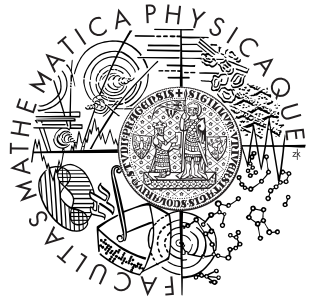
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Vadym Omelchenko

Stabilní rozdělení a finanční aplikace

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Prof. Lev Klebanov, DrSc.

Studijní program: Matematika

Studijní obor: pravděpodobnost, matematická statistika a ekonometrie

2007

Děkuji vedoucímu práce Prof. DrSc. Lvu Klebanovovi za mnoho cenných rad a odbornou pomoc při vedení této práce a za mnoho životních zkušeností získaných při společných konzultacích.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 6 srpna 2007, Vadym Omelchenko

Název práce: Stabilní rozdělení a finanční aplikace
Autor: Vadym Omelchenko
pravdipodobnosti a matematické statistiky
Vedoucí diplomové práce: Prof. Lev Klebanov, DrSc.
e-mail vedoucího: Lev.Klebanov@mff.cuni.cz

Abstrakt: Tato práce se zabývá teorií stabilních rozdělení, metodami odhadu jejich parametrů a jejich finanční aplikací. Byly zmíněny všeobecně známé odhady a navrženy metody odhadu parametrů na základě charakteristické funkce a projekční metody, která je modifikací metody maximální věrohodnosti. Kvalita odhadů se zjišťovala s pomocí simulací náhodného výběru ze stabilního rozdělení se známými parametry a porovnání odhadů parametrů s jejich skutečnými hodnotami. Jádrem této práce jsou odhady parametrů stabilních rozdělení, což je aplikovatelné pro modifikace modelu typu ARCH/ GARCH se stabilní inovací.

Klíčová slova: stabilní rozdělení, ARCH/GARCH modely, odhady založené na charakteristické funkci (CF), odhady založené na projekční metodě, příbuzné odhadům ML (MLP).

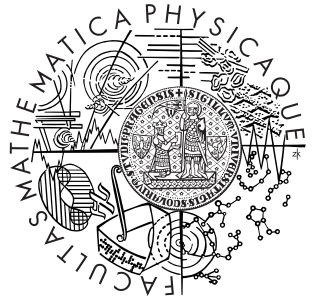
Title: Stable distributions and application to finance
Author: Vadym Omelchenko
Department: Department of Probability and Mathematical Statistics
Supervisor: Prof. Lev Klebanov, DrSc.
Supervisor's e-mail address: Lev.Klebanov@mff.cuni.cz

Abstract: This work deals with the theory of the stable distributions, their parameter estimation, and their financial application. There are given the methods of characteristic function and method of projections, which is relative to ML-methodology, for estimation of the parameters of stable distributions. We compare these methods with the conventional estimators. The quality of estimators is verified by the simulation of the sample having stable distribution with known parameters and comparing the estimates of these parameters with their real values. The aim of this work is estimation of parameters of the stable laws which is applicable for modification of ARCH/GARCH models with stable innovations.

Keywords: stable distribution, ARCH/GARCH models, characteristic function (CF) based estimators, maximum likelihood projection (MLP) estimators.

Charles University in Prague
Mathematical Physical Faculty

DIPLOMA THESIS



Vadym Omelchenko

Stable distributions and application to finance

Department of probability and mathematical statistics

Supervisor: Prof. Lev Klebanov, DrSc.

Study program: Mathematics

Branch of study: probability, mathematical statistics and econometrics

2007

Contents

Chapter 1

Introduction

Stable distributions play an important role in financial practice. First of all, they can generalize ARCH/GARCH model and their modifications. The classical ARCH/GARCH models are based on assumption that the random factor or the innovation is normally distributed, but the practice show that it is not always true. The random factor is normally distributed when the market volatility is low, and when there isn't such fast growth as in some Asian economies. But when we deal with assets whose price can quickly jump or drop, we can't apply ARCH/GARCH models with normal distribution, because they have a large volatility. Therefore, it is reasonable to apply random factors with large or infinite variance and stable distributions belong to them. They have many advantages opposed to other distributions, and they have been shown to be effective in prediction. The core of this work will be the parameter estimation of the stable distributions, because this theme is not well developed, and many researchers use it for commercial purposes, not disclosing their approach. There are many methods how to recover the time series of random factors from modification of ARCH/GARCH models, and they are described in many publications. But we will explore how to estimate the unknown parameters from the stable distribution which can be received in practice by extracting from the corresponding time series like modification of ARCH/GARCH models. Any stable distribution is uniquely represented by its four parameters:
1) α the most important parameter which is called the tail index, $\alpha \in (0, 2]$, if the parameter α of the stable random variable X belongs to $(0, 2)$ then $EX^a < \infty$ if $0 < a < \alpha$ and $EX^a = \infty$ if $a \geq \alpha$. From this follows that if $\alpha < 2$ the stable distribution has an infinite variance. If $\alpha = 2$ then

we deal with normal distribution, which is a particular case of the stable distributions.
2) $\beta \in [-1, 1]$ has the properties of skewness,
3) $\sigma > 0$ has properties of the variance,
4) $\mu \in R$ has properties of the mathematical expectation if $\alpha \leq 1$ and it is a mathematical expectation if $\alpha > 1$. The most important advantage of the stable distributions is the fact that they are preserved under convolution, and under some conditions that will be considered later the sum of random variables converges to the stable distributions. For example, sums of Pareto distribution with $\alpha \in (0, 2]$.

In practice there are popular such models like ARCH/GARCH and their modifications. They contain random factor which is normally distributed. Very often the prices of assets have such jumps and drops that are not typical for the models with the normal random factor. They show that the random factor must have a large variance. But if we take other random factors we can lose the advantages that the classical ARCH/GARCH models possess. But the more similar new random factor to the normal law the less properties of classical ARCH/GARCH models lost. But the stable distributions are the best candidates to replace the normal law. They have many properties of normal laws, e.g. preserving under convolution, it is known from mathematical statistics that the linear combination of normal laws is a normal law and the same holds for any stable distribution. When α is close to 2 the density function is very similar to the normal one but the variance of the corresponding random variable is infinite. The problem of estimating parameters of the stable distributions is very important, that why we aim to find the best estimate of parameters of the stable laws. First chapter is devoted to ARCH/GARCH models and their properties. The following chapters to the stable distributions and their modifications. From eighth chapter we explore methods of the parameter estimation of the stable distributions. The parameters will be estimated by two methods:
1) Method based on characteristic function
2) Method based on projections and Maximum Likelihood methodology.
Most of theorems will be without proofs but they can be found in referred literature.

Chapter 2

ARCH and GARCH models

ARCH and GARCH models were introduced by Engle in 1982. These models describe behavior of financial assets.

ARCH(p) model (Auto-Regressive Conditional Heteroskedastic. (See [1])

$$h_n = \ln \frac{S_n}{S_{n-1}}, \quad \sigma_n = h_n \varepsilon_n$$

$$\sigma_n{}^2 = a_0 + \sum_{k=1}^{p} h^2{}_{n-k} a_k$$

$$a_k \geq 0; a_0 > 0$$

and $h_0 = h_0(\omega)$ is a random variable independent of $(\varepsilon_n)_{n\geq 1}$

GARCH(p,q) model, (Generalized ARCH)

$$\sigma_n = h_n \varepsilon_n$$

$$\sigma_n{}^2 = a_0 + \sum_{k=1}^{p} h^2{}_{n-k} a_k + \sum_{k=1}^{p} \sigma^2{}_{n-k} b_k$$

$$a_k \geq 0; a_0 > 0; b_k \geq 0$$

The parameter ε_n is a random factor which has standard normal distribution i.e. $\varepsilon_n \sim N(0,1)$

2.1 ARCH

We see that in this model σ_n are predictable functions of $h_{n-1}^2, h_{n-2}^2, ..., h_{n-p}^2$ and it is clear that large values of h_k^2 imply large values of σ_n^2 and vice versa what explains "cluster phenomenon", so if the value of h_n isn't large it can be explained by the behavior of the random factor.

Let's consider several properties of ARCH model, for simplicity let' s assume p=1, i.e. let's consider following model:

$$\sigma_n^2 = \alpha_0 + a_1\sigma_{n-1}^2$$

following simple properties are obvious:

$$Eh_n = 0, \quad Eh_n^2 = \alpha_0 + a_1 Eh_{n-1}^2,$$
$$E(h_n^2 \mid F_n) = \sigma_n^2 = \alpha_0 + a_1 h_{n-1}^2$$

If $0 < a_1 < 1$ then

$$Eh_n^2 \equiv \alpha_0 + a_1 Eh_{n-1}^2 = \alpha_0 + a_1(\alpha_0 + a_1 Eh_{n-2}^2) = ...$$

$$... = \alpha_0(1 + a_1 + a_1^2 + a_1^3 + a_1^4...) = \frac{\alpha_0}{1 - a_1}$$

Next, simple calculations show that

$$Eh_n^4 = E\sigma_n^4 E\varepsilon_n^4 = 3E\sigma_n^4 = 3E(\alpha_0 + h_{n-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0 a_1 Eh_{n-1}^2 + a_1 Eh_{n-1}^4) =$$

$$\frac{3\alpha_0^2(1 + a_1)}{1 - a_1} + 3a_1^2 Eh_{n-1}^2$$

Hence assuming that $0 < a_1 < 1$ and $3a_1^2 < 1$ we can obtain the following stationary solution

$$Eh_n^4 = \frac{3\alpha_0^2(1 + a_1)}{(1 - a_1)(1 - 3a_1)}$$

The previous calculations enable us to compute the "stationary" value of kyrtosis:

$$K \equiv \frac{Eh_4^2}{(Eh_n^2)^2} - 3 = \frac{6a_1^2}{1 - 3a_1^2}$$

K must be equal to zero for the normal distribution. The kyrtosis can be estimated by the following statistic:

$$\hat{K} = \frac{\frac{1}{N}\sum_{k=1}^{N}(h_k - \overline{h}_N)^4}{(\frac{1}{N}\sum_{k=1}^{N}(h_k - \overline{h}_N))^2} - 3$$

where $\bar{h}_n = \frac{1}{N}(h_1 + .. + h_N)$

The reason of the usage of other innovation is that in many cases the kyrtosis is very different from zero. In most cases it is positive. For $0 < a_1 < 1$ the sequence h=(h_n) is a square integrable martingale difference, therefore it is a sequence of orthogonal variables:

$$Cov(h_n, h_m) = 0, m \neq n$$

of course it doesn't mean that they are independent. The dependence of h_n and h_m is worthy to be explored. To do this we can calculate the covariation between $\mid h_n \mid$ and $\mid h_m \mid$ or between h_n^2 and h_m^2

$$Dh_n^2 = E(h_n^4) - (Eh_n^2)^2 = \frac{3\alpha_0^2(1+a_1)}{(1-a_1)(1-3a_1^2)} - \left(\frac{\alpha_0}{1-a_1}\right)^2 = \frac{2}{1-3a_1^2}\left(\frac{\alpha_0}{1-a_1}\right)^2$$

and

$$Eh_n^2 h_{n-1}^2 = \frac{1+3a_1^2}{1-3a_1^2}\frac{\alpha_0^2}{1-a_1}$$

therefore

$$\rho(1) = Corr(h_n^2, h_{n-1}^2) = \frac{Cov(h_n^2, h_{n-1}^2)}{\sqrt{Dh_n^2 Dh_{n-1}^2}} = a_1.$$

Further

$$\begin{aligned} Eh_n^2 h_{n-k}^2 &= E[h_{n-k}^2 E(h_n^2 \mid F_{n-1})] = E[h_{n-k}^2 E(\sigma_n^2 \varepsilon_n^2 \mid F_{n-1})] \\ &= E[h_{n-k}^2(\alpha_0 + a_1 h_{n-1}^2)] = \alpha_0 Eh_{n-k}^2 + a_1 Eh_{n-1}^2 h_{n-k}^2 \end{aligned}$$

that gives us a simple recursion relation for $\rho(k)$. In the 'stationary case'

$$\rho(k) = \alpha_0 Eh_{n-k}^2 + a_1 \rho(k-1)$$

so that

$$\rho(k) = a_1^k$$

ARCH-models are connected with autoregressive models and h_n^2 is an AR(p) model with the innovation ν_n which is a martingale difference and which equals $\nu_n = h_n^2 - \sigma_n^2$.

There ARCH(1) model will become more complicated if we assume another dependence in h as follows:

$$h_n = b_0 + b_1 h_{n-1} + \sqrt{a_0 + a_1 h_{n-1}}\varepsilon_n$$

In this case h=(h_n) is governed by AR(1)/ARCH(1) model or it satisfies Autoregressive scheme with ARCH noise

$$(\sqrt{a_0 + a_1 h_{n-1}} \cdot \varepsilon_n), \quad n \geq 1$$

.

This model is conditionally Gaussian, therefore we can represent the density $p_\theta(h_1, ..., h_n)$ of the joint distribution P_θ of the variables $h_1, ..., h_n$ for a fixed value of the parameter $\theta = (a_0, a_1, b_0, b_1)$ as follows ($h_0 = 0$)

$$p_\theta(h_1, ..., h_n) = (2\pi)^{-n/2} \prod_{k=1}^{n} \frac{1}{\sqrt{a_0 + a_1 h_{k-1}^2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n} \frac{(h_k - b_0 - b_1 h_{k-1})^2}{a_0 + a_1 h_{k-1}^2}\right\}$$

As an example of the application of this method for the parameter estimations we will consider the problem of the estimation of b_1 under assumption that all other parameters are known. We will use the ML estimation. So we have the maximum when

$$\frac{dp_\theta(h_1, ..., h_n)}{db_1} = 0$$

so

$$\log(p_\theta(h_1, ..., h_n)) = \log((2\pi)^{-n/2}) + \log\left(\prod_{k=1}^{n}(a_0 + a_1 h_{k-1}^2)^{-0.5}\right) +$$

$$+ \log\left(\exp\left\{-\frac{1}{2}\sum_{k=1}^{n} \frac{(h_k - b_0 - b_1 h_{k-1})^2}{a_0 + a_1 h_{k-1}^2}\right\}\right)$$

The argmax of this function is the same as one of: $-\frac{1}{2}\sum_{k=1}^{n}\frac{(h_k-b_0-b_1h_{k-1})^2}{a_0+a_1h_{k-1}^2}$

So deriving we get:

$$\sum_{k=1}^{n}\frac{(h_k-b_0-b_1h_{k-1})h_{k-1}}{a_0+a_1h_{k-1}^2}=0$$

Hence

$$\sum_{k=1}^{n}\frac{(h_k-b_0)h_{k-1}}{a_0+a_1h_{k-1}^2}-b_1\sum_{k=1}^{n}\frac{h_{k-1}^2}{a_0+a_1h_{k-1}^2}=0$$

So we obtain:

$$\hat{b}_1=\frac{\sum_{k=1}^{n}\frac{(h_k-b_0)h_{k-1}}{a_0+a_1h_{k-1}^2}}{\sum_{k=1}^{n}\frac{h_{k-1}^2}{a_0+a_1h_{k-1}^2}}$$

and $\hat{b}_1 = b_1 + \frac{M_n}{\langle M_n\rangle}$ where $M_n = \sum_{k=1}^{n}\frac{h_{k-1}\varepsilon_k}{\sqrt{a_0+a_1h_{k-1}^2}}$ is a martingale and $\langle M_n\rangle = \sum_{k=1}^{n}\frac{h_{k-1}^2}{\sqrt{a_0+a_1h_{k-1}^2}}$ is its quadratic characteristics.

$\frac{M_n}{\langle M_n\rangle} \longrightarrow 0$ by the strong law of large numbers therefore this estimation is consistent.

2.2 GARCH

The success of ARCH(p) model and its ability to explain "cluster property" created an avalanche of generalizations of this model, first of them was GARCH(p,q) model. GARCH(p,q) model is defined as follows:

$$\sigma_n = h_n\varepsilon_n$$

$$\sigma_n{}^2 = a_0 + \sum_{k=1}^{p} h^2{}_{n-k}a_k + \sum_{k=1}^{p}\sigma^2{}_{n-k}b_k$$

$$a_k \geq 0; a_0 > 0; b_k \geq 0$$

The parameter ε_n is a random factor which has standard normal distribution i.e. $\varepsilon_n \sim N(0,1)$ GARCH(p,q) model is advantages as against its forefather ARCH(p) model, because it doesn't require large values of p and

q. In ARCH(p) model the parameter p is often pretty large. We will demonstrate the properties of GARCH model on the easiest GARCH(1,1) model, in which

$$h_n = \sigma_n \varepsilon_n$$

and

$$\sigma_n^2 = a_0 + a_1 h_{n-1}^2 + b_1 \sigma_{n-1}^2$$

with $a_0 > 0, a_1 \geq 0, b_1 \geq 0$. It is clear that

$$Eh_n^2 = a_0 + (a_1 + b_1) Eh_{n-1}^2$$

and the 'stationary' value exists when $a_1 + b_1 < 1$ and it is equal to:

$$Eh_n^2 = \frac{a_0}{1 - a_1 - b_1}$$

If $3a_1^2 + 2a_1 b_1 + b_1^2 < 1$ then we have well-defined 'stationary' value

$$Eh_4^2 = \frac{3a_0^2(1 + a_1 + b_1)}{(1 - a_1 - b_1)(1 - b_1^2 - 2a_1 b_1 - 3a_1^2)}$$

And therefore for the 'stationary' kyrtosis we obtain

$$K = \frac{Eh_n^4}{(Eh_n^2)^2} - 3 = \frac{6a_1^2}{(1 - b_1^2 - 2a_1 b_1 - 3a_1^2)}$$

The auto-covariance function can be easily found the same way we did it finding the one for ARCH model.

$$\rho(1) = \frac{a_1(1 - a_1 b_1 - b_1^2)}{1 - 2a_1 b_1 - b_1^2}$$

$$\rho(k) = a_1 + {b_1}^{k-1} \rho(1), k > 1.$$

Finally we point out that that we can generalize the prediction for ARCH model into GARCH as follows:

$$\widehat{h}_{n+m}^2 = \widehat{\sigma}_{n+m}^2 = E(\sigma_{n+m}^2 | F_n^h)$$

$$= a_0 \frac{1 - \gamma^m}{1 - \gamma} + \gamma^{m-1}(a_1 h_n^2 + b_1 \sigma_n^2)$$

where $\gamma = a_1 + b_1$. There are many other generalizations of ARCH models, like TGARCH, NGARCH, HARCH, some of them will be considered in the next chapter.

2.3 Necessity of the generalization

As it was said above, ARCH model can explain many phenomena which take place on the market, but it explains only part of it. Therefore, there was a need to make some modifications of ARCH models which would have all the advantages of ARCH models and that would explain other phenomena, which classical ARCH model can't explain. We can modify ARCH/GARCH model according to time, ARCH model is discrete, according to the formula of the dependence of volatilities and logarithm of price changes, and according to the type of the innovation, which rules the model. But we have to be aware that the new innovation has to be similar to the normal innovation, therefore it is reasonable to replace the normal innovation by the stable one, because many properties that are held for normal distribution hold for stable one. There other alternatives like hyperbolic distributions and others but they aren't similar to the normal distribution.

2.4 Generalization according to the kind of innovation

If we use ARCH(p) model with an innovation having infinite variance we can get the model which would explain 'cluster property' and unexpected jumps but this model isn't precise, moreover it can collapse as we saw from the previous example therefore we should modify the whole model to suit the innovation and to make it function. There exist stable-ARCH/GARCH models which will be considered in the next chapters.

Chapter 3

Stable Distributions

Definition of the stable distribution

There are 4 equivalent definitions of the stable distributions: (See [4],[6],[1])

1.1

A random variable X is said to have a stable distribution if for any positive numbers A and B, there is a positive number C and a real number D such that

$$AX_1 + BX_2 =_d CX + D$$

where $X1$ and $X2$ are independent copies of X and $=_d$ is equality in distribution.

1.2

A random variable is said to have a stable distribution if for any $n \geq 2$, there is a positive number C_n and a real number D_n such that:

$$X_1 + X_2 + ...X_n = C_nX + D_n$$

where $X_1, X_2, .., X_n$ are independent copies of X.

Both the first and second definition specify the preserving of the properties under convolution.

1.3

A random variable X is said to have a stable distribution if it has a that 1.4 domain of attraction, i.e., if there is sequence of i.i.d. random variable $Y_1, Y_2, ..., Y_n$ and two sequences of constants (d_n) and (a_n) such that

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \longrightarrow_d X$$

where $\longrightarrow_d$ means convergence in distribution. The Cauchy distribution is a stable distribution, as it is known from the mathematical statistics the average of variables having Cauchy distribution C(0,1) is Cauchy distribution C(0,1) what suits the second definition where $C_n = \frac{1}{n}$ and $D_n = 0$ and the third where $d_n = n$ and $a_n = 0$. The third definition specifies the limit properties of the stable distributions. Pareto and geometric stable distributions, that will be considered later, belong to the stable distributions.

1.4

A random variable X is said to have a stable distribution if there are four parameter $\alpha \in (0, 2]$, $-1 \leq \beta \leq 1, \sigma > 0$ and $\mu \in R$ such that its characteristic function has a form:

$$ch(t) = E \exp[itX] = \exp\left[-\sigma^\alpha |\theta|^\alpha \left(1 - i \cdot \beta \cdot \text{sgn}(\theta) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu\theta\right]$$

if $\alpha \neq 1$ and

$$ch(t) = E \exp[itX] = \exp\left[-\sigma|\theta| \left(1 - i\beta\frac{2}{\pi} \cdot \text{sgn}(\theta) \ln|\theta|\right) + i\mu\theta\right].$$

If $\alpha = 1$. The parameter α is the index of stability. Parameter β is a parameter showing skewness or asymetry of the model, parameter σ is a scale parameter and parameter μ is a location parameter. The fourth definition specifies characteristic function.

3.1 Properties of the stable laws

Theorem 1

For any stable random variable X there exist number $\alpha \in (0, 2]$, two independent copies X_1 and X_2 and constants A, B, D and C such that $AX_1 + BX_2 =$

$CX + D$ where $C^\alpha = A^\alpha + B^\alpha$. If D=0 then the distribution is strictly stable.

Theorem 2.

For any stable random variable X there exists a number $\alpha \in (0, 2]$ such that C_n in the definition 2. equals $n^{1/\alpha}$. In financial practice α always belong to the interval (1,2], that guarantees finiteness of the expected value. The next examples demonstrate why this is impossible. When $\alpha < 1$ and the distribution is strictly stable then there are some interesting properties of this distribution. Suppose that $\alpha = 0.1$ and we have three i.i.d. stable random variables X_1, X_2 and X_3. Suppose that X is an independent copy of that two random variables. Then according to the previous theorem

$$X_1 + X_1 + X_3 =_d 3^{1/0.1}X = 3^{10}X = 59049X$$

What about ten independent copies $X_1, ..., X_{10}$? Then

$$X_1 + X_2 + .. + X_{10} =_d 10^{10}X (= 10000000000X)$$

What about $\alpha = 0.01$? If we have only two copies of the stable distribution then

$$X_1 + X_2 =_d 2^{(1/0.01)}X = 2^{100}X (= 1267650600228229401496703205376X)$$

That is

$$X_1 + X_2 =_d 1267650600228229401496703205376X$$

It is absolutely clear that such dependence is impossible in economic and financial practice. The stable distributions with $\alpha \in (0, 1)$ have wide application in physics, biology and other disciplines but not in finances. The stable distributions having α close to 2 are very similar the normal distributions.

If X is a stable random variable with index of stability α, scale parameter σ, location parameter μ and skewness β then we denote it as follows: $X \sim S_\alpha(\sigma, \mu, \beta)$.

Theorem 3.

Let X_1, X_2 are i.i.d. $\sim S_\alpha(\sigma_i, \mu_i, \beta_i)$, i=1,2 then $X_1 + X_2 \sim S_\alpha(\sigma, \mu, \beta)$ where

$$\sigma = (\sigma_1^\alpha + \sigma_1^\alpha)^{1/\alpha}$$

$$\beta = \frac{\beta_1\sigma_1^2 + \beta_2\sigma_2^2}{\sigma_1^\alpha + \sigma_1^\alpha}$$
$$\mu = \mu_1 + \mu_2$$

The location parameter is independent of α of the convolution of two α-stable random variables.

Theorem 4.

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ then $X + a \sim S_\alpha(\sigma, \beta, \mu + a)$

Theorem 5.

