On the Fitting of Generalized Hyperbolical Distributions to Financial Data

Fredrik Jonsson

June 28, 2004

Abstract

In this master's thesis, we study the performance of the *Generalized Hyperbolic* distribution (GH), when it comes to fitting it to historical data, from the field of mathematical finance. We have based our work on master's thesises, written in the field of mathematical finance in Sweden the last three years. Our aim has been to see if we can get better results, when fitting the GH distribution to the data sets featuring in those thesises, instead of the parametric distribution originally used there, for an as great variety of data sets as possible.

To carry this out, we had to get hold of the data sets from different master's thesises, repeat the work done with the data there, and then fit the GH distribution to the data. Due to the high number of parameters of the GH distribution, we experienced great numerical problems, and had to limit our study to a fewer number of data sets than first intended.

We found that the GH distribution outperformed the distributions originally used in the master's thesises selected, in almost every case. We also got very good results when making use of modified versions of the GH distribution, which we fitted to supposedly extreme value distributed data.

Our main conclusion is that, if you want to achieve good fitting of parametric distributions, to data sets from the field of mathematical finance, then it is worthwhile to try to fit a GH distribution. It is also possible to use modified versions of the GH distribution, that are numerically more tractable, less time consuming to fit, and that also perform very well.

Hopefully, there will be a continued interest in the study of GH distributions in the future.

Acknowledgements

I want to thank my supervisor Patrik Albin, for everything, and the following persons, for beeing very helpful: Viktor Olsbo, Mattias Bengtsson, Erik Brodin, Johan Tykesson, Emma Söderberg, Daniel Andersson, Magnus Lernvik, Ulrika Trolle, Anna Widerberg, Cilla Persson, Katarina Åselius, Gunnar Karlsson at KTH and Lotta Persson.

Contents

1 Introduction

The main task of this master's thesis, is to examine properties of *Generalized Hyperbolic* distributions (GH).

First of all, we have to admit, that the time we had to spend on solving the numerical and computational problems we faced, when estimating the parameters of the GH distribution, made it necessary to study fewer datasets than we originally had intended to. Thus our results might be somewhat less convincing than they could have been.

We started out our work by collecting virtually all master's thesises, that have been written in the field of mathematical finance in Sweden the last three years. Many of these thesises claimed, that according to investigation of historical data, the Normal distribution is not always a good model. Erik Brodin, in his master's thesis [7], tried to improve fittings to two different stock price data sets, by using various distributional assumptions. He found that results were best for the GH distribution. We then decided to investigate how well the GH distribution would do in other cases, with other types of data sets.

To carry out our task, we decided to choose master's thesises corresponding to data from different subfields within mathematical finance. The process in each case, in essence, was to repeat the analyzis of the master's thesises selected, and then see if the GH distribution did perform better.

The disposition of this master's thesis is as follows: First we state some theory that have been used. Then we introduce the datasets that we have studied, and present the results obtained for each case. Last we make some concluding remarks. Computer programs that were used for our analysis are given in appendices.

2 Theory

There are some fundamentals in this study, that we have used, and that we think should be stated so, that nothing is unclear about which definitions we use, and which information we build our statements on.

2.1 Asset Pricing

According to the *Bachelier-Samuelson* model (B-S), the relation between a price process $S = (S_t)_{t\geq 0}$ and a return process $X = (X_t)_{t\geq 0}$, is given by

$$S_t = S_0\,\mathrm{e}^{X_t} \quad \text{for} \quad 0 \leq t \leq T, \quad \text{where} \quad X_t = \mu t + \sigma B_t$$

(see e.g. [5]). Here $B = (B_t)_{t\geq 0}$ is a standard Brownian motion, while the drift coefficient $\mu \in \mathbb{R}$ and the volatility $\sigma > 0$ are parameters.

The basic idea of this thesis, is to follow the thoughts of master's thesises written the last years, emanating from the insight that the B-S process model does not very well describe the behaviour of many financial assets. There are two important assumptions that are often questioned. The first is that consequtive log-returns are *Independent and Identically Distributed* (IID), and the other that the log-returns are *Normal* distributed.

There are several suggestions about how to deal with the two questions raised. Following, for example, Erik Brodin [7], we can modify the B-S model, by replacing the Browninan motion B with a more general Lévy process $L = (L_t)_{t\geq 0}$, so that $X_t = \mu t + \sigma_t L_t$. Here the problem, that log-returns are not IID is taken care of by allowing the volatility to be a stochastic process $\sigma = (\sigma_t)_{t\geq 0}$. The problem that log-returns are not Normal distributed, is taken care of by letting L be another Lévy processes than the Brownian motion, thus allowing other distributions for increments than the Normal.

The main interest for us, in this master's thesis, lies in the second of the above mentioned issues. Thus we want to examine which distributions that are best to model historical data, from the field of mathematical finance. Possibly, we could find a family of distributions that could be successful in some generality in such applications.

2.2 Extreme Value Theory and Risk

After having read the different master's thesises, we realized that, even if the most attention is paid to the distribution of financial assets, there is also great interest in other fields of mathematical finance. One example is the insurance industry, where the calculation of *risk* is in focus. This means that the mathematical models that feature tend to build on *Extreme Value Theory* (EVT) and the *Generalized Pareto* distribution (GP) (see e.g. [9]).

Now the question is: Even if natural, since supported by EVT, is GP always the best distributional assumption, in these contexts? We have made some studies which indicate that better results can be achieved with other distributions. One reason for this might be that the convergence to the limit laws of EVT is slow.

2.3 Distributions

There are two distributions that feature repeatedly, in the different master's thesises. They are the Normal distribution, and the GP distribution. To fix notation, we introduce these two distributions here, together with the GH distribution, and modifications thereof, that will be our main tools. Further distributions will be introduced later, as they are needed.

2.3.1 Normal Distribution

The *standard Normal* distribution ($\mathrm{N}(0,1)$) has probability density function

$$f_{\mathrm{N}}(x;0,1) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2} \quad \text{for} \quad x \in \mathbb{R}.$$

The *Normal* distribution ($N(\mu,\sigma^2)$) has probability density function

$$f_{\mathrm{N}}(x;\mu,\sigma^2) = \frac{1}{\sigma}\,f_{\mathrm{N}}\Big(\frac{x-\mu}{\sigma};0,1\Big) \quad \text{for} \quad x \in \mathbb{R},$$

where the parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are the expected value and the variance, respectively.

2.3.2 Generalized Pareto Distribution

The *Generalized Pareto* distribution (GP) has probability density function

$$f_{\mathrm{GP}}(x;u,\sigma,\xi) = \frac{1}{\sigma}\Big(1+\xi\frac{x-u}{\sigma}\Big)^{-1/\xi-1} \quad \text{for} \quad x > u \tag{1}$$

(see e.g. [9]). Here $u \in \mathbb{R}$ is a threshold, $\xi \geq 0$ a shape parameter, and $\sigma > 0$ a scale parameter.

The GP distribution function is given by

$$F_{\mathrm{GP}}(x;u,\sigma,\xi) = 1-\Big(1+\xi\frac{x-u}{\sigma}\Big)^{-1/\xi} \quad \text{for} \quad x > u. \tag{2}$$

When $\xi = 0$, (1) should be interpreted as the limit of (1) as $\xi \downarrow 0$. This gives the following probability density function

$$f_{\mathrm{GP}}(x;u,\sigma,0) = \frac{1}{\sigma}\,\mathrm{e}^{-(x-u)/\sigma} \quad \text{for} \quad x > u. \tag{3}$$

Of course, this is an exponential distribution, shifted $u \in \mathbb{R}$ to the right, and with variance $\sigma^2 > 0$.

2.3.3 Generalized Hyperbolic Distribution

The probability density function for the *Generalized Hyperbolic* distribution (GH), is given by

$$f_{\mathrm{GH}}(x;\lambda,\alpha,\beta,\delta,\mu) = \frac{(\alpha^2-\beta^2)^{\lambda/2}\,(\delta^2+(x-\mu)^2)^{(\lambda-\frac{1}{2})/2}}{\sqrt{2\pi}\,\alpha^{\lambda-1/2}\,\delta^{\lambda}\,K_\lambda(\delta\sqrt{\alpha^2-\beta^2})} \times K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})\,\mathrm{e}^{\beta(x-\mu)} \quad \text{for} \quad x\in\mathbb{R}, \tag{4}$$

where K_λ is the modified Bessel function of the third kind. The permitted values of the parameters are as follows:

$$\lambda,\beta,\mu\in\mathbb{R} \quad \text{with} \quad \begin{array}{llll} \delta\geq 0 & \text{and} & |\beta|<\alpha & \text{if} \quad \lambda>0, \\ \delta>0 & \text{and} & |\beta|<\alpha & \text{if} \quad \lambda=0, \\ \delta>0 & \text{and} & |\beta|\leq\alpha & \text{if} \quad \lambda<0. \end{array} \tag{5}$$

When $\alpha=\beta=0$ with $\lambda<0$, (4) should be interpreted as the limit of (4) as $\alpha\downarrow 0$ with $\beta=0$. The representation (21) of the Bessel function K_λ in Appendix A easily gives

$$\lim_{x\downarrow 0}\frac{K_\lambda(x)}{x^{|\lambda|}} = 2^{|\lambda|-1}\,\Gamma(|\lambda|) \quad \text{for} \quad \lambda\neq 0, \tag{6}$$

where Γ is the Gamma function. Hence we have the following probability density function:

$$f_{\mathrm{GH}}(x;\lambda,0,0,\delta,\mu) = \frac{\Gamma(1/2-\lambda)}{\sqrt{\pi}\,\delta\,\Gamma(-\lambda)}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{\lambda-1/2} \quad \text{for} \quad x\in\mathbb{R} \tag{7}$$

(see also [10]). Of course, this is Student's $t(-2\lambda)$ distribution, shifted $\mu\in\mathbb{R}$ to the right, and rescaled with the factor $\delta>0$.

The distribution function of the t distribution can be obtained from equation (10) below.

Information on various properties of the GH distributionas can be found in, for example, [10] and [17]. We have listed a few properties, that we appeal to, in Appendix A.

