ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

ALİ İHSAN GENÇ

ROBUST STATISTICAL ANALYSIS BASED ON THE GENERALIZED T DISTRIBUTION FAMILIES

MATEMATİK ANABİLİM DALI

ADANA, 2003

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ROBUST STATISTICAL ANALYSIS BASED ON THE GENERALIZED T DISTRIBUTION FAMILIES

ALİ İHSAN GENÇ

DOKTORA TEZİ
MATEMATİK ANABİLİM DALI

Bu Tez 03./01/2003 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oy Birliği/ ~~Oy Çokluğu~~ İle Kabul Edilmiştir.

İmza
Doç. Dr. Olcay ARSLAN
Danışman

İmza
Prof. Dr. Fikri AKDENİZ
Üye

İmza
Prof. Dr. Refik BURGUT
Üye

İmza
Prof. Dr. Fikri ÖZTÜRK
Üye

İmza
Doç. Dr. Sadullah SAKALLIOĞLU
Üye

Bu tez Enstitümüz Matematik Anabilim Dalında hazırlanmıştır.

Kod No: 708

T.C. FEN BİLİMLERİ ENSTİTÜSÜ ÇUKUROVA ÜNİVERSİTESİ 1973

Prof. Dr. Fikri AKDENİZ
Enstitü Müdürü
İmza ve Mühür

CONTENTS

ABSTRACT

Ph.D. THESIS

ROBUST STATISTICAL ANALYSIS BASED ON THE GENERALIZED T DISTRIBUTION FAMILIES

ALİ İHSAN GENÇ

DEPARTMENT OF MATHEMATICS
INSTITUTE OF NATURAL AND APPLIED SCIENCES
ÇUKUROVA UNIVERSITY

Supervisor: Assoc. Prof. Dr. Olcay ARSLAN
Year: 2003, Pages: 119
Jury: Assoc. Prof. Dr. Olcay ARSLAN
Prof. Dr. Fikri AKDENİZ
Prof. Dr. Refik BURGUT
Prof. Dr. Fikri ÖZTÜRK
Assoc. Prof. Dr. Sadullah SAKALLIOĞLU

In this thesis, we consider three different forms of the generalized t distributions defined as the alternatives to the normal and the Student's t distributions in modeling data, which may have longer than normal tails, skewness, or multimodality. We investigate the existence, the uniqueness, and the robustness properties of the maximum likelihood estimators for the parameters of these distributions. We show that the maximum likelihood estimators for the parameters of the generalized t distributions can be alternative robust estimators for the location and scale estimates of a dataset. We also show that the likelihood estimating equations cannot be explicitly solved to obtain the estimates; a numerical computation method should be used to compute the estimates. To overcome the computation problem of the estimates we propose a simple iterative reweighting algorithm, and show that this algorithm is an EM algorithm. As the natural extension of the location and scale problem, the *GT* distribution is used as an alternative model for the error term in regression.

Key Words: influence function, breakdown point, maximum likelihood, regression, EM algorithm.

ÖZ

DOKTORA TEZİ

GENELLEŞTİRİLMİŞ T DAĞILIMLARI AİLESİNE DAYALI DAYANIKLI İSTATİSTİKSEL ANALİZ

ALİ İHSAN GENÇ

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MATEMATİK ANABİLİM DALI

Danışman: Doç. Dr. Olcay ARSLAN
Yıl: 2003, Sayfa: 119
Jüri: Doç. Dr. Olcay ARSLAN
Prof. Dr. Fikri AKDENİZ
Prof. Dr. Refik BURGUT
Prof. Dr. Fikri ÖZTÜRK
Doç. Dr. Sadullah SAKALLIOĞLU

Bu çalışmada gözlemleri modellemede normal ve t dağılımına birer alternatif olarak tanımlanmış normalden daha kalın, çarpık ve hatta çok modlu olabilen üç farklı genelleştirilmiş t dağılımının her birinin konum ve ölçek parametrelerinin maksimum olabilirlik tahmin edicilerinin parametre uzayında varlığı, tekliği ve dayanıklılık (robustness) özellikleri incelenmiştir. Bir genelleştirilmiş t dağılımı olan *GT* dağılımı için bulunan tahmin edicilerin kapalı formda olması nedeniyle kestirimlerin hesaplanması için sayısal yöntem olarak tekrar ağırlıklandırmalı en küçük kareler algoritması önerilmiş ve bu yöntemin EM algoritmasına denk olduğu gösterilmiştir. Son olarak regresyonda hata terimini modellemede *GT* dağılımının kullanılabilirliği incelenmiştir.

Anahtar Kelimeler: etki fonksiyonu, bozulma noktası, maksimum olabilirlik, regresyon, EM algoritması.

ACKNOWLEDGEMENTS

I thank my supervisor Doç. Dr. Olcay Arslan for her help, encouragement, suggestions, remarks, patience and time over the period of this thesis.

I would also like to thank Prof. Dr. Fikri Akdeniz and Prof. Dr. Refik Burgut for their interests and suggestions over the period of this research.

I gratefully acknowledge the financial support provided by Niğde University during my study at the Department of Mathematics in Çukurova University.

Finally, I thank all the faculty, postgraduate students and staff in the Department of Mathematics in Cukurova University for creating a pleasant working environment.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

BT	Box-Tiao Distribution
EM	Expectation and Maximization
GT	Generalized T Distribution
GET	Generalized Student T Distribution
GST	Generalized Student Family
IF	Influence Function
IR	Iteratively Reweighted
Log-lik	The value of the log likelihood function
LMS	Least Median of Squares
LS	Least Squares
LS*	Least Squares based on outlier-free data
LS**	Least Squares based on outlier group only
MVE	Minimum Volume Ellipsoid
PIV	Pearson Type IV Distribution
R	The set of real numbers
Stdres	Standardized Residual

1. INTRODUCTION

The classical approach in deriving statistical theories (estimation and inference) usually assumes the existence of some exact parametric model, and derives statistical procedures based on it. The most widely used model is the normal model probably due to the belief that measurement errors often follow the normal distribution. However, in real life we often observe the violation of the normality assumption. The data may come from a longer-tailed, skewed, or multimodal distribution and the procedures based on the normality assumption will be inoperative.

An important point for the normal distribution is its mathematical tractability. This property of the normal distribution facilitated the development of the procedures based on normality assumption. For example, consider the location estimation problem of a univariate dataset. The sample mean is no doubt the most widely known estimator to solve this problem. The sample mean can be easily obtained under the assumption that the dataset is normally distributed with unknown location parameter. By the Gauss-Markov Theorem, the mean is the best estimator in the case of independent and identically distributed normal observations. However, since all observations enter with the same weight into the computation of the mean, its value is adversely affected by outliers, i.e. points whose position is unusually far from the majority of the data. It is also well-known that (see e.g. Lye and Martin, 1993a) outliers may produce heavier tailed distributions than normal, and the mean remains a poor estimate of location for heavy-tailed distributions. Therefore, since the statistical inference based on the normal distribution is very sensitive to outliers and/or departure from the normality assumptions the possible misleading results reveals the need for resistant procedures in case of the violation of the assumptions.

As an alternative to the statistical procedure based on the normal distribution the robust procedures, which are insensitive to the outliers and/or departures from model assumptions, have been introduced by Huber (1964). They enable to obtain valid conclusions for the data even in the presence of outliers and/or departure from the normality assumptions.

Since the mathematical derivation and justification of a statistical procedure requires the beauty of certain distribution assumption, a statistician has been looking for some alternative robust distributions to the normal distribution to model data which may have different tails structure, such as longer or sorter tails than the normal distribution, or may contain some unusual points or clusters. A number of symmetric long tailed distributions were introduced as alternatives to the normal distributions. The Student's t distribution is one of the first distributions that is used as an alternative to the normal. It provides a useful extension of the normal distribution for statistical modeling of data sets which have longer tails than normal (see e.g. Lange et al. 1989 and Arslan, 1992). However, it has been argued by Hoek et al. (1995) that the use of the t distribution provides some protection against outliers but if there are many outliers the other robust methods should be used. Some other robust distributions used in the literature are the double exponential distribution (see e.g. Bloomfield and Steiger, 1983), the Box-Tiao (*BT*) (see e.g. Klein and Spady, 1984), power exponential distribution (see e.g. Zeckhauser and Thompson, 1970) and some generalized version of the known distributions. A distribution is usually generalized by mixing different distributions. The resulting generalized distribution is reach in having shape parameters and has the property that it nests the ordinary distribution as well as some other important non-normal distributions for special or limiting cases of the shape parameters.

One of the distributions that is generalized is the Student's t distribution. We come across three variants of the generalized t distributions in the literature. The first one was introduced by McDonald and Newey (1988). They called it as the generalized t (*GT*) distribution and used it to develop a robust partially adaptive estimation method in regression problem. The second one is considered by Tiku and Suresh (1992) and Vaughan (1992) to estimate the location and scale of a univariate data set. Johnson, Kotz and Balakrishnan (1995) also mention this distribution as the generalized Student family, for short we will use the notation *GST* for this family. The multivariate version of this distribution was intruduced by Arellano-Valle and Bolfarine (1995). The third version of the generalized t distributions is in fact a subfamily of the generalized exponential distributions. Lye and Martin (1993a)

pointed out that outliers can produce not only fat-tailed distributions but also skewed and even multimodal distributions, and they suggest using this subfamily which is called the *GET* distributions family to stand these effects of outliers (see also Cobb et al., 1983). Theodossiou (1998) developed a skewed extension of the *GT* distribution and called it the skewed *GT* distribution. He used the skewed *GT* distribution as the model for some financial data. Recently, Arslan (2002c) defined a multivariate version of the generalized t distributions family, which includes the multivariate t distribution, the multivariate version of *GST*, and the multivariate version of *GT*.

Although the generalized t (*GT*, *GET* and *GST)* distributions have been widely used as an alternative robust modeling distribution in econometrics, they have not been handled much in statistical literature. The main bulk of this thesis will be devoted to the generalized t distributions in general. We will try to make a unifying approach to the generalized t distributions and investigate the properties of these distributions in details. We will use these families as robust alternative distributions, and investigate the robustness properties of the estimators obtained from the maximum likelihood estimation of the parameters of the generalized t distributions.

The thesis is organized as follows. In Chapter 2 we introduce and explore the properties of the generalized t distributions. In particular, we show that the *GT* distribution introduced by McDonald and Newey (1988) can be defined as the distribution of the ratio of two different distributions. Suppose we have two independent random variables distributed as a *BT,* and an inverse generalized gamma (*IGG*) distributions, respectively, then we show that the distribution of the ratio of these two random variables is the generalized t (or *GT*) distribution.

In Chapter 3, we model our data with each of the generalized t distributions and find the maximum likelihood estimators of location and scale parameters. We show that, when the shape parameters of the *GT* distribution are kept fixed and finite, the maximum likelihood estimators for the location and scale parameters of the *GT* distribution can be considered as the M estimators with bounded influence functions so that the effect of outlying observations on the estimates will be reduced. Since the shape parameters will control the robustness of the estimators, we will treat the shape parameters as the robustness tuning constants. Another example for a shape

parameter that has been treated as a robustness tuning constant is the degrees of freedom parameter of a Student's t distribution, (see, e.g., Yuh and Hogg, 1988, Lange et al., 1989 and Lucas, 1997). Similar considerations also hold for others. Next, we turn our attention to the existence and the uniqueness of the maximum likelihood estimators for the location and the scale parameters of the generalized t distributions with known shape parameters. We give some conditions under which the maximum likelihood estimates of the location and scale parameters exist and are unique in the parameter space. We also show that for some values of the shape parameters the likelihood function cannot have any maximum points in the parameter space, and for some values of the shape parameters it may have many local maximum points in the parameter space.

In Chapter 4 we investigate the robustness properties of the M estimators based on the generalized t distributions in detail. We mainly deal with the local and global robustness of the estimators. That is; we obtain the breakdown points and the influence functions for each case. We show that when the shape parameters are estimated along with the location and scale parameters, the estimators have unbounded influence functions, and hence poor local robustness properties. Therefore, to obtain estimators with good local robustness properties we have to hold the shape parameters fixed and treat them as the robustness tuning constants. We also show that we can have high breakdown point estimators if the density function of the *GT* distribution properly tuned; that is, if the shape parameters are properly chosen. Further, we show that the M estimators based on the generalized t distributions may have multiple solutions, in other words, the likelihood functions of the generalized t distributions may have multiple local maximum. This behavior of the *GT* likelihood function is very useful in recent robustness studies. This behavior will help us to identify the multiple structure in the data if the dataset has some multiple structures such as clusters.

In Chapter 5 we tackle the computation problem of the estimates based on the *GT* distributions. We show that we can obtain the maximum likelihood estimators but because of the adaptive future of the estimating equations we cannot explicitly solve the estimating equations. Therefore, we have to use some numerical algorithms

to compute the estimates. Many different algorithms can be proposed to compute the estimators, but we observe that the estimating equations can be easily reformulated to define an iteratively reweighting algorithm (IR) to calculate the estimates. We show that this iteratively reweighting (IR) algorithm can be identified as an EM algorithm, which is widely used in computational statistics. In this chapter comparative examples to determine the performance of the estimators on some problematic data sets are also presented.

For the complementary purposes we consider the regression setting, which is the natural extension to the location scale problem, in Chapter 6. We assume that the error term in the linear regression model is *GT*-distributed and find the estimators of the regression parameters. Some comparative examples are given.

Finally, the concluding remarks, open problems and the future studies and projects are given in Chapter 7.

2. GENERALIZED T DISTRIBUTIONS

This chapter introduces the three different generalized t distributions as well as their skewed versions. It is our goal in this chapter to collect and highlight some properties of each of the three different generalized t distributions. These distributions are commonly used in statistical economics to model financial data and error term in regression.

2.1. The Generalized T (GT) Distribution

McDonald and Newey (1988) introduced the GT distribution as an alternative to the normal and Student's t distribution for modeling errors in the regression. They used the GT distribution to develop robust partially adaptive estimation procedure. This procedure includes least squares, LAD (Least Absolute Deviation), L_p and several other estimators as special cases.

Statistical distributions of returns on financial instruments such as stocks, bonds or options play a central role in much of the financial literature. The GT family applies to symmetric, fat-tailed distributions and therefore adjusts for the leptokurtosis of the non-normal security returns. McDonald and Nelson (1989) and Butler, McDonald, Nelson and White (1990) applied the GT-based partially adaptive estimation procedure for the robust estimation of the market model. McDonald and Nelson (1993) compared the GT distribution with their new estimating procedure of the market model. Partially adaptive estimates of ARMA time series models based on the GT distribution were given in McDonald (1989), and applications of the GT to U.S. stock index returns were presented in Bollerslev, Engle and Nelson (1994, pp. 3017-3027).

In this section, we will show that the GT distribution can be redefined as a scale mixture of a Box-Tiao (BT) (Box and Tiao, 1962) and the generalized gamma (GG) distributions (McDonald and Butler, 1987). A Bayesian approach was also given by (Butler et al.,1990). They considered the BT distribution and assumed that its scale parameter is a random variable that has an inverse generalized gamma distributed (IGG). Here we show that the GT distribution can be obtained as the

ratio of two independent random variables. Theorem 2.1 states the main result of this chapter.

The density function of the *GT* distribution is

$$f(x;\mu,\sigma,p,q)=\frac{p}{2\sigma q^{1/p}B(1/p,q)(1+\frac{|x-\mu|^{p}}{q\sigma^{p}})^{q+1/p}}\quad , \qquad (2.1.1)$$

(McDonald and Newey , 1988), where $-\infty<x<\infty$, $\sigma>0$ is a scale parameter, $-\infty<\mu<\infty$ is a location parameter and $p>0$ and $q>0$ are shape parameters that control the shape of the density. Larger values of p and q are associated with thinner tails of the density. Similarly, smaller values of p and q correspond to thicker tails. Figure 2.1 displays some of the density functions of the *GT* distributions. In this figure we can see the tail behavior of the density from thicker to thinner with different choices of the values of the shape parameters. Note also that we have cusps at the center of the symmetry of the distribution for p=1 and p=.6.

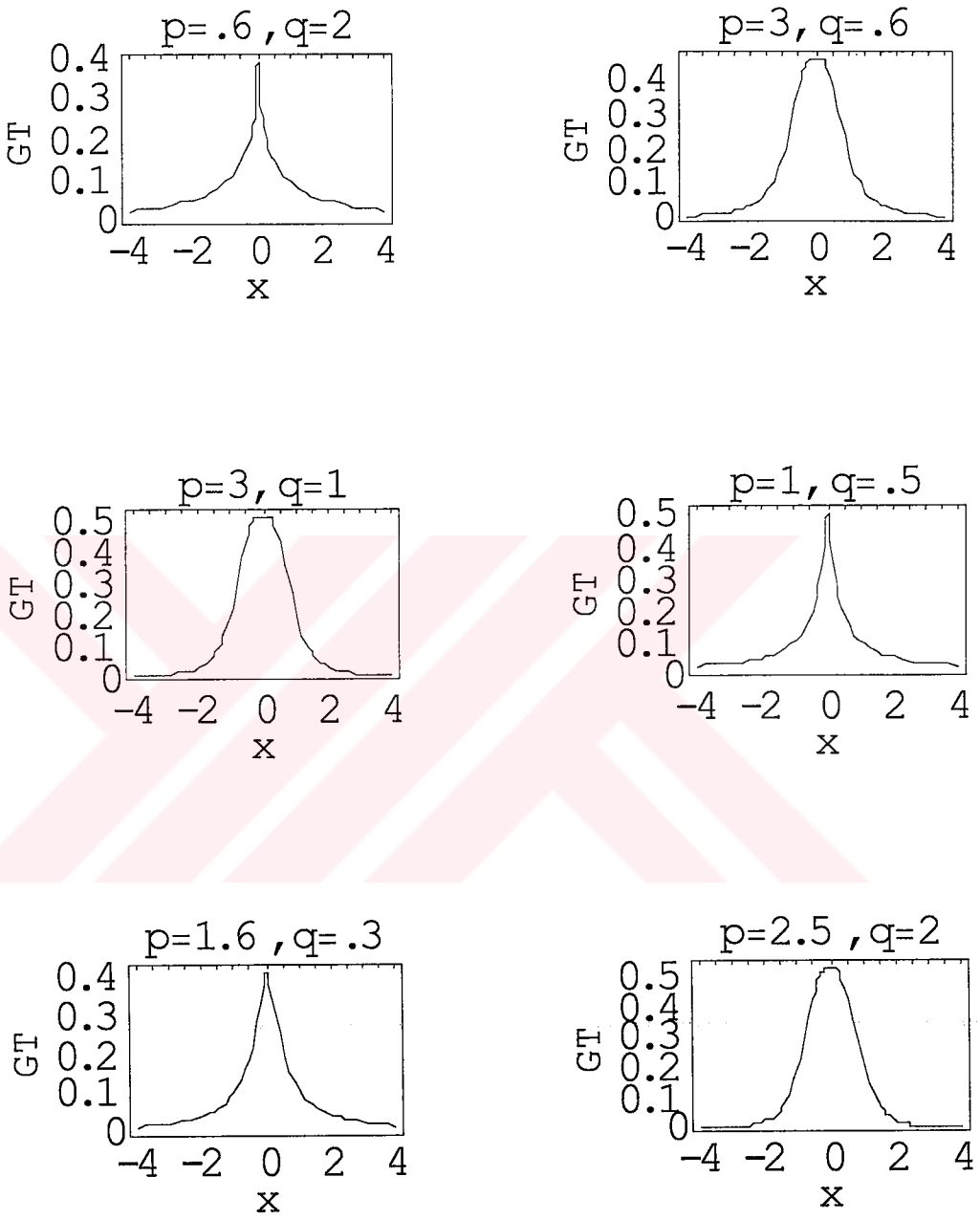


Figure 2.1. Density functions of the *GT* distributions for different values of the shape parameters.

Theorem 2.1. Let the random variable U have a $BT(u; p)$ distribution with the shape parameter $p > 0$ and let the random variable T have a $GG(t; p/2,1, q)$ distribution with the shape parameters $p/2$, 1 and $q > 0$. Assume that U and T are independent. Then the random variable

$$X = \mu + q^{1/p}\sigma U T^{-1/2} \tag{2.1.2}$$

has a $GT(x;\mu,\sigma,p,q)$ distribution with the density function given (2.1.1) where $\sigma > 0$ and $-\infty < \mu < \infty$.

Proof. The density function of U is

$$g(u;p) = p/(2\Gamma(1/p))e^{-|u|^p}, \tag{2.1.3}$$

where $-\infty < u < \infty$, and the density function of T is

$$h(t;p/2,1,q) = p/(2\Gamma(q))t^{pq/2-1}e^{-t^{p/2}}, \tag{2.1.4}$$

where $t \geq 0$. Since U and T are independent the joint density function will be

$$f(u,t) = p^2/(4\Gamma(1/p)\Gamma(q))t^{pq/2-1}e^{-|u|^p - t^{p/2}}. \tag{2.1.5}$$

From equation (2.1.2) we have $u = (x-\mu)/(q^{1/p}t^{-1/2}\sigma)$, and using this the right hand side of (2.1.5) reduces to

$$p^2/(4\Gamma(1/p)\Gamma(q)q^{1/p}\sigma)t^{(pq-1)/2}e^{-t^{p/2}((|x-\mu|^p/(q\sigma^p))+1)}. \tag{2.1.6}$$

Integrating (2.1.6) with respect to scale random variable t gives the marginal density of X as follows.

$$f(x;\mu,\sigma,p,q) = p/(2\sigma q^{1/p}B(1/p,q))(1+|x-\mu|^p/(q\sigma^p))^{-q-1/p}, \tag{2.1.7}$$

which is the probability density function of the *GT* distribution. □

2.1.1. Properties of the *GT* Distribution

a) The *GT* family includes several important distributions as special or limiting cases (McDonald and Newey, 1988). For example,

i) When $p=2$ we get the usual t distribution with $\frac{\sigma}{\sqrt{2}}=\alpha$ scale parameter and $r=2q$ degrees of freedom.

ii) Since as $q\to\infty$,

$$\frac{q^{-1/p}}{B(1/p,q)}=\frac{\Gamma(1/p+q)}{\Gamma(q)q^{1/p}}\frac{1}{\Gamma(1/p)}\to\frac{1}{\Gamma(1/p)} \tag{2.1.8}$$

and

$$(1+\frac{(\frac{|x|}{\sigma})^p}{q})^{-q-1/p}\to e^{-(\frac{|x|}{\sigma})^p}. \tag{2.1.9}$$

Thus the *GT* becomes the *BT*.

iii) When $p=1$ and as q approaches 0 we get the Laplace (double exponential) distribution with density function

$$\frac{1}{2\sigma}e^{-\frac{|x|}{\sigma}}. \tag{2.1.10}$$

iv) When $p=2$, $\frac{\sigma}{\sqrt{2}}=\alpha$,and as q approaches 0, we get the normal distribution with mean 0 and variance α^2.

v) In (i), for $q=1/2$ we get the Cauchy distribution.

vi) Since $\frac{p}{\Gamma(1/p)} \to 1$ as $p \to \infty$, we get the uniform distribution with probability density function $1/(2\sigma)$ when $|x| < \sigma$ and 0 otherwise.

b) When $p \leq 1$ we have cuspidate distributions (i.e. distributions with densities having a corner on which they are not differentiable).

c) Since the GT has a symmetric probability density function, its odd ordered moments are equal to zero. When n is even, the n th moment of this distribution is

$$E(X^n) = \frac{\sigma^n q^{n/p} \Gamma(\frac{n+1}{p}) \Gamma(q - \frac{n}{p})}{\Gamma(1/p)\Gamma(q)}, (n < pq) \tag{2.1.11}$$

(McDonald and Newey,1988). This follows immediately from the formula

$$\int_0^\infty x^r (1+qx^k)^{-m} \, dx = \frac{1}{k} q^{-\frac{r+1}{k}} \Gamma(\frac{r+1}{k}) \Gamma(m - \frac{r+1}{k}) \frac{1}{\Gamma(m)} \tag{2.1.12}$$

where $0 < \frac{r+1}{k} < m$, $m \neq 0$.

d) The variance of the GT distribution is

$$Var(X) = \frac{\sigma^2 q^{2/p} \Gamma(3/p) \Gamma(q - \frac{2}{p})}{\Gamma(1/p)\Gamma(q)}, (pq > 2). \tag{2.1.13}$$

2.2. The Generalized Student Family (GST)

In this section we will consider the Pearson Type VII distributions family that is also known as the generalized student family (GST). This family of distributions has been widely studied. Some of the references are Tiku and Suresh (1992), Johnson, Kotz, and Balakrishnan (1995) and Vaughan (1992).

The density function of the GST distribution is

$$f(x;\mu,\sigma,p) = \frac{1}{\sigma k^{1/2} B(\frac{1}{2}, p - \frac{1}{2})} [1 + \frac{(x-\mu)^2}{k\sigma^2}]^{-p}, \tag{2.2.1}$$

where $p > 1/2$, $k > 0$, $\sigma > 0$. Tiku and Suresh (1992) especially chose the values of p and k so that

$$k = \begin{cases} 2p-3 & if \ \ p \geq 2, \\ 1 & if \ \ 1 \leq p < 2 \end{cases}. \tag{2.2.2}$$

Arellano-Valle and Bolfarine (1995) discuss three different characterizations of a multivariate generalized t distribution within the class of the elliptical distributions. In the univariate case they used the *GST* with $p = (v+1)/2$.

It is true that the *GST* is a scale mixture of a normal distribution. Andrews and Mallows (1974) gave the proof of it for the usual Pearson Type VII distributions.

Theorem 2.2. (Andrews and Mallows, 1974) Let the random variable U have a standard normal $N(0, 1)$ distribution and let the random variable Q have a gamma distribution $GA(p\text{-}1/2, 1/2)$ with the shape parameters p-1/2, 1/2. Assume that U and Q are independent. Then the random variable

$$X = \mu + k^{1/2}\sigma U Q^{-1/2} \tag{2.2.3}$$

has a $GST(x; \mu, \sigma, p, k)$ distribution.

Proof. The joint density of U and Q is

$$h(u,q) = \frac{1}{\sqrt{\pi}\Gamma(p-1/2)2^p} e^{-\frac{1}{2}(q+u^2)} q^{p-3/2}. \tag{2.2.4}$$

With the transformation $u = \dfrac{x-\mu}{k^{1/2}\sigma q^{-1/2}}$ from (2.2.3), the density in (2.2.4) becomes

$$f(x,q) = \frac{1}{2^p\Gamma(p-1/2)\sqrt{\pi k}\sigma} e^{-\frac{q}{2k}(\frac{x-\mu}{\sigma})^2 - \frac{q}{2}} q^{p-1}, \tag{2.2.5}$$

and integrating (2.2.5) with respect to q we get the *GST* density:

$$\int_0^\infty f(x,q)dq = \frac{1}{\sigma k^{1/2} B(\frac{1}{2}, p-\frac{1}{2})}[1+\frac{(x-\mu)^2}{k\sigma^2}]^{-p}. \qquad (2.2.6)$$

□

The inverse gamma (*IGA*) distribution has the density

$$g(x;a,b) = \frac{b^a}{\Gamma(a)} e^{-b/x}(1/x)^{a+1}. \qquad (2.2.7)$$

where $x>0$, $a>0$. Note that when V is inverse gamma, $IGA(p-1/2, k/2)$, distributed and Z is normal, $N(0,\sigma^2)$, distributed and they are independent then the random variable $X = \mu + V^{1/2}Z$ is $GST(\mu,\sigma,p,k)$ distributed.

An interesting derivation of the inverse gamma distribution is resulted from the reciprocal of another –chi-square- distribution; that is, the random variable k/χ^2_{2p-1} is $IGA(z; p-\frac{1}{2}, \frac{k}{2})$ distributed.

2.2.1 Properties of the *GST* Distribution

a) When $p=1$ we get the Cauchy distribution.

b) When $p=(k+1)/2$ we get the usual t distribution with k degrees of freedom.

c) With the choice of (2.2.2), as p approaches ∞ we have

$$\lim_{p\to\infty} \frac{1}{k^{1/2} B(\frac{1}{2}, p-\frac{1}{2})} = \lim_{y\to\infty} \frac{\Gamma(y+\frac{1}{2})}{2^{1/2}(y-1)^{1/2}\Gamma(y)} = \frac{1}{\sqrt{2}}, \qquad (2.2.8)$$

where $y = \frac{k+2}{2}$, by using the formula $\lim_{n\to\infty} \frac{\Gamma(x+n)}{\Gamma(n)n^x} = 1$ for $0 \le x < \infty$. And also

$$\lim_{p\to\infty}[1+\frac{(x-\mu)^2}{k\sigma^2}]^{-p} = e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}. \qquad (2.2.9)$$

Thus we get the normal distribution with mean μ and variance σ^2 as p approaches ∞.

d) In (2.2.3) the expected value of U is zero by symmetry and thus the expected value of the GST distribution is μ.