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be a non-zero real constant. Then

$$aX \sim S_\alpha(\mid a \mid \sigma, \operatorname{sgn}(a)\beta, a\mu)$$

if $\alpha \neq 1$ and

$$a \cdot X \sim S_\alpha\left(\mid a \mid \sigma, \operatorname{sgn}(a)\beta, a\mu - \frac{2}{\pi}a \ln |a|\sigma\beta\right)$$

if $\alpha = 1$

Theorem 6.

For any $0 < \alpha < 2$, $X \sim S_\alpha(\sigma, \beta, \mu) \iff -X \sim S_\alpha(\sigma, -\beta, -\mu)$

Theorem 7.

X is symmetric if and only if $\beta = 0$ and $\mu = 0$. X is symmetric about μ if $\beta = 0$.

Theorem 8.

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \neq 0$. Then X is strictly stable if and only if $\mu = 0$. Hence If $\alpha \neq 1$ then $X - \mu$ is strictly stable.

Theorem 9.

$X \sim S_1(\sigma, \beta, 0)$ is strictly stable if and only if $\beta = 0$.

Theorem 10.

Fix $0 < \alpha < 1, \delta > 0$ and let N_δ be a Poisson random variable with mean $EN_\delta = \delta^{-\alpha}$ and let $Y_{\delta,k}$,k=1,2,.. be i.i.d. positive random variables independent of N_δ with distribution

$$P(Y_{\delta,k} > \gamma) = \delta^\alpha \gamma^{-\alpha}$$

if $\gamma > \delta$

$$(Y_{\delta,k} > \gamma) = 1$$

if $\gamma < \delta$.Then the compound of random variables

$$X_\delta = \sum_{k=1}^{N_\delta} Y_{\delta,k}$$

converges in distribution as $\delta \longrightarrow 0$ to the stable distribution $X \sim S_\alpha(\sigma, 1, 0)$ with

$$\sigma^\alpha = \Gamma(1-\alpha)\cos\left(\pi\frac{\alpha}{2}\right)$$

Theorem 11.

Let X have distribution $X \sim S_\alpha(\sigma, \beta, 0)$ with $\alpha < 2$. Then there exist two random variables i.i.d. Y_1 and Y_2 with common distribution $X \sim S_\alpha(\sigma, 1, 0)$ such that

$$X =_d \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_1 - \left(\frac{1-\beta}{2}\right)^{1/\alpha} Y_2$$

if $\alpha \neq 1$

$$X =_d \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_1 - \left(\frac{1-\beta}{2}\right)^{1/\alpha} Y_2 + \sigma\frac{1+\beta}{\pi}\ln\left(\frac{1+\beta}{2}\right) - \frac{1-\beta}{\pi}\ln\left(\frac{1-\beta}{2}\right)$$

if $\alpha = 1$.

Theorem 12.

For all $\alpha < 1$ and fixed σ the family of the distribution $S_\alpha(\sigma, \beta, 0)$ is stochastically ordered in $\beta, -1 \leq \beta \leq 1$. If $X \sim S_\alpha(\sigma, \beta, 0)$ and $\beta_1 \leq \beta_2$ then

$$P(X_{\beta_1} \geq x) \leq P(X_{\beta_2} \geq x)$$

for all x.

Theorem 13.

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 1$. Then

$$\lim_{\gamma \longrightarrow \infty} \gamma^\alpha P\{X > \gamma\} = C_\alpha \frac{1+\beta}{2} \sigma^\alpha$$

and

$$\lim_{\gamma \longrightarrow \infty} \gamma^\alpha P\{X < -\gamma\} = C_\alpha \frac{1-\beta}{2} \sigma^\alpha$$

Where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1}$$

And this is a reason why the stable distributions are called Pareto-type distributions. The proofs of these theorem can be found in [4].

Remark 1.

Very often we use symmetric stable distributions with zero μ and β i.e. $S_\alpha(\sigma, 0, 0)$. These distributions will be denoted by $S\alpha S$.

Remark 2.

$S\alpha S$ distributions have relatively simple characteristic function

$$Eexp(i\theta X) = e^{-\sigma^\alpha |\theta|^\alpha}$$

Remark 3.

If $S\alpha S$ distribution has scale parameter equal to 1 then it is called a standardized $S\alpha S$-distribution.

Remark 4.

If $\alpha = 2$ then $\sigma^2 = \frac{1}{2} Var(X)$, e.g. $N(0,1) = S_2(\frac{1}{\sqrt{2}}, 0, 0)$

Chapter 4

Explicit Density of the Stable Distributions

There only three notoriously known stable distributions whose density function is explicitly known:

Normal Distribution

$\alpha = 2$, $X \sim S_2(\sigma, 0, \mu)$

Cauchy Distribution

$\alpha = 1$, $X \sim S_1(\sigma, 0, \mu)$

Levy Distribution

$\alpha = 0.5$, $X \sim S_{0.5}(\sigma, 1, \mu)$
This enables very easily to estimate their parameters e.g. by means of method of maximum likelihood. In general the density of the stable random variable with arbitrary parameters isn't known and this fact complicates the parameter estimation of them.

Chapter 5

Simulation

1.

Let $X \sim S_1(\sigma, 0, 0)$. It is the Cauchy distribution and it can be simulated by $\sigma \tan(\gamma)$ where $\gamma \sim U(-\pi/2, \pi/2)$. $X \sim S_1(\sigma, 0, \mu)$ can be simulated by $\sigma \tan(\gamma) + \mu$.

2.

Let $X \sim S_{1/2}(\sigma, 1, \mu)$. Then we can simulate it by $\sigma Z^{-2} + \mu$ where $Z \sim N(0, 1)$.

3.

Suppose that we have to simulate $S_{1/2^m}(1, 1, 0)$ and we have m random variables $X_1, X_2, .., X_m \sim N(0, 1)$. Then:

$$S_{1/2^m}(1, 1, 0) =_d \prod_{j=1}^{m} X_j^{-2j}.$$

Levy distribution is a particular case of this relation.

4.

Suppose that γ has uniform distribution $U(-\pi/2.\pi/2)$ and that W is exponentially distributed with unit mean. Let us also suppose that γ and W are

independent, and define:

$$X = \frac{\sin(\gamma\alpha)}{(\cos(\gamma))^{1/\alpha}} \left(\frac{\cos((1-\alpha)\gamma)}{W} \right)^{(1-\alpha)/\alpha}$$

is $S_\alpha(1,0,0)$. If $X \sim S_\alpha(1,\beta,0)$ then $\sigma X + \mu \sim S_\alpha(\sigma,\beta,\mu)$ if $\alpha \neq 1$ and $\sigma X + \mu + \frac{2}{\pi}\beta\sigma\ln(\sigma) \sim S_\alpha(\sigma,\beta,\mu)$ if $\alpha = 1$ so it enables to simulate any random variable with the distribution $S_\alpha(\sigma,0,\mu)$

5.

For the simulation of the stable random variable with arbitrary parameters one can apply some approximations. We can use a series representation. But we must realize that the series that we use for approximation converges very slowly therefore there is a need to use a large number of summands.

Usage of the Series Representation for simulation of an arbitrary stable distribution

Before this method was out of use because of the duration of the computation but the progress in computers has enabled us to use it nowadays. Notation:

$$\Gamma_j = \sum_{k=1}^{j} e_k$$

where $e_k \sim exp(1)$

Let $\{W_1, W_2, ..\}$ and $\{\Gamma_1, \Gamma_2..\}$ be two independent sequences of random variables and $\{W_1, W_2, ..\}$ is an i.i.d. sequence of random variables with finite absolute $\alpha + \delta th$ moment ($\delta > 0$). For the simulation of a stable random variable with arbitrary parameters we will use the following theorem:

Theorem (*)

The series

$$\sum_{j=1}^{\infty} (\Gamma_j^{-1/\alpha} W_j - k_j^{(\alpha)})$$

with $k_j^{(\alpha)} = 0$ if $0 < \alpha < 1$, $E(W_1 \int_{|W_1|/j}^{|W_1|/(j-1)} x^{-2}\sin(x)dx)$ if $\alpha = 1$ and $k_j^{(\alpha)} = \frac{\alpha}{1-\alpha}(j^{(\alpha-1)/\alpha} - (j-1)^{(\alpha-1)/\alpha})EW_1$ if $\alpha > 1$ converges almost surely to

a $S_\alpha(\sigma, \beta, \mu)$ random variable with

$$\sigma^\alpha = \frac{E|W_1|^\alpha}{c_\alpha}$$

$$c_\alpha = \int_0^\infty \frac{\sin(x)}{x^\alpha} dx$$

$$\beta = \frac{E|W_1|^\alpha sign(W_1)}{E|W_1|^\alpha}$$

Moreover in case $\alpha = 1$, the series

$$\sum_{j=1}^{\infty} (\Gamma_j^{-1/\alpha} W_j - k_j^{(1)})$$

converges almost surely to a $S_1(\sigma, \beta, \mu)$ random variable, where σ and β are as above and $\mu = -EW_1 \log |W_1|$. The theorem can be reformulated that any random variable $X \sim S_\alpha(\sigma, \beta, \mu)$, $0 < \alpha < 2$ can be represented as

$$X =_d \sum_{j=1}^{\infty} (\Gamma_j^{-1/\alpha} W_j - k_j^{(\alpha)})$$

with above defined parameters. We need to attain the random variable with a unit skewness and a zero mean what enables us to use the formula defined in the Theorem 13.

$$\beta = \frac{E|W_1|^\alpha sign(W_1)}{E|W_1|^\alpha}$$

so if $W_1 > 0$ almost surely than $\beta = 1$ almost surely. We can choose normally distributed random variables with such variance and mean that they would be larger than zero almost surely according to the law of 3 sigmas for the normal distribution. We can choose e.g. and $\sigma^2 = 4$ so the probability that this random variable is smaller than zero is very low and in this case $\beta = 1$a.s. So if we want to simulate a random variable $X \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$ we can do it as follows. We take a large number of summand for the series representation, e.g. 10000, simulate 10000 random numbers governed by Normal Distribution N(45,4) and ten thousand random numbers governed by Poisson Distribution and put it to the series defined in the theorem. The sum Y converges to the random variable $S_\alpha(\sigma, \beta, \mu)$. Having simulated two random numbers Y_1, Y_2 with the distribution $S_\alpha(\sigma, 1, 0)$ and using the Theorem 13. we get

$$X =_d ((1+\beta)/2)^{\frac{1}{\alpha}} Y_1 - ((1-\beta)/2)^{\frac{1}{\alpha}} Y_2$$

$$X \sim S_\alpha(\sigma, \beta_1, 0)$$
$$Z = \frac{\sigma_1}{\sigma} X \sim S_\alpha(\sigma_1, \beta_1, 0)$$
$$FX = Z + \mu_1 \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$$

So this is a way how we can get an approximation of a random variables with arbitrary parameters for instance with $\sigma = \sigma_1, \beta = \beta_1, \mu = \mu_1$. This operation can be conducted in MATHEMATICA 5. It will be conducted only for $\alpha > 1$ because it happens is very when the innovation has an infinite expected value.

7.

There is a formula for generation of random numbers from the stable distribution $S_\alpha(\tilde{\sigma}, \beta, \mu)$ with known α and β but with unknown σ. The formula is following:

$$Y_A(\alpha, \beta) = \frac{2}{\pi}\left[\left(\frac{\pi}{2}\beta\phi\right)\tan\phi - \beta \ln\left(\frac{(\pi/2)E\cos\phi}{\pi/2 + \beta\phi}\right)\right], \ \ \alpha = 1$$

and

$$Y_A(\alpha, \beta) = [1 + \beta^2 \tan^2(\alpha\pi/2)]^{1/(2\alpha)} \frac{\sin(\alpha(\phi + b))}{(\cos\phi)^{1/\alpha}} \cdot$$
$$\cdot \left[\frac{\cos(\phi - \alpha(\phi + b))}{E}\right]^{(1-\alpha)/\alpha}, \ \ \alpha \neq 1, \ \ (*)$$

where $\phi \sim U[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $E \sim \exp(1)$ The Kolmogorov Smirnov test rejected that $\sigma = 1$. It can be tested as follows:
Having simulated a sample $X_1, X_2, ..., X_N$ from $S_\alpha(\sigma, \beta, 0)$by means of $(*)$ we can make a new sample $X_1 - X_2, X_2 - X_4, ..., X_{N-1} - X_N$ which is a sample from $S_\alpha(\sigma, 0, 0)$. We know how to simulate symmetric stable random variable $S_\alpha(1, 0, 0)$. Having simulated a sample $Y_1, Y_2, .., Y_{\lfloor N/2 \rfloor}$ we can compare their empirical distribution functions and conduct Kolmogorov Smirnov test. But it refuses the hypothesis even for $(1 - \alpha) = 60\%$.
Therefore, we propose the following way how to simulate $S_\alpha(\sigma, \beta, \mu)$:
1. We simulate a sample $X = (X_1, X_2, ..., X_N)$ by means of (*).
2. We transform it to the sequence $X_1 - X_2, X_2 - X_4, ..., X_{N-1} - X_N \sim S_\alpha(\sigma, 0, 0)$ and estimate unknown σ. For the estimation of σ we can use method based on characteristic function that is described in the next chapters.

Chapter 6

Generalization of stable distributions

6.1 Geometric stable distributions

In this chapter we consider a special case of ν-stable distributions, where $\nu = \nu_p$ is geometrically distributed random variable, $p \in \Delta = (0, 1)$. Consider the ch.f. given by $\psi(t) = \varphi(-\log(f(t)))$ where φ is the solution of the Poincare equation:

$$\varphi(t) = P(\varphi(p_0 t))$$

$f(t)$ is a ch.f. of the stable law, the solution of this equation is $\varphi(t) = \frac{1}{1-t}$ so $\psi(t) = (1 - f(t))^{-1}$. Probability distributions with ch.f.

$$\psi(t) = \frac{1}{1 - f(t)}$$

are known as geometric stable distributions.

Definition

A random variable Y is said to be geometric stable with respect to the summation scheme if there exists a sequence of i.i.d. random variables $X_1, X_2, ...$, a geometric random variable ν_p independent of all X_j and con-

stants $a = a(p) > 0$ and $b = b(p) \in R$ such that

$$a(p)\sum_{j=1}^{\nu_p}(X_j + b(p)) \longrightarrow_d Y, as, p \longrightarrow 0$$

The geo-stable random variable is symmetric if for any Borel set $A \in R$

$$P(Y \in A) = P(-Y \in A).$$

(see [6])

Definition (equivalent to the previous)

A random variable Y is said to have GS-distribution if its characteristic function has a form:

$$\psi(t) = (1 + \sigma^\alpha |t|^\alpha w(t, \alpha, \beta) - i\mu t)^{-1}$$

where

$$w(t, \alpha, \beta) =$$

$$1 - \beta \tan(\pi\alpha/2)\mathrm{sgn}(t)$$

if $\alpha \neq 1$ and

$$= 1 + \beta\frac{2}{\pi}\log|t|\mathrm{sgn}(t)$$

if $\alpha = 1$
and

$$0 < \alpha \leq 2, -1 \leq \beta \leq 1, \sigma \geq 0, \mu \in R$$

The parameters $\alpha, \sigma, \beta, \mu$ are unique, the parameter β is irrelevant when $\alpha = 2$. The most important parameter is the index of stability α. β is skewness parameter and σ and μ are scale and location parameters respectively.

Remark.

Densities and distribution functions of GS-distributions are not known in closed form, except for a few special cases: exponential distribution ($\sigma = 0$) and Laplace distribution ($\alpha = 2$). If $\sigma = 0$ then we say that this distribution is improper and we call it proper otherwise. We shall denote the

geometric distribution with parameters $\sigma, \beta, \mu, \alpha$ (they are functions of p) by $GS_\alpha(\sigma, \beta, \mu)$ and write

$$Y \sim GS_\alpha(\sigma, \beta, \mu)$$

for GS distributed random variables. Stable and geometric stable distributions are very similar to each other.

Theorem

Every proper geometric stable distribution is absolutely continuous with the following representation of the distribution function GS and the density gs. For $\alpha \neq 1$

$$GS(x; \alpha, \sigma, \beta, \mu) = \int_0^\infty S\left(\frac{x - \mu z}{\sigma z^{1/\alpha}}, \alpha, \beta\right) e^{-z} dz,$$

$$gs(x; \alpha, \sigma, \beta, \mu) = \int_0^\infty \sigma^{-1} z^{-1} s\left(\frac{x - \mu z}{\sigma z^{1/\alpha}}, \alpha, \beta\right) e^{-z} dz.$$

For $\alpha = 1$

$$GS(x; \alpha, \sigma, \beta, \mu) = \int_0^\infty S\left(\frac{x - \mu z - z\sigma\beta(2/\pi)\log(\sigma z)}{\sigma z^{1/\alpha}}, \alpha, \beta\right) e^{-z} dz,$$

$$gs(x; \alpha, \sigma, \beta, \mu) = \int_0^\infty \sigma^{-1} z^{-1} s\left(\frac{x - \mu z - z\sigma\beta(2/\pi)\log(\sigma z)}{\sigma z^{1/\alpha}}, \alpha, \beta\right) e^{-z} dz.$$

S and s mean the distribution function and the density of the corresponding stable random variable with the same parameters respectively.

Corollary of the previous theorem

Let $Y \sim GS_\alpha(\sigma, \beta, \mu)$, $X \sim S_\alpha(\sigma, \beta, \mu)$, $Z \sim \exp(1)$. Let X and Z be independent. Then

$$Y = \mu Z + Z^{1/\alpha}\sigma X, \text{ if } \alpha \neq 1$$

and

$$Y = \mu Z + Z^{1/\alpha}\sigma X + \sigma Z\beta(2/\pi)\log(Z\sigma), \text{ if } \alpha = 1$$

This corollary is very important because this formula can be applied for the simulation of $GS_\alpha(\sigma, \beta, \mu)$ random variable. To do this we need to be able to simulate the stable distribution $S_\alpha(\sigma, \beta, \mu)$ and the exponential distribution

with the unit parameter.

We know how the stable random variable can be approximated by virtue of the series representation. The GS random variables can also be approximated by means of the series representation what follows from the next theorem.

Theorem

Let $e_0, e_1, e_2, ..., e_n, ...$ be a sequence of i.i.d. r.v. having standard exponential distribution. Suppose that $R_1, R_2, ..., R_n, ...$ are i.i.d. random variables, independent of the sequence $\{e_i\}$. If the series

$$\sum_{k=1}^{\infty} \left(\frac{e_0}{e_1 + e_2 + ... + e_k} \right)^{1/\alpha} R_k$$

converges almost surely, then it converges to a strictly geo-stable random variable. As it was mentioned above, the geometric stable laws belong to the domain of attraction of the stable laws. It can be proven by means of characteristic function. We will prove it only for $GS_\alpha(1, 0, 0)$. When we have other parameters the proof will be analogous:
1) If $X \sim GS_\alpha(1, 0, 0)$ then its characteristic function equals: $\psi(t) = \frac{1}{1+|t|^\alpha}$. If we have n i.i.d. random variables $X_1, X_2, ..., X_n$ having $GS_\alpha(1, 0, 0)$ then for $X = \sum_{j=1}^{n} X_j$ holds:

$$\mathrm{E} \exp(i \cdot t \cdot X) = \prod_{j=1}^{n} \mathrm{E} \exp(i \cdot t \cdot X_j) = \left(\frac{1}{1 + |t|^\alpha} \right)^n$$

and the characteristic function of $\frac{1}{n^{1/\alpha}} X$ equals:

$$\mathrm{E} \exp\left(i \cdot t \cdot \frac{X}{n^{1/\alpha}} \right) = \prod_{j=1}^{n} \mathrm{E} \exp\left(i \cdot t \cdot \frac{X_j}{n^{1/\alpha}} \right) = \prod_{j=1}^{n} \frac{1}{1 + \frac{|t|^\alpha}{n}} = \frac{1}{\left(1 + \frac{|t|^\alpha}{n}\right)^n}$$

2) The characteristic function of $Y \sim S_\alpha(1, 0, 0)$ equals $\exp(-|t|^\alpha)$
3)

$$\lim_{n \to \infty} \frac{1}{\left(1 + \frac{|t|^\alpha}{n}\right)^n} = \exp(-|t|^\alpha)$$

therefore, $\frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} X_j$ is a domain of attraction of $Y \sim S_\alpha(1, 0, 0)$ and converges to that in distribution.

6.2 Smoothly Truncated α-Stable Distributions

We introduce a special class of truncated distributions which were baptized as smoothly truncated distributions (STS-distributions).The name is due to the special form of tail truncation, which guarantees a continuously differentiable distribution function for the truncated α-stable distribution. (See [2]) Formally we have:

Definition
Let g_θ denote the density of an α-stable distribution with parameter $\theta = (\alpha, \sigma, \beta, \mu)^{\mathrm{T}}$ and $h_i, i = 1, 2$ with mean ν_i and standard deviation τ_i, $i = 1, 2$. Furthermore, let $a, b \in R$ be two real numbers with $a \leq m \leq b$, where m denotes the mode of g_θ. The density of STS-distribution is defined by:

$$f(x) = \begin{cases} h_1(x) & for \quad x < a \\ g_\theta(x) & for \quad x \in [a, b] \\ h_2(x) & for \quad x > b \end{cases}$$

In order to guarantee a well-defined continuous probability density, the following conditions are imposed:

$$h_1(a) = g_\theta(a), h_2(b) = g_\theta(b)$$

and

$$p_1 := \int_{-\infty}^{a} h_1(x)dx = \int_{-\infty}^{a} g_\theta(x)dx$$

and

$$\int_{b}^{\infty} h_2(x)dx = \int_{b}^{\infty} g_\theta(x)dx =: p_2$$

The family of STS distributions will be denoted by $\mathcal{S}$, the subclass of standardized STS distributions by $\mathcal{S}_0$. Element of $\mathcal{S}$ are denoted by $S_\alpha^{[a,b]}(\sigma, \beta, \mu)$. $p_1 = G_\theta(a)$ and $p_2 = 1 - G_\theta(b)$ denote the "cut-off-probabilities". Let the density and the cumulative distribution function of the standard normal distribution are denoted by φ and Φ respectively. Then the parameters (ν_i, τ_i) of two normal distributions describing the tails of the STS distributions can be obtained from the following two equations:

$$\tau_1 = \frac{\varphi(\Phi^{-1}(p_1))}{g_\theta(a)}, \nu_1 = a - \tau_1 \Phi^{-1}(p_1)$$

$$\tau_2 = \frac{\varphi(\Phi^{-1}(p_2))}{g_\theta(b)}, \nu_1 = b + \tau_2 \Phi^{-1}(p_2)$$

So the kernel of this method lies in the replacement of the tails the stable distributions with the normal tail of the same square with such density functions which equal the value of the stable density function at the border points. The normal distribution is characterized by two parameters and we can choose and set such expected values and variances that they will suit the requirements on the STS-distribution. Therefore we can't use many other types of tails, because many of them cannot suit the assumptions. The STS-distributions have many advantages because all their moments exist and can be used for their characteristics and to the characteristics of the model that they rule. A useful property of α-stable distributions - and normal in particular - is their scale and translation invariance, which is transmitted to the class of STS distributions: For $c, d \in R$ and $X \sim S_\alpha^{[a,b]}(\sigma, \beta, \mu)$ we have that the random variable $Y = cX + d$ as an affine transform of the variable X is again STS distributed, i.e.

$$Y \sim S_\alpha^{[\tilde{a},\tilde{b}]}(\tilde{\sigma}, \tilde{\beta}, \tilde{\mu}) \in \mathcal{S}$$

where

$$\tilde{a} = ca + d, \tilde{b} = cb + d, \tilde{\sigma} = |c|\sigma$$

$\tilde{\mu} = c\mu + d$ if $\alpha \neq 1$ and $\tilde{\mu} = c\mu - \frac{2}{\pi} c \log|c|\sigma\beta + d$ if $\alpha = 1$.

Pareto Distribution.