2.3.4 Modified Generalized Hyperbolic Distribution

We will use a *Modified Generalized Hyperbolic* distribution (MGH), the probability density function of which is given by, with obvious notation,

$$f_{\mathrm{MGH}}(x;\lambda,\alpha,\beta,\delta,\mu,\rho) = \frac{(\alpha^2-\beta^2)^{\lambda/2}\,(\delta^2+|x-\mu|^{2/\rho})^{(\lambda-\frac{1}{2})/2}}{\sqrt{2\pi}\,\alpha^{\lambda-1/2}\,\delta^{\lambda}\,K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})}$$
$$\times\frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2+|x-\mu|^{2/\rho}})\,\mathrm{e}^{\beta\,\mathrm{sign}(x-\mu)\,|x-\mu|^{1/\rho}}}{\rho\,\mathbf{E}\{|\mathrm{GH}(\lambda,\alpha,\beta,\delta,\mu)|^{\rho-1}\}}\quad\text{for}\quad x\in\mathbb{R}. \tag{8}$$

The permitted values of the parameters are

$$\rho>0\quad\text{and}\quad\lambda,\beta,\mu\in\mathbb{R}\quad\text{with}\quad\begin{array}{c}\delta>0\quad\text{and}\quad|\beta|<\alpha\\ \text{or}\\ \delta=0\quad\text{and}\quad|\beta|<\alpha\quad\text{and}\quad\lambda>0,\\ \text{or}\\ \delta>0\quad\text{and}\quad|\beta|=\alpha\quad\text{and}\quad 2\lambda+\rho<1.\end{array} \tag{9}$$

Using (6), we get as in (7),

$$f_{\mathrm{MGH}}(x;\lambda,0,0,\delta,\mu,\rho) = \frac{\Gamma(\frac{1}{2}-\lambda)\,(1+|(x-\mu)/\delta|^{2/\rho})^{\lambda-1/2}}{2\,\delta\,\Gamma(1+\frac{1}{2}\rho)\,\Gamma(\frac{1}{2}-\frac{1}{2}\rho-\lambda)}\quad\text{for}\quad x\in\mathbb{R}. \tag{10}$$

The corresponding distribution function is given by

$$F_{\mathrm{MGH}}(x;\lambda,0,0,\delta,\mu,\rho)$$
$$=\begin{cases}\dfrac{1}{2}-\dfrac{y\,\Gamma(\frac{1}{2}-\lambda)\,{}_2F_1(\frac{1}{2}-\lambda,\frac{1}{2}\rho,1+\frac{1}{2}\rho;-(-x/\delta)^{2/\rho})}{2\,\delta\,\Gamma(1+\frac{1}{2}\rho)\,\Gamma(\frac{1}{2}-\lambda-\frac{1}{2}\rho)} & \text{for}\quad x<0,\\[2ex] \dfrac{1}{2}+\dfrac{y\,\Gamma(\frac{1}{2}-\lambda)\,{}_2F_1(\frac{1}{2}-\lambda,\frac{1}{2}\rho,1+\frac{1}{2}\rho;-(x/\delta)^{2/\rho})}{2\,\delta\,\Gamma(1+\frac{1}{2}\rho)\,\Gamma(\frac{1}{2}-\lambda-\frac{1}{2}\rho)} & \text{for}\quad x>0,\end{cases} \tag{11}$$

where ${}_2F_1$ is the hypergeometric function.

2.4 Devolatilization

Several times, we assume that the basic model for the price process S_t of an asset, is described by the following Lévy process based model (see e.g. [12]):

$$S_t = S_0\,\mathrm{e}^{\mu t+\sigma_t L_t}\quad\text{for}\quad t\geq 0.$$

Here L_t is the Lévy process, while σ_t is a stochastic process modelling the volatility of the price process. We know that log-returns cannot be assume

IID without questioning. Following [4], we can see that the dependency structure can be captured by a stochastic volatility.

Assuming that σ_t moves slowly compared to L_t, we have, for small increments,

$$\log S_t - \log S_{t-1} = \mu t + \sigma_t L_t - \mu (t-1) - \sigma_{t-1} L_{t-1} \approx \mu + \sigma_t \Delta L_t.$$

Here $\Delta L_t = L_t - L_{t-1}$ are the IID increments of the Lévy process. Thus we can fit an infinitely divisible distribution to the *devolatilized* log-returns ΔL_t of the price process S_t, by estimating μ and σ_t, and calculate $\Delta L_t = (X_t - \mu)/\sigma_t$.

In this master's thesis, where the aim is to compare the fit of the GH distribution to different data sets given by already written thesises, we have followed the work processes done by the different authors. In some cases we have not devolatilized data at all and in some cases we have had to adopt different volatility models used in that particular case.

2.5 Goodness of Fit

One important task, when estimating the parameters for an distribution, is to evaluate how good the fit is to the data set. There are many ways to do this, and the choice depends on what is the main purpose of the fit. If modeling risk, it might be most interesting to get a good fit of the tails. In other cases, it might be more to the point to measure the fit of the whole data set.

Since, in the field of mathematical finance, there is often at least some focus on the tails of distributions, for example, in connection with evaluation of risk, we have decided to pay special attention to the fit of tails. Standard graphical procedures, such as PP-plots and QQ-plots are good tools to get an idea of the quality of the fit (see e.g. [19]), but in order to get a tool to compare different fits, we want to measure the distance between the observed empirical distribution of the data set (see below), and the parametric distribution fitted to it.

One well-known statistic is the *Kolmogorov-Smirnov* distance (K-S) (see e.g. [20]), given by

$$D = \max_{x \in \mathbb{R}} |F_{\text{emp}}(x) - F_{\text{fit}}(x)|. \tag{12}$$

Here F_{fit} is the fitted distribution function, while

$$F_{\text{emp}}(x) = \frac{\text{number of observations} \leq x}{\text{total number of observations}}$$

is observed *empirical* distribution function. To calculate the K-S distance, we use the following formula from [1]:

$$\max_{x \in \mathbb{R}} |F_{\text{emp}}(x) - F_{\text{est}}(x)| = \max_{i \leq n} \Big(\max \{ |(i-1)/n - F(X_{(i)})|, |i/n - F(X_{(i)})| \} \Big), \tag{13}$$

where $X_{(1)} \leq \ldots \leq X_{(n)}$ is the ordered data set. As can be seen, this is an objective absolute measure of the actual maximum distance over the whole distribution. The problem is that as the values far out in the tails are small the use of (12) will not give an indication of how well the tails fit. The contribution from the values of (12) in the tails will be so relatively small that they are barely noticable compared to the relatively larger values from the parts where the distribution function takes larger values.

So, if your interest lies rather in the tails than in the middle part the K-S distance are not a very suitable tool.

Another measure of the distance, that do not surpress the importance of the tails, are the *Kuiper* statistic (K), which is given by (see [18])

$$K = \max_{x \in [0,1]} (\hat{F}_U(x) - x) + \max_{x \in [0,1]} (x - \hat{F}_U(x)). \tag{14}$$

Here $\hat{F}_U$ denotes the empirical distribution function of $(\hat{F}_t(\Delta X_t))_{t=1,\ldots,T} = (U_t)_{t=1,\ldots,T}$, where $\hat{F}$ is the estimated distribution function. We can get an asymptotic large sample p-value for this test, given by by

$$p = 2 \sum_{j=1}^{\infty} (4j^2\lambda^2 - 1)\, \mathrm{e}^{-2j^2\lambda^2},$$

with

$$\lambda = K \left(\sqrt{T} + 0.155 + \frac{0.24}{\sqrt{T}} \right).$$

Yet another distance measure statistic, used by e.g. [17], was introduced by Anderson and Darling [2] in 1954,. This statistic pays even more attention to the tails emphasizing values by a heavy weight. The *Anderson & Darling* statistic (A-D) is given by

$$AD = \max_{x \in \mathbb{R}} \frac{|F_{\mathrm{emp}}(x) - F_{\mathrm{est}}(x)|}{\sqrt{F_{\mathrm{est}}(x)(1 - F_{\mathrm{est}}(x))}}. \tag{15}$$

Here the values of the tails are blown up by the denominator, which the A-D statistic useful to measure goodness of fit in the tails.

Which of the above statistics to use, is, of course, a matter of what application one has in focus. Therefore, we give the values of all three statistics for our fittings. We give the p-value for the K statistic.

2.6 Numerical Estimation

We just want to mension a little about how we have estimated the parameters. We have use two well known computer programs: Matlab and Mathematica. Matlab is faster and numerically stronger while Mathematica's analytical properties is of great importance when we want to evaluate the function expressions and adjust them for different properties. Through the whole period we have had great problem to get good estimations and have had to spend a lot of time on the computational programming. When estimating GH parameters with Maximum Likeklihood methods it is crucial to have good start values, and to achieve this we have often combined the different skills of Mathematica and Matlab. We have put some of the program code that we thin it could be of the readers value in the appendices and hopefully this would help you understand the procedures. When estimating the parameters with Maximum Likelihood in Matlab we have mainly used the following head program:

```
tmp=importsiemens;

opts=optimset('MaxFunEvals',10000,'MaxIter',
           2000,'TolX',1e-6,'TolFun',1e-6,'Display','iter');

%The loglikelihood function is supplied in the m-file loglik.
%The result of the minimization is put in res-object.
%x0 is a vector of start values

[res.par_ests,res.funval,res.terminated,res.details]=
    fminsearch('loglik',x0,opts,v);

[res.par_ests,res.funval,res.terminated,res.details]=
    fminunc('loglik',res.par_ests,opts,v);
```

3 Fitting of Empirical Data to Distributions

3.1 ABB Stocks

In his master's thesis [24], Johan Tykesson examines different aspects of Lévy processes in mathematical finance. He starts with a discussion of alternative processes that could model the log-returns better than the Brownian motion. Concentrating on the distributional aspects of Johan's work, he produce a comparison of the fit of the Normal and the NIG and Meixner distributions, introduced below, to the log-returns of financial data.

The *Normal Inverse Gaussian* distribution (NIG) is the special case $\lambda = -1/2$ of GH (4).

The *Meixner* distribution has the following probability density function

$$f_{\mathrm{MXN}}(x;a,b,d,m) = \frac{(2\cos(\frac{b}{2}))^{2d}}{2a\pi\,\Gamma(2d)}\,\mathrm{e}^{b(x-m)/a}\left|\Gamma\!\left(d+\frac{i(x-m)}{a}\right)\right|^2 \quad \text{for} \quad x \in \mathbb{R} \tag{16}$$

(see [21]). Here the parameters yield to the restrictions $a > 0$, $-\pi < b < \pi$, $d > 0$ and $m \in \mathbb{R}$.

One of the data sets examined by Johan is the ABB stock listed daily from September 19, 2000 to September 17, 2002. He found that both the NIG and the Meixner distribution outperformed the Normal distribution. The difference between NIG and Meixner was small, with the NIG slighty better.

We obtained ABB data for the period from [28], and estimated the parameters for NIG, Meixner, Normal and GH distributions. The result are displayed in Table 1.

