Estimation of Stable parameters using Spacings

S. Rao Jammalamadaka [a] Gaoyuan Tian[a]

SUMMARY

An alternative estimation method for stable family of distributions is presented here, based on spacings–the gaps between the observations. We briefly describe several existing estimation methods for the same problem, and compare them with this new method which has some built-in flexibility. Extensive simulation studies indicate that the proposed spacing based estimation is quite a practical alternative approach in large sample estimation, and performs well in dealing with the stable family of distributions.

1 Introduction

Stable distributions were described by Paul Levy (1924) in a study of normalized sums of independently and identically distributed (iid) random variables. Stable distributions are found to be useful for many reasons. First, there are some theoretical reasons for using stable distributions, e.g. hitting times for a Brownian motion yield a Levy distribution. Secondly, stable distributions turn out to be limits of normalized sums of iid random variables – a property that is considered to be one of the main reasons that these distributions are viewed as suitable for describing stock-returns, since a stock price may be considered the result of random instantaneous arrival of information. Mandelbrot (1963) was among the first to apply the stable law to stock-return data. Finally, many practical data sets exhibit heavy tails and skewness which stable distributions are able to capture.

2 Definition, parameterization and basic properties

Since there is no closed form of the density function, a stable distribution is defined through its characteristic function. A random variable $X \sim S(\alpha, \beta, \sigma, \mu)$, if its characteristic function has the form

$$\varphi(t) = Ee^{itX} = \begin{cases} exp(-\sigma^{\alpha}|t|^{\alpha}(1 - i\beta\frac{t}{|t|}tan(\frac{\pi\alpha}{2})) + i\mu_0 t), \alpha \neq 1 \\ exp(-\sigma|t|^{\alpha}(1 + i\beta\frac{2}{|\pi|}\ln(t)) + i\mu_0 t), \alpha = 1 \end{cases} \tag{2.1}$$

[a]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, 93106, USA

Here $0 < \alpha \leq 2$ measures the tail thickness, whereas $-1 \leq \beta \leq 1$ determines the skewness. And $\mu_0 \in \mathbb{R}$ and $\sigma > 0$ are location and scale parameters in the sense that $\frac{X-\mu}{\sigma} \sim S(\alpha, \beta, 1, 0)$. This particular parameterization, denoted as P_0 (Samorodnitsky and Taqqu (1994)) has the advantage that the parameters are easy to interpret in terms of location and scale. But there is a disadvantage, namely that when it comes to numerical or statistical work, it is discontinuous at $\alpha = 1$ and $\beta \neq 1$.

An alternative parametric representation (denoted as P_1) to overcome this problem, has the characteristic function: $\varphi_1(t) =$

$$\begin{cases} exp(-|\sigma t|^{\alpha} + i\sigma t\beta(|\sigma t|^{\alpha-1} - 1)\tan(\frac{\pi\alpha}{2}) + i\mu_1 t), \alpha \neq 1 \\ exp(-|\sigma t| + i\sigma t\beta\frac{2}{\pi}\ln|\sigma t| + i\mu_1 t), \alpha = 1 \end{cases} \tag{2.2}$$

where $0 < \alpha \leq 2$,$-1 \leq \beta \leq 1$, $\sigma > 0$ and $\mu_1 \in \mathbb{R}$.

Basic properties of stable distributions can be found in Samorodnitsky and Taqqu (1994). One important property is that: asymptotic tail behavior of stable laws is Pareto, namely, as $x \to \infty$,

$$P(X > x) \sim \sigma^{\alpha} C_{\alpha}(1 + \beta)x^{-\alpha}$$

where $C_{\alpha} = \sin\frac{\pi\alpha}{2}\Gamma(\alpha)/\pi$.By the reflection property, $P(X < -x) \sim \sigma^{\alpha}C_{\alpha}(1-\beta)x^{-\alpha}$. For this reason, the non-Gaussian stable distribution is also called stable Pareto.