The density of the random variable $R=1/Q$ is

$$f(r;p)=\frac{1}{2^{p-1/2}\Gamma(p-1/2)}e^{-\frac{1}{2r}}(\frac{1}{r})^{p+1/2}, \tag{2.2.10}$$

and its expected value is

$$E(R)=\frac{1}{2^{p-1/2}\Gamma(p-1/2)}\int_0^\infty re^{-\frac{1}{2r}}(\frac{1}{r})^{p+1/2}dr$$

$$=\frac{1}{2p-3}. \tag{2.2.11}$$

Thus the variance of X is

$$Var(X)=E(X^2)-[E(X)]^2$$

$$=\mu^2+k\sigma^2E(U^2)E(R)-\mu^3$$

$$=\frac{k\sigma^2}{2p-3} \tag{2.2.12}$$

With the choice (2.2.2), the expected value and the variance of the GST distribution is μ and σ^2, respectively.

d) The n th moment about μ is 0 if n is odd and is

$$E[X^n]=\frac{k^{n/2}\Gamma(\frac{n+1}{2})\Gamma(p-\frac{n+1}{2})}{\sqrt{\pi}\Gamma(p-1/2)} \tag{2.2.13}$$

if n is even and if $n+1<2p$. With the choice of (2.2.2) we have

$$E[X^n]=\frac{k^{n/2}\Gamma(\frac{n+1}{2})\Gamma(p-\frac{n+1}{2})}{\sqrt{\pi}\Gamma(p)} \qquad (2.2.14)$$

e) With the choice of (2.2.2), if X is the GST distributed then $(v/k)^{1/2}Z$ is Student t distributed with $v=2p-1$ degrees of freedom where $Z=\frac{X-\mu}{\sigma}$.

In Figure 2.2 some GST densities are plotted. From this figure we see that for small values of p and k the density has thinner tails.

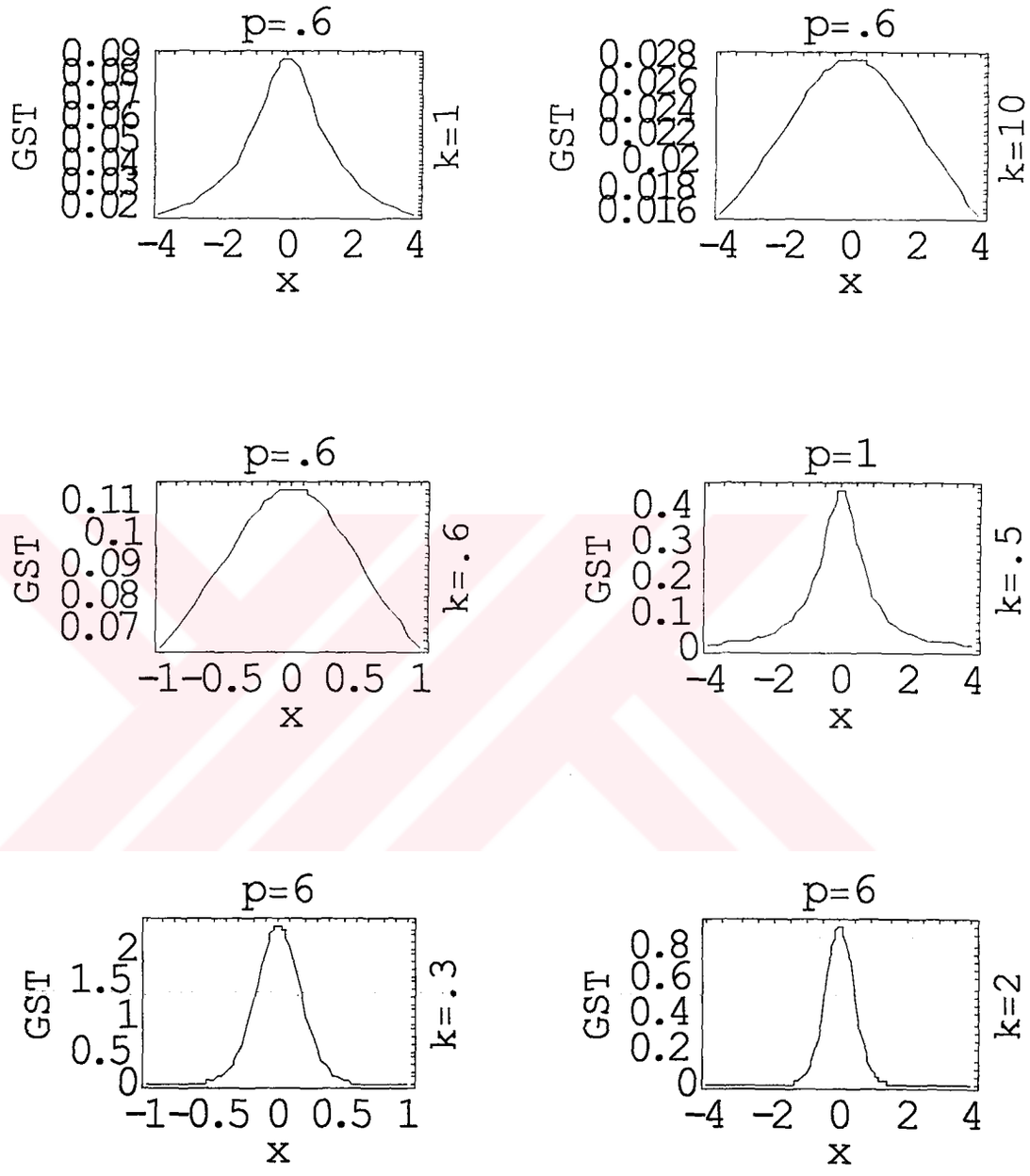


Figure 2.2. Density functions of the *GST* distributions for different values of the shape parameters.

2.3. The Generalized Student t (*GET*) Distribution

The *GET* represents a generalization of the standard unimodal, symmetric Student t distribution. The *GET* distributions family is a subclass of the generalized exponential distributions family.

The generalized exponential family provides an extension of the exponential family to the case where distributions can exhibit multimodality (Lye and Martin, 1993a).

The generalized exponential family can be derived by extending the Pearson differential equation as follows:

$$\frac{df}{du} = -\frac{g(u)f(u)}{h(u)} \tag{2.3.1}$$

where $g(u)$ and $h(u)$ are polynomials in the random variable in U and $f(u)$ is the density function of U. In the standard Pearson system $\deg[g(u)] \leq 1$ and $\deg[h(u)] \leq 2$.

The general solution of (2.3.1) is

$$f(u) = e^{-\int^{u} \frac{g(s)}{h(s)} ds - \eta}, (u \in D), \tag{2.3.2}$$

where the normalizing constant is given by

$$\eta = \log \int e^{-\int^{u} \frac{g(s)}{h(s)} ds} du \tag{2.3.3}$$

and the domain D of $f(u)$ in (2.3.2) is the open interval where $h(u)$ is positive.

The differential equation given by (2.3.1) can be extended in two ways. First, $\deg[g(u)]$ and $\deg[h(u)]$ can be greater than 1 or 2. Second, instead of using polynomials, $g(u)$ and $h(u)$ can be allowed to include more general functions such as logarithms, trigonometric expressions, and so forth.

The *GET* distributions family are obtained with the following choices of the functions g and h :

$$g(u) = \sum_{i=0}^{M-1} \alpha_i u^i , \tag{2.3.4a}$$

$$h(u) = \gamma^2 + u^2 , \tag{2.3.4b}$$

where γ^2 is referred to as the "degrees of freedom" parameter. Substituting (2.3.3) into (2.3.2) yields the *GET* . For the case $M = 6$, the *GET* is

$$f(u;\gamma,\theta_i) = e^{\theta_1 \tan^{-1}(u/\gamma)+\theta_2 \log(\gamma^2+u^2)+\sum_{i=3}^{6}\theta_i u^{i-2}-\eta} , -\infty < u < \infty \tag{2.3.5}$$

where the normalizing constant is

$$\eta = \log \int_{-\infty}^{\infty} e^{\theta_1 \tan^{-1}(u/\gamma)+\theta_2 \log(\gamma^2+u^2)+\sum_{i=3}^{6}\theta_i u^{i-2}} du , \tag{2.3.6}$$

and $\theta_1 = -\frac{\alpha_0}{\gamma} + \gamma\alpha_2 - \gamma^3\alpha_4$, $\theta_2 = \frac{-\alpha_1 + \gamma^2\alpha_3 - \gamma^4\alpha_5}{2}$, $\theta_3 = -\alpha_2 + \gamma^2\alpha_4$,

$\theta_4 = \frac{-\alpha_3 + \gamma^2\alpha_5}{2}$, $\theta_5 = -\alpha_4/3$, $\theta_6 = -\alpha_5/4$.

Fong (1997) used data on monthly returns of twenty-two stocks listed on the Singapore Stock Exchange and found that the *GET* provides a significantly better fit to the data than the normal distribution or the symmetric Student's t distribution. Based on a small out-of-sample experiment the *GET* was also found to outperform ordinary least squares and Student's t betas in forecasting ability.

A form of the *GET* distribution which represents both normal and the Student's t distributions is given as

$$f(x;\mu,\sigma,\gamma) = C\frac{1}{\sigma}[\gamma^2 + (\frac{x-\mu}{\sigma})^2]^{-\frac{1+\gamma^2}{2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} , \tag{2.3.7}$$

where $-\infty < x < \infty$, location parameter $-\infty < \mu < \infty$, scale parameter $\sigma > 0$, the shape parameter $\gamma > 0$, and the normalizing constant C. The density function in (2.3.7) is symmetric.

We note that the normalizing constants involving the *GET* distributions family are difficult to obtain analytically. This drawback of the family makes it difficult to do further analysis about the family (see Chapter 3).

2.3.1. Properties of the *GET* Distribution

a) The *GET* distribution includes several important distributions as special cases of its parameter values (Lye and Martin,1993a). For example,

i) The unimodal Student t distribution with γ^2 degrees of freedom is obtained when $\theta_2 = -\frac{1+\gamma^2}{2}$, $\theta_1 = \theta_3 = \theta_4 = \theta_5 = \theta_6 = 0$. This is true because if we put the parameter values in (2.3.4) above, then

$$f(u;\gamma) = \frac{(\gamma^2+u^2)^{\frac{1+\gamma^2}{2}}}{\int_{-\infty}^{\infty}(\gamma^2+u^2)^{\frac{1+\gamma^2}{2}}du}$$

$$= \frac{1}{(\gamma^2)^{1/2}(1+\frac{u^2}{\gamma^2})^{\frac{\gamma^2+1}{2}}} \frac{1}{(\gamma^2)^{\gamma^2/2}\int_{-\infty}^{\infty}(\gamma^2+u^2)^{\frac{1+\gamma^2}{2}}du}$$

$$= \frac{1}{(\gamma^2)^{1/2}(1+\frac{u^2}{\gamma^2})^{\frac{\gamma^2+1}{2}}} \frac{1}{2\int_0^{\infty}(1-\frac{u^2}{\gamma^2+u^2})^{\frac{\gamma^2}{2}-1}(1-\frac{u^2}{\gamma+u^2})(\gamma^2+u^2)^{-1/2}du}, \qquad (2.3.8)$$

change the variable u into the variable v by letting $v = \frac{u^2}{u^2+\gamma^2}$, then we have

$$= \frac{1}{(\gamma^2)^{1/2}(1+\frac{u^2}{\gamma^2})^{\frac{\gamma^2+1}{2}}} \int_0^1 (1-v)^{\frac{\gamma^2}{2}-1} v^{-1/2} dv$$

$$= \frac{1}{(\gamma^2)^{1/2}(1+\frac{u^2}{\gamma^2})^{\frac{\gamma^2+1}{2}}} \frac{\Gamma(\frac{\gamma^2+1}{2})}{\Gamma(1/2)\Gamma(\gamma^2/2)}. \qquad (2.3.9)$$

ii) When $\gamma = 1$ the Cauchy distribution occurs.

iii) The *GET* also contains the unimodal normal distribution as a special case when $\theta_1 = \theta_2 = \theta_5 = \theta_6 = 0$. For these parameter values we have

$$f(u;\theta_3,\theta_4) = e^{\theta_3 u + \theta_4 u^2 - \eta}$$

$$= \frac{e^{-\alpha_2 u - \frac{\alpha_3}{2}u^2}}{\int_{-\infty}^{\infty} e^{-\alpha_2 u - \frac{\alpha_3}{2}u^2} du}$$

$$= e^{-\alpha_2 u - \frac{\alpha_3}{2}u^2} \frac{1}{\int_{-\infty}^{\infty} e^{\frac{\alpha_2^2}{2\alpha_3}} e^{-\frac{\alpha_3}{2}(u+\frac{\alpha_2}{\alpha_3})^2} du}, \qquad (2.3.10)$$

changing the variable u into y by $y = u + \frac{\alpha_2}{\alpha_3}$, we get

$$= e^{-\alpha_2 u - \frac{\alpha_3}{2}u^2} \frac{1}{\int_{-\infty}^{\infty} e^{\alpha_2^2/2\alpha_3} e^{-\frac{\alpha_3}{3}y^2} dy}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1/\alpha_3}} e^{-\frac{\alpha_3}{2}(u+\frac{\alpha_2}{\alpha_3})^2}. \qquad (2.3.11)$$

iv) The generalized normal distribution (GEN) is obtained when $\theta_1 = \theta_2 = 0$.

v) The Pearson Type IV (PIV) distribution is obtained when $\theta_3 = \theta_4 = \theta_5 = \theta_6 = 0$ (See, Section 2.3.2).

2.2.2. Moment Recursion Formula For the *GET* Distribution

Lye and Martin (1993b) gave the moment recursion formula for the *GET* distribution by following the steps below:

Multiply both sides of $h(u)df = -g(u)f(u)du$ by u^k and then integrate:

$$\int u^k h(u)df = -\int g(u)f(u)u^k du . \tag{2.3.12}$$

The right hand side of (2.3.12) can be written as

$$-\int g(u)f(u)u^k du = -\sum_{i=0}^{M-1} \alpha_i \mu'_{i+k} , \tag{2.3.13}$$

where μ'_j represents the j th moment around zero. Using integration by parts, the left hand side of (2.3.12) can be written as

$$\int u^k h(u)df = u^k h(u)f(u) \Big|_a^b - \int \frac{d}{du}[u^k h(u)]f(u)du$$

$$= -\int \frac{d}{du}[u^k h(u)]f(u)du . \tag{2.3.14}$$

Thus the general expression for the recurrence formula is

$$\int \frac{d}{du}[u^k h(u)]f(u)du = \sum_{i=0}^{M-1} \alpha_i \mu'_{i+k} . \tag{2.3.15}$$

For the *GET* distribution, the moment recursion formula is

$$\int \frac{d}{du}[u^k (\gamma^2 + u^2)]f(u)du = \sum_{i=0}^{M-1} \alpha_i \mu'_{i+k} , \tag{2.3.16}$$

or

$$\sum_{i=0}^{M-1} \alpha_i \mu'_{i+k} = k\gamma^2 \mu'_{k-1} + (k+2)\mu'_{k+1} \quad , (k = 0,1,2,....,M-1). \qquad (2.3.17)$$

In Figure 2.3 some *GET* densities are plotted. We see from this figure that the *GET* densities can have different shapes for various choices of functions, from symmetric unimodal to skewed and multimodal.

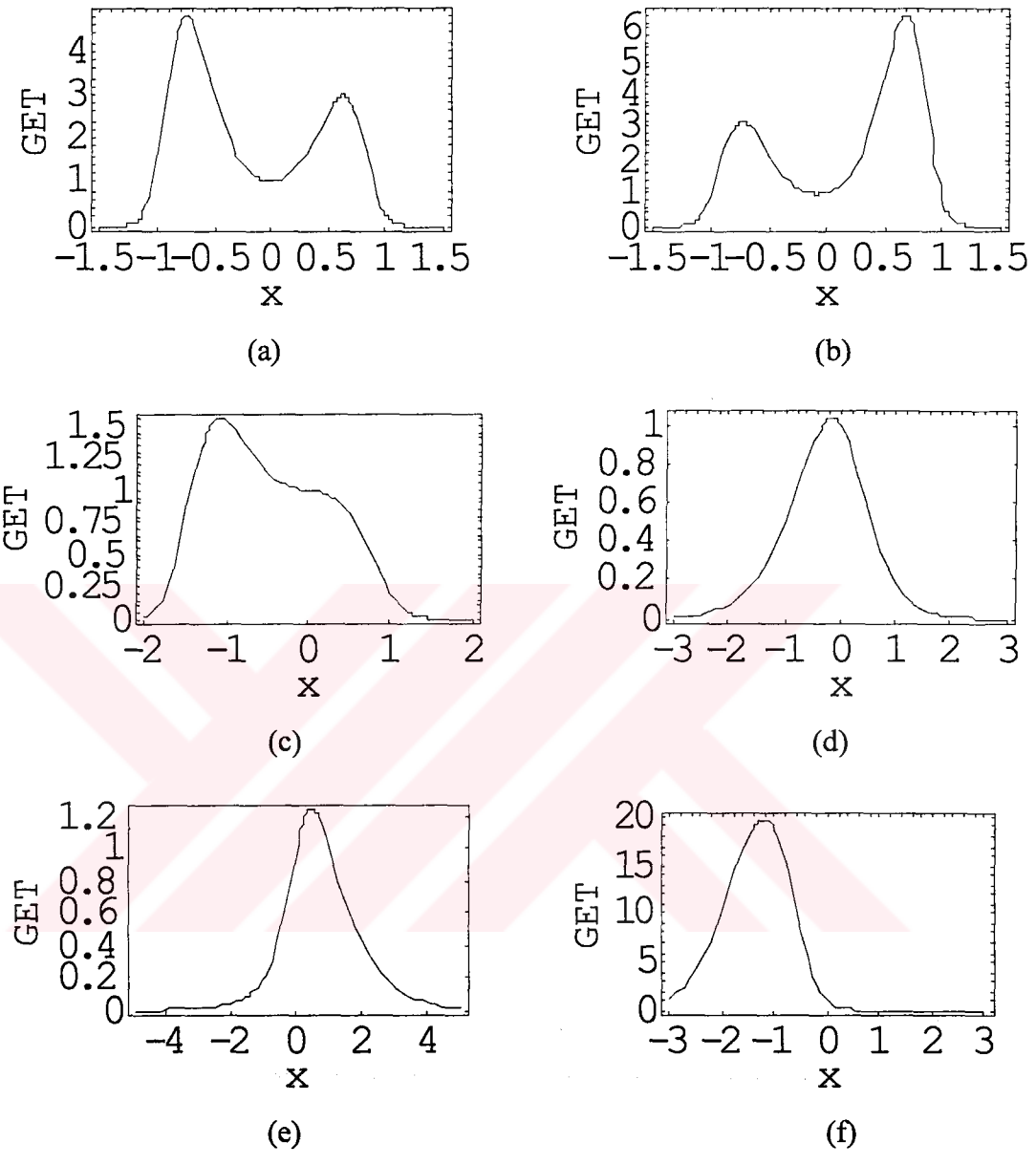


Figure 2.3. Density functions of the *GET* distributions for different values of the shape parameters.

(a) $(1/(1+x^2)^{1/2})\exp(-\tan^{-1}x + x + 6x^2 - x^3 - 6x^4 - \eta)$;

(b) $\exp(x + 6x^2 - x^3 - 6x^4 - \eta)$;

(c) $(1/(1+x^2))\exp(-.1x + .998x^2 - .9x^3 - .874x^4 - \eta)$;

(d) $(1/(1+x^2))\exp(-.5x - .5x^2 - \eta)$;

(e) $(1/(1+x^2))\exp(-.1x^2 + \tan^{-1}x - \eta)$;

(f) $\exp(-.5x^2 - 3.3\tan^{-1}(1.67x) + .08\log(.36 + x^2) - \eta)$.

2.4. Skewed Generalized T Distributions

In the previous sections we have seen that the generalized t distributions have symmetric densities except for the *GET* distributions family which can be more flexible, from symmetric to skewed and from unimodal to multimodal. However, the data in economics are usually skewed and some skewed distribution is needed to model them. In the literature we see a new skewed distribution, the skewed *GT* distribution, which is alternatively defined to others.

2.4.1. Skewed *GT* Distribution

Theodossiou (1998) developed a skewed version of the *GT* distribution and used it to model some financial data. The density of the skewed *GT* is

$$f(x;k,n,\lambda,\sigma^2)=\begin{cases} f_1 = c[1+(k/(n-2))\theta^{-k}(1-\lambda)^{-k}|x/\sigma|^k]^{-\frac{n+1}{k}} & x<0 \\ f_2 = c[1+(k/(n-2))\theta^{-k}(1+\lambda)^{-k}|x/\sigma|^k]^{-\frac{n+1}{k}} & x\geq 0 \end{cases} \quad (2.4.1)$$

where $k>0$, $n>2$, $-1<\lambda<1$, and c and θ are normalizing constants ensuring that $f(.)$ is a proper density and k,n,λ are scaling parameters and σ^2 is the variance of X. The parameters k and n control the height and tails of the density. The skewness parameter λ controls the rate of descent of the density around $x=0$. For example, in the case of positive skewness, $\lambda>0$, the random variable X in f_1 is weighted by a value greater than unity and in f_2 is weighted by a value less than a unity. The parameter n has the degrees of freedom interpretation in case $\lambda=0$ and $k=2$.

Theorem 2.3. (Theodossiou, 1998) For (2.4.1) we have

i) $$c=.5kB^{-3/2}(1/k,n/k)B^{1/2}(3/k,(n-2)/k)S(\lambda)\sigma^{-1}, \quad (2.4.2)$$

$$\theta=(k/(n-2))^{1/k}B^{1/2}(1/k,n/k)B^{-1/2}(3/k,(n-2)/k)S^{-1}(\lambda), \quad (2.4.3)$$

$$S(\lambda)=[1+3\lambda^2-4\lambda^2B^2(2/k,(n-1)/k)B^{-1}(1/k,n/k)B^{-1}(3/k,(n-2)/k)]^{1/2}, \tag{2.4.4}$$

where $B(.)$ is the beta function.

ii) The r th noncentered moments of X is

$$E(X^r)=.5[(-1)^r(1-\lambda)^{r+1}+(1+\lambda)^{r+1}]S^{-r}(\lambda)B((r+1)/k,(n-r)/k)$$
$$B^{-1+r/2}(1/k,n/k)B^{-r/2}(3/k,(n-2)/k)\sigma^r$$
$$B^{-1+r/2}(1/k,n/k)B^{-r/2}(3/k,(n-2)/k)\sigma^r \tag{2.4.5}$$

and hence the expected value of X is

$$E(X)=\mu=2\lambda S^{-1}(\lambda)B(2/k,(n-1)/k)B^{-1/2}(1/k,n/k)B^{-1/2}(3/k,(n-2)/k)\sigma \tag{2.4.6}$$

iii) The expected value of $|X|$ is

$$E(|X|)=(1+\lambda^2)S^{-1}(\lambda)B(2/k,(n-1)/k)B^{-1/2}(1/k,n/k)B^{-1/2}(3/k,(n-2)/k)\sigma \tag{2.4.7}$$

iv) The probabilities

$$P(X<0)=\int_{-\infty}^{0}f_1dx=(1-\lambda)/2, \tag{2.4.8}$$

$$P(X>0)=\int_{0}^{\infty}f_2dx=(1+\lambda)/2. \tag{2.4.9}$$

Now one can easily find the skewness and kurtosis measures from the following formulas

$$\text{Skewness}=\frac{E[(X-\mu)^3]}{\sigma^3}=\frac{E(X^3)-3\mu\sigma^2-\mu^3}{\sigma^3}=\frac{m_3}{\sigma^3} \tag{2.4.10}$$

$$\text{Kurtosis}=\frac{E[(X-\mu)^4]}{\sigma^4}=\frac{E(X^4)-4\mu m_3-6\mu^2\sigma^2-\mu^4}{\sigma^4} \tag{2.4.11}$$

Note that for $\lambda=0$, the skewed GT will be the symmetric GT distribution.

2.4.2. Skewed *GET* Distributions

Lye and Martin (1993a) used this form to model nonnormalities in the Martin-Marietta data set, which is given in Butler et al. (1990). They assumed that the disturbances are distributed as GET and they also compared the results with previous methods. They used

$$f(u;\sigma,\gamma,\theta_1)=e^{\theta_1\tan^{-1}(\frac{u/\sigma}{\gamma})-\frac{\gamma^2+1}{2}\log(\gamma^2+\frac{u^2}{\sigma^2})-\eta} \tag{2.4.12}$$

where $-\infty<u<\infty$, $\sigma>0$, $\gamma>0$.

The density allows for leptokurtosis through the Student t distribution and for skewness through the arctan term.

The location-scale form of (2.4.12) is

$$f(u;\gamma,\theta_1,\mu,\sigma)=Ce^{\theta_1\tan^{-1}(\frac{u-\mu}{\sigma\gamma})}\frac{1}{[\gamma^2\sigma^2+(u-\mu)^2]^{\frac{\gamma^2+1}{2}}}, \tag{2.4.13}$$

where C is normalizing constant, μ is the location parameter and σ is the scale parameter (See Figure 2.2 for the graph of the density function).

Theorem 2.4. If the random variable U has a density given in (2.4.13), then

i) the normalizing constant is

$$C = \frac{\Gamma(\frac{\gamma^2 + i\theta_1 + 1}{2})\Gamma(\frac{\gamma^2 - i\theta_1 + 1}{2})(\sigma\gamma)^{\gamma^2}}{\Gamma(\frac{\gamma^2+1}{2})\Gamma(\gamma^{2/2})\sqrt{\pi}}, \tag{2.4.14}$$

where i indicates an imaginary unit.

ii) the expected value is

$$E(U) = \mu + \frac{\theta_1 \gamma \sigma}{\gamma^2 - 1}, (\gamma^2 > 1). \tag{2.4.15}$$

iii) the variance is

$$Var(U) = \frac{(\sigma\gamma)^2}{\gamma^2 - 2}[1 + (\frac{\theta_1}{\gamma^2 - 1})^2], (\gamma^2 > 2). \tag{2.4.16}$$

iv) skewness and kurtosis are given respectively by

$$\frac{E\{[U - E(U)]^3\}}{[Var(U)]^{3/2}} = \frac{(\gamma^2 - 2)^{1/2}\theta_1 4}{(\gamma^2 - 1)(\gamma^2 - 3)[1 + (\frac{\theta_1}{\gamma^2 - 1})^2]^{1/2}}, (\gamma^2 > 3) \tag{2.4.17}$$

and

$$\frac{E\{[U - E(U)]^4\}}{[Var(U)]^2} = \frac{3(\gamma^2 - 2)}{\gamma^2 - 4} \frac{1 + \frac{\gamma^2 + 5}{\gamma^2 - 3}(\frac{\theta_1}{\gamma^2 - 1})^2}{[1 + (\frac{\theta_1}{\gamma^2 - 1})^2]}, (\gamma^2 > 4). \tag{2.4.18}$$

Proof: The proof can be found in Nagahara (1999) for the general Pearson IV distributions. □

Lim et al. (1998) formulated an empirical model of the distribution of exchange rate returns based on a combination of the *GET* and conditional variance

specifications. They applied this model to four daily bilateral exchange rates over the period 1984 to 1991. They used the form

$$f(u;\gamma,\theta_1,\theta_4)=(\gamma^2+u^2)^{\frac{1+\gamma^2}{2}}e^{\theta_4 u^2}e^{\theta_1\tan^{-1}(u/\gamma)-\eta}, \tag{2.4.19}$$

where $\theta_4 < 0$. The Student's t distribution is represented by the power term, the first term and the normal distribution is represented by the exponential term, the second term. It is the power term that gives the distribution its fat tails as it results in the distribution decaying relatively more slowly for larger absolute values of u than the exponential decay of the normal distribution. When $\theta_1 = 0$, the distribution is symmetric, with the degrees of freedom parameter γ^2, controlling the degree of fatness in the tails of the distribution. For values of $\theta_1 \neq 0$, the distribution exhibits skewness.

3. MAXIMUM LIKELIHOOD ESTIMATION FOR THE PARAMETERS OF THE GENERALIZED T DISTRIBUTION FAMILIES

In this chapter we will use the generalized t distributions as the alternative distributions to the normal and the standard t distribution to find alternative robust estimates for the location and scale parameters of a one-dimensional data set. We will model our sample by a generalized t distribution and try to find the maximum likelihood estimates for the parameters.

We will also investigate the existence and the uniqueness of the local maximum of the likelihood functions obtained from three different cases; i.e., location only, scale only and location-scale only cases. This will lead us to the conditions which ensure the existence and uniqueness of the solutions to a set of redescending M-estimating equations obtained from the generalized t distributions.

As we have seen in Chapter 2, normalizing constants concerning the *GET* distributions are difficult to find and even if we accomplish it they can be in a very complex form (See, Theorem 2.4). Therefore we will mainly focus on the *GT* and the *GST* distributions from this chapter on. However, we considered an important sub distribution of the *GET* family that includes both normal and the Student's t distribution through their terms. A form including these terms and an additional term is also used by Fong (1997) (See Section 2.2). Also, we will not consider the maximum likelihood estimation for the skewed generalized t distributions since they require a special treatment and thus it is not scope of the thesis (See Chapter 7).