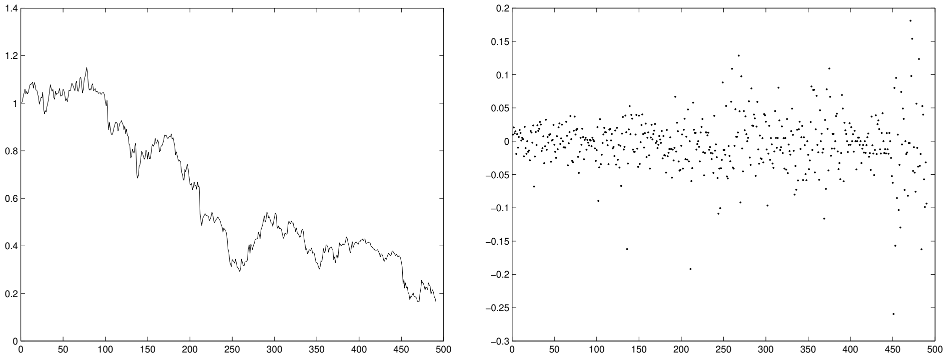


Figure 1: *ABB data. The stock ratings (left) and the corresponding log-returns (right)*

In Table 2, we give the K statistics for the fits, together with their corresponding p-values. We also give the A-D statistics for the fits. Clearly,

	α	β	δ	μ	λ
GH	11.50	-1.183	0.03508	-0.001652	-0.9206
NIG	17.03	-1.262	0.02883	-0.001564	$-1/2$
Meixner	a	b	d	m	
	0.1276	-0.1646	0.2065	-0.001531	
Normal	μ	σ			
	0.003710	0.04158			

Table 1: *Estimated parameters for ABB log-returns*

GH, NIG and Meixner all outperfom the Normal. The GH distribution performed best, but the difference to NIG is not that large.

	A-D	K	K p-value
GH	0.0655	0.0450	0.818
NIG	0.0772	0.0460	0.767
Meixner	0.0982	0.0491	0.347
Normal	67.6	0.175	0.000

Table 2: *A-D and K statistics for fitting to devolatilized ABB log-returns.*

As have been mentioned, NIG is a special case of GH ($\lambda = -1/2$), and therefore cannot perform better than GH. In view of the considerable extra time needed to estimate parameters for GH compared to NIG, the small difference in the Goodness of Fit could make one queastion if the extra work really is worthwile.

The Meixner also fits the data well. But as it is harder to deal with numerically than the NIG, due to the imaginary unit i in the Gamma function in (16), it does not seem worthwile to use.

Notice that one argument to stay with the GH distribution, including special cases thereof, is that this gives freedom, depending on the actual application, to choose between better fit or less time consumtion for computations, and still stay with a one single computer program.

3.2 LF Group, Insurance Costs

An important field of interest is the insurance industry and the importance of accurate measures of risk. One example is the distribution of the size and the frequency of natural disasters which is examined by Cilla Persson in her master's thesis [16]. Cilla is creating a model for the distribution of costs, emanating from natural disasters, within the LF Insurance Group. To do this, she develops a model where the size of the disasters is distributed using a *Peaks over Threshold* model resulting in a Generalized Pareto distriubtion and the frequency is estimated by a Poisson process. In our case, it would be interesting to see if the size of the insurance costs due to natural disaster could be better modeled using the GH distribution. By Cillas kindness, we got the data she used in her analysis. It is a dataset containing all the claims for the LF Insurance Group exceeding 500 000 SEK, from 1990-1999.

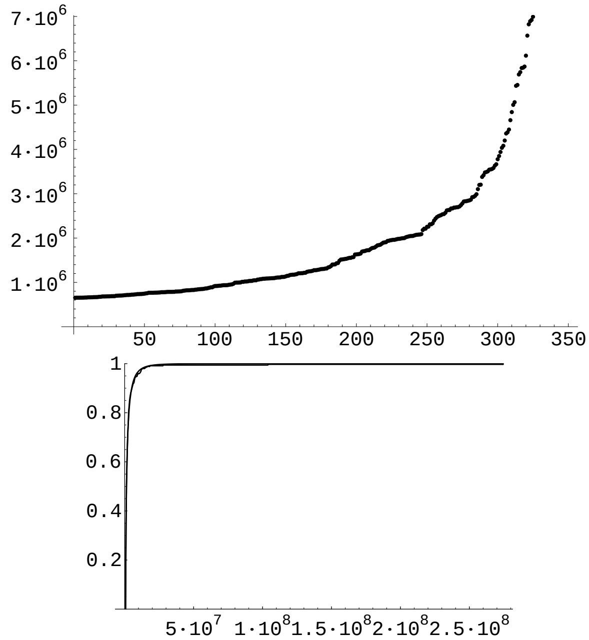


Figure 2: *LF-data (left) and the fitted GH-distribution to the empirical distribution function (right)*

The data consisted of 450 claims and these were first corrected for inflation by use of consumer price index obtained from [26], where all data was discounted to the value of the last year of data (1999). A problem was that the threshold 500 000 sek also had to be corrected for inflation and this means that, for example, a claim 1999 on 520 000 SEK would av been worth say 480 000 sek 1990 when corrected for inflation and then it would not have been part of the statistics for that year. This problem was taken care of by discounting the threshold from the earliest year (1990) to the value of 1999 giving us a maximum threshold of 652261 SEK and then sorting out all discounted claims exceeding this threshold. This give us a dataset of 348 claims that would be comparable.

Next problem is now that these data is not independent due to the fact that several claims originates from the same natural disaster. One of the problems araising from this lack of independency is the possibility to use Maximum Likelihood methods to estimate the parametes of the distribution. In her thesis Cilla solved this problem by grouping together the data emanating from the same disaster. Unfortunately we did not have access to the kind of information making such a grouping possible why we had to find out another way to estimate the parameters. As the goal of the estimation is to fit the assumed distribution as well as possible to the historical data, a natural estimation method would be to minimize the distance between the empirical distribution function of the data set and the distribution function itself. As our dataset consists of extremes we are already situated in the tail and a distance measure that pays special attention to the tails would give as a measure of the fit of the tail of the tail, and therefore it is natural to use (13).

It is then time to start the estimation procedure. To be able to make a valid comparison we first have to estimate the parameters for the Generalized Pareto distribution (2). The advantage of minimizing the distance is that the result of the minimization procedure and the goodness of fit measure will be the same, so we can directly see how well the estimation is performed. The result of the estimations is given in Table 3

Now, the GH distribution is defined on the whole real line, so that, with obvious notation, we have to consider the conditional distribution

$$\mathbf{P}\{X > x | X > u\} = \frac{1 - F_{\mathrm{GH}}(x)}{1 - F_{\mathrm{GH}}(u)} \quad \text{for} \quad x > u.$$

During the estimation we met two problems. The first was due to the size of the specific data in the dataset. As we were examinating a really small quantile of the distribution it was hard for the optimization procedure in the computer software to converge to reasonable values. This problem was solved by rescaling some of the parameters of the distribution. The other problem was that we reached a limit of GH where the parameters are permitted. Using the limit distribution defined in equation 3.10 in[10] and rescaling it and defining it only for $x > 0$, we end up with a distribution with probability density function given by

$$f(x;\delta,\beta,\xi,\lambda,\sigma) = \frac{\delta^{2(\lambda/\xi-1/\beta)}\,\xi^{1/\beta}\,\Gamma(\lambda/\xi)}{\sigma\,\Gamma(1+1/\beta)\,\Gamma(\lambda/\xi-1/\beta)}\left(\delta^2+\xi\Big(\frac{x}{\sigma}\Big)^{\beta}\right)^{-\lambda/\xi} \tag{17}$$

for $x > 0$. This limit distribution can more or less be seen as a limit case of the original GH distributions, when the original parameters $\alpha = \beta \to 0$ (see [10]), where β, ξ, σ are new parameters introduced to scale the distribution in order to obtain good fits to the data set. This density function has the advantage that it is possible to integrate analytically, so that we can obtain an analytic expression for the distribution function, which in turn is of great value when minimizing the K-S distance. The distribution function is then given by

$$\mathbf{P}\{X > x | X > u\} = \frac{\Gamma(\lambda/\xi)\,\mathrm{B}(-\xi\,((x-u)/\sigma)^{\beta}/\delta^2, 1/\beta, 1-\lambda/\xi)}{(-1)^{1/\beta}\,\Gamma(1/\beta)\,\Gamma(\lambda/\xi-1/\beta)} \tag{18}$$

for $x > u$, where B is the Beta function.

The minimizing procedure was this time carried out using Mathematica, and the corresponding notebook is displayed in Appendix C. The results can be seen in Table 3.

modified GH	λ	δ	β	ξ	σ	K-S distance
	1.489	0.9326	0.4590	0.07329	0.6423	0.0260
GP	ξ	σ				
	0.9871	0.9959				0.03446

Table 3: *The GP parameter σ is scaled with the value of the threshold u, so that the correct value for the minimized distance is $\sigma \cdot u$.*

3.3 Siemens

The issue at hand for Mattias Bengtsson and Viktor Olsbo, in their master's thesis [4], is to measure market risk using *Value at Risk* (VaR). They are doing this in several different ways, using several volatility models and a couple of underlying distribution assumptions. In their very ambitious work they fit the models to three different data sets. One of these sets are the Siemens stock prices from Mars 29, 1995 to May 23, 2002. We decided to work with this set and obtained it from [28], see Figure 3.

When dealing with VaR the fit in the lower part of the data set is the most important, and Mattias och Viktor examines four different distribution assumptions: the classic Normal distribution, extreme value theory with the General Pareto distribution (GP), the Hyperbolic distribution (HYP) (sublcass of GH with $\lambda = 1$), and the *Pearson VII* distribution. When considering the lower tail, they found that both HYP, GP and Pearson fits the data very well. Our aim is now to produce a fit of GH to the data and see how well it performs.

The *Pearson VII* distribution may need a further explanation. This distribution is one-sided, with probability density function given by

$$f_{\text{Pearson}}(x; m, c) = \frac{2\Gamma(m)}{c\Gamma(m - \frac{1}{2})\sqrt{\pi}} \left(1 + \left(\frac{x}{c}\right)^2\right)^{-m} \quad \text{for} \quad x > 0.$$

Here $m > 1/2$ is a shape parameter and $c > 0$ a scale parameter. Note that this density function is that of a one-sided t distribution, with $\nu = 2m - 1$ degrees of freedom, and scaled with a factor c.

