.661]	[0.378E+02]	[3.843]	[0.207E+04]
11	-0.006	-0.206	0.995	1.567	2.386	34.902
	[5.105]	[0.299E+05]	[5.605]	[0.120E+05]	[17.809]	[0.145E+07]
12	0.004	1.638	1.006	1.329	3.266	65.861
	[8.954]	[0.432E+06]	[9.799]	[0.922E+06]	[31.050]	[0.209E+08]

Table 6.4: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of θ_2, θ_3 and θ_4 in multiple linear regression model where random errors are assumed to come from a long tailed symmetric family; $n = 50$.

Model	$\hat{\theta}_2$	$\tilde{\theta}_2$	$\hat{\theta}_3$	$\tilde{\theta}_3$	$\hat{\theta}_4$	$\tilde{\theta}_4$
1	1.000	1.000	1.000	1.000	0.999	0.999
	[1.067]	[1.074]	[1.191]	[1.196]	[1.058]	[1.058]
2	1.004	1.003	0.999	0.998	1.001	1.000
	[0.946]	[1.011]	[0.976]	[1.037]	[0.993]	[1.068]
3	0.996	0.995	1.005	1.006	1.003	1.003
	[0.958]	[1.081]	[1.008]	[1.157]	[0.934]	[1.065]
4	1.001	1.000	0.999	0.999	0.999	0.999
	[0.812]	[1.033]	[0.772]	[1.000]	[0.878]	[1.134]
5	1.000	1.005	0.998	1.000	1.002	1.001
	[0.618]	[1.107]	[0.620]	[0.973]	[0.642]	[1.075]
6	1.003	1.003	1.001	1.000	0.998	0.998
	[0.905]	[0.971]	[1.071]	[1.140]	[1.302]	[1.485]
7	0.995	0.996	0.996	0.994	0.995	0.994
	[0.595]	[0.988]	[0.754]	[1.348]	[1.030]	[2.051]
8	0.998	0.996	1.002	1.003	0.997	0.998
	[1.031]	[1.118]	[1.055]	[1.150]	[1.038]	[1.121]
9	1.003	1.005	1.004	1.004	0.997	0.997
	[0.614]	[1.031]	[0.619]	[1.050]	[0.649]	[1.103]
10	0.997	0.981	0.998	1.031	0.999	0.999
	[2.516]	[0.389E+02]	[2.342]	[0.396E+02]	[2.591]	[0.219E+02]
11	1.011	0.683	1.005	0.698	0.992	0.921
	[5.719]	[0.334E+05]	[6.129]	[0.352E+05]	[5.271]	[0.200E+05]
12	1.008	2.221	0.983	-0.410	1.001	-1.680
	[9.216]	[0.193E+06]	[9.570]	[0.706E+06]	[10.117]	[0.678E+06]

As in simple linear regression, MMLEs are observed to be considerably more efficient than the LSEs except for the normal distribution in which case they are a little less efficient. For models (10)-(12), the variances of the LSEs are much larger than the MMLEs. Regarding the bias, MMLEs and LSEs of θ_0 and θ are both unbiased for models (1)-(10). However, the LSEs are not even unbiased for models (11) and (12). MMLEs are unbiased (almost) for all distributions (1)-(12) and have finite variances. This is because they have bounded influence functions; see Appendix C.

6.2 Generalized Logistic Distributions

Consider now the situation when errors have one of the distributions in the family

$$f(e)=\frac{b}{\sigma}\frac{\exp(-e/\sigma)}{\{1+\exp(-e/\sigma)\}^{b+1}}, \qquad -\infty<e<\infty,$$

where σ is scale and b is shape parameter.

The maximum likelihood equations are

$$\frac{\partial \ln L}{\partial \theta_0}=\frac{n}{\sigma}-\frac{(b+1)}{\sigma}\sum_{i=1}^{n} g(z_i)=0, \tag{6.2.1}$$

$$\frac{\partial \ln L}{\partial \theta_j}=\frac{1}{\sigma}\sum_{i=1}^{n} u_{ij}-\frac{(b+1)}{\sigma}\sum_{j=1}^{n} u_{ij} g(z_i)=0 \quad (1\le j\le q) \tag{6.2.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma}\sum_{i=1}^{n} z_i-\frac{(b+1)}{\sigma}\sum_{i=1}^{n} z_i g(z_i)=0; \tag{6.2.3}$$

$$z_i=\left(y_i-\theta_0-\sum_{j=1}^{n}\theta_j u_{ij}\right)\Big/\sigma, \quad g(z)=e^{-z}/(1+e^{-z}), \quad u_{ij}=(x_{ij}-\bar{x}_j)/s_j,$$

$$\bar{x}_j=(1/n)\sum_{i=1}^{n} x_{ij} \text{ and } s_j=\sqrt{(1/n)\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2} \quad (1\le i\le n, 1\le j\le q).$$

To work out the MMLEs, we use as usual the linear approximation

$$g(z_{(i)})\cong \alpha_i+\beta_i z_{(i)} \quad (1\le i\le n); \tag{6.2.4}$$

$$\alpha_i=(1+e^t+te^t)/(1+e^t)^2, \ \beta_i=e^t/(1+e^t)^2, \ t=t_{(i)}=E\{z_{(i)}\}, \ 1\le i\le n.$$

Since we do not know the value of b, we estimate $(b+1)$ by $1/(1-\tilde{\tilde{w}}_i)$ $(1\le i\le n)$ where

$$\tilde{\tilde{w}}=(1/n)\sum_{i=1}^{n}\tilde{w}_i, \quad \tilde{w}_i=e^{\tilde{t}_i}/(1+e^{\tilde{t}_i})=1/(1+e^{-\tilde{t}_i}) \quad 1\le i\le n. \tag{6.2.5}$$

Refer to section 3.2 for details.

As in case of long tailed symmetric family, we assume $\theta_j \ (1 \le j \le q)$ are all equal to find an initial estimate of $t_{(i)}$:

$$\tilde{t}_{(i)} = \frac{y_{[i]} - T_0 - T_1 v_{[i]}}{S_0} \ ; \ v_i = \sum_{j=1}^{q} u_{ij} \ \left(1 \le i \le n\right) \qquad (6.2.6)$$

where

$$T_1 = median\{r_\ell\}; \ r_\ell = \frac{y_{\ell+1} - y_\ell}{v_{\ell+1} - v_\ell} \quad (1 \le \ell \le n-1) \qquad (6.2.7)$$

is the initial estimator of $\dot{\theta}$,

$$T_0 = median\{\tilde{w}_i\}, \ \tilde{w}_i = y_i - T_1 v_i \ \ (1 \le i \le n), \qquad (6.2.8)$$

is the initial estimator of the intercept θ_0 and

$$S_0 = 1.483 \, median\{|\, \tilde{w}_i - T_0 \,|\} \qquad (6.2.9)$$

is the initial estimator of σ; $\left(y_{[i]}, v_{[i]}\right)$ are the concomitants of the ordered variates $z_{(i)} = \left(y_{[i]} - \dot{\theta}\, v_{[i]} - \theta_0\right)/\sigma \ \ (1 \le i \le n)$.

Dropping the ordering symbol,

$$\tilde{t}_i = \frac{y_i - T_0 - T_1 v_i}{S_0}. \qquad (6.2.10)$$

Replacing t_i by $\tilde{t}_i$, we obtain the initial estimates of α_i and β_i, namely, $\tilde{\alpha}_i$ and $\tilde{\beta}_i$.

The solutions obtained by using the linear approximation $g(z_i) \cong \tilde{\alpha}_i + \tilde{\beta}_i z_i \ (1 \le i \le n)$ are the following MMLEs:

$$\hat{\theta}_0 = \hat{\bar{y}} - \sum_{j=1}^{q} \hat{\theta}_j \hat{\bar{u}}_j + (\Delta/m)\hat{\sigma}, \ \hat{\theta} = K + D\hat{\sigma} \text{ and} \qquad (6.2.11)$$

$$\hat{\sigma} = \left\{B + \sqrt{\left(B^2 + 4nC\right)}\right\} \Big/ \left(2\sqrt{n(n-q-1)}\right) \qquad (6.2.12)$$

where

$$B=(b+1)\sum_{i=1}^{n}\Delta_i\left\{y_i-\hat{\bar{y}}-\sum_{j=1}^{q}K_j\mathrm{M}_{ij}\right\},\ C=(b+1)\sum_{i=1}^{n}\beta_i\left\{y_i-\hat{\bar{y}}-\sum_{j=1}^{q}K_j\mathrm{M}_{ij}\right\}^2,$$

$$K=(\mathrm{M}'\theta\mathrm{M})^{-1}(\mathrm{M}'\theta Y)=\left(K_j\right)_{qx1},\ D=(\mathrm{M}'\theta\mathrm{M})^{-1}(\mathrm{M}'\Lambda Y)=\left(D_j\right)_{qx1},$$

$$\mathrm{M}_{ij}=u_{ij}-\hat{\bar{u}}_j,\ \mathrm{M}=\left(\mathrm{M}_{ij}\right)_{nxq},\ \beta=diag(\beta_i)_{nxn},$$

$$\Delta_i=(b+1)^{-1}-\alpha_i,\ \Delta=\sum_{i=1}^{n}\Delta_i,\ \Lambda=diag\left(\Delta_i\right)_{nxn}$$

$$\hat{\bar{y}}=\sum_{i=1}^{n}\beta_i y_i\Big/\sum_{i=1}^{n}\beta_i \text{ and } \hat{\bar{u}}_j=\sum_{i=1}^{n}\beta_i u_{ij}\Big/\sum_{i=1}^{n}\beta_i\ \left(1\le i\le n, 1\le j\le q\right).$$

Here, $\hat{\theta}_0$ is estimating

$$\tau=\theta_0+\text{scaled median of the error distribution.}$$

In order to be able to compare LSEs and MMLEs, we define the LSEs of θ_0 and σ respectively as follows:

$$\tilde{\tau}=\bar{y}-\{\Psi(b)-\Psi(1)\}\tilde{\sigma} \text{ and } \tilde{\sigma}=s_e\Big/\sqrt{\Psi'(b)+\Psi'(1)}. \qquad (6.2.13)$$

Therefore, $\hat{\tau}$ and $\tilde{\tau}$ are estimating -0.777, 0, 1.056, 2.174, 2.819, and 3.268 for b = 0.5, 1, 2, 4, 6 and 8, respectively.

We calculate the MMLEs in three steps: First, we obtain the initial MMLEs $\hat{\theta}_{00}$, $\hat{\theta}_{j0}$ $\left(1\le j\le q\right)$ and $\hat{\sigma}_0$ by assuming that all θ_j $\left(1\le j\le q\right)$ are equal. Second, we use these initial MMLEs to calculate

$$\tilde{t}_i=\frac{y_i-\hat{\theta}_{00}-\sum_{j=1}^{q}\hat{\theta}_{j0}u_{ij}}{\hat{\sigma}_0}$$

and the new MMLEs. Finally, we use these new MMLEs to carry out one more iteration and obtain the final MMLEs.

6.2.1 Simulations

To study the efficiency and robustness of the MMLEs (6.2.11)-(6.2.12), we generated $N = [100{,}000/n]$ (integer value) samples of independently distributed random errors e_i of size n from generalized logistic distribution for the shape parameter b = 0.5, 1, 2, 4, 6 and 8. Remember that different values of b characterize different types of distributions. When b<1 and b>1, the distribution becomes negatively skewed and positively skewed, respectively, while for b=1, the distribution becomes symmetric and is the well known logistic distribution. The random errors e_i were multiplied by $[2\Psi'(1)/\{\Psi'(b)+\Psi'(1)\}]^{1/2}$. Without loss of generality, we assume that $\theta_0 = 0, \theta_j = 1\ (1 \le j \le q)$ and $\sigma = 1$. We generated the nonstochastic design variables x_{ij}'s $(1 \le i \le n, 1 \le j \le q)$ from a uniform distribution and, for illustration, $q = 4$ explanatory variables are considered. The LSEs (6.4) and (6.2.13) are also calculated for different values of shape parameter b. The means and variances of the resulting N values of MMLEs and LSEs are given in Table 6.5-Table 6.6 with $n = 20$ and Table 6.7-Table 6.8 with $n = 50$.

