Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

Inference From Stable Distributions

Hammou El Barmi and Paul I. Nelson
Kansas State University

ABSTRACT

We consider linear regression models of the form $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where the components of the error term have symmetric stable ($S\alpha S$) distributions centered at zero with index of stability α in the interval (0,2). The tails of these distributions get progressively heavier as α decreases and their densities have known closed form expressions in only two special cases: $\alpha = 2$ corresponds to the normal distribution and $\alpha = 1$ to the Cauchy distribution. The $S\alpha S$ family of distributions has moments of order less than α. Therefore, for $\alpha \leq 1$, the components of $X\boldsymbol{\beta}$ are viewed as location parameters. The usual theory of optimal estimating functions does not apply since variances of the components of $\mathbf{Y}$ are not finite. We study the behavior of estimators of $\boldsymbol{\beta}$ based on 3 types of estimating equations: (1) least squares, (2) maximum likelihood and (3) optimal norm. The score function from these stable models can also be used to consistently estimate $\boldsymbol{\beta}$ for a general class of variance mixture error models.

Key Words: Stable distribution, regression, estimating function, consistency, constrained minimization, variance mixture.

1 Introduction

Statistical analyses of regression type models of the form

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.1}$$

typically assume that the error terms have independent normal distributions with common variance and that the components of the full rank design matrix X are constants. Here, we generalize and allow the components of $\boldsymbol{\epsilon}$ to have independent symmetric stable distributions with infinite variance. A stable distribution symmetric about μ has a log-characteristic function of the form

$$\psi(t) = -|\sigma t|^{\alpha} + i\mu t, \tag{1.2}$$

where $\sigma \geq 0$ is a scale parameter and α is called the index of stability, $0 < \alpha \leq 2$. Affine transformations of independent copies of a stable random variable also have stable distributions. This closure property has proved useful in applications to economics and astronomy. The tails of these distributions get progressively heavier as α decreases. Except for $\alpha = 2$, stable distributions only have moments of order less than α. The normal distribution with variance $2\sigma^2$ corresponds to $\alpha = 2$ and the Cauchy distribution to $\alpha = 1$. For $\alpha < 1$, the components of $X\boldsymbol{\beta}$ should be viewed as location parameters. We use $V \sim S\alpha S(\sigma)$ (called symmetric alpha stable) to denote that V has the distribution given in (1.2) with $\mu = 0$. The recent text by Samorodnitsky and Taqqu(1994) is an excellent source of information on stable distributions and processes. We will need one special type of skewed stable distribution with index of stability $\delta < 1$, having log-characteristic function of the form:

$$\psi(t) = -\sigma^\delta |t|^\delta (1 - isgn(t)Tan(\pi\delta/2)), \tag{1.3}$$

where $sgn(t)$ denotes the sign of t. Random variables having such distributions are supported on the positive axis and called stable subordinators of index δ with scale parameter σ. From (1.2) and (1.3) it follows that if $V \sim S\alpha S(\sigma), \alpha < 2$, then, in distribution,

$$V = \sqrt{A}Z, \tag{1.4}$$

where $Z \sim N(0, 2\sigma^2)$, A is a stable subordinator of index $\alpha/2$ having scale parameter $cos(\pi\alpha/4)$ and A is independent of Z. See Samorodnitsky and Taqqu(1994). Thus, every $S\alpha S$ distribution is a variance mixture of normals and in particular we take the components $\{\epsilon_i\}$ of the error term in (1.1) to be independent $S\alpha S(1)$ random variables. The representation of a stable subordinator A given on page 29 of Samorodnitsky and Taqqu(1994) leads to $E(1/A^k) < \infty$, for $k > -\alpha/2$.

A method for estimating the $p \times 1$ vector of parameters $\boldsymbol{\beta}$ when the error terms $\{\epsilon_i\}$ do not have finite variances based on transforming the observations $\mathbf{Y}$ into bounded complex random variables $\exp(ity_j), j = 1, 2, ..., n$, is given in Chambers and Heathcote(1981) and Paulson and Delehanty(1985). Also see Merkouris(1991) and McLeish and Small(1991). Here, we investigate the performance of estimators obtained from three estimating equations: (1) maximum likelihood, (2) minimum α norm, (3) least squares. Our analyses assume that the index of stability α is known. The value of σ is not needed to estimate the regression parameters $\boldsymbol{\beta}$, but would be to construct confidence intervals. In practice both α and σ could be iteratively estimated from residuals. Our simulation study, presented in Section 5, illustrates how this can easily be done for α. The representation given in (1.4) leads to the observation that the form of the score function for $S\alpha S$ errors can also be used to consistently estimate $\boldsymbol{\beta}$ for a general class of error distributions modeled as variance mixtures.

We use boldface to denote column vectors, $\mathbf{v}^T$ to indicate the transpose of $\mathbf{v}$ and $|\mathbf{v}|$ its length. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ denote the *ith* column of the transpose of the design matrix X, $i = 1, 2, \ldots, n$. Let P denote the underlying probability and $P_{\boldsymbol{\beta}}$ the probability generated by shifting the random vector ϵ by an amount $X\boldsymbol{\beta}$. Unsubscripted probabilities and expectations are with respect to P. Note that $P \equiv P_0$. We follow the common practice of omitting the qualifier "a.e." between almost every where equal random variables when context allows.

2 Maximum Likelihood

The lack of closed form expressions for their probability densities makes the use of maximum likelihood with stable distributions very difficult. Zolotarev (1966) provides an integral representation of symmetric stable densities which was used by Brorsen and Yang(1990) to find the maximum likelihood estimates (mle's) of the parameters α, σ and μ. DuMouchel (1971) uses a multinomial approximation to the likelihood equation to estimate these parameters. Feuerverger and McDunnough(1981) employ a fast Fourier transform of the empirical characteristic function to obtain an approximate likelihood.