We will close this chapter with some examples which show the behaviors of the generalized t based likelihood functions on some data sets.

3.1. The *GT* Distribution Case

Consider fitting a $GT(x;\mu,\sigma,p,q)$ distribution to the data $x_1, x_2, ..., x_n$ in R with the log-likelihood function

$$l(\mu,\sigma,p,q) = n[\log p + \log\Gamma(q+1/p) - \log\sigma - \log 2 - (1/p)\log q$$

$$-\log\Gamma(1/p)-\log\Gamma(q)]-(\frac{pq+1}{p})\sum_{i=1}^{n}\log(1+\frac{|x_i-\mu|^p}{q\sigma^p}). \quad (3.1.1)$$

We will refer to a maximum likelihood estimate as any point in the parameter space at which the likelihood function has a maximum. Taking the derivatives of (3.1.1) with respect to μ, σ, p and q, and setting to zero yield the following likelihood equations

$$l_\mu(\mu,\sigma,p,q)=(\frac{pq+1}{q\sigma^p})\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-1}\,sign(x_i-\mu)}{1+\frac{|x_i-\mu|^p}{q\sigma^p}}=0 \text{ if } p>1, \quad (3.1.2)$$

$$l_\sigma(\mu,\sigma,p,q)=-\frac{n}{\sigma}+\frac{pq+1}{q\sigma^{p+1}}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{1+\frac{|x_i-\mu|^p}{q\sigma^p}}=0, \quad (3.1.3)$$

$$l_p(\mu,\sigma,p,q)=n[\frac{1}{p}-\frac{1}{p^2}\gamma(q+1/p)+\frac{1}{p^2}\gamma(1/p)+\frac{1}{p^2}\log q]$$
$$+\frac{1}{p^2}\sum_{i=1}^{n}\log(1+\frac{|x_i-\mu|^p}{q\sigma^p})-\frac{(pq+1)}{pq\sigma^p}\sum_{i=1}^{n}\frac{|x_i-\mu|^p\log(|x_i-\mu|/\sigma)}{1+\frac{|x_i-\mu|^p}{q\sigma^p}}=0, \quad (3.1.4)$$

$$l_q(\mu,\sigma,p,q)=n[\gamma(q+1/p)-\gamma(q)-1/(pq)]-\sum_{i=1}^{n}\log(1+\frac{|x_i-\mu|^p}{q\sigma^p})$$
$$+\frac{(pq+1)}{pq^2\sigma^p}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{1+\frac{|x_i-\mu|^p}{q\sigma^p}}=0, \quad (3.1.5)$$

where $\gamma(.)$ is the digamma function $\gamma(.)=\Gamma'(.)/\Gamma(.)$.

The maximum likelihood estimates simultaneously solve the four estimating equations given in (3.1.2)-(3.1.5). Since the estimating equations cannot be obtained in an explicit form of the parameters, an iterative method such as Newton-Raphson is needed to find the solutions to the likelihood equations.

The maximum likelihood equations given above can be considered as the M-estimating equations. It is known that bounded influence function M-estimators are considered as the estimators that have good local robustness properties (see, e.g. Hampel et al., 1986). For our estimation problem, let $\theta = (\mu, \sigma, p, q)$ be the parameter vector, and $\hat{\theta}$ be the maximum likelihood estimator of θ. The influence function for $\hat{\theta}$ is $IF(x;\hat{\theta},F) = -(E(\psi'(x)))^{-1}\psi(x)$, where $\psi(x) = \frac{\partial}{\partial\theta}(-\log f(x))$ is the score function, and the expectations are taken over the underlying distribution of the data (e.g., Hampel et al., 1986, p. 86). Thus, one can easily see that the influence function of $\hat{\theta}$ is proportional to the score function $\psi(x) = \frac{\partial}{\partial\theta}(-\log f(x))$. Since the score functions for μ and σ are both bounded functions of the residuals for fixed, finite values of p and q, it is clear that the influence functions for the location and scale estimators are bounded. However, it can be easily seen from the likelihood equations (3.1.2) and (3.1.5) that the score functions for the shape parameters p and q are unbounded functions of the residuals. Thus, the influence functions for the estimators of the shape parameters will be unbounded, and hence will distort the robustness of the location and scale estimators when they are simultaneously estimated with the location and the scale parameters. Therefore, for the sake of robustness, we will assume that the shape parameters are fixed and finite, and we will only focus on the estimation problem for the location and scale parameters of the data. A similar discussion about the Student's t distribution is given in the paper by Lucas (1997). It is given in that paper that the M-estimators based on the Student's t distribution for the location and the scale parameters of a data set have bounded influence function if the degrees of freedom parameter is kept fixed. We will treat the shape parameters as the user specified robustness tuning constants.

If we assume that the shape parameters p and q are known, then the estimating equations (3.1.2) and (3.1.3) can be rewritten as

$$\hat{\mu} = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i , \qquad (3.1.6)$$

$$\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2 , \tag{3.1.7}$$

where

$$w_i = \frac{(pq+1)(|x_i - \hat{\mu}|/\hat{\sigma})^{p-2}}{q + (|x_i - \hat{\mu}|/\hat{\sigma})^p}. \tag{3.1.8}$$

Here the weight function $w(s) = (pq+1)s^{p-2}/(q+s^p)$ with $s = |x-\hat{\mu}|/\hat{\sigma}$ is a nonnegative, decreasing function of s so that the outlying observations will receive very small weights and the effect of outlying observations on the estimators will be reduced. Further, since $\lim_{s\to\infty} w(s)s = 0$ the equation given in (3.1.6) gives a redescending M-estimator for μ. Also, since $\lim_{s\to\infty} w(s)s^2 = pq+1$ the influence function for the scale parameter σ is bounded. Some properties of this function are as follows:

i) $\lim_{s\to\infty} w(s) = 0$.

ii) $$\lim_{s\to 0} w(s) = \begin{cases} \infty & if\ p<2 \\ 0 & if\ p>2 \\ (2q+1)/q & if\ p=2 \end{cases}.$$

iii) $w(s)$ has a local maximum at $s^* = [(p-2)q/2]^{1/p}$ and this maximum value is $\frac{(p-2)(pq+1)}{p}[\frac{2}{(p-2)q}]^{2/p}$. Therefore, the observations in the neighborhood $\hat{\mu} \pm \hat{\sigma}[\frac{(p-2)q}{2}]^{1/p}$ takes the greatest weights.

Note that the equations (3.1.6) and (3.1.7) can be viewed as an adaptively weighted sample mean and sample covariance where the weights are adaptively dependent on the estimates $\hat{\mu}$ and $\hat{\sigma}$. A simple iteratively reweighting algorithm, which can also be identified as an EM (expectation-maximization) algorithm, can be used to find the solutions of these estimating equations (See Chapter 5).

3.1.1. Existence and Uniqueness of the Maximum Likelihood Estimators With Known Shape Parameters

In this section we will focus on the existence and the uniqueness of the solutions to the estimating equations for the location and the scale parameters, assuming that the shape parameters are known. We will consider the existence and the uniqueness of a maximum of the *GT*-likelihood function for location-only case, scale-only case and location-scale case separately because the existence and the uniqueness of a maximum of the likelihood function will be different in these three cases.

In McDonald and Newey's paper (1988) asymptotic properties of the regression estimators based on the *GT* distributed errors were given after fixing the shape and scale parameters at some finite values with some constraints. We clarify that if those constraints are not imposed on the shape parameters p and q the likelihood function may have multiple local maxima in the parameter space.

The following two lemmas are useful when studying the existence and uniqueness of the maximum of the likelihood and log-likelihood functions. These lemmas have been given by Mäkeläinan et al. (1981).

Lemma 3.1. Let $L(\phi)$ be a twice continuously differentiable function with ϕ varying in a connected open subset $\Omega \subset R^p$.

i) If $\lim_{\phi \to \partial\Omega} L(\phi) = 0$ and ,

ii) if the Hessian matrix of second partial derivatives is negative definite at every point $\phi \in \Omega$ for which the gradient vector $\nabla L = \{L_\phi\}$ vanishes

then there is a unique maximum likelihood estimate $\hat{\phi} \in \Omega$, and the likelihood function attains no other maxima, no minima or other stationary points in Ω. Furthermore, its infimum value 0 is on the boundary $\partial\Omega$ and nowhere else.

The next lemma which is especially useful when working with log-likelihood functions is

Lemma 3.2. Let $l(\phi)$ be a twice continuously differentiable log-likelihood function with ϕ varying in a connected open subset $\Omega \subset R^p$.

i) If $\lim_{\phi \to \partial\Omega} l(\phi) = c$ where c is either a real number or $-\infty$, and,

ii) if the Hessian matrix of the second derivatives is negative definite at every point $\phi \in \Omega$ for which the gradient vector vanishes,

then $l(\phi)$ has a unique maximum and no other stationary points. Furthermore, $l(\phi) > c$ for every $\phi \in \Omega$.

In these lemmas, condition (i) ensures the existence of the solutions to the likelihood equations in the parameter space, and condition (ii) ensures that the solution is a maximum and it is unique.

3.1.1.1. Location-only Case

We first discuss the existence and uniqueness problem when only μ is the parameter of interest. Note that the results obtained for the location only case will hold for the regression case when the shape and the scale parameters are kept fixed. The results we obtained here will help us to understand why the constraints are imposed on the shape parameters in (McDonald and Newey, 1988).

We assume, without loss of generality, that σ=1. Then the log-likelihood function is

$$l(\mu) = -((pq+1)/p)\sum_{i=1}^{n} \log(q + |x_i - \mu|^p), \tag{3.1.9}$$

where $-\infty < \mu < \infty$. First, since $\lim_{\mu \to \pm\infty} l(\mu) = -\infty$, the likelihood function always has a local maximum in the parameter space R according to the Lemma 3.2. The first two derivatives of (3.1.9) are

$$l'(\mu) = (pq+1)\sum_{i=1}^{n} \frac{|x_i - \mu|^{p-1} sign(x_i - \mu)}{q + |x_i - \mu|^p}, \quad p>1 \tag{3.1.10}$$

$$l''(\mu) = -(pq+1)\sum_{i=1}^{n} \frac{(p-1)q|x_i-\mu|^{p-2} - |x_i-\mu|^{2p-2}}{(q+|x_i-\mu|^p)^2}. \qquad (3.1.11)$$

We can easily see that the second derivative of the likelihood function given in (3.1.11) may be negative, positive or zero at a solution of the equation $l'(\mu)=0$. Therefore the likelihood function may have local maxima, local minima and inflection points.

In the following theorem we give the properties of the likelihood function for the location-only case.

Theorem 3.1. Let $x_1, x_2, ..., x_n$ be a data set in R, modeled by the $GT(x;\mu,p,q)$ distribution with known shape parameters p,q and unknown location parameter μ.

i) If the shape parameter q is chosen small enough then the likelihood function given in (3.1.9) has n local maxima, each close to one sample point, and $n-1$ local minima.

ii) When $p \le 1$, the likelihood function given in (3.1.9) has a local maximum at each observation for any values of q, and at these maximum points the likelihood function has cusps.

Proof. i) Suppose that the observations are ordered as $x_1 < x_2 < ... < x_n$. When $\mu < x_1$ then $l'(\mu) > 0$ and when $\mu > x_n$ then $l'(\mu) < 0$. So the estimating equation $l'(\mu)=0$ has at least one root between x_1 and x_n. This root is a local maximum. Now let us form the intervals $I_j = (x_j - q^{1/p}, x_j + q^{1/p})$, $j = 1,2,3,...,n$, and choose q in such a way that $I_i \cap I_{i+1} = \varnothing$, $i = 1,2,..,n-1$; that is, choose q small to make the intervals disjoint. Note that at the end points of I_j, $j = 1,2,3,...,n$ we have

$$l'(x_j - q^{1/p}) = \frac{1}{2q^{1/p}} + \sum_{i \neq j}^{n} \frac{\left|x_i - x_j + q^{1/p}\right|^{p-2}(x_i - x_j + q^{1/p})}{q + \left|x_i - x_j + q^{1/p}\right|^p}, \qquad (3.1.12)$$

$$l'(x_j + q^{1/p}) = -\frac{1}{2q^{1/p}} + \sum_{i \neq j}^{n} \frac{\left|x_i - x_j - q^{1/p}\right|^{p-2}(x_i - x_j - q^{1/p})}{q + \left|x_i - x_j - q^{1/p}\right|^{p}}. \qquad (3.1.13)$$

Since q is chosen small enough, the term $1/(2q^{1/p})$ will be large so that $l'(x_j - q^{1/p}) > 0$ and $l'(x_j + q^{1/p}) < 0$. Thus $l'(\mu)=0$ will have a root in the interval I_j, for $j = 1,2,3,...,n$ and this root will be a local maximum. Similarly, we can show that the likelihood function will have a local minimum in each interval $(x_i + q^{1/p}, x_{i+1} - q^{1/p})$, for $i = 1,2,..,n-1$.

ii) Let $x_k < \mu < x_{k+1}$, $k = 1,2,...,n-1$. We have

$$l(\mu) = -(\frac{pq+1}{p})\{\sum_{i=1}^{k} \log[q + (\mu - x_i)^p] + \sum_{i=k+1}^{n} \log[q + (x_i - \mu)^p]\}, \qquad (3.1.14)$$

and

$$l'(\mu) = -(pq+1)[\sum_{i=1}^{k} \frac{1}{q(\mu - x_i)^{1-p} + (\mu - x_i)} - \sum_{i=k+1}^{n} \frac{1}{q(x_i - \mu)^{1-p} + (x_i - \mu)}]. \qquad (3.1.15)$$

It can be easily seen that as μ approaches x_k from the right $l'(\mu)$ will be $-\infty$ and as μ approaches x_{k+1} from the left $l'(\mu)$ will be ∞ so that $l'(\mu)$ has a root between x_k and x_{k+1}. This means that $l(\mu)$ will have a critical point in the interval (x_k, x_{k+1}). Since

$$l''(\mu) = (pq+1)\{\sum_{i=1}^{k} \frac{(1-p)q(\mu - x_i)^{-p} + 1}{[q(\mu - x_i)^{1-p} + (\mu - x_i)]^2} + \sum_{i=k+1}^{n} \frac{(1-p)q(x_i - \mu)^{-p} + 1}{[q(x_i - \mu)^{1-p} + (x_i - \mu)]^2}\} \geq 0, \qquad (3.1.16)$$

for $p \le 1$, that critical point should be a local minimum. It can be shown that as μ approaches x_k from the left, the likelihood function is increasing and it is decreasing as μ approaches x_k from the right. So it should have a local maximum at $\mu = x_k$. Since the likelihood function is continuous but it is not differentiable at each x_i, $i = 1,2,...,n$, this maximum should be a cusp. The likelihood function is also increasing for $\mu < x_1$ and decreasing for $x_n < \mu$. As a result, the likelihood function has a cusp at each x_i, $i = 1,2,...,n$. □

Arslan (1992) also obtained the same result as in (i) for the Student's t distribution. For the triangular distribution Oliver (1972) obtained the same result as in (ii) (see also Rohatgi, 1976, p.380).

3.1.1.2. Scale-only Case

Next the existence and uniqueness of the solutions to the likelihood function are discussed when the location parameter μ is known. We assume, without loss of generality, that $\mu = 0$. Then the likelihood function is

$$L(\sigma) = \sigma^{npq} \prod_{i=1}^{n} [q\sigma^p + |x_i|^p]^{-(pq+1)/p}, \qquad (3.1.17)$$

where $\sigma > 0$. In the following theorem we show that this likelihood function always has a unique maximum.

Theorem 3.2. Let $x_1, x_2, ..., x_n$ be a data set in R which is modeled by the $GT(x;\sigma,p,q)$ distribution with known shape parameters p,q and unknown scale parameter σ, and let $m = \#\{i : x_i = 0\}$. If $m/n < pq/(pq+1)$, the likelihood function for the scale parameter given in (3.1.17) has a unique maximum in the parameter space $(0,\infty)$.

Proof. To show the existence of a local maximum in the parameter space $(0,\infty)$ according to Lemma 3.1, we have to show that $\lim L(\sigma)=0$ as $\sigma\to 0$ and as $\sigma\to\infty$ (0 and ∞ are the boundaries of the parameter space).

Since $L(\sigma)\le\sigma^{npq}(q\sigma^{p})^{-(pq+1)n/p}$, we have $\lim L(\sigma)=0$ as $\sigma\to\infty$. Now let $m=\#\{i:x_i=0\}$ and assume $x_1=x_2=...=x_m=0$. Then

$$L(\sigma)=\frac{\sigma^{npq-m(pq+1)}}{\prod_{i=m+1}^{n}[q\sigma^{p}+|x_i|^{p}]^{(pq+1)/p}}, \tag{3.1.18}$$

and we will have the following three cases as $\sigma\to 0$:

i) If $npq-m(pq+1)>0$; i.e. $m/n<pq/(pq+1)$, $\lim_{\sigma\to 0}L(\sigma)=0$,

ii) If $npq-m(pq+1)<0$; i.e. $m/n>pq/(pq+1)$, $\lim_{\sigma\to 0}L(\sigma)=\infty$,

iii) If $npq-m(pq+1)=0$; i.e. $m/n=pq/(pq+1)$,

$$\lim_{\sigma\to 0}L(\sigma)=\prod_{i=m+1}^{n}|x_i|^{-(pq+1)}.$$

Thus when $m/n<pq/(pq+1)$ the likelihood function $L(\sigma)$ has a local maximum in the parameter space $(0,\infty)$. From the case (ii), when $m/n>pq/(pq+1)$, the likelihood function increases and it tends to ∞ on the boundary. Thus it does not have any local maximum in the parameter space but it reaches its global maximum on the boundary of the parameter space. Case (iii) gives that the likelihood function is finite on the boundary. It reaches its global maximum on the boundary.

Now we will turn our attention to the uniqueness of the maximum in the parameter space. For the scale-only case taking the first derivative of the log-likelihood function we get

$$l'(\sigma)=\frac{npq}{\sigma}-\frac{pq+1}{p}\sum_{i=1}^{n}\frac{pq\sigma^{p-1}}{q\sigma^{p}+|x_i|^{p}}, \tag{3.1.19}$$

and at the solution of $l'(\sigma)=0$ we have

$$l''(\sigma)=-(pq+1)\sum_{i=1}^{n}\frac{pq\sigma^{p-2}|x_i|^p}{(q\sigma^p+|x_i|^p)^2}, \tag{3.1.20}$$

which is always negative. Thus if the likelihood function has a local maximum it must be unique. □

3.1.1.3. Location-scale Case

Finally, we will focus on the location and scale parameter case and jointly estimate these two parameters. Location-only case and scale-only case will be impotent to investigate the behavior of location-scale likelihood.

The following theorem summarizes the behavior of the likelihood function when the location and scale parameters are unknown and the shape parameters are known.

Theorem 3.3. Let $x_1,x_2,....,x_n$ be a data set in R which is modeled by the $GT(x;\mu,\sigma,p,q)$ distribution with known shape parameters p,q and unknown location and scale parameters μ,σ. Let $m/n<pq/(pq+1)$, where m is the number of coincident observations. If $p\geq 2$ and $pq\geq 1$, then the likelihood function for the location and the scale parameters of the GT distribution always has a unique maximum in the parameter space RxR^+.

Proof. The location-scale likelihood function up to a scalar constant is

$$L(\mu,\sigma)=(1/\sigma^n)\prod_{i=1}^{n}[q+(|x_i-\mu|/\sigma)^p]^{-(pq+1)/p}, \tag{3.1.21}$$

and the boundary of the parameter space is the line $\sigma=0$ and $\mu=\pm\infty$. To show the existence of a local maximum according to Lemma 3.1, we will show that when

$\theta = (\mu, \sigma) \in RxR^+$ tends to the boundary of the parameter space, the likelihood tends to zero. There are three cases:

1) If σ is fixed then $L(\mu, \sigma) \to 0$ as $\mu \to \pm\infty$.

2) Since $L(\mu, \sigma) \le \sigma^{-n} q^{-(pq+1)/p}$, then we have $L(\mu, \sigma) \to 0$ as $\sigma \to \infty$.

3) Let $\sigma \to 0$. If $\mu \to \pm\infty$ then clearly $L(\mu, \sigma) \to 0$. Suppose now that μ is bounded and there are m coincident observations, say $x_1 = x_2 = ... = x_m$. Then at $\mu = x_1$ the likelihood function becomes

$$L(x_1, \sigma) = \sigma^{npq - m(pq+1)} \prod_{i=m+1}^{n} [q\sigma^p + |x_i - x_1|^p]^{-(pq+1)/p}. \qquad (3.1.22)$$

We obtain the following cases as $\sigma \to 0$:

i) When $m/n < pq/(pq+1)$ the likelihood function has a local maximum in the parameter space.

ii) When there are exactly $npq/(pq+1)$ coincident observations, the likelihood function will be bounded on the boundary. Further it will not have any local maximum in the parameter space, but its global maximum will appear on the boundary of the parameter space.

iii) When $m/n > pq/(pq+1)$ the likelihood function will be arbitrarily large as σ tends to zero along the path $\mu = x_1$. Thus the point $(x_1, \sigma = 0)$ can be taken as global maximum.

To show the uniqueness of the maximum according to Lemma 3.2 we should look at the second derivative matrix of the log-likelihood function. The second derivatives will be as follows:

$$l_{\mu\mu}(\mu, \sigma) = (pq+1) \sum_{i=1}^{n} \frac{-(p-1)q\sigma^p |x_i - \mu|^{p-2} + |x_i - \mu|^{2p-2}}{(q\sigma^p + |x_i - \mu|^p)^2}, \qquad (3.1.23)$$

$$l_{\sigma\sigma}(\mu, \sigma) = -\frac{npq}{\sigma^2} - (pq+1) \sum_{i=1}^{n} \frac{q(p-1)\sigma^{p-2} |x_i - \mu|^p - q^2\sigma^{2p-2}}{(q\sigma^p + |x_i - \mu|^p)^2}, \qquad (3.1.24)$$

$$l_{\mu\sigma}(\mu,\sigma) = -(pq+1)pq\sigma^{p-1}\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-1}sign(x_i-\mu)}{(q\sigma^p+|x_i-\mu|^p)^2}. \tag{3.1.25}$$

When $l_\sigma(\mu,\sigma)=0$, we have

$$l_{\sigma\sigma} = -(pq+1)pq\sigma^{p-2}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{(q\sigma^p+|x_i-\mu|^p)^2}, \tag{3.1.26}$$

which is always negative. When $l_\mu(\mu,\sigma)=0$, we have

$$l_{\mu\sigma} = (\frac{pq+1}{\sigma})\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}(x_i-\mu)[|x_i-\mu|^p-(p-1)q\sigma^p]}{(q\sigma^p+|x_i-\mu|^p)^2}. \tag{3.1.27}$$

Further, $l_{\mu\mu}(\mu,\sigma)$ can be rewritten as

$$l_{\mu\mu} = -\frac{(pq+1)}{pq\sigma^p}\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]^2}{(q\sigma^p+|x_i-\mu|^p)^2} + (\frac{pq+1}{pq\sigma^p})\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]}{q\sigma^p+|x_i-\mu|^p}. \tag{3.1.28}$$

Now let D^2l denote the second derivative matrix of the log-likelihood function. From above we have the following determinant of D^2l at a solution of the estimating equations :

$$\det(D^2l) = l_{\mu\mu}(\mu,\sigma)l_{\sigma\sigma}(\mu,\sigma)-[l_{\mu\sigma}(\mu,\sigma)]^2$$

$$= (\frac{pq+1}{\sigma})^2\{\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]^2}{(q\sigma^p+|x_i-\mu|^p)^2}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{(q\sigma^p+|x_i-\mu|^p)^2} - [\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}(x_i-\mu)[|x_i-\mu|^p-(p-1)q\sigma^p]}{(q\sigma^p+|x_i-\mu|^p)^2}]^2\}$$

$$-(\frac{pq+1}{\sigma})^2\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]}{q\sigma^p+|x_i-\mu|^p}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{(q\sigma^p+|x_i-\mu|^p)^2}. \tag{3.1.29}$$

The term

$$\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]^2}{(q\sigma^p+|x_i-\mu|^p)^2}\sum_{i=1}^{n}\frac{|x_i-\mu|^p}{(q\sigma^p+|x_i-\mu|^p)^2}$$
$$-[\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}(x_i-\mu)[|x_i-\mu|^p-(p-1)q\sigma^p]}{(q\sigma^p+|x_i-\mu|^p)^2}]^2, \quad (3.1.30)$$

is always positive from the Cauchy-Schwarz inequality. Since $l_{\sigma\sigma}(\mu,\sigma)$ is always negative at the solutions of the estimating equations the behavior of the $\det(D^2l)$ depends on the behavior of $l_{\mu\mu}(\mu,\sigma)$. The first term of $l_{\mu\mu}(\mu,\sigma)$ is always negative, therefore the behavior of $l_{\mu\mu}(\mu,\sigma)$ at a solution of the estimating equations will depend only on the last term

$$(\frac{pq+1}{pq\sigma^p})\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]}{q\sigma^p+|x_i-\mu|^p}. \quad (3.1.31)$$

If the above term (3.1.31) is negative, $l_{\mu\mu}(\mu,\sigma)$ will be negative, and hence $\det(D^2l)$ will be positive. Thus the critical point (the solution of the estimating equations at which we computed D^2l) of the likelihood function will be a local maximum.

Note that if $l_\sigma(\mu,\sigma)=0$, we have

$$\sum_{i=1}^{n}\frac{|x_i-\mu|^p-(1/p)\sigma^p}{|x_i-\mu|^p+q\sigma^p}=0. \quad (3.1.32)$$

When $p\geq 2$, we have

$$\sum_{i=1}^{n}\frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(1/p)\sigma^p]}{|x_i-\mu|^p+q\sigma^p}\leq 0, \quad (3.1.33)$$

from the Theorem 22 of Hardy, Littlewood and Polya (1997), and when $pq \geq 1$ we get the following inequality

$$\sum_{i=1}^{n} \frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(p-1)q\sigma^p]}{|x_i-\mu|^p+q\sigma^p} \leq \sum_{i=1}^{n} \frac{|x_i-\mu|^{p-2}[|x_i-\mu|^p-(1/p)\sigma^p]}{|x_i-\mu|^p+q\sigma^p} \leq 0. \quad (3.1.34)$$

This result tells us that when $p \geq 2$ and $pq \geq 1$ then $l_{\mu\mu}(\mu,\sigma)$ is always negative. Therefore when $p \geq 2$ and $pq \geq 1$, $\det(D^2 l)$ is positive and so the uniqueness of the maximum is obtained from the Lemma 3.2. □

Theorem 3.4. If $p \geq 2$ or $pq \geq 1$ does not hold, then the likelihood function may have local maxima or saddle points, but it cannot have any local minima in the parameter space.

Proof. If $p \geq 2$ or $pq \geq 1$ does not hold, $\det(D^2 l)$ may not be positive any longer. Therefore the uniqueness of the maximum cannot be guaranteed, and the likelihood function may have other critical points : local maxima or saddle points. However, we will never get any local minima since $l_{\sigma\sigma}(\mu,\sigma)$ is always negative at the solutions of the estimating equations. □

3.2. The *GST* Distribution Case

Let $x_1, x_2, ..., x_n$ be modeled with the *GST* distribution. Then the log-likelihood function is

$$l(\mu,\sigma,p,k) = -n\log\sigma - \frac{n}{2}\log k + n\log\Gamma(p) - n\log\Gamma(p-1/2) - p\sum_{i=1}^{n}\log[1+\frac{(x-\mu)^2}{k\sigma^2}]. \quad (3.2.1)$$

The likelihood equations are

$$l_\sigma(\mu,\sigma,p,k) = -\frac{n}{\sigma} + \frac{2p}{k\sigma^3}\sum_{i=1}^{n}\frac{(x-\mu)^2}{1+\frac{(x-\mu)^2}{k\sigma^2}} = 0, \tag{3.2.2}$$

$$l_\mu(\mu,\sigma,p,k) = \frac{2p}{k\sigma^2}\sum_{i=1}^{n}\frac{x-\mu}{1+\frac{(x-\mu)^2}{k\sigma^2}} = 0, \tag{3.2.3}$$

$$l_p(\mu,\sigma,p,k) = n\psi(p) - n\psi(p-1/2) - \sum_{i=1}^{n}\log[1+\frac{(x-\mu)^2}{k\sigma^2}] = 0, \tag{3.2.4}$$

$$l_k(\mu,\sigma,p,k) = -\frac{n}{2k} + \frac{p}{\sigma^2k^2}\sum_{i=1}^{n}\frac{(x-\mu)^2}{1+\frac{(x-\mu)^2}{k\sigma^2}} = 0. \tag{3.2.5}$$

The maximum likelihood estimates are obtained by computing (3.2.2)-(3.2.5) simultaneously. However, for the same reason in the *GT* distribution case we will treat the shape parameters being fixed and concentrate on the location and scale parameters.