Table 6.5: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of τ, θ_1 and σ in a multiple linear regression model for generalized logistic distribution with shape parameter b; $n = 20$.

b	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
0.5	-0.853	-0.906	0.982	0.985	0.945	0.948
	[3.319]	[3.430]	[3.103]	[3.418]	[0.890]	[1.030]
1	-0.003	-0.004	0.996	0.997	0.987	0.977
	[3.434]	[3.471]	[3.249]	[3.442]	[0.897]	[0.940]
2	1.090	1.126	0.995	0.997	1.000	0.990
	[3.629]	[3.616]	[3.745]	[3.968]	[0.970]	[1.037]
4	2.245	2.299	0.985	0.997	0.999	0.991
	[3.764]	[3.785]	[3.284]	[3.513]	[0.983]	[1.101]
6	2.885	2.945	1.001	1.005	0.987	0.985
	[3.749]	[3.790]	[3.113]	[3.364]	[0.978]	[1.146]
8	3.350	3.412	1.002	0.993	0.990	0.989
	[3.664]	[3.705]	[3.893]	[3.963]	[1.031]	[1.203]

Table 6.6: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of θ_2, θ_3 and θ_4 in a multiple linear regression model for generalized logistic distribution with shape parameter b; $n = 20$.

b	$\hat{\theta}_2$	$\tilde{\theta}_2$	$\hat{\theta}_3$	$\tilde{\theta}_3$	$\hat{\theta}_4$	$\tilde{\theta}_4$
0.5	0.989	0.986	1.005	0.998	0.994	0.997
	[2.981]	[3.327]	[2.990]	[3.356]	[3.042]	[3.417]
1	1.004	1.002	0.994	0.994	0.995	0.993
	[3.585]	[3.825]	[3.279]	[3.409]	[3.473]	[3.661]
2	0.996	0.996	1.004	1.004	1.011	1.010
	[3.353]	[3.522]	[3.776]	[4.071]	[3.362]	[3.587]
4	0.996	0.995	1.002	1.000	1.011	1.009
	[4.630]	[4.983]	[3.557]	[3.822]	[5.128]	[5.541]
6	0.998	1.005	1.003	1.011	0.996	0.993
	[3.205]	[3.454]	[3.334]	[3.613]	[3.127]	[3.405]
8	0.988	0.994	0.990	0.991	1.001	1.005
	[3.989]	[4.364]	[5.757]	[5.577]	[4.940]	[5.036]

Table 6.7: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of τ, θ_1 and σ in a multiple linear regression model for generalized logistic distribution with shape parameter b; $n = 50$.

b	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\hat{\sigma}$	$\tilde{\sigma}$
0.5	-0.804	-0.886	0.984	0.982	0.948	0.966
	[3.371]	[3.391]	[2.728]	[3.174]	[0.767]	[1.023]
1	0.006	0.003	1.007	1.005	0.987	0.985
	[3.780]	[3.741]	[3.465]	[3.700]	[0.775]	[0.839]
2	1.089	1.143	1.008	1.009	1.010	1.007
	[3.930]	[3.773]	[3.109]	[3.379]	[0.830]	[0.995]
4	2.204	2.290	0.996	0.999	1.008	1.013
	[3.750]	[3.693]	[3.031]	[3.400]	[0.866]	[1.108]
6	2.853	2.949	0.998	0.996	1.002	1.010
	[3.727]	[3.769]	[3.626]	[4.011]	[0.867]	[1.135]
8	3.291	3.391	1.011	1.009	1.002	1.014
	[3.605]	[3.519]	[3.127]	[3.552]	[0.900]	[1.187]

Table 6.8: Simulated values of $(1/\sigma)$Mean and (n/σ^2)Variance (in brackets) of MMLEs and LSEs of θ_2, θ_3 and θ_4 in a multiple linear regression model for generalized logistic distribution with shape parameter b; $n = 50$.

b	$\hat{\theta}_2$	$\tilde{\theta}_2$	$\hat{\theta}_3$	$\tilde{\theta}_3$	$\hat{\theta}_4$	$\tilde{\theta}_4$
0.5	0.995	0.997	0.997	0.998	0.994	0.997
	[2.804]	[3.264]	[3.017]	[3.511]	[3.200]	[3.686]
1	1.007	1.007	0.999	1.000	1.004	1.004
	[3.085]	[3.439]	[3.118]	[3.367]	[3.502]	[3.754]
2	0.999	0.997	0.997	1.000	0.996	0.997
	[3.243]	[3.533]	[2.975]	[3.114]	[3.249]	[3.420]
4	1.003	1.003	1.002	1.004	0.998	0.998
	[2.992]	[3.415]	[2.976]	[3.300]	[3.115]	[3.422]
6	1.010	1.007	1.000	1.004	0.996	0.993
	[3.033]	[3.382]	[3.832]	[4.186]	[4.624]	[5.136]
8	0.998	0.994	0.995	0.994	1.005	1.001
	[2.976]	[3.321]	[3.010]	[3.389]	[3.157]	[3.596]

The results match those for simple linear regression model. Both the MMLEs and LSEs are unbiased (almost). However, the MMLEs are more efficient. They have also bounded influence functions; see Appendix C.

CHAPTER 7

APPLICATIONS

In this chapter, we consider a few real life data sets to illustrate the usefulness of the MMLEs developed in this thesis. We examined a large number of data sets given in Hand et al. (1994); some of them are reproduced in Tiku and Akkaya (2004). We found very few data sets which can appropriately be modeled by a normal distribution. This agrees with the findings of Pearson (1931), Geary (1947), Elveback et al. (1970) and Spjøtvoll and Aastveit (1980). We consider data sets which have specifically long-tailed symmetric or skew distributions. It is known that locating the underlying distribution exactly from a sample is not possible. However, by locating a distribution in reasonable proximity to the true distribution, MMLEs are known to give very good results for estimating the parameters of the underlying distribution (Tiku, 1967, 1968a,b, 1980; Tiku et. al, 1986, Islam et. al, 2001; Tiku et al, 2001). Constructing Q-Q plots or using goodness-of-fit tests help in identifying the shape parameter of the underlying distribution. Such techniques have been used in Tiku and Akkaya (2004).

Example 7.1: ***Cushny and Peebles prolongation of sleep data***

The data of Cushny and Peebles (1905) measure the prolongation of sleep by two soporofic drugs as ordered differences:

$$y: 0.0 \quad 0.8 \quad 1.0 \quad 1.2 \quad 1.3 \quad 1.3 \quad 1.4 \quad 1.8 \quad 2.4 \quad 4.6.$$

Under the normality assumption, the ideal estimates of the population mean and standard deviation are

$$\bar{y} = 1.58 \text{ and } s = 1.230.$$

However, Tiku and Akkaya (2004) show that normality assumption is not valid for this data. Both Shapiro-Wilk test and Q-Q plot support this conclusion. Since the Shapiro-Wilk test statistic is calculated as

$$W = 0.781$$

and is smaller than the 5% significance level 0.830, normality assumption is not appropriate. Surucu's (2008) test of normality is in agreement with that of Shapiro-Wilk. With the help of Q-Q plots, Tiku and Akkaya (2004) concluded that the Generalized Logistic with $b = 8$ is a plausible model for this data.

We simply assume that the data comes from the family of Generalized Logistic distributions and calculate the new MMLEs $\hat{\mu}$ and $\hat{\sigma}$ given in Chapter 3. Note that $\hat{\mu}$ is estimating the median

$$median = \left\{\mu - \ln(2^{1/b} - 1)\sigma\right\}$$

while $\hat{\sigma}$ is estimating the scale parameter σ.

We also calculated the corresponding LSEs as follows:

$$\tilde{\mu} = \bar{y} - \left\{\Psi\left(\hat{b}\right) - \Psi(1)\right\}\tilde{\sigma} \text{ and } \tilde{\sigma} = s\Big/\sqrt{\Psi'\left(\hat{b}\right) + \Psi'(1)}$$

where $\hat{b}$ is the estimator of the shape parameter b given by $\left\{\left(1 - \tilde{\bar{w}}\right)^{-1} - 1\right\}$ (see Section 3.2 for details).

The results are given below:

$\hat{\mu}$	$\tilde{\mu}$	$\hat{\sigma}$	$\tilde{\sigma}$
1.301	1.427	0.615	0.710

It is observed that the LSEs and MMLEs are in league. We have already illustrated that $\hat{\mu}$ has smaller bias than $\tilde{\mu}$ and the MMLEs are jointly more efficient tahn the LSEs.

Example 7.2: *Box-Cox data*

The Box and Cox (1964, p.220) data forms a 3x4 factorial experiment with four observations x_{ijl} $(1 \le i \le 3, 1 \le j \le 4, 1 \le l \le 4)$ for each combination of two factors. In this experiment, the survival times (10 hour units) of 48 animals exposed to two factors are recorded. One factor has 3 levels depending on the type of poison used while the other represents a treatment with 4 different levels. The allocation of animals is completely randomized.

Table 7.1: Box-Cox data.

	Treatment			
Poison	A	B	C	D
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

The Box-Cox data is known to have interaction (Schrader and McKean, 1977, p.889 and Brown, 1975). However, the F statistic based on the LSEs does not reject the hypothesis of "no interaction effects" although the data is known to have interaction. This perhaps results from the wrongful assumption of normality.

With the help of a Q-Q plot of the residuals, Senoglu and Tiku (2001) concluded that the underlying distribution is Generalized Logistic with shape parameter $b = 0.5$. In addition, they applied multi-sample goodness-of-fit test based on sample spacings of Tiku (1988, pp. 2382–83) and concluded that the assumption of the Generalized Logistic with $b = 0.5$ is not rejected at 10% significance level.

However, we do not assume any particular pre-determined value of *b*. We estimate it from the data. The assumed two-way classification fixed effects model is

$$y_{ijl} = \mu + \gamma_i + \delta_j + \tau_{ij} + e_{ijl} \quad (1 \le i \le 3, 1 \le j \le 4, 1 \le l \le 4).$$

The new MMLEs given in section 4.2.3 and the LSEs of the parameters in this two-way classification fixed effects model are calculated and given below:

Table 7.2: The Adaptive MMLEs and the LSEs of Box-Cox data.

$\hat{\mu}$	$\tilde{\mu}$	$\hat{\sigma}$	$\tilde{\sigma}$
0.476	0.475	0.092	0.084

Treatment	$\hat{\delta}_j$	$\tilde{\delta}_j$
1	-0.145	-0.153
2	0.178	0.184
3	-0.091	-0.089
4	0.056	0.058
$\sum_{j=1}^{4} \delta_j =$	0.000	0.000

Poison	$\hat{\gamma}_i$	$\tilde{\gamma}_i$
1	0.144	0.142
2	0.050	0.057
3	-0.199	-0.199
$\sum_{i=1}^{3} \gamma_i =$	-0.002	0.000

	Treatment							
	A		B		C		D	
Posion	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\tau}$	$\tilde{\tau}$
I	-0.031	-0.034	0.046	0.058	0.013	0.021	-0.027	-0.045
II	-0.055	-0.056	0.057	0.067	-0.059	-0.068	0.047	0.057
III	0.085	0.090	-0.120	-0.125	0.049	0.048	-0.013	-0.013
$\sum_{i=1}^{3} \tau_{ij} =$	-0.001	0.000	-0.017	0.000	0.003	0.000	0.007	0.000

Again, the modified maximum likelihood and the least squares estimates are close to one another. The former are, however, more precise since thay are jointly more efficient.