Here, we allow the location parameter μ to depend on covariates and use the score function directly to develop asymptotic properties of the mle of the vector of regression parameters $\boldsymbol{\beta}$. Let the components of the error term $\boldsymbol{\epsilon}$ given in (1.1) be independent $S\alpha S(1), 0 < \alpha < 2$. From (1.4) we have that:

$$\epsilon_i = Z_i \sqrt{A_i}, \tag{2.1}$$

$i = 1, 2, ..., n$, where the stable subordinators $\{A_i\}$ of index $\alpha/2$ and the mean zero normal random variables $\{Z_i\}$ with variance 2 are all jointly independent. The probability density f of Y_i in (1.1) and the likelihood $L_n(\boldsymbol{\beta})$ are given by:

$$f(y_i|\boldsymbol{\beta}) = \int_0^\infty (2a)^{-1/2} \phi((y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sqrt{2a}) g(a) da, \tag{2.2}$$

$$L_n(\boldsymbol{\beta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\beta}),$$

where g is the density of a stable subordinator as described in (1.3) with $\delta = \alpha/2$, $\sigma = cos(\pi\alpha/4)$ and ϕ is the standard normal density. From Proposition 1.31 of Samorodnitsky and Taqqu(1994), $f(y|\boldsymbol{\beta})$ is a Cauchy density for $\alpha = 1$. Hence, $f(y|\boldsymbol{\beta})$, which is a mixture of log-concave functions of $\boldsymbol{\beta}$, is not necessarily log- concave as a function of $\boldsymbol{\beta}$. However, we give conditions in

Theorem 2.1 which ensure that as $n \to \infty, L_n(\boldsymbol{\beta})$ a.e. has a local maximum in every neighborhood of the true regression parameter. Since $E(1/A^{1/2})$ is finite and ϕ is bounded, differentiation with respect to the components of $\boldsymbol{\beta}$ can be passed through the integral in (2.2). The score function $l_n(\boldsymbol{\beta})$ then has the form of a weighted least squares estimating function:

$$l_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} x_i(y_i - \mathbf{x}_i^T\boldsymbol{\beta})w(y_i - \mathbf{x}_i^T\boldsymbol{\beta})/2, \tag{2.3}$$

where the weights are a.e. $P_{\boldsymbol{\beta}}$ given in terms of conditional expectations of the form:

$$\begin{aligned} w(y_i - \mathbf{x}_i^T\boldsymbol{\beta}) &= (1/\sqrt{2})\int_0^\infty a^{-1.5}\phi((y_i - \mathbf{x}_i^T\boldsymbol{\beta})/\sqrt{2a})g(a)da/f(y_i|\boldsymbol{\beta}) \\ &= E_{\boldsymbol{\beta}}(1/A_i|y_i - \mathbf{x}_i^T\boldsymbol{\beta}). \end{aligned} \tag{2.4}$$

In a simulation study summarized in Section 5 we were able to effectively find roots of $l_n(\boldsymbol{\beta})$ by using Monte Carlo Integration to approximate the likelihood function $L_n(\boldsymbol{\beta})$ and a grid search to find its local maximum. This process avoids the more difficult task of computing and finding the root of the score function.

The score function given (2.3) can also be used to form an estimating function for a general class of variance mixture models. Suppose that $\{Z_i\}$ in (2.1) are i.i.d. according to a baseline location-scale family with continuous pdf ϕ^* having mean zero and finite variance, say variance = 2 to be in conformity with the stable case. Further assume that $\{A_i\}$ are i.i.d. according to any distribution with pdf g^* supported on the positive axis such that (2.4) exists. Let $G_n(\boldsymbol{\beta})$ represent $l_n(\boldsymbol{\beta})$ as given in (2.3) with ϕ replaced by ϕ^* and g replaced by g^* :

$$G_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}_i^T\boldsymbol{\beta})E^*_{\boldsymbol{\beta}}(1/A_i|y_i - \mathbf{x}_i^T\boldsymbol{\beta})/2, \tag{2.5}$$

where "$E^*_{\boldsymbol{\beta}}(.|.)$" denotes conditional expectation under ϕ^* and g^*. The estimating function $G_n(\boldsymbol{\beta})$ may be motivated as follows. Conditional on $\{A_i\}, Q_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}^T\boldsymbol{\beta})/2A_i$ is a Godambe optimal estimating function and hence optimality holds unconditionally, if second moments are finite. However, the variables $\{A_i\}$ are not observable so that $Q_n(\boldsymbol{\beta})$ cannot be used to estimate $\boldsymbol{\beta}$. The estimating function $G_n(\boldsymbol{\beta})$ then results by conditioning on the data, $G_n(\boldsymbol{\beta}) = E(Q_n(\boldsymbol{\beta})|y_i, i = 1, 2, ..., n)$. In Theorems 2.1 and 2.2 we give conditions under which roots of $G_n(\boldsymbol{\beta}) = \mathbf{0}$ yield consistent estimators of $\boldsymbol{\beta}$. We now delete the superscript $*$ from the expectation operator.

To simplify the derivation of asymptotic properties of roots of $G_n(\boldsymbol{\beta}) = \mathbf{0}$, assume without loss of generality that the true $\boldsymbol{\beta} = 0$. Using a result given in Silvey and Aitchison(1957), the asymptotic existence and strong consistency as $n \to \infty$ of roots of (2.5) hold under conditions which guarantee that for all sufficiently small $\delta > 0$, a.e. P,

$$\limsup_{n\to\infty}\{\sup \boldsymbol{\beta}^T G_n(\boldsymbol{\beta}), |\boldsymbol{\beta}| = \delta\} < 0. \tag{2.6}$$

Obtaining the uniform upper bound on $\boldsymbol{\beta}^T G_n(\boldsymbol{\beta})$ required in (2.6) can be difficult. Key tools of our approach are strong laws of large numbers. First, Neveu(1975) shows that if $\{U_i\}$ are jointly independent zero mean random variables with finite variances $\{\sigma_i^2\}$, then $S_n \equiv \sum_{i=1}^{n} \sigma_i^2 \to \infty$ implies that a.e.:

$$\sum_{i=1}^{n} U_i/h(S_n) \to 0, \tag{2.7}$$

for h a nondecreasing positive function with:

$$\int_0^\infty 1/(1 + h(t))^2 dt < \infty. \tag{2.8}$$

Chung(1974, page 130) states that if the random variables $\{U_i\}$ are i.i.d with $E(U_i) = 0$, then for any sequence of uniformly bounded constants $\{c_n\}$, .a.e.