The equations given in (3.2.2) and (3.2.3) can be rewritten as follows:

$$\hat{\mu} = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i, \tag{3.2.6}$$

and

$$\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2, \tag{3.2.7}$$

where

$$w_i = \frac{2p}{k + (\frac{x_i - \mu}{\sigma})^2}. \tag{3.2.8}$$

The weight function $w(d)=\frac{2p}{k+d}$, where $d=(\frac{x_i-\mu}{\sigma})^2$, is a positive function. Some properties of this function are

i) At the solutions of (3.2.6) and (3.2.7) we have $\sum_{i=1}^{n} w_i / n = (2p-1)/k$.

ii) $w(d)$ is a decreasing function of d so that the outlying observations are downweighted. Thus, the estimators given in equations (3.2.6) and (3.2.7) exhibit resistance to the outliers in the dataset.

3.2.1. Existence and Uniqueness of the Maximum Likelihood Estimators With Known Shape Parameters

In this section we investigate under which conditions the maximum likelihood estimates exist and are unique.

3.2.1.1. Location-Only Case

Let $x_1, x_2, ..., x_n$ be modeled with the *GST* distribution with known p and k. For the location parameter μ, the log-likelihood function up to a scalar constant is

$$l(\mu) = -p\sum_{i=1}^{n} \log[k+(x_i-\mu)^2]. \qquad (3.2.9)$$

The boundary of the parameter space is $\mp\infty$. Since $\lim_{\mu\to\pm\infty} l(\mu) = -\infty$, by Lemma 3. 2 $l(\mu)$ always has a local maximum in R.

Now the second derivative is

$$l''(\mu) = 2p\sum_{i=1}^{n} \frac{(x_i-\mu)^2-k}{[(x_i-\mu)^2+k]^2}. \qquad (3.2.10)$$

This equation (3.2.10) may be negative or positive at a solution of the estimating equation. So the likelihood function may have local maxima and local minima.

A similar result as in Theorem 3.1(i) can be obtained for the *GST* distribution. The proof is similar to that of Theorem 3.1 and therefore is omitted. We have

Theorem 3.5. Let p be given. If k is chosen small enough then the estimating equation will have $2n-1$ solutions consisting n local maxima, each close to one sample point, and $n-1$ local minima.

3.2.1.2. Scale-Only Case

Let $x_1, x_2, ..., x_n$ be modeled with the *GST* distribution with known p and k. Without loss of generality, we may assume $\mu = 0$. For the scale parameter we have the following likelihood function up to a scalar constant :

$$L(\sigma) = \sigma^{(2p-1)n} \prod_{i=1}^{n} (k\sigma^2 + x_i^2)^{-p}, \tag{3.2.11}$$

where $\sigma > 0$.

Since $L(\sigma) \le \sigma^{-n} k^{-pn}$, we have $\lim_{\sigma \to \infty} L(\sigma) = 0$.

Now let $m = \#\{i : x_i = 0\}$. Suppose that $x_1 = x_2 = ... = x_m = 0$. So the likelihood function becomes,

$$\sigma^{(2p-1)n-2pm} \prod_{i=m+1}^{n} (k\sigma^2 + x_i^2)^{-p}, \tag{3.2.12}$$

giving three cases to investigate as $\sigma \to 0$:

i) If $\frac{m}{n} < \frac{2p-1}{2p}$ then $\lim_{\sigma \to 0} L(\sigma) = 0$.

ii) If $\frac{m}{n} > \frac{2p-1}{2p}$ then $\lim_{\sigma \to 0} L(\sigma) = \infty$.

iii) If $\frac{m}{n} = \frac{2p-1}{2p}$ then $\lim_{\sigma \to 0} L(\sigma) = \prod_{i=m+1}^{n} (x_i^2)^{-p}$.

Thus, when $\frac{m}{n} < \frac{2p-1}{2p}$ the scale likelihood function has a local maximum in the parameter space $(0,\infty)$, by Lemma 3.1. By the case (ii), the likelihood function does not have any local maximum in the parameter space $(0,\infty)$ but it tends to ∞, and the case (iii) gives that the likelihood function is finite at the boundary, and it reaches its global maximum at the boundary. So $\frac{m}{n} < \frac{2p-1}{2p}$ is a necessary and sufficient condition for getting a local maximum in the parameter space $(0,\infty)$.

Now the log-likelihood function is

$$l(\sigma) = (2p-1)n\log\sigma - p\sum_{i=1}^{n}\log(k\sigma^2 + x_i^2). \tag{3.2.13}$$

Its first derivative is

$$l'(\sigma) = ((2p-1)n/\sigma) - 2pk\sigma\sum_{i=1}^{n}\frac{1}{k\sigma^2 + x_i^2}. \tag{3.2.14}$$

If $l'(\sigma) = 0$ then

$$l''(\sigma) = -4pk\sum_{i=1}^{n}\frac{x_i^2}{(k\sigma^2 + x_i^2)^2}, \tag{3.2.15}$$

which is always negative. Thus, if the likelihood function has a local maximum it must be unique.

As a result, we can say that if $\frac{m}{n} < \frac{2p-1}{2p}$ the likelihood function has a unique maximum in the parameter space $(0,\infty)$.

3.2.1.3. Location-scale case

The likelihood function up to a scalar constant is

$$L(\mu,\sigma)=\sigma^{2np-n}\prod_{i=1}^{n}[k\sigma^2+(x_i-\mu)^2]^{-p}, \tag{3.2.16}$$

where $\theta=(\mu,\sigma)\in RxR^+$ and the boundary of the parameter space is the line $\sigma=0$ and $\mu=\pm\infty$.

For the existence of a local maximum in the parameter space, it is enough to show that when θ tends to the boundary, the likelihood tends to zero.

i) Let $\mu\to\pm\infty$. When σ is bounded $L(\mu,\sigma)$ tends to zero.

ii) Let $\sigma\to\infty$. Since $L(\mu,\sigma)\le\sigma^{-n}k^{-np}$, $L(\mu,\sigma)$ tends to zero.

iii) Let $\sigma\to 0$. If $\mu\to\pm\infty$ then clearly $L(\mu,\sigma)\to 0$. Suppose now that μ is bounded and there are m coincident observations, say $x_1=x_2=...=x_m$. Then at $\mu=x_1$ the likelihood function becomes,

$$L(x_1,\sigma)=\sigma^{2np-n-2pm}\prod_{i=m+1}^{n}\{k\sigma^2+(x_i-x_1)^2\}^{-p}. \tag{3.2.17}$$

For the limit

$$\lim_{\sigma\to 0}L(\mu,\sigma)=\lim_{\sigma\to 0}\{\sigma^{2np-n-2pm}\prod_{i=m+1}^{n}[k\sigma^2+(x_i-x_1)^2]^{-p}\}, \tag{3.2.18}$$

we obtain the same cases (i),(ii),and (iii) which we obtained in the scale-only case. That is, when $\frac{m}{n}<\frac{2p-1}{2p}$ the likelihood function has a local maximum in the parameter space $(0,\infty)$, by Lemma 3.1. If $\frac{m}{n}>\frac{2p-1}{2p}$, the likelihood function will be arbitrarily large as σ tends to zero along the path $\mu=x_1$. Thus the point $(x_1,\sigma=0)$ can be taken as global maximum.

Vaughan (1992) also studied the uniqueness of the maximum likelihood estimates for the location-scale case with the special choice of the parameters p and k mentioned in Chapter 2. Here we search for the uniqueness condition removing this restriction

The log-likelihood function is

$$l(\mu,\sigma) = (2p-1)n\log\sigma - p\sum_{i=1}^{n}\log[k\sigma^2 + (x_i-\mu)^2]. \tag{3.2.19}$$

Then we have the following partial derivatives.

$$l_\mu(\mu,\sigma) = 2p\sum_{i=1}^{n}\frac{x_i-\mu}{k\sigma^2+(x_i-\mu)^2}, \tag{3.2.20}$$

$$l_\sigma(\mu,\sigma) = (2p-1)\frac{n}{\sigma} - 2pk\sigma\sum_{i=1}^{n}\frac{1}{k\sigma^2+(x_i-\mu)^2}, \tag{3.2.21}$$

$$l_{\mu\mu}(\mu,\sigma) = 2p\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2}{[k\sigma^2+(x_i-\mu)^2]^2}, \tag{3.2.22}$$

$$l_{\mu\sigma}(\mu,\sigma) = -4pk\sigma\sum_{i=1}^{n}\frac{x_i-\mu}{[k\sigma^2+(x_i-\mu)^2]^2}, \tag{3.2.23}$$

$$l_{\sigma\sigma}(\mu,\sigma) = -(2p-1)\frac{n}{\sigma^2} - 2kp\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2}{[k\sigma^2+(x_i-\mu)^2]^2}. \tag{3.2.24}$$

When $l_\sigma(\mu,\sigma)=0$ we have

$$l_{\sigma\sigma}(\mu,\sigma) = -4kp\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{[k\sigma^2+(x_i-\mu)^2]^2}, \tag{3.2.25}$$

which is always negative.

If $l_\mu(\mu,\sigma)=0$ we have

$$\sum_{i=1}^{n}\frac{x_i-\mu}{[k\sigma^2+(x_i-\mu)^2]^2} = -\frac{1}{2k\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)[(x_i-\mu)^2-k\sigma^2]}{[k\sigma^2+(x_i-\mu)^2]^2}, \tag{3.2.26}$$

and hence,

$$l_{\mu\sigma}(\mu,\sigma) = \frac{2p}{\sigma}\sum_{i=1}^{n}\frac{(x_i-\mu)[(x_i-\mu)^2-k\sigma^2]}{[k\sigma^2+(x_i-\mu)^2]^2}. \tag{3.2.27}$$

Finally $l_{\mu\mu}(\mu,\sigma)$ can be rewritten as

$$l_{\mu\mu}(\mu,\sigma) = -\frac{p}{k\sigma^2}\sum_{i=1}^{n}\frac{[(x_i-\mu)^2-k\sigma^2]^2}{[k\sigma^2+(x_i-\mu)^2]^2}+\frac{p}{k\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2}{k\sigma^2+(x_i-\mu)^2}. \tag{3.2.28}$$

Let D^2l denote the second derivative matrix of the log-likelihood function. From above we have the following determinant of D^2l at a solution of the estimating equations,

$$\det(D^2l) = l_{\mu\mu}(\mu,\sigma)l_{\sigma\sigma}(\mu,\sigma)-[l_{\mu\sigma}(\mu,\sigma)]^2$$

$$=\frac{4p^2}{\sigma^2}\{\sum_{i=1}^{n}\frac{[(x_i-\mu)^2-k\sigma^2]^2}{[k\sigma^2+(x_i-\mu)^2]^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{[k\sigma^2+(x_i-\mu)^2]^2}-$$
$$[\sum_{i=1}^{n}\frac{(x_i-\mu)((x_i-\mu)^2-k\sigma^2)}{[k\sigma^2+(x_i-\mu)^2]^2}]^2\}$$
$$-\frac{4p^2}{\sigma^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2}{k\sigma^2+(x_i-\mu)^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{[k\sigma^2+(x_i-\mu)^2]^2}. \tag{3.2.29}$$

Since $l_{\sigma\sigma}(\mu,\sigma)$ is negative at the solutions of the estimating equations so the behavior of the $\det(D^2l)$ depends on the behavior of $l_{\mu\mu}(\mu,\sigma)$. The first term of $l_{\mu\mu}(\mu,\sigma)$ is always negative, therefore the behavior of $l_{\mu\mu}(\mu,\sigma)$ at a solution of the estimating equations will depend on the last term of it. So if the last term is negative $l_{\mu\mu}(\mu,\sigma)$ will be negative and hence $\det(D^2l)$ will be positive, and the solutions will be local maxima of the likelihood function.

From the Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n}\frac{[(x_i-\mu)^2-k\sigma^2]^2}{[k\sigma^2+(x_i-\mu)^2]^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{[k\sigma^2+(x_i-\mu)^2]^2}-$$
$$[\sum_{i=1}^{n}\frac{(x_i-\mu)((x_i-\mu)^2-k\sigma^2)}{[k\sigma^2+(x_i-\mu)^2]^2}]^2 \tag{3.2.30}$$

is positive. So the first part of $\det(D^2 l)$ is positive.

Now for the second part if $l_\sigma(\mu,\sigma)=0$ then we have

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2/(2p-1)}{(x_i-\mu)^2+k\sigma^2}=0. \tag{3.2.31}$$

When $p\geq 1$ then

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2}{(x_i-\mu)^2+k\sigma^2}\leq\sum_{i=1}^{n}\frac{(x_i-\mu)^2-k\sigma^2/(2p-1)}{(x_i-\mu)^2+k\sigma^2}=0, \tag{3.2.32}$$

giving $l_{\mu\mu}(\mu,\sigma)$ negative. Therefore, when $p\geq 1$ and $\frac{m}{n}<\frac{2p-1}{2p}$ the likelihood function for location and scale parameters of the *GST* distribution always has a unique maximum in the parameter space RxR^+.

However, when $p\geq 1$ is not satisfied, $\det(D^2 l)$ may not be positive any longer. Therefore the uniqueness of the maximum cannot be guaranteed, and the likelihood function may have other stationary points. Clearly they must be local maxima and saddle points (no local minima) since $l_{\sigma\sigma}(\mu,\sigma)$ is always negative at the solutions of the estimating equations.

3.3. The *GET* Distribution Case

Let $x_1,x_2,...,x_n$ be modeled with the *GET* distribution which is given by the probability density function

$$f(x;\gamma,\mu,\sigma)=C\frac{1}{\sigma}\{\gamma^2+(\frac{x-\mu}{\sigma})^2\}^{-\frac{1+\gamma^2}{2}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}. \tag{3.3.1}$$

This is the form of the *GET* distribution which represents both normal by its second term and the Student's t distribution by its first term.

The log-likelihood function except for a scalar constant is

$$l = n\log(\frac{1}{\sigma}) - \frac{1}{2}\sum_{i=1}^{n}\{(\frac{x_i-\mu}{\sigma})^2 + (\gamma^2+1)\log[\gamma^2 + (\frac{x_i-\mu}{\sigma})^2]\}. \tag{3.3.2}$$

The maximum likelihood estimating equations are

$$l_\mu(\mu,\sigma) = \sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma^2})[1 + \frac{1+\gamma^2}{\gamma^2 + (\frac{x_i-\mu}{\sigma})^2}] = 0, \tag{3.3.3}$$

$$l_\sigma(\mu,\sigma) = -\frac{n}{\sigma} + \sum_{i=1}^{n}[(x_i-\mu)^2\sigma^{-3} + (\gamma^2+1)\frac{(x_i-\mu)^2\sigma^{-3}}{\gamma^2 + (\frac{x_i-\mu}{\sigma})^2}] = 0. \tag{3.3.4}$$

So the ML estimators are

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \tag{3.3.5}$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2, \tag{3.3.6}$$

where

$$w_i = 1 + \frac{\gamma^2+1}{\gamma^2 + (\frac{x_i-\mu}{\sigma})^2}. \tag{3.3.7}$$

The weight function $w(d) = 1 + \frac{\gamma^2+1}{\gamma^2+d}$ is a decreasing function of $d = (\frac{x_i-\hat{\mu}}{\hat{\sigma}})^2$.

And thus the outlying observations downweighted by the corresponding weights.

3.3.1. Existence and Uniqueness of the Maximum Likelihood Estimators With Known Shape Parameters

In this section we will try to obtain a solution for the existence and uniqueness problem for the *GET* distribution under consideration.

3.3.1.1. Location-only case

Let $x_1, x_2, ..., x_n$ be modeled with the *GET* distribution with known γ. For the location parameter μ, the log-likelihood function up to a scalar constant is

$$l(\mu) = -\frac{1}{2}\sum_{i=1}^{n}\{(x_i-\mu)^2 + (1+\gamma^2)\log[\gamma^2 + (x_i-\mu)^2]\}, \tag{3.3.8}$$

where $-\infty < \mu < \infty$.

Then

$$l'(\mu) = \sum_{i=1}^{n}(x_i-\mu)[1 + \frac{1+\gamma^2}{(x_i-\mu)^2+\gamma^2}] = 0, \tag{3.3.9}$$

implies that

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \tag{3.3.10}$$

where

$$w_i = 1 + \frac{1+\gamma^2}{(x_i-\mu)^2+\gamma^2}. \tag{3.3.11}$$

Since $\lim_{\mu\to\pm\infty} l(\mu) = -\infty$ $l(\mu)$ always has a local maximum in R by Lemma 3. 2.

The second derivative is

$$\begin{aligned} l''(\mu) &= -n + (1+\gamma^2)\sum_{i=1}^{n}\frac{(x_i-\mu)^2-\gamma^2}{[(x_i-\mu)^2+\gamma^2]^2} \\ &= -2\gamma^4\sum_{i=1}^{n}\frac{1}{[\gamma^2+(x_i-\mu)^2]^2} - \sum_{i=1}^{n}\frac{(x_i-\mu)^4}{[\gamma^2+(x_i-\mu)^2]^2} \\ &\quad -\gamma^2\sum_{i=1}^{n}\frac{1}{[\gamma^2+(x_i-\mu)^2]^2} + (1-\gamma^2)\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{[\gamma^2+(x_i-\mu)^2]^2} \end{aligned} \tag{3.3.12}$$

Note that when $\gamma^2 \geq 1$, l^* will always be negative. Thus when $\gamma^2 \geq 1$ any critical point is a relative maximum for $l(\mu)$.

A similar result as in Theorem 3.1(i) can be obtained for the *GET* distribution. The proof is similar to that of Theorem 3.1 and therefore is omitted. We have

Theorem 3.6. If the shape parameter γ is chosen small enough then the estimating equation will have $2n-1$ solutions consisting n local maxima, each close to one observation, and $n-1$ local minima.

3.3.1.2. Scale-only Case

Let $x_1, x_2, ..., x_n$ be our dataset modeled with the *GET* distribution with known γ. Without loss of generality, we may assume that $\mu = 0$. For the scale parameter we have the following likelihood function up to a scalar constant

$$L(\sigma) = \sigma^{n\gamma^2} \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2} (\gamma^2\sigma^2 + x_i^2)^{-\frac{\gamma^2+1}{2}}, \tag{3.3.13}$$

where $\sigma > 0$.

Since

$$L(\sigma) \leq \gamma^{-n(1+\gamma^2)} \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2}, \tag{3.3.14}$$

we have $\lim_{\sigma \to \infty} L(\sigma) = 0$.

Next let $m = \#\{i : x_i = 0\}$ and suppose that $x_1 = x_2 = ... = x_m = 0$. For this case the likelihood function becomes

$$L(\sigma) = \frac{\sigma^{n\gamma^2 - (\gamma^2+1)m}}{\prod_{i=m+1}^{n} e^{\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2} (\gamma^2\sigma^2 + x_i^2)^{\frac{\gamma^2+1}{2}}}$$

$$= \frac{\sigma^{n\gamma^2 - m(\gamma^2+1)}}{e^{\frac{1}{2\sigma^2}\sum_{i=m+1}^{n} x_i^2} \prod_{i=m+1}^{n} (\gamma^2\sigma^2 + x_i^2)^{\frac{1+\gamma^2}{2}}} \tag{3.3.15}$$

When $\frac{m}{n} \leq \frac{\gamma^2}{\gamma^2+1}$ we have $\lim_{\sigma\to 0} L(\sigma) = 0$. If $\frac{m}{n} > \frac{\gamma^2}{\gamma^2+1}$ we can still have $\lim_{\sigma\to 0} L(\sigma) = 0$. Thus without imposing any condition on the data set we will always have a local maximum point of the likelihood function in the parameter space (Lemma 3.1).

Next we will turn our attention to the uniqueness of the maximum point for the scale only case. The scale only log-likelihood function can be obtained as

$$l(\sigma) = n\gamma^2 \log\sigma - \frac{1}{2}\sum_{i=1}^{n}[(\frac{x_i}{\sigma})^2 + (\gamma^2+1)\log(\gamma^2\sigma^2 + x_i^2)]. \tag{3.3.16}$$

The first and the second derivatives of the log-likelihood function are

$$l'(\sigma) = \frac{n\gamma^2}{\sigma} + \sum_{i=1}^{n}[\frac{x_i^2}{\sigma^3} - (\gamma^2+1)\frac{\gamma^2\sigma}{x_i^2 + \gamma^2\sigma^2}] \tag{3.3.17}$$

$$l''(\sigma) = -2\sum_{i=1}^{n}[x_i^2\sigma^{-4} + \frac{\gamma^2(\gamma^2+1)x_i^2}{(x_i^2 + \gamma^2\sigma^2)^2}]. \tag{3.3.18}$$

We can observe that the second derivative of the log-likelihood function is always negative. Thus the likelihood function has a unique maximum in the parameter space $(0,\infty)$.

3.3.1.3. Location-Scale Case

Now we can explore the location-scale case. In this case the likelihood function up to a scalar constant is

$$L(\mu,\sigma)=\sigma^{n\gamma^2}\prod_{i=1}^{n}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2}[\gamma^2\sigma^2+(x_i-\mu)^2]^{-\frac{1+\gamma^2}{2}}. \qquad (3.3.19)$$

where $(\mu,\sigma)\in RxR^+$.

For the existence of a local maximum in the parameter space it is enough to show that when (μ,σ) tends to the boundary of the parameter space the likelihood function tends to zero.

There are three cases:

i) Let $\mu\to\pm\infty$. When σ is bounded, $L(\mu,\sigma)$ tends to zero.

ii) Let $\sigma\to\infty$. Since $L(\mu,\sigma)<\gamma^{-n(1+\gamma^2)}\sigma^{-n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$, we have $L(\mu,\sigma)\to 0$ as $\sigma\to\infty$.

iii) Let $\sigma\to 0$. Then clearly $\lim_{\mu\to\mp\infty}L(\mu,\sigma)=0$. Suppose that μ is bounded and there are m coincident observations, say $x_1=x_2=...=x_m$. Then at $\mu=x_1$ the likelihood function becomes

$$L(x_1,\sigma)=\sigma^{n\gamma^2-m(\gamma^2+1)}\prod_{i=m+1}^{n}[\gamma^2\sigma^2+(x_i-x_1)^2]^{-\frac{1+\gamma^2}{2}}e^{-\frac{1}{2}(\frac{x_i-x_1}{\sigma})^2}. \qquad (3.3.20)$$

and taking the limit as $\sigma\to 0$ we always have $\lim_{\sigma\to 0}L(x_1,\sigma)=0$. Thus the likelihood function for location and scale has always a local maximum in the parameter space by Lemma 3.1.

To investigate the uniqueness of the maximum we need the second partial derivatives of the log likelihood function given below:

$$l_{\mu\mu}(\mu,\sigma)=-\frac{n}{\sigma^2}+\sum_{i=1}^{n}\frac{(\gamma^2+1)[(x_i-\mu)^2-\sigma^2\gamma^2]}{[(x_i-\mu)^2+\sigma^2\gamma^2]^2}, \qquad (3.3.21)$$

$$l_{\sigma\sigma}(\mu,\sigma)=-\frac{n\gamma^2}{\sigma^2}+\sum_{i=1}^{n}\{-3(x_i-\mu)^2\sigma^{-4}-\frac{(1+\gamma^2)\gamma^2[(x_i-\mu)^2-\sigma^2\gamma^2]}{[(x_i-\mu)^2+\sigma^2\gamma^2]^2}\} \qquad (3.3.22)$$

$$l_{\mu\sigma}(\mu,\sigma) = -2\sum_{i=1}^{n}\{(x_i-\mu)\sigma^{-3} + \frac{(1+\gamma^2)\gamma^2\sigma(x_i-\mu)}{[(x_i-\mu)^2+\sigma^2\gamma^2]^2}\}, \tag{3.3.23}$$

We will now evaluate these partial derivatives at a solution of the estimating equations. If $l_\sigma = 0$ we have $-\frac{n\gamma^2}{\sigma^2} = \sum_{i=1}^{n}[(x_i-\mu)^2\sigma^{-4} - \frac{(1+\gamma^2)\gamma^2}{(x_i-\mu)^2+\sigma^2\gamma^2}]$, and hence

$$l_{\sigma\sigma}(\mu,\sigma) = -2\sum_{i=1}^{n}\{(x_i-\mu)^2\sigma^{-4} + \frac{(1+\gamma^2)\gamma^2(x_i-\mu)^2}{[(x_i-\mu)^2+\sigma^2\gamma^2]^2}, \tag{3.3.24}$$

which is always negative. When $l_\sigma = 0$ we can also obtain the following inequality:

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2-\sigma^2}{(x_i-\mu)^2+\sigma^2\gamma^2} = -\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2\gamma^2} < 0. \tag{3.3.25}$$

Using above inequality we can easily see that $l_{\mu\mu}$ is always negative for $\gamma^2 \geq 1$. Thus when $\gamma^2 \geq 1$ the second derivative of the likelihood function will both negative at the solution of the likelihood equations. This condition will ensure the uniqueness of the maximum under the condition that the determined of the second derivative matrix is positive at the solution of the likelihood equations. However, proving that the determinant is positive at the solutions of the likelihood equations is not an easy problem for this case and we leave it as an open problem.

3.4. Examples

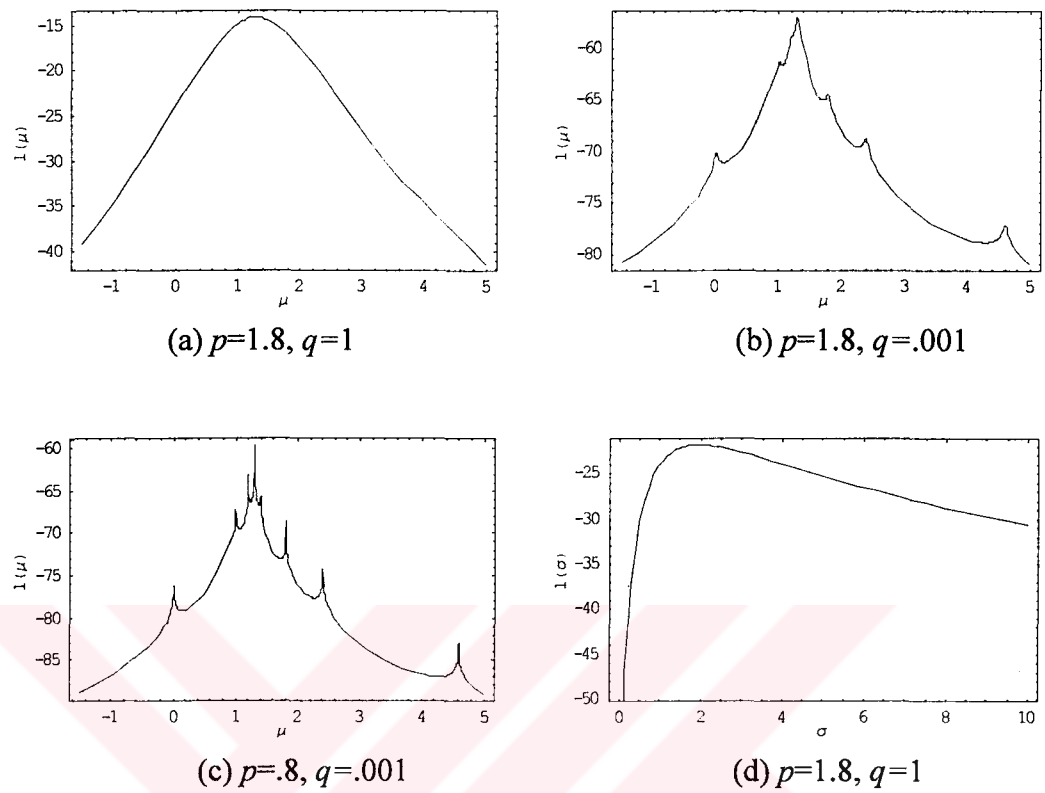
In the following examples we will use some data sets to demonstrate the generalized t based likelihood functions' behavior for different values of the shape parameters.

Example 3.1. The first one is Cushny and Peebles (1905) data set. This is a commonly used real data set in the literature (e.g. Hampel et al. (1986)). The data

represent the prolongation of sleep by means of two drugs. The differences between drug effects for ten objects are 0.0, 0.8, 1.0, 1.2, 1.3, 1.3, 1.4, 1.8, 2.4, 4.6.

Figures 3.1(a)-3.1(d) show the plots of the location-only likelihood functions for the different values of the shape parameters. For the values of p=1.8 and q=1 the likelihood function has a unique maximum point occurring at $\hat{\mu}$=1.29 (See Chapter 5 and Figure 3.1(a)).

For the same data set if we keep the p=1.8 and set q=0.001 we can see that the likelihood function has multiple maximum points in the parameter space (See Figure 3.1(b)). When we set p=0.8 (the case $p \leq 1$) the likelihood function has cusps at each observation (See Figure 3.1(c)). If $m/n < pq/(pq+1)$ is hold,(where n= sample size=10, m= number of coincident observations =2) regardless of the values of the shape parameter we can see that the scale likelihood function is unimodal (See Figure 3.1(d)). One can easily verify that when p=1.8 and q=1, we have $m/n = 0.2 < pq/(pq+1) = 0.6$. The maximum is attained at $\hat{\sigma} = 2.18$ (See Chapter 5 and Figure 3.1(d)).