3 Some common estimation methods for stable parameters

McCulloch (1986) obtained consistent estimators for four parameters in stable distributions based on five sample quantiles. The main estimating algorithm is described as below:

Step 1 Estimating α and β. X_p is the p-th quantile if $F(x_p) = p$, where $F(x)$ is the cdf. $\hat{x_p}$ is the sample quantile if $F_n(\hat{x_p}) = p$, where $F_n(x)$ is empirical cdf. Define two functions of theoretical quantiles:

$$\begin{cases} v_{\alpha} = \dfrac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} = \phi_1(\alpha, \beta) \\ v_{\beta} = \dfrac{(x_{0.95} - x_{0.5}) - (x_{0.5} - x_{0.05})}{x_{0.95} - x_{0.05}} = \phi_2(\alpha, \beta) \end{cases} \tag{3.3}$$

Replace v_{α} and v_{β} with their sample counterparts $\hat{v_{\alpha}}$ and $\hat{v_{\beta}}$, we get estimators

$$\begin{cases} \hat{\alpha} = \varphi_1(\hat{v_{\alpha}}, \hat{v_{\beta}}) \\ \hat{\beta} = \varphi_2(\hat{v_{\alpha}}, \hat{v_{\beta}}) \end{cases} \tag{3.4}$$

with $\hat{v}_{\alpha} = \frac{\hat{x}_{0.95} - \hat{x}_{0.05}}{\hat{x}_{0.75} - \hat{x}_{0.25}}$ and $\hat{v}_{\beta} = \frac{(\hat{x}_{0.95} - \hat{x}_{0.5}) - (\hat{x}_{0.5} - \hat{x}_{0.05})}{\hat{x}_{0.95} - \hat{x}_{0.05}}$

Step 2 Estimating scale parameter σ. Let us first define v_{σ} as $v_{\sigma} = \frac{x_{0.75} - x_{0.25}}{\sigma} = \phi_3(\alpha, \beta)$. The estimator $\hat{\sigma}$ is obtained by replacing (α, β) with $(\hat{\alpha}, \hat{\beta})$, thus $\hat{\sigma} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\phi_3(\hat{\alpha}, \hat{\beta})}$.

Step 3 Estimating location parameter μ. Let us define $\frac{x_{0.5}-\mu}{\sigma} = \phi_4(\alpha, \beta)$. The estimator $\hat{\mu}$ is obtained by replacing (α, β, σ) with $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$, thus $\hat{\mu} = \phi_4(\hat{\alpha}, \hat{\beta})\hat{\sigma} + \hat{x}_{0.5}$.

The main idea is to use difference of quantiles in order to get rid dependence on location and scale parameters. Then, two functions on alpha and beta are numerically calculated form sample quantiles and inverted to get the corresponding parameter estimates. The advantage of the method is that the computing speed is fast since we only use five quantiles. It is not sensitive to the extreme observations. Maybe one drawback is that the estimation procedure reflects little tail information of the distribution.

Since there is a closed form of characteristic function, the estimator based on empirical characteristic function can be developed. The regression -type estimation of Koutrouvellis (1980) starts with an initial estimate(In practice, we usually choose quantile estimate) and proceeds iteratively until some convergence criterion is satisfied.

Directly from the convenient form of the logarithm of the cf, we have the following linear equations

$$\ln(-\Re(\ln \phi(t))) = \alpha \ln \sigma + \alpha \ln |t| \tag{3.5}$$

and

$$\Im(\ln \phi(t)) = \mu_1 t + \beta\sigma t(|\sigma t|^{\alpha-1} - 1)\tan(\frac{\pi\alpha}{2}) \tag{3.6}$$

The algorithm is as follows,

Step 1 Given a sample of iid observations $x_1, x_2, ..., x_n$ first we find preliminary estimates σ_0 and μ_{01} by the quantile method of McCulloch and we normalize the observations as $x'_j = \frac{x_j - \hat{\mu}_{01}}{\hat{\sigma}_0}$ for $j = 1, 2, ..., n$.