$$\sum_{i=1}^{n} c_i U_i/n \to 0. \tag{2.9}$$

Theorem 2.1 on consistency will show, for example, that by taking $h(t) = t^q$, $0.5 < q < 1$, asymptotically a strongly consistent root $\hat{\boldsymbol{\beta}}_n$ of $Gn(\boldsymbol{\beta})$ exists a.e. if for a positive definite matrix Σ and a constant c, as $n \to \infty$,

$$X^T X/n \to \Sigma \tag{2.10}$$

and

$$\sum_{j=1}^{p} \sum_{j=1}^{n} x_{ij}^4/n \to c.$$

Theorem 2.1 *Let $e_n(min)$ and $e_n(max)$ denote the minimum and maximum eigenvalues of $X^T X$ and let $h_1(t)$ and $h_2(t)$ be functions of the type given in (2.8). Then, asymptotically a strongly consistent root of $G_n(\boldsymbol{\beta})$ exists a.e. P, if as $n \to \infty$,*

$$e_n(min) \to \infty, \tag{2.11}$$

$$\rho \equiv \limsup_{n\to\infty}(h_1(e_n(max))/e_n(min)) < \infty,$$

$$\limsup_{n\to\infty}(h_2(\max\{\sum_{i=1}^{n} x_{ij}^4, j \le p\})/e_n(min)) < \infty,$$

$$E(1/A_1^2) < \infty.$$

If in addition (2.10) holds, letting $\hat{\boldsymbol{\beta}}_n$ denote the consistent root, we have that as $n \to \infty, |\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}| = O_p(1/\sqrt{e_n(min)})$.

Proof. Let $|\boldsymbol{\beta}| = \delta > 0$. From (2.5) we have that:

$$\begin{aligned} 2\boldsymbol{\beta}^T G_n(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \boldsymbol{\beta}^T \mathbf{x}_i y_i E(1/A_i|\epsilon_i) - \sum_{i=1}^{n} \boldsymbol{\beta}^T \mathbf{x}_i \mathbf{x}_i^T \boldsymbol{\beta} E(1/A_i|\epsilon_i) \quad (2.12) \\ &\equiv C_n - D_n. \end{aligned}$$

Under P_0, C_n is a sum of independent, zero mean random variables with $\sigma_n^2 \equiv$Variance$(C_n) = 2E(E^2(1/A_1|\epsilon_1))\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \equiv \tau \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \ge \tau\delta^2 e_n(min)$ $\to \infty$ by hypothesis, where $\tau = 2E(E^2(1/A_1|\epsilon_1))$. From the second line of (2.11), we then have that a.e., for large n,

$$\begin{aligned} |C_n|/\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \le |C_n|/\delta^2 e_n(min) &\le 2\rho|C_n|/\delta^2 h_1(e_n(max)) \\ &\le 2\rho|C_n|/\delta^2 h_1(\sigma_n^2/\tau\delta^2). \end{aligned}$$

Since C_n is linear in $\boldsymbol{\beta}$, from (2.7) as $n \to \infty$, we have that uniformly in $\boldsymbol{\beta}, |\boldsymbol{\beta}| = \delta, C_n/\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \to 0$ a.e. The second quantity on the last line of (2.12) can be written as $D_n = \boldsymbol{\beta}^T \sum_{i=1}^{n} U_i \boldsymbol{\beta} + E(1/A_1)\boldsymbol{\beta}^T X^T X \boldsymbol{\beta}$, where the $p \times p$ matrix $U_i \equiv (U_i(j,k))$ with $U_i(j,k) = x_{ij}x_{ik}[E(1/A_i|\epsilon_i) - E(1/A_i)], j,k = 1,2,\ldots,p$. For each j,k pair, $W_n(j,k) \equiv \sum_{i=1}^{n} U_i(j,k)$ is a weighted sum of jointly independent, mean zero random variables with Variance$(W_n(j,k)) = \gamma\sum_{i=1}^{n} x_{ij}^2 x_{ik}^2 \le \gamma[\sum_{i=1}^{n} x_{ij}^4 \sum_{i=1}^{n} x_{ik}^4]^{1/2} \le \gamma \max\{\sum_{i=1}^{n} x_{im}^4, m \le p\}, \gamma =$ Variance $(E(1/A_1|\epsilon_1))$. If $\sigma_n^2(j,k) \equiv$ Variance$(W_n(j,k)/\sqrt{\gamma})$ is bounded in n, $\{W_n(j,k)/\sqrt{\gamma}\}$ converges a.e. to a finite limit as $n \to \infty$. Hence, from (2.7) and (2.11),

$$\{W_n(j,k)\}/e_n(min) = \{h_2(\sigma_n^2(j,k))W_n(j,k)\}/[h_2(\sigma_n^2(j,k))e_n(min)] \to 0$$

a.e. as $n \to \infty, j,k = 1,2,\ldots,p$. It then follows that

$$\boldsymbol{\beta}^T \sum_{i=1}^{n} U_i \boldsymbol{\beta}/\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \le \boldsymbol{\beta}^T \sum_{i=1}^{n} U_i \boldsymbol{\beta}/\delta^2 e_n(min) \to 0$$

a.e., uniformly in $\boldsymbol{\beta}$, $|\boldsymbol{\beta}| = \delta$. Finally, uniformly in $\boldsymbol{\beta}$, $\boldsymbol{\beta}^T G_n(\boldsymbol{\beta})$ is a.e. asymptotically equivalent to $\boldsymbol{\beta}^T X^T X \boldsymbol{\beta}(-E(1/A_1)+o(1)) \leq \delta^2 e_n(min)((-E(1/A_1)+o(1)) < 0$, which using (2.6) completes the proof of consistency. The bound on the rate of convergence is obtained by taking $\delta = 1/\sqrt{e_n(min)}$.

Chung's (1974) strong law as given in (2.9) allows a weakening of the conditions in (2.11) at the expense of placing a uniform bound on the elements of the design matrix and requiring that $e_n(min)/n$ converge to a positive constant.

Theorem 2.2 *Let* $E(1/A_1) < \infty, e_n(min)/n \to c$, *a positive constant, and* $\sup\{|x_{ij}(n)|, i, j \leq p, n = 1, 2, \ldots\ldots\} \leq M < \infty$, *where* $x_{ij}(n)$ *is the element in row* i, *column* j *of the design matrix* X *based on* n *observations. Then, a.e there is asymptotically a consistent root of* $G_n(\boldsymbol{\beta})$.

Proof:
The proof parallels the one given for Theorem 2.1 and is omitted.