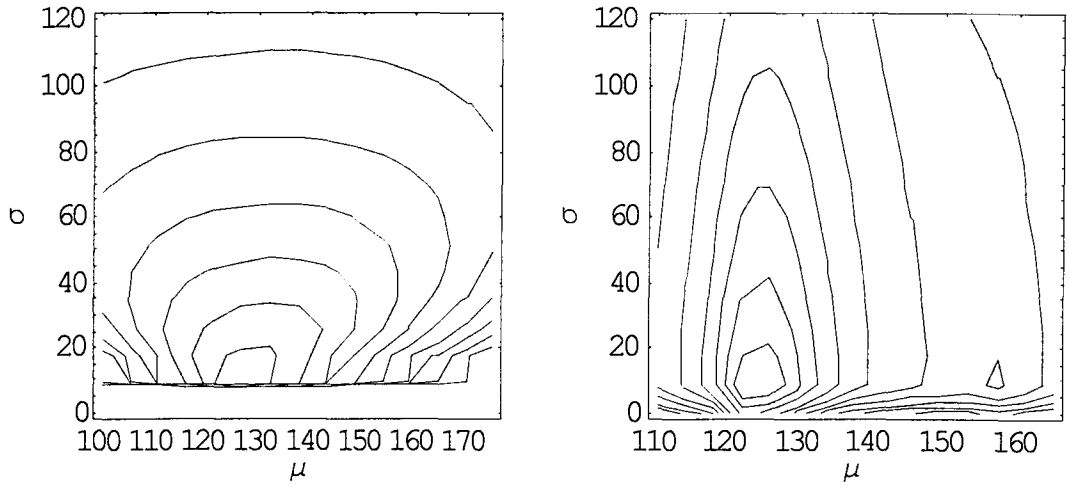
(a) p=1.8, q=1 (b) p=1.8, q=.001

(c) p=.8, q=.001 (d) p=1.8, q=1

Figure 3.1. Behaviors of the *GT* likelihood functions for Cushny and Peebles Data.

Example 3.2. In this example we use the Derby Data from Petruccelli et al. (1999). The data set consists of Kentucky Derby winning times (in seconds) between the years 1875 and 1980. In this data set the existence of two densely scattered observations groups at different locations catches one's eye.

When p=2.5 and q=1 the likelihood function has a unique maximum at $(\hat{\mu}, \hat{\sigma}) = (125.69, 7.42)$ (See Chapter 5 and Figure 3.2(a)). However when we choose smaller values of the shape parameters, for example when p=1.2, q=.05, the likelihood function has two local maximum points in the parameter space (See Figure 3.2(b)). The maximum points occur at $(\hat{\mu}, \hat{\sigma}) = (125, .70)$ and $(\hat{\mu}, \hat{\sigma}) = (157.50, 3.69)$, respectively (See Chapter 5).

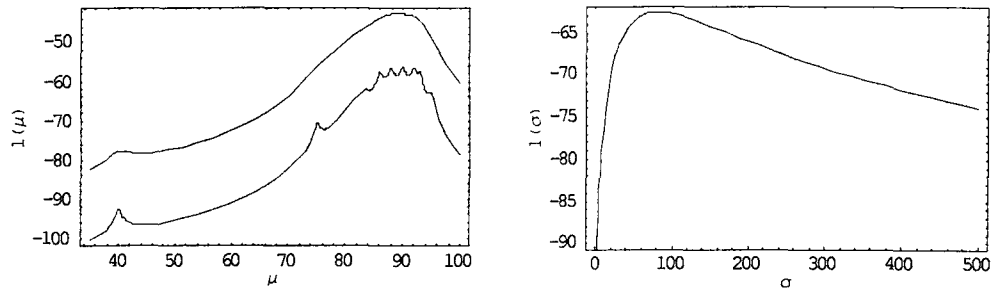


(a) p=2.5, q=1 (b) p=1.2, q=.05

Figure 3.2. Contour plots of the *GT* likelihood function for the location and scale parameters on Derby Data.

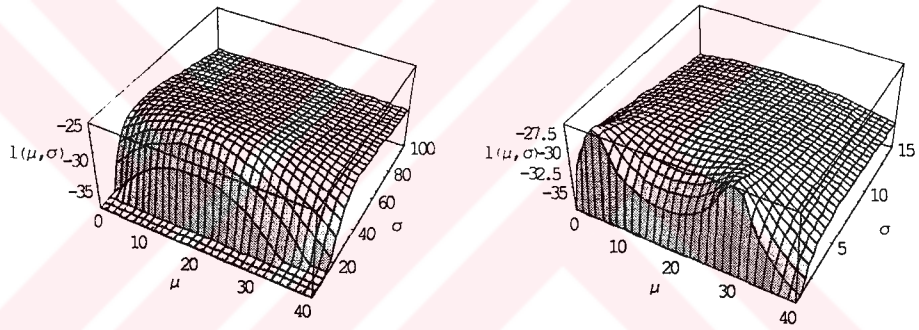
Example 3.3 In this example we investigate the behavior of the likelihood functions for the *GST* distribution on the Rosner Data (1975), {90, 93, 86, 92, 95, 83, 75, 40, 88, 80} (See Chap. 5). The graphs of the likelihood function for the location parameter are in the following Figure 3.3 (a). The upper graph is for p=1 and k=2 and the lower one is for p=1, k=.05. Note that we get multimodality of the likelihood function choosing k sufficiently small. In Figure 3.3 (b) we see a graph of unimodal scale likelihood function for p=k=1.

To see the behavior of the likelihood function well for location and scale we use a simple artificial data set containing two subgroups, i.e. {2, 4, 4, 27, 29, 30}. In Figure 3.4 (a) the likelihood surface is unimodal for p=5 and k=3. However, when we choose p less than 1 we can catch the multimodality, as in Figure 3.4 (b). In that plot we chose p=.7 and k=3.



(a) **(b)**

Figure 3.3: Location-only and scale-only *GST* likelihood functions for the dataset given in Example 3.3.



(a) **(b)**

Figure 3.4: Location-Scale *GST* likelihood surfaces for the dataset given in Example 3.3.

4. ROBUSTNESS PROPERTIES OF THE M ESTIMATORS BASED ON THE GENERALIZED T DISTRIBUTIONS

In Chapter 3 we obtained and investigated the maximum likelihood estimators for the parameters of the generalized t distributions holding the shape parameters fixed. We have shown that these estimators can be regarded as the M estimators for the location and scale parameters of a univariate dataset. In this chapter we will explore the robustness properties of these M estimators in details. We will give the influence functions and finite sample breakdown points for the M estimators obtained from the three different generalized t distributions. Here we will use the log-density functions of the generalized t distributions as alternative ρ functions used in robust statistical analysis. It should be noted that ρ functions play a key role at the robustness properties of an estimator and they always has a tuning constant to tune the robustness of the estimators.

Our ρ functions, the log-density functions of the generalized t distributions, here have distributional shape parameters and we will show that these parameters can serve as the tuning parameters. We can therefore tune them to obtain estimators with good robustness properties. One may think that the shape parameters should be estimated simultaneously with the location and scale parameters from the data. However, we have shown in Chapter 3 that estimating the shape parameters simultaneously will challenge the robustness of the location and the scale estimators. Therefore, we will hold the shape parameters fixed. In Chapter 5 we will explain how we can find the estimates for the shape parameters alone with the location and the scale parameters in case we want to estimate them from the data. This procedure is called adaptive estimation in general and the literature on the *GT* distribution usually deals with the adaptive estimation for the parameters of the *GT* distribution (See e.g. McDonald and Newey, 1988).

4. 1. Basic Definitions

We will start with the definition of an M estimator introduced by Huber (1964).

Definition 4.1.1. Any estimate $T_n(x_1, x_2, ..., x_n)$ defined by minimization $\min \sum_{i=1}^{n} \rho(x_i; T_n)$ or by an implicit equation $\sum_{i=1}^{n} \psi(x_i; T_n) = 0$, where ρ is an arbitrary function and $\psi(x, \theta) = \frac{\partial}{\partial \theta} \rho(x; \theta)$, is called an *M-estimate* (maximum likelihood type estimate).

Note that $\rho(x; \theta) = -\log f(x; \theta)$ gives the ordinary maximum likelihood estimate. So the maximum likelihood estimators can be regarded as the M-estimators for the parameters of the generalized t distributions.

Definition 4.1.2. Let $x_1, x_2, ..., x_n$ be observed values of the random sample $X_1, X_2, ..., X_n$ from distribution F. Then the function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \le x\}, \; x \in R, \tag{4.1.1}$$

where I is an indicator function, is called the *empirical distribution function.*

The empirical distribution function assigns a probability $1/n$ to each observation. The function $F_n(.)$ is a distribution function when considered as a function of x for fixed sample $X_1, X_2, ..., X_n$. On the other hand $F_n(.)$ is a random variable when considered as a function of the sample for fixed x.

Since we never know the underlying distribution from which our data come, the theoretical construction of robustness concepts ascends on the functionals. A *functional* is a function from the space of distribution functions, including empirical distribution function, to real numbers. We represent estimators as functionals which are applied to empirical distribution function. For example, the mean functional is defined as $T(F) = \int x dF(x)$. Here the integral is of Riemann-Stieltjes type (See Staudte and Sheater, 1990). At the empirical distribution function F_n the mean functional becomes

$$T(F_n)=\int x dF_n(x)=\sum_x x p_{F_n}(x)=\sum_{i=1}^{n} X_i/n=\bar{X}_n. \tag{4.1.2}$$

Definition 4.1.3. A *mixture model* G is a weighted average of two different models; that is, $G=(1-p)F+pH$, where $0<p<1$.

Now consider the mixture model $F_{x,\varepsilon}=(1-\varepsilon)F+\varepsilon\Delta_x$, where a large amount of the observations comes from F with probability $1-\varepsilon$ and the small amount from x with probability ε. Here Δ_x is the distribution of x.

Definition 4.1.4. The *influence function* of the estimator T at the distribution F is defined as

$$IF(x)=\lim_{\varepsilon\to 0^+}\frac{T(F_{x,\varepsilon})-T(F)}{\varepsilon}. \tag{4.1.3}$$

The local robustness property of an estimator is measured with the influence function. An estimator which has a bounded influence function is regarded as locally robust.

A useful equivalent expression to (4.1.3) is given as follows. Let $g(\varepsilon)=T(F_{x,\varepsilon})$. Suppose that the function $g(\varepsilon)$ has a continuous derivative on $[0,\varepsilon_0)$ for some $\varepsilon_0>0$. Then $IF(x)=\lim_{\varepsilon\to 0^+} g'(\varepsilon)$.

For example the influence function of the mean can be obtained as follows:

$$\lim_{\varepsilon\to 0^+}\frac{T(F_{x,\varepsilon})-T(F)}{\varepsilon},$$

$$=\lim_{\varepsilon\to 0^+}\frac{(1-\varepsilon)T(F)+x\varepsilon-T(F)}{\varepsilon},$$

$$=x-T(F),\ x\in R \tag{4.1.4}$$

As we can see from (4.1.4) the influence function for the mean is unbounded. Thus the mean is a non-robust estimator of location.

A sample can be contaminated in many different ways. But in robust statistical analysis the *ε-replacement* type contamination is often used. Replacing any arbitrary *m* observations in the sample by *m* arbitrary values (*m* outliers) produces the *ε-replacement* contamination. New sample will have a fraction of $\varepsilon = \frac{m}{n}$ outliers.

Definition 4.1.5. The *maximum bias* is defined by $b(\varepsilon; X, T) = \sup_{X'} \left| T(X') - T(X) \right|$, where X' denotes the contaminated sample.

Definition 4.1.6. The *breakdown point* of an estimator T is defined by $\varepsilon^*(X,T) = \inf\{\varepsilon : b(\varepsilon; X, T) = \infty\}$.

The breakdown point is the smallest fraction of contamination that can render the estimator meaningless. We require an estimator to have as high as possible breakdown point. The theoretically reachable highest breakdown point is .5.

The breakdown point property is global in the sense that it is independent of the model (distribution) *F*. The breakdown point property is therefore considered as the global robustness property of an estimator.

4.2. Influence Functions of the Generalized t based M-Estimators

In this section we will derive the influence functions of the generalized t based M-estimators.

4.2.1. The *GT* Distribution Case

The estimating equations $\hat{\mu} = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i$ and $\hat{\sigma}^2 = \sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2$ can be rewritten as

$$\sum_{i=1}^{n}\psi(x_i,\theta)/n=0 \tag{4.2.1}$$

where $\theta=(\mu,\sigma)$. $\psi=\begin{pmatrix}\psi_1(x,\theta)\\ \psi_2(x,\theta)\end{pmatrix}$ is the function of $\psi_1(x,\theta)=w(s)(x-\mu)$ and $\psi_2(x,\theta)=w(s)(x-\mu)^2-\sigma^2$, $s=|x-\mu|/\sigma$ and $w(s)=(pq+1)s^{p-2}/(q+s^p)$ is the weight function.

The functional form of the equation given in (4.2.1) at a distribution function F is given as

$$\int\psi(x;T(F))dF(x)=0 \tag{4.2.2}$$

where $T(F)$ will be the parameters of the underlying distribution F. Now if we insert the ε-contaminated model $(1-\varepsilon)F+\varepsilon\delta_y$ into (4.2.2) we get the following equation.

$$\int\psi(x,T((1-\varepsilon)F+\varepsilon\delta_y))d((1-\varepsilon)F+\varepsilon\delta_y))(x)=0 \tag{4.2.3}$$

where $\delta_y(x)=\begin{cases}1 & if\ x\geq y\\ 0 & otherwise\end{cases}$ is the distribution function of point mass 1 at y.

$$(1-\varepsilon)\int\psi(x,T((1-\varepsilon)F+\varepsilon\delta_y))dF(x)+\varepsilon\psi(y,T((1-\varepsilon)F+\varepsilon))=0 \tag{4.2.4}$$

Taking derivative of (4.2.4) with respect to ε we get

$$(-1)\int\psi(x,T((1-\varepsilon)F+\varepsilon\delta_y))dF(x)+(1-\varepsilon)\frac{\partial}{\partial\varepsilon}\left[\int\psi(x,T((1-\varepsilon)F+\varepsilon\delta_y))dF(x)\right]$$
$$+\psi(y,T((1-\varepsilon)F+\varepsilon)+\varepsilon\frac{\partial}{\partial\varepsilon}\psi(y,T((1-\varepsilon)F+\varepsilon))=0 \tag{4.2.5}$$

Setting $\varepsilon=0$, (4.2.5) becomes

$$(-1)\int \psi(x,T(F))dF(x)+\frac{\partial}{\partial\varepsilon}\left[\int \psi(x,T((1-\varepsilon)F+\varepsilon\delta_Y))dF(x)\right]_{\varepsilon=0}+\psi(y,T(F))=0 \tag{4.2.6}$$

Since $\int \psi(x,T(F))dF(x)=0$, (4.2.6) reduces

$$\int (\frac{d}{d\theta}\psi(x,\theta)\Big|_{T(F)})\frac{dT}{d\varepsilon}\Big|_{\varepsilon=0}dF(x)+\psi(y,T(F))=0 \tag{4.2.7}$$

$$IF(y;T,F)=\frac{dT}{d\varepsilon}\Big|_{\varepsilon=0}=\frac{-\psi(y,T(F))}{\int (\frac{d}{d\theta}\psi(x,\theta))\Big|_{T(F)}dF(x)} \tag{4.2.8}$$

Let $\Omega=\int (\frac{d}{d\theta}\psi(x,\theta))\Big|_{T(F)}dF(x)=E_F[\frac{d}{d\theta}\psi(x,\theta)]$. Then

$$\Omega=\begin{bmatrix}\int (\frac{d}{d\mu}\psi_1(x,\theta))\Big|_{T(F)}dF(x) & \int (\frac{d}{d\mu}\psi_2(x,\theta))\Big|_{T(F)}dF(x)\\ \int (\frac{d}{d\sigma}\psi_1(x,\theta))\Big|_{T(F)}dF(x) & \int (\frac{d}{d\sigma}\psi_2(x,\theta))\Big|_{T(F)}dF(x)\end{bmatrix} \tag{4.2.9}$$

4.2.2. The *GST* Distribution Case

Since the estimating equations involving the *GST* distribution are $\hat{\mu}=\sum_{i=1}^{n}w_ix_i/\sum_{i=1}^{n}w_i$ and $\hat{\sigma}^2=(1/n)\sum_{i=1}^{n}w_i(x_i-\hat{\mu})^2$, they can be rewritten similarly as in (4.2.1). In this case we have the weight function $w(s)=\frac{2p}{k+s^2}$, where $s=\frac{x-\mu}{\sigma}$. As a result, we have the same form of the influence function as in (4.2.8).

4.2.3. The *GET* Distribution Case

The form of the *GET* distribution we have used in Chapter 3 has the estimating equations $\hat{\mu} = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i$ and $\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2$, where the weight function is $w(s) = 1 + \frac{1+\gamma^2}{\gamma^2 + s^2}$, in which $s = \frac{x-\mu}{\sigma}$. Thus we will have the same form of influence function as in (4.2.8).

4.3. Influence Functions of the Generalized t Based M-Estimators When the Underlying Distribution Belongs to the Spherical Family

Definition 4.3.1. If a random vector $\mathbf{X}$ has the density of the form $g(\| \mathbf{x} \|)$ for some function $g(u) \geq 0$, $u \geq 0$, where $\| \mathbf{x} \| = (\mathbf{x}'\mathbf{x})^{1/2}$ then $\mathbf{X}$ is said to have a *spherical distribution.*

Spherical family is a subfamily of the elliptically contoured family of distributions. The density of the elliptically contoured family has the form

$$| \Sigma |^{-1/2} \, g((\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu)), \tag{4.3.1}$$

for some function $g(u) \geq 0$, $u \geq 0$ where $\mathbf{x} \in R^n$, μ is the location vector and Σ is a positive definite dispersion matrix. The multivariate normal and Student's t and the contaminated normal distributions are typical examples of the spherical distributions. Spherical distributions are obtained for $\mu = 0$ and $\Sigma = \alpha \mathbf{I}$ in (4.3.1), where $\mathbf{I}$ is the identity matrix and α is a constant. So for a spherical distributed random variable $\mathbf{X}$ we have $E(\mathbf{X}) = 0$ and $Cov(\mathbf{X}) = \alpha \mathbf{I}$. The contours of a spherical distribution consist of hyperspheres.

Theorem 4.3.1. The influence function of the *GT*-based M-estimators at the spherical distribution F_s is given by

$$IF(y;\mu,F_s)=-\frac{1}{a}\frac{(pq+1)|y|^{p-2}y}{q+|y|^p}, \tag{4.3.2}$$

$$IF(y;\sigma,F_s)=-\frac{1}{b}[\frac{(pq+1)|y|^p}{q+|y|^p}-1], \tag{4.3.3}$$

where

$$a=E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^{p-2}-2|x|^{2p-2}}{(q+|x|^p)^2}]\}-E_{F_s}[\frac{(pq+1)|x|^{p-2}}{q+|x|^p}], \tag{4.3.4}$$

$$b=E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^p-2|x|^{2p}}{(q+|x|^p)^2}]\}-2. \tag{4.3.5}$$

Proof. We will first compute the matrix Ω in (4.2.9) at F_s.

$$\frac{\partial\psi_1}{\partial\mu}\Big|_{\substack{\mu=0\\\sigma=1}}=[\frac{\partial w(s)}{\partial s}\frac{\partial s}{\partial\mu}(x-\mu)-w(s)]\Big|_{\substack{\mu=0\\\sigma=1}}$$

$$=-(pq+1)[\frac{(p-2)q|x|^{p-2}-2|x|^{2p-2}}{(q+|x|^p)^2}]-\frac{(pq+1)|x|^{p-2}}{q+|x|^p}, \tag{4.3.6}$$

$$\int[\frac{\partial\psi_1(x;\mu,\sigma)}{\partial\mu}]\Big|_{\substack{\mu=0\\\sigma=1}}dF_s=$$

$$E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^{p-2}-2|x|^{2p-2}}{(q+|x|^p)^2}\}-E_{F_s}[\frac{(pq+1)|x|^{p-2}}{q+|x|^p}] \tag{4.3.7}$$

$$\frac{\partial\psi_2}{\partial\sigma}\Big|_{\substack{\mu=0\\\sigma=1}}=[\frac{\partial w(s)}{\partial s}\frac{\partial s}{\partial\sigma}(x-\mu)^2-2\sigma]\Big|_{\substack{\mu=0\\\sigma=1}}$$

$$= -(pq+1)[\frac{(p-2)q|x|^p - 2|x|^{2p}}{(q+|x|^p)^2}] - 2, \quad (4.3.8)$$

$$\int [\frac{\partial \psi_2(x;\mu,\sigma)}{\partial \sigma}]\Big|_{\substack{\mu=0\\ \sigma=1}} dF_s = E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^p - 2|x|^{2p}}{(q+|x|^p)^2}]\} - 2 \quad (4.3.9)$$

$$\frac{\partial \psi_2}{\partial \mu}\Big|_{\substack{\mu=0\\ \sigma=1}} = [\frac{\partial w(s)}{\partial s}\frac{\partial s}{\partial \mu}(x-\mu)^2 - 2w(s)(x-\mu)]\Big|_{\substack{\mu=0\\ \sigma=1}}$$

$$= -(pq+1)[\frac{(p-2)q|x|^{p-2} - 2|x|^{2p-2}}{(q+|x|^p)^2}]x - -(pq+1)\frac{2|x|^{p-2}x}{q+|x|^p}, \quad (4.3.10)$$

is an odd function of x. Thus $\int [\frac{\partial \psi_2(x;\mu,\sigma)}{\partial \mu}]\Big|_{\substack{\mu=0\\ \sigma=1}} dF_s = 0$. Also,

$$\frac{\partial \psi_1}{\partial \sigma}\Big|_{\substack{\mu=0\\ \sigma=1}} = [\frac{\partial w}{\partial s}\frac{\partial s}{\partial \sigma}(x-\mu)]\Big|_{\substack{\mu=0\\ \sigma=1}}$$

$$= -(pq+1)[\frac{(p-2)q|x|^{p-2} - 2|x|^{2p-2}}{(q+|x|^p)^2}]x, \quad (4.3.11)$$

is an odd function of x and thus in $\int [\frac{\partial \psi_1(x;\mu,\sigma)}{\partial \sigma}]\Big|_{\substack{\mu=0\\ \sigma=1}} dF_s = 0$.

Therefore the matrix Ω has the form $\Omega = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, where

$$a = E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^{p-2} - 2|x|^{2p-2}}{(q+|x|^p)^2}]\} - E_{F_s}[\frac{(pq+1)|x|^{p-2}}{q+|x|^p}], \quad (4.3.12)$$

$$b = E_{F_s}\{-(pq+1)[\frac{(p-2)q|x|^p - 2|x|^{2p}}{(q+|x|^p)^2}]\} - 2. \quad (4.3.13)$$

The influence function, therefore, becomes

$$IF(y;T,F_s) = -\Omega^{-1}\psi(y;0,1)$$

$$= -\begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix} \begin{bmatrix} \dfrac{(pq+1)|y|^{p-2}y}{q+|y|^p} \\ \dfrac{(pq+1)|y|^p}{q+|y|^p} - 1 \end{bmatrix}$$

$$= \begin{bmatrix} -\dfrac{1}{a}\dfrac{(pq+1)|y|^{p-2}y}{q+|y|^p} \\ -\dfrac{1}{b}[\dfrac{(pq+1)|y|^p}{q+|y|^p} - 1] \end{bmatrix}. \qquad (4.3.14)$$

□

We note that the first component of (4.3.14) is a constant times the function $\psi_1(y,\theta)$. Similarly, the second component is a constant times $\psi_2(y,\theta)$. We also note that

i) The influence function for the scale $IF(y;\sigma^2,F_s) = -\dfrac{1}{b}[\dfrac{(pq+1)|y|^p}{q+|y|^p} - 1]$ is bounded.

ii) The influence function for the location $IF(y;\mu,F_s) = -\dfrac{1}{a}\dfrac{(pq+1)|y|^{p-2}y}{q+|y|^p}$ is redescending, i.e. $\lim_{|y|\to\infty} IF(y;\mu,F_s) = 0$.

iii) $|IF(y;\mu,F_s)|$ has a maximum at $y^* = [q(p-1)]^{1/p}$, $p>1$ and is increasing over $(0,y^*)$ and is decreasing over (y^*,∞).

The following Theorems 4.3.2 and 4.3.3 give the influence functions for the *GST* and *GET* –based M-estimators of location and scale in the spherical family of

distributions. Their proofs are similar to that of Theorem 4.3.1 and are therefore omitted.

Theorem 4.3.2. The influence function of the *GST*-based M-estimators at the spherical distribution F_s is given by

$$IF(y;\mu,F_s)=-\frac{1}{a}\frac{2py}{k+y^2}, \tag{4.3.15}$$

$$IF(y;\sigma,F_s)=-\frac{1}{b}[\frac{2py^2}{k+y^2}-1], \tag{4.3.16}$$

where

$$a=E_{F_s}[2p\frac{x^2-k}{x^2+k}], \tag{4.3.17}$$

$$b=E_{F_s}[4p\frac{x^4}{(x^2+k)^2}]-2. \tag{4.3.18}$$

Remark 4.3.1. Note that the influence function for the scale is bounded and the influence function for the location is redescending. Thus the *GST* distribution has good local robustness properties.

Theorem 4.3.3. The influence function of the *GET*-based M-estimators at the spherical distribution F_s is given by

$$IF(y;\mu,F_s)=-\frac{1}{a}(1+\frac{1+\gamma^2}{\gamma^2+y^2})y, \tag{4.3.19}$$

$$IF(y;\sigma,F_s)=-\frac{1}{b}[(1+\frac{1+\gamma^2}{\gamma^2+y^2})y^2-1], \tag{4.3.20}$$

where

$$a = E_{F_s}[\frac{(1+\gamma^2)2x^2}{(\gamma^2+x^2)^2} - 1 - \frac{1+\gamma^2}{\gamma^2+x^2}], \tag{4.3.21}$$

$$b = E_{F_s}[\frac{2(1+\gamma^2)x^4}{(\gamma^2+x^2)^2}] - 2. \tag{4.3.22}$$

Remark 4.3.2. We note that the influence functions for both location and scale are unbounded. Thus the location and scale estimators found from the *GET* distribution do not have good local robustness properties.

4.4. Breakdown Point Properties of the Generalized T Based M-Estimators

4.4.1. The *GT* Distribution Case

Huber (1984) gave sufficient conditions for an M-estimator of location to have a breakdown point .5. We verified these conditions for the *GT*-based M-estimator of location:

We have

$$\rho(x) = (q+1/p)\log(q+|x|^p), \tag{4.4.1}$$

and thus the ψ-function is

$$\psi(x) = \frac{(pq+1)|x|^{p-1} sign(x)}{q+|x|^p} \tag{4.4.2}$$

1) Since $\rho(-x) = \rho(x)$ ρ is a symmetric function.

2) Since $\lim_{|x|\to\infty} \rho(x) = \infty$ ρ is an unbounded function.

3) $\lim_{|x|\to\infty} \frac{\rho(x)}{|x|} = 0$.

4) ψ is an odd continuous function.

5) $\lim_{x\to 0}\psi(x)=0$ for finite p>1 and q.

6) Since $\lim_{|x|\to\infty}\psi(x)=0$ ψ redescends.

7) The ψ-function has a maximum at $x^*=[(p-1)q]^{1/p}$, p>1 and this maximum is

$$\psi(x^*)=\frac{(q+1/p)(p-1)^{(p-1)/p}}{q^{1/p}}.$$

Thus the *GT*-based M-estimator of location has breakdown point .5 by Theorem 4.1. of Huber (1984).

We can also find the breakdown point for the scale estimator. If the contamination takes the value of the estimator to the boundaries of the parameter space, i.e. to ∞ or to 0, then the estimator breaks down. These kinds of breakdown are called explosion and implosion respectively (Huber, 1981).

We showed in Chapter 3 that a necessary and sufficient condition for the existence of *GT*-based M-estimator for scale in the parameter space is $\frac{m}{n}<\frac{pq}{pq+1}$. This means that the fraction of contamination must not greater than $\frac{pq}{pq+1}$.Otherwise, the explosion will occur when $\frac{m}{n}>\frac{pq}{pq+1}$.

Tyler (1986) showed that the breakdown point of a scatter M-estimator is $\varepsilon^*=\min\{\frac{1}{K_2},1-\frac{p'}{K_2}\}$, where $K_2=\sup\{w(s')s'\}$ and p' is the dimension of the sample space. In the univariate *GT* case we have $p'=1$ and $K_2=\sup\{w(s)s^2\}=\sup\{\frac{(pq+1)s^p}{q+s^p}\}=pq+1$. Thus,

$$\varepsilon^*=\min\{\frac{1}{pq+1},\frac{pq}{pq+1}\}=\begin{cases}\frac{1}{pq+1} & ,if\quad pq>1\\ \frac{pq}{pq+1} & ,if\quad pq<1\\ 1/2 & ,if\quad pq=1\end{cases} \qquad (4.4.3)$$

When $pq=1$ the joint breakdown point for the location-scale estimator is .5. When $pq>1$ since the breakdown point for the scale i.e. $1/(pq+1)$, is less than that of the location, i.e. .5, so the scale estimator first breaks down. Thus the joint breakdown point is $\frac{1}{pq+1}$. When $pq<1$ the breakdown value for the scale is $\frac{pq}{pq+1}$ and it is less than that of the location. Therefore in this case the joint breakdown point is $\frac{pq}{pq+1}$.