Example 7.3: ***Brownlee's stack loss data***

Brownlee (1965, p. 454) presented the following data obtained from 21 days of operation of a plant for the oxidation of ammonia to nitric acid.

Table 7.3: Brownlee's stack loss data.

Air Flow x_1	Water Temperature x_2	Acid Concentration x_3	Stack loss y
80	27	89	42
80	27	88	37
75	25	90	37
62	24	87	28
62	22	87	18
62	23	87	18
62	24	93	19
62	24	93	20
58	23	87	15
58	18	80	14
58	18	89	14
58	17	88	13
58	18	82	11
58	19	93	12
50	18	89	8
50	18	86	7
50	19	72	8
50	19	79	8
50	20	80	9
56	20	82	15
70	20	91	15

The response variable *y*, called "stack loss", is 10 times the percentage of the ingoing ammonia to the plant that is lost. A linear model is assumed, that is,

$$y_j = \theta_0 + \theta_1 u_{1j} + \theta_2 u_{2j} + \theta_3 u_{3j} + e_j \quad (1 \le j \le 20),$$

$$u_{ij} = \left(x_{ij} - \bar{x}_i\right) / s_i \quad (i = 1,2,3);$$

x_1 is "air flow" representing the rate of operation of the plant, x_2 is the temperature of the cooling water circulated through the coils in the absorption tower for the nitric acid, and x_3 is the concentration of acid circulating.

Examining the Q-Q plot of the residuals obtained by using the LSEs (see Andrews, 1974, p.530), it is seen that the smallest residual corresponding to the observation $\left(y = 15, x_1 = 70, x_2 = 20, x_3 = 91\right)$ is grossly anomalous. In fact, Andrews (1974) stated that after exclusion of this observation, the probability plot of residuals does not show any significant anomalies. Therefore, it is decided not to include this observation and base the estimation on the remaning n = 20 observations. The Q-Q plot of the 20 residuals indicates a long-tailed symmetric distribution. As a plausible value for the shape parameter p, Tiku and Akkaya (2004) suggest p = 2 which maximizes $\ln \hat{L}$, where $\hat{L}$ is the value of L with parameters equated to the corresponding MMLEs. We do not assume any particular value of p and calculate the MMLEs (6.1.11)-(6.1.12) and LSEs (6.4)-(6.5) given in Chapter 6. They are given below:

Table 7.4: The Adaptive MMLEs and the LSEs of Browlee's stack loss data.

	MMLE	LSE
θ_0	17.436	17.650
θ_1	8.077	7.915
θ_2	2.611	2.573
θ_3	-0.771	-0.562
σ	2.392	2.569

It can be seen that the LSEs and MMLEs are close to one another. However, the MMLE of θ_3 indicates a more potent effect of the acid concentration. The latter are, however, more precise as shown in Chapter 6.

CHAPTER 8

CONCLUSION

In this thesis, following Tiku and Surucu (2009), we have given a new innovation to MMLEs so that we can use them for machine data processing. They have bounded influence functions and we will call them “new” MMLEs. The question arises how good are these estimators as compared to the “old” MMLEs given below for ready reference.

For the long-tailed symmetric family (2.1), the old MMLEs of μ and σ are

$$\hat{\mu}_0 = (1/m)\sum_{i=1}^{n} \beta_i x_{(i)} \quad \left(m = \sum_{i=1}^{n} \beta_i \right) \text{ and } \hat{\sigma}_0 = \left\{ B + \sqrt{(B^2 + 4nC)} \right\} \Big/ 2n; \quad (8.1)$$

$$B = (2p/k)\sum_{i=1}^{n} \alpha_i (x_{(i)} - \hat{\mu}_0) \text{ and } C = (2p/k)\sum_{i=1}^{n} \beta_i \left(x_{(i)} - \hat{\mu}_0\right)^2 .$$

The coefficients α_i and β_i are given by

$$\alpha_i = (2/k)t_{(i)}^3 \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2 \quad \text{and} \quad \beta_i = \left\{1 - (1/k)t_{(i)}^2\right\} \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2 . \quad (8.2)$$

If $\beta_1 < 0$, α_i and β_i are replaced by α_i^* and β_i^* respectively:

$$\alpha_i^* = (1/k)t_{(i)}^3 \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2 \quad \text{and} \quad \beta_i^* = 1 \Big/ \left\{1 + (1/k)t_{(i)}^2\right\}^2 ; \quad (8.3)$$

this is done to ensure that $\hat{\sigma}$ is always real and positive. Also, the divisor $2n$ in (8.1) may be replaced by $2\sqrt{n(n-1)}$ as a bias correction. Here, we have used the divisor $2n$ as in Tiku et. al (2009).

Tiku and his collaborators (Tiku and Akkaya, 2010; Tiku et al., 2009; Islam and Tiku, 2009; Tiku and Senoglu, 2009; Tiku et al., 2008; Akkaya and Tiku, 2008) contend that for an assumed distribution (having finite variance) and its plausible alternatives, the old MMLEs have no or negligible bias and are highly efficient (in terms of having smaller variances). Assume that the underlying distribution is long-tailed symmetric (2.1) with $p = 3.5$. Out of the twelve models considered in section 2.2, plausible alternatives would be

$$\text{(1), (2), (3) is the true model, (4), (6) and (8).} \qquad (8.4)$$

Given in Table 8.1 are the simulated values (based on $[100{,}000/n]$ Monte Carlo runs) of the means and variances of the MMLEs of μ and σ. From these values, we conclude that for an assumed distribution (having finite variance) and its plausible alternatives, one should prefer the old MMLEs because (i) both $\hat{\mu}_0$ and $\hat{\sigma}_0$ have no or negligible bias and (ii) $\hat{\mu}_0$ is as efficient as the new $\hat{\mu}$. The new $\hat{\sigma}$ has of course smaller variance but, unfortunately, it inherits substantial bias. Correcting for bias would pose no problem if it was known but it is not because the exact distribution is not known.

Table 8.1: Means and *n*xVariances of the new and old MMLEs.

		μ				σ			
		Mean		Variance		Mean		Variance	
	Model	New	Old	New	Old	New	Old	New	Old
$n = 10$	1	-0.01	-0.01	1.06	1.03	0.93	1.04	0.58	0.64
	2	0.00	0.00	0.95	0.94	0.90	1.03	0.66	0.80
	3	0.00	0.00	0.90	0.90	0.87	1.01	0.69	0.94
	4	0.00	-0.01	0.77	0.81	0.81	0.97	0.69	1.20
	6	0.00	0.00	0.95	0.94	0.89	1.02	0.59	0.73
	8	0.00	0.00	0.95	0.95	0.89	1.02	0.63	0.79
$n = 20$	1	-0.01	-0.01	1.01	1.01	0.96	1.06	0.56	0.64
	2	0.00	0.00	0.94	0.93	0.92	1.03	0.65	0.77
	3	0.00	0.00	0.90	0.90	0.89	1.01	0.64	0.81
	4	0.00	0.00	0.75	0.77	0.83	0.97	0.65	1.10
	6	0.00	0.00	0.98	0.97	0.92	1.03	0.57	0.67
	8	0.00	0.00	0.91	0.91	0.92	1.03	0.64	0.77
$n = 50$	1	0.00	0.00	1.08	1.08	0.97	1.08	0.54	0.63
	2	0.00	0.00	0.94	0.94	0.93	1.05	0.66	0.82
	3	0.00	0.00	0.87	0.89	0.90	1.03	0.67	0.96
	4	0.00	0.00	0.73	0.78	0.84	0.99	0.68	1.32
	6	0.00	0.00	0.96	0.95	0.93	1.05	0.56	0.71
	8	0.00	0.00	0.93	0.93	0.93	1.05	0.60	0.78
$n = 100$	1	0.00	0.00	1.01	1.00	0.97	1.07	0.53	0.63
	2	0.00	0.00	0.88	0.88	0.94	1.05	0.61	0.72
	3	0.00	0.00	0.92	0.93	0.91	1.02	0.67	0.86
	4	0.00	0.00	0.70	0.73	0.84	0.98	0.68	1.29
	6	0.00	0.00	0.96	0.96	0.94	1.04	0.54	0.64
	8	0.00	0.00	0.90	0.89	0.94	1.04	0.60	0.71

Consider now the models (5), (7), (9), (10), (11), and (12) which represent strong deviations from the assumed distribution ($p = 3.5$ in (2.1)). Given in Table 8.2 are the simulated values similar to those in Table 8.1 . We reproduce the values only for $n = 10$ and 50 for conciseness. Realize that the models (5), (7) and (9) have finite variances, model (10) has finite mean but nonexistent variance and models (11) and (12) have nonexistent means and variances. It is clear that the old MMLEs should not be used for models (10)-(12). For models (5), (7) and (9), the old MMLEs may be used: $\hat{\mu}_0$ is unbiased and $\hat{\sigma}_0$ has negligible bias although, as compared to the new MMLEs, $\hat{\mu}_0$ has somewhat larger variances and $\hat{\sigma}_0$ has substantially larger variances. But, the new MMLE $\hat{\sigma}$ has substantial bias. For machine data processing, however, the new MMLEs should be used with the clear understanding that $\hat{\sigma}$ can have substantial bias particularly for distributions of extreme type. If efficient estimation of only the location parameter μ is intended, the new MMLE $\hat{\mu}$ should always be used. It is indeed pleasing to note that the new MMLEs are overall more efficient than Huber M-estimators.

Table 8.2: Means and *n*xVariances of the new and old MMLEs for models representing strong deviations from the assumed distribution.

		μ				σ			
		Mean		Variance		Mean		Variance	
	Model	New	Old	New	Old	New	Old	New	Old
$n = 10$	5	0.00	0.00	0.58	0.67	0.71	0.90	0.62	1.95
	7	0.00	0.00	0.55	0.64	0.72	0.94	0.47	1.19
	9	0.00	0.00	0.59	0.70	0.72	0.92	0.68	1.64
	10	0.00	0.01	2.18	4.80	1.41	2.17	3.34	*
	11	0.01	-0.12	4.77	*	2.07	16.21	14.15	*
	12	0.00	0.51	9.53	*	2.85	25.05	23.97	*
$n = 50$	5	0.00	0.00	0.57	0.68	0.73	0.94	0.55	1.89
	7	0.00	0.00	0.53	0.61	0.75	0.95	0.42	1.03
	9	0.00	0.00	0.57	0.65	0.75	0.94	0.59	1.63
	10	0.00	-0.01	2.00	3.10	1.44	2.24	2.85	*
	11	0.00	0.02	3.59	*	1.95	22.23	9.44	*
	12	0.01	0.17	7.15	*	2.77	29.77	16.66	*

** Represents a very large value.*

To test the null hypothesis $H_0 : \mu = 0$, the test statistics based on the old and new MMLEs are

$$T_0 = \sqrt{m}\, \hat{\mu}_0 / \hat{\sigma}_0 \text{ and } T = \sqrt{m}\, \hat{\mu} / \hat{\sigma} \qquad (8.5)$$

respectively; σ^2 / m is the minimum variance bound for estimating μ; for $p = 3.5$, $m = 1.167n$. Large values of T_0 (and T) lead to the rejection of H_0 in favor of $H_1 : \mu > 0$.

Since $\hat{\mu}_0$ is a linear function of order statistics and $\hat{\sigma}_0$ converges to its expected value $E(\hat{\sigma}_0)$ as *n* becomes large, the asymptotic null distribution of T_0 is normal with mean zero. Its asymptotic standard deviation is

$$SD_0 = \sqrt{1.167 \frac{nVar(\hat{\mu}_0)}{\{E(\hat{\sigma}_0)\}^2}}. \tag{8.6}$$

The asymptotic null distribution of T is also normal because $\hat{\sigma}$ converges to its expected value and $\hat{\mu}$ is the mean of bounded iid variables. Under H_0, $E(T) = 0$. The standard deviation of T is

$$SD = \sqrt{\frac{nVar(\hat{\mu})}{\{E(\hat{\sigma})\}^2}}. \tag{8.7}$$

The percentage points of the null distributions of T_0 and T can be approximated by those of normal $N(0,1)$ if SD_0 and SD are 1 (or close to 1). Given below are the values of the standard deviations of T and T_0, respectively.