Step 2 Consider the regression equation constructed above $y_k = b + \alpha\omega_k + \epsilon_k, k = 0, 1, 2, ..., 9$, where $y_k = \ln(-\Re(\ln \hat{\phi}(t)))$, $\omega_k = \ln |t_k|$, $t_k = 0.1 + 0.1k$ and ϵ_k denotes the error term. We find $\hat{\alpha}$ and $\hat{b}$ according to the method of least squares using the normalized sample $x'_1, x'_2, ...x'_n$. The estimator $\hat{\sigma_1}$ of the scale parameter of the normalized sample is $\hat{\sigma} = exp(\frac{\hat{b}}{\hat{\alpha}})$. The sample cf is defined as

$$\hat{\phi}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itx_j} = (\frac{1}{n}\sum_{j=1}^{n}\cos tx_j) + i(\frac{1}{n}\sum_{j=1}^{n}\sin tx_j), t \in \mathbb{R} \tag{3.7}$$

Step 3 Estimators $\hat{\beta}$ and $\hat{\mu_{11}}$ of the skewness parameter and the modified location parameter respectively are derived from the second regression equation based on (3.6): $z_k = \mu_{11}t_k + \beta\nu_k + \eta_k$, where $z_k = \Im(\ln \hat{\phi}(t))$, $\nu_k = \hat{\sigma}_1 t_k(|\hat{\sigma}t|^{\hat{\alpha}-1} - 1)\tan(\frac{\pi\hat{\alpha}}{2})$, $t_k = 0.1 + 0.1k$ and η_k is the error term.

Step 4 Compute the final estimates $\hat{\sigma} = \hat{\sigma}_0\hat{\sigma}_1$ and $\hat{\mu}_1 = \hat{\mu}_{01}\hat{\sigma}_0 + \hat{\mu}_{11}$. If we aim at estimating the location parameter μ, we need to take advantage of the connection between the two parametric forms P_0 and P_1:

$$\hat{\mu} = \hat{\mu}_1 - \hat{\beta}\hat{\sigma}\tan\frac{\pi\hat{\sigma}}{2} \tag{3.8}$$

Repeat step1 to step 4 until some convergence criterion is met.

The method of maximum likelihood is very attractive because of the good asymptotic properties of the estimates, provided that the likelihood function obeys certain general conditions. The likelihood function is $L(x_1, x_2, ..., x_n|\theta) = \prod_{k=1}^{n} f(x_k|\theta)$, where $x_1, x_2, ..., x_n$ is a sample of iid observations of a random variable X, $f(x|\theta)$ is the pdf of X and θ is a vector of parameters. In the case of stable distributions, $\theta = (\alpha, \beta, \sigma, \mu)$. Maximum likelihood estimates are found by searching for those parameter values which maximize the likelihood function, or equivalently, the log-likelihood function $\hat{\theta}_n = argmax \log(L(x_1, x_2, ..., x_n|\theta))$. Maximum likelihood estimation is theoretically most efficient estimating method when the sample size is big. But it is computationally intensive: The density function and the maximum searching procedure have to be both carefully numerically evaluated.

4 Spacings based estimation

This is a new method we propose: Given an iid random sample $x_1, \dots, x_n$ from a univariate distribution with cdf $F(x;\theta)$. Let $x_{(1)}, \dots, x_{(n)}$ be the corresponding order statistics. Define spacings as the 'gaps' between the values of the distribution function at adjacent ordered points, $D_i = F(x_{(i)};\theta) - F(x_{(i-1)};\theta), i = 1, \dots, n+1$. And we denote $F(x_{(0)};\theta) = 0, F(x_{(n+1)};\theta) = 1$. Then for any convex function $h : (0, \infty) \to \Re$, minimize the quantity $T(\theta) = \frac{1}{n}\sum_{i=1}^{n} h(nD_i(\theta))$. The resulting minimizer $\hat{\theta}$ is called Generalized Spacing Estimator (GSE) of θ.