4.4.2. The *GST* Distribution Case

For the *GST* distribution we have

$$\rho(x) = p\log(k+x^2), \tag{4.4.4}$$

and so the ψ-function is

$$\psi(x) = \frac{2xp}{k+x^2}. \tag{4.4.5}$$

It is easily verified that the functions (4.4.4) and (4.4.5) satisfy all the sufficient conditions given in Theorem 4.1 of Huber (1984). Thus the *GST*-based M-estimator of location has breakdown point .5.

In this case the explosion will occur when $\frac{m}{n} > \frac{2p-1}{2p}$ for the scale estimator. We have

$$K_2 = \sup\{w(s)s^2\} \tag{4.4.6}$$

$$= \sup\{\frac{2ps^2}{k+s^2}\} \tag{4.4.7}$$

$$=2p, \tag{4.4.8}$$

and the breakdown for the scale estimator is

$$\varepsilon^* = \min\{\frac{1}{2p}, \frac{2p-1}{2p}\} \tag{4.4.9}$$

$$= \begin{cases} \frac{1}{2p}, & if \quad p > 1 \\ \frac{2p-1}{2p}, & if \quad p < 1 \\ 1/2, & if \quad p = 1 \end{cases} \tag{4.4.10}$$

When $p=1$ the joint breakdown point for the location-scale estimator is .5. When $p>1$ since the breakdown point for the scale, $\frac{1}{2p}$, is less than that of location, i.e. .5, in this case the scale estimator first breaks down. Thus the joint breakdown point is $\frac{1}{2p}$. When $p<1$ the breakdown value for the scale is $\frac{2p-1}{2p}$ and it is less than that of the location. Therefore in this case the joint breakdown point is $\frac{2p-1}{2p}$. Note that we should not have $p \leq 1/2$.

4.4.3. The *GET* Distribution Case

For the *GET* distribution we have

$$\rho(x) = (\frac{1+\gamma^2}{2})\log(\gamma^2 + x^2) + \frac{1}{2}x^2, \tag{4.4.11}$$

and so the ψ-function is

$$\psi(x) = \frac{(1+\gamma^2)x}{\gamma^2 + x^2} + x. \tag{4.4.12}$$

The functions (4.4.11) and (4.4.12) do not satisfy all the sufficient conditions to have breakdown point of .5 for location estimator by Theorem 4.1 of Huber (1984). For example, $\lim_{|x|\to\infty} \frac{\rho(x)}{|x|} \neq 0$ and $\lim_{|x|\to\infty} \psi(x) = \infty$. Also, we can show that the location estimators obtained from the *GET* distribution is a function of the sample mean, so we cannot expect that the location estimator can have high breakdown point. In fact, theoretical breakdown point will be zero because of the sample mean (Chapter 3, Equation 3.3.5).

For the scale case we have

$$K_2 = \sup\{w(s)s^2\} \tag{4.4.13}$$

$$= \sup\{(1 + \frac{1+\gamma^2}{\gamma^2 + s^2})s^2\} \tag{4.4.14}$$

$$= \infty, \tag{4.4.15}$$

and thus $\varepsilon^* = 0$, from which it follows that the scale estimator from the *GET* distribution is not robust. Also, we can show that the scale estimator for the *GET* distribution is a function of the sample variance, and it is well known that the breakdown point of the sample variance is zero, and hence the scale estimators based on the *GET* distribution is inheriting the breakdown point of the sample variance. This is the result of the *GET* density having a term associated with the normal distribution.

As we have seen in Section 4.3 the influence functions for both location and scale estimators are unbounded. Thus the use of the *GET* distribution as a robust alternative is skeptical. However, it is known that (See e.g. Fong, 1997) the *GET* distribution can be used as an alternative model for the financial datasets.

5. ALGORITHMS TO COMPUTE THE ESTIMATES FOR THE PARAMETERS OF THE GENERALIZED T DISTRIBUTIONS

In this chapter we will give an iteratively reweighting algorithm to compute the estimates for the parameters of the generalized t distributions. We will establish the connection between the iterative reweighting algorithm and EM (Expectation and Maximization) algorithm. We will first start with the iteratively reweighting algorithm to compute the *GT*-based M-estimates. In literature it is often stated that an iterative reweighting algorithm can be used to compute the estimates for the parameters of the *GT* distribution (e.g., McDonald and Newey, 1988) but connection between the suggested iterative reweighting algorithm and the EM algorithm is not known. The main aim of this chapter is to show that this iteratively reweighting algorithm is in fact an EM algorithm.

5.1. Iteratively Reweighting (IR) Algorithm for the *GT* Distribution

Recall that the estimating equations for the location and the scale parameters of the *GT* distribution with known shape parameters are

$$\hat{\mu} = \sum_{i=1}^{n} w_i x_i / \sum_{i=1}^{n} w_i \quad \text{and} \quad \hat{\sigma}^2 = \sum_{i=1}^{n} w_i (x_i - \hat{\mu})^2 / n, \tag{5.1.1}$$

where

$$w_i = \frac{(pq+1)\left(\frac{|x_i - \hat{\mu}|}{\hat{\sigma}}\right)^{p-2}}{q + \left(\frac{|x_i - \hat{\mu}|}{\hat{\sigma}}\right)^{p}}. \tag{5.1.2}$$

Since the weights depend on the parameters to be estimated the estimating equations cannot be solved explicitly. Therefore, some numerical methods such as Newton-Raphson, are needed to find an approximate solution to the estimating equations. However, the nature of the estimating equations suggest an updating equation that leads us for using an iteratively reweighting algorithm. The estimating equations given above suggest the following IR algorithm.

$$\hat{\mu}^{(m+1)} = \sum_{i=1}^{n} w_i^{(m)} x_i \Big/ \sum_{i=1}^{n} w_i^{(m)} \tag{5.1.3}$$

$$\hat{\sigma}^{2(m+1)} = \sum_{i=1}^{n} w_i^{(m)} (x_i - \hat{\mu}^{(m+1)})^2 / n \tag{5.1.4}$$

where,

$$w_i^{(m)} = \frac{(pq+1)\left(\dfrac{\left|x_i - \hat{\mu}^{(m)}\right|}{\hat{\sigma}^{(m)}}\right)^{p-2}}{q + \left(\dfrac{\left|x_i - \hat{\mu}^{(m)}\right|}{\hat{\sigma}^{(m)}}\right)^{p}}, \quad m = 0,1,2,.... \tag{5.1.5}$$

At each iteration we substitute the current estimates for μ and σ^2 into the weight function (5.1.5) to find the new weights. Then we keep these weights fixed, and compute a new estimate for μ from the updating equation (5.1.3). Next, substituting the present weights and the new estimate for μ into the updating equation (5.1.4) we find a new estimate for σ^2.

5.2. Relation Between the IR and EM Algorithms for the *GT* Distribution

The EM algorithm is an iterative procedure for maximizing a likelihood function. Dempster, Laird and Rubin (1977) introduced this algorithm for computing maximum likelihood estimates from incomplete data. (The data which we observe is called *incomplete data* and they together with missing data is called the *complete data*). They termed EM because each iteration in this algorithm consists of two steps called the expectation step (E-step) and the maximization step (M-step).

The main concept in the theory of the EM algorithm is missing data. Data can be missing in the ordinary sense of a failure to record certain observations on certain cases. Data can also be missing in a theoretical sense.

In the E-step of the algorithm the conditional expectations of the missing data are found given the observed data and the current estimates of the parameters. These expectations are then substituted for the missing data and used to complete the data.

In the M-step, maximum likelihood estimation of the parameters is performed in the usual manner using the completed data. The estimated parameters are then used to reestimate the missing data, which in turn lead to new parameter estimates. The procedure can be carried out until convergence is achieved.

Theorem 5.1. The iterative reweighting algorithm to compute the *GT* based M-estimates

$\hat{\mu}^{(m+1)} = \sum_{i=1}^{n} w_i^{(m)} x_i / \sum_{i=1}^{n} w_i^{(m)}$, $\hat{\sigma}^{2(m+1)} = \sum_{i=1}^{n} w_i^{(m)} (x_i - \hat{\mu}^{(m+1)})^2 / n$, where

$$w_i^{(m)} = \frac{(pq+1)\left(\frac{\left|x_i - \hat{\mu}^{(m)}\right|}{\hat{\sigma}^{(m)}}\right)^{p-2}}{q + \left(\frac{\left|x_i - \hat{\mu}^{(m)}\right|}{\hat{\sigma}^{(m)}}\right)^{p}}, \quad m = 0,1,2,\ldots$$ can be identified as an EM algorithm.

Proof. Since we observe X directly, the random variable T in $X = \mu + q^{1/p}\sigma U T^{-1/2}$ can be regarded as missing information. Thus the observations $x_1, x_2, \ldots, x_n$ can be viewed as incomplete data with the missing components $t_1, t_2, \ldots, t_n$. The combined data (x_i, t_i) will be regarded as the complete data with the joint density function

$$f_c(x,t;\mu,\sigma,p,q) = \frac{p^2}{4\Gamma(1/p)\Gamma(q)q^{1/p}\sigma} e^{-\frac{t^{\frac{p}{2}}}{2}(1+|r|^p)} t^{\frac{pq-1}{2}}, \qquad (5.2.1)$$

where $r = \frac{x-\mu}{q^{1/p}\sigma}$. To perform the E-step of the algorithm we need the complete data log-likelihood function which is

$$l_c(\mu,\sigma) = \log \prod_{i=1}^{n} f_c(x_i, t_i; \mu, \sigma, p, q)$$

$$= -n\log\sigma - \sum_{i=1}^{n} t_i^{p/2}(1+|r_i|^p) + \frac{pq-1}{2}\sum_{i=1}^{n}\log t_i$$

$$= -n\log\sigma - \sum_{i=1}^{n} t_i^{p/2}|r_i|^p + \sum_{i=1}^{n}\left\{\frac{pq-1}{2}\log t_i - t_i^{p/2}\right\} \qquad (5.2.2)$$

Now the conditional expectation of (5.2.2) given $x, \mu^{(m)}$ and $\sigma^{2(m)}$ will give the E-step of the EM algorithm. Further, to find the conditional expectation we need the conditional density function of t given r. After some straightforward calculation the conditional density of t given r can be obtained as follows:

$$h(t \mid r) = \frac{f_c(x,t;\mu,\sigma^2,p,q)}{f(x;\mu,\sigma^2,p,q)}$$

$$= \frac{pe^{-t^{p/2}(1+|r|^p)} t^{\frac{pq-1}{2}} (1+|r|^p)^{\frac{pq+1}{p}}}{2\Gamma(\frac{pq+1}{p})} \qquad (5.2.3)$$

Now we can find the conditional expectation of the complete data log likelihood function. Since the last term does not involve μ and σ it can be ignored. Then the conditional expectation is

$$E[l_c(\mu,\sigma) \mid \mu^{(m)},\sigma^{(m)}] = -n\log\sigma - \sum_{i=1}^{n} E[t_i^{p/2} \mid r_i]\left|\frac{x_i-\mu}{\sigma q^{1/p}}\right|^p, \qquad (5.2.4)$$

where $E(t_i^{p/2} \mid r_i) = \frac{q+1/p}{1+|r_i|^p}$.

To perform the M step of the algorithm we have to maximize this function with respect to μ and σ, but we can see that finding the maximum points of this function is not an easy task. To perform the M step of the algorithm we will use the method given in Lange and Sinsheimer (1993) and Arslan (2002b). Let us construct the following function:

$$Q(\mu,\sigma \mid \mu^{(m)},\sigma^{(m)}) = -\log\sigma - \frac{1}{n}\sum_{i=1}^{n} w_i^{(m)} \frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2 \tag{5.2.5}$$

where $w_i^{(m)}$ are given in (5.1.5), and they are computed at the current estimate of μ and σ. This function is a quadratic function and it has a unique maximum point. Now we can perform the M-step. It can be easily seen that this function and the conditional expectation of the complete data likelihood function will have the same stationary points.

In the M-step of the algorithm, we maximize $Q(\mu,\sigma^2 \mid \mu^{(m)},\sigma^{2(m)})$ with respect to μ and σ. The solution of the equations

$$\frac{\partial Q}{\partial \mu} = \frac{1}{n}\sum_{i=1}^{n} w_i^{(m)}(x-\mu)\sigma^{-2} = 0, \tag{5.2.6}$$

$$\frac{\partial Q}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{n}\sum_{i=1}^{n} w_i^{(m)}(x-\mu)^2\sigma^{-3} = 0 \tag{5.2.7}$$

are

$\hat{\mu}^{(m+1)} = \sum_{i=1}^{n} w_i^{(m)} x_i / \sum_{i=1}^{n} w_i^{(m)}$ and $\hat{\sigma}^{2(m+1)} = \sum_{i=1}^{n} w_i^{(m)}(x_i - \hat{\mu}^{(m+1)})^2 / n$, respectively.

These solutions are the same equations given in (5.1.3) and (5.1.4). Therefore we have obtained that the IR algorithm derived from the *GT* estimating equations is an EM algorithm.

5.3. Adaptive Robust Estimation for the *GT* Distribution

In this section we will consider the estimation problem of the *GT* distribution under the assumption that the shape parameters are unknown. This estimation problem will lead us to the adaptive estimation procedure in which the shape parameter, mainly the distribution, will also be estimated from the data.

A number of procedures can be proposed to obtain estimates, but a general approach to estimate the four parameters can be described as fallows: First estimate

the shape parameters p and q, and then use the estimates for p and q to find the estimates for μ and σ.

One can use the maximum likelihood estimating equations for the four parameters of the *GT* distribution. This procedure will be an iterative procedure. To start the procedure we have to start some initial estimates for the parameters. These initial estimates will be inserted in the maximum likelihood estimating equations to obtain new estimates for the shape parameters. Then using the new estimates for the shape parameters one can find the new estimates for the location and scale parameters from the maximum likelihood equations given in (3.1.2) and (3.1.3). This procedure will be repeated until convergence.

Another procedure to find the estimates for the shape parameter is to use the moments of the *GT* distribution. Recall the even central moments of the *GT* distribution

$$\mu_j = E[(X-\mu)^j] = \frac{\sigma^j q^{j/p}\Gamma((j+1)/p)\Gamma(q-j/p)}{\Gamma(1/p)\Gamma(q)}, \tag{5.3.1}$$

for j even. The population kurtosis is given by

$$\frac{\mu_4}{\mu_2^2} = \frac{\Gamma(5/p)\Gamma(q-4/p)\Gamma(1/p)\Gamma(q)}{\Gamma^2(3/p)\Gamma^2(q-2/p)} \tag{5.3.2}$$

A moment estimator of p and of q can be obtained by solving the following equations simultaneously for p and q :

$$\frac{\Gamma(5/p)\Gamma(q-4/p)\Gamma(1/p)\Gamma(q)}{\Gamma^2(3/p)\Gamma^2(q-2/p)} = \frac{n\sum_{i=1}^{n}(x_i-\bar{x})^4}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^2} \tag{5.3.3}$$

$$\frac{\Gamma(9/p)\Gamma(q-8/p)\Gamma(1/p)\Gamma(q)}{\Gamma^2(5/p)\Gamma^2(q-4/p)} = \frac{n\sum_{i=1}^{n}(x_i-\bar{x})^8}{[\sum_{i=1}^{n}(x_i-\bar{x})^4]^2} \tag{5.3.4}$$

where the right hand sides of (5.3.3) and (5.3.4) involve the sample moments. Now we can choose the values of p and q which are solutions to (5.3.3) and (5.3.4), and use them in the likelihood equations (3.1.4) and (3.1.5) to find the estimates for μ and σ.

5.4. IR Algorithm for the *GST* Distribution and Its Relation with the EM Algorithm

In this section we will give the IR algorithm for the *GST* distribution and its relation with the EM algorithm which have already been given in the literature (e.g., see Dempster, Laird and Rubin, 1980). From the estimating equations we can formulate the following IR algorithm

$$\hat{\mu}^{(m+1)} = \sum_{i=1}^{n} w_i^{(m)} x_i / \sum_{i=1}^{n} w_i^{(m)} \tag{5.4.1}$$

$$\hat{\sigma}^{2(m+1)} = \sum_{i=1}^{n} w_i^{(m)} (x_i - \hat{\mu}^{(m+1)})^2 / n \tag{5.4.2}$$

$$w_i^{(m)} = \frac{2p}{k + (\frac{x_i - \hat{\mu}^{(m)}}{\hat{\sigma}^{(m)}})^2} \tag{5.4.3}$$

for $m = 0,1,2,....$

Let $x_1, x_2, ..., x_n$ be a data set modeled by the *GST* distribution. Suppose that we observe X directly but Q is assumed missing in $X = \mu + k^{1/2}\sigma U Q^{-1/2}$, where U has a standard normal distribution and Q has gamma $GA(p\text{-}1/2,1/2)$ distribution. Thus the data $x_1, x_2, ..., x_n$ can be viewed as missing. So the complete data is $y = (x, q)$ (see Dempster, Laird, Rubin, 1980).

Let $R=\frac{X-\mu}{k^{1/2}\sigma}$. So the joint density function of the random variables of R (or X) and Q is

$$f_c(x,q;\mu,\sigma,k,p)=\frac{e^{-q(\frac{1+r^2}{2})}q^{p-1}}{\sqrt{\pi}\Gamma(p-1/2)\sigma k^{1/2}2^p}. \tag{5.4.4}$$

The complete data log-likelihood function is

$$l_c(\mu,\sigma)=\log\prod_{i=1}^{n}f_c(x_i,q_i;\mu,\sigma,k,p)$$

$$=-n\log\sigma-(1/2)\sum_{i=1}^{n}q_i(1+r_i^2)+(p-1)\sum_{i=1}^{n}\log q_i \tag{5.4.5}$$

except for an additive constant.

The conditional expectation of $l_c(\mu,\sigma)$ given x and $\mu=\mu^{(m)}$, $\sigma=\sigma^{(m)}$ is

$$Q(\mu,\sigma\mid\mu^{(m)},\sigma^{(m)})=E[l_c(\mu,\sigma)\mid\mu^{(m)},\sigma^{(m)}]$$

$$=-n\log\sigma-(1/2)\sum_{i=1}^{n}(1+r_i^2)E(q_i\mid r_i) \tag{5.4.6}$$

We have clearly $E(q\mid r)=2p/(1+r^2)$. Now letting $w_i^{(m)}=E(q_i\mid r_i)=2p/(1+r_i^2)$, we obtain

$$Q(\mu,\sigma\mid\mu^{(m)},\sigma^{(m)})=-n\log\sigma-(1/2)\sum_{i=1}^{n}w_i^{(m)}(1+r_i^2)$$

$$=-n\log\sigma-(1/2)\sum_{i=1}^{n}w_i^{(m)}[1+\frac{1}{k}(\frac{x_i-\mu}{\sigma})^2] \tag{5.4.7}$$

In the M-step we maximize $Q(\mu,\sigma \mid \mu^{(m)},\sigma^{(m)})$ with respect to μ and σ. The equations $\frac{\partial Q}{\partial \mu}=0$, $\frac{\partial Q}{\partial \sigma}=0$ yields the same estimates as in the (5.4.1) and (5.4.2). Therefore the IR and the EM algorithms are equivalent.

5.5. Examples

In this section we will study the real data sets whose likelihood functions were sketched in Chapter 3. These data sets are useful in demonstrating the performance of the generalized t distributions for finding the robust location and scale estimates of some data sets containing possible outliers and subgroups. In Example 5.1 we use the Cushny and Peebles Data. This data set is widely used in robust statistical analysis to evaluate the performance of the proposed method. In Example 5.2 we use a data set that has two subgroups. We can see in the example that for some values of the shape parameters the likelihood function has two local maximum points one for each group. This behavior of the *GT* and *GST* likelihood functions is a desirable behavior in robust statistical analysis to identify the different structures in the data. In Example 5.3 we compare the *GT*-based M-estimators with other robust location and scale estimators on two real data sets.

The estimates for the *GT* and *GST* distributions have been calculated using the IR algorithm given above Sections 5.1 and 5.4.

Example 5.1. (Cushny and Peebles Data) Hampel et al. (1986) used some robust methods to compute the location of the data since the observation 4.6 is clearly an outlier. Their results are summarized as follows: Median is 1.3, %10 trimmed mean is 1.4, %20 trimmed mean is 1.33 and mean of data set without the observation 4.6 is 1.24 (Hampel et al., 1986).

Figures 3.1(a)-3.1(d) show the plots of the location-only likelihood functions for the different values of the shape parameters. For the values of p=1.8 and q=1 the likelihood function has a unique maximum point occurs at $\hat{\mu}$=1.29 (See Figure 3.1(a)). We can see that unlike the sample mean, 1.58, the location estimate obtained from the *GT* distribution is very close to the median.

If $m/n < pq/(pq+1)$ is hold,(where n= sample size=10, m= number of coincident observations =2) regardless of the values of the shape parameter we can see that the scale likelihood function is unimodal (See Figure 3.1(d)). One can easily verify that when p=1.8 and q=1, we have $m/n = 0.2 < pq/(pq+1) = 0.6$. The maximum is attained at $\hat{\sigma} = 2.18$ (See Figure 3.1(d)).

Example 5.2. (Derby Data) The sample mean and the sample standard deviation for this data set are 131.87 and 14.25, respectively. The median is 125.2, and the robust scale estimate MAD (Median Absolute Deviation) is 4.74. The differences between the median and the sample mean, and also sample standard deviation and MAD suggest that the sample mean and the sample standard deviation are affected from the observations belonging to the interval 1875-1895 which are far from the main bulk of the data. We can also find the univariate version of MVE (Minimum Volume Ellipsoid) location estimate as 124.72 with the corresponding scale estimate 8.68. Further, the sample mean and the standard deviation of the observations without the points between 1875 and 1895 are 125.02 and 3.32, respectively.

Similar robust estimates for the location and scale parameters of the data set can also be obtained by the M-estimators based on the *GT* distribution. For example, when p=2.5 and q=1 the likelihood function has a unique maximum at $(\hat{\mu}, \hat{\sigma}) = (125.69, 7.42)$ (See Figure 3.2(a)). One can see that the maximum point occurs very close to the median and the robust estimates. Also, the maximum point is very close to the sample mean without the subgroup.

We treated above the observations between the years 1875-1895 as an outlier group in the data. However, one can think that these points come from a different distribution, and may want to find the estimates for this subgroup of the data. The sample mean and the sample standard deviation of this small group (without the main bulk of the data) are 159.60 and 3.74, respectively. The robust location estimates for this group are median=158.25 and MVE=158.40 with the scale MAD =1. 85 and 3.11, respectively.

We can use the multimodality property of the *GT* likelihood function to detect these two groups in the data. For example, when p=1.2, q=.05 the likelihood function

has two local maximum points in the parameter space (See Figure 3.2(b)). We can further observe that each maximum point corresponds to a group in the data. The maximum points occur at $(\hat{\mu}, \hat{\sigma}) = (125, .70)$ and $(\hat{\mu}, \hat{\sigma}) = (157.50, 3.69)$, respectively.

The multimodality of the likelihood function is also caught by the estimators from the *GST* distribution. When we choose p less than 1 we may obtain multiple roots of the *GST* based likelihood function. For example, for p=.55, k=.6 we have $(\hat{\mu}, \hat{\sigma})$=(124.06, .31) and $(\hat{\mu}, \hat{\sigma})$=(157.49, 1.37).

Example 5.3. In this example we use two real data sets to compare the generalized t based M-estimators with other robust location-scale estimators. They are Cushney and Peebles and Rosner (1975) data sets, which are often used in the literature to illustrate various robust estimators of location. Cushny and Peebles data were given in Example 3.1 (See also Figure 5.1). Rosner data consist of 10 monthly diastolic blood pressure measurements (Figure 5.2).

Table 5.1. Location and scale estimates of two real data sets (mve= minimum volume ellipsoid estimate for the univariate data).

	Cushny & Peebles		*Rosner*	
Distribution	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\mu}$	$\hat{\sigma}^2$
Normal	1.5800	1.3616	82.2000	232.3600
Cauchy	1.2628	0.1089	88.2266	23.1408
$T(\upsilon = 2)$	1.2806	0.2380	87.4224	37.3632
$T(\upsilon = .5)$	1.2834	0.0254	88.8448	13.7284
GT(p=1.6,q=1)	1.2820	0.2924	87.9741	55.4380
GT(p=1.6,q=.3)	1.2971	0.0406	88.3984	25.2477
GT(p=2.6,q=.4)	1.2433	0.2151	88.0170	51.2241
GT(p=2.8,q=.3)	1.2387	0.1499	88.3003	43.2078
GT(p=2.9,q=.1)	1.2800	0.0147	89.1015	16.2000
GST(p=5,k=1)	1.42	7.51	85.12	1034.13
GST(p=5,k=3)	1.42	2.50	85.12	344.77
GST(p=1,k=.2)	1.26	.46	88.23	115.70
GST(p=1,k=1)	1.26	.09	88.23	23.14
mve	1.2571	0.0995	88.3750	26.5536

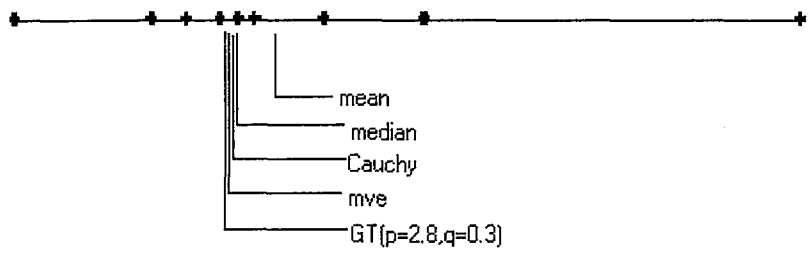


Figure 5.1. Cushny & Peebles Data and location estimates.

Figure 5.2. Rosner Data $x_i = \{90,93,86,92,95,83,75,40,88,80\}$ and location estimates.

As we can see from Table 5.1 and Figs. 5.1 and 5.2 above, the location and scale estimates based on the *GT* distribution may be more robust than the other robust estimators. From the Table 5.1 we can observe that for some values of p and q we can obtain very robust estimates of the location and scale parameters of a univariate data set. If p and q are properly chosen the location estimate based on the *GT* distribution can be better than the median which is known as the most robust location estimate and than the Cauchy.

Rosner data set has only one outlier. In this case the location and scale estimates based on the *GT* distribution can be as good as the mve estimates for location and scale, which are known as the most robust estimates (Fig. 5.2).

We observe that the estimates of the *GST* distribution become poor as we increase the value of p. Especially, for small k the scale estimate gets large for fixed

p. For the Rosner data Figure 3.3(a) shows the plots of the location-only likelihood functions for the different values of the shape parameters. For the values of p=1 and k=2 the likelihood function has a unique maximum point occurs at $\hat{\mu}$=89.54 (See Figure 3.3(a)). This estimate is very close to the median. Also Figure 3.3(b) shows the plot of the scale likelihood function for p=1=k. The estimate is $\hat{\sigma} = 80.97$.

6. REGRESSION ESTIMATION BASED ON THE *GT* DISTRIBUTION

In this chapter we will explore the robust regression estimation based on the *GT* distribution with known shape parameters. We will assume that the error term $\mathbf{u}$ in the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ has a *GT* distribution with known shape parameters. Then we will use the maximum likelihood estimation method to estimate the regression and scale parameters. We show that the estimation problem can be reduced to the weighted least squares case with the weights adjusted iteratively, and the regression estimate is the M-estimate with a redescending influence function. Because of the redescending property of the estimate the outliers will be downweighted, and hence the estimates is little affected by the outlying observations.

The performance of the robust regression estimators based on the *GT* distribution will be compared with other robust regression estimators that are well-known and widely used in regression analysis. The performance will be compared via the examples. The datasets used in examples are widely used in robust statistical analysis to assess the robustness of the estimators. Some of the data sets are real, and some of them are artificially generated data sets. In the examples our intention is to illustrate the ability of the *GT* based regression M estimators to handle the outliers in the regression setting. We show in the examples that the *GT* based regression M estimating equations often have a solution which gives good fit to the data, ignoring the outlying observations, if they are properly tuned (that is, if the shape parameters are properly chosen). Also, we observe from the examples that if the data set has multiple structure, such as different clusters, then the *GT* based M estimators can have multiple solutions for the smaller values of the shape parameters. Using this property we can find one fit to each cluster in the data. This behavior of the *GT* M estimators is desirable in recent robustness studies. Mainly, we want to identify the different structures in the data if there are some (see Arslan, 2002a).