Table 8.3: The values of the standard deviations of T and T_0; SD and SD_0, respectively.

	$n = 20$		$n = 50$		$n = 100$	
Model	SD	SD_0	SD	SD_0	SD	SD_0
1	1.05	1.02	1.07	1.04	1.04	1.01
2	1.05	1.01	1.04	1.00	1.00	0.97
3	1.07	1.01	1.04	0.99	1.05	1.02
4	1.04	0.98	1.02	0.96	1.00	0.94
5	1.01	0.92	1.03	0.95	1.00	0.92
6	1.08	1.03	1.05	1.00	1.04	1.02
7	1.00	0.90	0.97	0.98	0.99	0.93
8	1.04	1.00	1.04	0.99	1.01	0.98
9	1.02	0.93	1.01	0.93	0.98	0.91
10	1.04	*	0.98	*	0.96	*
11	0.98	*	0.97	*	0.94	*
12	0.99	*	0.97	*	0.96	*

The normal distribution $N(0,1)$ does indeed provide accurate approximations to the percentage points for $n \geq 20$. For $n < 20$, Student's t distribution with $n-1$ degrees of freedom provides accurate approximations to the percentage points of the null distributions of T_0 and T; see also Tiku and Surucu (2009).

Simulations and asymptotic mathematics reveals that for an assumed long-tailed symmetric distribution (having finite variance) and its plausible alternatives, the T_0 test has somewhat higher power than T test. For others (e.g., those considered in Table 8.2), the T test is somewhat more powerful. In machine data processing, however, the T test should be used.

For the skew family of Generalized Logistic distributions, we have also given a new innovation to the method of MML estimation. The method includes estimation of the shape parameter from bounded emprical functions. We have shown that the resulting estimators are more efficient than the least squares estimators. This approach can perhaps be extended to other families of skew distributions. That will be the subject matter of future research.

Besides single sample estimation and hypothesis testing, we have extended the above methods to experimental design (one-way and two-way classification) and linear and multiple linear regression. We believe that the methods can be extended to more complex data structures, e.g., time series, autoregression, multivariate data, etc.

REFERENCES

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

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics* 16, 523-531.

Andrews, D. F., Bickel, P. J., Hampel, F., Huber, P., Rogers, W., and Tukey, J. W. (1972). *Robust Estimation of Location: Survey and Advances.* Princeton University press: Princeton, N. J.

Balakrishnan, N. and Leung, W. Y. (1988). Means, variances and covariances of order statistics, BLUE for the Type I generalized logistic distribution and some applications. *Commun. Stat.- Simula.* 17, 51-84.

Barnett, V. D. (1966). Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots. *Biometrika* 53, 151–165.

Barton, D. E. and Dennis, K. E. (1952). The condition under which Gram Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 39, 425-427.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* 16, 147-186.

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.

Birch, J. B. and Myers, R. H. (1982). Robust analysis of covariance. *Biometrics* 38, 699-713.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with Discussion). *Journal of the Royal Statistical Society* B26, 211–252.

Brown, M. B. (1975). Exploring interaction effects in the ANOVA. *Journal of the Royal Statistical Society* C24, 288–298.

Brownlee, K. A. (1965) *Statistical Theory and Methodology in Science and Engineering* (2nd ed.). John Wiley: New York.

Cushny, A. R. and Peebles, A. R. (1905). The action of optical isomers. II. Hyoscines. *J. Physiol.* 32, 501–510.

Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *J. Amer. Stat. Assoc.* 63, 660-667.

Dunnett, C. W. (1982). Robust multiple comparisons. *Commun. Stat.-Theory Meth.* 11, 2611-2629.

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.

Gayen, A. K. (1949). The distribution of Student's t in random samples of any size drawn from non-normal universe. *Biometrika* 36, 353-369.

Gayen, A. K. (1950). The distribution of the variance-ratio of random samples of any size drawn from non-normal universes. *Biometrika* 37, 236-255.

Geary, R. C. (1947). Testing for normality. *Biometrika* 34, 209-220.

Gross, A.M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J.Amer.Stat.Assoc*. 71, 409-416.

Hampel, F.R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* 69, 383–394.

Hampel, F. R., Ronchetti, E. M. and Rousseeuw, P. J. (1986). *Robust Statistics*. John Wiley: New York.

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

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals. Math. Stat*. 35, 73-101.

Huber, P. J. (1981). *Robust Statistics*. John Wiley: New York.

Islam, M. Q., Tiku, M. L. & Yildirim, F. (2001). Non-normal regression, I: Skew distributions. *Commun. Stat.-Theory Meth.* 30, 993-1020.

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

Islam, M. Q. and Tiku, M. L. (2009). Multiple linear regression model with stochastic design variables. *J. Applied Stat*. DOI: 10.1080/02664760902939612 (to be published).

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

Kendall, M. G. and Stuart, A. (1968). *The Advanced Theory of Statistics*, Vol. 1. McMillan: New York.

Lee, A. F. S., Kapadia, C. H. and Dwight, B. B. (1980). On estimating the scale parameter of Rayleigh distribution from doubly censored samples. *Statist. Hefte* 21, 14–29.

More, J. J., Garbow, B. S. and Hillstrom, K. E. (1980). *User guide for MINPACK-1*, Argonne National Labs Report ANL-80-74, Argonne, Illinois.

Oral, E. (2006). Binary regression with stochastic covariates. *Commun. Stat.-Theory Meth*. 35, 1429-1447.

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

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.

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

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 Universtiy: Ankara.

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

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

Schrader, R. M. and McKean, J. W. (1977). Robust analysis of variance. *Commun. Stat. Theory Meth.* A6, 879–894.

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

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

Senoglu, B. and Tiku, M. L. (2002). Linear contrast in experimental design with non-identical error distributions. *Biometrical J.* 44, 359-374.

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

Srivastava, A. B. L. (1958). Effect of non-normality on the power function of t-test. *Biometrika* 45, 421-429.

Srivastava, A. B. L. (1959). Effect of non-normality on the power of the analysis of variance test. *Biometrika* 46, 114-122.

Surucu, B. (2008). A power comparison and simulation study of goodness-of-fit tests, *Computers and Mathematics with Applications* 56, 1617-1625.

Tan, W. Y. and Wong, S. P. (1977). On the Roy-Tiku approximation to the distribution of sample variances from non-normal universes. *J. Amer. Stat. Assoc*. 72, 875-881.

Tan, W. Y. and Tabatabai, M. A. (1988). A modified Winsorized regression procedure for linear models. *Journal of Statistical Computation and Simulation* 30, 299-313.

Tan, W. Y. and Tiku, M. L. (1999). *Sampling Distributions in terms of Laguerre Polynomials with Applications*. New Age International (Formerly, Wiley Eastern): New Delhi.

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions. *Biometrika* 51, 83-95.

Tiku, M. L. (1965). Laguerre series forms of non-central χ^2 and F distributions. *Biometrika*. 52, 415-427.

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

Tiku, M. L. (1968a). Estimating the parameters of log-normal distribution from censored samples. *J. Amer. Stat. Assoc.* 63, 134-140.

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

Tiku, M. L. (1971a). Student's t distribution under non-normal situations. *Australian J. Stat.* 13, 142-148.

Tiku, M. L. (1971b). Power function of the F-test under non-normal situations. *J. Amer. Stat. Assoc.* 66, 913-916.

Tiku, M. L. (1980). Robustness of MML estimators based on censored samples and robust test statistics. *J. Stat.Plann. Inf.* 4, 123-143.

Tiku, M. L. (1985a). *Noncentral Chi-square Distribution*. Encyclopedia of Statistical Sciences, Vol. 6. (Eds., Johnson, N. L. And Kotz, S.). John Wiley: New York.

Tiku, M. L. (1985b). *Noncentral F Distribution*. Encyclopedia of Statistical Sciences, Vol. 6. (Eds., Johnson, N. L. And Kotz, S.). John Wiley: New York.

Tiku, M. L. (1988). Order statistics in goodness-of-fit tests. *Commun. Stat.-Theory Meth.* 17, 2369-2387.

Tiku, M. L. (1989). *Modified Likelihood Estimation*. In: Kotz, S. and Johnson, N.L. (Eds), Encyclopedia of Statistical Sciences, Supplement Volume.

Tiku, M. L. and Kumra, S. (1981). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). *Selected Tables in Mathematical Statistics* 8, 141-270. American Mathematical Society: Providence, RI, USA.

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

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.

Tiku, M. L., Islam, M. Q. and Selcuk, A. (2001). Non-normal regression, II. Symmetric distributions. *Commun. Statist.-Theory Meth.* 30, 1021-1045.

Tiku, M. L. and Akkaya, A. D. (2004). *Robust Estimation and Hypothesis Testing*. (1st ed.). New Age International (Wiley Eastern): New Delhi.

Tiku, M. L., Islam, M. Q. and Sazak, H. S. (2008). Estimation in bivariate nonnormal distributions with stochastic variance functions. *Comp. Stat. & Data Analysis* 52, 1728-1745.

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

Tiku, M. L. and Senoglu, B. (2009). Estimation and hypothesis testing BIB design and robustness. *Comp. Stat. & Data Analysis* 53, 3439-3451.

Tiku, M. L. and Surucu, B. (2009). MMLEs are as good as M-Estimators or better. *Statistics & Probability Letters* 79, 984-989.

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

Vaughan, D.C. (1992). On the Tiku-Suresh method of estimation. *Commun. Stat.-Theory Meth.* 21, 451–469.

Vaughan, D. C. (2002). The generalized secant hyperbolic distribution and its properties. *Commun. Stat.-Theory Meth*. 31, 219-238.

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

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S.* (4th ed.). Springer: New York.

APPENDIX A

THE ASYMPTOTIC PROPERTIES OF $\hat{\mu}_x$ AND $\hat{\sigma}_x$

To evaluate the asymptotic properties of $\hat{\mu}_x$, we first note that as n tends to infinity, w/n converges to its expected value (assuming that the variance of X is finite). To evaluate the expected value of w and the variance of $\hat{\mu}_x$, we first note that

$$\int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-j}dz = \sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(j-\frac{1}{2}\right)\Big/\Gamma(j) \quad (j \geq 1). \qquad \text{(A.1)}$$

Asymptotically, $T_0 \cong E(T_0) = \mu$ for symmetric distributions. Consider the situation when asymptotically $S_0 \cong E(S_0) = \sigma$. For the family (1.6.1),

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}E\sum_{i=1}^{n}w_i &= \frac{\Gamma(p)}{\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(p-\frac{1}{2}\right)}\int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-(p+2)}dz \\ &= \left(p-\frac{1}{2}\right)\left(p+\frac{1}{2}\right)\Big/p(p+1).\end{aligned} \qquad \text{(A.2)}$$

Clearly, $E(\hat{\mu}_x) = \mu$. This follows from symmetry.