The choice of different $h(x)$ yields different criterions of spacing estimation. If we choose $h(x) = -\lg(x)$, $T_n(\theta) = \sum_{i=1}^{n+1} \log D_i(\theta)$, which is called the maximum product of spacing (criterion1). If we choose $h(x) = (x-1)^2$, $T_n(\theta) = G_n(\theta) = \sum_{i=1}^{n+1}(D_i - \frac{1}{n})^2$, which is called Green-wood statistics (criterion2). If we choose $h(x) = |x-1|$, $T_n(\theta) = \sum_{i=1}^{n+1} |D_i - \frac{1}{n}|$, which is called Rao-statistic (cirterion3).

We will compare this three criterions of their performances in estimating parameters in stable distributions. Here we are interested in estimating (α, β) in the range $(1, 2] \times [-1, 1]$.

4.1 Parameter estimation in symmetric stable distribution

Firstly, we consider estimating alpha when $\beta = 0$ (symmetric case). In some cases, the practitioner strongly believes the distribution is symmetric. Sample $(x_1, \dots, x_{1000})$ is simulated from $S(\alpha = 1.5, 0, 1, 0)$ using algorithm of Chambers et al (1976).

The pdf and the cdf of stable laws can be very accurately evaluated numerically by Simpson's rule, with the help of special integral representations derived by Zolotarev in parameterization P_1. For $x > \xi$, they can be expressed as,

$$f(x; \alpha, \beta, P_1) = c_2(x; \alpha, \beta) \int_{\theta_0}^{\pi/2} g(\theta; x, \alpha, \beta) exp(-g(\theta; x, \alpha, \beta)) d\theta \tag{4.9}$$

and

$$F(x;\alpha,\beta,P_1) = c_1(\alpha,\beta) + c_3(\alpha)\int_{\theta_0}^{\pi/2} exp(-g(\theta;x,\alpha,\beta))d\theta \tag{4.10}$$

where for the case $\alpha \neq 1$, the parameters take values as following,
$c_1(\alpha,\beta) = \frac{1}{\pi}(\frac{\pi}{2} - \theta_0)$ for $\alpha < 1$, and 1 for $\alpha > 1$,
$c_2(x;\alpha,\beta) = \frac{\alpha}{\pi|\alpha-1|(x-\xi)}$,
$c_3(\alpha) = \frac{sign(1-\alpha)}{\pi}$,
$g(\theta;x,\alpha,\beta) = (x-\xi)^{\frac{\alpha}{\alpha-1}}V(\theta;\alpha,\beta)$,
$\xi = \xi(\alpha,\beta) = -\beta\tan(\frac{\pi\alpha}{2})$,
$\theta_0 = \theta_0(\alpha,\beta) = \frac{1}{\alpha}arctan(\beta\tan(\frac{\pi\alpha}{2}))$,
$V(\theta;\alpha,\beta) = (\cos(\alpha\theta_0))^{\frac{1}{1-\alpha}}(\frac{\cos\theta}{\sin\alpha(\theta_0+\theta)})^{\frac{\alpha}{\alpha-1}}\frac{\cos(\alpha\theta_0+(\alpha-1)\theta)}{\cos\theta}$,

and for the case $\alpha = 1$, they reduce to
$c_1(\alpha,\beta) = 0$,
$c_2(x;\alpha,\beta) = \frac{1}{2|\beta|}$,
$c_3(\alpha) = \frac{1}{\pi}$,
$g(\theta;x,\alpha,\beta) = exp(\frac{\pi x}{2\beta})V(\theta;\alpha,\beta)$,
$\xi = 0$,
$\theta_0 = \frac{\pi}{2}$,
$V(\theta;\alpha,\beta) = \frac{2}{\pi}(\frac{\frac{\pi}{2}+\beta\theta}{\cos\theta})exp(\frac{1}{\beta}(\frac{\pi}{2}+\beta\theta)\tan\theta)$.