The distribution with probability density function and distribution function

$$f(x) = \frac{ab^a}{x^{a+1}}$$

$$F(x) = 1 - \left(\frac{b}{x}\right)^a$$

defined over the interval $x \geq b$.(See [6]) If $a < 1$ then this distribution has an infinite expected value and an infinite variance. If $1 < a < 2$ then there exists the expected value and is equal to $\frac{ab}{a-1}$ but the variance is infinite. As it was noted above, the stable distributions are Pareto type distributions and

their tails behave almost the same way. If $a > 2$ then there exists the second raw moment and is equal to:

$$\mu_2^{'} = \frac{ab^2}{a-2}.$$

For $a > n$ the nth central moment is

$$\mu_n = ab^n \Gamma(a-n)\tilde{F}_1\left(a-n, -n, 1+a-n; \frac{a}{a-1}\right) =$$

$$= (1-a)^{a-n}(-a)^{n-a}ab^n B\left(\frac{a}{a-1}; a-n, n+1\right)$$

for $a > n$ and where $\Gamma(z)$ is a gamma function, $\tilde{F}_1\left(a, b, c; d\right)$ is a regularized hypergeometric function, and $B(z; a, b)$ is a beta function.
The mean, variance, skewness, and kurtosis are therefore

$$\mu = \frac{ab}{a-1}$$

$$\sigma^2 = \frac{ab^2}{(a-1)^2(a-2)}$$

$$\gamma_1 = \sqrt{\frac{a-2}{a}}\frac{a(a+1)}{a-3}$$

$$\gamma_2 = \frac{6(a^3+a^2-6a-2)}{a(a-3)(a-4)}$$

provided that the corresponding moments exist. The description of the Hill estimator can found in Appendix A. This estimator has been shown to be unappropriate estimator of α in case of stable law but it estimates well the parameter α of the Pareto distribution with parameters 1 and α. The results are summarized in the following table:

α	Mean	Variance
1.1	1.156	0.242
1.2	1.251	0.229
1.3	1.361	0.248
1.4	1.458	0.293
1.5	1.526	0.296
1.6	1.679	0.304
1.7	1.799	0.362
1.8	1.883	0.326
1.9	1.944	0.372

Any Pareto distribution has an explicit expression of its density therefore it enables to estimate its parameters by virtue of ML methodology. In the later chapters are considered characteristic function based estimators and estimators based on the method of projections that also requires knowledge of explicit form of characteristic function. These estimates cannot be applied for the Pareto distributions because we don't possess explicit form of its characteristic function.

Chapter 7

Modified models based on the GARCH(p,q) model

7.1 Symmetric GARCH-stable Processes

A sequence of random variables $Y_n, n \in Z$ is said to be a stable $GARCH(\alpha, p, q)$ if:

1.$Y_n = \sigma_n S_n$, where S_n are i.i.d. r.v.'s with standard $S\alpha S$-distribution, $1 < \alpha \leq 2$,
2. there exist nonnegative constants α_i, $i = 1, ..., q$ and β_j, $j = 1, ..., p$ and $\delta > 0$, such that

$$\sigma_n = \delta + \sum_{i=1}^{p} \alpha_i |Y_{n-i}| + \sum_{j=1}^{p} \beta_j \sigma_{n-j}, n \in Z.$$

Our assumption that $\alpha > 1$ is not very restrictive because most of financial time series have finite mean. (See [7])

7.2 Asymmetric GARCH-stable Processes

Generalizing the stable GARCH process to the asymmetric case, sequence $y_t, t \in Z$, is said to be a stable Paretian Asymmetric GARCH process if

$$y_t = \mu_t + c_t\varepsilon_t, \varepsilon_t \sim S_\alpha(1, \beta, 0)$$

$$c_t = \alpha_t + \sum_{i=1}^{p} \alpha_i |y_{t-i} - \mu_{t-i}| + \sum_{j=1}^{q} \beta_j c_{t-j}$$

where $S_{\alpha\beta}$ means the stable Paretian distribution with the index of stability α and the skewness index $\beta \in [-1, 1]$, zero location parameter and a unit scale parameter.
More generally the expression of c_t looks like

$$|c_t|^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i |y_{t-i} - \mu_{t-i}|^\delta + \sum_{j=1}^{q} \beta_j |c_{t-j}|^\delta$$

The number δ in experiment in most cases equals 1,2 or α. It can be in the interval $(1, \alpha)$ because if $\delta \geq \alpha$ then c_t and y_t have an infinite variance.

An $S_{\alpha,\beta}GARCH(r, s)$ process with $1 < \alpha < 2$ has unique strictly stationary solution if $\alpha_i > 0, i = 0, .., r, \beta_j > 0, j = 1, .., s$, and

$$\gamma_{\alpha,\beta} \sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j \leq 1$$

where $\gamma_{\alpha,\beta} := E|\varepsilon_t|$ is driven by

$$\gamma_{\alpha,\beta} = \frac{2}{\pi}\Gamma\left(1 - \frac{1}{\alpha}\right)(1 + \tau_{\alpha,\beta}^2)^{\frac{1}{2\alpha}} cos\left(\frac{1}{\alpha} arctan(\tau_{\alpha,\beta})\right)$$

if $1 < \alpha < 2$ and

$$\gamma_{\alpha,\beta} = \sqrt{2/\pi}$$

if $\alpha = 2$ with $\tau_{\alpha,\beta} = \beta \cdot tan\frac{\alpha \cdot \pi}{2}$. (See [7])

7.3 A Generalized NGARCH Option Pricing Model (See [2])

In this section we introduce a general option pricing models containing most of the features of GARCH process. Formally the log-returns of the underlying asset are assumed to follow the following dynamic under the objective probability measure P:

$$\log S_t - \log S_{t-1} = r_t - s_t + \gamma_t \sigma_t - g(\sigma_t) + \sigma_t \varepsilon_t, t \in N, \varepsilon_t \sim F, \varepsilon_t \sim iid, (*)$$

S_t denotes the price of the underlying ex dividend at date t and r_t and d_t denote the continuously compounded risk free rate of return and dividend rate respectively for the period $[t-1, t]$. Both quantities as well as γ_t are assumed to be predictable, but can in general be modeled by separate stochastic process. F denotes the marginal distribution of the innovation process and we assume F is a standardized continuous probability distribution whose support equals the whole real line R and whose moment generating function is m is finite. g represents the logarithmic moment generating function of F, i.e. we have $g(u) = \log \int \exp(ux) dF$. The conditional variance σ_t^2 is assumed to follow an asymmetric NGARCH(1,1)-process:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\varepsilon_{t-1} - \gamma)^2 + \beta_1 \sigma_{t-1}^2, t \in N (**)$$

where we assume $\alpha_1(1+\gamma^2) + \beta_1 < 1$ to guarantee the existence of a strong stationary solution with the finite unconditional mean.
From the definition (*) we can deduce some characteristic properties of the process dynamic of the underlying. First we mention, that if the distribution of the innovations F equals the standard Gaussian distribution and if we assume constants r, d and $\bar{\gamma}$ then the equation (*) reduces to Duan option pricing model. For $\alpha_1 = \beta_1 = 0$ the model boils down to the discrete time Black-Scholes model.
Let's now introduce some of modifications of the models:

DGMB:

The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, standard normally distributed innovations, and constant conditional variance $\alpha_1 = \beta_1 = 0$

Gaussian GARCH:

The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, with standard normally distributed innovations and symmetric GARCH conditional variance ($\gamma = 0$).

Gaussian NGARCH:

The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, with standard normally distributed innovations and asymmetric NGARCH conditional variance.

GED-NGARCH:

The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, with standardized GED-distributed innovations, and asymmetric NGARCH conditional variance.

Skewed-t-NGARCH:

The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, with standardized skewed t-distributed innovations, and asymmetric NGARCH conditional variance. The logarithmic moment distribution function in (*) is replaced by one of standard normal distribution. **STS-NGARCH:**The model as given by equations (*) and (**) with constant price of risk $\bar{\gamma}$, with standardized STS-distributed innovations, and asymmetric NGARCH conditional variance. This case is subdivided. First we estimate the STS-distribution with fixed parameters (STS-NGARCH fixed) and in the second case estimate a standardized STS-distribution
(STS-NGARCH estimated).

All the models are estimated by a numerical maximum likelihood routine. Let us assume for a moment, that the parameters of the standardized distribution F governing the innovation process are known. If we denote by l the logarithm of the corresponding density function, then the standard argument leads to the following conditional log-likelihood function for the NGARCH

stock-price model with innovation distribution F:

$$\tilde{L}_{(x,y)}(\bar{\gamma}, \alpha_0, \alpha_1, \beta_1, \gamma) = \sum_{t=1}^{T} l\left(\frac{y_t - r_t + d_t - \bar{\gamma}\sigma_t + g(\sigma_t)}{\sigma_t}\right) - \frac{1}{2}\log(\sigma_t).$$

The conditional variance is recursively obtained from:

$$\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2(\varepsilon_{t-1} - \gamma)^2 + \beta_1\sigma_{t-1}^2, t \in N(**).$$

$y_t = S_t - S_{t-1}$ denotes a series of log-returns. As usual the conditional likelihood function depends on the choice of the starting values ε_0 and σ_0 but for a large sample the impact of the starting values is negligible. Maximizing the conditional likelihood function leads to estimates for the unknown parameters $(\bar{\gamma}, \alpha_0, \alpha_1, \beta_1, \gamma)$. From the estimates we can recursively recover the time series of of empirical residuals $(\varepsilon_t)_{t=1,..,T}$. The empirical distribution $\hat{F}$ of these residuals can now be used to update the distributional assumption F by estimating new distributional parameters. The estimation of the model parameters can now be repeated with the updated distribution and this gives raise to an iterative procedure. After each iterative step we calculate the Kolmogorov-Smirnov distance $d(F, \hat{F})$ between the distributional assumption and the empirical distribution of the residuals. We finish the iteration procedure and the estimation procedure ends as soon as this distance stops decreasing. We judge the appropriateness of the tests by means of the Kolmogorov-Smirnov statistic and the p-value. The practice shows that in many cases the STS-NGARCH model fits the data the best way. Due to its modeling flexibility, the class of STS distributions turns out to be a viable alternative to other popular heavy-tailed distributions.
The following example will emphasize this fact. It deals with one of the most problematic events from a statistical viewpoint in recent decades: The October crash in 1987. This event has the reputation to be "unexplainable" by every reasonable time series model. Christian Menn and Svetlozar T. Rachev in their book "Smoothly Truncated Stable Distributions, GARCH-Models and Option Pricing", June 2005, studied the ability of different models to forecast the October crash. They compared all six models (DGMB,Gaussian GARCH and NGARCH, GED-NGARCH, skewed-t-NGARCH and STS-GARCH). The data are fitted to the same data series consisting of 1000 S&P 500 log-returns preceding the crash and ending with the observation on Friday October 16, 1987. On the next trading day, namely

on the 19 October, the S&P 500 dropped by more than 20%. Having estimated a specific time series model, we can express the drop which occurred on Black Monday in terms of some implied realization for the residual $\hat{\varepsilon}_{Oct.19}$. Given the value of the "crash residual", we can derive the model dependent of implicit probability $\hat{p} = P(\varepsilon \leq \hat{\varepsilon}_{Oct.19})$ for such an event. Having compared the log-result and using Kolmogorov-Smirnov and Anderson-Darling statistic they concluded that that crash could have been forecasted by the STS-NGARCH model the best. The mean time of occurrence of such event, according to their calculations is approximately 25 years.

Chapter 8

CF based estimators of parameters of stable distributions

8.1 Description of CF based estimates

Let's have a sample $X_1, X_2, ..., X_N \sim S_\alpha(\sigma, \beta, \mu)$. The problem lies in estimation of the parameter $\theta = (\alpha, \sigma, \beta, \mu)$. The most difficult is an estimation of α. Therefore, we will first of all look for the most effective estimate for α of the standard symmetric distribution $S_\alpha(1, 0, 0)$ and since then we will consider more complicated problems. There are lots of conventional methods for α estimations, e.g. Hill estimate, Pickand estimate, MUP estimate and others. In appendix A there are described properties of Hill and Pickand estimators. In this work one of methods that we will apply is the method based on the characteristic function. The characteristic function is explicitly known for any stable distribution therefore we will look for estimate of α comparing theoretical and empirical distribution characteristic function. Let us denote the estimates based on characteristic function as CF-estimates. What problems can arise when we apply this method? First of all empirical characteristic function doesn't converge uniformly on R as against empirical distribution function, which converges to theoretical one uniformly, according to Glivenko-Cantelli theorem. Moreover empirical characteristic function behaves like periodic function. This problem can be solved if we apply Scőrgö theorem: (See [8],[3])

Scőrgö theorem

For any d-variate characteristic function C, if $\lim_{n\to\infty}(T_n)/n = 0$ then

$$\lim_{n\to\infty} \Delta_n(T_n) = 0$$

almost surely, where

$$\Delta_n = \sup_{|t|\le T_n} |C_n(t) - C(t)|,$$

$$C(t) = \int_{R^d} \exp(i\langle t, x\rangle) dF(x)$$

$$C_n(t) = \int_{R^d} \exp(i\langle t, x\rangle) dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{\langle t, X_j\rangle}$$

Proof

Let $\varepsilon > 0$ be arbitrarily small, $\varepsilon \le 2$, and choose $K = K(\varepsilon, F)$ so large that

$$\int_{|x|\ge K} dF(x) < \frac{\varepsilon}{8}$$

Writing $D_n(t) = B_n(t) - B(t)$, we have

$$\Delta_n(T_n) = \sup_{|t|\le T_n} |D_n(t)| + \sup_{|t|\le T_n} |B_n(t) - B(t)| + \sup_{|t|\le T_n} |B(t) - C(t)|$$

with the truncated integrals

$$B(t) = \int_{|x|\le K} e^{i\langle t,x\rangle} dF(x)$$

$$B_n(t) = \int_{|x|\le K} e^{i\langle t,x\rangle} dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{i\langle t, X_j\rangle}\chi(|X_j| \le K)$$

where $\chi(A)$ denotes the indicator of the event A. The second term is

$$\frac{1}{n}\sup_{|t|\le T_n}\left|\sum_{j=1}^{n} e^{i\langle t,X_j\rangle}\chi(|X_j| > K)\right| \le \frac{1}{n}\sup_{|t|\le T_n}\left|\sum_{j=1}^{n}\chi(|X_j| > K)\right|$$

and these bounds converge almost surely to

$$\int_{|x|>K} dF(x)$$

which is also a bound for the third term.
Let us cover the cube $[-T_n, T_n]^d$ by $N_n = \left([8Kd^{3/2}T_n/\varepsilon] + 1\right)^d$ disjoint small cubes $\Lambda_1, \Lambda_2, ..., \Lambda_{N_n}$, the edges of which are of the length $\varepsilon/(4Kd^{3/2})$, and let $t_1, t_2, ..., t_{N_n}$ be the centers of these cubes. Then

$$\sup_{|t|\le T_n} |D_n(t)| \le \max_{1\le k\le N_n} |D_n(t_k)| + \max_{1\le k\le N_n} \sup_{t\in\Lambda_k} |D_n(t) - D_n(t_k)| \le$$

$$\max_{1\le k\le N_n} |D_n(t_k)| + \frac{\varepsilon}{4}$$

for

$$|D_n(s) - D_n(t)| \le |B_n(s) - B_n(t)| + |B(s) - B(t)| \le$$

$$\le \frac{1}{n}\sum_{j=1}^{n} |\langle s-t, X_j\rangle|\chi(|X_j| \le K) + \int_{|x|\le K} |\langle s-t, x\rangle| dF(x) \le 2dK|s-t|, \;\; s,t \in R^d$$

Summing up:

$$\Delta_n(T_n) \le \max_{1\le k\le N_n} |D_n(t_k)| + \frac{\varepsilon}{2}$$

almost surely for large enough n, the threshold depending on ω. Now

$$p_n = P\left\{\max_{1\le k\le N_n} |D_n(t_k)| > \frac{\varepsilon}{2}\right\} \le N_n \sup_{t\in R^d} P\left\{|D_n(t)| > \frac{\varepsilon}{2}\right\} \le$$

$$\le MT_n^d \sup_{t\in R^d}\left(P\left\{\frac{1}{n}\left|\sum_{j=1}^{n} R_j(t)\right| > \frac{\varepsilon}{4}\right\} + P\left\{\frac{1}{n}\left|\sum_{j=1}^{n} I_j(t)\right| > \frac{\varepsilon}{4}\right\}\right)$$

with some constant $M = M(\varepsilon, F, d)$, where the random variables

$$R_j(t) = (\cos\langle t, X_j\rangle)\chi(|X_j| \le K) - \int_{|x|\le K} \cos\langle t, x\rangle dF(x), \;\; j = 1, .., n$$

are independent, $|R_j(t)| \le 2, ER_j(t) = 0$, and

$$v^2(t) = \int_{|x|\le K} \cos^2\langle t, x\rangle dF(x) - \left(\int_{|x|\le K} \cos\langle t, x\rangle dF(x)\right)^2 \le 1$$

The random functions are defined with the cosine functions replaced by sine, and hence these all are also i.i.d. with $|I_j(t)| \leq 2, EI_j(t) = 0$ and $EI_j^2(t) \leq 1$. Therefore the Bernstein equality gives

$$P\left\{\frac{1}{n}\left|\sum_{j=1}^{n} R_j(t)\right| > \frac{\varepsilon}{4}\right\} = 2\exp\left(-\frac{\varepsilon n}{32}\right) \; if \; \varepsilon \geq 2v^2(t)$$

$$P\left\{\frac{1}{n}\left|\sum_{j=1}^{n} R_j(t)\right| > \frac{\varepsilon}{4}\right\} = 2\exp\left(-\frac{\varepsilon^2 n}{64v^2(t)}\right) \; if \; \varepsilon \leq 2v^2(t)$$

Since $v^2(t) \leq 1$ and $\varepsilon \leq 2$, the probability in question is not greater than $2\exp(-\varepsilon^2 n/64)$, and the same holds for other one with the I_j's. Thus

$$p_n \leq 4MT_n^d \exp\left(-\frac{\varepsilon^2 n}{64}\right)$$

Let $\delta < \varepsilon^2/(64d)$. Then for large enough n, $T_n \leq \exp(\delta n)$, and hence

$$\sum_{j=1}^{\infty} p_n < \infty.$$

The Borel-Cantelli lemma and

$$\Delta_n(T_n) \leq \max_{1 \leq k \leq N_n} |D_n(t_k)| + \frac{\varepsilon}{2}$$

give the desired result.QED.
When estimating α from the sample $X_1, .., X_n$ from $S_\alpha(1, 0, 0)$ we have the following representations of empirical characteristic functions and theoretical one:

$$C(t, \alpha) = \exp(-|t|^\alpha)$$

and

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n} \cos(tX_j)$$

because in symmetric case the imaginary part of the characteristic function is zero, therefore in

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itX_j) = \frac{1}{n}\sum_{j=1}^{n} \cos(tX_j) + i\frac{1}{n}\sum_{j=1}^{n} \sin(tX_j)$$

so

$$\Re(C_n(t)) \to C(t)$$

and

$$\Im(C_n(t)) \to 0$$

for $t \in [-T_n, T_n]$ such that:

$$\lim_{n\to\infty} \frac{\ln(T_n)}{n} = 0$$

In our case

$$\Delta_n(T_n) = \sup_{|t|\le T_n} |C_n(t) - C(t,\alpha)|, \quad \lim_{n\to\infty} \frac{\ln(T_n)}{n} = 0$$

So if α is unknown then we can solve the problem:

$$(1) \qquad \min_{\alpha} \sup_{|t|\le T_n} |C_n(t) - C(t,\alpha)|,$$

The function's $f_n(a) = \sup_{|t|\le T_n} |C_n(t) - C(t,a)|$, $a \in (0,2]$ minimum possible value is evidently zero. Moreover

$$\lim_{n\to\infty} f_n(\alpha) = 0$$

according to Scőrgö theorem. So if we denote $\hat{\alpha}_n$ the solution of the problem (1) then α is the solution of the limiting problem. Let's prove that this solution is unique. Suppose that $\tilde{\alpha}$ is another solution of the problem

$$\lim_{n\to\infty} \min_{\alpha} \sup_{|t|\le T_n} |C_n(t) - C(t,\alpha)|$$

such that $\tilde{\alpha} \in (0,2]$ and $\alpha \neq \tilde{\alpha}$. But in this case $C_n(t)$ converges to another characteristic function $C(t) = \exp(-|t|^{\tilde{\alpha}})$ but this contradicts Scőrgö theorem. Moreover the function $|\exp(-{c_1}^a) - c_2|$ where $c_1 > 0, c_2 \in R$, (c_1 and c_2 are constants) is monotone in R, therefore our problem can't have two solutions. From this follows that α is unique solution of the problem

$$\lim_{n\to\infty} \min_{\alpha} \sup_{|t|\le T_n} |C_n(t) - C(t,\alpha)|$$

and

$$\alpha = \lim_{n\to\infty} \arg \min_{\tilde{\alpha}} \sup_{|t|\le T_n} |C_n(t) - C(t,\tilde{\alpha})|$$

From this and the uniform convergence follows that $\hat{\alpha}_n \to \alpha$. Solving this problem is computationally very difficult. Therefore instead of looking for minimum of the supreme we can look for the the minimum of the sum of absolute differences. That is:

$$\min_{\tilde{\alpha}} \sum_{j=-m}^{m} |C(t_j, \tilde{\alpha}) - C_n(t)|$$

or

$$\min_{\tilde{\alpha}} \sum_{j=1}^{m} |C(t_j, \tilde{\alpha}) - C_n(t)|.$$

From

$$\lim_{n\to\infty} \sup_{|t|\leq T_n} |C_n(t) - C(t, \alpha)| = 0$$

follows that

$$\lim_{n\to\infty} |C_n(t_k) - C(t_k, \alpha)| = 0$$

for any $t_k \in [-T_n, T_n]$. Therefore if we consider the function

$$f_{m,n}(a) = \sum_{j=1}^{m} |C(t_j, \tilde{\alpha}) - C_n(t)|$$

then evidently $\lim_{n\to\infty} f_{m,n}(\alpha) = 0$, for any $t_k \in [-T_n, T_n]$ and any $m \in N$. And this solution is unique what follows from the same reasons as for the minimum of the supreme. If m=1, i.e. $f_{1,n}(a)$ then our problem looks like

$$\min_{\tilde{a}} f_{m,n}(\tilde{a}) \Leftrightarrow |C(t_1, a) - C_n(t_1)| = 0 \Leftrightarrow C(t_1, a) - C_n(t_1) = 0$$

$$\exp(-|t_1|^a) = C_n(t_1)$$

For large enough n, $C_n(t_1)$ will be a positive number and this equation in a will have a solution.

$$\exp(-|t_1|^a) = C_n(t_1) \Leftrightarrow -|t_1|^a = \ln(C_n(t_1)) \Leftrightarrow |t_1|^a = \ln\left(\frac{1}{C_n(t_1)}\right)$$

$$\Leftrightarrow a\ln(|t_1|) = \ln\left(\ln\left(\frac{1}{C_n(t_1)}\right)\right)$$

Hence

$$\hat{\alpha}_n = \frac{\ln\left(\ln\left(\frac{1}{C_n(t_1)}\right)\right)}{\ln(|t_1|)}$$

Checking of the unbiasedness of this estimator is difficult, but its consistency can be easily proven. $t_1 \in [-T_n, T_n]$ therefore $C_n(t)$ converges uniformly to $\exp(-|t_1|^\alpha)$ so for large enough n, we can replace $C_n(t)$ with $C(t)$ i.e.

$$\hat{\alpha}_n = \frac{\ln\left(\ln\left(\frac{1}{C_n(t_1)}\right)\right)}{\ln(|t_1|)} \approx \frac{\ln\left(\ln\left(\frac{1}{C(t_1)}\right)\right)}{\ln(|t_1|)} =$$

$$= \frac{\ln\left(\ln\left(\frac{1}{\exp(-|t_1|^\alpha)}\right)\right)}{\ln(|t_1|)} = \frac{\ln\left(-\ln\left(\exp(-|t_1|^\alpha)\right)\right)}{\ln(|t_1|)} = \frac{\ln(-(-|t_1|^\alpha))}{\ln(|t_1|)} = \alpha\frac{\ln(|t_1|)}{\ln(|t_1|)} = \alpha.$$

That is $\hat{\alpha}_n \to \alpha$ as $n \to \infty$ i.e. this estimate is consistent. The quality of the estimator depends on t_1 because $\hat{\alpha}_n$ is a function of t_1. Therefore it is better to take more than one point to comprise as much of the information as possible. α is a tail index, the more points are concentrated in the tails, the more information about α we possess. But the information contained in the tails of the density function corresponds to one in the vicinity of zero of the characteristic function. Therefore the choice of points can be following:

$$t_1 = (1\Delta)^2, t_2 = (2\Delta)^2, t_3 = (3\Delta)^2, ..., t_m = (m\Delta)^2$$

where

$$\Delta = \frac{\sqrt{\ln n}}{m}$$

so that $t_m = \ln n$ $T_n = \ln n$ follow all the assumptions of Scőrgö theorem, therefore such choice is appropriate. We chose a power as a concentrator of point in the vicinity of zero as one of possible alternatives. If m is large enough then the points $t_j, j = 1, .., m$ will be concentrated in the vicinity of zero and part of them will also comprise an information about the distribution in other parts of the characteristic function. We will also apply other choice of points, such that:

$$t_1 = (1\Delta)^8, t_2 = (2\Delta)^8, t_3 = (3\Delta)^8, ..., t_m = (m\Delta)^8$$

where

$$\Delta = \frac{(\ln n)^{1/8}}{m}$$

so that $t_m = \ln n$ what implies a larger concentration of point around zero. When $m > 1$ we can't solve the problem $f_{m,n}(a) = 0$ because it is a sum of absolute values and it requires that every summand equals to zero but this is impossible in general. Therefore we will look for a minimum of

$$f_{m,n}(a) = \sum_{j=1}^{m} |C(t_j, a) - C_n(t_j)|$$

as it was proposed in the beginning of the chapter. We've chosen $T_n = \ln n$ because such choice is appropriate and corresponds to all assumptions and requirements of Scőrgö theorem. The algorithm is following:

1.
$$T_n = \ln n, x = 2 \ or \ x = 8, \Delta = \frac{(\ln n)^{1/x}}{m}$$

2.
$$t_1 = (1\Delta)^x, t_2 = (2\Delta)^x, ..., t_m = (m\Delta)^x$$

3.
$$f_{m,n}(a) = \sum_{j=1}^{m} |C(t_j, a) - C_n(t_j)|$$

4.
$$\hat{\alpha}_n = \arg\min_a f_{m,n}(a)$$

As it was proven above $\hat{\alpha}_n$ is consistent. The lack of explicit expression for α-estimate makes it difficult to verify its unbiasedness and normality. Therefore, we should work with the empirical distribution function of $\hat{\alpha}_n$ and verify unbiasedness and normality by means of mean of sample $(\hat{\alpha}_n)_k, k = 1, .., T$, (where T is a number of simulations) and Kolmogorov-Smirnov test respectively. Besides this estimator we can apply other method also based on the characteristic function. Having the sample $X_1, X_2, .., X_n$ we construct $C_n(t)$ and solve the following problems:

$$C_n(X_j) = \exp(-|X_j|^a) \Rightarrow (\hat{\alpha}_n)_j = \frac{\ln\left(\ln\left(\frac{1}{C_n(X_j)}\right)\right)}{\ln(|X_j|)}$$

Hence we calculate the mean of the sample $(\hat{\alpha}_n)_j, j = 1, .., n$

$$\bar{\hat{\alpha}}_n = \frac{(\hat{\alpha}_n)_1 + (\hat{\alpha}_n)_2 + ... + (\hat{\alpha}_n)_j}{n}$$

If $|X_j| > T_n$ we simply exclude it from the sample. Any summand in the sample mean converges to α therefore $\bar{\hat{\alpha}}_n$ is also a consistent estimator.