Before estimating the parameters for GH, we have to devolatize the log-returns to make them IID. Mattias and Viktor use four different volatility

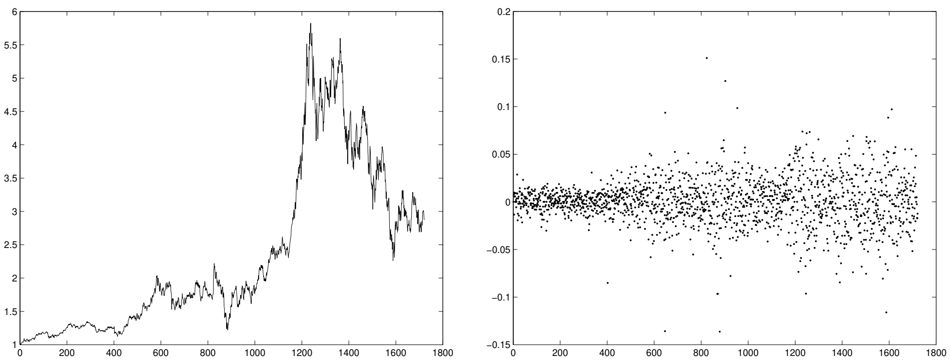


Figure 3: *Siemens stock data. The stock ratings (left) and the corresponding log-returns (right)*

models, and found that the nonparametric *Nadarya-Watson* (see e.g. [23], [15] and [25]) model performed best. The volatility is then estimated by

$$\hat{\sigma}_t^2 = \frac{\sum_{i=t-n}^{t-1} K_h(i-t)Y_i^2}{\sum_{i=t-n}^{t-1} K_h(i-t)}, \tag{19}$$

where $Y_t = X_t - \hat{\mu}$, $\hat{\mu} = \frac{1}{t}\sum_{j=1}^{t} X_j$ and K_h is a scaled normal kernel $K_h(x) = \frac{1}{h}\mathrm{e}^{-x^2/h}$. Apparently, the volatility estimation is based on the last n values for every log-return X_t, and next task is to find how many (in our case with daily closing prices) *days* this would be, as well as to find an apropriate value on the parameter h, to achieve the best devolatilization possible. To do this we used a pseudo-scientific method, where the Ljung-Box statistic were checked for which values of h and n the log-returns are least correlated (see [14]). Using this statistic as described in Appendix B, we found that for $n = 14$ and $h = 50$ the dataset $(X_t)_{t=15}^{n}$ is devolatilized in a way that makes the dependency structure of the log-returns sufficiently low.

In order to perform the comparison the next step is to fit the probability distributions mentioned above to the devolatilized dataset. After that, we analyze the goodness of fit using both the K and A-D statistics. As have been mentioned, the fit of the lower taile of the data that is most interesting when dealing with VaR, so the Goodness of Fit of the lower tails have also been analyzed.

The parameter estimations kan be seen in Table 4, and the test statistics in Table 5.

As the p-values here are high for all distributions, the comparison does not really give any differentiating results. In addition, we have not been consequent in our work. GH, HYP and Normal are fitted to the whole dataset. Pearson is fitted to the half set and GP for all values above the

	α	β	δ	μ	λ
GH	0.06432	-0.06432	2.442	0.08736	-3.397
HYP	1.603	-0.06983	0.9875	0.09436	1
Pearson	m	c			
	3.214	2.118			
GP	ξ	σ	u		
	$8 \cdot 10^{-14}$	0.7548	1.400		
Normal	μ	σ			
	0.006320	1.160			

Table 4: *Estimated parameters for Siemens log-returns*

	D-A	Kuiper	p-value
GH	0.0939	0.0105	0.999
HYP	0.160	0.0091	0.999
Pearson	0.0534	0.0254	0.992
GP	0.0420	0.0131	0.999
Normal	0.740	0.0244	0.807

Table 5: *A-D and Kuiper test statistics for lower tails of Siemens log-returns.*

threshold 1.4 and as we see in Table 5 the results are a little strange. We want to refine this work and as HYP is a special case of GH and Normal is outperformed by the others (according to the work of Mattias Bengtsson and Viktor Olsbo), we concentrate on an (as far as possible) unbiased comparison between, GH, Pearson and GP. To do this right, we first have to calculate a conditional distribution that dependend on the threshold. As we already have done this in the preceeding case with the LF Insurance Group we use the expression (18) for the conditional GH distribution, and we also use a conditional Pearson distribution, where we move the distribution, so that it live on the interval $(1.4, \infty)$.

As our primary goal is to minimize the K-S distance to the empirical distribution, we estimate the parameters by minimizing (13). We also use the possibility to produce the fit for both the lower tail and the higer tail to get stronger results. The result is presented in Table 6, and from this we can see that the modified conditional GH distribution outperforms both the GP and the conditional Pearson.

From the table we can also see that the parameter ξ takes a very small value for both positive and negative tails. What does this mean? The GP distribution function is given by

$$F(x|x > u) = 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi} \quad \text{for} \quad x > u,$$

GH	λ	α	β	δ	σ	distance
positive	-1.562	0.4004	0.9275	0.6091	1.700	0.04006
negative	-0.6048	0.3427	1.263	0.2066	1.998	0.03094
Pearson	m	σ				
positive	1.508	0.6724				0.04668
negative	3.553	1.789				0.04942
GP	ξ	σ				
positive	$13 \cdot 10^{-8}$	0.4070				0.04599
negative	$13 \cdot 10^{-8}$	0.5399				0.07240

Table 6: *Estimation parameters and minimal distance for the positive and the negative tail of the Siemens stock log-returns.*

and thus have polynomial tails. We know that

$$\lim_{n \to 0} (1+n)^{1/n} = \mathrm{e},$$

and as $\xi \downarrow 0$, this indicates that the tails should decrease exponentially, which is the caracteristic of GH.

From this we conclude that GH with its great flexibility could be adjusted to be used even in this kind of conditional distribution with great sucess. Thus, hitherto we have seen that GH is a better choice when it comes to studies both of the whole distribution and when special attention is given to the tails. The Mathematica programs solving this problem can be seen in Appendix D. When choosing distribution assumption it is always a question of time against precision and it should be clear that though it seem simple, we have worked a long time and tried a lot of difference methods to estimate the parameters of GH, with more or less sucess before this terminal adjusted GH distribution was developed and sucessfully determinated.

3.4 The Spread SHB-SEB

In her master's thesis [29], Katarina Åselius has developed a time series model for the spread SHB-SEB, where she uses the t distribution as the underlying distribution assumption. She motivates why she studies a spread and why she chooses these two stocks, but we are satisfied with the fact that spreads are widely used in financial business and, as a consequense of that, are of great interest.

A spread is the relative value of two stocks making the actual spread calculated as SHB/SEB. By the very kindness of Katarina Åselius, we could use her dataset plus extra data until april this year, giving us a dataset starting 1993-10-27 consisting of 2726 trading days. Following the work of Katarina the dataset was first adjusted to achieve stationarity. This was

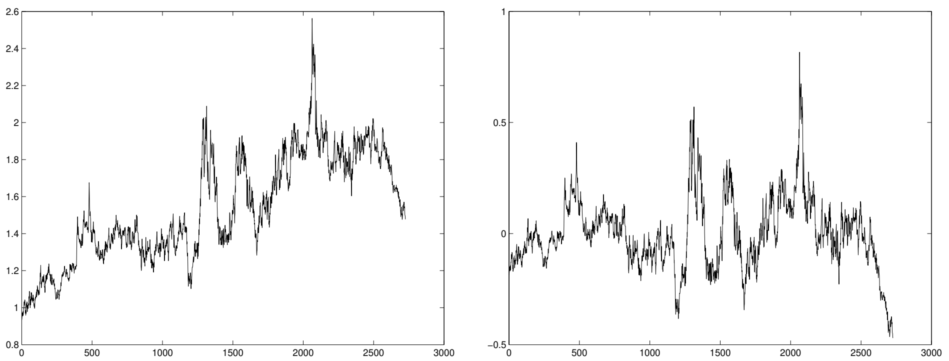


Figure 4: *The spread SHB-SEB (left), and the same spread adjusted for the linear trend (right).*

done removing a linear trend $y = 1.12033 + 0.0003034x$, and this adjusted dataset can be seen in Figure 4

The time series model that Katarina developed is a so called ARARCH process (see [6]), where the volatility is catched in the ARCH part, and the distribution assumption is a modified t distribution, scaled to have unit variance:

$$X_t = \phi X_{t-1} + Z_t, \quad \text{where} \quad Z_t = \sqrt{h_t} v_t, \tag{20}$$

and

$$v_t \sim \text{ modified } t(\nu) \quad \text{and} \quad h_t = \xi + \alpha Z_{t-1}^2.$$

To estimate the parameters of this model from our dataset we use Maximum Likelihood method, and to use it properly we have to calculate the likelihood function for the devolatilized data. The only problem with this is that as we assume that the data are not Normal distributed we have to relocate and rescale the distribution so that the expectation becomes zero and the variance one. The density for the ARARCH$(1,1)$-devolatilized modified t distributed data is then in our case (for a definition of the t distribution see e.g. [20])

$$f(x_t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\,(\nu-2)\,h_t}\,\Gamma(\nu/2)} \left(1 + \frac{(x_t - \phi x_{t-1})^2}{h_t(\nu-2)}\right)^{-(\nu+1)/2}.$$

The estimation of the parameters for this case, as well as for the Normal distribution, were carried out in Mathematica, and the program can be found in Appendix E. Now, of course we want to examine how well the GH distribution will perform in comparison to the t distribution, but this was not easily done. There were a lot of initial probabilities that we hade to overcome. The first problem was due to the fact that ARARCH$(1,1)$ model with GH-distribution assumed result in 3+5 paramaters to estimate and this in addition to the relatively large dataset (2726 data) made it nearly impossible for us to get convergence. Looking at a few data at the time we got preliminary data that indicated that we were dealing with the same limiting case as for the LF insurance data described above. Therefore we decided to use the distribution given by (17), and scale it to get expectation zero and variance one. Then we succeeded to estimate the parameters using Maximum Likelihood estimation in Mathematica. These calculations can be found in Appendix F. The parameter estimation for this dataset is presented in Table 7, where the series model parameters [see (20)] are presented next to the Maximum Likelihood score, featuring as a measure on the Goodness of Fit.

	α	ϕ	ξ	ML-score
$N(0,1)$	0.2444	0.9867	0.0006646	5812
modified t	0.3546	0.9878	0.0006461	5940
MGH	0.3412	0.9920	0.0006068	5948

Table 7: *Estimated ARARCH series parameters for the different distribution.*

As we can see, the performance of the modified GH distribution is better than both Normal and the t distribution which we expected as the two later are possible to derive as limiting cases from GH. Notice that according to their Maximum Likelihood values, the difference between the t distribution and GH are not that large. This could be due to the fact that the modified GH in this case will take the numerical form in the nearness of the t distribution. In this case it probably would not be worth using the GH instead of the t distribution as it is much harder to estimate, with its many parameters. However, this is just one case and GH would possibly perform evidently better in other corresponding cases, which may have a moore tricky structure which would be captured better by the flexiblitiy of GH. The distribution parameters of the Maximum Likelihood estimation can be found, for the $t(\nu)$ distribution in Appendix E, and for GH in Appendix F.