The case $x < \xi$ can be treated by taking advantage of the relations

$$f(x;\alpha,\beta;P_1) = f(-x;\alpha,-\beta;P_1) \tag{4.11}$$

and

$$F(x;\alpha,\beta;P_1) = 1 - F(-x;\alpha,-\beta;P_1) \tag{4.12}$$

The estimated value of spacing estimator that minimizes Greenwood-statistic is $\hat{\alpha} = 1.5126$.

Now we use Monte Carlo methods to compare three criterions. Suppose we have M samples based on the data $(x_1,\ldots,x_N)$ generated from $S(\alpha = 1.5, 0, 1, 0)$. In each sample i, we have spacing estimator $\hat{\alpha}_i$, $i = 1,\ldots,M$. And $E(\hat{\alpha}) \approx \frac{1}{M}\sum_{i=1}^{M}\hat{\alpha}_i$, $MSE(\hat{\alpha}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\alpha}_i - 1.5)^2$ for large M.

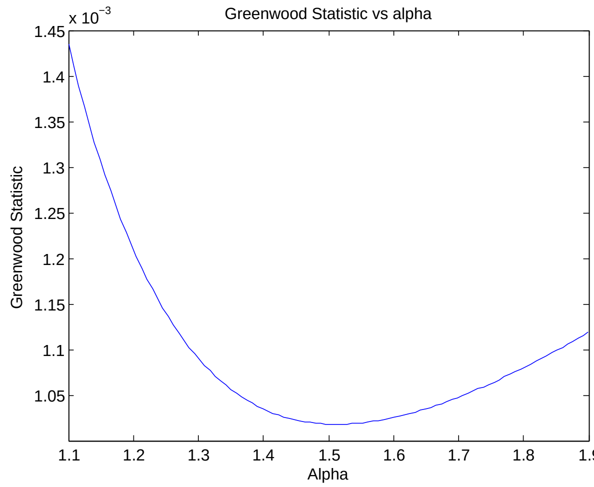

Figure 1: Greenwood statistic vs.α

N=200, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4639	1.4397	1.4649
MSE of $\hat{\alpha}$	0.0110	0.0214	0.0162
N=500, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4786	1.4677	1.4783
MSE of $\hat{\alpha}$	0.0051	0.0088	0.0072
N=1000, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4892	1.4824	1.4938
MSE of $\hat{\alpha}$	0.0023	0.0038	0.0036

Table 1:Monte Carlo MSE of $\hat{\alpha}$ with various sample size

Similar to MLE, one advantage of spacing based estimator is the asymptotic normality of the estimator. Ghosh and Jammalamadaka(2001) show, under some regularity conditions on the density and $h(\cdot)$, $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\text{d}} N(0, \sigma_h^2/I(\theta_0))$, where $I(\theta_0)$ is the Fisher Information in one observation from the true distribution and $\sigma_h^2 = \frac{E(Wh'(W))^2 - 2EWh'(W)Cov\{Wh'(W), W\}}{[EW^2h''(W)]^2}$, with $W \sim Exp(1)$.

If $h(x) = (x-1)^2$, $\sqrt{n}(\hat{\alpha}-\alpha) \to N(0, \frac{2}{I(\alpha)})$. To illustrate this, take $M = 1000$, $N = 1000$ and we plot the histogram and kernel density of the estimator $\hat{\alpha}$.