1.

σ and α are unknown, $\mu = 0$, $\beta = 0$. $C(t) = \exp(-\sigma^\alpha |t|^\alpha)$ i.e. the characteristic function is real-valued therefore we apply only real part of the empirical characteristic function i.e.

$$C_n(t) = \frac{1}{n} \sum_{j=1}^{n} \cos(tX_j)$$

Let us define the function

$$f_{m,n}(\tilde{\alpha}, \tilde{\sigma}) = \sum_{j=1}^{m} |C_n(t_j) - C(t_j, \tilde{\alpha}, \tilde{\sigma})|$$

$t_j \in [-T_n, T_n]$, $\lim_{n\to\infty} \ln T_n/n = 0$ According to Scőrgö theorem

$$\lim_{n\to\infty} f_{m,n}(\alpha, \sigma) = 0$$

for any $m \in N$ and any $t_j \in [-T_n, T_n]$. This solution is also unique, any other solution will be in contradiction with Scőrgö theorem. Therefore the solutions $\hat{\alpha}_n$ and $\hat{\sigma}_n$ of the problem

$$\min_{\tilde{\alpha},\tilde{\sigma}} f_{m,n}(\tilde{\alpha}, \tilde{\sigma}), \quad \tilde{\alpha} \in (0, 2], \quad \sigma > 0$$

are consistent because of uniform convergence on $[-T_n, T_n]$. Most of the information about σ is concentrated around mean in the density function in our case around zero. Therefore we need more points lying in the tails of the characteristic function. For this reason we only increase T_n to $T_n = 2 \ln n$ to comprise more points lying further from zero. Or we can look for a minimum of function

$$f_{m_1+m_2,n}(\tilde{\alpha}, \tilde{\sigma}) = \sum_{j=1}^{m_1} |C_n(t_j) - C(t_j, \tilde{\alpha}, \tilde{\sigma})| + \sum_{i=1}^{m_2} |C_n(s_j) - C(s_j, \tilde{\alpha}, \tilde{\sigma})|$$

where $t_j = (j\Delta)^8$, $j = 1, 2, ..., m_1$, $\Delta = (\ln n)^{1/8}/m_1$, $t_{m_1} = \ln n$, $s_i = \frac{i \ln n}{m_2}$, $s_{m_2} = \ln n$. So we can comprise large and small values, that improve the quality of α- and σ-estimates respectively. If m=2 then we can solve the problem $f_{2,n} = 0$ which is equivalent to the system:

$$(1) \qquad \exp(-\sigma^\alpha |t_1|^\alpha) = C_n(t_1)$$

$$(2) \qquad \exp(-\sigma^\alpha |t_2|^\alpha) = C_n(t_2)$$

$$-\sigma^\alpha |t_1|^\alpha = \ln(C_n(t_1))$$

$$-\sigma^\alpha |t_2|^\alpha = \ln(C_n(t_2))$$

$$\frac{-\sigma^\alpha |t_1|^\alpha}{-\sigma^\alpha |t_2|^\alpha} = \left|\frac{t_1}{t_2}\right|^\alpha = \frac{\ln(C_n(t_1))}{\ln(C_n(t_2))}$$

Hence

$$\hat{\alpha}_n = \ln\left(\frac{\ln(C_n(t_1))}{\ln(C_n(t_2))}\right) / \ln\left|\frac{t_1}{t_2}\right|$$

Hence

$$-(\sigma|t_1|)^{\hat{\alpha}_n} = \ln C_n(t_1)$$

$$(\sigma|t_1|)^{\hat{\alpha}_n} = \ln\left(\frac{1}{C_n(t_1)}\right)$$

$$\sigma|t_1| = \left(\ln\left(\frac{1}{C_n(t_1)}\right)\right)^{1/\hat{\alpha}_n}$$

Hence

$$\hat{\sigma}_n = \frac{\left(\ln\left(\frac{1}{C_n(t_1)}\right)\right)^{1/\hat{\alpha}_n}}{|t_1|}$$

Both estimates are consistent. $C_n(t) \to C(t)$ uniformly on $[-T_n, T_n]$ and:

$$\hat{\alpha}_n = \ln\left(\frac{\ln(C_n(t_1))}{\ln(C_n(t_2))}\right) / \ln\left|\frac{t_1}{t_2}\right| \approx \ln\left(\frac{\ln(C(t_1))}{\ln(C(t_2))}\right) / \ln\left|\frac{t_1}{t_2}\right| =$$

$$\ln\left(\frac{\ln(\exp(-\sigma^\alpha |t_1|^\alpha))}{\ln(\exp(-\sigma^\alpha |t_2|^\alpha))}\right) / \ln\left|\frac{t_1}{t_2}\right| = \alpha \ln\left|\frac{\sigma t_1}{\sigma t_2}\right| / \ln\left|\frac{t_1}{t_2}\right| = \alpha$$

where $\approx$ means that for large n they are almost equal or limit limit equality. Applying the same approach to $\hat{\sigma}_n$ we can prove consistency of that estimator.

2.

α and σ are known and we estimate μ and β. In financial practice α is always larger than one, but we can't estimate its expected value by $EX = \mu$, therefore we will also estimate it by the method based on characteristic function. The empirical characteristic function for $\beta \neq 0$ and $\mu \neq 0$ is complex-valued.

$$F_{m,n}(\tilde{\beta}, \tilde{\mu}) = \sum_{j=1}^{m} \left| \exp\left(\mu i - \sigma^{\alpha}|t_j|^{\alpha}\left(1 + i\beta\frac{t}{|t|}\tan\left(\frac{\pi\alpha}{2}\right)\right)\right) - C_n(t_j)\right|$$

According to Scőrgö theorem

$$\lim_{n\to\infty} F_{m,n}(\beta, \mu) = 0$$

and the solutions $\hat{\beta}_n$ and $\hat{\sigma}_n$ of the problem

$$\min_{b,\tilde{\mu}} F_{m,n}(b, \tilde{\mu})$$

are consistent. For the location parameter nor the skewness parameter there is no need to explore the tails therefore we don't have to concentrate the point around zero. When choosing the points $t_j, j = -m, -m+1, .., -1, 0, 1, 2, .., m$ we will make equal distance between adjacent points. $T_n = 2\ln n$,

$$t_j = sign(j)\frac{2\ln n \cdot j}{m}$$

Let us have a function with $t > 0$:

$$f_{2,m}(\beta, \mu) = \left| \exp\left(\mu i - \sigma^{\alpha}|t|^{\alpha}\left(1 + i\beta\frac{t}{|t|}\tan\left(\frac{\pi\alpha}{2}\right)\right)\right) - C_n(t)\right| +$$

$$+ \left| \exp\left(\mu i - \sigma^{\alpha}|-t|^{\alpha}\left(1 + i\beta\frac{-t}{|t|}\tan\left(\frac{\pi\alpha}{2}\right)\right)\right) - C_n(-t)\right| =$$

$$= \left|\exp\left(\mu i - \sigma^\alpha |t|^\alpha \left(1 + i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) - C_n(t)\right| +$$
$$+ \left|\exp\left(\mu i - \sigma^\alpha |t|^\alpha \left(1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) - C_n(-t)\right|$$

We can solve $f_{2,n}(b, m) = 0$ as follows:
The problem

$$f_{2,n}(b, m) = 0$$

is equivalent to the system:

$$\exp\left(\mu i - \sigma^\alpha |t|^\alpha \left(1 + i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) = C_n(t)$$

$$\exp\left(\mu i - \sigma^\alpha |t|^\alpha \left(1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) = C_n(-t)$$

$$\mu i - \sigma^\alpha |t|^\alpha \left(1 + i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right) = \ln(C_n(t))$$

$$\mu i - \sigma^\alpha |t|^\alpha \left(1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right)\right) = \ln(C_n(-t))$$

$$\ln\left(\frac{C_n(t)}{C_n(-t)}\right) = -\sigma^\alpha |t|^\alpha 2i\beta \tan(\pi\alpha/2)$$

Hence follows

$$\hat{\beta}_n = \Re \frac{\ln\left(\frac{C_n(-t)}{C_n(t)}\right)}{\sigma^\alpha |t|^\alpha 2i \tan(\pi\alpha/2)}$$

$C_n(-t)$ and $C_n(t)$ converge to $C(-t)$ and $C(t)$ respectively. Therefore

$$\hat{\beta}_n = \frac{\ln\left(\frac{C_n(-t)}{C_n(t)}\right)}{\sigma^\alpha |t|^\alpha 2i \tan(\pi\alpha/2)} \approx \frac{\ln\left(\frac{C(-t)}{C(t)}\right)}{\sigma^\alpha |t|^\alpha 2i \tan(\pi\alpha/2)} = \beta$$

i.e. this estimate is consistent.

$$i\mu = \ln(C_n(t)) + \sigma^\alpha |t|^\alpha \left(1 - i\hat{\beta}_n \tan\left(\frac{\pi\alpha}{2}\right)\right)$$

Hence:

$$\hat{\mu}_n = \Re\left(\frac{1}{i}\left(\ln(C_n(t)) + \sigma^\alpha |t|^\alpha \left(1 - i\hat{\beta}_n \tan\left(\frac{\pi\alpha}{2}\right)\right)\right)\right)$$

From the same considerations as for the estimation of other parameters follows that $\hat{\mu}_n$ is also consistent estimate of μ.

3.General case

Let's have a sample

$$(X_1, X_2, ..., X_n)$$

from $S_\alpha(\sigma, \beta, \mu)$ where all parameters are unknown. Then

$$(X_1 - X_2, X_3 - X_4, X_5 - X_6, ..., X_{n-1} - X_n)$$

is a sample from $S_\alpha(\sigma, 0, 0)$.

(If $X \sim S_\alpha(\sigma, \beta, \mu)$ and $Y \sim S_\alpha(\sigma, \beta, \mu)$ and $X \parallel Y$ then for $\alpha \neq 1$ $X - Y \sim S_\alpha(\sigma, 0, 0)$)

To find α and σ we will use the empirical characteristic function

$$ch_n(t) = \frac{1}{n} \sum_{i=1}^{n} e^{i \cdot t \cdot X_i}$$

The parameters α and σ can be estimated as follows: we choose appropriate points $t_1, t_2, ..., t_m,\ m \in N$ and calculate $|f(t_i; \alpha, \sigma) - ch_n(t_i)|^2$, $i = 1, 2, ..., m$ The parameters will be found by minimizing the function

$$F(\alpha, \sigma) = \sum_{k=1}^{m} |ch_n(t_k) - f(t_k; \alpha, \sigma)|^2$$

When we find the estimations $\hat{\alpha}$ and $\hat{\sigma}$ we can estimate the rest the same way. We will substitute $\hat{\alpha}$ and $\hat{\sigma}$ into the expression of the characteristic function

$$\hat{f}(t, \mu, \beta, \hat{\alpha}, \hat{\sigma}) = \exp\left(-\hat{\sigma}^{\hat{\alpha}} |t|^{\hat{\alpha}} (1 - i\beta(sign(t)) \tan\frac{\pi\hat{\alpha}}{2}) - i\mu t\right)$$

and μ and β will be estimated by minimizing the function

$$\hat{F}(\mu, \beta) = \sum_{k=1}^{m_1} |ch_n(s_k) - \hat{f}(s_k; \mu, \beta, \hat{\alpha}, \hat{\sigma})|^2$$

The found numbers $\hat{\alpha}_n, \hat{\sigma}_n, \hat{\beta}_n, \hat{\mu}_n$ are the consistent estimates of $\theta = (\alpha, \sigma, \beta, \mu)$.

8.2 Testing of the estimators

Notation:

$(\hat{\alpha}_{n,m,z})$ is an CF estimator of α gained by the solution of the problem:

$$\min_{\alpha} \sum_{j=1}^{m} |\exp(-t_j^{\alpha}) - \hat{C}_n(t_j))|$$

where z is the concentrating power.
N=1000, Number of simulations = 1500. Number of summands = 100. The probability that $a_j \sim N[\mu, \sigma]$ is in $[\mu - 3\sigma, \mu + 3\sigma]$ equals $\Phi(3) - \Phi(-3) = 0.9973$, is in $\mu - 1.6\sigma, \mu + 1.6\sigma$: $\Phi(1.6) - \Phi(-1.6) = 0.89$:

α	$\bar{\alpha}$	Var[α]=σ^2	$\pm 3\sigma$	$\pm 1.6\sigma$
1.1	1.1025	0.0017	(0.9757,1.2292)	(1.0348,1.1701)
1.2	1.1989	0.0020	(1.0617,1.3342)	(1.1253,1.2706)
1.3	1.3015	0.0022	(1.1583,1.4446)	(1.2251,1.3778)
1.4	1.4064	0.0025	(1.2540,1.5587)	(1.3251,1.4876)
1.5	1.5004	0.0026	(1.3461,1.6553)	(1.4179,1.5820)
1.6	1.6040	0.0028	(1.4430,1.7649)	(1.5181,1.6898)
1.7	1.7073	0.0028	(1.5432,1.8641)	(1.6181,1.7885)
1.8	1.8053	0.0027	(1.6481,1.9625)	(1.7214,1.8891)
1.9	1.9012	0.0023	(1.7550,2.0473)	(1.8231,1.9791)

Table 8.1: ch.f. estimator of alpha with 1000 observations

The probability that $a_j \sim N[\mu, \sigma]$ is in $[\mu - 1.1\sigma, \mu + 1.1\sigma]$ equals $\Phi(1.1) - \Phi(-1.1) = 0.72$, is in $\mu - 0.8\sigma, \mu + 0.8\sigma$: $\Phi(0.8) - \Phi(-0.8) = 0.57$:

As we can observe, the intervals for the estimation of the adjacent α-s cover if we take 3 σ. The continue covering if we take 1.6 σ but at less rate and they don't cover at all if we take the interval 54% interval: $[\mu - 0.8\sigma, \mu + 0.8\sigma]$ The problem can be solved as follows:

1. Better choice of sample size (Increase the number of observations)
2. Better choice of
3. Better choice of the concentration

α	$\bar{\alpha}$	Var[α]=σ^2	$\pm 1.1\sigma$	$\pm 0.8\sigma$
1.1	1.1025	0.0017	(1.0560,1.1489)	(1.0687,1.1363)
1.2	1.1989	0.0020	(1.1480,1.2479)	(1.1616,1.2343)
1.3	1.3015	0.0022	(1.2490,1.3539)	(1.2633,1.3396)
1.4	1.4064	0.0025	(1.3505,1.4622)	(1.3657,1.4470)
1.5	1.5004	0.0026	(1.4435,1.5564)	(1.4589,1.5410)
1.6	1.6040	0.0028	(1.5449,1.6630)	(1.5610,1.6469)
1.7	1.7073	0.0028	(1.6441,1.7618)	(1.6602,1.7457)
1.8	1.8053	0.0027	(1.7476,1.8629)	(1.7633,1.8472)
1.9	1.9012	0.0023	(1.8476,1.9547)	(1.8622,1.9401)

Table 8.2: ch.f. estimator of alpha with 1000 observations

If we take 30000 observation and number of summands 100 and concentrate the points t_j as follows: $t_j = (\Delta j)^8$, $\Delta = \ln n^{1/8}/m$ then we will have very appropriate estimate of alpha. The following table, where we observe 30 independent estimates $\hat{\alpha}_{30000,100,8}$ of α, shows its properties:

α	$\bar{\alpha}$	Var[α]=σ^2	$\bar{\alpha} \pm 3\sigma$	$\bar{\alpha} \pm 1.66\sigma$
1.1	1.1005	0.000057	(1.07785,1.12315)	(1.0879,1.1130)
1.2	1.1977	0.000053	(1.17586,1.21954)	(1.1856,1.2097)
1.3	1.2974	0.000086	(1.26958,1.32522)	(1.2820,1.3127)
1.4	1.4005	0.000081	(1.37350,1.42750)	(1.3855,1.4154)
1.5	1.4990	0.000106	(1.46811,1.52989)	(1.4819,1.5160)
1.6	1.5994	0.000106	(1.56851,1.63029)	(1.5823,1.6164)
1.7	1.6953	0.000078	(1.66880,1.72180)	(1.6806,1.7099)
1.8	1.8037	0.000077	(1.77737,1.83002)	(1.7891,1.8182)
1.9	1.9012	0.000054	(1.87753,1.92166)	(1.8873,1.9118)

The following tables shows any particular estimate:

1) For $\alpha = 1.1$:
1.1082, 1.1106, 1.09801, 1.09351, 1.10202, 1.10637, 1.10645, 1.10321, 1.10141, 1.09503, 1.08941, 1.1138, 1.08915, 1.09104, 1.08847, 1.08957, 1.10886, 1.10434, 1.10892, 1.1053, 1.09825, 1.09933, 1.09321, 1.1055, 1.10385, 1.10196, 1.08802, 1.11072, 1.10393, 1.09877

2) For $\alpha = 1.2$:
1.20496, 1.20249, 1.19517, 1.19604, 1.20167, 1.19494, 1.18743, 1.20151, 1.19428, 1.19804, 1.18794, 1.20512, 1.196, 1.18698, 1.20481, 1.19169, 1.21027, 1.19788, 1.20406, 1.19547, 1.20514, 1.19805, 1.19988, 1.19651, 1.2037, 1.18148, 1.2028, 1.21141, 1.1872, 1.1893

3) For $\alpha = 1.3$:
1.2935, 1.3151, 1.29559, 1.29847, 1.30572, 1.29071, 1.30592, 1.30491, 1.29851, 1.2842, 1.30382, 1.2773, 1.30333, 1.28713, 1.2845, 1.28981, 1.29872, 1.30761, 1.30348, 1.3049, 1.28578, 1.29711, 1.29223, 1.29298, 1.29124, 1.29064, 1.29361, 1.30026, 1.3111, 1.31368

4) For $\alpha = 1.4$:
1.39429, 1.37332, 1.41154, 1.37893, 1.39565, 1.4033, 1.40493, 1.4009, 1.4029, 1.39903, 1.40091, 1.40502, 1.40194, 1.40726, 1.40317, 1.39165, 1.41048, 1.39977, 1.40614, 1.4095, 1.41253, 1.4081, 1.40045, 1.39079, 1.39246, 1.4052, 1.40406, 1.40874, 1.40441, 1.39019

5) For $\alpha = 1.5$:
1.50065, 1.50034, 1.50638, 1.49664, 1.49629, 1.49527, 1.50285, 1.50114, 1.49815, 1.52136, 1.49376, 1.49189, 1.49457, 1.49533, 1.48902, 1.49971, 1.5083, 1.50469, 1.49852, 1.50014, 1.49522, 1.50339, 1.47766, 1.51644, 1.49559, 1.51264, 1.49487, 1.48305, 1.47853, 1.52043

6) For $\alpha = 1.6$:
1.60539, 1.59677, 1.58844, 1.61145, 1.62125, 1.58101, 1.58802, 1.60896, 1.60007, 1.59911, 1.61678, 1.59863, 1.59, 1.59516, 1.59729, 1.59834, 1.59634, 1.59175, 1.61957, 1.59433, 1.59859, 1.59841, 1.58783, 1.59833, 1.59295, 1.59413, 1.59617, 1.61032, 1.61868, 1.58904

7) For $\alpha = 1.7$:
1.69735, 1.70082, 1.70857, 1.69944, 1.69913, 1.68713, 1.68803, 1.70391, 1.70805, 1.69703, 1.69646, 1.68646, 1.70104, 1.69849, 1.68931, 1.70219, 1.68206, 1.69366, 1.671, 1.71233, 1.70166, 1.68737, 1.70235, 1.68343, 1.70086, 1.69522, 1.69474, 1.69236, 1.6924, 1.68796

8) For $\alpha = 1.8$:
1.79663, 1.81661, 1.80159, 1.8116, 1.79709, 1.78594, 1.80199, 1.78413, 1.80843,

1.81448, 1.82001, 1.79595, 1.81403, 1.80046, 1.79043, 1.79372, 1.80484, 1.80637, 1.79332, 1.81044, 1.79858, 1.80538, 1.80499, 1.80583, 1.80995, 1.80269, 1.80935, 1.81111, 1.80589, 1.80978

9) For $\alpha = 1.9$:
1.90411, 1.89804, 1.89732, 1.90743, 1.91036, 1.89394, 1.89956, 1.89235, 1.89556, 1.89975, 1.90654, 1.9116, 1.90313, 1.88928, 1.89956, 1.88236, 1.89519, 1.89911, 1.89334, 1.9165, 1.89402, 1.90574, 1.8989, 1.90076, 1.89882, 1.89375, 1.90758, 1.90284, 1.90196, 1.88842

So this estimator is very effective despite we applied only 30 independent estimators. No $\mu \pm 3\sigma$ interval of the estimates of α covers any other, moreover they are small enough. We took such relatively small number because it takes a plenty of time to calculate it. Disadvantage of this estimator lies in time which we need to calculate it. If we reduce the number of summands from 100 to 20 then we will have an estimator whose value will be calculated more than 5 times quicker.

Time needed to calculate $(\hat{\alpha}_{30000,100,8})$: 71.132 seconds
Time needed to calculate $(\hat{\alpha}_{30000,20,4})$: 14.371 seconds

The table 9.3 shows the properties of $(\hat{\alpha}_{30000,20,4})$.