Now, assume that $E(1/A_1^2) < \infty$. Let $\dot{G}_n(\boldsymbol{\beta})$ denote the matrix of partial derivatives of $G_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and suppose that $E(\dot{G}_n) = -E(G_n G_n^T)$, which is the case when $G_n(\boldsymbol{\beta})$ is the score function $l_n(\boldsymbol{\beta})$. The "information" matrix $J(n, \alpha)$ of n is then given by:

$$\begin{aligned} J(n,\alpha) &= X^T X E(\epsilon_1^2 E^2(1/A_1|\epsilon_1))/4 \\ &\equiv X^T X v(\alpha). \end{aligned} \tag{2.13}$$

If the second order term in the expansion of $G_n(\boldsymbol{\beta})$ is suitably well behaved (a matter we have not been able to resolve), it follows from the Linderberg-Feller central limit theorem that if a consistent root of $G_n(\boldsymbol{\beta})$ exists and $\max\{|x_i|, i \leq n\}/e_n(min) \to 0$, we then have for large n, approximately,

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (X^T X)^{-1}/v(\alpha)) \tag{2.14}$$

Finally, if (2.10) holds, $l_n(\boldsymbol{\beta})/\sqrt{n}$ will converge weakly to a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $v(\alpha)\Sigma(\boldsymbol{\beta})$. Therefore, a test for $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ can be based on $T_n = l_n^T(\boldsymbol{\beta}_0)\Sigma^{-1}(\boldsymbol{\beta}_0)l_n(\boldsymbol{\beta}_0)/nv(\alpha)$, which has asymptotically a chi-square distribution with p degrees of freedom under H_0.

3 Minimum α Norm Estimators

Here, we seek estimators of linear compounds $\gamma = \boldsymbol{\lambda}^T \boldsymbol{\beta}$ of the form $\hat{\gamma}(\mathbf{c}) = \mathbf{c}^T \mathbf{Y}$, for a vector of constants $\mathbf{c}$, which are unbiased for γ, so that:

$$\mathbf{c}^T X = \boldsymbol{\lambda}^T, \tag{3.1}$$

and "close" to γ. Instead of using the variance of $\hat{\gamma}$, which is not finite, we define $\hat{\gamma}$ to be close to γ if the scale parameter of $\hat{\gamma}-\gamma$ is small. Blattberg and Sargent(1971) introduced this concept for the one predictor case ($p=1$).

To further develop and extend these minimum α norm estimators, we first briefly describe what is called the covariation between jointly α stable, symmetric random variables. See Samorodnitsky and Taqqu(1994) for a full treatment of this concept.

For $\alpha \in (0,2]$, the random variables $\mathbf{U}=\{U_i, i=1,2,\ldots,n\}$ are said to be jointly symmetric α stable, denoted $S\alpha S$, if their log joint characteristic function is given by:

$$\psi(t) = -\int_S |\mathbf{t}^T\mathbf{s}|^\alpha \Gamma(d\mathbf{s}), \tag{3.2}$$

where Γ is a finite measure, called the spectral measure, on the surface of the unit n-sphere S centered at the origin in $\mathcal{R}^n$. For $\mathbf{U}$ having the distribution specified by (3.2), $1<\alpha\leq 2$, define the covariation of U_i on U_j by $[U_i,U_j]_\alpha = \int_S s_i s_j^{<\alpha-1>}\Gamma(ds)$, where $x^{<q>} = sgn(x)|x|^q$. Covariation is not in general symmetric; $[aU_i,bU_j]_\alpha = ab^{<\alpha-1>}[U_i,U_j]$ and for $\alpha=2, [U_i,U_j]_2 = \text{Covariance}(U_i,U_j)/2$. Covariation leads to the α norm $||U||_\alpha$ of a scalar $S\alpha S(\sigma)$ random variable U defined by $||U||_\alpha = [U,U]_\alpha^{1/\alpha}$. Note that $||cU||_\alpha = |c|||U||_\alpha$ and that $||U||_\alpha = 0$ if and only if $U=0$ a.e. If the independent random variables $U_i \sim S\alpha S(\sigma), i=1,2,...,n$, then, for a vector of constants $\mathbf{c}$, we have that $||\mathbf{c}'\mathbf{U}||_\alpha = (\sum_{i=1}^{n}|c_i\sigma_i|^\alpha)^{1/\alpha}$. Thus, for $\mathbf{c}$ satisfying (3.1), $||\hat{\gamma}(\mathbf{c})-\gamma||_\alpha^\alpha = \sum_{i=1}^{n}|c_i|^\alpha$.

Now, for $1<\alpha\leq 2$, define an estimator $\hat{\gamma}(\mathbf{b}^*)$ to be best linear unbiased α-norm (BLUαN) estimator of $\gamma=\boldsymbol{\lambda}^T\boldsymbol{\beta}$ if $\mathbf{b}^*$ satisfies (3.1) and $||\hat{\gamma}(\mathbf{b}^*)-\gamma||_\alpha \leq ||\gamma(\mathbf{c})-\gamma||_\alpha$ for all vectors $\mathbf{c}$ that satisfy (3.1). For $\alpha=2$, BLU2N and BLUE (best linear unbiased estimation) are identical concepts. However, for $\alpha<2$, it is not even necessarily so that $\hat{\gamma}(b^*) = \sum_{i=1}^{n}\lambda_i\hat{\beta}_i$, where $\hat{\beta}_i$ are BLUαN for $\{\beta_i\}$. Unlike the mle, the BLUαN is defined only for α in (1,2]. A method for computing the BLUαN will be given below.

In the scalar case, where β is a single unknown parameter, the BLUαN may be viewed as the solution of an optimal estimating equation in the following sense. Consider the family of estimating functions of the form $G(\mathbf{k},\beta) = \sum_{i=1}^{n}(y_i - x_i\beta)k_i$. Call $G(\mathbf{k}^*,\beta)$ optimal if within this class it minimizes $||G(k,\beta)||_\alpha^\alpha/|E(\partial(G(k,\beta)/\partial\beta)|^\alpha = \sum_{i=1}^{k}|k_i|^\alpha/|\sum_{i=1}^{n}k_ix_i|^\alpha$. This is equivalent to minimizing $\sum_{i=1}^{n}|k_i|^\alpha$ subject to $\sum_{i=1}^{n}k_ix_i = 1$, which leads to the BLUαN.