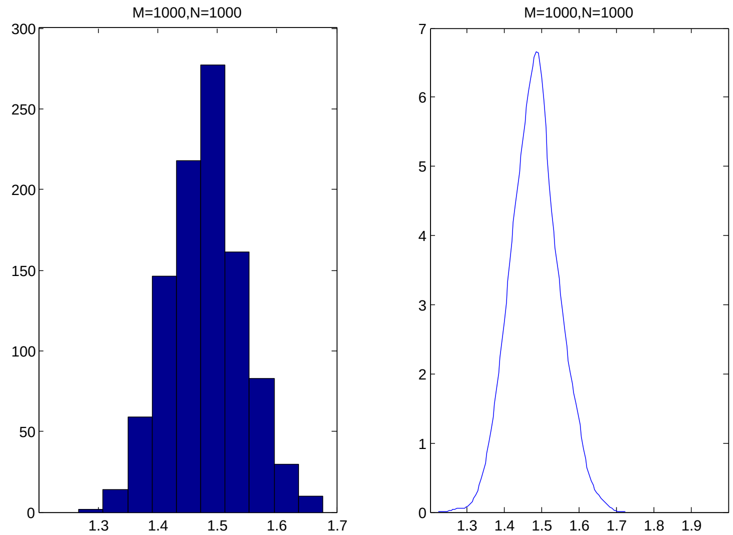

Figure 2: Histogram and kernel density estimate of $\hat{\alpha}$

4.2 Parameter estimation in asymmetric stable distribution

Consider the asymmetric case where α and β are both unknown. The sample is the same as above, 1000 realization from $S(\alpha = 1.5, \beta = 0, 1, 0)$. Following the similar estimating procedure as above (under criterion 2), we get this two-dimensional estimator $(\hat{\alpha}, \hat{\beta}) = (1.5158, 0.0572)$. Figure 3 shows this estimator is the global minimal point.

A comparison of the three criterions are listed at Table 2. It seems criterion 3 has little bias while criterion 1 has smallest mean square error.

N=1000, M=100			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha}, \hat{\beta})$	(1.4861,0.1940)	(1.4823,0.1909)	(1.4874,0.1989)
MSE of $(\hat{\alpha}, \hat{\beta})$	(0.0025,0.0029)	(0.0051,0.0066)	(0.0037,0.0064)

N=500, M=100			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha}, \hat{\beta})$	(1.4816,0.1941)	(1.4675,0.1824)	(1.4848,0.2071)
MSE of $(\hat{\alpha}, \hat{\beta})$	(0.0039,0.0046)	(0.0086,0.0117)	(0.0062,0.0125)

Table 2: Monte Carlo MSE of $(\hat{\alpha}, \hat{\beta})$ with various sample size

A monte carlo evaluation to compare different methods is listed in the following table.

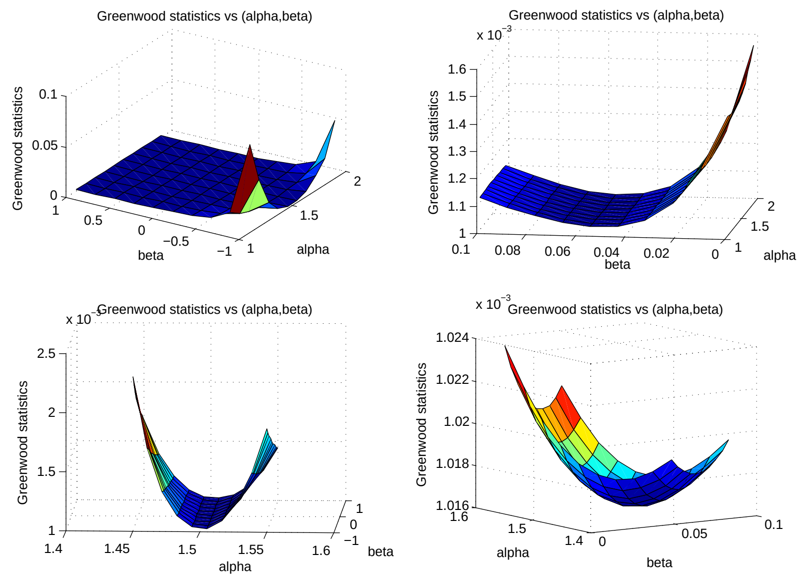

Figure 3: Greenwood statistics vs. (α, β)