α	$\bar{\alpha}$	Var[α]=σ^2	$\bar{\alpha} \pm 3\sigma$
1.1	1.0993	0.000072	(1.07390,1.12486)
1.2	1.1968	0.000094	(1.16774,1.22598)
1.3	1.3009	0.000117	(1.26836,1.33345)
1.4	1.3991	0.000087	(1.37107,1.42718)
1.5	1.5018	0.000069	(1.47693,1.52681)
1.6	1.5972	0.000076	(1.57092,1.62355)
1.7	1.7007	0.000097	(1.67105,1.73039)
1.8	1.7969	0.000100	(1.76692,1.82693)
1.9	1.8995	0.000089	(1.87119,1.92798)

Table 8.3: Table of properties of estimator $\alpha^*_{30000,20,4}$

If we compare this estimate with the previous it has all advantages of

$$(\hat{\alpha}_{30000,100,8})$$

but it takes much less time to calculate. Although the variance is a bit larger on the one hand such difference is not significant on the other hand the mean of all values is better for $(\hat{\alpha}_{30000,20,4})$ if we compare them by means of sum of absolute differences. For $(\hat{\alpha}_{30000,100,8})$ it equals 0.0171 and for $(\hat{\alpha}_{30000,20,4})$ its value is 0.0151. These both estimates despite their quality have one important disadvantage: it is too difficult in practice to possess 30000 observations. Therefore we must explore from what number of observations it begins to work. There are two factors which complicate the computation. They are the number of summands and the number of observations. The power doesn't complicate the calculation, but if we take very large power then all points except for the last point will be concentrated in such vicinity of zero and they will differ so little that the computer will not remark the difference between $ch(t_{i+1}) - ch(t_i)$ i.e. it will be a computer zero and the values in the close vicinity of zero, will be so close to one that it will give no information about α and we will possess only the information in the point $\ln n$ which is far from zero, and contains only little information about the tail index. The power should be larger than 1 to concentrate the points around zero well but it has some border whose crossing means the behavior we've just described. Let us examine the samples with 5000 observations. Such number of observations is usual for practical purposes. We will explore for what power and number of summands it estimates the best.

z=1					
	1.1	1.2	1.3	1.4	1.5
avg	1.09785	1.20023	1.30106	1.40048	1.5027
var	0.00058	0.00057	0.00077	0.00067	0.0009
$\pm 3\sigma$	[1.025,1.170]	[1.128,1.272]	[1.217,1.384]	[1.32,1.47]	[1.40,1.59]
$\pm 2\sigma$	[1.049,1.146]	[1.152,1.248]	[1.245,1.356]	[1.34,1.45]	[1.44,1.56]
1.5σ	[1.061,1.134]	[1.164,1.236]	[1.259,1.342]	[1.36,1.43]	[1.45, 1.54]
	1.6	1.7	1.8	1.9	1.99
avg	1.6018	1.7046	1.7958	1.8947	1.9861
var	0.00138	0.00094	0.00105	0.0011	0.0008
$\pm 3\sigma$	[1.490,1.713]	[1.612,1.796]	[1.698,1.893]	[1.79,1.99]	[1.90,2.07]
$\pm 2\sigma$	[1.527,1.676]	[1.643,1.766]	[1.730,1.860]	[1.82,1.96]	[1.92,2.04]
1.5σ	[1.546,1.657]	[1.658,1.750]	[1.747,1.844]	[1.84,1.94]	[1.94,2.02]

Table 8.4: estimate with power z=1, ($\alpha_{5000,5,1}$)

($\alpha_{5000,5,1.4}$):

z=1.4					
	1.1	1.2	1.3	1.4	1.5
avg	1.10348	1.9891	1.3032	1.4028	1.4993
var	0.00047	0.00049	0.00069	0.00055	0.00066
$\pm 3\sigma$	[1.037,1.169]	[1.131,1.265]	[1.224,1.382]	[1.33,1.47]	[1.42,1.57]
$\pm 2\sigma$	[1.059,1.147]	[1.154,1.243]	[1.250,1.355]	[1.35,1.44]	[1.44,1.55]
1.5σ	[1.070,1.136]	[1.165,1.232]	[1.263,1.342]	[1.36,1.43]	[1.46,1.53]
	1.6	1.7	1.8	1.9	1.99
avg	1.5995	1.7013	1.7999	1.9006	1.9894
var	0.00060	0.00053	0.00056	0.00044	0.00021
$\pm 3\sigma$	[1.525,1.673]	[1.632,1.770]	[1.728,1.871]	[1.84,1.96]	[1.94,2.03]
$\pm 2\sigma$	[1.550,1.648]	[1.655,1.747]	[1.752,1.847]	[1.86,1.94]	[1.96,2.01]
1.5σ	[1.562,1.636]	[1.666,1.735]	[1.764,1.835]	[1.86,1.93]	[1.97,2.01]

$(\alpha_{5000,5,1.5})$:

z=1.5					
	1.1	1.2	1.3	1.4	1.5
avg	1.10209	1.202	1.3024	1.4010	1.4976
var	0.00035	0.00048	0.00049	0.00054	0.00060
$\pm 3\sigma$	[1.045,1.1583]	[1.136,1.267]	[1.235,1.369]	[1.33,1.47]	[1.42,1.57]
$\pm 2\sigma$	[1.064,1.139]	[1.158,1.245]	[1.257,1.347]	[1.35,1.45]	[1.45,1.54]
1.5σ	[1.073,1.1302]	[1.169,1.234]	[1.268,1.335]	[1.36,1.43]	[1.46,1.53]
	1.6	1.7	1.8	1.9	1.99
avg	1.6014	1.7034	1.7988	1.8988	1.9896
var	0.00066	0.00062	0.00055	0.00032	0.00017
$\pm 3\sigma$	[1.525,1.678]	[1.653,1.778]	[1.728,1.869]	[1.84,1.95]	[1.95,2.02]
$\pm 2\sigma$	[1.549,1.653]	[1.653,1.753]	[1.751,1.845]	[1.86,1.93]	[1.96,2.01]
1.5σ	[1.562,1.640]	[1.665,1.740]	[1.763,1.834]	[1.87,1.92]	[1.97,2.00]

$(\alpha_{5000,5,2.2})$:

z=2.2					
	1.1	1.2	1.3	1.4	1.5
avg	1.10154	1.206	1.301	1.4036	1.50262
var	0.00078	0.00097	0.00096	0.00128	0.00152
$\pm 3\sigma$	[1.017,1.185]	[1.112,1.300]	[1.208,1.394]	[1.29,1.51]	[1.38,1.61]
$\pm 2\sigma$	[1.045,1.157]	[1.143,1.268]	[1.239,1.363]	[1.33,1.47]	[1.42,1.58]
1.5σ	[1.059,434]	[1.159,1.253]	[1.254,1.347]	[1.34,1.45]	[1.44,1.56]
	1.6	1.7	1.8	1.9	1.99
avg	1.60753	1.69643	1.80522	1.90187	1.9871
var	0.00232	0.00266	0.00191	0.00134	0.00034
$\pm 3\sigma$	[1.462,1.752]	[1.541,1.851]	[1.673,1.936]	[1.79,2.01]	[1.93,2.04]
$\pm 2\sigma$	[1.511,1.703]	[1.593,1.799]	[1.717,1.892]	[1.82,1.97]	[1.95,2.04]
1.5σ	[1.535,1.679]	[1.619,1.773]	[1.739,1.870]	[1.84,1.95]	[1.95,2.01]

Among these estimates $(\alpha_{5000,5,1.5})$ is the best. To verify this we made 5000 simulations of this estimates and the following table shows properties of this estimator for larger sample.

z=1.5 with 5000 simulations					
	1.1	1.2	1.3	1.4	1.5
avg	1.0997	1.2005	1.3000	1.3997	1.5002
var	0.00040	0.00047	0.00052	0.00055	0.00058
$\pm 3\sigma$	[1.039,1.160]	[1.134,1.266]	[1.231,1.368]	[1.32,1.47]	[1.42,1.57]
$\pm 2\sigma$	[1.059,1.140]	[1.156,1.244]	[1.254,1.345]	[1.35,1.44]	[1.45,1.54]
1.5σ	[1.167,1.233]	[1.169,1.234]	[1.265,1.334]	[1.36,1.43]	[1.46,1.53]
	1.6	1.7	1.8	1.9	1.99
avg	1.6009	1.6999	1.8006	1.9001	1.9902
var	0.00060	0.00056	0.00049	0.00037	0.00017
$\pm 3\sigma$	[1.527,1.674]	[1.628,1.771]	[1.733,1.845]	[1.84,1.95]	[1.95,2.02]
$\pm 2\sigma$	[1.551,1.650]	[1.652,1.747]	[1.755,1.845]	[1.86,1.93]	[1.96,2.01]
1.5σ	[1.564,1.637]	[1.664,1.735]	[1.767,1.834]	[1.87,1.92]	[1.97,2.00]

5000 simulations confirm the quality of this estimator. Therefore if we have 5000 observations of the random variable $X \sim S_\alpha(1,0,0)$ we propose $(\hat{\alpha}_{5000,5,1.5})$ because this estimate doesn't require much time for calculation and is precise enough as it is shown in previous tables. According to Csőrgö theorem they are also consistent therefore, if we possess more than 5000 observations we can also apply this estimate. The following table shows the properties of $(\alpha_{6000,5,1.5})$ for 100 simulations:

z=1.5 with 100 simulations and 6000 observations					
	1.1	1.2	1.3	1.4	1.5
avg	1.1025	1.2048	1.2963	1.4060	1.5054
var	0.00045	0.00044	0.00036	0.00050	0.00052
$\pm 3\sigma$	[1.038,1.166]	[1.141,1.268]	[1.239,1.353]	[1.33,1.47]	[1.43,1.57]
$\pm 2\sigma$	[1.059,1.145]	[1.162,1.247]	[1.258,1.334]	[1.36,1.45]	[1.45,1.54]
1.5σ	[1.070,1.134]	[1.173,1.236]	[1.267,1.324]	[1.37,1.44]	[1.47,1.53]
	1.6	1.7	1.8	1.9	1.99
avg	1.6018	1.7006	1.7985	1.8975	1.9943
var	0.00042	0.00053	0.00044	0.00022	0.00014
$\pm 3\sigma$	[1.539,1.663]	[1.631,1.769]	[1.735,1.861]	[1.85,1.94]	[1.96,2.03]
$\pm 2\sigma$	[1.560,1.643]	[1.654,1.746]	[1.756,1.840]	[1.86,1.92]	[1.97,2.01]
1.5σ	[1.570,1.632]	[1.665,1.735]	[1.766,1.830]	[1.87,1.92]	[1.98,2.01]

Variance in this case is always less than the one in $(\hat{\alpha}_{5000,5,1.5})$ except for $\alpha = 1.1$ but the difference is insignificant and can be explained by a low number of simulations. If we take 15000 observations then the estimator $\hat{\alpha}_{15000,5,1.5}$ can be described by the following table:

z=1.5 with 100 simulations and 15000 observations					
	1.1	1.2	1.3	1.4	1.5
avg	1.1018	1.1978	1.2989	1.4018	1.4983
var	0.00013	0.00014	0.00013	0.00016	0.00013
$\pm 3\sigma$	[1.066,1.136]	[1.161,1.233]	[1.263,1.334]	[1.36,1.43]	[1.46,1.53]
$\pm 2\sigma$	[1.078,1.125]	[1.173,1.221]	[1.275,1.322]	[1.37,1.42]	[1.47,1.52]
1.5σ	[1.084,1.119]	[1.179,1.215]	[1.281,1.316]	[1.38,1.42]	[1.48,1.51]
	1.6	1.7	1.8	1.9	1.99
avg	1.5990	1.6992	1.7986	1.8997	1.9887
var	0.00016	0.00019	0.00016	0.00010	0.00006
$\pm 3\sigma$	[1.560,1.637]	[1.657,1.740]	[1.760,1.836]	[1.87,1.93]	[1.96,2.01]
$\pm 2\sigma$	[1.573,1.624]	[1.671,1.727]	[1.773,1.823]	[1.88,1.92]	[1.97,2.00]
1.5σ	[1.579,1.618]	[1.678,1.720]	[1.779,1.817]	[1.88,1.91]	[1.98,2.01]

This estimator is qualitative, the interval 3σ is small enough to say that the vicinity of the precise value is very small for this estimation. Two previous tables empirically confirm the the theoretical consistency of this estimator following from Csőrgö theorem. Based on the previous results we conclude:

If we have the observations $X_1, X_2, ..., X_N$ from $S_\alpha(1, 0, 0)$ and $N \geq 5000$ then we propose to apply $(\bar{\alpha}_{N,5,1.5})$ with $t_j = (j \cdot \Delta)^{1.5}$, Tn=log$(N)$, $\Delta = \frac{\sqrt{Tn}}{10 \cdot m}$, m=5.

If we take N=100 then the average of large samples of the estimates of α based on characteristic function and minimizing the sum is close to the real value of the parameter, but the variance is so large that $\bar{\alpha} \pm 3\sigma$ interval covers for any $\alpha \in [1, 2]$ the whole interval [1,2] therefore, we can't rely on any estimator of α if we possess such a small number of observations. If we have N=4000 then the estimator $(\alpha_{4000,5,1.4})$ then the interval $\pm 1.5\sigma$ is thin yet and this estimator works with 0.86% probability.

When we have 1000 observations we can look for the best estimate as follows. Having chosen number of summands and the points t_j we will look for such power, that minimizes the corresponding intervals. The width of the intervals depends on the variance. Therefore having estimated $\bar{\bar{\alpha}}$ and $var(\hat{\alpha})$ of $\alpha = 1.1, 1.2, 1.3, ..., 1.8, 1.9$ and 1.99 we will calculate the sum of variances for any estimate (or the average) and choose such z that minimizes the sum. We can do this by two methods:
1) we will calculate that sum for a large number of z and choose such z for which that sum is minimal.
2) we will conduct regression analysis of the dependence of the sum on the power and will choose the appropriate minimum of the corresponding regression function.

8.3 First Method of Choosing the Power

We simulate 100 estimators ($\alpha_{1000,15,z}$) for 18 values of z. The results described in the following table:

z	Sum
0.500	0.296817
1.000	0.037670
1.300	0.023993
1.360	0.023610
1.380	0.022058
1.390	0.226670
1.395	0.023160
1.400	0.022040
1.410	0.021772
1.415	0.022900
1.420	0.025020
1.440	0.022511
1.500	0.025991
1.600	0.026501
2.200	0.688980
3.000	0.161500
6.000	1022.950

The minimum sum is attained for $z = 1.41$ therefore let us simulate $\alpha_{1000,15,1.41}$ 1500 times to have reliable intervals for $\bar{\alpha}$. The following table shows the results and compares $\alpha_{1000,100,8}$ with $\alpha_{1000,15,1.41}$

α	$\bar{\alpha}$	Var[α]=σ^2	$\pm 3\sigma$,($\alpha_{1000,100,8}$)	$\pm 3\sigma$,($\alpha_{1000,15,1.41}$)
1.1	1.10258	0.0021	(0.9757,1.2292)	(0.9650,1.2401)
1.2	1.20261	0.0022	(1.0617,1.3342)	(1.0608,1.3444)
1.3	1.3026	0.0024	(1.1583,1.4446)	(1.1538,1.4514)
1.4	1.4023	0.0026	(1.2540,1.5587)	(1.2985,1.5060)
1.5	1.5031	0.0028	(1.3461,1.6553)	(1.3430,1.6632)
1.6	1.6030	0.0029	(1.4430,1.7649)	(1.4403,1.7657)
1.7	1.7012	0.0027	(1.5432,1.8641)	(1.5442,1.8583)
1.8	1.8012	0.0024	(1.6481,1.9625)	(1.6532,1.9491)
1.9	1.9022	0.0016	(1.7550,2.0473)	(1.7794,2.0250)

The intervals from $\alpha = 1.1$ to $\alpha = 1.6$ are thiner for $\alpha_{1000,100,8}$ but beginning with $\alpha = 1.7$ the estimator $\alpha_{1000,15,1.41}$ has thinner $\bar{\alpha} \pm 3\sigma$ intervals. But the difference of the width of both estimators is insignificant but $\alpha_{1000,15,1.41}$ can be calculated more than 5 times quicker. Therefore we can recommend this estimate but it will estimate α with the precision 0.01 with probability 0.72 because it corresponds to the interval $\pm 1.1\sigma$.

8.4 Second Method of Choosing the Power

We conduct regression analysis based on the value from the table of z and corresponding sums. From the observations we see that the sum increases from 1.41 to 2.2 and decreases from 0.5 to 1.41. Therefore the dependence cannot be linear. Suppose that the dependence is polynomial. There is one outlier. For z=6 the sum equals 1022.9 therefore we will analyze both models where we exclude this observation and when we take it into the model.
I. The regression model when we exclude the outlier:
First model: $Sum \sim \{1, z, z^2\}$:

Parameter Table	Estimate	SE	TStat	R-Squared	p-Value
1	0.481415	0.0478379	10.1278	0.88	0
x	-0.523574	0.0543815	-9.6278		0
x^2	0.141504	0.0144629	9.78389		0

The R-Squared can be increased by changing the power of the polynomial. If we take the regression dependence $Sum \sim \{1, z, |z|^{1.1}\}$ then we will have the following table:

Parameter Table	Estimate	SE	TStat	R-Squared	p-Value
1	0.71098	0.0491	14.471	0.93	0
x	-4.4568	0.321	-13.867		0
$\|x\|^{1.2}$	3.834	0.276	13.848		0

I chose such power because it maximizes the R-squared among other choices. From here we get the following dependence:

$$Sum = 0.710 - 4.456z + 3.834|z|^{1.1} + \epsilon$$

This function attains the minimum value in 1.73. If we include the outlier into the model polynomial regression we will have the determination coefficient equal to 0.99 and the function:

$$Sum = 0.710 - 4.456z + 3.834|z|^{1.1}$$

It attains minimum at 1.73. If we apply the quadratic polynomial we will have the following results:

$$R - Squared = 0.98, \;\; Sum = 145.882 - 190.219z + 55.741z^2$$

whose minimum is attained for z=1.706.

As it is clear from the table, this estimator is worse than $(\alpha_{1000,15,1.41})$ therefore when we possess 1000 simulations we propose to use this estimate. It will work for any N larger than 1000 because of consistency of any estimator of this kind, but if $N \geq 5000$ than the estimator $(\alpha_{N,5,1.5})$ can be calculated faster with suitable precision, therefore based on just conducted research we propose the following estimators for α:

1. If $N \geq 5000$ we propose $(\alpha_{N,5,1.5})$
2. If $5000 \geq N \geq 1000$ we propose $(\alpha_{N,15,1.41})$
3. If $N < 1000$ but it is close to this value then we can apply $(\alpha_{N,15,1.41})$ or $(\alpha_{N,m,1.41})$ where $m \geq 15$. If N is small in comparison to 1000 than no CF estimator is reliable.

α	$\bar{\alpha}$	Var[α]=σ^2	$\pm 3\sigma$,($\alpha_{1000,100,8}$)	$\pm 3\sigma$,($\alpha_{1000,15,1.73}$)
1.1	1.10016	0.0020	(0.9757,1.2292)	(0.9637,1.2365)
1.2	1.20355	0.0027	(1.0617,1.3342)	(1.0462,1.3608)
1.3	1.30554	0.0029	(1.1583,1.4446)	(1.1412,1.4698)
1.4	1.4061	0.0033	(1.2540,1.5587)	(1.2314,1.5807)
1.5	1.5055	0.0039	(1.3461,1.6553)	(1.3178,1.69322)
1.6	1.6009	0.0040	(1.4430,1.7649)	(1.4094,1.7925)
1.7	1.6994	0.0103	(1.5432,1.8641)	(1.3938,2.0051)
1.8	1.8014	0.0032	(1.6481,1.9625)	(1.6295,1.9734)
1.9	1.9039	0.0024	(1.7550,2.0473)	(1.7554,2.0524)

Figure 8.1: Table of CF estimates

8.5 Checking the Normality of the Estimators

1. First we will check the normality of ($\alpha_{5000,5,1.5}$).
When $\alpha = 1.1$ we simulate 5000 estimates of this parameter. The average of the sample equals 1.09977, the variance equals 0.00040. We can compare the the cumulative distribution function of $N(1.09977, \sqrt{0.0040})$ ($N(\mu, \sigma)$) with the empirical distribution function based on 5000 thousand observations:8.2

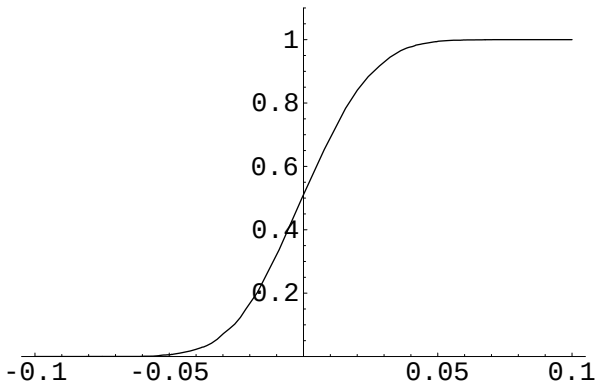


Figure 8.2: Graph of the empirical distribution function

And let us take a look at the comparison of the graph of CDF of the corresponding normal distribution with the empirical distribution function, see the figure 10.3:

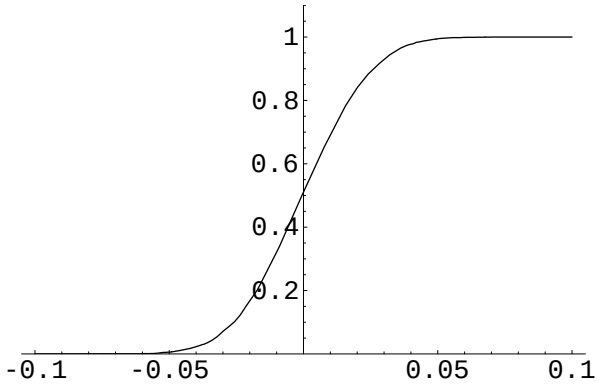


Figure 8.3: Graph of the CDF of $N(1.09977, \sqrt{0.0040})$ and the empirical distribution function, they are so similar that their graphs intersect and the difference is invisible

Kolmogorov Smirnov test didn't refuse the hypothesis of normality for $(1-\alpha) = 98\%$. The empirical distribution function and the distribution function of the normal distribution with mean equal to the average of the sample and the variance equal to the empirical variance of the sample cover each other and K-S test doesn't refuse normality for large enough values of $(1-\alpha)$. The same situation holds if we have 1000 observations. The more simulations the more evident normality from the graph. There isn't explicit expression of any $(\alpha_{N,m,z})$ therefore we have to rely only on empirical data. In Appendix D there is a graph showing how $\bar{\hat{\alpha}}_{5000,5,1.5}$ (i.e. mean of the estimates of α of $X \sim S_{1.3}(1,0,0)$) converges to 1.3.

8.6 Comparison of the variance of $\alpha_{30000,5,1.5}$ with Fisher information with 50 independent estimates

In the following table we compare the variance of the estimator $\alpha_{30000,5,1.5}$ with the Fisher information calculated by the following formula:

$$I(\alpha) = [4(2-\alpha)|\ln(2-\alpha)|]^{-1}$$

when α is close to 2. This theorem and its proof can be found in "CHANCE and STABILITY Stable Distributions and their Applications" by Vladimir

V. Uchaikin and Vladimir M. Zolotarev, VSP 1999, page 249.(book: [10])

α	$\mathrm{Var}(\alpha_{30000,5,1.5})$	$1/(30000 \cdot I(\alpha))$	ratio %
1.7	0.000084365	0.0000481589	57 %
1.8	0.000074794	0.0000429183	57 %
1.9	0.000068985	0.0000307011	44 %
1.93	0.000058339	0.0000248198	42 %
1.96	0.000043466	0.0000171673	39 %
1.99	0.000034410	$6.14023 \cdot 10^{-6}$	17 %
1.9996	0.000037898	$4.17282 \cdot 10^{-7}$	1.1 %

In the last two chapters we propose the method based on projections invented by Kagan and developed for the stable distributions by Klebanov and Melamed. Here are results obtained by method based on projections:

α	$\mathrm{Var}(\alpha_{30000,5,1.5})$	$1/(30000 \cdot I(\alpha))$	ratio %
1.7	0.000084365	0.000067	79 %
1.8	0.000074794	0.000056	75 %
1.9	0.000068985	0.000038	55 %
1.93	0.000058339	0.000030	52 %
1.96	0.000043466	0.000020	46 %
1.99	0.000034410	$7.4791 \cdot 10^{-6}$	21 %
1.9996	0.000037898	$5.8594 \cdot 10^{-7}$	1.5 %

From both tables we can see that the function obtained by Zolotarev and Uchaikin converges to $I(\alpha)$ for the values which are "too" close to 2 and their deviation from 2 is insignificant. Moreover the values of α that are so close to 2 have no place in financial practice. If we approximate the value of the Fisher information by this formula then it will always be exaggerated.