Now

$$\begin{aligned}V(\hat{\mu}_x) &\cong \frac{\sigma^2}{n[E(w/n)]^2}E\left(z_i^2\, w_i^2\right) \\ &= \frac{\sigma^2}{n}\frac{p^2(p+1)^2}{\left(p-\frac{1}{2}\right)^2\left(p+\frac{1}{2}\right)^2}\frac{\Gamma(p)}{\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(p-\frac{1}{2}\right)}\int_{-\infty}^{\infty}z^2\left(1+\frac{1}{k}z^2\right)^{-(p+4)}dz\end{aligned}$$

and

$$\int_{-\infty}^{\infty} z^2\left(1+\frac{1}{k}z^2\right)^{-(p+4)} dz = k\int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-(p+3)} dz - k\int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-(p+4)} dz$$

$$= \frac{k}{2}\frac{\left(p-\frac{1}{2}\right)\left(p+\frac{1}{2}\right)\left(p+\frac{3}{2}\right)}{p(p+1)(p+2)(p+3)}.$$

Thus,

$$V(\hat{\mu}_x) = \frac{\sigma^2}{n}\frac{p(p+1)}{(p+2)(p+3)}\frac{\left(p^2-\frac{9}{4}\right)}{\left(p^2-\frac{1}{4}\right)} < \frac{\sigma^2}{n}. \qquad \text{(A.3)}$$

We conclude that $\hat{\mu}_x$ is more efficient than the sample mean $\bar{x}$ for the family (1.6.1). This is a very interesting result. Since B^2 is much smaller than nC and $B/\sqrt{nC} \cong 0$, $\hat{\sigma}_x^2 \cong C/n$. Therefore,

$$E(\hat{\sigma}_x^2) \cong \sigma^2\frac{2p}{k}E(w_i\ z_i^2)$$

$$= \sigma^2\, 2p\frac{\Gamma(p)}{\sqrt{k}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(p-\frac{1}{2}\right)}\left[\int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-(p+1)} dz - \int_{-\infty}^{\infty}\left(1+\frac{1}{k}z^2\right)^{-(p+2)} dz\right]$$

$$= \frac{p-1/2}{p+1}\sigma^2\ ;$$

$$E(\hat{\sigma}_x) \cong \sqrt{\frac{p-1/2}{p+1}}\,\sigma$$ (lower bound since $\hat{\sigma}_x^2$ is some what greater than C/n). (A.4)

For $p = 16.5$, (A.4) assumes the value 0.96σ.

When $E(S_0) \neq \sigma$, the expressions for $V(\hat{\mu}_x)$ and $E(\hat{\sigma}_x^2)$ can be obtained from the equation although the algebra is a little bit involved,

$$g(S_0) \cong g(\sigma) + (S_0 - \sigma)\{g'(S_0)\}_{S_0=\sigma}\ . \qquad \text{(A.5)}$$

For Tiku - Surucu estimator (2.1.1),

$$E(\hat{\sigma}_x) \cong \sqrt{\frac{\left(p-\frac{1}{2}\right)}{(p+1)} \frac{p(p+1)}{\left(p-\frac{1}{2}\right)\left(p+\frac{1}{2}\right)}}\,\sigma = \sqrt{\frac{p}{p+\frac{1}{2}}}\,\sigma\,. \qquad \text{(A.6)}$$

For $p = 16.5$, (A.6) assumes the value 0.99σ.

APPENDIX B

EMPRICAL INFLUENCE FUNCTION OF THE MMLEs

To show that the estimators developed in Chapter 3 have bounded influence functions, we simply have to show that the i^{th} terms in the expressions for m, Δ_i, K, D, B and C tend to zero as x_i (equivalently, $\tilde{t}_i$) tends to infinity.

From equation (3.3.2), it immediately follows that $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ tend to zero as $\tilde{t}_i$ tends to infinity, Now,

$$\Delta_i = \sum_{i=1}^{n} \left\{ \frac{1}{n} - \frac{1}{n} \tilde{w}_i \right\} - \alpha_i \to 0 \text{ as } \tilde{t}_i \to \infty \text{ because } \tilde{w}_i \to 1.$$

Because of equation (3.3.2), and $\tilde{\tilde{w}}_i$ being bounded between 0 and 1, all the i^{th} terms mentioned above tend to zero as $\tilde{t}_i$ tends to ∞ (or $-\infty$).

Given in *Figure B.1* is the emprical influence function of $\hat{\mu}_x$ for illustration. It confirms the high breakdown of $\hat{\mu}_x$.

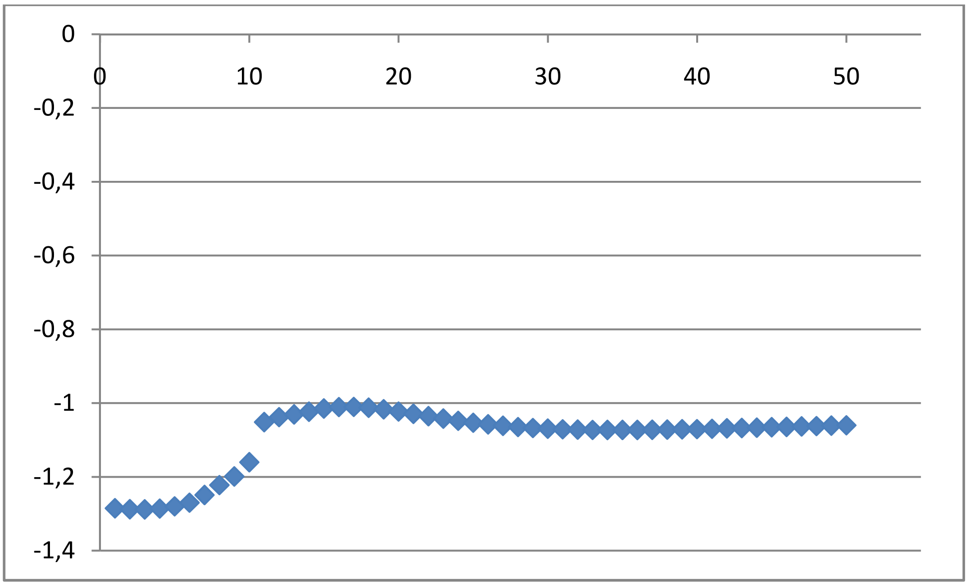

Figure B.1: Emprical influence function of $\hat{\mu}_x$, b = 0.5.

Remark**:** It may be noted that T_0 and S_0 also have bounded influence functions. However, because of the low joint efficiency, they cannot be prefered over the MMLEs developed in Chapter 3.

APPENDIX C

THE ASYMPTOTIC PROPERTIES OF THE MMLEs

***Lemma*:** For large *n*, the MMLEs are unbiased.

***Proof*:** If $\hat{\theta}_1$ and $\hat{\theta}_2$ are statistics calculated from a random sample of size *n*, then for large *n* (Kendall and Stuart, 1968)

$$E\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) \cong \frac{E\left(\hat{\theta}_1\right)}{E\left(\hat{\theta}_2\right)}. \tag{C.1}$$

Consider the MMLEs given in (5.1.11) and note that for large $p\ (\geq 10)$

$$T_0 \cong \mu \text{ and } S_0 \cong \sigma, \tag{C.2}$$

because long tailed symmetric *p*-family is indistinguishable from normal $N(0,1)$ for $p \geq 10$. Now,

$$E\left(\hat{\bar{u}}\right) \cong \frac{1}{E(m)} \sum_{i=1}^{n} u_i E(\beta_i), \quad E(m) = \sum_{i=1}^{n} E(\beta_i).$$

Since $E(\beta_i)$ $(1 \leq i \leq n)$ are all equal, it follows that $E\left(\hat{\bar{u}}\right) \cong \bar{u} = 0$. Similarly,

$$E\left(\hat{\theta}_0\right) \cong E\left(\hat{\bar{y}}\right) \cong \theta_0,\ E(D) \cong 0, E\left(\hat{\theta}_1\right) \cong E(K) \cong \theta_1 \tag{C.3}$$

and (see equation A.4 in Appendix A)

$$E(\hat{\sigma}) \cong \sqrt{(2p/k)E(C)/n} = \sigma\sqrt{(2p/k)E\left(w_i z_i^2\right)} \quad \left(w_i = \left(1 + z^2/k\right)^{-2},\ z_i = e_i/\sigma\right)$$

$$\cong \sigma\sqrt{\frac{p - 1/2}{p+1}} = 0.96\sigma \quad \text{for } p = 16.5\,;$$

this is, in fact, a lower bound since $\hat{\sigma}^2 \geq (2p/k)C/n$.

***Lemma*:** For large *n*, the variance of $\hat{\theta}_1$ is

$$\left[\frac{\dfrac{(p-3/2)(p-1/2)(p+1/2)(p+3/2)}{p(p+1)(p+2)(p+3)}\sum_{i=1}^{n}u_i^2}{\dfrac{(p-1/2)(p+1/2)}{p(p+1)}\sum_{i=1}^{n}u_i^4+\dfrac{(p-1/2)^2}{p^2}\sum\sum_{i\neq j}u_iu_j}\right]\sigma^2. \qquad \text{(C.4)}$$

***Proof*:** For large *n*,

$$E\left(\hat{\theta}_1\right)\cong E\left(\sum_{i=1}^{n}\beta_iu_iy_i\right)\Big/E\left(\sum_{i=1}^{n}\beta_iu_i^2\right)$$

since $\hat{\bar{u}}=\frac{1}{m}\sum_{i=1}^{n}\beta_i u_i$ converges to its expected value $\bar{u}=0$. Now,

$$E\sum_{i=1}^{n}\beta_iu_iy_i=E\sum_{i=1}^{n}\beta_iu_i(\theta_0+\theta_1u_i+\sigma z_i)\ \ (z_i=e_i/\sigma)$$

$$\cong\theta_1\sum_{i=1}^{n}\beta_iu_i^2+\sigma\sum_{i=1}^{n}\beta_iu_iz_i\quad\text{since}\ \ (1/n)\sum_{i=1}^{n}\beta_iu_i=\hat{\bar{u}}\cong\bar{u}=0$$

and

$$E\sum_{i=1}^{n}\beta_i\left(u_i-\hat{\bar{u}}\right)^2\cong E\sum_{i=1}^{n}\beta_i u_i^2\ . \qquad \text{(C.5)}$$

Therefore,

$$E\left(\hat{\theta}_1^2\right)\cong E\left(K^2\right)=E\left[\theta_1+\sigma\frac{\sum_{i=1}^{n}\beta_iu_iz_i}{\sum_{i=1}^{n}\beta_iu_i^2}\right]^2$$

$$=\theta_1^2+\sigma^2E\left[\frac{\left(\sum_{i=1}^{n}\beta_iu_iz_i\right)^2}{\left(\sum_{i=1}^{n}\beta_iu_i^2\right)^2}\right]+2\theta_1\sigma E\left[\frac{\sum_{i=1}^{n}\beta_iu_iz_i}{\sum_{i=1}^{n}\beta_iu_i^2}\right].$$

Using the fact that $\beta_i z_i$ is a symmetric function over $(-\infty,\infty)$ and, for large *n*, $E\left(\hat{\theta}_1/\hat{\theta}_2\right)\cong E\left(\hat{\theta}_1\right)/E\left(\hat{\theta}_2\right)$ we have

$$E\left(K^2\right)\cong\theta_1^2+\sigma^2\left\{E\left(\sum_{i=1}^{n}\beta_iu_iz_i\right)^2\Big/E\left(\sum_{i=1}^{n}\beta_iu_i^2\right)^2\right\}. \qquad \text{(C.6)}$$

Thus,

$$Var\left(\hat{\theta}_1\right) \cong \sigma^2 \left\{ E\left(\sum_{i=1}^{n} \beta_i u_i z_i\right)^2 \Big/ E\left(\sum_{i=1}^{n} \beta_i u_i^2\right)^2 \right\} \tag{C.7}$$

$$= \left[\frac{\dfrac{(p-3/2)(p-1/2)(p+1/2)(p+3/2)}{p(p+1)(p+2)(p+3)} \sum_{i=1}^{n} u_i^2}{\dfrac{(p-1/2)(p+1/2)}{p(p+1)} \sum_{i=1}^{n} u_i^4 + \dfrac{(p-1/2)^2}{p^2} \sum\sum_{i \neq j} u_i u_j} \right] \sigma^2$$

from equations (A.2)-(A.6) in Appendix A.