The concept of James orthogonality can be used to characterize BLUαN estimators. For two jointly symmetric stable random variables U and V, $1 < \alpha < 2$, V is said to be James orthogonal to U if for all real τ, $||\tau U+V||_\alpha \geq ||V||_\alpha$. Samorodnitsky and Taqqu(1994) prove that V is James orthogonal to U if and only if $[U, V]_\alpha = 0$. Now, let $\hat{\gamma}(\mathbf{b})$ and $\hat{\gamma}(\mathbf{c})$ be two linear unbiased estimators of γ as described above. Taking $U = \hat{\gamma}\mathbf{c}) - \hat{\gamma}(\mathbf{b})$ and $V = \hat{\gamma}(\mathbf{b}) - \gamma$, we have that $||\hat{\gamma}(\mathbf{c}) - \gamma||_\alpha = ||U+V||_\alpha \geq ||\hat{\gamma}(\mathbf{b}) - \gamma||_\alpha$ if $[U, V]_\alpha = 0$. Since the components of $\boldsymbol{\epsilon}$ are iid $S\alpha S(1)$, letting $\mathbf{k} = \mathbf{c} - \mathbf{b}$, $[U, V]_\alpha = [\mathbf{k}^T\boldsymbol{\epsilon}, \mathbf{b}^T\boldsymbol{\epsilon}]_\alpha = \sum_{i=1}^{n}[k_i\epsilon_i, b_i\epsilon_i]_\alpha = \sum_{i=1}^{n}k_ib_i^{<\alpha-1>} = 0$ holds if:

$$\sum_{i=1}^{n}|b_i|^\alpha = \sum_{i=1}^{n}c_ib_i^{<\alpha-1>}. \tag{3.3}$$

Thus, $\hat{\gamma}(\mathbf{b}^*)$ is BLUαN if $\mathbf{b} = \mathbf{b}^*$ satisfies (3.1) and (3.3) holds for all $\mathbf{c}$ that satisfy (3.1), or equivalently if for all $\mathbf{k}$ in the null space of X^T, $\sum_{i=1}^{n}k_i(b_i^*)^{<\alpha-1>} = 0$.

Finding the BLUαN of $\gamma = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ requires obtaining that vector $\mathbf{b}^*$ which minimizes $\sum_{i=1}^{n}|b_i|^\alpha$ subject to the constraint given in (3.1). In addition to using (3.3), this can be accomplished by using a Fenchel duality type theorem (Rockafeller(1970)) to characterize $\mathbf{b}^*$. For $\mathbf{b}$ satisfying (3.1), let $\mathbf{z} = \mathbf{b} - \mathbf{z}_0$, where $\mathbf{z}_0 = X(X^TX)^{-1}\boldsymbol{\lambda} = (z_{01}, z_{02}, \ldots, z_{0n})^T$ and define the convex function $f_\alpha(\mathbf{z})$ by:

$$f_\alpha(\mathbf{z}) = \sum_{i=1}^{n}|z_i + z_{0i}|^\alpha/\alpha. \tag{3.4}$$

It then follows that $\mathbf{b}^* = \mathbf{z}^* + \mathbf{z}_0$, where $\mathbf{z}^*$ minimizes $f_\alpha(\mathbf{z})$ subject to $\mathbf{z} \in N(X^T)$, the null space of X^T. To find $\mathbf{z}^*$, consider the convex conjugate of f_α given by (Rockafellar(1970)):

$$f_\alpha^*(\mathbf{y}) = \sup[\sum_{i=1}^{n}z_iy_i - f_\alpha(\mathbf{z}), \mathbf{z} \in N(X^T)]. \tag{3.5}$$

Since for all $\mathbf{y}, \mathbf{z}$, $f_\alpha^*(\mathbf{y}) + f_\alpha(\mathbf{z}) \geq \sum_{i=1}^{n}y_iz_i$, we have that $\inf[f_\alpha(\mathbf{z}), \mathbf{z} \in N(X)] + \inf[f_\alpha^*(\mathbf{y}), \mathbf{y} \in N^+(X^T)] \geq \inf[\sum_{i=1}^{n}y_iz_i, \mathbf{z} \in N(X^T), \mathbf{y} \in N^+(X^T)] = 0$, where

$N^+(X^T)$ is the orthogonal complement of $N(X^T)$. Therefore, if we can find $\mathbf{y}^* \in N^+(X^T)$ such that $f_\alpha(\mathbf{z}^*) + f_\alpha^*(\mathbf{y}^*) \leq 0$, then $\mathbf{z}^*$ and $\mathbf{y}^*$ must respectively solve the problems of finding the infima of $[f_\alpha(\mathbf{z}), \mathbf{z} \in N(X^T)]$, called the primal problem, and of $[f_\alpha^*(\mathbf{y}), \mathbf{y} \in N^+(X^T)]$, called the dual problem. Specifically we would have $f_\alpha(\mathbf{z}^*) = \inf[f_\alpha(\mathbf{z}), \mathbf{z} \in N(X^T)] = -f_\alpha^*(\mathbf{y}^*) = \inf[f_\alpha^*(\mathbf{y}), \mathbf{y} \in N^+(X^T)]$. The values $\mathbf{z}^*, \mathbf{y}^*$ and $\mathbf{b}^*$ can be found with the aid of the following lemma.

Lemma 3.1 *The convex conjugate in (3.4) can be expressed as:*

$$f_\alpha^*(\mathbf{y}) = -\sum_{i=1}^{n} z_{0i} y_i + [(\alpha-1)/\alpha] \sum_{i=1}^{n} |y_i|^{\alpha/(\alpha-1)}. \tag{3.6}$$

Moreover, $\mathbf{y}^*, \mathbf{z}^*$ *and* $\mathbf{b}^*$ *must satisfy:*

$$y_i^* = (z_i^* + z_{0i})^{<\alpha-1>} = b_i^{*<\alpha-1>}, i = 1, 2, \ldots, n, \tag{3.7}$$

where $a^{<q>}$ *is as defined above.*

The proof of Lemma 3.1 follows from Luenberger(1969, p 196) and is omitted.