8.7 Testing of the estimators of unknown (α,σ) with zero μ and β

If we apply the estimator $(\hat{\alpha}, \hat{\sigma})_{30000,20,8}$ of the parameters from the distribution $S_\alpha(2, 0, 0)$ then we can observe its properties from the following table:

σ	α	$\hat{\alpha}$	$\mathrm{Var}[\hat{\alpha}]=\sigma^2$	$\hat{\sigma}$	$\mathrm{Var}[\hat{\sigma}]$
	1.1	1.09932	0.000063	2.12512	0.0002288
	1.2	1.20263	0.000097	2.12158	0.0003309
	1.3	1.30006	0.000114	2.11718	0.0002135
	1.4	1.40343	0.000193	2.12500	0.0001914
	1.5	1.50146	0.000061	2.12082	0.0002073
	1.6	1.60420	0.001510	2.12370	0.0002163
	1.7	1.70547	0.000817	2.12042	0.0002876
	1.8	1.80084	0.000106	2.12632	0.0001501
	1.9	1.90067	0.000049	2.12014	0.0001103

8.8 general case where all parameters are unknown

As it was mentioned when we have a sample $X_1, X_2, ..., X_N \sim S_\alpha(\sigma, \beta, \mu)$ we have to make differences $X_1 - X_2, X_2 - X_4, ..., X_{N-1}, X_N \sim S_\alpha(\sigma, 0, 0)$ and estimate the tail and scale parameter. Then having substituted the computed values into the corresponding functions, we estimate β and μ. We tested this method by virtue of estimating parameters of the sample from $S_{1.4}(1, 0.23, 0.1)$ with 30000 thousand numbers, applying 2000 simulations. We simulated the samples applying the formula:

$$Y_A(\alpha, \beta) = \frac{2}{\pi}\left[\left(\frac{\pi}{2}\beta\phi\right)\tan\phi - \beta\ln\left(\frac{(\pi/2)E\cos\phi}{\pi/2+\beta\phi}\right)\right], \ \ \alpha = 1$$

and

$$Y_A(\alpha, \beta) = [1+\beta^2\tan^2(\alpha\pi/2)]^{1/(2\alpha)}\frac{\sin(\alpha(\phi+b))}{(\cos\phi)^{1/\alpha}}\left[\frac{\cos(\phi-\alpha(\phi+b))}{E}\right]^{(1-\alpha)/\alpha}$$

if $\alpha \neq 1$, $(*)$ where $\phi \sim U[-\pi/2, \pi/2]$ and $E \sim \exp(1)$. After estimation of the scale parametr from the sample and making elementary algebraic transformations we have $S_\alpha(\sigma, \beta, \mu)$ simulated. Having simulated $S_\alpha(\sigma, \beta, \mu)$ we can commence estimating of their parameters. We tested this method on $S_{1.44}(1, 0.23, 0)$. The results are following:

α	$\bar{\alpha}$	SD of α	σ	$\bar{\sigma}$	SD of σ	β	$\bar{\beta}$	SD of β
1.44	1.4405	0.011	1.00	1.004	0.007	0.23	0.25	0.01

Where SD means standard deviation. When estimating the parameters there were used 2000 simulations.

8.9 Test of Stability

Let us have a sample $X_1, X_2, ..., X_N$. If we want to check stability of the given sample we have to do this in two steps:

1. First we estimate parameters of the model by virtue of the methods represented in previous chapters. Suppose that the corresponding values of parameters are equal to ($\hat{\alpha}$,$\hat{\sigma}$,$\hat{\beta}$,$\hat{\mu}$)

2. Having estimated the parameters of the sample under hypothesis that it is stably distributed we simulate lots of Samples from $S_{\hat{\alpha}}(\hat{\sigma}, \hat{\beta}, \hat{\mu})$ and check the equality in distribution by virtue of Kolmogorov Smirnov test.

If the estimate of α based on characteristic function is larger than 2 for a large ($\geq$ 5000) sample of observation then this is also a reason to reject stability. Therefore we recommend to go to the second step only if the estimate $\hat{\theta} = (\hat{\alpha}, \hat{\sigma}, \hat{\beta}, \hat{\mu})$ belongs to the set $(0, 2) \times (0, \infty) \times [-1, 1] \times R$

8.10 Comparison of the CF estimates with Pickand estimate

α	Type of estimator	Mean	Variance	$\pm 3\sigma$
1.1	$\alpha_{3000,5,1.5}$	1.1001	0.00007	[1.0747,1.1255]
	Pickand	1.1034	0.00133	[0.9938,1.2130]
1.2	$\alpha_{3000,5,1.5}$	1.2002	0.00008	[1.1724,1.2279]
	Pickand	1.2275	0.00103	[1.1311,1.3239]
1.3	$\alpha_{3000,5,1.5}$	1.3000	0.00007	[1.2732,1.3267]
	Pickand	1.3652	0.00077	[1.2817,1.4487]
1.4	$\alpha_{3000,5,1.5}$	1.4003	0.00008	[1.3723,1.4283]
	Pickand	1.4904	0.00056	[1.4191,1.5618]
1.5	$\alpha_{3000,5,1.5}$	1.5004	0.00009	[1.4713,1.5295]
	Pickand	1.6135	0.00087	[1.5250,1.7021]
1.6	$\alpha_{3000,5,1.5}$	1.5995	0.00008	[1.5713,1.6277]
	Pickand	1.7270	0.00102	[1.6308,1.8232]
1.7	$\alpha_{3000,5,1.5}$	1.7004	0.00008	[1.6728,1.7279]
	Pickand	1.8342	0.00088	[1.7451,1.9233]
1.8	$\alpha_{3000,5,1.5}$	1.7997	0.00007	[1.7738,1.8257]
	Pickand	1.8542	0.00095	[1.7616,1.9467]
1.9	$\alpha_{3000,5,1.5}$	1.9006	0.00006	[1.8759,1.9254]
	Pickand	1.9225	0.00123	[1.8169,2.0282]

Table 8.5: Comparison of Unconditional Pickand Estimator with $\alpha_{30000,5,1.5}$

When we apply unconditional Pickand estimator we face the problem of finding the best k. In Appendix A this estimator is described. Based on Monte Carlo simulations I've concluded that the quality of the estimates based on the choice of k depends on α. But we should eliminate this de-

pendence because α is unknown. To do this we simulated the large number of samples with 30000 elements with $\alpha = 1.1, 1.2, 1.3, ..., 1.9$ for any α and any sample we found the best k, and from the samples of the best k I took the average for any α. Let denote k_1 the average best k for $\alpha = 1.1$, k_2 the average best k for $\alpha = 1.2$, etc. k_9 the best k for $\alpha = 1.9$. When applying k_j for the estimation of $\alpha = 1 + \frac{j}{10}$ we mentioned that the variance of this estimate is relatively low therefore we used following method. Having the sample $X_1, X_2, .., X_30000$ we substitute k_1 into the estimate at first. If the calculated Pickand estimate is larger than 1.2 then we take the following k (k_2), calculate the Pickand estimate with k=k_2 and if the value of the Pickand estimate is larger then 1.3 we take the following k and continue this operation until it is larger than 2 and in this case we take the Pickand estimate to be equal to 2. As only one of the previous conditions doesn't hold we take the calculated value of the Pickand estimate as an estimate of α. The results of this approach are summarized in the previous table.

The reason why we don't propose to use average as an estimate of μ in general

If we have a sample $X_1, X_2, .., X_N$ and X i.i.d. from $S_\alpha(\sigma, \beta, \mu)$ and $\alpha > 1$ as it is always in financial practice, then $\bar{X}$ is a consistent estimator of μ. It follows from elementary properties of stable law:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_N}{N} =_d \frac{N^{1/\alpha} X}{N} =$$

$$= \frac{1}{N^{1-1/\alpha}} X \sim S_\alpha \left(\frac{1}{N^{1-1/\alpha}} \sigma, \beta, \frac{1}{N^{1-1/\alpha}} \mu \right)$$

The consistency follows from the relations:

$$\bar{X} =_d \frac{1}{N^{1-1/\alpha}} X, \quad \lim_{N \longrightarrow \infty} \frac{1}{N^{1-1/\alpha}} X = 0.$$

But $\bar{X}$ is a stably distributed i.e. has an infinite variance therefore it is not reasonable to apply such estimator but there is one exception: it is clear that the quality of the estimate depends on α. To empirically test the quality of this estimator we simulated samples with 30000 elements from $S_{1.1}(1, 0, 0)$, $S_{1.5}(1, 0, 0)$ and $S_{1.9}(1, 0, 0)$. The desirable value of $\bar{X}$ is zero. We made 5000

α	average	variance
1.1	0.307396	161.495
1.5	0.001902	77.5086
1.9	0.000335	0.00015

Table 8.6: Table of estimates of μ for $\alpha = 1.1, 1.5, 1.9$

simulations. The table 9.6 shows the results of computation: From the table we can see that when α is small then this estimator doesn't work even for samples with 30000 elements. But when it is close to 2 then it works well. The largest value in the sample of 5000 estimates of μ when $\alpha = 1.9$ is equal to 0.2175 and the smallest value equals -0.11. Therefore we propose to use this estimator only when we know α and when it is larger than 1.9. In any other case we propose to estimate the location parameter by virtue of other methods e.g. by CF method.

8.11 Parameter Estimation of Geometric Stable Distribution

The same technique that we applied for estimation of the parameters of stable distributions we can apply for the estimation of the parameter of geometric stable laws. If we know how to simulate the stable law $X \sim S_\alpha(\sigma, \beta, \mu)$ then the geometric stable law can be simulated by means of the following formula:

$$Y =_d \mu Z + Z^{1/\alpha} X, \quad Y \sim GS_\alpha(\sigma, \beta, \mu), \quad \alpha \neq 1$$

The characteristic function of the geometric stable law equals:

$$\psi(t) = (1 + \sigma^\alpha |t|^\alpha \exp(-i\beta(\pi/2)K(\alpha))\mathrm{sgn}(t) - i\mu t)^{-1},$$

$$K(\alpha) = \alpha - 1 + \mathrm{sgn}(1 - \alpha)$$

We consider only the cases when $\alpha > 1$ therefore $K(\alpha) = \alpha - 2$. If $\beta = 0$ then

$$\psi(t) = \frac{1}{1 + \sigma^\alpha |t|^\alpha}$$

This gives us possibility to estimate parameters by virtue of the same technique as for the stable laws. The following table shows properties of $\alpha^*_{15000,5,1.5}$:

α	$\bar{\alpha}$	$\mathrm{Var}[\alpha]=\sigma^2$	$\bar{\alpha} \pm 3\sigma$
1.1	1.0983	0.000185	(1.05751,1.13919)
1.2	1.2003	0.000213	(1.15661,1.24418)
1.3	1.3019	0.000209	(1.25849,1.34543)
1.4	1.3979	0.000267	(1.34896,1.44703)
1.5	1.5007	0.000289	(1.44967,1.55175)
1.6	1.5994	0.000264	(1.55064,1.64826)
1.7	1.6989	0.000228	(1.65358,1.74433)
1.8	1.8001	0.000238	(1.75384,1.84648)
1.9	1.9000	0.000171	(1.86074,1.93926)

Table 8.7: Table of properties of estimator $\alpha^*_{15000,5,1.5}$ of α of $GS_\alpha(1,0,0)$

If $X, Y \sim GS_\alpha(\sigma,\beta,\mu)$ and i.i.d. then the characteristic function of $X-Y$ equals:

$$\psi(t) = E\exp(it(X-Y)) = E\exp(itX)E\exp(-itY) = \frac{1}{1+|t|^\alpha}\frac{1}{1+|-t|^\alpha}$$

$$= \left(\frac{1}{1+|t|^\alpha}\right)^2$$

But this is not a geometric stable law. From this follows that geometric stable laws aren't preserved under convolution. Therefore we can't apply the method of differences, that we applied for stable law. Therefore when we don't know all four parameters we have to estimate them by solving the minimum problem of the corresponding function of four parameters α, σ, β, and μ. The technique of the estimation is the same as one of the stable distributions. It is enough to replace the characteristic function of $S_\alpha(\sigma,\beta,\mu)$ by the one of $GS_\alpha(\sigma,\beta,\mu)$ and apply to it the same methodology, taking into account that the geometric stable distribution isn't preserved under convolution including difference.

Chapter 9

Fisher Information and MLP Estimators of α of $S_\alpha S$

9.1 Projections

As it was mentioned in previous chapters the density function of any general stable distribution doesn't have any explicit form, therefore the calculation of Fisher information and ML-estimator are very problematic and we can't conduct it without approximations. The idea of the method described below was given by Kagan (See [5]) and the method itself was developed by Klebanov and Melamed. There are some important identities of the ML-estimates:

$$J(x,\alpha) = \frac{\left(\frac{\partial p(x,\alpha)}{\partial \alpha}\right)}{p(x,\alpha)}$$

$$I(\alpha) = \int_{-\infty}^{\infty} J^2(x,\alpha) p(x,\alpha) dx$$

$$\hat{\alpha}_{ML} = \left\{ \alpha : \sum_{j=0}^{n} J(x,\alpha) = 0 \right\}$$

where $p(x,\alpha)$ is the density function. The method that was just mentioned enables to calculate approximately the function $J(x,\alpha)$ by projection of it to the space $\{1, \exp(it_1 x), \exp(it_2 x), \exp(it_3 x), ..., \exp(it_k x)\}$, that is

$$J_k(x,\alpha) = \sum_{j=0}^{k} a_j \exp(it_j x) = \sum_{j=0}^{k} a_j \cos(t_j x) + i \sum_{j=0}^{k} a_j \sin(t_j x)$$

Where $a_j, j = 1, 2, .., k$ are unknown coefficients that will be calculated from the linear equations and $t_j, j = 0, 1, 2, .., k$ are chosen known points. We project to the space with scalar multiplication given as follows:

$$X \sim S_\alpha(1, 0, 0)$$

$$\langle \exp(it_m X), \exp(it_n X) \rangle = E \exp(it_m X) \cdot \exp(it_n X) = E \exp(iX(t_m + t_n)) =$$
$$= \int_{-\infty}^{\infty} p(x, \alpha) \exp(ix(t_m + t_n))dx = \exp\left(-|t_m + t_n|^\alpha\right)$$

For any projection $J_k(x, \alpha)$ holds:

$$(J_k(x, \alpha) - J(x, \alpha)) \perp \exp(it_j x), j = 1, 2, .., k$$

in other words:

$$\langle (J_k(X, \alpha) - J(X, \alpha)), \exp(it_j X) \rangle = 0, j = 1, 2, .., k$$

or

$$E((J_k(X, \alpha) - J(X, \alpha)) \cdot \exp(it_j X)) = 0, j = 1, 2, .., k$$

From this follows:

$$\int_{-\infty}^{\infty} (J_k(x, \alpha) - J(x, \alpha)) p(x, \alpha) \exp(it_j x)dx = 0$$

Hence

$$\int_{-\infty}^{\infty} J_k(x, \alpha) p(x, \alpha) \exp(it_j x)dx = \int_{-\infty}^{\infty} J(x, \alpha) p(x, \alpha) \exp(it_j x)dx$$

Let us calculate both integrals from the above equality separately:

$$\int_{-\infty}^{\infty} J_k(x, \alpha) p(x, \alpha) \exp(it_j x)dx = \int_{-\infty}^{\infty} \sum_{v=0}^{k} a_v \exp(it_v x) p(x, \alpha) \exp(it_j x)dx =$$

$$= \sum_{v=0}^{k} a_v \int_{-\infty}^{\infty} \exp(it_v x) p(x, \alpha) \exp(it_j x)dx = \sum_{v=0}^{k} a_v \int_{-\infty}^{\infty} p(x, \alpha) \exp(ix(t_j + t_v))dx =$$

$$= \sum_{v=0}^{k} a_v \exp\left(-|t_v + t_j|^\alpha\right), j = 1, 2, .., k$$

We could change sum and integral because the sum is finite. The second integral will be calculated as follows:

$$\int_{-\infty}^{\infty} J(x,\alpha)p(x,\alpha)\exp(it_jx)dx = \int_{-\infty}^{\infty} \frac{\left(\frac{\partial p(x,\alpha)}{\partial \alpha}\right)}{p(x,\alpha)} p(x,\alpha)\exp(it_jx)dx$$

$$= \int_{-\infty}^{\infty} \frac{\partial p(x,\alpha)}{\partial \alpha}\exp(it_jx)dx = \frac{\partial}{\partial \alpha}\int_{-\infty}^{\infty} p(x,\alpha)\exp(it_jx) = \frac{\partial}{\partial \alpha}\exp\left(-|t_j|\right)^{\alpha}$$

The integral and derivative could be replaced because $p(x,\alpha) \geq 0$. Hence follows the system of linear equations, where a_v, $v = 0, 1, .., k$ are unknown:

$$\sum_{v=0}^{k} a_v \exp\left(-|t_v + t_j|\right)^{\alpha} = \frac{\partial}{\partial \alpha}\exp\left(-|t_j|\right)^{\alpha}, j = 1, 2, .., k$$

that can be written in the form:

$$\frac{\partial}{\partial \alpha}\exp\left(-|t_j|\right)^{\alpha} = -\exp\left(-|t_j|\right)^{\alpha} \cdot |t_j|^{\alpha} \cdot \ln|t_j|$$

$$\sum_{v=0}^{k} a_v \exp\left(-|t_v + t_j|^{\alpha}\right) = -\exp\left(-|t_j|\right)^{\alpha} \cdot |t_j|^{\alpha} \cdot \ln|t_j|, j = 1, 2, .., k$$

In matrix form the system looks as folows:

$$\begin{pmatrix} 1 & e^{-|t_1|^{\alpha}} & ... & e^{-|t_k|^{\alpha}} \\ e^{-|t_1|^{\alpha}} & e^{-|t_1+t_1|^{\alpha}} & ... & e^{-|t_1+t_k|^{\alpha}} \\ ... & ... & & ... \\ ... & ... & & ... \\ ... & ... & & ... \\ e^{-|t_k|^{\alpha}} & e^{-|t_k+t_1|^{\alpha}} & ... & e^{-|t_k+t_k|^{\alpha}} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ .. \\ .. \\ a_k \end{pmatrix} = \begin{pmatrix} 0 \\ -|t_1|^{\alpha}\ln|t_1|e^{-|t_1|^{\alpha}} \\ -|t_2|^{\alpha}\ln|t_2|e^{-|t_2|^{\alpha}} \\ .. \\ .. \\ -|t_k|^{\alpha}\ln|t_k|e^{-|t_k|^{\alpha}} \end{pmatrix}$$

Kagan was making projections to the space $\{1, x, x^2, x^3, x^4, ...\}$ but in case of stable distributions such method can't be applied because stable distributions have finite moments beginning with 2. Therefore, we have chosen trigonometric base.

Notation

For the estimates based on Maximum likelihood estimator we will need the following notation:

$$A(\alpha) = \begin{pmatrix} 1 & e^{-|t_1|^\alpha} & ... & e^{-|t_k|^\alpha} \\ e^{-|t_1|^\alpha} & e^{-|t_1+t_1|^\alpha} & ... & e^{-|t_1+t_k|^\alpha} \\ ... & ... & & ... \\ ... & ... & & ... \\ ... & ... & & ... \\ e^{-|t_k|^\alpha} & e^{-|t_k+t_1|^\alpha} & ... & e^{-|t_k+t_k|^\alpha} \end{pmatrix}$$

$$b(\alpha) = \begin{pmatrix} 0 \\ -|t_1|^\alpha \ln|t_1| e^{-|t_1|^\alpha} \\ -|t_2|^\alpha \ln|t_2| e^{-|t_2|^\alpha} \\ .. \\ .. \\ -|t_k|^\alpha \ln|t_k| e^{-|t_k|^\alpha} \end{pmatrix}$$

The coefficients a_j and we will have the following representation of J_k:

$$t = (0, t_1, t_2, t_3, ..., t_k)^T$$

$$F(t,x) = \begin{pmatrix} 1 \\ 2\cos(t_1 x) \\ 2\cos(t_2 x) \\ 2\cos(t_3 x) \\ .. \\ .. \\ 2\cos(t_k x) \end{pmatrix}$$

9.2 The MLP estimates

Let us denote the estimates based on projections and method of maximum likelihood as MLP estimates.

$$J_k(x,\alpha) = \sum_{j=1}^{k} \hat{a}_j \exp(i t_j x)$$

For J_k holds the following limit equality:

$$\lim_{k\longrightarrow\infty} J_k(x,\alpha) = J(x,\alpha)$$

The proof of this assertion can be found in the mentioned publication of Kagan for general case. From that equality follows:

$$\lim_{k\longrightarrow\infty} \Re(J_k(x,\alpha)) = J(x,\alpha)$$

and

$$\lim_{k\longrightarrow\infty} \Im(J_k(x,\alpha)) = 0$$

Let us denote the approximation of the Fisher information received by the projections as $I_k(\alpha)$ and the estimate of α received the same way as $\hat{\alpha}_k$. Then

$$I_k(\alpha) = EJ_k^2(x,\alpha)$$

and

$$\hat{\alpha}_k = \left\{ \alpha : \sum_{j=1}^{n} J_k(X_j,\alpha) = 0 \right\}$$

where $X_j, j = 1,2,..,n$ is a random choice from $S_\alpha(1,0,0)$. The calculation of Fisher information is easier because in this case α is known but when we estimate the parameter α then any quotient $\hat{a}_j$ of J_k is a function of α i.e. $\hat{a}_j = \hat{a}_j(\alpha)$. Hence

$$J_k(x,\alpha) = \sum_{j=1}^{k} \hat{a}_j(\alpha)\exp(-it_jx) = \left((A(\alpha))^{-1}\cdot b(\alpha)\right)^T \cdot F(t,x)$$

Hence

$$J_{k,n}(\alpha) = \sum_{j=1}^{n} J_k(X_j,\alpha) = \sum_{j=1}^{n} \left((A(\alpha))^{-1}\cdot b(\alpha)\right)^T \cdot F(t,X_j) =$$

$$= \left((A(\alpha))^{-1}\cdot b(\alpha)\right)^T \sum_{j=1}^{n} F(t,X_j)$$

This function is computationally complicated, particularly when k=500 or more. Therefore the calculation of the root or Minimum of $J_{k,n}(\alpha)$ takes up a long time, therefore it is reasonable to calculate $J_{k,n}(\alpha_j)$ for N points

α_j where $\alpha_j = 1 + j/N$, $j = 1, 2, .., N$. (We don't consider the case when $\alpha \leq 1$). From these α_j we choose the one for which $|J_{k,n}(\alpha_j)|$ has a minimum value. Let us denote such estimator of alpha as $\alpha_{k,N}$. It is evident that $\lim_{N\to\infty} \alpha_{k,N} = \alpha_k$. Otherwise when k is comparably small we can directly calculate the root of $J_{k,n}(\alpha)$ and it will be α_k. From the fact that

$$\lim_{k\to\infty} J_k = J$$

follows that

$$\lim_{k\to\infty} \alpha_k = \alpha_{ML}$$

From mathematical statistics it is known that:

$$\sqrt{n}(\alpha_{ML} - \alpha) \to N\left(0, \frac{1}{I(\alpha)}\right), \quad \text{as} \quad n \to \infty$$

where α_{ML} is a maximum likelihood estimator of α from the sample with n elements.

$$I_k(\alpha) \to I(\alpha), \quad \text{as } k \to \infty$$

therefore for large k it holds that

$$\sqrt{n}(\alpha_k - \alpha) \sim N\left(0, \frac{1}{I_k(\alpha)}\right)$$

and this identity will be used to testify the quality of the estimator α_k. The calculation of the estimate of α can be made faster by means of the fact that both $J_k(x, \alpha)$ and $J_{k,n}(\alpha)$ are monotone in α therefore we can accelerate calculation by virtue of dividing the interval.

The Choice Of $t = 0, t_1, .., t_k$ That We Recommend

Let us choose the following representation of J:

$$J^*(x, \alpha) = \sum_{j=-k}^{k} a_j \exp(it_j x), \text{where } a_{-j} = a_j \text{ and } t_{-j} = -t_j, j = 1, 2, .., k$$

Then

$$J^*(x, \alpha) = a_0 + \sum_{j=1}^{k} a_j(\exp(it_j x) + \exp(-it_j x)) = a_0 + 2\sum_{j=1}^{k} a_j \cos(t_j x)$$

From this representation we will get the system:

$$\sum_{v=0}^{k} a_v(\exp\left(-|t_v - t_j|\right)^{\alpha} + \exp\left(-|t_v + t_j|\right)^{\alpha}) = -|t_j|^{\alpha} \ln|t_j| \exp(-|t_j|^{\alpha}),$$

j=0,1,..,k. And, solving this system we will get the vector of coefficients $(\hat{a}_j), j = 0, 1, .., k$. The advantage of this function is that it is real, moreover it shows to converge quicker than general function. Having chosen such shape of $J(x, \alpha)$ we propose to choose the points $t = (0, t_1, .., t_k)^{\mathrm{T}}$ as follows:

$$t_j = j \cdot \Delta,$$

$$\Delta = \frac{1}{\sqrt{k+1}}.$$

The proposition is based on applicable principle i.e. my estimates of α and Fisher information are of the biggest quality under such choice.