For $p = \infty$, (C.4) reduces to

$$\sigma^2 \Big/ \sum_{i=1}^{n} u_i^2 = \sigma^2 / n .$$

For $p = 16.5$, (C.4) reduces to

$$\frac{0.705 \sum_{i=1}^{n} u_i^2}{0.9420 \sum_{i=1}^{n} u_i^4 + 0.9403 \sum\sum_{i \neq j} u_i u_j} \sigma^2 \cong \frac{0.705 \sum_{i=1}^{n} u_i^2}{0.940 \left(\sum_{i=1}^{n} u_i^2 \right)^2} \sigma^2 = 0.75 \frac{\sigma^2}{n} .$$

The estimator $\hat{\theta}_1$ is considerably more efficient than the LSE $\tilde{\theta}_1$, particulary for large *n*.

APPENDIX D

FORTRAN PROGRAM CALCULATING THE MMLEs AND THE PROPOSED LSEs OF GENERALIZED LOGISTIC DISTRIBUTION FOR ONE SAMPLE CASE

```
c************************************************************************
c       Written by  Ayca DONMEZ, 2009, Ankara
c************************************************************************
c------------------------------------------------------------------------
c       Calculates the LSEs and MMLEs of GL (Generalized Logistic) distribution
c       for different b values, where b is unknown. The MML estimators of GL
c       with one block case is studied. The sample size is n. The samples are
c       generated from GL with mu=0.0, sigma=1.0 and b=0.5, 1,2,4,6,8.
c------------------------------------------------------------------------
       use numerical_libraries

c     Declaration of the variables and arrays

       parameter n=10
       parameter nn=floor(100000/(n*1.0))
       parameter sigma=1.0, mu=0.0
       parameter iteration=5
       parameter dpsii1=1.6449

       real y(n), resultMML(3)
       real T0,S0, preS0(n),median_preS0
       real mu_0, sigma_0
       real correction

       real muMML(nn), sigmaMML(nn)
       real mean_muhatMML, mean_sigmaMML
       real var_muhatMML, var_sigmaMML

       real xbardot(nn), s2(nn)
       real mutilda(nn), sigmatilda(nn)
       real mean_mutilda, mean_sigmatilda
       real var_mutilda, var_sigmatilda

       real estb, estbponeinvm(nn)
       real Ezestb, Vzestb

c     Declaration of the functions
      real GL_rnd, GL_invcdf, GL_MML_iteration
      real mymean, mysort, mymedian, variance, psidash

       do l=1,6

              if (l==1) then
                     b=0.5
              elseif (l==2) then
                     b=1.0
              elseif (l==3) then
```

```
                b=2.0
        elseif (l==4) then
                b=4.0
        elseif (l==5) then
                b=6.0
        elseif (l==6) then
                b=8.0
        endif

        Ez=psi(b)-psi(1.0)
        Vz=psidash(b)+psidash(1.0)
        correction=sqrt(Vz/(2.0*psidash(1.0)))

c       Loop for simulation runs, 100 k
        do 100 k=1,nn

                call GL_rnd(mu,sigma,b,n,y)

c               T0 is equal to median of y_i`s:
                call mymedian(n,y,T0)

c               S0=1.483*median{|y_i-T0|}
                do j=1,n
                  preS0(j)=abs(y(j)-T0)
                enddo

                call mymedian(n,preS0,median_preS0)

                S0=1.483*median_preS0

                mu_0=T0
                sigma_0=S0

c               Iteration loop 200, ii
                do 200 ii=1,iteration

                        call GL_MML_iteration(y,n,mu_0,sigma_0,resultMML)

                        mu_0=resultMML(1)
                        sigma_0=resultMML(2)

c               End of iteration loop 200, ii
200             continue

                muMML(k)=resultMML(1)/correction
                sigmaMML(k)=resultMML(2)/correction

                estbponeinvm(k)=resultMML(3)
c               Note that estbponeinvm is the estimator of 1/(b+1).
c               However, we need estimator of b:
                estb=1/estbponeinvm(k)-1.0
                Ezestb=psi(estb)-psi(1.0)
                Vzestb=psidash(estb)+psidash(1.0)

                xbardot(k)=mymean(y,n)
                s2(k)=variance(y,n,0)

                sigmatilda(k)=sqrt(s2(k)/Vzestb)/correction
                mutilda(k)=xbardot(k)-Ezestb*sigmatilda(k)
                mutilda(k)=mutilda(k)/correction

c       End of simulation runs, 100 k
100     continue

        mean_muhatMML=mymean(muMML,nn)
        mean_sigmaMML=(1/sigma)*mymean(sigmaMML,nn)
        mean_mutilda=mymean(mutilda,nn)
        mean_sigmatilda=(1/sigma)*mymean(sigmatilda,nn)
        mean_estbponeinv=mymean(estbponeinvm,nn)

        var_muhatMML=(n/(sigma**2.0))*variance(muMML,nn,1)
        var_sigmaMML=(n/(sigma**2.0))*variance(sigmaMML,nn,1)
        var_mutilda=(n/(sigma**2.0))*variance(mutilda,nn,1)
        var_sigmatilda=(n/(sigma**2.0))*variance(sigmatilda,nn,1)
```

```
1000   FORMAT (A6,F7.3,A5,A6,F7.3,A5)
1111   FORMAT (A10,I3,A7,F3.1)

       print 1111, 'n  = ',n,'  b = ',b
       print*, '-------------------------------------------'
       print*, '_____muhatMML_________sigmahatMML________'
       print 1000,  ' ',mean_muhatMML,' ','     ', mean_sigmaMML
       print 1000,  '  [',var_muhatMML,']  ', '  [',var_sigmaMML,']   '
       print*, '-------------------------------------------'
       print*, '_____mutildaLSE______sigmatildaLSE_______'
       print 1000,  ' ',mean_mutilda,' ','     ', mean_sigmatilda, ' '
       print 1000, '  [',var_mutilda,']  ', '  [',var_sigmatilda,']  '
       print*, '*******************************************'

c      End of l loop for b.
       enddo

       stop
       end

c-----------------------------------------------------------------------
c      Calculates the MML estimator of mu for GL (Generalized Logistic)
c      distribution. y is assumed to be a column vector.
c-----------------------------------------------------------------------
       subroutine GL_MML_iteration(y,n,mu_0,sigma_0,output)

       integer n

       real y(n), output(3), mu_0, sigma_0
       real sory(n), t(n), tt, te
       real bet(n), alf(n), del(n)
       real mm,KK,DD,BB,CC
       real SIGdot, MUdot, estbponeinv

       estbponeinv=0.0
       do 21 i=1,n
              t(i)=(y(i)-mu_0)/sigma_0
              tt=t(i)
              te=exp(t(i))
              estbponeinv=estbponeinv+(te/(1.0+te))
              alf(i)=(1.0+te+tt*te)/((1.0+te)**2.0)
              bet(i) = te/((1.0+te)**2.0)
21     continue

c      estbponeinv is the estimator of b/(b+1)
       estbponeinv=estbponeinv/(n*1.0)
c      To make estb be the estimator of 1/(b+1), we need:
       estbponeinv=(1.0-estbponeinv)

       do j=1,n
              del(j)=alf(j)-estbponeinv
       enddo

       mm=sum(bet)
       DD=sum(del)
       KK=dot_product(bet,y)

       KK=KK/mm
       DD=DD/mm

       BB=0.0
       CC=0.0
       do 25 j=1,n
              BB=BB+(1.0/estbponeinv)*del(j)*(y(j)-KK)
              CC=CC+(1.0/estbponeinv)*bet(j)*((y(j)-KK)**2.0)
25     continue

       SIGdot=-BB+sqrt(BB**2.0+4.0*n*CC)
       SIGdot=SIGdot/(2.0*sqrt(n*(n-1.0)))
       MUdot=KK-DD*SIGdot

       output(1)=MUdot
       output(2)=SIGdot
       output(3)=estbponeinv
```

```
        return
        end
c-----------------------------------------------------------------------
c       Calculates the derivative of psi(x) function with respect to x. It can
c       be estimated by a summation over i from zero to nfinity of the component
c        1/(i+x-1)^2
c       Note that psi is already defined in Fortran with a real input. e.g. psi(2.0)
c-----------------------------------------------------------------------
        real function psidash(x)

        real x
        integer n

        psidash=0.0
        do i=1,10000
               psidash=psidash+1/((i+x-1.0)**2.0)
        enddo

        return
        end
c-----------------------------------------------------------------------
c       Calculates mean of the array x of size n
c-----------------------------------------------------------------------
        real function mymean(x,n)

        real x(n), sumx
        integer n

        sumx=sum(x)
        mymean=sumx/(1.0*n)

        return
        end
c-----------------------------------------------------------------------
c       Calculates the variance of the array x of size n
c-----------------------------------------------------------------------
        real function variance(x,n,true)
c       t=1 for dividing n, and t=0 for dividing n-1.

        real x(n),mu, ss
        integer n, true
        real mymean

        mu=mymean(x,n)
        ss=0.0
        do i=1,n
           ss=ss+((x(i)-mu)**2.0)
        enddo
        variance=ss/(1.0*n-1.0)

        if (true==1) then
        variance=ss/(1.0*n)
        endif

        return
        end
c-----------------------------------------------------------------------
c       Sorts the data in array 'x' as ascending order and stores this
c       sorted data in 'sortedx'.
c-----------------------------------------------------------------------
        subroutine mysort(n,x,sortedx,true)
c       t=1 for descending order t=0 for ascending order
        real x(n),sortedx(n), ascen(n)
        integer n, true

      do i=1,n
         sortedx(i)=x(i)
        enddo

c       Ascending order
      do i=1,n
         do j=i+1,n
           if(sortedx(i)>=sortedx(j)) then
             dummy=sortedx(i)
               sortedx(i)=sortedx(j)
```

```
            sortedx(j)=dummy
         endif
        enddo
      enddo

      if (true==1) then
c     Descending order
     do i=1,n
       ascen(i)=sortedx(n-i+1)
      enddo
      sortedx=ascen
      endif

      return
      end
c----------------------------------------------------------------------
c     Calculates the median of the data in array 'x' and stores this
c     median in 'med'.
c----------------------------------------------------------------------
      subroutine mymedian(n,x,med)

      real x(n),sortedx(n),med
      integer n,half
     real mysort

      call mysort(n,x,sortedx,0)

      half = floor(n/2.0)
      med = sortedx(half+1)

      if (2*half == n ) then
      med = (sortedx(half)+med)/2.0
      endif

      return
      end
c----------------------------------------------------------------------
c     Generates n number ofrandom numbers from GL (Generalized Logistic
c     Distribution) with parameters mu, sigma and b.
c----------------------------------------------------------------------
      subroutine GL_rnd(mu,sigma,b,n,y)

      real mu,sigma,b,u(n),y(n)
      integer n

      call rnun(n,u)

      do i=1,n
             y(i)=-sigma*log((1.0/u(i))**(1.0/b)-1.0)+mu
      enddo

      return
      end
c----------------------------------------------------------------------
c     Generates the inverse cdf of a n number of a GL (Generalized Logistic
c     Distribution) where F(y) = alpha. The parameters of the distribution
c     are mu, sigma and b.
c----------------------------------------------------------------------
      real function GL_invcdf(mu,sigma,b,alpha)

      real mu,sigma,b,alpha,y

      y=mu-sigma*log(alpha**(-1.0/b)-1.0)
      GL_invcdf=y

      return
      end
c----------------------------------------------------------------------
```