3.5 Portfolio Analysis

To study the development in time of a portfolio of stocks is of more practical interest than trying to model the series of one singel stock. Anna Widerberg studied a portfolio of swedish stocks in her master's thesis, see [27], where her aim was to calculate VaR. It is natural to decrease the risk of an investement by use of a portfolio instead of singel stocks. She analyze the data using Extreme Value Theory, and then fitted the a GP distribution [see (2) for a definition] to the log-returns.

The portfolio that Anna study, and that we by her kindness could use in our study, consists of 15 swedish stocks, see Table 8, January 29, 2001 to

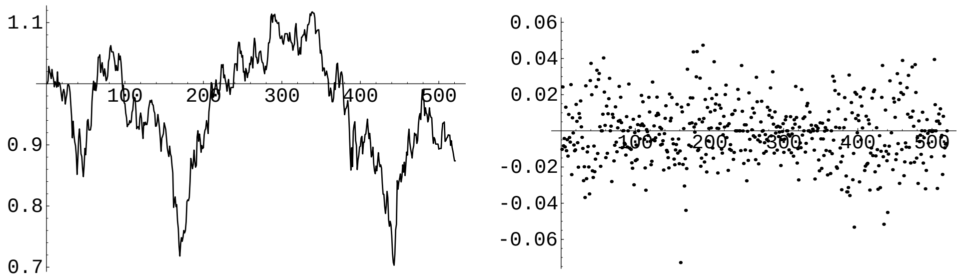


Figure 5: *The portfolio time series (left) and the corresponding log-returns (right).*

Assa Abloy	Atlas Copco	Electrolux
Ericsson	Gunnebo	Kinnevik
Lindex	Nobel Biocare	Nokia
Sandvik	SEB C	SHB B
SKF	SSAB	Volvo

Table 8: *The 15 stocks in the portfolio analyzed by Anna Widerberg.*

January 27, 2003. The stocks in the portfolio were initially equally weighted, i.e. each stock corresponded to a 1/15 of the total value of the portfolio, but due to a constraint that no stock alone should consitute more than 10% of the total value changed the portfolio values a little along the way. How this correction was done can be seen in Appendix G.

Our aim in this case is to repeat the work of Anna and estimate the parameters for the Generalized Pareto distribution and then try to fit GH to the data and see if we could get some improved results. Starting out with the estimation of the parameters for the GP we have to decide on a value for the threshold and we do this by studying the mean excess function and try to find the point where this function turns linear. The mean excess function can be seen in Appendix G, and after studying the graph we decided on a threshold of the value 0.2. As we are interested to achieve a fit that is as good as possible, the natural method to estimate the parameters would to minimize the K-S distance between the empirical distribution and the estimated distribution, using (13).

To estimate the parameters for GH, we use the same method and as for GP, and the Mathematica program code for this can be seen in Appendix H. The result of the estimation procedure can be seen in Table 9. Notice that the GH distribution failed to perform better than the GP distribution. This is peculiar, as GP is a limiting case of GH. As this was our last case study and as we were running out of time we didn't have the possibility to analyze this suprising result more deeply.

GH	λ	α	β	δ	μ	distance
	-4.195	2.700	-0.8574	0.01870	-0.05450	0.0460
GP	ξ	σ				
	0.1241	0.009300				0.0459

Table 9: *The estimated parameters and the minimal K-S distance for the portfolio data.*

In this case there is aparantly no point trying to fit the GH distribution instead of the GP, as the calculations are much more time consuming and the results are worse, but as we said this is no concluding remark. The result is indeed strange and several more studies of different portfolios would be necessary to get a more general conclusion. It would also be of use to fit some limiting cases of GH to see why we got this result, but maybe eventhough we will get a better fit for GH than for the GP, the difference may be so small that it is not worth the effort. We now leave this to be investigated by the interested reader.

4 Conclusion

As we have seen, the GH distribution seem to outperform other parametric models, when it comes to fitting to historical data sets, from the field of mathematical finance. It would then, for example, be of great interest, to use the GH distribution in a modified Bachelier-Samuelson model.

The main problem that we have met, with the GH distribution, is that it is tricky to use, as its many parameters make it numerically demanding. On the other hand, we have shown that, in some cases, it is possible to use simplified versions of the GH distribution, to get better numerical properties.

If one wants to use the GH distribution in a certain field, it could be advantageous to develop a case-specific variant of it. In fact, possibly, one could develop a library of various modifications of the GH distribution, like, for example, the MGH distribution we used for applications to extreme value theory.

Of course, in numerical applications, one often faces a trade-off between time and precision. Our experience is, that if it is important with high precision, then it could be very fruitful to consider the GH or MGH distribution. On the other hand, if the focus primarily lies on time, and a quick fit is needed, then GH and MGH might be too slow. However, special cases of GH, like, for example, the NIG distribution, might be considered. In addition, one might consider using case-specific simplifications of GH and MGH.

We have to admit, that numerical and computational problems when estimating the parameters took a lot of the our time available, so that the number of datasets that we could study had to be reduced to about half the number originally intended. Because of this, our results might be somewhat less convincing than they could have been. Nevertheless, our experience from the worh with this master's thesis, is that the GH distribution is very versatile and useful, and that it should be worthwile to consider the use of it in various fields of application.

References

[1] P. Albin, *Föreläsningar och laborationer i statistisk databehandling*, Departement of Mathematical Statistics, Gothenburg, 2003 [In Swedish].

[2] T.W. Anderson, D.A. Darling, *A test of goodness of fit*, Journal of the American Statistical Association, 49, pp. 765-769, 1954.

[3] O.E. Barndorff-Nielsen, *Exponentially decreasing distributions for the logarithm of particle size*, Proceedings of the Royal Society London, 353, pp. 401-419, 1977.

[4] M. Bengtsson, V. Olsbo, *Value at Risk Using Stochastic Volatility Models*, Master's thesis, Departement of Mathematical Statistics, Gothenburg, 2003.

[5] C. Borell, *Finansiella derivat och stokastisk analys*, Departement of Mathematics, Gothenburg, 2002 [In Swedish].

[6] M. Borkovec, *Asymptotic behavious of the sample autocovariance function of $AR(1)$ process with $ARCH(1)$ errors*, Bernoulli 7, pp. 847-872, 2001.

[7] E. Brodin, *On the logreturns of empirical financial data*, Master's thesis, Departement of Mathematical Statistics, Gothenburg, 2002.

[8] P. Burns, *Robustness of the Ljung-box Test and its Rank Equivavlent*, Preprint, 2002.

[9] S. Coles, *An Introduction to Statistical Modelling of Extreme Values*, Springer, Berlin, 2001.

[10] E. Eberlein, E.A. von Hammerstein, *Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting cases and Approximation of Processes*, Preprint, University of Freiburg, 2002.

[11] E. Eberlein, U. Keller, *Hyperbolic distribution in finance*, Bernoulli 1, pp. 281-299, 1995.

[12] E. Eberlein, J. Kallsen, J. Kristen, *Risk Management Based on Stochastic Volatility*, FDM Preprint 72, University of Freiburg, 2001.

[13] G.M. Ljung, G.E.P. Box, *On A Measure of Lack of Fit in Time Series Models'*, Biometrika, 65, pp. 297-303, 1978.

[14] T.C. Mills, *The Econometric Modelling of Financial Time Series, 2^{nd} Ed.*, Cambridge University Press, 1999.

[15] E.A. Nadaraya, *On estimating regression*, Theory of Probability and its Applications, 9, pp. 141-142, 1964.

[16] C. Persson, *Distributing Reinsurance Costs within the LF Insurance Group*, Master's thesis, Departement of Mathematical Statistics, Gothenburg, 2001.

[17] K. Prause, *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*, Ph.D. thesis, University of Freiburg, 1999.

[18] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.

[19] J.A. Rice, *Mathematical Statistics and Data Analysis,* 2^{nd} *Ed.*, Duxbury Press, 1995.

[20] L. Råde, B. Westergren, *Mathematics Handbook for Science and Engineering BETA*, Studentlitteratur, Lund, 1995.

[21] W. Shoutens, *Meixner Processes in Finance*, EURANDOM Report 2001-2002, EURANDOM, Eindhoven, 2001.

[22] W. Shoutens, *Lévy Processes in Finance, Pricing Financial Derivatives*, Wiley, 2003.

[23] C. Starica, H. Drees, *A simple non-stationary model for stock returns*, Preprint, 2002.

[24] J. Tykesson, *Some aspects of Lévy processes in finance*, Master's thesis, Departement of Mathematical Statistics, Gothenburg, 2003.

[25] G.S. Watson, *Smooth regression analysis*, Sankhya, Series A 26, pp. 359-372, 1964.

[26] Website of Statistics Sweden, SCB, http://www.scb.se.

[27] A. Widerberg, *Value at risk : Gaussian vs extreme value methods*, Master's thesis, Departement of Mathematical Statistics, Gothenburg, 2003.

[28] Yahoo Finance Internet site: http://finance.yahoo.com.

[29] K. Åselius, *Tidsserieanalys tillämpad på spreaden SHB-SEB med införande av autoregressiv heteroskedastisk process för variansen*, Master's thesis, Departement of Matematics, Royal Institute of Technology, Stockholm, 2003 [In Swedish].

A Properties of the GH Distribution

Ole-Einar Barndorff-Nielsen [3] introduced the GH distribution in the seventies, to model wind blowned sand. In 1995, Eberlein and Keller [11] proposed GH as a possible distribution for log-increments of asset price processes.

The main advantage with the GH distribution, is that the many parameters makes it very flexible and suitable in many different contexts. The GH distribution, as used and modified by us, contains many important distributions as special cases. Examples of such distributions are the Normal, GP, t, NIG and Hyperbolic distribution (HYP).

From the literature and our own experience within, the field of mathematical finance, it seems clear that it can be very useful to make different reparametrizations of the GH distribution, to gain numerical and computational benefits. These benefits can be of importance, as the GH distribution can be a little hard to deal with, as given in (4). This in turn is in part due to the Bessel function that features in (4), and that the order of that function is one of the parameters to estimate. Another reason is the exponential behaviour of the tails of the GH distribution, which causes problems when dealing with data that do not display such exponential decay.

There are several examples of recent studies of the GH distribution with focus on applications in mathematical finance. Argubly, Ernst Eberlein and Karsten Prause have contributed the most to this field. In particular, Prauses Ph.D. thesis [17] is an important source for information about the GH distribution. See also e.g. [10].