9.3 MLP estimates for α of $GS_{\alpha}(1, 0, 0)$

Geometric stable distributions have an explicit form of the characteristic function. And this fact gives us the possibility to calculate $J_k(x, \alpha)$ for geometric stable distributions. Let us have the following representation of J_k for $GS_{\alpha}(1, 0, 0)$:

$$J_k(x, \alpha) = \sum_{j=0}^{k} a_j \exp(it_j x)$$

Then

$$E(J_k(X, \alpha) - J(X, \alpha)) \exp(it_j X) = 0$$

if $J_k(x, \alpha)$ is projection. Then

$$\int_{-\infty}^{\infty} J_k(x, \alpha) \exp(it_j x) p(x, \alpha) dx = \int_{-\infty}^{\infty} \sum_{v=0}^{k} a_v \exp(it_v x) \exp(it_j x) p(x, \alpha) dx =$$

$$= \sum_{v=0}^{k} a_v \int_{-\infty}^{\infty} \exp(it_v x) \exp(it_j x) p(x, \alpha) dx = \sum_{v=0}^{k} a_v \int_{-\infty}^{\infty} \exp(i(t_v + t_j)x) p(x, \alpha) dx =$$

$$= \sum_{v=0}^{k} \frac{a_v}{1 + |t_v + t_j|^{\alpha}}$$

and

$$\int_{-\infty}^{\infty} J(x,\alpha)\exp(it_j x)p(x,\alpha)dx = \frac{\partial}{\partial\alpha}\left(\frac{1}{1+|t_j|^{\alpha}}\right) = -\frac{|t_j|^{\alpha}\ln|t_j|}{(1+|t_j|^{\alpha})^2}$$

Hence follows the system of linear equations:

$$\sum_{v=0}^{k}\frac{a_v}{1+|t_v+t_j|^{\alpha}} = -\frac{|t_j|^{\alpha}\ln|t_j|}{(1+|t_j|^{\alpha})^2}, \quad j=0,1,2,..,k$$

whose solution gives us the representation of $J_k(x,\alpha)$ for geometric stable distribution $Ge_{\alpha}(1,0,0)$.
When

$$J_k(x,\alpha) = a_0 + 2\sum_{j=1}^{k} a_j\cos(t_j x)$$

then we have the following system of the linear equations:

$$\sum_{v=0}^{k} a_v\left(\frac{1}{1+|t_v+t_j|^{\alpha}} + \frac{1}{1+|t_v-t_j|^{\alpha}}\right) = -\frac{|t_j|^{\alpha}\ln|t_j|}{(1+|t_j|^{\alpha})^2}, \quad j=0,1,2,..,k$$

In this work we will not demonstrate the ML-like estimates of α of the standard geometric stable distributions because the core of their estimation is the same as one of the standard $S_{\alpha}S$ distributions.

Chapter 10

Comparison of α_k and $\alpha_{k,N}$ with other estimators

α	Type of estimator	Mean	Variance	$\pm 2\sigma$
1.1	$\hat{\alpha}_{150,50}$	1.1371	0.0015	[0.9997,1.1840]
	Hill	0.9709	0.6425	[-0.6322,2.5740]
1.2	$\hat{\alpha}_{150,50}$	1.2313	0.0019	[1.1426,1.3199]
	Hill	0.8279	0.6385	[-0.7701,2.4261]
1.3	$\hat{\alpha}_{150,50}$	1.3292	0.0021	[1.2368,1.4215]
	Hill	0.8157	0.4968	[-0.5939,2.2255]
1.4	$\hat{\alpha}_{150,50}$	1.4150	0.0021	[1.3228,1.5079]
	Hill	0.6272	0.4305	[-0.6850,1.9395]
1.5	$\hat{\alpha}_{150,50}$	1.5243	0.0024	[1.4259,1.6226]
	Hill	0.6207	0.4179	[-0.6721,1.9136]
1.6	$\hat{\alpha}_{150,50}$	1.6164	0.0016	[1.5345,1.6982]
	Hill	0.6991	0.4851	[-0.6939,2.0922]
1.7	$\hat{\alpha}_{150,50}$	1.7152	0.0021	[1.6227,1.8077]
	Hill	0.5346	0.3888	[-0.7125,1.7818]
1.8	$\hat{\alpha}_{150,50}$	1.8016	0.0017	[1.7188,1.8844]
	Hill	0.5013	0.3377	[-0.6608,1.6635]
1.9	$\hat{\alpha}_{150,50}$	1.8997	0.0013	[1.8265,1.9731]
	Hill	0.4258	0.3483	[-0.7546,1.6063]

Table 10.1: Estimators of alpha with 1000 observations. For Hill estimate we took k=2

α	Type of estimator	Mean	Variance	$\pm 2\sigma$
1.1	$\hat{\alpha}_{150,50}$	1.13710	0.00153	[0.99976,1.18400]
	Pickand	0.96719	0.09602	[0.34742,1.58695]
1.2	$\hat{\alpha}_{150,50}$	1.23130	0.001965	[1.14263,1.31997]
	Pickand	1.06231	0.09158	[0.45706,1.66757]
1.3	$\hat{\alpha}_{150,50}$	1.32920	0.00213	[1.23680,1.42150]
	Pickand	1.17840	0.12286	[0.47734,1.87945]
1.4	$\hat{\alpha}_{150,50}$	1.41505	0.00215	[1.32208,1.50792]
	Pickand	1.26820	0.12629	[0.55752,1.97907]
1.5	$\hat{\alpha}_{150,50}$	1.52433	0.00241	[1.42590,1.62260]
	Pickand	1.34944	0.12255	[0.64927,2.04960]
1.6	$\hat{\alpha}_{150,50}$	1.61649	0.00167	[1.53455,1.69825]
	Pickand	1.41827	0.14474	[0.65737,2.17917]
1.7	$\hat{\alpha}_{150,50}$	1.71526	0.00213	[1.62269,1.80771]
	Pickand	1.51282	0.14024	[0.76382,2.26181]
1.8	$\hat{\alpha}_{150,50}$	1.80165	0.00171	[1.71885,1.88445]
	Pickand	1.58867	0.13059	[0.86592,2.31142]
1.9	$\hat{\alpha}_{150,50}$	1.89979	0.00134	[1.82648,1.97311]
	Pickand	1.64280	0.15890	[0.84563,2.44014]

Table 10.2: This table shows advantages of $\alpha_{150,100}$ opposed to Pickand estimate

α	Type of estimator	Mean	Variance	$\pm 2\sigma$
1.1	$\hat{\alpha}_{150,100}$	1.1052	0.0003	[1.070,1.139]
	$\hat{\alpha}_{5000,5,1.5}$	1.0997	0.0004	[1.059,1.140]
1.2	$\hat{\alpha}_{150,100}$	1.2054	0.00035	[1.167,1.243]
	$\hat{\alpha}_{5000,5,1.5}$	1.2005	0.00047	[1.059.1.140]
1.3	$\hat{\alpha}_{150,100}$	1.3000	0.00035	[1.262,1.337]
	$\hat{\alpha}_{5000,5,1.5}$	1.3000	0.00052	[1.254,1.345]
1.4	$\hat{\alpha}_{150,100}$	1.4074	0.00049	[1.363,1.451]
	$\hat{\alpha}_{5000,5,1.5}$	1.3997	0.00055	[1.350,1.440]
1.5	$\hat{\alpha}_{150,100}$	1.5010	0.00046	[1.458,1.543]
	$\hat{\alpha}_{5000,5,1.5}$	1.5002	0.00058	[1.450,1.540]
1.8	$\hat{\alpha}_{150,100}$	1.8004	0.00033	[1.764,1.836]
	$\hat{\alpha}_{5000,5,1.5}$	1.8006	0.00044	[1.755,1.845]
1.9	$\hat{\alpha}_{150,100}$	1.8988	0.00023	[1.868,1.929]
	$\hat{\alpha}_{5000,5,1.5}$	1.9000	0.00037	[1.860,1.931]

Table 10.3: This table compares the CF estimator $\hat{\alpha}_{5000,5,1.5}$ with $\hat{\alpha}_{150,100}$ when we possess 5000 observations. Let us remind that $\hat{\alpha}_{5000,5,1.5}$ is CF-based estimator of α with 5 summands, and z equal to 1.5

α	Type of estimator	Mean	Variance	$\pm 2\sigma$
1.1	$\hat{\alpha}_{150,50}$	1.1371	0.0015	[0.9997,1.1840]
	$\hat{\alpha}_{1000,100,8}$	1.10255	0.0021	[1.0110,1.1942]
1.2	$\hat{\alpha}_{150,50}$	1.2313	0.0019	[1.1426,1.3199]
	$\hat{\alpha}_{1000,100,8}$	1.2026	0.0022	[1.1088,1.2964]
1.3	$\hat{\alpha}_{150,50}$	1.3292	0.0021	[1.2368,1.4215]
	$\hat{\alpha}_{1000,100,8}$	1.3026	0.0024	[1.2046,1.4006]
1.4	$\hat{\alpha}_{150,50}$	1.4150	0.0021	[1.3221,1.5079]
	$\hat{\alpha}_{1000,100,8}$	1.4023	0.0026	[1.3003,1.5042]
1.5	$\hat{\alpha}_{150,50}$	1.5243	0.0024	[1.4259,1.6226]
	$\hat{\alpha}_{1000,100,8}$	1.5031	0.0028	[1.3972,1.6089]
1.6	$\hat{\alpha}_{150,50}$	1.6164	0.0016	[1.5345,1.6982]
	$\hat{\alpha}_{1000,100,8}$	1.6030	0.0029	[1.4953,1.7107]
1.7	$\hat{\alpha}_{150,50}$	1.7152	0.0021	[1.6227,1.8077]
	$\hat{\alpha}_{1000,100,8}$	1.7012	0.0027	[1.5973,1.8051]
1.8	$\hat{\alpha}_{150,50}$	1.8016	0.0017	[1.7188,1.8844]
	$\hat{\alpha}_{1000,100,8}$	1.8012	0.0024	[1.7032,1.8991]
1.9	$\hat{\alpha}_{150,50}$	1.8997	0.0013	[1.8264,1.9731]
	$\hat{\alpha}_{1000,100,8}$	1.9022	0.0016	[1.8222,1.9822]

Table 10.4: From the this table we can see that the variance is lower for $\hat{\alpha}_{150,50}$ and for $\alpha = 1.9, 1.8$ and 1.6 its interval $\mu \pm 2\sigma$ lies totally in the same interval of $\alpha_{1000,100,8}$. For other points it insignificantly exceeds that interval.

α	Type of estimator	Mean	Variance	$\pm 2\sigma$
1.1	$\hat{\alpha}_{150}$	1.11584	0.00056	[1.0684,1.1632]
	$\hat{\alpha}_{1000,100,8}$	1.10255	0.0021	[1.0109,1.1942]
1.2	$\hat{\alpha}_{150}$	1.20275	0.00174	[1.1191,1.2863]
	$\hat{\alpha}_{1000,100,8}$	1.2026	0.0022	[1.1087,1.2964]
1.3	$\hat{\alpha}_{150}$	1.29915	0.00213	[1.2067,1.3915]
	$\hat{\alpha}_{1000,100,8}$	1.30260	0.0024	[1.2046,1.4005]
1.4	$\hat{\alpha}_{150}$	1.39966	0.00181	[1.3145,1.4847]
	$\hat{\alpha}_{1000,100,8}$	1.4023	0.0026	[1.3003,1.5042]
1.5	$\hat{\alpha}_{150}$	1.49584	0.00206	[1.4049,1.5867]
	$\hat{\alpha}_{1000,100,8}$	1.5031	0.0028	[1.3972,1.6089]
1.6	$\hat{\alpha}_{150}$	1.59134	0.00151	[1.5135,1.6691]
	$\hat{\alpha}_{1000,100,8}$	1.6030	0.0029	[1.4953,1.7107]
1.7	$\hat{\alpha}_{150}$	1.70204	0.00187	[1.6154,1.7886]
	$\hat{\alpha}_{1000,100,8}$	1.7012	0.0027	[1.5972,1.8051]
1.8	$\hat{\alpha}_{150}$	1.79173	0.00159	[1.7118,1.8717]
	$\hat{\alpha}_{1000,100,8}$	1.8012	0.0024	[1.7032,1.8991]
1.9	$\hat{\alpha}_{150}$	1.89991	0.00113	[1.8326,1.9671]
	$\hat{\alpha}_{1000,100,8}$	1.9022	0.0016	[1.8222,1.9822]

Table 10.5: This table compares α_{150} with the CF estimator when we possess 1000 observations.

From both previous tables we can see that the variance is the lowest for α_{150}. The second lowest variance has $\alpha_{150,50}$. The CF estimator has a higher variance than ML-like estimates but there isn't such a great difference. The next table summarizes the variances of all considered estimators and compares them with the variance of ML-estimate which equals:

$$\frac{1}{I(\alpha)1000}$$

if the number of observations is 1000.

α	Hill	Pickand	$\hat{\alpha}_{1000,100,8}$	$\hat{\alpha}_{150,50}$	$\hat{\alpha}_{150}$	$\frac{1}{I(\alpha)1000}$
1.1	0.6425	0.0967	0.0021	0.0015	0.0006	0.0016
1.2	0.6385	0.0915	0.0022	0.0019	0.0017	0.0018
1.3	0.4968	0.1228	0.0024	0.0021	0.0021	0.0020
1.4	0.4305	0.1262	0.0026	0.0021	0.0018	0.0021
1.5	0.4117	0.1225	0.0028	0.0024	0.0020	0.0022
1.6	0.4851	0.1447	0.0029	0.0016	0.0015	0.0021
1.7	0.3888	0.1402	0.0027	0.0021	0.0018	0.0020
1.8	0.3377	0.1306	0.0024	0.0017	0.0015	0.0016
1.9	0.4258	0.1589	0.0016	0.0016	0.0011	0.0011

Table 10.6: Table of comparison different estimator by their variance

We can see from the table that the variance of the estimator $\hat{\alpha}_{150}$ is even lower than the one calculated by the Fisher information but *can be explained by biasedness of this estimator*, a low number of estimates, (we have taken only 100 of them and calculated variance of that sample) or by low number of observations. If we increase the number of observations to 5000 then the variance of $\hat{\alpha}_{150}$ will be higher, e.g. when $\alpha = 1.5$ then the Variance of the sample with 100 estimates of α equals 0.00047 and the variance of ML-estimate equals 0.00044, if $\alpha = 1.6$ then the variance of the sample with 100 elements equals to 0.0004752 and the one of ML-estimate is equal to 0.00043, and so on. Moreover the oscillations of α_{150} are statistically insignificant. Among all considered estimators we propose to use α_k, k can be 150 or higher. This estimator begins to work well from number of observations equal to 500. When $\alpha = 1.6$ then the Mean of the sample of 100 estimates is equal to 1.60993 and the standard deviation equals to 0.076, if $\alpha = 1.5$ then mean = 1.48994 and the standard deviation = 0.074. The Fisher information is the least in the points 1.5 and 1.6 therefore, the variance will be lower for the rest of points. The deviation less than 0.1 is appropriate for practical purposes therefore, we conclude that from 500 observation the estimator α_{500} is applicable in practice. Although this is the estimator for the distribution $S_\alpha(1, 0, 0)$ this method can be expanded to the problem of estimating multivariate parameters e.g. (α, σ) or $(\alpha, \sigma, \beta, \mu)$ and instead of the Fisher information we will consider the Fisher information matrix. We will only outline the case when other parameters are unknown. If σ and α

are unknown and $\beta = \mu = 0$. Then the characteristic function will be of the form:

$$\psi(t, \alpha) = \exp(-\sigma^\alpha |t|^\alpha),$$

and

$$\frac{\partial \psi(t,\alpha)}{\partial \alpha} = \exp\left(-\sigma^\alpha |t|^\alpha\right)(-\sigma^\alpha |t|^\alpha \ln(\sigma) - \sigma^\alpha |t|^\alpha \ln|t|) =$$
$$= \exp\left(-\sigma^\alpha |t|^\alpha\right)(-\sigma^\alpha |t|^\alpha \ln(\sigma|t|))$$

If we project $J(x, \alpha, \sigma)$ to $\{1, \exp(it_1 x), \exp(it_2 x), ..., \exp(it_k x)\}$ we will have the following scalar multiplication:

$$\langle \exp(it_m X), \exp(it_n X) \rangle = \exp(-\sigma^\alpha |t_m + t_n|^\alpha)$$

and the system of equations:

$$\sum_{v=0}^{k} a_v \exp(-\sigma^\alpha |t_v + t_j|^\alpha) = \exp\left(-\sigma^\alpha |t_j|^\alpha\right)(-\sigma^\alpha |t_j|^\alpha \ln(\sigma|t_j|)), j = 0, 1, 2, .., k$$

From this system we will calculate $a_m, m = 0, 1, 2, .., k$ and substitute it to the expression of $J_k(x, \alpha.\sigma)$. Then we substitute this sample into the expression of $J_{k,n}(x, \alpha, \sigma)$ i.e.

$$J_{k,n}(\alpha, \sigma) = \sum_{j=1}^{n} J_k(X_j, \alpha, \sigma)$$

and find the root of the problem

$$J_{k,n}(\alpha, \sigma) = 0.$$

The ML-like estimates of α and σ will be as follows:

$$\{(\alpha, \sigma) : J_{k,n}(\alpha, \sigma) = 0\}$$

The function $J_{k,n}(\alpha, \sigma)$ is monotone in α and σ therefore, the solution of that problem will be unique. Suppose the next problem when α and σ are known and we have to estimate β and μ. Then making the same transformations we will have analogous results for μ and β.
If we have a sample $(X_1, X_2, .., X_{2n})$ from $S_\alpha(\sigma, \beta, \mu)$ where all 4 parameters are unknown then transforming it to $(X_1 - X_2, X_3 - X_4, .., X_{2n-1} - X_{2n})$ we will get a sample from $S_\alpha(\sigma, 0, 0)$ with n elements. The estimation of parameters of this problem was just described. Taking into account the estimates ($\hat{\alpha}$ and $\hat{\sigma}$) of α and σ we will get the problem where α and σ are known and β and μ are unknown. This problem was also just described. So we get the following estimator of $\theta = (\alpha, \sigma, \beta, \mu)^{\mathrm{T}}$ which equals $(\alpha_{k,n}, \sigma_{k,n}, \beta_{k,2n}, \mu_{k,2n})$.

Chapter 11

Methods of calculation of the Fisher Information

There are many methods of calculation of the Fisher information. To do this we can apply projections. I'll shows two method of the calculation:

First method:

Using the notation shown in the previous section we can calculate the Fisher information as follows:

$$I_k(\alpha) = (A(\alpha) \cdot c(\alpha))^T \cdot c(\alpha)$$

The figure 12.1 shows the convergence of $I_k(\alpha)$ to the Fisher information. It is enough to take only 100 summands. If we compare the graphs with 100 and 500 summands they will simply intersect so much that it will be impossible to differ them visually.

If

$$J_k(x, \alpha) = a_0 + \sum_{j=1}^{k} a_j \cos(x \cdot t_j)$$

then its integration is very easy task.

$$I_k(\alpha) = \int_{-\infty}^{\infty} J_k^2(x, \alpha) p(x, \alpha) dx$$

$J_k^2(x, \alpha)$ can be expressed as a sum of sines and cosines. Then, knowledge of the characteristic function and oddness of sine enables to precisely calculate $I_k(\alpha)$

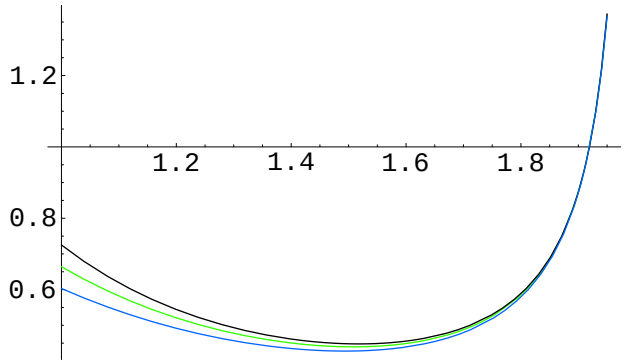


Figure 11.1: The graphs with 10 (blue line), 20 (green line), 400 (black line) summands

Chapter 12

Conclusion

The stable laws have wide application in financial practice, particularly when there is necessity to describe the value of assets whose price can dramatically jump or drop. The examples are following: price of petrol, price of currency of countries having crisis or very dynamically developing economy. The model NGARCH with smoothly truncated stable innovation, has been proven by Christian Menn and Svetlozar Rachev to be the best predictor of Black Monday. The technique of estimation of the parameters is based on setting initial value of parameters of the stable law. Based on given parameters of the stable law, we estimate the parameters of the modification of GARCH-model and from them we can recover the time series of residuals and estimate its parameters and then based on these parameters estimate the parameters of the model, continuing this until Kolmogorov-Smirnov or some other criterion distance stop decreasing. For this method it is very important to have quick and precise estimates of the parameters of the parameters of stable innovation therefore we devoted so much strength to solving this problem. The theory created by Kagan has been shown to applicable very well for parameter estimation of the stable distributions. Before some researchers were estimating parameters of the stable distributions by means of approximation of $J(x, \alpha)$ by knowledge of the characteristic function and integral expression of the density, but method, similar to ML methodology, was very slow and not precise enough. The estimates of α gained by the method of projections have the least variance among all known estimates of α. This methodology can be applied in modification of ARCH/GARCH models for the parameter estimation of the stable distributions and grant more precise estimates than before. When we have a sample about which we suppose that it has stable

distribution, we propose to conduct test of stability, described in previous chapters, to be sure that the innovation is stable, otherwise we can apply modifications of stable laws like geometric stable and others.

Chapter 13

APPENDICES

13.1 APPENDIX A

Hill estimator

This estimator is based on the fact that the stable distributions are Pareto-type i.e. for large numbers it holds:

$$1 - F(x) \approx cx^{-\alpha_P}, (\alpha_P > 0, c > 0)$$

(See [8]) Alpha can be measured by virtue of this estimator but not precisely what will be explained later. Hill estimate is given explicitly by:

$$\hat{\alpha}_{\text{Hill}} = \frac{1}{(1/k)\sum_{j=1}^{k} \ln X_{(n+1-j)} - \ln X_{(n-k)}}$$

with the error

$$STD(\hat{\alpha}_{\text{Hill}}) = \frac{k\hat{\alpha}_{\text{Hill}}}{(k-1)\sqrt{(k-2)}},$$

$k > 2$.
Where $X_{(j)}$ is $j-$th order statistic. This estimation is asymptotically normally distributed:

$$(\hat{\alpha}_{\text{Hill}}^{-1} - \alpha^{-1})k^{1/2} \sim N(0, \alpha^{-2})$$

for large values of n and appropriate k.
As it is said above this estimator is not so good for measuring index of stability. We publish it only to demonstrate theoretical properties of the stable

laws. The densities of table laws are asymptotically Pareto-densities, i.e. for very large numbers they behave as Pareto-density. Suppose that a_0 and b_0 are the numbers such that in the intervals $(-\infty, a_0)$ and (b_0, ∞) the density of the stable law behaves as a Pareto density function. In practice most of observations lie outside the intervals (b_0, ∞) and $(-\infty, a_0)$ and sometimes we need to simulate more than million observations to have one in the set

$$(b_0, \infty) \cup (-\infty, a_0)$$

That's why this estimator is not appropriate in spite of intuitive feeling. Moreover many times a Hill estimate of α exceeds 2 and can even reach 20, 30, etc. The reason is simple. Most of observations lie in the interval (a_0, b_0) where the density doesn't behave like Pareto-type density. In that interval the slope is relatively high and if we estimate α under assumption that the density function is Pareto then the estimate of α will be higher than α. Therefore α is almost always overestimated when we use Hill estimator.