Finally, since $N^+(X^T) = \{X\mathbf{d}, \mathbf{d} \in \mathcal{R}^p\}$, from the definition of $\mathbf{z}_0$, we have that $\mathbf{y}^* = X\mathbf{d}$, where $\mathbf{d}$ achieves the minimization:

$$\inf[-\boldsymbol{\lambda}^T \mathbf{d} + [(\alpha-1)/\alpha] \sum_{i=1}^{n} |\mathbf{x}_i^T \mathbf{d}|^{\alpha/(\alpha-1)}, \mathbf{d} \in \mathcal{R}^p]. \tag{3.8}$$

A direct computation yields that $\mathbf{d}$ is the implicit solution to the system of equations,

$$\lambda_j = \sum_{i=1}^{n} x_{ij} (\sum_{i=1}^{n} x_{ij} d_j)^{<1/(\alpha-1)>}, j = 1, 2, \ldots, p. \tag{3.9}$$

The solutions to (3.9) can explicitly be found in 2 special cases. For $\alpha = 2, \mathbf{d} = (X^T X)^{-1} \boldsymbol{\lambda}$ and using (3.7) we obtain $\mathbf{b}^* = X(X^T X)^{-1} \boldsymbol{\lambda}$, the usual least squares estimator. For $p = 1, \gamma = \beta, d = (\sum_{i=1}^{n} |x_i|^{\alpha/(\alpha-1)})^{1-\alpha}$ and $b_i^* = x_i^{<1/(\alpha-1)>} / \sum_{i=1}^{n} |x_i|^{\alpha/(\alpha-1)}$.

The BLUαN of γ is optimal within the class of linear unbiased estimators in the sense of having maximal probability of being close to γ. Specifically, let $\hat{\gamma}(\mathbf{b}^*)$ be BLUαN, $\hat{\gamma}(\mathbf{c})$ be a linear unbiased estimator of γ and ϵ be a $S\alpha S(1)$ scalar random variable. For any linear unbiased estimator, $(\hat{\gamma}(\mathbf{c}) - \gamma)/[\sum_{i=1}^{n} |c_i|^\alpha]^{1/\alpha}$ is distributed as ϵ. Hence, for $\delta > 0, P(|\hat{\gamma}(c) - \gamma| \leq \delta) =$

$P(|\epsilon| \leq \delta/[\sum_{i=1}^{n}|c_i|^{\alpha}]^{1/\alpha}) = P(\hat{\gamma}(\mathbf{b}) - \gamma| \leq \delta[\sum_{i=1}^{n} b_i^{\alpha}/\sum_{i=1}^{n}|c_i|^{\alpha}]^{1/\alpha}) \leq P(|\hat{\gamma}(\mathbf{b}^*) - \gamma| \leq \delta)$.

Conditions for the weak consistency of the BLUαN $= \hat{\gamma}(\mathbf{b}^*)$, can be given in terms of the rate at which $e_n(min)$, the minimum eigenvalue of $X^T X$, diverges to infinity.

Theorem 3.1 *The BLUαN $\hat{\gamma}(\mathbf{b}^*)$ of $\gamma = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ converges to γ in probability as $n \to \infty$ if $n^{(2-\alpha)}/(e_n(min))^{\alpha} \to 0$.*

Proof:

Since $\hat{\gamma}(\mathbf{b}^*) - \gamma \sim S\alpha S([\sum_{i=1}^{n}|b_i^*|^{\alpha}]^{1/\alpha})$, it suffices to show that $\sum_{i=1}^{n}|b_i^*|^{\alpha} \to 0$. The least squares estimator of γ is given by $\hat{\gamma}(\mathbf{d})$ for $\mathbf{d}^T = \boldsymbol{\lambda}^T(X^T X)^{-1}X^T$, which satisfies (3.1). Hence, since $\hat{\gamma}(\mathbf{b}^*)$ is BUαN and the usual L_p norms are non-decreasing in p,

$$\sum_{i=1}^{n}|b_i^*|^{\alpha} \leq \sum_{i=1}^{n}|d_i|^{\alpha} \leq n^{(2-\alpha)/2}[\sum_{i=1}^{n} d_i^2]^{\alpha/2}$$

$$= n^{(2-\alpha)/2}[\boldsymbol{\lambda}^T(X^T X)^{-1}\boldsymbol{\lambda}]^{\alpha/2} \leq |\boldsymbol{\lambda}|^{\alpha} n^{(2-\alpha)/2}/(e_n(min))^{\alpha/2} \to 0,$$

by hypothesis, which completes the proof.

4 Least Squares

The usual least squares estimator (LS) of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}}_{LS} = (X^T X)^{-1}X^T\mathbf{Y}$, which has the advantages of simplicity and not requiring that α be known or estimated. Our simulation study, presented below, indicates that least squares estimator performs reasonably well compared to the BLUαN for $\alpha > 1$. For $\alpha > 1, \boldsymbol{\beta}_{LS}$ is weakly consistent for $\boldsymbol{\beta}$ under the conditions of Theorem 3.1. Note that $\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}$ has a $S\alpha S$ distribution for all sample sizes n.

Least squares also plays a role in a special case of the joint symmetric α stable distributions determined by (3.3). Suppose instead of (2.1), $\epsilon_i = \sqrt{A}Z_i, i = 1, 2, \ldots, n$, where now a single stable subordinator A , distributed as specified in (1.3) with $\delta = \alpha/2$, is multiplied by all the components of $\boldsymbol{\epsilon}$. The random variable A is independent of $\{Z_i\}$. Such joint distributions are called subGaussian. By conditioning on A, it is easily seen that the least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ is the mle in this setting. Further, if we let $\hat{A} = |\mathbf{Y} - X\hat{\boldsymbol{\beta}}_{LS}|/(n-p)$, then $\gamma^T(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})/\sqrt{\hat{A}}\boldsymbol{\lambda}^T(X^T X)^{-1}\boldsymbol{\lambda}$ has a t-distribution with $n - p$ degrees of freedom ,which can be used to construct confidence intervals for $\boldsymbol{\lambda}^T\boldsymbol{\beta}$.

5 Simulation

We carried out a simulation study to investigate and compare the performance of the mle, BLUαN and least squares estimators in the one predictor case $(p = 1), y_i = x_i\beta + \epsilon_i$, where $\{\epsilon_i\}$ are iid as defined by (2.1). Samorodnitsky and Taqqu(1994) provide a formula for transforming uniform random variables into symmetric stable random variables and a series expansion in terms of gamma variates which may be used to generate stable subordinators.

We used simulation based on the law of large numbers to approximate the integral with respect to the distribution of A given in (2.1) in order to form the likelihood $L_n(\beta)$. The likelihood turned out to be quadratic in shape in a neighborhood of $\hat{\beta}$ and a grid search was effective in finding the mle. In this case where $p = 1$; there are explicit expressions for the BLUαN and the familiar least squares estimator, denoted $\hat{\beta}_{LS}$.