The GH distribution is infinitely divisible (see e.g. [17]), so that it can serve as marginal distribution of a Lévy process.

The GH probability density function is given by (4). We complete this formula by giving a representation of the modified Bessel function of the third kind:

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} \mathrm{e}^{-x(y+1/y)/2} dy \quad \text{for} \quad x > 0. \tag{21}$$

It is mainly the properties of this Bessel function that affect the tractability of the GH distribution, and chosing $\lambda = 1$ or $\lambda = -1/2$ make calculations especially favourable. This is the main reason that these special cases are sorted out, among the various possible choices of λ.

Setting $\lambda = -1/2$ in (4), we get the *Normal Inverse Gaussian* distribution (NIG), with probability density function

$$f_{\mathrm{NIG}}(x;\alpha,\beta,\delta,\mu) = \frac{\alpha\delta}{\pi}\,\mathrm{e}^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)}\,\frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\sqrt{\delta^2+(x-\mu)^2}} \quad \text{for} \quad x \in \mathbb{R}.$$

The restrictions on the parameters are the same as for GH (9).

Another important special case of the GH distribution is $\lambda = 1$, which gives the *Hyperbolic* distribution (HYP), with probability density function

$$f_{\mathrm{HYP}}(x;\alpha,\beta,\delta,\mu) = \frac{1}{2\delta\alpha}\,\mathrm{e}^{\beta(x-\mu)-\alpha\sqrt{\delta^2+(x-\mu)^2}}\,\frac{\sqrt{\alpha^2-\beta^2}}{K_1(\delta\sqrt{\alpha^2-\beta^2})} \quad \text{for} \quad x \in \mathbb{R}.$$

Often parameter estimations are done by the Maximum Likelihood method. In principle, this is not complicated, but due to the many parameters, GH might be very flat in some dimensions. Then it can be favourable to calculate the derivatives of the log-likelihood function. We used this possibility occasionally, to find starting values for optimization. See e.g. [17] on closed formulas for these derivatives.

To facilitate parameter estimation, one might want to solve the equation for some parameters. This could be done using the method of moments. For example, by means of estimating the expected value with the sample mean, the parameter μ could be solved from the following expression for the expected value of GH:

$$\mathbf{E}\{\mathrm{GH}(\lambda,\alpha,\beta,\delta,\mu)\} = \mu + \frac{\delta\,\beta}{\sqrt{\alpha^2-\beta^2}}\,\frac{K_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}.$$

Here, in addition to (9), it is required that $\lambda < -1$ if $|\beta| = \alpha$ [see (6)].

Similarly, one might estimate the variance with the sample variance, and make use of the following expression for the variance of GH:

$$\begin{aligned}\mathbf{Var}\{\mathrm{GH}(\lambda,\alpha,\beta,\delta,\mu)\} = {} & \frac{\delta\,K_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{\sqrt{\alpha^2-\beta^2}\,K_\lambda(\delta\sqrt{\alpha^2-\beta^2})} \\ & + \frac{\delta^2\beta^2}{\alpha^2-\beta^2}\left[\frac{K_{\lambda+2}(\delta\sqrt{\alpha^2-\beta^2})}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})} - \left(\frac{K_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}\right)^2\right].\end{aligned}$$

Here, in addition to (9), it is required that $\lambda < -2$ if $|\beta| = \alpha$ [see (6)].

Expression for the moment generating and characteristic functions of the GH distribution, to compute more moments, can be found in e.g. [17].

As stated before, the analytical handling of the GH distribution can be hard, due to the complexity of the density functions. Therefore it is important to check the literature concerning. However, we have discovered, that in some cases, there are some typing errors in that literature, so we give tha advice to, whenever possible, double check all formulas with a few independent sources. This, on the other hand, is not always easy to find.

B The Ljung-Box Statistic

To decide the parameters of the Nadaraya-Watson devolatilization structure according to devolatilize log-returns for the Siemens stock we used the Ljung-Box statistics, described in [14]. From (19), we have that the parameters to estimate are the width of the window n and the parameter to the normal kernel h.

The test of Ljung and Box is commonly used to test the quality of fit of a time seried model. The idea is that for different values on n and h use the Ljung-Box statistic to check if the model is correct, i.e. that the data are uncorrelated. Ljung and Box ([13]) argue that

$$Q(k) = T(T+2)\sum_{i=1}^{k}(T-i)^{-1}r_i^2$$

should be χ_k^2 distributed, if the model is correct. This means that, if no signifcant autocorrelation is found from the residuals, then the model pass the test. As in our case, we want independence and therefore no autocorrelation this test would work. The *sample autocorrelation function* r_k is described by

$$r_k = \frac{\sum_{t=k+1}^{T}(x_t-\bar{x})(x_{t-k}-\bar{x})}{Ts^2} \quad \text{for} \quad k = 1, 2, \dots ,$$

where $(x_t)_{t=1}^T$ are our devolatized log-returns, $\bar{x} = \sum_{t=1}^T x_t$ the sample mean and $s^2 = \frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2$ the sample variance.

One problem that occurs when doing the calculation, is how many lags k that should be used. A rule of thumb, is that $0 << k << T$ – usually $k \approx \frac{T}{5}$), which in our case with $T = 1700$ means $k = 340$. On the other hand, according to Patrick Burns ([8]), the number of lags should be no more than 5% of the length of the series, which would give us $k \leq 85$. We handled this problem by calculating the Ljung-Box statistic for $k = 1 \dots 100$, and then studied which k that produced the lowest p-value $\mathbf{P}\{\chi_k^2 > Q(k)|\text{model is correct}\}$, giving the k corresponding to the highest correlation for the specific n and h beeing tested.

In our case, the test were then carried out in Matlab using a double for-loop looping partwise over $h = 1 \dots 100$ and $n = 1 \dots 100$. For every choice of h and n the whole dataset is devolatilized using equation (19). A loop for $k = 1 \dots 100$ is then constructed and the Ljung-Box statistic p-value is calculated for every k. The smallest p-value (corresponding to the highest correlation as stated before) is then placed on location (h, n) in a 100-dimensional square matrix. The only thing left was now to find the maximum p-value in the matrix corresponding to the n and h giving the least amount of autocorrelation. Doing this four our dataset of log-returns of the Siemens stock described in Section 3.3, we found that the least autocorrelation was found four $n = 14$ and that the p-values corresponding to increasing h converges for $h > 40$. As h is a parameter that does not affect the number of calculation we choosed $h = 50$ as te be sure of the convergence behaviour.

The Matlab code for this is delivered here:

```
function q=ljungbox
s=importsiemens;clear q;
for h = 1:100
  for n = 1:100
    devolx=0;

    %The devolatilization is carried out for every choice of
    %h and n, using the function volest. The vector devolx
    %will now contain the devolatized values for this h,n.

    for i = (n+1):length(s)
      devolx(i)=(s(i)-mean(s))/sqrt(volest(i,h,n));
    end
      ldx=length(devolx);pmin=1;
    for k = 1:100
    ria=0;
        for j = 1:k
              ria = ria+(sacf(j,devolx))^2/(ldx-j);
        end
          qk=ldx*(ldx-2)*ria;
        if (1-chi2cdf(qk,k))<pmin
             pmin=(1-chi2cdf(qk,k));
        end

     %The p-value for the Ljung-Box statistic is calculated
     %for each k. The smallest p is element (h,n) in matrix q.

    end
      q(h,n)=pmin;
  end
end
```

C LF Parameter Estimation

```
utbetalningar = Import["nylansforsakring.dat"];
index = Import["nykpi.dat"]; $RecursionLimit = $IterationLimit = 20000;
nyindex = Table[Table[1 / (index[[i, j + 1]] / index[[1, 13]]), {j, 1, 12}], {i, 1, 10}];
cutoff = Max[nyindex] * 500000; nyutbetalningar = Table[utbetalningar[[i]][[5]] *
    nyindex[[10 - (utbetalningar[[i]][[2]] - 1990)]][[(utbetalningar[[i]][[3]])]],
   {i, 1, 450}]; For[nynyutbetalningar = {}; i = 1, i ≤ 450, i++,
 If[nyutbetalningar[[i]] ≥ cutoff, AppendTo[nynyutbetalningar, nyutbetalningar[[i]]]]];
data = Sort[nynyutbetalningar];
logdata = Log[data] - Table[Log[cutoff], {k, 1, Length[data]}]; lt = Length[data]

348

%% The raw data is collected from the file "nylansforsakring.dat" and is corrected
 for inflasion. The vector "data" now contains all the data we want to use.

Clear[Femp]; Femp[data_, x_] := Sum[If[data[[i]] ≤ x, 1, 0], {i, 1, lt}] / lt;

Clear[ftest2]; ftest2[x_, delta_, beta_, ksi_, lambda_, sigma_] :=
```

$$(\text{delta}^\wedge 2 + \text{ksi} * (x / \text{sigma})^\wedge(\text{beta}))^\wedge(-\text{lambda}/\text{ksi}) \Bigg/ \left(\frac{\text{delta}^{\frac{2}{\text{beta}}-\frac{2\,\text{lambda}}{\text{ksi}}}\ \text{ksi}^{-1/\text{beta}}\ \text{sigma Gamma}\left[1+\frac{1}{\text{beta}}\right]\text{Gamma}\left[-\frac{1}{\text{beta}}+\frac{\text{lambda}}{\text{ksi}}\right]}{\text{Gamma}\left[\frac{\text{lambda}}{\text{ksi}}\right]}\right);$$

```
%% Femp is the empirical distributionfunction
  and ftest2 the density of our adjusted GH - distribution.

Clear[Ftest1]; Ftest1[y_, delta_, beta_, ksi_, lambda_, sigma_] :=
```