APPENDIX E

FORTRAN PROGRAM CALCULATING THE MMLEs OF THE PARAMETERS OF MULTIVARIATE LINEAR REGRESSION FOR LONG-TAILED SYMMETRIC DISTRIBUTION

```
c**************************************************************************
c       Written by  Ayca DONMEZ, 2009, Ankara
c**************************************************************************
c--------------------------------------------------------------------------
c       Calculates the MMLEs and LSEs of multiple linear regression
c       Y=theta0+theta*U+e where e is assumed to come from the
c       distribution family of Long-Tailed Symmetric (LTS) with p>=2.
c       To evaluate the efficiency and robustness, different secenarios
c       are considered:
c     Scenario No.1: N(0,sigma^2)
c     Scenario No.2: LTS(mu, sigma, p) p=5.0
c     Scenario No.3: LTS(mu, sigma, p) p=3.5
c     Scenario No.4: LTS(mu, sigma, p) p=2.5
c     Scenario No.5: LTS(mu, sigma, p) p=2.0
c       Outlier Models:
c     Scenario No.6: (n-r) xi come from N(0,sigma^2) and
c                             r(we do not know which) come from N(0,4*sigma^2)
c     Scenario No.7: (n-r) xi come from N(0,sigma^2) and
c                             r(we do not know which) come from N(0,16*sigma^2)
c       Mixture Models:
c     Scenario No.8: 0.90*N(0,sigma^2)+0.10*N(0,4*sigma^2)
c     Scenario No.9: 0.90*N(0,sigma^2)+0.10*N(0,16*sigma^2)
c     Scenario No.10: Student's t distr. with 2 dof.
c     Scenario No.11: Cauchy distribution
c     Scenario No.12: Slash (Normal/Uniform) distribution
c--------------------------------------------------------------------------
        use numerical_libraries

c     Declaration of the variables and arrays
        parameter n = 20
        parameter nn = floor(100000/(n*1.0))
        parameter mu = 0.0, sigma = 1.0
        parameter q = 4
        parameter rtheta0 = 0.0
        parameter iteration = 2

        real rtheta(q,1), mUrtheta(n,1),Urtheta(n)
        real ybar, xj(n), xbar(q), s2(q), s(q)
        real su(n), msu(n,1)
        real y(n), my(n,1), x(n,q), u(n,q), er(n)
        real r(n-1), w(n)

        integer scenario
        real p, cc, rr, urn(n), nor(n), correction
        real S0, preS0(n),median_preS0
        real resultMML(q+2)
```

```
      real tu(q,n),tuy(q,1),uu(q,q), invuu(q,q)

      real theta0,sigma_0,thetaa, theta(q)
      real mtheta(q,1),tmtheta(1,q),mUtheta(n,1), Utheta(n)
      real theta0MML(nn), thetaMML(nn,q), sigmaMML(nn)
      real mean_theta0MML,mean_thetaMML(q),mean_sigmaMML
      real var_theta0MML,var_thetaMML(q),var_sigmaMML

      real mUthetaLSE(n,1),UthetaLSE(n),presigmaLSE(n)
      real theta0tildaLSE,thetatildaLSE(q),sigmatildaLSE
      real mthetatildaLSE(q,1)
      real theta0LSE(nn), thetaLSE(nn,q), sigmaLSE(nn)
      real mean_theta0LSE,mean_thetaLSE(q),mean_sigmaLSE
      real var_theta0LSE,var_thetaLSE(q),var_sigmaLSE

c     Declaration of the functions
      real LTS_rnd, LTS_invcdf, LTS_Regmulti_MML_iteration
      real mymean, mysort, mymedian, variance


1111  FORMAT (A5,I3,A6,I3,A15,I2)
      print 1111, '  n =',n,' q =',q,'  iteration =',iteration
      print*, '---------------------------------------------------'

      do jj=1,q

            rtheta(jj,1)=1.0

            call rnun(n,xj)
            x(:,jj)=xj
            xbar(jj)=sum(xj)/(n*1.0)
            s2(jj)=variance(xj,n,1)
            s(jj)=sqrt(s2(jj))

                  do j=1,n
                        u(j,jj)=(x(j,jj)-xbar(jj))/s(jj)
                  enddo
      enddo

      su=sum(u,2)
      do j=1,n
            msu(j,1)=su(j)
      enddo
      mUrtheta=matmul(u,rtheta)
      do j=1,n
            Urtheta(j)=mUrtheta(j,1)
      enddo

      do k=1,12

      scenario=k
      correction=0.0

      do 100 h=1,nn

            if (scenario.EQ.1) then
c           'Scenario No.1: N(0,sigma^2)'
              call rnnor(n,er)
              do j=1,n
                  er(j)=mu+sigma*er(j)
              enddo

            elseif (scenario.EQ.2) then
c           'Scenario No.2: LTS(mu, sigma, p) p=5.0'
                  p=5.0
                  call LTS_rnd(mu,sigma,p,n,er)
            elseif (scenario.EQ.3) then
c           'Scenario No.3: LTS(mu, sigma, p) p=3.5'
                  p=3.5
                  call LTS_rnd(mu,sigma,p,n,er)
            elseif (scenario.EQ.4)then
c           'Scenario No.4: LTS(mu, sigma, p) p=2.5'
                  p=2.5
                  call LTS_rnd(mu,sigma,p,n,er)
            elseif (scenario.EQ.5) then
```

```
c              'Scenario No.5: LTS(mu, sigma, p) p=2.0'
                      p=2.0
                      call LTS_rnd(mu,sigma,p,n,er)
               elseif (scenario.EQ.6) then
c              'Scenario No.6: (n-r) ei come from N(0,sigma^2) and
c                    r(we do not know which) come from N(0,4*sigma^2)'

c                 To generate outliers first r of n observations are rescaled
c                 by multiplying sample units with a constant c (c=2 or c=4)
c                 to have observations from N(0,c*sigma). Note that the selection
c                 of the first r units does not matter since the sample is
c                 originally generated randomly.

                 call rnnor(n,er)
                 do j=1,n
                      er(j)=mu+sigma*er(j)
                 enddo

                      cc=2.0
                      rr=int(0.5+0.1*n)
                        do j=1,rr
                             er(j)=cc*er(j)
                        enddo
                      correction=(rr*(cc**2.0)+(n-rr))/(n*(1.0))

               elseif (scenario.EQ.7) then
c              'Scenario No.7: (n-r) xi come from N(0,sigma^2) and
c                     r(we do not know which) come from N(0,16*sigma^2^)'

                 call rnnor(n,er)
                 do j=1,n
                      er(j)=mu+sigma*er(j)
                 enddo

                      cc=4.0
                      rr=int(0.5+0.1*n)
                        do j=1,rr
                             er(j)=cc*er(j)
                        enddo
               correction=(rr*(cc**2.0)+(n-rr))/(n*(1.0))

               elseif (scenario.EQ.8) then
c              'Scenario No.8: 0.90*N(0,sigma^2)+0.10*N(0,4*sigma^2)'

                      call rnnor(n,er)
                      call rnun(n,urn)
                      cc=2.0
                        do j=1,n
                             if (urn(j).GT.0.90) then
                                    er(j)=cc*er(j)
                             endif
                        enddo
                      correction=(0.90*n+0.10*n*(cc**2.0))/(n*(1.0))

               elseif (scenario.EQ.9) then
c              'Scenario No.9: 0.90*N(0,sigma^2)+0.10*N(0,16*sigma^2)'

                      call rnnor(n,er)
                      call rnun(n,urn)
                      cc=4.0
                        do j=1,n
                             if (urn(j).GT.0.90) then
                                    er(j)=cc*er(j)
                             endif
                        enddo
                      correction=(0.90*n+0.10*n*(cc**2.0))/(n*(1.0))

               elseif (scenario.EQ.10) then
c              'Scenario No.10: Student s t with 2 dof'
                             call rnstt(n,2.0,er)

               elseif (scenario.EQ.11) then
c              'Scenario No.11: Cauchy distribution'

                      call rnchy(n,er)
```

```
              elseif (scenario.EQ.12) then
c             'Scenario No.12: Slash (Normal/uniform) distribution'

c              To generate slash distribution, a random number 'nor'
c              from N(0,1) and a random number 'u' from U(0,1)
c              are generated. The desired random number will be the
c              result of nor divided by u. This operation will be
c              repeated n times.
                     call rnnor(n,nor)
                     call rnun(n,urn)
                            do j=1,n
                                   er(j)=nor(j)/urn(j)
                            enddo

              endif

c             The outlier and mixture models should be bias corrected
c             for sigma. 'correction' variable is defined as 0.0 at the
c             begining of the loop of scenarios (k) and it is only used
c             in outlier and mixture models (Scenario No.6 to No.9).
c             The corrections for sigmaMML are:
              if (correction.GT.0.0) then
                     do j=1,n
                            er(j)=er(j)/sqrt(correction)
                     enddo
              endif

              do j=1,n
                     y(j)=rtheta0+Urtheta(j)+er(j)
                     my(j,1)=y(j)
              enddo

              do j=1,(n-1)
                     r(j)=(y(j+1)-y(j))/(su(j+1)-su(j))
              enddo

c             Initial estimator of theta:
              call mymedian(n-1,r,thetaa)

              theta=thetaa

              do j=1,n
                     w(j)=y(j)-thetaa*su(j)
              enddo

              call mymedian(n,w,theta0)

              do j=1,n
                preS0(j)=abs(w(j)-theta0)
              enddo

              call mymedian(n,preS0,median_preS0)

c             S0=1.483*median{|y_j-theta1*x_j-theta0|}
              S0=1.483*median_preS0

              sigma_0=S0

              do ll=1,iteration

                     call LTS_Regmulti_MML_iteration(y,u,n,q,theta0,theta,
     &               sigma_0,resultMML)

                     theta0=resultMML(1)
                     theta=resultMML(2:(q+1))
                     sigma_0=resultMML(q+2)

              enddo

                     theta0MML(h)=resultMML(1)
                     thetaMML(h,:)=resultMML(2:(q+1))
                     sigmaMML(h)=resultMML(q+2)

              ybar=sum(y)
```

```
            ybar=ybar/(n*1.0)

            tu=transpose(u)
            uu=matmul(tu,u)
            tuy=matmul(tu,my)

            CALL LINRG (q,uu,q,invuu,q)
            mthetatildaLSE=matmul(invuu,tuy)
            mUthetaLSE=matmul(u,mthetatildaLSE)

            UthetaLSE=mUthetaLSE(:,1)
            thetatildaLSE=mthetatildaLSE(:,1)

            theta0tildaLSE=ybar

            sigmatildaLSE=0.0
            do j=1,n
                   sigmatildaLSE=sigmatildaLSE+
     &          (y(j)-ybar-UthetaLSE(j))**2.0
            enddo

            sigmatildaLSE=sqrt(sigmatildaLSE/(n-q-1.0))

                   theta0LSE(h)=theta0tildaLSE
                   thetaLSE(h,:)=thetatildaLSE
                   sigmaLSE(h)=sigmatildaLSE

c      End of simulation runs nn
100    continue

       mean_theta0MML=mymean(theta0MML,nn)
       do jj=1,q
              mean_thetaMML(jj)=mymean(thetaMML(:,jj),nn)
              var_thetaMML(jj)=(n/(sigma**2.0))*variance(thetaMML(:,jj),nn,1)
       enddo
       mean_sigmaMML=(1/sigma)*mymean(sigmaMML,nn)

       var_theta0MML=(n/(sigma**2.0))*variance(theta0MML,nn,1)
       var_sigmaMML=(n/(sigma**2.0))*variance(sigmaMML,nn,1)


       mean_theta0LSE=mymean(theta0LSE,nn)
       do jj=1,q
              mean_thetaLSE(jj)=mymean(thetaLSE(:,jj),nn)
              var_thetaLSE(jj)=(n/(sigma**2.0))*variance(thetaLSE(:,jj),nn,1)
       enddo
              mean_sigmaLSE=(1/sigma)*mymean(sigmaLSE,nn)

       var_theta0LSE=(n/(sigma**2.0))*variance(theta0LSE,nn,1)
       var_sigmaLSE=(n/(sigma**2.0))*variance(sigmaLSE,nn,1)

       print*, '_______________theta0_______________theta1_________
     &______sigma______'
      print*, 'Model_____MML_______LSE________MML________LSE____
     &___MML_______LSE__'

1100   FORMAT (I5,F10.3,F10.3,F11.3,F11.3,F10.3,F10.3)
1200   FORMAT (I5,F10.3,F11.3,F13.3,F13.3,F12.3)
2100   FORMAT (A9,F6.3,A1,A4,F5.3,A1,A4,F6.3,A1,A4,F6.3,A1,A4,F5.3,A1,
     &A4,F5.3,A1)
2200   FORMAT (A9,F6.3,A1,A4,E9.3,A1,A4,F6.3,A1,A4,E9.3,A1,A4,F6.3,A1,
     &A4,E9.3,A1)

       if (k.LT.10) then
       print 1100,k,mean_theta0MML,mean_theta0LSE,mean_thetaMML(1),
     &mean_thetaLSE(1),mean_sigmaMML,mean_sigmaLSE
       print 2100, '[',var_theta0MML,']','[',var_theta0LSE,']',
     &'[',var_thetaMML(1),']','[',var_thetaLSE(1),']','[',var_sigmaMML
     &,']','[',var_sigmaLSE,']'
       else
       print 1200,k,mean_theta0MML,mean_theta0LSE,mean_thetaMML(1),
     &mean_thetaLSE(1),mean_sigmaMML,mean_sigmaLSE
       print 2200, '[',var_theta0MML,']','[',var_theta0LSE,']',
     &'[',var_thetaMML(1),']','[',var_thetaLSE(1),']','[',var_sigmaMML
     &,']','[',var_sigmaLSE,']'
```

```
      endif
      print*, '-----------------------------------------'


1110  FORMAT (A5,F11.3,F11.3,F11.3)

      print*,'______________theta2_______________theta3_________
     &______theta4______'
      print*, 'Model_____MML_______LSE________MML________LSE____
     &___MML_______LSE__'

      if (k.LT.10) then
      print 1100,k,mean_thetaMML(2),mean_thetaLSE(2),mean_thetaMML(3),
     &mean_thetaLSE(3),mean_thetaMML(4),mean_thetaLSE(4)
      print 2100, '[',var_thetaMML(2),']','[',var_thetaLSE(2),']',
     &'[',var_thetaMML(3),']','[',var_thetaLSE(3),']',
     &'[',var_thetaMML(4),']','[',var_thetaLSE(4),']'
      else
      print 1200,k,mean_thetaMML(2),mean_thetaLSE(2),mean_thetaMML(3),
     &mean_thetaLSE(3),mean_thetaMML(4),mean_thetaLSE(4)
      print 2200, '[',var_thetaMML(2),']','[',var_thetaLSE(2),']',
     &'[',var_thetaMML(3),']','[',var_thetaLSE(3),']',
     &'[',var_thetaMML(4),']','[',var_thetaLSE(4),']'
      endif
      print*, '-----------------------------------------'

c     End of scenario loop k
      enddo

      stop
      end

c----------------------------------------------------------------------
c     Calculates the MML estimators of mu and sigma for LTS (Long-Tailed
c     Symetric) distribution with p>=2.
c----------------------------------------------------------------------
      subroutine LTS_Regmulti_MML_iteration(y,u,n,q,theta0,theta
     &,sigma_0,output)

      integer n, q
      real pp, kk,output(q+2)
      real theta0, theta(q), sigma_0
      real mtheta(q,1),mUtheta(n,1), Utheta(n)
      real y(n), u(n,q)
      real t(n),bet(n),alf(n),m,betxy(n),betuu(q)
      real ybar, ubar(q)
      real sB, sC, BB, CC, thetaMMLubar
      real theta0MML,thetaMML(q), sigmahatMML

      real my(n,1), mbet(n,n), malf(n,n), mone(n,1)
      real stu(n,q),tstu(q,n), tumbet(q,n), tumbetu(q,q)
      real tumbety(q,1), invtumbetu(q,q)
      real tumalf(q,n), tumalfmone(q,1)
      real msK(q,1), msD(q,1), sK(q), sD(q)
      real presigmaa(n,q), presigma(n)

c     In order to make MML estimation free of p, we put p=16.5, k=30,
      pp= 16.5
      kk = 30

      mbet=0.0
      malf=0.0

      do j=1,n
             my(j,1)=y(j)
             mone(j,1)=1.0
      enddo

      mtheta(:,1)=theta
      mUtheta=matmul(u,mtheta)

             do j=1,n
                    Utheta(j)=mUtheta(j,1)
             enddo
```

```
do j=1,n
    t(j) = (y(j)-theta0-Utheta(j))/sigma_0
        bet(j) = 1.0/((1+t(j)**2.0/kk)**2.0)
        alf(j) =((1.0/kk)*t(j))/((1.0+((t(j)**2.0)/kk))**2.0)
        betxy(j)=bet(j)*y(j)
enddo

m = sum(bet)
do jj=1,q
        betuu(jj)=dot_product(bet,u(:,jj))
        ubar(jj)=betuu(jj)/m
enddo

  ybar=sum(betxy)/m

  do j=1,n
         mbet(j,j)=bet(j)
         malf(j,j)=alf(j)
  enddo


do jj=1,q
do j=1,n
stu(j,jj)=u(j,jj)-ubar(jj)
enddo
enddo


tstu=transpose(stu)
tumbet=matmul(tstu,mbet)
tumbetu=matmul(tumbet,stu)
tumbety=matmul(tumbet,my)

CALL LINRG (q,tumbetu,q,invtumbetu,q)

msK=matmul(invtumbetu,tumbety)
tumalf=matmul(tstu,malf)
tumalfmone=matmul(tumalf,mone)
msD=matmul(invtumbetu,tumalfmone)

do jj=1,q
        sK(jj)=msK(jj,1)
        sD(jj)=msD(jj,1)
enddo

do jj=1,q
do j=1,n
        presigmaa(j,jj)=sK(jj)*stu(j,jj)
enddo
enddo

presigma=sum(presigmaa,2)

  sB=0.0
  sC=0.0

  do j=1,n
        sB=sB+alf(j)*(y(j)-ybar-presigma(j))
        sC=sC+bet(j)*(y(j)-ybar-presigma(j))**2.0
  enddo

BB=((2*pp)/kk)*sB
CC=((2*pp)/kk)*sC

sigmahatMML = (BB+sqrt((BB**2.0)+4*n*CC))/(2.0*sqrt(n*(n-q-1.0)))

do jj=1,q
        thetaMML(jj)=sK(jj)+sD(jj)*sigmahatMML
enddo

thetaMMLubar=dot_product(thetaMML,ubar)
theta0MML=ybar-thetaMMLubar

output(1)=theta0MML
do jj=1,q
```

```
        output(jj+1)=thetaMML(jj)
        enddo

        output(q+2)=sigmahatMML

        return
        end
c----------------------------------------------------------------------
c       Generates n number of random numbers from LTS (Long-Tailed Symmetric
c       Distribution) with parameters mu, sigma and p where it is assumed
c       that p>=2
c----------------------------------------------------------------------
        subroutine LTS_rnd(mu,sigma,p,n,y)

        real mu,sigma,p,kk,vv,x(n),y(n)
        integer n

        kk=2*p-3.0
        vv=2*p-1.0

        call rnstt(n,vv,x)

        do i=1,n
                y(i)=sigma*sqrt(kk/vv)*x(i)+mu
        enddo

        return
        end
c----------------------------------------------------------------------
c       Calculates mean of the array x of size n
c----------------------------------------------------------------------
        real function mymean(x,n)

        real x(n), sumx
        integer n

        sumx=sum(x)
        mymean=sumx/(1.0*n)

        return
        end
c----------------------------------------------------------------------
c       Calculates the variance of the array x of size n
c----------------------------------------------------------------------
        real function variance(x,n,true)
c       t=1 for dividing n, and t=0 for dividing n-1.
c      Declaration of the variables and arrays

        real x(n),mu, ss
        integer n, true

c      Declaration of functions
        real mymean

        mu=mymean(x,n)

        ss=0.0
        do i=1,n
                ss=ss+((x(i)-mu)**2.0)
        enddo

        variance=ss/(1.0*n-1.0)

        if (true==1) then
                variance=ss/(1.0*n)
        endif

        return
        end
c----------------------------------------------------------------------
c       Sorts the data in array 'x' as ascending order and stores this
c       sorted data in 'sortedx'.
c----------------------------------------------------------------------
        subroutine mysort(n,x,sortedx,true)
c       t=1 for descending order t=0 for ascending order
```

```
c     Declaration of the variables and arrays
       real x(n),sortedx(n), ascen(n)
       integer n, true

      do i=1,n
        sortedx(i)=x(i)
       enddo

c      Ascending order
      do i=1,n
        do j=i+1,n
          if(sortedx(i)>=sortedx(j)) then
            dummy=sortedx(i)
             sortedx(i)=sortedx(j)
             sortedx(j)=dummy
          endif
         enddo
       enddo

       if (true==1) then
c      Descending order
      do i=1,n
        ascen(i)=sortedx(n-i+1)
       enddo
       sortedx=ascen
       endif

       return
       end
c-----------------------------------------------------------------------
c      Calculates the median of the data in array 'x' and stores this
c      median in 'med'.
c-----------------------------------------------------------------------
      subroutine mymedian(n,x,med)
c     Declaration of the variables and arrays
       real x(n),sortedx(n),med,half
       integer n

c     Declaration of the functions
      real mysort

       call mysort(n,x,sortedx,0)

       half = floor(n/2.0);
       med = sortedx(half+1);

       if (2*half == n ) then
       med = (sortedx(half)+med)/2.0
       endif

       return
       end
c-----------------------------------------------------------------------
c      Generates the inverse cdf of LTS (Long-Tailed Symmetric Distribution)
c      where F(y) = alpha. The parameters of the distribution are
c      mu, sigma and p where p>=2.
c-----------------------------------------------------------------------
       real function LTS_invcdf(mu,sigma,p,alpha)

       real mu,sigma,p,alpha,y

       kk=2.0*p-3.0
       vv=2.0*p-1.0

       Tinv=tin(alpha,vv)
       y=((sigma*Tinv)/sqrt(vv/kk))+mu
       LTS_invcdf=y

       return
       end
c-----------------------------------------------------------------------
```