Starting at the least squares estimator, which doesn't require knowledge of α, we formed residuals $\hat{\epsilon}_i = y_i - x_i\hat{\beta}_{LS}, i = 1, 2, ..., n$ in order to estimate α. McCulloch(1981) gives expected values $\{m(\alpha)\}$ of the observed ratio r of specified spacings of iid symmetric α stable random variables expressed as functions of α. Inverting the tabled values $\{m(\alpha)\}$, for $\alpha \geq 0.5$, based on an observed r obtained from the residuals $\{\hat{\epsilon}_i\}$, provided a quick moderately effective way to estimate α. Initially we iterated this scheme, but abandoned this refinement because it was time consuming and did not yield significantly better results. More sophisticated methods for estimating α from iid observations are available. See Arad(1980) for example. Additional study of the problem of estimating α in the context of regression models needs to be carried out.

We generated the predictor variables $\{x_i\}$ from a uniform distribution on the interval (0,5). We simulated data with n = 10, 20 and 50 and several values of α in order to cover a wide range of possible parameter settings. Since the BLUαN is not defined for $\alpha \leq 1$, whenever the estimated α was less than or equal to 1 we used $\alpha = 1.05$ in the formula for computing the BLUαN.

We first assess the performance of the 3 estimators in terms of their estimated mean. Table 5.1 contains (rounded to 2 places) sample means and sample mean absolute deviations, $MAD = \sum |\hat{\beta}_i - \beta|/1000$, of the three $\hat{\beta}$'s across the 1000 iterations. For $\alpha \leq 1$, the responses do not have means and the least squares and BLUαN estimators are highly unstable, sometimes swinging from plus to minus with magnitudes of several thousand. Entries where the MAD is larger than the mean are consequently omitted from Table 5.1. The last row, corresponding to $\alpha = 2$ where all 3 estimators are identical, is included to provide a basis of comparison for the other α's.

Keep in mind that all estimators (except when $\alpha = 2$) used an estimated value of α and may not retain properties such as unbiasedness. Table 5.1 indicates that all 3 estimators have little bias, get better as the sample size n increases for fixed α and are relatively unaffected by the value of $\alpha, \alpha > 1$ for the BLUαN and LS, all α for the mle. The mle is a clear winner for $\alpha \leq 1$.

In order to get a more detailed picture of the behavior of the 3 estimators, Table 5.2 presents percentages (rounded to 2 places) of times out of 1000 iterations that the estimators were within the specified distances of β. Two standard errors of the entries are no larger than 0.032. Table 5.2 reveals that the mle performs substantially better than both the least squares and BLUαN estimators for the smaller values of α, especially for the interval $\beta \pm \delta_1$. Note that the mle is best for the small values of α. We find the relatively good performance of the least squares estimator compared to the BLUαN for $\alpha > 1$ in Table 5.2 surprising and comforting in view of its wide use. However, as noted above, the least squares estimator performs poorly for $\alpha < 1$.

To check limiting normality, Table 5.3 compares sample percentiles of simulated mle's to corresponding percentiles based on the asymptotic normality of the mle as given in (2.14). The symbol "S" denotes a sample percentile obtained from simulation and the symbol "A" an asymptotic percentile. Since we generated the carriers $\{x_i\}$ from a uniform distribution, we globally approximated $\sum_{i=1}^{n} x_i^2$ by $nE(x^2)$. For $n = 10$ and $\alpha = 0.5$, and 1.0, the "S" column of Table 5.3 indicates that the distribution is skewed right. Otherwise, the sample percentiles are approximately symmetric around the sample median and reasonably close to the asymptotic percentiles. As expected, the normal approximation improves as n increases. More work needs to be done on assessing the role of α on the rate of convergence to normality.

6 Conclusions

Regression models whose error terms have mixture distributions with infinite variance have the potential for important applications. Our study of the case where the errors have symmetric stable distributions indicates that estimation via maximum likelihood is better than least squares and BLUαN, unless it is known that α is close to 2. Future studies should investigate the performance of roots of (2.5) as estimators of $\boldsymbol{\beta}$ for other families of variance mixture models.

TABLE 5.1
Means and Mean Absolute Deviations of the 3 Estimators
MAD's in Parentheses
True Value $\beta = 5$

		n		
α	Estimator	10	20	30
0.5	MLE	5.05 (0.21)	5.05 (0.10)	5.00 (0.04)
0.7	MLE	5.06 (0.20)	5.04 (0.11)	5.00 (0.05)
1	MLE	5.02 (0.19)	5.01 (0.11)	5.00 (0.06)
	BLUαN	5.33 (1.62)	6.48 (2.19)	5.22 (1.38)
	LS	4.99 (1.30)	5.62 (2.06)	5.10 (1.42)
1.5	MLE	5.02 (0.16)	5.00 (0.11)	5.00 (0.07)
	BLUαN	5.01 (0.32)	5.01 (0.17)	5.01 (0.16)
	LS	5.00 (0.24)	4.98 (0.23)	5.01 (0.12)
1.7	MLE	5.01 (0.16)	5.00 (0.11)	5.00 (0.06)
	BLUαN	4.99 (0.18)	4.98 (0.15)	5.00 (0.10)
	LS	4.99 (0.17)	4.99 (0.13)	5.00 (0.09)
2	ALL	5.00 (0.13)	5.00 (0.09)	5.00 (0.06)

TABLE 5.2
Percentages of Times Estimators were within δ of β

$\delta_1 = 0.025, \quad \delta_2 = 0.100, \quad \delta_3 = 0.175, \quad \delta_4 = 0.250, \quad \delta_5 = 0.325$

		n=10					n=20				
α	Estimator	δ_1	δ_2	δ_3	δ_4	δ_5	δ_1	δ_2	δ_3	δ_4	δ_5
0.5	MLE	24	59	74	81	85	39	78	88	93	94
	BLUαN	5	17	26	31	35	4	13	21	26	31
	LS	0	4	7	9	12	0	3	4	6	7
0.7	MLE	17	48	68	80	87	27	70	86	93	96
	BLUαN	6	23	35	43	49	5	20	32	39	46
	LS	1	8	16	23	29	2	8	13	19	24
1.0	MLE	14	46	68	82	88	22	61	84	93	96
	BLUαN	8	28	45	57	64	6	26	43	53	61
	LS	5	21	32	41	49	6	20	34	43	52
1.1	MLE	17	48	70	82	88	20	61	84	93	97
	BLUαN	7	29	45	57	66	7	32	51	64	72
	LS	6	23	37	49	58	7	26	40	52	60
1.3	MLE	16	48	70	84	90	21	57	81	93	97
	BLUαN	8	36	52	66	75	9	37	57	71	79
	LS	8	29	47	63	71	9	35	53	66	76
1.5	MLE	17	46	66	82	90	23	59	82	94	98
	BLUαN	10	38	58	74	83	12	44	66	82	88
	LS	10	36	56	69	79	13	45	65	78	87
1.7	MLE	13	42	64	81	89	21	59	82	93	98
	BLUαN	11	44	64	79	88	13	53	78	89	84
	LS	13	44	67	80	89	13	53	76	90	95
2.0	ALL	15	50	73	88	95	23	65	89	98	100