$$\frac{(-1)^{-1/\text{beta}}\ \text{Beta}\left[-\frac{\text{ksi}\left(\frac{y}{\text{sigma}}\right)^{\text{beta}}}{\text{delta}^2},\ \frac{1}{\text{beta}},\ 1-\frac{\text{lambda}}{\text{ksi}}\right]\text{Gamma}\left[\frac{\text{lambda}}{\text{ksi}}\right]}{\text{Gamma}\left[\frac{1}{\text{beta}}\right]\text{Gamma}\left[-\frac{1}{\text{beta}}+\frac{\text{lambda}}{\text{ksi}}\right]};$$

```
Clear[Ftest2]; Ftest2[y_, delta_, beta_, ksi_, lambda_, sigma_] :=
 Ftest1[y - cutoff, delta, beta, ksi, lambda, sigma * cutoff];

%% ftest2 is the density function and Ftest2
 is the distribution function used in the estimation

Clear[Fpareto];
Fpareto[x_, ksi_, sigma_] := 1 - (1 + ksi * (x - cutoff) / (cutoff * sigma)) ^ (-1 / ksi);

%% The patero distribution function
Clear[maxtestdistance]; maxtestdistance[delta_, beta_, ksi_, lambda_, sigma_] :=
 Max[Table[Max[Abs[Ftest2[data[[i]], delta, beta, ksi, lambda, sigma] - i / lt],
    Abs[Ftest2[data[[i]], delta, beta, ksi, lambda, sigma] - (i - 1) / lt]], {i, 1, lt}]]

{delta0, beta0, ksi0, lambda0, sigma0} =
 {delta0, beta0, ksi0, lambda0, sigma0} /. {beta0 → 0.45909862847752564`,
   delta0 → 0.9325795620072715`, ksi0 → 0.07325907557972827`,
   lambda0 → 1.4885845339431245`, sigma0 → 0.6423128953369146`};
Clear[delta, beta, ksi, lambda, sigma]; FindMinimum[
 maxtestdistance[delta, beta, ksi, lambda, sigma],
 {{lambda, lambda0}, {beta, beta0}, {delta, delta0}, {ksi, ksi0}, {sigma, sigma0}}]

{0.0260494,
 {lambda → 1.48857, beta → 0.459042, delta → 0.932614, ksi → 0.0732931, sigma → 0.642325}}

%% maxtestdistance is the kolmogorovdistance formula. And this is the corrected GH -
 distribution formula.

Clear[maxpardistance];
maxpardistance[ksi_, sigma_] := Max[Table[Max[Abs[Fpareto[data[[i]], ksi, sigma] - i / lt],
    Abs[Fpareto[data[[i]], ksi, sigma] - (i - 1) / lt]], {i, 1, lt}]]

Clear[ksi, sigma];
NMinimize[{maxpardistance[ksi, sigma], ksi > 0, sigma > 0}, {ksi, sigma}]

{0.034456, {ksi → 0.987084, sigma → 0.995859}}

%% Here is the minimization of the distance
 when the distribution function assumpotion is the GPD
```

D Siemens Parameter Estimation

```
$RecursionLimit = $IterationLimit = Infinity; indata = Import["devolsiemens.txt", "List"];
cutoff = 1.4; indata = -indata; lt = Length[indata]; For[negdata = {}; i = 1,
 i <= lt, i++, If[indata[[i]] >= cutoff, AppendTo[negdata, indata[[i]]]]];
negdata = Sort[negdata]; nynegdata = negdata; PrependTo[nynegdata, cutoff];
neglt = Length[negdata]; Clear[gh]; gh[x_, lambda_, alfa_, beta_, delta_, sigma_] :=
 If[beta == 1, 1, beta * ((x - cutoff) / sigma) ^ (beta - 1)] *
   (alfa^(1/2) delta^-lambda (delta^2 + ((x - cutoff) / sigma)^(2*beta))^(1/2 (-1/2+lambda))
     BesselK[-1/2 + lambda, alfa Sqrt[delta^2 + ((x - cutoff) / sigma)^(2*beta)]]) /
   (Sqrt[2 π] BesselK[lambda, alfa * delta] * sigma); Unset[maxnegdist[lambda_?NumberQ,
  alfa_?NumberQ, beta_?NumberQ, delta_?NumberQ, sigma_?NumberQ]]; Clear[maxnegdist];
maxnegdist[lambda_?NumberQ, alfa_?NumberQ, beta_?NumberQ,
  delta_?NumberQ, sigma_?NumberQ] :=
 {coeff = NIntegrate[gh[x, lambda, alfa, beta, delta, sigma], {x, cutoff, 100 * cutoff}],
   For[vector = {}; int = 0; i = 1, i <= neglt, i++,
    int = int + NIntegrate[gh[x, lambda, alfa, beta, delta, sigma],
        {x, nynegdata[[i]], nynegdata[[i + 1]]}] / coeff; AppendTo[vector, int]],
   nymax = Max[{Max[Abs[vector - Table[(i - 1) / neglt, {i, 1, neglt}]]],
      Max[Abs[vector - Table[i / neglt, {i, 1, neglt}]]]}], nymax}[[4]];
Unset[cutgh[x_?NumberQ, lambda_?NumberQ, alfa_?NumberQ, beta_?NumberQ,
  delta_?NumberQ, sigma_?NumberQ]]; Clear[cutgh];
cutgh[x_?NumberQ, lambda_?NumberQ, alfa_?NumberQ,
  beta_?NumberQ, delta_?NumberQ, sigma_?NumberQ] :=
 gh[x, lambda, alfa, beta, delta, sigma] / NIntegrate[
   gh[y, lambda, alfa, beta, delta, sigma], {y, cutoff, 100 * cutoff}]; NMaximize[
 {Sum[Re[Log[cutgh[negdata[[i]], -0.4, alfa, beta, delta, sigma]]], {i, 1, neglt}],
  sigma > 0.1, delta > 0.1, alfa > 0.0001, beta > 0.1}, {alfa, beta, delta, sigma}]

{-98.6632, {alfa → 0.71937, beta → 1.30047, delta → 0.115683, sigma → 2.47353}}

maxnegdist[-0.4, alfa, beta, delta, sigma] /. %[[2]]

0.0369418
```

E Spread SHB-SEB, t and Normal Distribution

```
data = Import["spread.txt", "List"]; $RecursionLimit = $IterationLimit = 20000;
<< Statistics`DescriptiveStatistics`; lt = Length[data]

2726

Clear[fn]; fn[x_] := (1 / Sqrt[2 * Pi]) * Exp[-x^2 / 2];

Clear[logAAn];
logAAn[alfa_, ksi_, phi_, length_] :=
  Sum[Log[fn[(data[[i]] - phi * data[[i - 1]]) / Sqrt[
        ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^2]] /
     Sqrt[ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^2]], {i, 3, length}];

NMaximize[{logAAn[alfa, ksi / 10000, phi, 500],
  0.20 ≤ alfa ≤ 0.45, 0.95 ≤ phi < 1, 1 ≤ ksi ≤ 15}, {alfa, phi, ksi}]

{1288.81, {alfa → 0.200543, ksi → 2.78693, phi → 0.98093}}

{alfa0, phi0, ksi0} = {alfa, phi, ksi} /. {alfa → Log[0.20054336646099336`],
   ksi → Log[2.7869266807340853`], phi → Log[0.9809299529400217`]};
Clear[alfa, phi, ksi]; FindMaximum[logAAn[Exp[alfa], Exp[ksi] / 10000, Exp[phi], lt],
 {{alfa, alfa0}, {phi, phi0}, {ksi, ksi0}}]

{5812.07, {alfa → -1.40888, phi → -0.0134107, ksi → 1.89396}}

{alfa0, phi0, ksi0} =
 {alfa, phi, ksi} /. {alfa → -1.4088824161023035`, phi → -0.013410658141120203`,
   ksi → 1.8939649855526997`}; {Exp[alfa0], Exp[phi0], Exp[ksi0] / 10000}

{0.244416, 0.986679, 0.000664567}

Clear[ft];
ft[ny_, x_] := (Gamma[(ny + 1) / 2] / (Sqrt[Pi * (ny - 2)] * Gamma[ny / 2])) *
   (1 + (x / Sqrt[ny - 2]) ^ 2) ^ (-(ny + 1) / 2);

Clear[logAAt];
logAAt[ny_, alfa_, ksi_, phi_, length_] :=
 Sum[Log[ft[ny, (data[[i]] - phi * data[[i - 1]]) /
      Sqrt[ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^2]] /
    Sqrt[ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^2]], {i, 3, length}]

{alfa0, phi0, ksi0} = {alfa, phi, ksi} /. {alfa → -1.4088824161023035`,
   phi → -0.013410658141120203`, ksi → 1.8939649855526997`}; Clear[alfa, phi, ksi];
FindMaximum[logAAt[Exp[ny] + 2, Exp[alfa], Exp[ksi] / 10000, Exp[phi], 500],
 {{ny, Log[2]}, {alfa, alfa0}, {phi, phi0}, {ksi, ksi0}}]

{1311.49, {ny → 0.769964, alfa → -1.6127, phi → -0.0158108, ksi → 1.09285}}

{ny0, alfa0, phi0, ksi0} = {ny, alfa, phi, ksi} /. %[[2]]; Clear[ny, alfa, phi, ksi];
FindMaximum[logAAt[Exp[ny] + 2, Exp[alfa], Exp[ksi] / 10000, Exp[phi], lt],
 {{ny, ny0}, {alfa, alfa0}, {phi, phi0}, {ksi, ksi0}}]

{5940.4, {ny → 0.833104, alfa → -1.03665, phi → -0.0123298, ksi → 1.86586}}

{ny0, alfa0, phi0, ksi0} = {ny, alfa, phi, ksi} /. %%%[[2]];
{Exp[ny0] + 2, Exp[alfa0], Exp[phi0], Exp[ksi0] / 10000}

{4.30045, 0.35464, 0.987746, 0.000646149}
```