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Dönmez, Ayça
Nationality: Turkish (TC)
Date and Place of Birth: 6 November 1980 , Ankara
email: donmezayca@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
BS	METU Statistics	2003
MA	Bilkent University Economics	2005

WORK EXPERIENCE

Year	Place	Enrollment
2005- Present	METU Department of Statistics	Teaching Assistant
2003-2005	Bilkent University Department of Economics	Research Assistant
2002	Central Bank of the Republic of Turkey	Intern Student

LANGUAGES

Turkish (mother tongue), English (Excellent), German (Beginners)

CONFERENCE PAPERS & PRESENTATIONS

Donmez, A., Batmaz, I., Kartal, E., Kuran, B., *Comparison of Linear and Robust Regression Methods for Metamodelling Complex Systems*, 23rd European Conference on Operational Research 5-8 July 2009, Bonn-Germany.

Donmez, A., Akinc, D., Surucu, B., *Robust Linear Regression with Censored Data*, COMPSTAT 2008 – International Conference on Computational Statistics, 24-29 August 2008, Porto – Portugal.

Donmez, A. and Surucu, B., *Robust Classification Procedures*, COMPSTAT 2006 – International Conference on Computational Statistics, 28 August-1 September 2006, Rome-Italy.

SCHOLARSHIPS & HONORS

- Bilkent University Scholarship for Master of Arts in the Department of Economics.
- Ranked third among 40 students and graduated with High Honor degree from the Department of Statistics at METU.
- Listed in 2000, 2001 (Fall & Spring semesters) and 2002 Spring semester President's High Honors Roll, Department of Statistics, METU.