Before we move on the regression estimation via the *GT* distribution it should be noted that in this chapter we only discuss the *GT* based regression M estimators, since the regression estimation methods based on the *GST* distributions have already given in the literature (see, e.g. Dempster, Laird, Rubin, 1980).

6.1. Parameter Estimation

Consider the following regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\ , \tag{6.1.1}$$

where, $\mathbf{y}$ is the $n\times 1$ response variable, $\mathbf{X}$ is the $n\times p$ matrix containing p explanatory variables including an intercept as the first column, $\boldsymbol{\beta}$ is the $p\times 1$ unknown regression parameters, and $\mathbf{u}$ is the $n\times 1$ vector of random errors.

Assume that the errors u_i in model (6.1.1) are independently and identically *GT*-distributed with zero mean, unknown scale parameter σ and known shape parameters p and q. If $\mathbf{x_i}$ denote the *i*th row of the matrix $\mathbf{X}$ then the density of the *i*th observation y_i is given by

$$f(y_i) = \frac{pq^q}{2\sigma B(1/p,q)}\left(q+\frac{|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^p}{\sigma^p}\right)^{-q-1/p},\ -\infty < y_i < \infty,$$
$$i=1,2,...,n. \tag{6.1.2}$$

After taking the derivatives of the log likelihood function

$$l(\boldsymbol{\beta},\sigma) = -n\log\sigma - (q+\frac{1}{p})\sum_{i=1}^{n}\log\left(q+\frac{|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^p}{\sigma^p}\right) \tag{6.1.3}$$

with respect to $\boldsymbol{\beta}$ and σ, and setting to zero yield the following estimating equations for $\boldsymbol{\beta}$ and σ

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = (q+\frac{1}{p})\sum_{i=1}^{n}\frac{\sigma^{-p}px_i'\,|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^{p-1}\ sign(y_i-\mathbf{x_i}\boldsymbol{\beta})}{q+\dfrac{|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^p}{\sigma^p}} = 0\,,\ p>1 \tag{6.1.4}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + (q+\frac{1}{p})\sum_{i=1}^{n}\frac{p\sigma^{-p-1}\,|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^p}{q+\dfrac{|\,y_i-\mathbf{x_i}\boldsymbol{\beta}\,|^p}{\sigma^p}} = 0\,. \tag{6.1.5}$$

These equations can be rewritten in the following forms:

$$\sum_{i=1}^{n} w_i \mathbf{x}_i' (\mathbf{x}_i \hat{\boldsymbol{\beta}} - y_i) = 0 \tag{6.1.6}$$

$$\sum_{i=1}^{n} w_i (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2 - n\hat{\sigma}^2 = 0 \tag{6.1.7}$$

where $w_i = \dfrac{(pq+1)\,|\,y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\,|^{p-2}\,\hat{\sigma}^{2-p}}{q+|\,y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\,|^{p}\,\hat{\sigma}^{-p}}$.

These estimating equations can be viewed as the M estimating equations for the regression parameters $\boldsymbol{\beta}$ and the scale parameter σ with the corresponding ψ function for the regression parameter

$$\psi(r) = \frac{(pq+1)\left|r/\sigma\right|^{p-2} r}{q+\left|r/\sigma\right|^{p}}, \tag{6.1.8}$$

where $r = y - \mathbf{x}\boldsymbol{\beta}$ is the residual.

To investigate the local robustness of the regression estimator obtained from the *GT* distribution we will first give the influence function. One can easily obtain the joint influence function of the regression and scale parameters under the assumption that the errors and the carriers are independent and the error distribution is symmetric about origin. Since our main interest is the regression parameter rather than the scale, we will only give the influence function of the regression parameter $\boldsymbol{\beta}$. The influence function for the regression M estimator for $\boldsymbol{\beta}$ given in equation (6.1.6) can be obtained as (see, e.g., Hampel et al. 1986)

$$IF(\mathbf{x}, y, \hat{\boldsymbol{\beta}}) = \psi(r).\mathbf{M}^{-1}\mathbf{x}', \tag{6.1.9}$$

where

$$\mathbf{M} = (E(\psi'(r)).(E(\mathbf{x}'\mathbf{x})), \tag{6.1.10}$$

and the expectations are taken with respect to the error and the carrier distributions. One can easily see that the influence function has two factors, the first factor $\psi(r)$ is only the function of the residual, and the second factor $\mathbf{M}^{-1}\mathbf{x}'$ is only the function of the carrirers. In general, the influence function seems unbounded in the carrier space but because of the redescending nature of $\psi(r)$ function (see Figure (6.1)) the outlying observations in **x**-direction will be downweighted as well. These regression estimators can be regarded as the redescending regression M estimators. Since the ψ-function depends on the shape parameters p and q the amount of the points to take small (or large) weights will increase or decrease for different values of these parameters. Therefore, since the shape parameters p and q will also play an important role in regression estimation problem we will treat them as the robustness tuning constants.

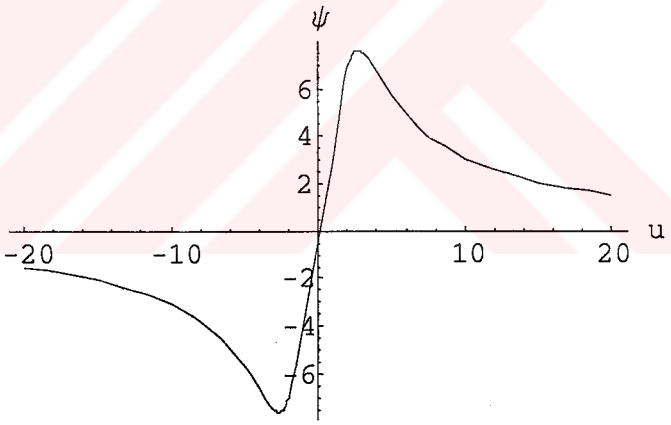


Figure 6.1. ψ-function (p=3, q=10)(McDonald and Newey, 1988)

Further, after rearranging the estimating equations (6.1.6) and (6.1.7) the estimators for $\boldsymbol{\beta}$ and σ can be obtained as the iteratively reweighted least squares form as follows

$$\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i y_i \mathbf{x}_i') \qquad (6.1.11)$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} w_i (y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})^2 . \qquad (6.1.12)$$

Since the weight function is a decreasing function of the residuals it will give small weights to points which are far from the bulk of the observations, i.e. outliers. So our estimators are not affected much from outliers. Note that the weights assigned to observations are not fixed and they change for each iteration in the estimation process. Since the weights depend on the estimators the estimators can not be explicitly stated and, therefore, some numerical algorithms should be used to compute the estimates. Again, many different algorithms can be proposed to calculate the estimates, but the equations (6.1.11) and (6.1.12) can be easily modified to obtain a simple iteratively reweighting algorithm. In the example given below this iterative reweighting algorithm will be used to compute the regression M estimates based on the *GT* distribution.

6.2. Examples

In this section we will present some examples to display the performance of the *GT* based M regression estimators by handling the outliers and/or the clusters or different structures in the data. In the examples we will use some of the widely used datasets in robust statistical analysis. We will mainly consider the following problems and the examples will help us to see how well the *GT* based regression M estimators can handle these problems: (i) Can we find a fit to the data that fits to the majority of the data and ignores the outliers? (ii) Can we use the IR algorithm to get that fit or fits? (iii) Do the convergence of the IR algorithm sensitive to the starting points? (iv) Can we find a fit to each cluster in the data if some clusters are present?

From our limited experiences based on the examples we can say that the *GT* based regression M estimators can produce a good fit to the data ignoring the outliers. The IR algorithm formulated from the estimating equations can be used to compute the estimates. Different starting points for the algorithm can produce different convergence points if the likelihood function has multiple maximum points in the parameter space. If the data contain clusters or subgroups (or a group of

outliers) we can be able to find these subgroups using the *GT* based M estimators. Each local maximum of the likelihood function may correspond to a cluster in the data, and produce a different fit to the data. Thus, one can consider all the fits to illustrate the complete structure of the data. Finding the multiple fits to the data is particularly very important in applied sciences such as computer vision.

Example 6.1. Pilot-Plant Data (Rousseeuw and Leroy, 1987).

In this data set the response variable corresponds to the acid content determined by titration, and the explanatory variable is the organic acid content determined by extraction and weighing. This data set has no outliers and there is a strong linear relationship between variables. So the least squares (LS) fit $y = 35.46 + 0.32x$ accords with the other robust fits. If the x-value of the 6^{th} observation is recorded as 370, this error produces an outlier in the x-direction,i.e. a leverage point. Now the LS estimate is no longer a good fit. Because the LS fit is pulled by the outlier. To remedy this problem a robust method is needed.

Table 6.1. Regression estimates for Pilot-Plant Data.
(standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
LS	58.94	.08	14.80	-82.27
	(6.61)	(.05)		
GT(p=2.1, q=1)	35.50	.32	1.55	-13.86
	(.64)	(.00)		
GT(p=2.4, q=1)	35.37	.32	1.65	-13.86
	(.58)	(.00)		
GT(p=4, q=.25)	35.26	.32	1.37	-33.14
	(.55)	(.00)		
GT(p=1.8, q=2)	35.54	.32	1.69	22.70
	(.65)	(.00)		
t_1	35.78	.32	.84	-45.63
	(.92)	(.01)		
t_2	35.55	.32	1.07	-46.73
	(.73)	(.01)		
LMS	36.34	.31	1.33	
Huber	54.34	.13		
Tukey	35.32	.32		
LS*	35.32	.32	1.22	-30.81

We use the iteratively reweighting least squares (IRLS) method to estimate the regression parameters. For a comparison we consider several robust methods (Table 6.1). We see that the LS and Huber fits are influenced by the outlier. However, the redescending M-estimators Tukey, *t*-estimators and *GT* estimators stay resistant to outlier (Figure 6.2). These fits are similar. If we delete the outlier and perform the LS, the resulting line LS*, will be the same as the robust fit (Table 6.1).

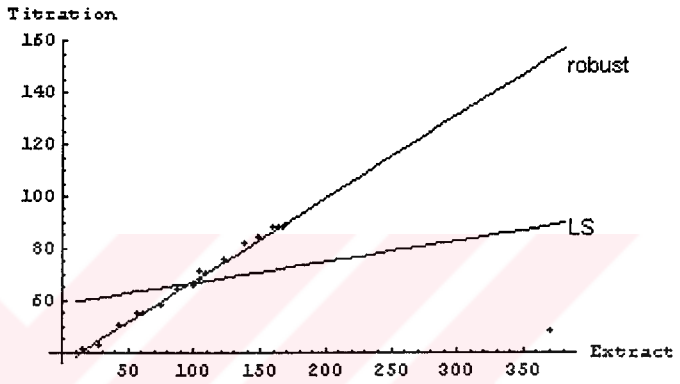


Figure 6.2. Regression lines for Pilot-Plant Data

Scatter plots of the observations can be used to detect outliers in simple linear regression. However, the robust procedures easily and automatically identify the outliers. In this example the identification of the 6^{th} observation as an outlier is resulted from having the smallest weight. For example, the weights given by the estimator $GT(p=4, q=.25)$ are ranges from $3.3x10^{-4}$ to 2 and the 6^{th} observation has the smallest weight.

Example 6.2. Star Data (Rousseeuw and Leroy, 1987).

This example comes from astronomy. The independent variable log temperature is the logarithm of the effective temperature at the surface of the star, and the logarithm of its light intensity corresponds to the dependent variable. In the scatter plot we see two groups of points: a major group in the center and a group of four observations in y-space (Figure 6.3.)

Table 6.2. Regression estimates for Star Data.
(standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
LS	6.79	-.41	.55	-38.81
	(1.24)	(.29)		
$GT(p=1.5, q=1)$	8.00	-.65	.50	-32.58
	(1.06)	(.25)		
$GT(p=1.3, q=1.5)$	8.04	-.66	.48	11.81
	(2.30)	(.53)		
$GT(p=1.8, q=2)$	7.07	-.46	.66	53.34
	(1.91)	(.44)		
$GT(p=1.8, q=4)$	6.88	-.42	.70	268.32
	(1.64)	(.38)		
$GT(p=3.1, q=1)$	6.93	-.45	.76	-32.58
	(1.51)	(.35)		
$GT(p=1.1, q=.05)$	8.46	-.78	.13	-176.30
	(.14)	(.03)		
t_1	7.95	-.64	.36	-50.67
	(1.35)	(.31)		
t_2	7.34	-.52	.44	-45.18
	(2.30)	(.53)		
LMS	-12.63	3.97	.36	
Huber	6.79	-.41		
Tukey	6.83	-.42		

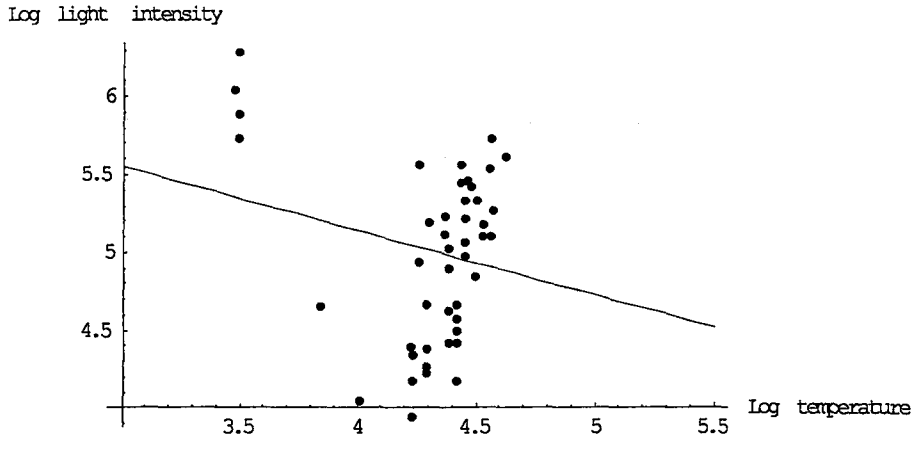


Figure 6.3. LS regression line for Star Data.

If we begin with the LS fit for our iteration we get the parameter estimates presented in Table 6.2. These estimates are all poor except for the LMS. The LMS

line nicely fits the main body of the data. If we use the LMS line as a robust starting value for the iteration we find two lines for different values of shape parameters (Table 6.3.) For very small values of them the *GT* estimators fit as well as LMS. For example, the estimator *GT*(p=1.1, q=.05). This estimator gives the observation 11, 20, 30 and 34 zero weight in the final step of the iteration.

Table 6.3. Regression estimates for Star Data with starting point LMS. (standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
GT(p=1.7, q=1)	1.99	.69	.54	-32.58
	(.74)	(.17)		
GT(p=1.6, q=1)	1.68	.76	.51	-32.58
	(1.00)	(.23)		
GT(p=1.3, q=2)	2.07	.67	.50	59.39
	(.88)	(.20)		
GT(p=4, q=.25)	-4.29	2.10	.45	-77.88
	(3.55)	(.82)		
GT(p=1.3, q=1.3)	-4.73	2.20	.42	-6.28
	(.00)	(.00)		
GT(p=1.1, q=.05)	-12.74	4.00	.13	-176.30
	(.00)	(.00)		

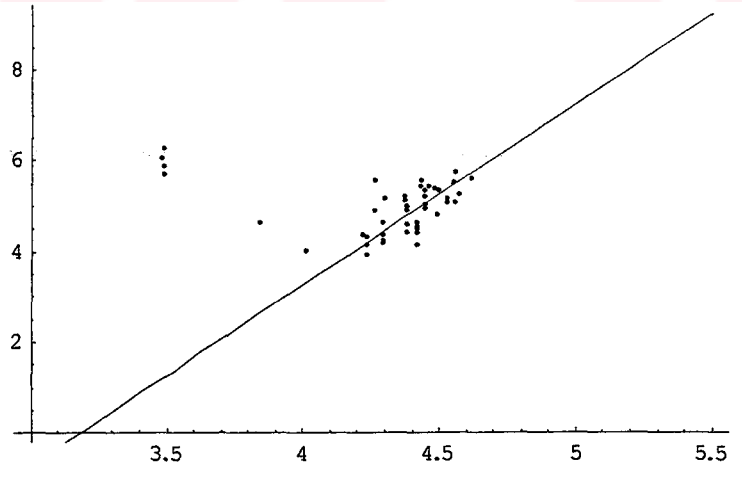


Figure 6.4. *GT*(p=1.1, q=.05) regression line for Star Data.

Example 6.3.

We now analyze an artificial data set generated by Arslan (2002a). It is advantageous to employ such an artificial data set because the position of the bad

points is known exactly and so the effectiveness of the method can be measured without much controversies as in real data. This data set consists of one hundred points and contains two subgroups. As it can be seen easily from the scatter plot, one group of forty lies in the x-space (Figure 6.5.) The parameter estimates are given in the following Table 6.4. (The initial value for the iteration is LS fit.)

Table 6.4. Regression estimates for the artificial data.(standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
LS	.34	-.17	2.50	-233.38
	(.32)	(.02)		
$GT(p=1.2, q=.01)$	.95	-.21	1.03	-516.52
	(.07)	(.00)		
$GT(p=1.8, q=.01)$	1.12	-.22	.46	-476.78
	(.02)	(.00)		
$GT(p=1.3, q=.1)$	1.13	-.22	.52	-300.61
	(.04)	(.00)		
$GT(p=1.01, q=1)$	.95	-.21	1.03	-69.31
	(.45)	(.02)		
t_1	.44	-.19	1.15	-234.94
	(.27)	(.01)		
t_2	.36	-.19	1.46	-228.46
	(.25)	(.01)		
LMS	1.82	-.24	1.68	
Huber	.39	-.18		
Tukey	.47	-.19		

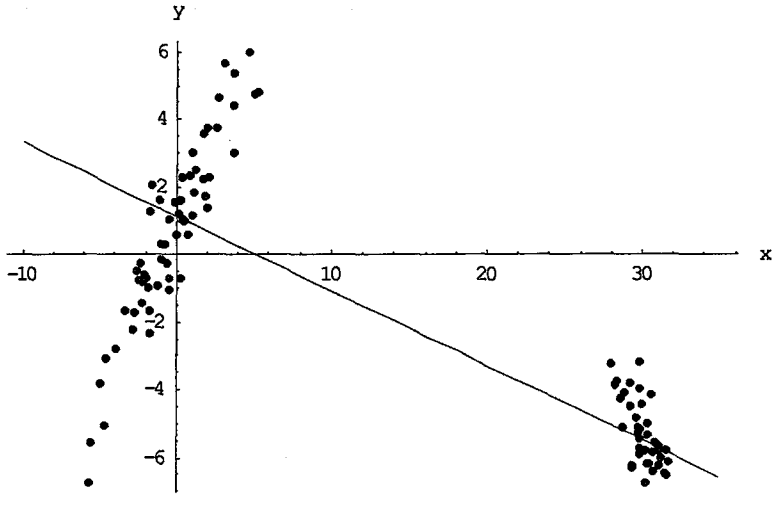


Figure 6.5. A fit exposed to the effect of the outliers group.

We note that Table 6.4 reflects similar fits. For example, $GT(p=1.8, q=.01)$ passing through the outliers group (Figure 6.5.)

Table 6.5. Regression estimates for the artificial data with starting point LS* (standard errors are in parentheses).

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
LS*	.98	.97	.99	-84.43
	(.13)	(.05)		
$GT(p=1.2, q=.01)$	.77	.90	.62	-516.52
	(.01)	(.00)		
$GT(p=1.8, q=.01)$	.87	.76	.38	-476.78
	(.04)	(.00)		
$GT(p=1.3, q=.1)$	.69	.63	.74	-300.61
	(.07)	(.00)		
$GT(p=1.01, q=1)$	-.65	-.16	1.04	-69.31
	(.53)	(.03)		
t_5	1.02	-.22	.76	-250.68
	(.18)	(.01)		
t_1	.45	-.19	1.15	-234.94
	(.27)	(.01)		
t_2	.36	-.19	1.46	-228.46
	(.25)	(.01)		
LMS*	1.17	.97	.98	
Huber	.39	-.18		
Tukey	.97	.97		

Table 6.6. Regression estimates for the artificial data with starting point LS** (standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	Log-lik
LS**	15.46	-.69	.71	-42.91
	(3.66)	(.12)		
$GT(p=1.2,q=.01)$	6.88	-.40	.47	-516.52
	(.05)	(.00)		
$GT(p=1.8,q=.01)$	1.35	-.23	.27	-476.78
	(.65)	(.03)		
$GT(p=1.3, q=.1)$	1.35	-.23	.45	-300.61
	(.58)	(.03)		
$GT(p=2.1, q=1)$	.35	-.19	2.13	-69.31
	(.26)	(.01)		

When we use the LS fit based on the outlier-free data as an initial value for our iteration we see that the *GT* estimators yield lines through good data points for

small p and q. (Table 6.5. and Figure 6.6.). On the other hand when we choose LS** (Least Squares estimates for outlier group) as an initial estimate we can obtain fits through the outlier group (Table 6.6 and Figure 6.7).

In this example we also see that the LMS estimator which is known to be a high-breakdown-point estimator is pulled by the outlier group (Figure 6.8.) Notice that the LMS fit is similar to the LS fit (Table 6.4.)

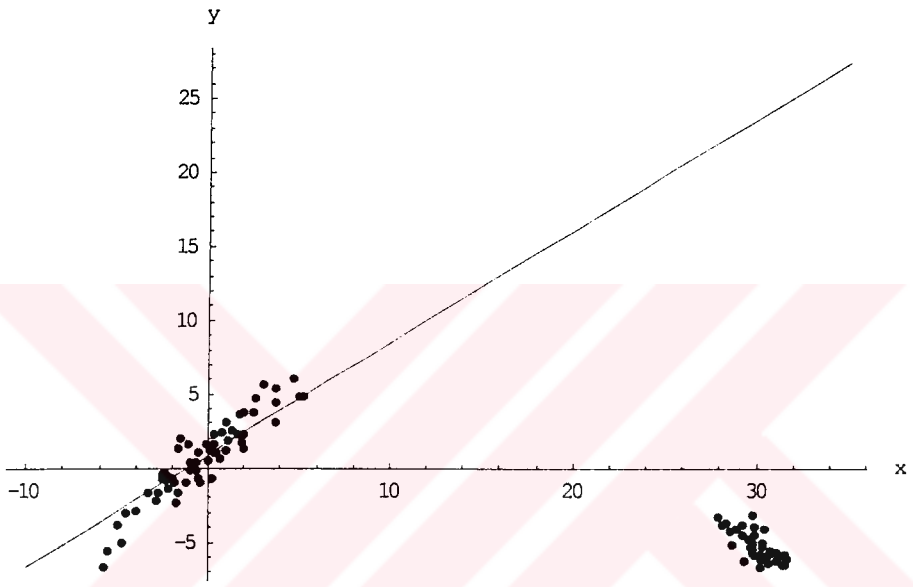


Figure 6.6. A fit through the group of good data points.

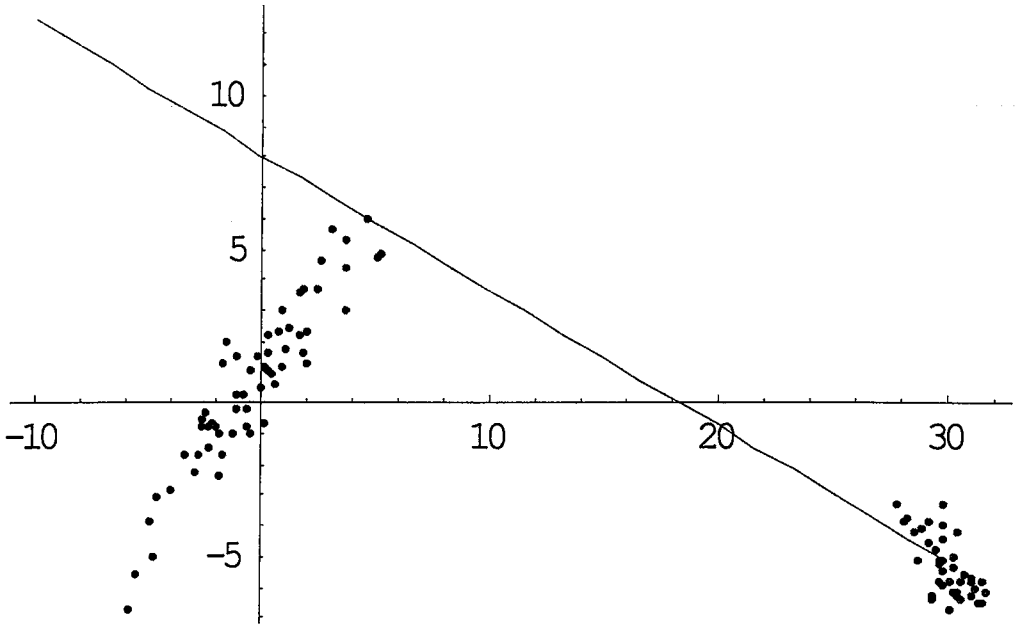


Figure 6.7. A fit through the group of outliers.

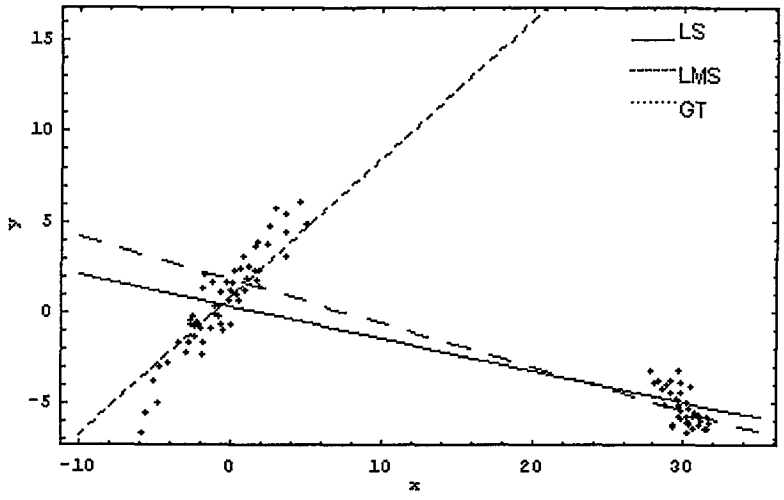


Figure 6.8. A comparison of the LS, LMS and *GT* fits.

Consequently this example shows that using the *GT* distributions family we can easily model data containing subgroups.

Example 6.4. Stackloss Data (Rousseeuw and Leroy, 1987).

This real data set describes the operation of a plant for the oxidation of ammonia to nitric acid and consist of 21 four-dimensional observations and has been analyzed by several authors such as Andrews (1974) and Lange et al. (1989).

The LS regression line for the data is $y = -39.92 + .72x_1 + 1.3x_2 - .15x_3$. In case a multiple regression model is considered, index plots can be used to detect outliers. These plots are based on standardized residuals and are reliable tools for identifying outliers when a difference occurs between the corresponding plots for the LS and the robust method.

The LS index plot is shown in Figure 6.9. In that figure the vertical axis stdres corresponds to the standard residuals. The standardization is performed by the division of the raw residuals $r_i = y_i - \hat{y}_i$ by the scale estimate corresponding to the fit. We notice that the LS standardized residuals lies within the horizontal bands –2.5 and 2.5.

If we start at the LS line we get the estimates in the following Table 6.7. For p=2.5 and q=1 the *GT* estimators give the observations 4 and 21 the smallest weights .4 and .2 respectively in the final step of the iteration. From the corresponding index plot, Figure 6.10., we see that the observations 4 and 21 lie outside the band formed by the LS standardized residuals. When we get small the values of p and q the number of observations having smallest weights increases. For example, when p=1.5 and q=.2 the observations 1, 3, 4, and 21 all take zero weight at the final iteration. The index plot associated with the *GT*(p=1.5,q=.2) estimator is given in Figure 6.11. We see from Figure 6.11 that the observations 1, 3, 4, and 21 stay away from the region bounded by the LS standardized residuals. These observations are outliers.