N=1000, M=1000				
	MLE	Quantile	Regression	Spacing(Criterion3)
$(\hat{\alpha}, \hat{\beta})$	(1.5017,0.2045)	(1.5049,0.2085)	(1.5042,0.2010)	(1.4913,0.1998)
MSE$(\hat{\alpha}, \hat{\beta})$	(0.0025,0.0092)	(0.0045,0.0126)	(0.0034,0.0143)	(0.0035,0.0058)

N=200, M=1000				
	MLE	Quantile	Regression	Spacing(Criterion3)
$(\hat{\alpha}, \hat{\beta})$	(1.5016,0.2159)	(1.5039,0.2242)	(1.5031,0.2058)	(1.4641,0.1747)
MSE $(\hat{\alpha}, \hat{\beta})$	(0.0138,0.0567)	(0.0250,0.0822)	(0.0179,0.0833)	(0.0173,0.0741)

Table 3: Monte Carlo MSE of $(\hat{\alpha}, \hat{\beta})$ with different methods

5 Conclusion

We introduced spacing based estimation for stable distributions. Compared with other estimation method, spacing-based estimation has the advantage that one could select certain criterion in a specific question. Also, one could also use the spacing ideal here to do a goodness of fit test.

Because of the asymptotic normality of spacing estimator, it is a candidate estimator in the large sample estimation. Even though Maximum likelihood estimation beat the spacing-based estimation, it is computational intensive, While spacing based estimation would improve the speed by some data reduction if we consider estimators using higher order spacings, namely m-step spacing $D_i(m) = F(X_{(j+m)}) - F(X_{(j)})$.

Acknowledgements

The programs for estimating via the ML method, Quantile method, and regression method are graciously provided by Professors J.P Nolan and Svetlozar Rachev. We would like to thank Professors Tomoyuki Ichiba and Emanuel Taufer for very useful discussions and suggestions.

References

1. Adler,R.,Feldman, R. and Taqqu, M.(eds.) (1998), A practical Guide to Heavy Tailed Data, Birkhauser,Boston,MA.

2. Bergstrom.H(1952), On Some Expansions of Stable Distribution, Akiv for Matematik II, 375-378.

3. Chambers, J.M., Mallows,C. and Stuck, B.W.(1976), A method for simulating stable random variables, JASA 71: 340-344

4. DuMouchel,W.H.(1971), Stable Distributions in Statistical Inference, Ph.D. dissertation, Department of Statistics, Yale University.

5. Koutrouvelis, I.(1980), Regression-type estimation of the parameters of stable laws, Journal of the American Statistical Association 75, 919-928.

6. McCulloch,J.H.(1986), Simple Consistent Estimator of Stable Distribution Parameters, Communication in Statistics. Simulation and Computation, 15, 1109-1136.

7. Nolan, J.P.(1997), Numerical computation of stable densities and distribution functions, Commun.Stat.: Stochastic Models 13: 759-774.

8. Nolan,J.P.(2002), Maximum likelihood estimation and diagnostics for stable distributions. American University, Washington.

9. Press,S.(1972), Estimation of univariate and multivariate stable distributions, Journal of the American Statistical Association 67,842-846.

10. Rachev,S.T. and Mittnik.S(2000), Stable Paretian Models in Finance, John Wiley & Sons Ltd.

11. Zolotarev,V.M.(1964), On the Representations of Stable Laws by Integrals, Selected Translations in Mathematical Statistics and Probability, 6, 84-88, American Mathematical Society, Providence, Rhode Island.

12. Zolotarev,V.M.(1995), On representation of densities of stable laws by special functions, Theory probab. Appl. 39:354-362.