Unconditional Pickand Estimator

book: [8]
We can construct this estimator using infinite Bergstroem series for α-stable distribution and applying order statistics:

$$S_\alpha(x) = 1 + \frac{1}{\pi} \sum_{m=1}^{\infty} (-1)^m \frac{\Gamma(\alpha m)}{m!} x^{-\alpha m} \sin \frac{\alpha m \pi}{2}$$

as $x \to \infty$.

$$S_\alpha(X_{(n-j+1)}) = 1 + \frac{1}{\pi} \sum_{m=1}^{\infty} (-1)^m \frac{\Gamma(\alpha m)}{m!} X_{(n-j+1)}^{-\alpha m} \sin \frac{\alpha m \pi}{2} = \frac{n-j+1}{n}$$

$$1 - S_\alpha(X_{(n-j+1)}) = 1 - \frac{n-j+1}{n} = \frac{j-1}{n}$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} (-1)^{m+1} \frac{\Gamma(\alpha m)}{m!} X_{(n-j+1)}^{-\alpha m} \sin \frac{\alpha m \pi}{2}$$

hence follows the equality:

$$\frac{j-1}{n} = \frac{1}{\pi} \sum_{m=1}^{\infty} (-1)^{m+1} \frac{\Gamma(\alpha m)}{m!} X_{(n-j+1)}^{-\alpha m} \sin \frac{\alpha m \pi}{2}$$

If we take only one number from this infinite series we'll get for $j = k$ and $j = 2k$

$$\frac{k-1}{n} \approx c_\alpha X_{(n-k+1)}^{-\hat{\alpha}}$$

$$\frac{2k-1}{n} \approx c_\alpha X_{(n-2k+1)}^{-\hat{\alpha}}$$

Further we have:

$$\frac{\frac{k-1}{n}}{\frac{2k-1}{n}} \approx \frac{c_\alpha X_{(n-k+1)}^{-\hat{\alpha}}}{c_\alpha X_{(n-2k+1)}^{-\hat{\alpha}}}$$

For large k holds:

$$2 \approx \frac{X_{(n-2k+1)}^{-\hat{\alpha}}}{X_{(n-k+1)}^{-\hat{\alpha}}}$$

$$\ln 2 \approx \hat{\alpha}(\ln X_{(n-k+1)} - \ln X_{(n-2k+1)})$$

Hence

$$\hat{\alpha}_{UP} = \frac{\ln 2}{\ln X_{(n-k+1)} - \ln X_{(n-2k+1)}}$$

This estimate is not Pareto-type and it takes into account approximation of the distribution function. Heavy tails which appear in the expression of the density and distribution functions endanger preciseness of the estimator. The reason is the same as for the Hill estimator but that factor which worsens precision is reduced because it is based on theoretical properties of stable laws, not on asymptotical as for Hill estimator. Hill estimator takes into account only heavy tails, unconditional Pickand estimator takes into account all the values which the stable random variable can reach but it takes into account only one number in the Bergstroem series but if we take more than one number then the estimator won't have an explicit form but it will have a better preciseness. If we take two summands then we will have Modified Unconditional Pickand Estimator. But even in this case such approximation of the distribution function isn't very effective. Even if we have a precise explicit expression of the distribution function where α is one of variables we also have a problem of estimating of the index of stability because if our sample isn't large enough we don't have observations lying in tails of the distribution i.e. lying in the set:

$$(b_1, \infty) \cup (-\infty, a_1)$$

And this will also affect preciseness. The estimator based on distribution and density functions requires that the sample is large enough to have a number of points lying in

$$(b_1, \infty) \cup (-\infty, a_1)$$

comparable to the theoretical probabilities, i.e. if we denote n_1 the number of points in that interval and N the sample size, than it must hold:

$$\frac{n_1}{N} \approx \int_{-\infty}^{a_1} f(x)dx + \int_{b_1}^{\infty} f(x)dx$$

It can be attained only for very large number of observations which exceeds 10000 and depends on α. The smaller α the larger number of observations we need.

13.2 APPENDIX B

The stable distributions have a wide application beyond the world of finances. They have applications in Physics, Astrophysics, Theory of stochastic fractals, Biology and other spheres of human activities. Stable distributions with $\alpha < 1$ have wide application in physics and biology. The applications of are perfectly described in work by Zolotarev V. and Uchaikin V. CHANCE and STABILITY Satble Distributions and their Applications (book: [9]).

13.3 APPENDIX C

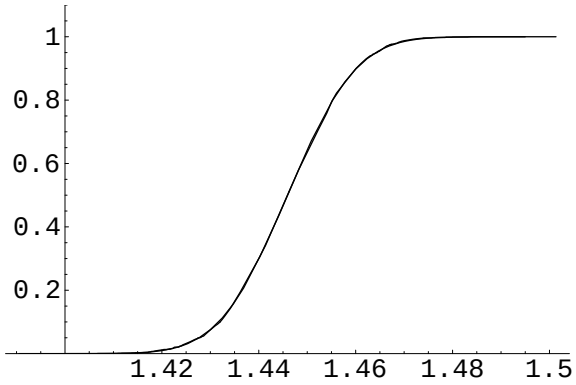


Figure 13.1: Checking the normality. Graph of the empirical distribution function for α as a part of the overall estimator $(\alpha_{30000,50,8}, \sigma_{30000,50,8}, \beta_{30000,50,1}, \mu_{30000,50,1})$and N$(\mu, \sigma)$where μ equals an average of the sample of the estimates of α and σ is their standard deviation and $(\alpha, \sigma, \beta, \mu) = (1.4, 1.05, 0.23, 0)$. For 2000 simulations the difference between theoretical and empirical distribution function is so small that the difference is invisible

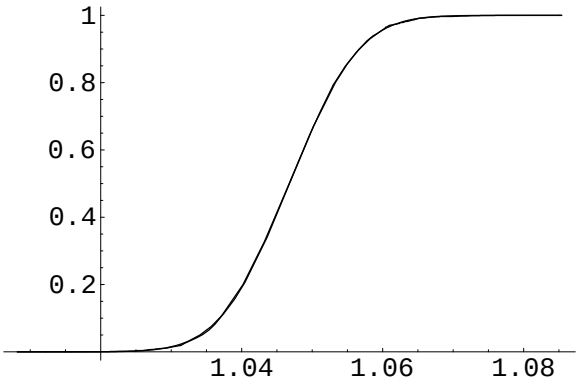


Figure 13.2: Checking the normality. Graph of the empirical distribution function for σ as a part of the overall estimator $(\alpha_{30000,50,8}, \sigma_{30000,50,8}, \beta_{30000,50,1}, \mu_{30000,50,1})$and $N(\mu, \sigma*)$where μ equals an average of the sample of the estimates of σ and $\sigma*$ is their standard deviation and $(\alpha, \sigma, \beta, \mu) = (1.4, 1.05, 0.23, 0)$. For 2000 simulations the difference between theoretical and empirical distribution function is so small that the difference is invisible

13.4 APPENDIX D

13.16:

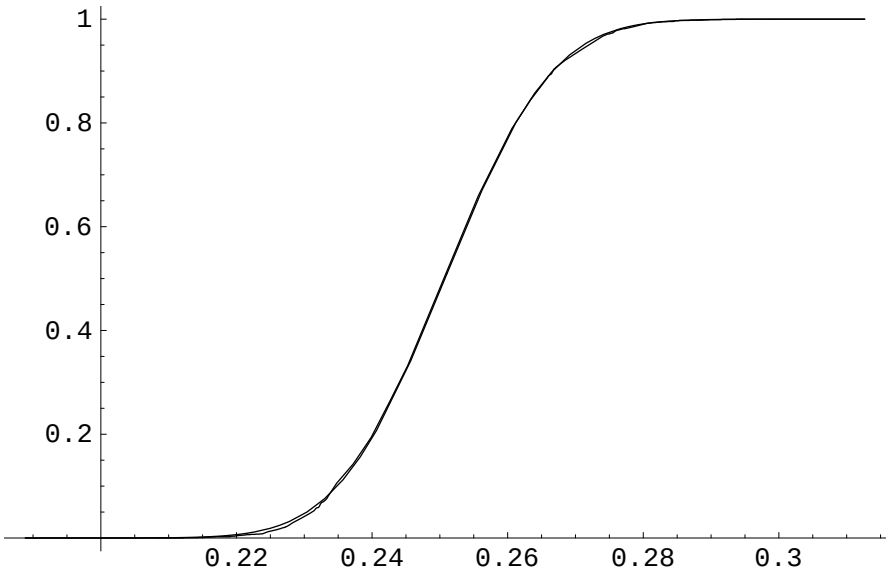


Figure 13.3: Checking the normality. Graph of the empirical distribution function for β as a part of the overall estimator $(\alpha_{30000,50,8}, \sigma_{30000,50,8}, \beta_{30000,50,1}, \mu_{30000,50,1})$

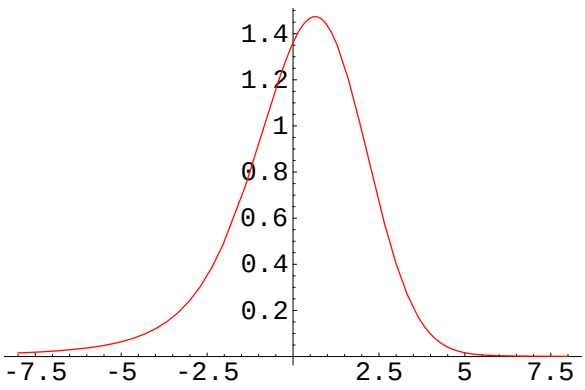


Figure 13.4: graph of the density function of $S_{1.7}(1.2, 0.9, 0)$

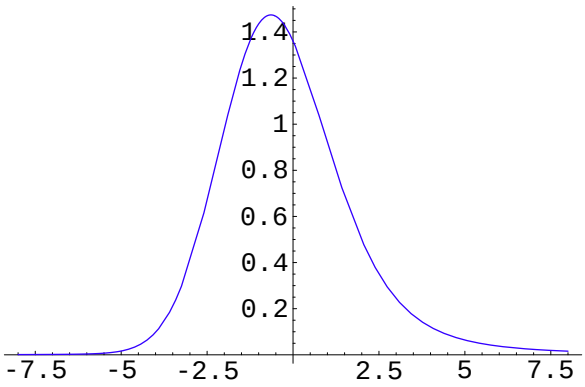


Figure 13.5: graph of the density function of $S_{1.7}(1.2, -0.9, 0)$

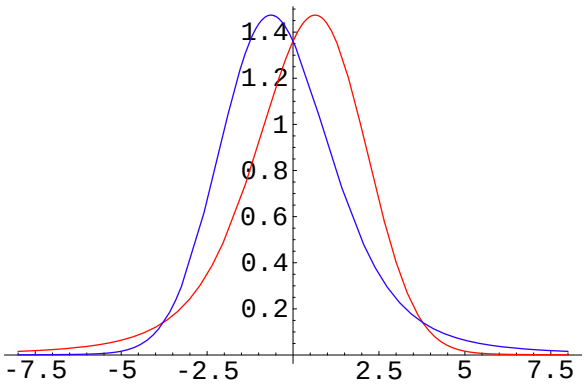


Figure 13.6: Two previous graphs on one picture. This is one more evident reason why we can use the differences $X_1 - X_2 \sim S_\alpha(\sigma, 0, 0)$, $X_1, X_2 \sim S_\alpha(\sigma, \beta, \mu)$ for parameter estimation along with the fact, that α is preserved under any convolution

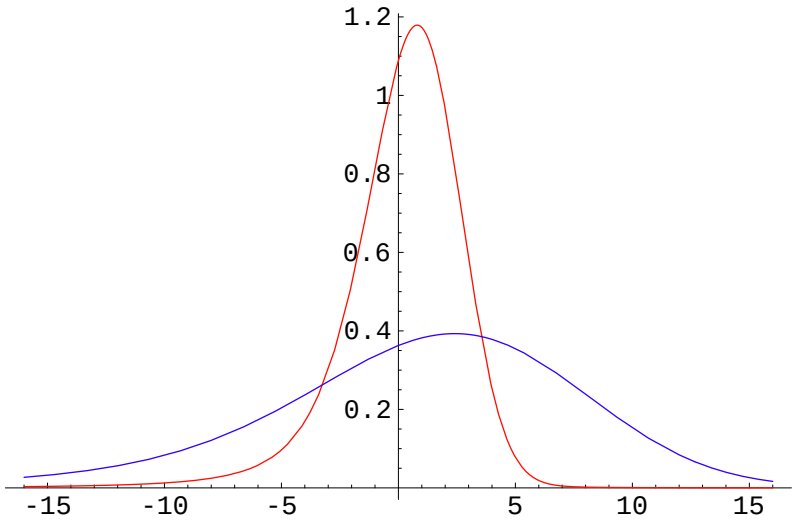


Figure 13.7: The role of the parameter σ. The red graph is the graph of the density function of $S_{1.7}(1.5, 0.9, 0)$ nad the blue graph is the graph of the density function of $S_{1.7}(4.5, 0.9, 0)$

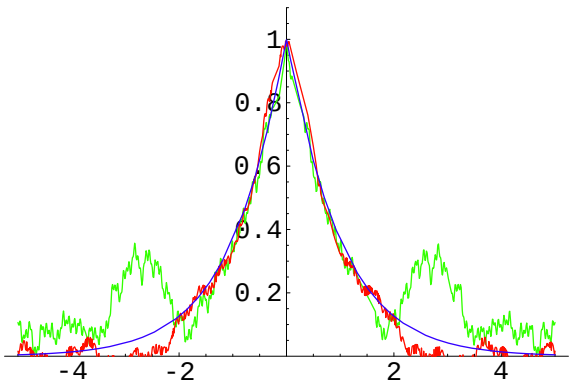


Figure 13.8: Theoretical characteristic function and real part of the empirical characteristic function of $S_{1.1}(1, 0, 0)$ when we possess only 70 observations. We have 2 independent samples with 70 numbers therefore here are 3 graphs. Blue graph is a graph of theoretic ch.f.(This is one more reason why we can't rely on any CF estimator of α when we possess only few observation)

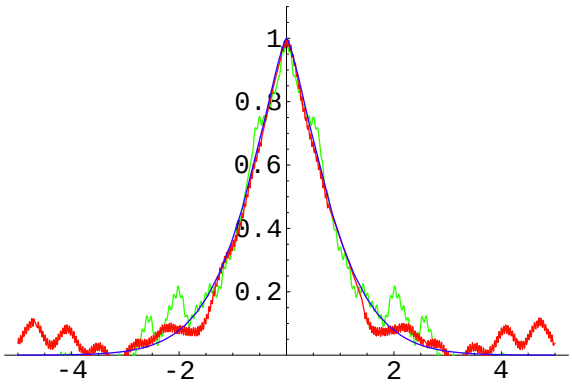


Figure 13.9: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.4}(1, 0, 0)$ when we possess only 70 observations.

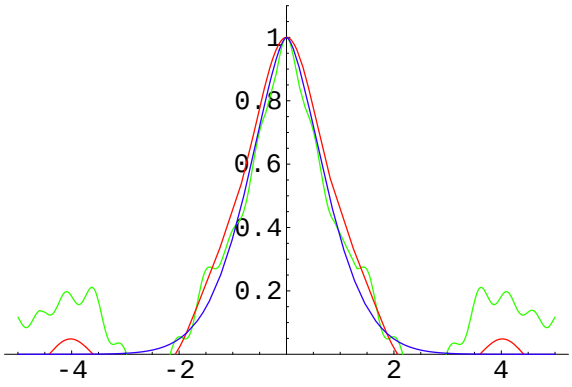


Figure 13.10: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.7}(1, 0, 0)$ when we possess only 70 observations.

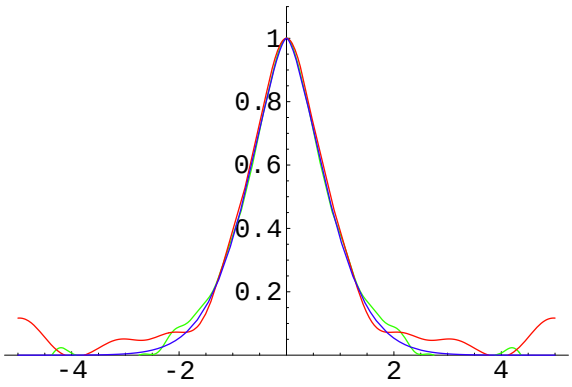


Figure 13.11: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.6}(1, 0, 0)$ when we possess only 140 observations.

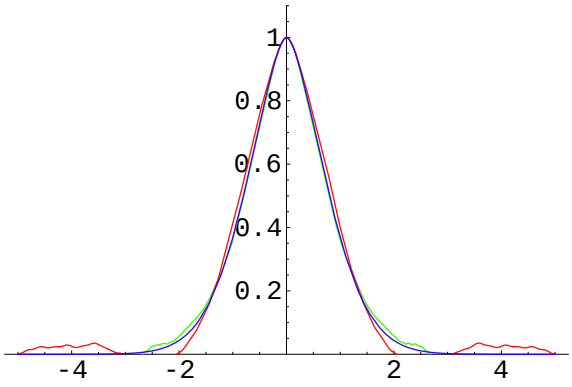


Figure 13.12: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.7}(1, 0, 0)$ when we possess 630 observations.

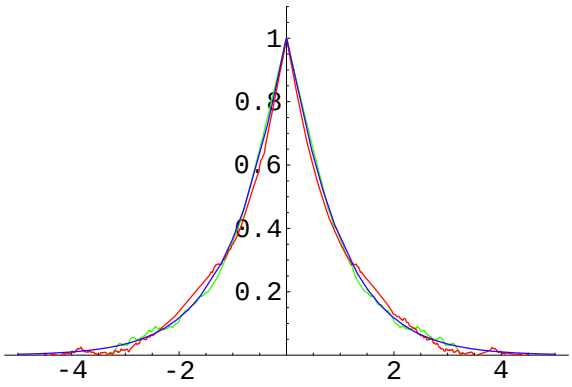


Figure 13.13: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.1}(1, 0, 0)$ when we possess 1050 observations.

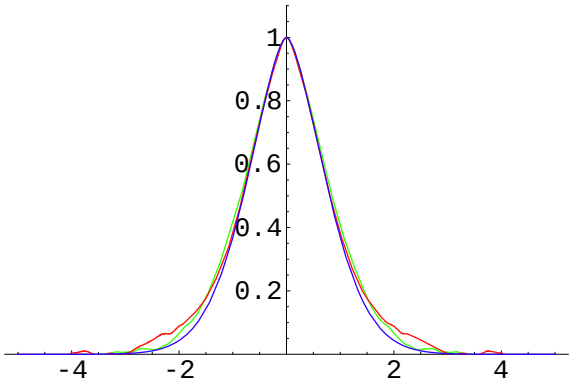


Figure 13.14: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.65}(1, 0, 0)$ when we possess 1050 observations.

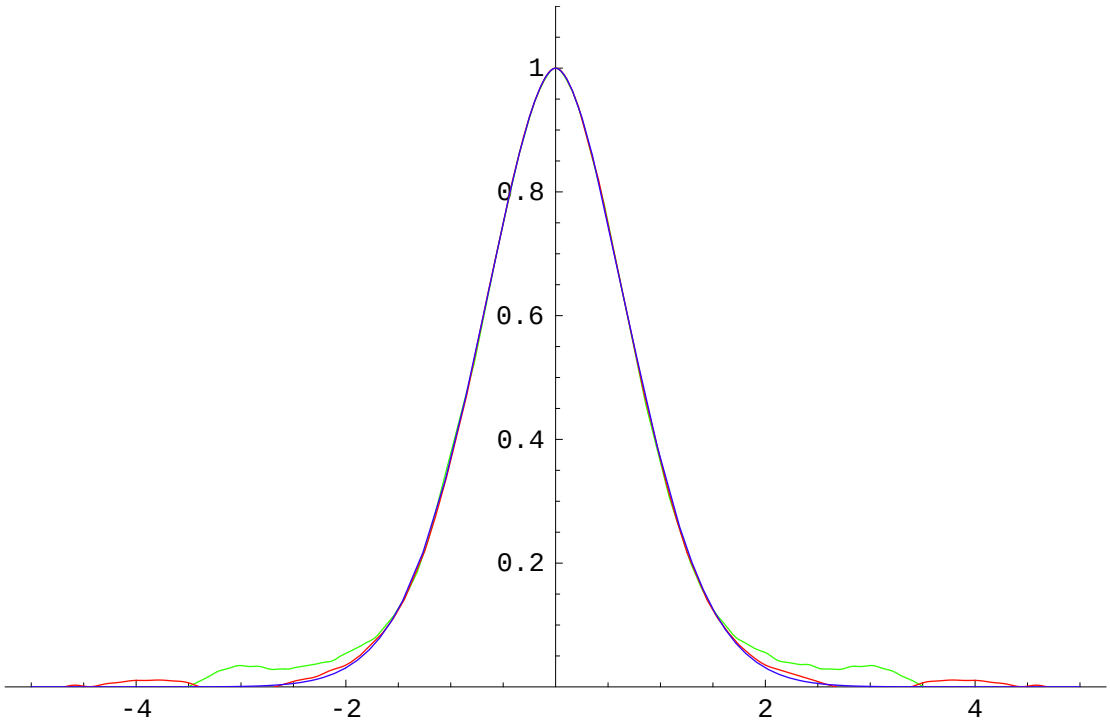


Figure 13.15: Theoretical characteristic function and real part of empirical characteristic function of $S_{1.9}(1, 0, 0)$ when we possess 1610 observations.

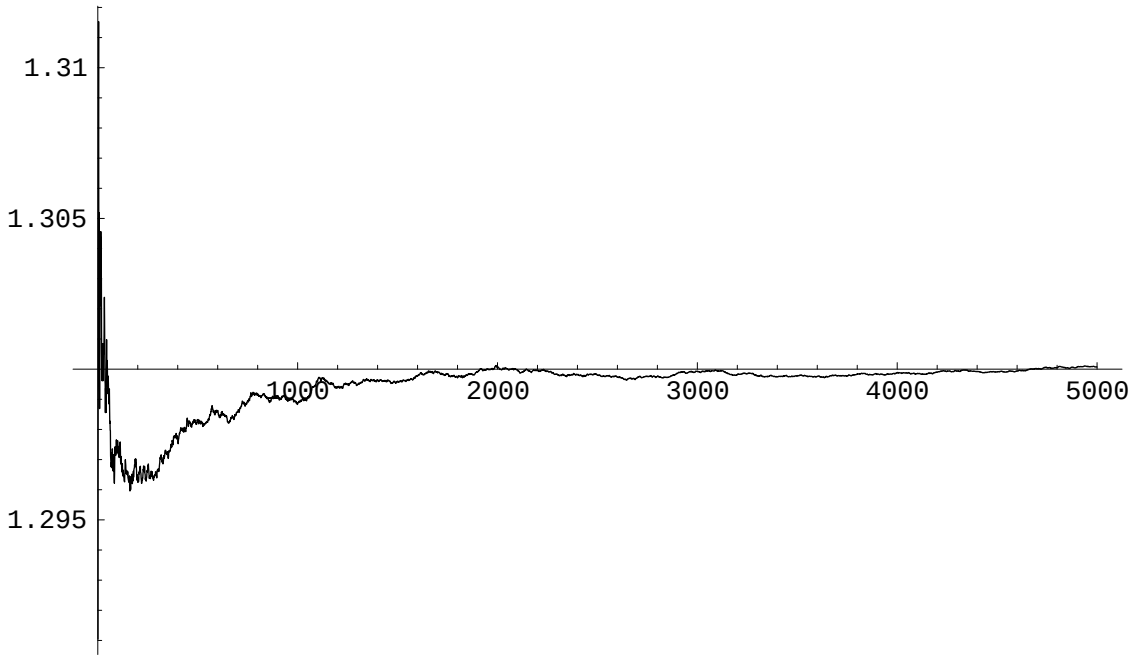


Figure 13.16: Graph of the averages of the sample of $\alpha_{5000,5,1.5}$ from 1 to 5000 sample sizes, where $\alpha = 1.3$

Bibliography

[1] Albert N. Shiryaev(1999): ESSENTIALS OF STOCHASTIC FINANCE. Facts, Models, Theory: World Scientific

[2] C. Menn, S.T. Rachev, "Smoothly Truncated Stable Distributions, GARCH-Models and Option Pricing", 2005

[3] G.P. Patil S. Kotz J.K. Ord(1975) Statistical Distributions in Scientific Work: Nato Advanced Study Institutes Series, Series C: Mathematical and Physical Sciences.Pages: 88-93

[4] Gennady Samorodnitsky Murad S. Taqqu(1994)Stable Non-Gaussian Random Processes: Chapman & Hall/CRC

[5] Kagan, A. M. Fisher information contained in a finite-dimensional linear space, and a properly formulated version of the method of moments. (Russian) Problemy Peredači Informacii 12 (1976), no. 2, 20–42. (Reviewer: S. Kotz)

[6] Lev B. Klebanov(2003) Heavy Tailed Distributions: MatfyzPress. Prague

[7] Rachev, S. and Mitnik S. Stable Paretian Models in Finance , John Wiley, Series in Financial Economics and Quatitative Analysis, Chechester, New York, 2000.

[8] S. Csorgo: Rates of uniform convergence for the empirical characteristic function. Acta Scientiarum Mathematicarum (Szeged) 48 (1985), 97–102. 87b: 60033.

[9] Uchaikin V.V., Zolotarev V.M. Chance and Stability. VSP, Netherlands, Utrecht, 1999.