Table 6.7. Regression estimates for Stackloss Data (standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Loglik
LS	-39.92	.72	1.30	-.15	2.97	-52.65
	(11.9)	(.13)	(.37)	(.16)		
LMS	-39.25	.75	.50	0.00	1.21	
Huber	-40.75	.76	1.17	-.14		
Tukey	-41.33	.83	.95	-.13		
t_1	-38.63	.85	.49	-.07	.76	-49.58
	(5.97)	(.07)	(.18)	(.08)		
t_2	-38.12	.85	.56	-.09	1.80	-50.31
	(6.86)	(.07)	(.21)	(.09)		
GT(p=2.5,q=1)	-39.96	.86	.69	-.11	2.52	-14.56
	(10.38)	(.11)	(.32)	(.14)		
GT(p=1.5,q=.5)	-40.47	.84	.55	-.05	.80	-33.28
	(4.09)	(.05)	(.13)	(.05)		
GT(p=1.5,q=.2)	-39.94	.83	.57	-.06	.09	-49.23
	(1.60)	(.02)	(.05)	(.02)		
GT(p=1.5,q=.05)	-39.99	.83	.56	-.06	0.00	-72.84
	(2.01)	(.02)	(.06)	(.03)		
GT(p=1.3,q=.5)	-40.09	.83	.56	-.06	.58	-34.34
	(6.92)	(.08)	(.21)	(.09)		

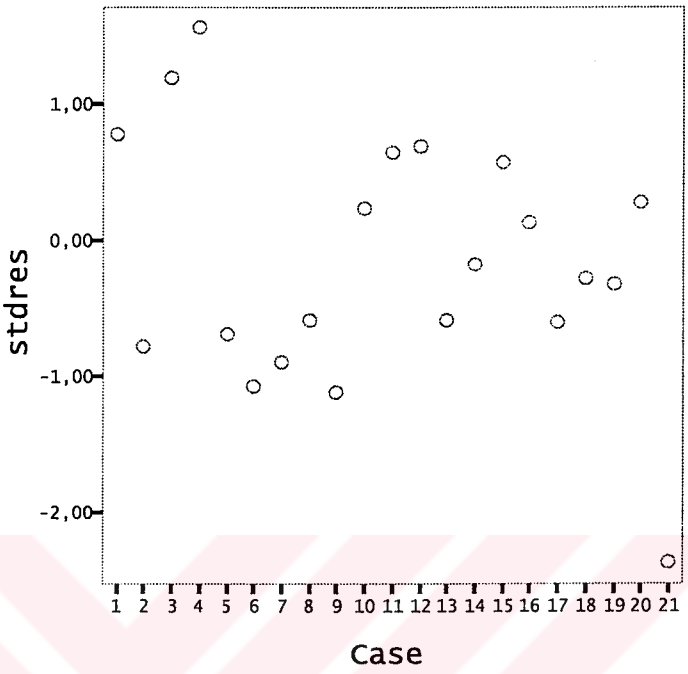


Figure 6.9. Index plot associated with LS.

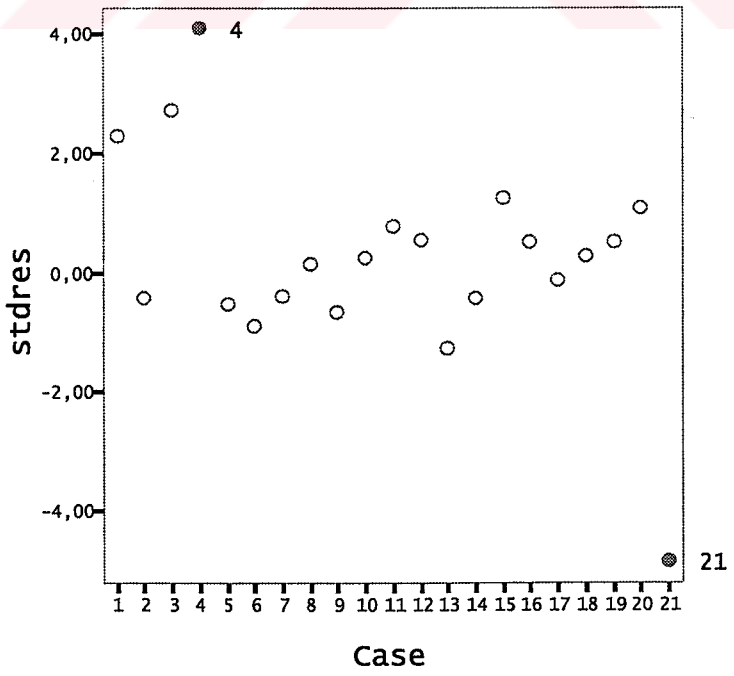


Figure 6.10. Index plot associated with the *GT*(p=2.5,q=1) fit.

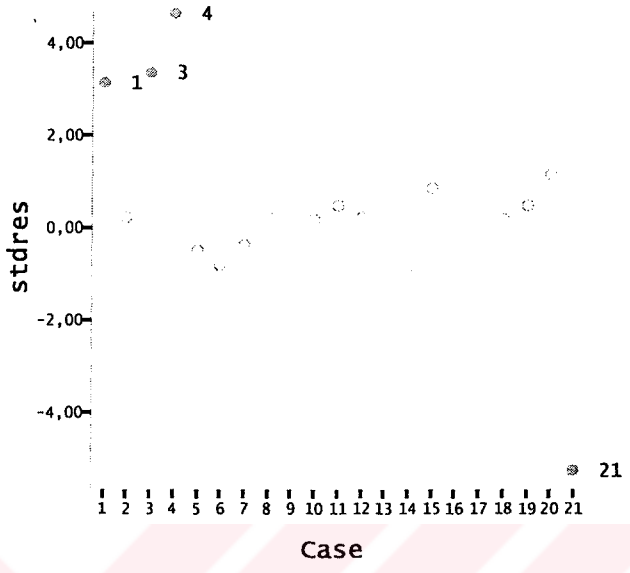


Figure 6.11. Index plot associated with the $GT(p=1.5,q=.2)$ fit.

Table 6.8. Regression estimates for Stackloss Data without four outliers. (standard errors are in parentheses)

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Loglik
LS*	-37.65	.80	.58	-.07	1.10	-25.70
	(4.73)	(.07)	(.17)	(.06)		
LMS	-34.25	.71	.36	0.00	.48	
Huber	-37.56	.79	.57	-.07		
Tukey	-37.30	.81	.54	-.07		
t_1	-36.69	.75	.38	.00	.23	-27.43
	(1.82)	(.02)	(.06)	(.02)		
t_2	-37.71	.84	.46	-.06	.57	-26.56
	(6.03)	(.07)	(.19)	(.08)		
$GT(p=2.5,q=1)$	-37.21	.83	.49	-.07	1.22	-11.78
	(5.63)	(.06)	(.17)	(.07)		
$GT(p=1.5,q=.5)$	-37.18	.82	.41	-.05	.57	-26.94
	(2.64)	(.03)	(.08)	(.03)		
$GT(p=1.5,q=.2)$	-35.91	.82	.44	-.07	0.08	-39.85
	(1.84)	(.02)	(.06)	(.02)		
$GT(p=1.5,q=.05)$	-35.94	.82	.44	-.07	0.00	-58.97
	(.41)	(.00)	(.01)	(.01)		

If we remove the four outliers from the data and apply our methods to the remaining observations we get the estimates given in Table 6.8. We note that the

estimates are all similar. We also note that the LS* is similar to *GT* estimates for small p and q given in Table 6.7.

Example 6.5. Hawkins-Bradu-Kass Data (Rousseeuw and Leroy, 1987).

This artificial data set has been generated by Hawkins, Bradu and Kass (1984). The data set consists of 75 observations in four dimensions. The first 10 observations are bad leverage points, and the next 4 points are good leverage points.

It is reported in Hawkins et al. (1984) that the bad leverage points are masked and the four good leverage points appear outlying by M-estimators.

The parameter estimates are given in the following Table 6.9. We note that the Tukey method and the t estimates are not much different. We also note that the $\hat{\beta}_2$ estimate for each of the methods LS, Huber and $GT(p=1.1,q=.05)$ is negative.

Table 6.9. Regression estimates for Hawkis-Bradu-Kass Data (Standard errors are in parentheses).

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Loglik
LS	-.39	.24	-.33	.38	2.19	-165.19
	(.42)	(.26)	(.16)	(.13)		
LMS	-.56	.18	.04	-.07	.62	
Huber	-.68	.19	-.12	.33		
Tukey	-.94	.14	.19	.18		
$t_{.5}$	-1.06	.07	.28	.16	.09	-125.15
	(.03)	(.02)	(.01)	(.01)		
t_1	-1.06	.12	.26	.16	.21	-115.46
	(.17)	(.10)	(.06)	(.05)		
t_2	-.99	.14	.21	.17	.34	-114.20
	(.22)	(.14)	(.08)	(.07)		
$GT(p=2.5,q=.5)$	-1.01	.15	.22	.16	.76	-106.96
	(.20)	(.12)	(.07)	(.06)		
$GT(p=1.8,q=2)$	-.95	.14	.19	.19	.93	85.12
	(.15)	(.09)	(.05)	(.05)		
$GT(p=1.5,q=.5)$	-1.04	.07	.28	.16	.49	-118.86
	(.05)	(.03)	(.02)	(.01)		
$GT(p=1.3,q=.2)$	-.61	-.10	.13	.30	.28	-183.08
	(.06)	(.04)	(.02)	(.02)		
$GT(p=1.1,q=.05)$	-.90	.35	-.16	.34	.11	-281.33
	(.01)	(.01)	(.00)	(.00)		

By the index plot of LS ,Figure 6.12, we see that the observations 11, 12 and 13 fall outside the ± 2.5 band; so they are outliers with respect to this plot. We know,

however, that they are good points. On the other hand, the bad leverage points having small standardized LS residuals have been masked. Hawkins et al. (1984) noted that the index plots associated with M-estimators are very similar to that of LS. One estimator in Table 6.9, the LMS estimator, however, singles out the bad leverage points perfectly (Figure 6.13.)

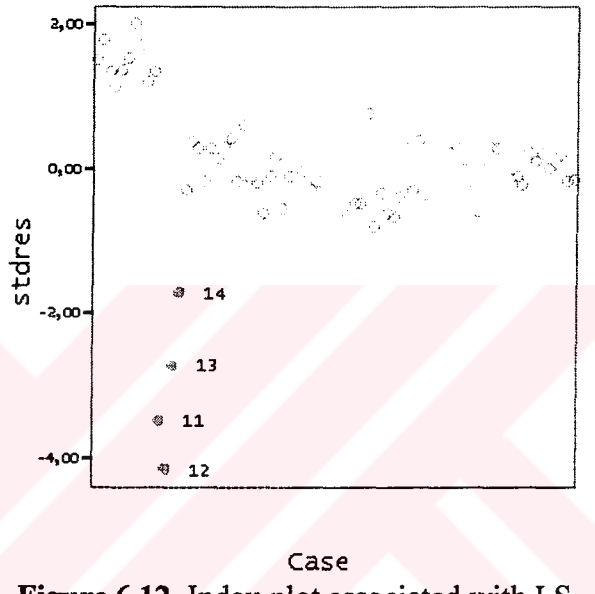


Figure 6.12. Index plot associated with LS.

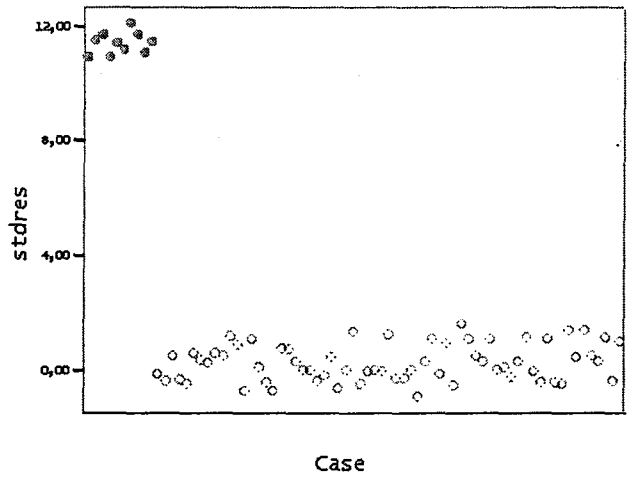


Figure 6.13. Index plot associated with LMS.

If we delete the first 10 outliers and fit the model, we obtain the LS* and the LMS* estimates shown in Table 6.10. We note that the LMS and the LMS* estimates are similar. If we start the iteration with LS* then we can capture the outliers by some

methods (Figure 6.14.) Tukey, $t_{.5}$, and the GT estimates except for $GT(p=1.8,q=1)$ are all robust. These estimates do not give much different results (Table 6.10.)

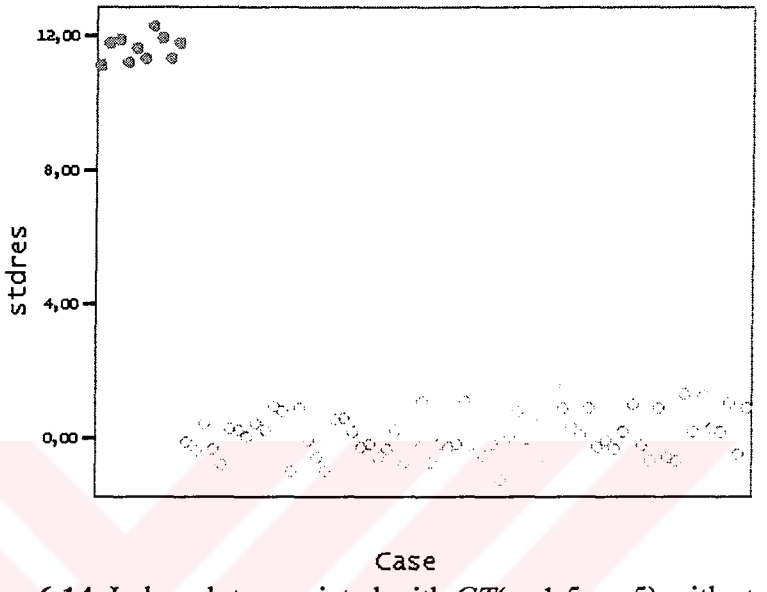


Figure 6.14. Index plot associated with $GT(p=1.5,q=.5)$ with starting point LS*.

Table 6.10. Regression estimates for Hawkis-Bradu-Kass data with starting point LS* (Standard errors are in parentheses).

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	Loglik
LS*	-.18	.08	.04	-.05	.54	-52.17
	(.10)	(.07)	(.04)	(.04)		
Tukey	-.18	.08	.04	-.05		
$t_{.5}$	-.42	.21	.04	-.09	.12	-141.72
	(.07)	(.04)	(.03)	(.02)		
t_1	-1.06	.12	.26	.16	.21	-115.46
	(.17)	(.10)	(.06)	(.05)		
$GT(p=1.8,q=1)$	-1.01	.13	.23	.17	.77	-51.99
	(.14)	(.09)	(.05)	(.04)		
$GT(p=1.8,q=.8)$	-.30	.14	.04	-.07	.75	-76.67
	(.16)	(.10)	(.06)	(.05)		
$GT(p=1.5,q=1)$	-.30	.15	.04	-.07	.72	-51.99
	(.21)	(.13)	(.08)	(.07)		
$GT(p=1.5,q=.5)$	-.41	.23	.04	-.09	.55	-118.86
	(.19)	(.12)	(.07)	(.06)		
$GT(p=1.3,q=.2)$	-.25	.17	.04	-.08	.29	-183.08
	(.02)	(.01)	(.01)	(.01)		

7. CONCLUSIONS AND FUTURE WORKS

In this thesis we have considered the three kinds of generalizations of the Student's t distribution in detail. A new scale mixture representation of the *GT* distribution has been given. We have used the generalized t distributions originally to obtain robust estimators for the location and the scale parameters of a univariate data set. We have shown that the local robustness of the estimators for the location and scale estimators can only be achieved if the shape parameters of the *GT* distribution are kept fixed. We have obtained some conditions on the shape parameters under which the likelihood function is unimodal in the parameter space. Also, we have observed that when we relax the conditions on the shape parameters the likelihood function may have multiple local maxima. This behavior of the likelihood function for location and scale can be useful in robust statistical analysis to detect presence of subgroups in data. It can be argued that each maximum point may correspond to a subgroup in data. A similar discussion for the regression case for the redescending M-estimators is given by Morgenthaler (1990) and Arslan (2002a). Since the results obtained for the location and scale estimation problem can be easily extended to the regression problem, it can be considered that each local maximum point of the likelihood function may correspond to some structure of interest in the data.

We have also given some examples to illustrate the properties of the likelihood functions obtained in the thesis. In the examples the values of the shape parameters have been chosen to get the appropriate display of the shape of the likelihood function. One can easily choose different values for the shape parameters as long as the conditions given in the theorems hold, and still can have the similar behavior of the likelihood function.

We have also used the *GT* distribution as an alternative error distribution in the linear regression case. The estimators we have found were redescending, so the effect of outliers in data reduced. We have observed from the examples that *GT* distribution can provide a good robust alternative error distribution in regression setting, and the estimators obtained from the *GT* distributions can fit the data well if they are properly tuned.

In regression setting the influence function of the estimators for the regression and the scale parameters have been mentioned, but we have not given the asymptotic properties of the estimators, such as asymptotic normality asymptotic efficiency. However, it should be noted that since the estimators have been obtained from the maximum likelihood estimation method the estimator $\hat{\boldsymbol{\beta}}$ is normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\Delta = \frac{\sigma^2 E[\psi^2(u)]}{\{E[\frac{d}{du}\psi(u)]\}^2}(\mathbf{X}'\mathbf{X})^{-1}$ (Huber, 1981). The breakdown point of the regression estimators based on the *GT* distribution has not been also given. A future work is planned to investigate the asymptotic properties of the regression estimators as well as the breakdown points.

In this thesis we have only investigated the generalized t distributions with symmetric density functions. A similar analysis can also be done for the skewed generalized t distributions mentioned in Chapter 2 (cf. Carroll and Welsh, 1988). In particular, the skew generalized t distributions can be introduced as a robust modeling distribution alternative to the traditional symmetric distributions.

Further in this thesis our main emphasis has only been given to the univariate versions of the generalized t distributions. Recently, Arslan (2002c) has defined a new family of the multivariate generalized t distributions that includes the multivariate t distribution, generalized t distribution defined by Arellano-Valle and Bolfarine (1995). Also, in Arslan (2002c) the multivariate version of the *GT* distribution has been defined. Multivariate generalized t distributions family defined by Arslan (2002c) will be a robust alternative distributions family to the multivariate normal and multivariate t distribution to model a multivariate data set with heavier tails that the normal and t distributions. The properties of this family have not been studied yet. One of our future projects is to study the properties of the multivariate *GT* distribution, and use it to find alternative robust estimates for the location and scatter parameters of a multivariate dataset. Also, we will use the multivariate generalized t distributions to find the estimates for the regression parameters in the multivariate regression setting.

REFERENCES

ANDREWS, D. F., 1974. A robust method for multiple linear regression. Technometrics, 16, 4, 523-531.

ANDREWS, D. F., MALLOWS, C. L., 1974. Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B, Vol:36, 99-102.

ARELLANO-VALLE, R. B., BOLFARINE, H., 1995. On some characterizations of the t distribution. Statistics and Probability Letters, 25, 79-85.

ARSLAN, O., 1992. Multivariate robust analysis based on the t distribution and the EM algorithm. Unpublished PhD thesis, Leeds University, Leeds, U.K.

_____, 2002(a). A simple test to identify good solutions to redescending M-estimating equations for regression. (R. Dutter, U. Gather, P. J. Rousseeuw and P. Filzmoser, Eds.). Development in Robust Statistics. Proceedings of ICORS 2001, 50-61.

_____, 2002(b). Convergence behavior of an iterative reweighting algorithm to compute multivariate M-Estimates for location and scatter. Journal of Statistical Planning and Inference. (In Press)

_____, 2002(c). Family of multivariate generalized t distributions. (Revised for Journal of Multivariate Analysis)

BLOOMFIELD, P., STEIGER, W. L., 1983. Least Absolute Deviations Theory: Applications and Algorithms. Boston: Birkhauser.

BOLLERSLEV, T., ENGLE, R. F., NELSON, D. B., 1994. ARCH models. (R. F. Engle and D. L. McFadden, Eds.) Handbook of econometrics 4. Elsevier Science, 2959-3038.

BOX, G. E. P., TIAO, G. C., 1962. A further look at robustness via Bayes Theorem. Biometrika, 49, 419-432.

BUTLER, R. J., McDONALD, J. B., NELSON, R. D., WHITE, S. B., 1990. Robust and partially adaptive estimation of regression models. The Review of Economics and Statistics, 72, 321-327.

CARROLL, R. J., WELSH, A. H., 1988. A note on asymmetry and robustness in linear regression. The American Statistician, Vol.: 42, No:4, 285-287.

COBB, L., KOPPSTEIN, P., CHEN, N. H., 1983. Estimation and moment recursion relations for multimodal distributions of the exponential family. Journal of the American Statistical Association, 78, 124-130.

CUSHNY, A. R., PEEBLES, A. R., 1905. The action of optical isomers II. Hyoscines. J. Physiol, 32, 501-510.

DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, Vol. 39, 1-39.

DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B., 1980. Iteratively reweighted least squares for linear regression when errors are normal/independent distributed. In Multivariate Analysis V, pp. 35-57, North Holland, Amsterdam-New York.

FONG, W. M., 1997. Robust beta estimation: Some empirical evidence. Review of Financial Economics, Vol. 6, 2, 167-197.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J., STAHEL, W. A., 1986. Robust statistics the approach based on influence functions. New York: Wiley.

HARDY, G. H., LITTLEWOOD, J. E., POLYA, G., 1997. Inequalities. 2 nd ed., Cambridge University Press.

HAWKINS, D. M., BRADU, D., KASS, G. V., 1984. Location of several outliers in multiple–regression data using elemental sets. Technometrics, 26, no: 3, 197-208.

HOEK, H., LUCAS, A., VAN DIJK, H. K., 1995. Classical and Bayesian aspects of robust unit root inference. Journal of Econometics, 69, no: 1, 27-59.

HUBER, P. J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics, 35, 73-101.

_____, 1981. Robust Statistics. Wiley. New York.

_____, 1984. Finite sample breakdown of M- and P-estimators. The Annals of Statistics, Vol.12, No. 1, 119-126.

JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N., 1995. Continuous Univariate Distributions. Vol. 2, 2 nd ed., New York: Wiley.

KLEIN, R., SPADY, R., 1984. Quasi-maximum likelihood as a parametric approach to robust estimation. Working paper, Bell Communication Research.

LANGE, K. L., LITTLE, J. A., TAYLOR, J. M. G., 1989. Robust statistical modeling using the t distribution. Journal of the American Statistical Association, 84, 881-896.

LANGE, K., SINSHEIMER, J. S., 1993. Normal/independent distributions and their applications in robust regression. Journal of Computational and Graphical Statistics, Vol.: 2, No: 2, 175-198.

LIM, G. C., LYE, J. N., MARTIN, G. M., MARTIN, V. L., 1998, The distribution of exchange rate returns and the pricing of currency options. Journal of International Economics, 45, 351-368.

LUCAS, A., 1997. Robustness of the Student t based M-estimator. Communications in Statistics, Series A–Theory and Methods, 26(5), 1165-1182.

LYE, J. N., MARTIN, V. L., 1993(a). Robust estimation, nonnormalities and generalized exponential distributions. Journal of the American Statistical Association, 88, 261-267.

LYE, J. N., MARTIN, V. L., 1993(b). A flexible parametric density estimator for multimodal distributions of test statistics. Communications in Statistics, Series A–Theory and Methods, 22(3), 813-830.

MAKELAINEN, T., SCHMIDT, K., STYAN, G. P. H., 1981. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. Annals of Statistics, 9, 758-767.

McDONALD, J. B., BUTLER, R. J., 1987. Some generalized mixture distributions with an application to unemployment duration. The Review of Economics and Statistics, 69, 232-240.

McDONALD, J. B., NEWEY, W. K., 1988. Partially adaptive estimation of regression models via the generalized t distribution. Econometric Theory, 4, 428-457.

McDONALD, J. B., NELSON, R. D., 1989. Alternative beta estimation for the market model using partially adaptive techniques. Communications in Statistics, Series A–Theory and Methods, 18(11), 4039-4058.

McDONALD, J. B., 1989. Partially adaptive estimation of ARMA time series models. International Journal of Forecasting, 5, 217-230.

McDONALD, J. B., NELSON, R. D., 1993. Beta estimation in the market model : skewness and leptokurtosis. Communications in Statistics, Series A–Theory and Methods, 22(10), 2843-2862.

MORGENTHALER, S., 1990. Fitting redescending M-estimators in regression. (K. D. Lawrence and J. L. Arthur, Eds.). Robust regression: analysis and applications. Dekker, New York, 105-128.

NAGAHARA, Y., 1999. The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. Statistics and Probability Letters, 43, 251-264.

OLIVER, E. H., 1972. A maximum likelihood oddity. The American Statistician, 26(3), 43-44.

PETRUCCELLI, J. D., NANDRAM, B., CHEN, M., 1999. Applied statistics for engineers and scientists, Prentice Hall.

ROHATGI, V. K., 1976. An introduction to probability theory and mathematical statistics. New York: Wiley.

ROSNER, B., 1975. On the detection of many outliers. Technometrics, 17, 221-227.

ROUSSEEUW, P. J., LEROY, A. M., 1987. Robust regression and outlier detection. John Wiley and Sons, New York.

STAUDTE, R. G., SHEATHER, S. J., 1990. Robust estimation and testing. John Wiley and Sons, New York.

THEODOSSIOU, P., 1998. Financial data and the skewed generalized t distribution. Management Science, Part 1, 12, 1650-1661.

TIKU, M. L., SURESH, R. P., 1992. A new method of estimation for location and scale parameters. Journal of Statistical Planning and Inference, 30, 281-292.

TYLER, D. E., 1986. Breakdown properties of the M-estimators of multivariate scatter. Technical report, Dept. Of Statistics, Rutgers University, New Jersey.

VAUGHAN, D. C., 1992. On the Tiku-Suresh method of estimation. Communications in Statistics, Series A–Theory and Methods, 21(2), 451-469.

YUH, L., HOGG, R. V., 1988. On adaptive M-regression. Biometrics, 44, 433-445.

ZECKHAUSER, R., THOMPSON, M., 1970. Linear regression with non-normal error terms. The Review of Economics and Statistics, 52, 280-286.

S-Plus Program for the Regression Method GT(p=1.5,q=.2)

```
y<-matrix(c(data.matrix(stackloss[1:21,4:4])),ncol=1)
x<-matrix(c(rep(1,21),data.matrix(stackloss[1:21,1:3])),ncol=4)
x1<-matrix(c(data.matrix(stackloss[1:21,1:3])),ncol=3)
data<-c(x1,y)
p<-1.5
q<-.2
old.beta<-matrix(c(-39.92,.72,1.30,-.15),ncol=1)
old.sigma<-sqrt(sum((data-mean(data))^2)/84)
n<-21
it<-0
repeat{
    k1<-matrix(0,21,4)
    k2<-matrix(0,1,16)
    wt<-((p*q+1)*((abs((y-x%*%old.beta)/old.sigma))^(p-2)))/
    (q+(abs((y-x%*%old.beta)/old.sigma))^p)
    i<-1
    while(i<22){
    add1<-(wt[i,]*y[i,]*as.matrix(x[i,],ncol=1))
    k1[i,]<-add1
    i<-i+1
    }
    a1<-matrix(colSums(k1),ncol=1)
    j<-1
    while(j<22){
        add2<-as.vector(wt[j,]*as.matrix(x[j,],ncol=1)%*%
        t(as.matrix(x[j,],ncol=1)))
        k2<-rbind(k2,add2)
        j<-j+1
    }
    a2<-matrix(colSums(k2),ncol=4)
    new.beta<-solve(a2)%*%a1
    new.sigma<-sqrt(sum(wt*((y-x%*%new.beta)^2))/n)
    lik<-n*log((p*(q^q)*gamma(q+1/p))/(2*gamma(1/p)
    *gamma(q)),base=exp(1))-n*log(new.sigma,base=exp(1))
    -(q+1/p)*sum(log(q+(abs(y-x%*%new.beta)
    /new.sigma)^p,base=exp(1)))
it<-it+1
if(it>80)
        break
        old.beta<-new.beta
        old.sigma<-new.sigma
        cat(it,new.beta,new.sigma,lik,"\n")
}
```

S-Plus Program for the Standard Errors of the Regression Method GT(p=1.5,q=.2)

```
y<-matrix(c(data.matrix(stackloss[1:21,4:4])),ncol=1)
x<-matrix(c(rep(1,21),data.matrix(stackloss[1:21,1:3])),ncol=4)
n<-21
pp<-4
p<-1.5
q<-.2
beta<-matrix(c(-39.94,.83,.57,-.06),ncol=1)
sigma<-.09
i<-1
r<-matrix(0,1,n)
inv<-solve(t(x)%*%x)
while(i<n+1){
    z<-y[i,]-x[i,] %*% beta
    r<-replace(r,i,z)
    i<-i+1
}
s<-median(abs(r))/.6745
w<-r/s
psisqr<-(((p*q+1)*(abs(w)^(p-1))*sign(w))/(q*(sigma^p)+abs(w)^p))^2
psider<-(p*q+1)*((p-1)*q*(sigma^p)*(abs(w)^(p-2))-abs(w)^(2*p-
    2))/((q*(sigma^p)+abs(w)^p)^2)
a<-(1/n)*sum(psider)
b<-(1/(n-pp))*sum(psisqr)
lambda<-1+(pp/n)*((1-a)/a)
varcov<-b*lambda*((s/a)^2)*inv
stderr<-sqrt(diag(varcov))
stderr
```

RESUME

I was born in 1973 in Erzurum. I graduated in mathematics form the Çukurova University and started at the position of research assistantship in Niğde University in 1996. I got an MSc degree in pure mathematics from the Çukurova University in 1998. At the same year I started the PhD program for statistics there. To complete my education in the name of Niğde University I was appointed temporarily to the same job by Higher Education Council.