TABLE 5.2 (Continued)
Percentages of Times Estimators were within δ of β

$\delta_1 = 0.025, \quad \delta_2 = 0.100, \quad \delta_3 = 0.175, \quad \delta_4 = 0.250, \quad \delta_5 = 0.325$

		n=50				
α	Estimator	δ_1	δ_2	δ_3	δ_4	δ_5
0.5	MLE	51	94	99	100	100
	BLUαN	1	7	10	13	15
	LS	0	2	2	3	4
0.7	MLE	39	92	99	100	100
	BLUαN	3	14	22	28	33
	LS	2	7	10	14	18
1.0	MLE	35	86	98	100	100
	BLUαN	8	26	40	51	60
	LS	6	21	34	44	52
1.1	MLE	33	84	97	100	100
	BLUαN	8	32	47	59	66
	LS	8	26	40	55	63
1.3	MLE	30	83	98	100	100
	BLUαN	13	44	66	77	84
	LS	12	43	65	77	84
1.5	MLE	35	84	98	100	100
	BLUαN	18	58	80	89	94
	LS	19	57	79	89	94
1.7	MLE	29	80	97	100	100
	BLUαN	21	69	90	95	97
	LS	22	69	90	95	98
2.0	ALL	35	88	99	100	100

TABLE 5.3
Asymptotic "A" and Sample "S" Percentiles of the MLE

		n					
		10		20		50	
α	Percentile	S	A	S	A	S	A
0.5	10	4.85	4.92	4.91	4.94	4.94	4.96
	20	4.96	4.97	4.98	4.98	4.98	4.98
	50	5.01	5.00	5.01	5.00	5.00	5.00
	80	5.09	5.03	5.05	5.02	5.03	5.02
	90	5.39	5.08	5.18	5.06	5.06	5.04
1.0	10	4.78	4.80	4.85	4.86	4.90	4.91
	20	4.91	4.92	4.94	4.94	4.96	4.96
	50	5.00	5.00	5.00	5.00	5.00	5.00
	80	5.09	5.08	5.08	5.06	5.04	5.04
	90	5.26	5.20	5.18	5.14	5.09	5.09
1.5	10	4.78	4.83	4.85	4.88	4.90	4.92
	20	4.91	4.93	4.94	4.95	4.96	4.97
	50	5.01	5.00	5.00	5.00	5.00	5.00
	80	5.11	5.07	5.08	5.05	5.04	5.03
	90	5.26	5.17	5.18	5.12	5.10	5.08
2.0	10	4.80	4.80	4.85	4.86	4.92	4.91
	20	4.93	4.92	4.93	4.94	4.96	4.96
	50	5.00	5.00	5.00	5.00	5.00	5.00
	80	5.09	5.08	5.05	5.06	5.04	5.04
	90	5.21	5.20	5.14	5.14	5.09	5.09

REFERENCES

Aitchison, J. and Silvey, S.D.(1958). "Maximum likelihood estimation of parameters subject to constraints." Ann.Math.Stat.,29, 813- 828.

Arad, R.W.(1980). "Parameter estimation for symmetric stable distributions." International Economic Review, 209-220.

Blattberg, R. and Sargent, T. (1971). "Regression with non-Gaussian stable disturbences: some sampling results," Econometrica, 39, 501-510.

Brorsen, W.B. and Yang, S.R.(1990) "Maximum likelihood estimators of symmetric stable distribution parameters." Commun. Statist.-Simula.,14 1464.

Chambers, R.L. and Heathcote, C.R.(1981) "On the estimation of the slope and the identification of outliers in linear regression." Biometrika, 21-33.

Chung, K.L.(1974)." A Course in Probability Theory, Second Edition," Academic Press, N.Y.

DuMouchel, W.H.(1971) "Stable Distributions in Statistical Inference." unpublished Ph.D. dissertation, Yale University.

Feuerverger, A. and McDunnough, P.(1981). "On efficient inference in symmetric stable laws and processes." In Statistics and Related Topics. Edited by Csorgo, M.,Dawson, M.A., Rao, J.N.K. and Saleh, A.K.Md.E., North Holland, Amsterdam.

McCulloch, J. H. (1986). "Simple consistent estimators of stable distribution parameters." Commun. Statist.-Simul., 1109-1136

McLeish,D.L. and Small, C.G.(1991). "A projected likelihood function for semi-parametric models." In Proceedings of a Symposium in Honour of Professor V.P. Godambe, University of Waterloo, Waterloo Ontario, Canada.

Luenberger, D.G.(1969) "Optimization by Vector Space Methods," Wiley, N.Y.

Merkouris, T.(1991) "A transform method of optimal estimation in stochastic processes: basic aspects." In Proceedings of a Symposium in Honour of V.P. Godambe, University of Waterloo, Waterloo, Ontario, Canada.

Neveu, J.(1975). " Discrete Parameter Martingales", North Holland, Amsterdam.

Paulson, A.S. and Delehanty, T.A.(1985). "Modified weighted square error estimation procedures with special emphasis on the stable laws." Commun. Statist.-Simula., 927-972.

Rockafellar, R.T.(1970) "Convex Analysis," Princeton University Press, Princeton, N.J.

Samorodnitsky, G. and Taqqu, M.S.(1994). "Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance." Chapman and Hall, New York.

Zolotarev, V.M.(1966). "On representation of the stable laws by integrals." Selected translations in Mathematical Statistics and Probability. Providence, R.I., American Mathematical Society, 6, 84-88.