F Spread SHB-SEB, GH Distribution

```
data = Import["spread.txt", "List"]; $RecursionLimit = $IterationLimit = 20000;
<< Statistics`DescriptiveStatistics`; lt = Length[data]
```

```
2726
```

```
Clear[ftest2]; ftest2[x_, delta_, beta_, ksi_, lambda_, sigma_] :=
 (delta^2 + ksi * (x / sigma) ^ (beta)) ^ (-lambda / ksi) /
  ((2 delta^(2/beta - 2 lambda/ksi) ksi^(-1/beta) sigma Gamma[1 + 1/beta] Gamma[-1/beta + lambda/ksi]) /
    Gamma[lambda/ksi])
```

```
Clear[ftest3]; ftest3[x_, delta_, beta_, ksi_, lambda_, sigma_] :=
 ftest2[x * Sqrt[(delta^(4/beta) ksi^(-2/beta) sigma^2 Gamma[3/beta] Gamma[-3/beta + lambda/ksi]) /
      (Gamma[1/beta] Gamma[-1/beta + lambda/ksi])],
   delta, beta, ksi, lambda, sigma] *
  Sqrt[(delta^(4/beta) ksi^(-2/beta) sigma^2 Gamma[3/beta] Gamma[-3/beta + lambda/ksi]) /
    (Gamma[1/beta] Gamma[-1/beta + lambda/ksi])]
```

```
2 * Integrate[ftest3[x, delta, beta, ksi, lambda, sigma], {x, 0, Infinity}, Assumptions →
   beta > 0 && lambda > 0 && ksi > 0 && sigma > 0 && delta > 0 && beta * lambda / ksi > 3]
```

```
1
```

```
2 * Integrate[x^2 * ftest3[x, delta, beta, ksi, lambda, sigma],
  {x, 0, Infinity}, Assumptions →
   beta > 0 && lambda > 0 && ksi > 0 && sigma > 0 && delta > 0 && beta * lambda / ksi > 3]
```

```
1
```

```
Clear[ftest4]; ftest4[x_, delta_, beta_, ksi_, lambda_, sigma_] :=
 ftest3[x, delta, beta + 3 * ksi / lambda, ksi, lambda, sigma]

Clear[logAAgh];
logAAgh[deltagh_, betagh_, ksigh_, lambdagh_, sigmagh_, alfa_, ksi_, phi_, length_] :=
 Sum[Log[ftest4[Abs[data[[i]] - phi * data[[i - 1]]] /
      Sqrt[ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^ 2],
     deltagh, betagh, ksigh, lambdagh, sigmagh] /
    Sqrt[ksi + alfa * (data[[i - 1]] - phi * data[[i - 2]]) ^ 2]], {i, 3, length}]

{deltagh0, betagh0, ksigh0, lambdagh0, sigmagh0} =
 {deltagh, betagh, ksigh, lambdagh, sigmagh} /.
  {betagh → 0.1684942151126499`, lambdagh → 5.12485596082017`,
   ksigh → -6.26095243398954`, sigmagh → -0.7984954912205737`,
   deltagh → -4.882041347310315`, alfa → -1.1444491538363653`,
   ksi → 1.4441481133893268`, phi → -0.013365442273153911`};
FindMaximum[logAAgh[Exp[deltagh], Exp[betagh], Exp[ksigh], Exp[lambdagh],
  Exp[sigmagh], Exp[alfa], Exp[ksi] / 10000, Exp[phi], lt],
 {{betagh, betagh0}, {lambdagh, lambdagh0}, {ksigh, ksigh0}, {sigmagh, sigmagh0},
  {deltagh, deltagh0}, {alfa, alfa0}, {ksi, ksi0}, {phi, phi0}}]
```

```
{5948.1, {betagh → 0.166061, lambdagh → 5.12142, ksigh → -6.26538, sigmagh → -0.798495,
  deltagh → -4.8818, alfa → -1.07543, ksi → 1.80304, phi → -0.00807647}}
```

```
{deltagh0, betagh0, ksigh0, lambdagh0, sigmagh0, alfa0, ksi0, phi0} =
 {deltagh, betagh, ksigh, lambdagh, sigmagh, alfa, ksi, phi} /. %%%[[2]];
{Exp[deltagh0], Exp[betagh0] + 3 * Exp[ksigh0] / Exp[lambdagh0], Exp[ksigh0],
 Exp[lambdagh0], Exp[sigmagh0], Exp[alfa0], Exp[ksi0] / 10000, Exp[phi0]}
```

```
{0.00758337, 1.18068, 0.00190099, 167.573, 0.450006, 0.34115, 0.000606807, 0.991956}
```

G Portfolio Weight Correction and Mean Excess Function

```
Clear[port]; porttemp = Transpose[{Import["assabloy.txt", "Table"][[All, 2]],
   Import["atlascopco.txt", "Table"][[All, 2]],
   Import["electrolux.txt", "Table"][[All, 2]],
   Import["ericsson.txt", "Table"][[All, 2]], Import["gunnebo.txt", "Table"][[All, 2]],
   Import["kinnevik.txt", "Table"][[All, 2]], Import["lindex.txt", "Table"][[All, 2]],
   Import["nobelbiocare.txt", "Table"][[All, 2]],
   Import["nokia.txt", "Table"][[All, 2]], Import["sandvik.txt", "Table"][[All, 2]],
   Import["sebc.txt", "Table"][[All, 2]], Import["shbb.txt", "Table"][[All, 2]],
   Import["skf.txt", "Table"][[All, 2]], Import["ssab.txt", "Table"][[All, 2]],
   Import["volvo.txt", "Table"][[All, 2]]}];

Clear[nyport]; For[nyport = {}; i = 1, i ≤ Length[porttemp], i++, AppendTo[nyport,
  (1 / 15 * Total[porttemp[[i, All]]]) / (1 / 15 * Total[porttemp[[1, All]]])]]

port = porttemp;

vikt = Table[(1^i) / 15, {i, 15}]; Clear[pos];
For[tot = {}; i = 1, i ≤ 521, i++, {AppendTo[tot, Sum[vikt[[j]] * port[[i, j]], {j, 1, 15}]];
  minj = vikt[[1]] * port[[i, 1]]; For[rest = 0; j = 1, j ≤ 15, j++,
   {If[vikt[[j]] * port[[i, j]] < minj, {minj = vikt[[j]] * port[[i, j]]; pos = j}];
    If[vikt[[j]] * port[[i, j]] > 0.1 * tot[[i]],
     {rest = vikt[[j]] * port[[i, j]] - 0.1 * tot[[i]];
      vikt[[j]] = (vikt[[j]] * port[[i, j]] - rest) / port[[i, j]]}]
   }];
  If[rest ≠ 0, vikt[[pos]] = (rest + vikt[[pos]] * port[[i, pos]]) / port[[i, pos]]]
 }]

vikt
```

$\left\{\frac{1}{15}, \frac{1}{15}, \frac{1}{15}, 1.21196, 0.0771071, 0.087652, 0.067632, 0.0227503, 0.0518069, 0.0548744, 0.0927271, \frac{1}{15}, 0.0547344, 0.0908442, \frac{1}{15}\right\}$

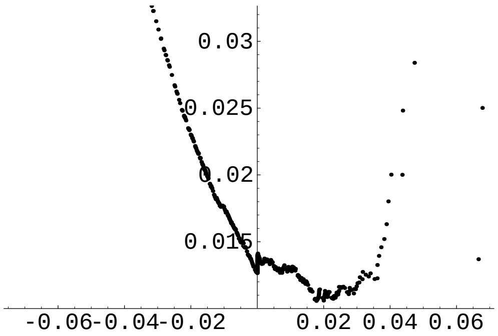
```
logport = Sort[logport];

Clear[mef]; For[mef = {}; i = 1, i < Length[logport], i++, {AppendTo[mef,
   Sum[logport[[j]] - logport[[i]], {j, i + 1, Length[logport]}] / (Length[logport] - i)]}]

templogport = logport[[Range[1, 519]]];

Clear[nymef]; For[nymef = {}; i = 1, i ≤ Length[templogport],
 i++, AppendTo[nymef, {templogport[[i]], mef[[i]]}]];

ListPlot[nymef]
```



- Graphics -

H Portfolio Parameter Estimation

```
// logoport contains the log - returns evaluated before

In[12]:= logport = Sort[logport];

cutoff = 0.02;

cutoff = 0.02; For[data = {}; i = 1, i ≤ Length[logport],
 i++, If[logport[[i]] ≥ cutoff, AppendTo[data, logport[[i]]]]];
data = Sort[data]; nydata = data; PrependTo[nydata, cutoff]; lt = Length[data]

61

Clear[Femp]; Femp[dat_, x_] := Sum[If[dat[[i]] ≤ x, 1, 0], {i, 1, lt}] / lt;

Clear[Fpareto]; Fpareto[x_, ksi_, sigma_] := 1 - (1 + ksi * (x - cutoff) / (sigma)) ^ (-1 / ksi);

D[Fpareto[x, ksi, sigma], x]
```

$$\frac{\left(1+\frac{\text{ksi}\,(-0.02+x)}{\text{sigma}}\right)^{-1-\frac{1}{\text{ksi}}}}{\text{sigma}}$$

```
Clear[ftest1]; ftest1[x_, delta_, beta_, ksi_, lambda_, sigma_] :=
 (delta^2 + ksi * (x / sigma) ^ (beta)) ^ (-lambda / ksi);

ftest1[x - cutoff, 1, 1, ksi, ksi + 1, sigma]
```

$$\left(1+\frac{\text{ksi}\,(-0.02+x)}{\text{sigma}}\right)^{\frac{-1-\text{ksi}}{\text{ksi}}}$$

```
Clear[maxpardistance];
maxpardistance[ksi_, sigma_] := Max[Table[Max[Abs[Fpareto[data[[i]], ksi, sigma] - i / lt],
    Abs[Fpareto[data[[i]], ksi, sigma] - (i - 1) / lt]], {i, 1, lt}]]

Clear[ksi, sigma];
NMinimize[{maxpardistance[ksi, sigma], ksi > 0, sigma > 0}, {ksi, sigma}]

{0.0459453, {ksi → 0.124135, sigma → 0.00927758}}
$RecursionLimit = $IterationLimit = Infinity; Clear[gh];
gh[x_, lambda_, alfa_, beta_, delta_, my_] :=
```

$$\left(\text{alfa}^{\frac{1}{2}-\text{lambda}}\,(\text{alfa}^2-\text{beta}^2)^{\text{lambda}/2}\,\text{delta}^{-\text{lambda}}\,e^{\text{beta}\,(-\text{my}+x)}\,(\text{delta}^2+(-\text{my}+x)^2)^{\frac{1}{2}\left(-\frac{1}{2}+\text{lambda}\right)}\,\text{BesselK}\left[-\frac{1}{2}+\text{lambda},\ \text{alfa}\sqrt{\text{delta}^2+(-\text{my}+x)^2}\right]\right)\Big/\left(\sqrt{2\pi}\,\text{BesselK}\left[\text{lambda},\ \sqrt{\text{alfa}^2-\text{beta}^2}\ \text{delta}\right]\right);$$

```
Unset[maxposdist[lambda_?NumberQ, alfa_?NumberQ,
  beta_?NumberQ, delta_?NumberQ, my_?NumberQ]]; Clear[maxposdist];
maxposdist[lambda_?NumberQ, alfa_?NumberQ, beta_?NumberQ,
  delta_?NumberQ, my_?NumberQ] :=
 {coeff = NIntegrate[gh[x, lambda, alfa, beta, delta, my], {x, cutoff, 10}],
   For[vector = {}; int = 0; i = 1, i ≤ lt, i++, int = int + NIntegrate[
        gh[x, lambda, alfa, beta, delta, my], {x, nydata[[i]], nydata[[i + 1]]}] / coeff;
    AppendTo[vector, int]], nymax = Max[{Max[Abs[vector - Table[(i - 1) / lt, {i, 1, lt}]]],
      Max[Abs[vector - Table[i / lt, {i, 1, lt}]]]}], nymax}[[4]]

NMinimize[{maxposdist[lambda, alfa, beta, delta, my], delta > 0, alfa > Abs[beta]},
 {lambda, alfa, beta, delta, my}]

{0.0460008,
 {alfa → 2.69949, beta → -0.857383, delta → 0.0186723, lambda → -4.1948, my → -0.0545497}}
```

J Jason 1965-2002