The Generalised Hyperbolic Skew Student's t-distribution

Kjersti Aas†

The Norwegian Computing Center, Oslo, Norway

Ingrid Hobæk Haff

The Norwegian Computing Center, Oslo, Norway

Abstract.
The empirical distribution of daily returns from financial market variables such as exchange rates, equity prices, and interest rates, is often skewed, having one heavy, and one semi-heavy, or more Gaussian-like tail. The NIG distribution, that has two semi-heavy tails, models skewness rather well, but only in cases where the tails are not too heavy. On the other hand, the skew Student's t-distributions presented in the literature have two polynomial tails. Hence, they fit heavy-tailed data well, but they do not handle substantial skewness.
In this paper, we argue for a special case of the generalised hyperbolic distribution that we denote the GH skew Student's t-distribution. This distribution has the important property that one tail has polynomial, and the other exponential behaviour. Further, it is the only subclass of the generalised hyperbolic distribution having this property. Although the GH skew Student's t-distribution has been previously proposed in the literature, it is not well known, and specifically, its special tail behaviour has not been addressed.
This paper presents empirical evidence of exponential/polynomial tail behaviour in skew financial data, and demonstrates the superiority of the GH skew Student's t-distribution with respect to data fit, compared with its competitors. Through VaR and expected shortfall calculations we show why the exponential/polynomial tail behaviour is important in practice. We also present a simple algorithm for computing the MLE estimators, using a mixture representation of the GH skew Student's t-distribution and the EM-algorithm.

1. Introduction

It is a well-known fact that returns from financial market variables such as exchange rates, equity prices, and interest rates, measured over short time intervals, i.e. daily or weekly, are characterized by non-normality. The empirical distribution of such returns is more peaked and has heavier tails than the normal distribution, which implies that very large changes in returns occur with a higher frequency than under normality. In addition it is often skewed, having one heavy, and one semi-heavy or more Gaussian-like tail.

One of the most promising distributions for such returns proposed in the literature, is the normal inverse Gaussian (NIG) distribution (Barndorff-Nielsen, 1997). The NIG distribution possesses a number of attractive theoretical properties, among others its analytical tractability. For these reasons, it has been used repeatedly for financial applications, both as the conditional distribution of a GARCH-model (Andersson, 2001; Jensen and Lunde, 2001; Forsberg and Bollerslev, 2002; Venter and de Jongh, 2002) and as the unconditional

†*Address for correspondence:* The Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway
E-mail: Kjersti.Aas@nr.no

return distribution (Bølviken and Benth, 2000; Eberlein and Keller, 1995; Lillestøl, 2000; Prause, 1997; Rydberg, 1997). The two tails of the NIG distribution behave differently, but they are both semi-heavy. One would therefore expect NIG to model skewness rather well, but only in cases where the tails are not too heavy.

An alternative set of distributions for modelling skew and heavy-tailed data is the skew extensions to the Student's t-distribution. Hansen (1994) was the first to propose a skew extension to the Student's t-distribution for modelling financial returns. Since then, several other papers have studied different skew t-type distributions for financial and other applications, see e.g. Azzalini and Capitanio (2003); Bauwens and Laurent (2002); Branco and Dey (2001); Fernandez and Steel (1998); Jones and Faddy (2003); Patton (2004); Sahu et al. (2003); Venter and de Jongh (2002). All these skew t-type distributions have two tails behaving as polynomials. This means that they fit heavy-tailed data well, but they do not handle substantial skewness.

The NIG distribution is a subclass of the generalised hyperbolic (GH) distribution (Barndorff-Nielsen, 1977). The GH distributions possess a number of attractive properties, e.g. they are closed under conditioning, marginalisation and affine transformations. They can be both symmetric and skew, and their tails are generally semi-heavy. While several specific subclasses, like the NIG and the hyperbolic distribution (Barndorff-Nielsen and Blæsild, 1981), have been applied in various situations, the GH distribution itself is seldom used in practical applications. This is probably due to the fact that it is not particularly analytically tractable, and that it even for very large sample sizes, may be hard to make a distinction between different values of the parameter determining the subclass. The latter is due to the flatness of the GH likelihood function in this parameter (Prause, 1999).

In this paper, we argue for a special case of the generalised hyperbolic distribution that we denote the GH skew Student's t-distribution. It is briefly mentioned by Prause (1999), Barndorff-Nielsen and Shepard (2001), Jones and Faddy (2003), Mencia and Sentana (2004) and Demarta and McNeil (2004). However, it not well known, and specifically, its special tail behaviour has not been addressed. Unlike any other member of the GH family of distributions, it has one tail determined by polynomial, and the other by exponential behaviour. This distribution is almost as analytically tractable as the NIG distribution. Moreover, maximum likelihood estimation of its parameters is quite straightforward using the EM-algorithm (Dempster et al., 1977), making it very useful for financial applications.

The remainder of this paper is organised as follows. Section 2 presents empirical evidence for the exponential/polynomial tail behaviour of skew financial data. Section 3 reviews other skew distributions with heavy or semi-heavy tails. Section 4 provides the definition of the GH skew Student's t-distribution, and Section 5 gives the details of the EM-algorithm for the estimation of its parameters. In Section 6, we fit the GH skew Student's t-distribution to the financial market variables presented in Section 2, and compare the results with the fit of the alternative distributions presented in Section 3. The practical importance of the exponential/polynomial tail behaviour of the GH skew Student's t-distribution is highlighted through VaR and expected shortfall calculations in Section 7. Finally, Section 8 contains some concluding remarks.

2. Data

The data set studied in this paper consists of four different kinds of market variables; the total index for Norwegian stocks (TOTX), the SSBWG hedged bond index for international

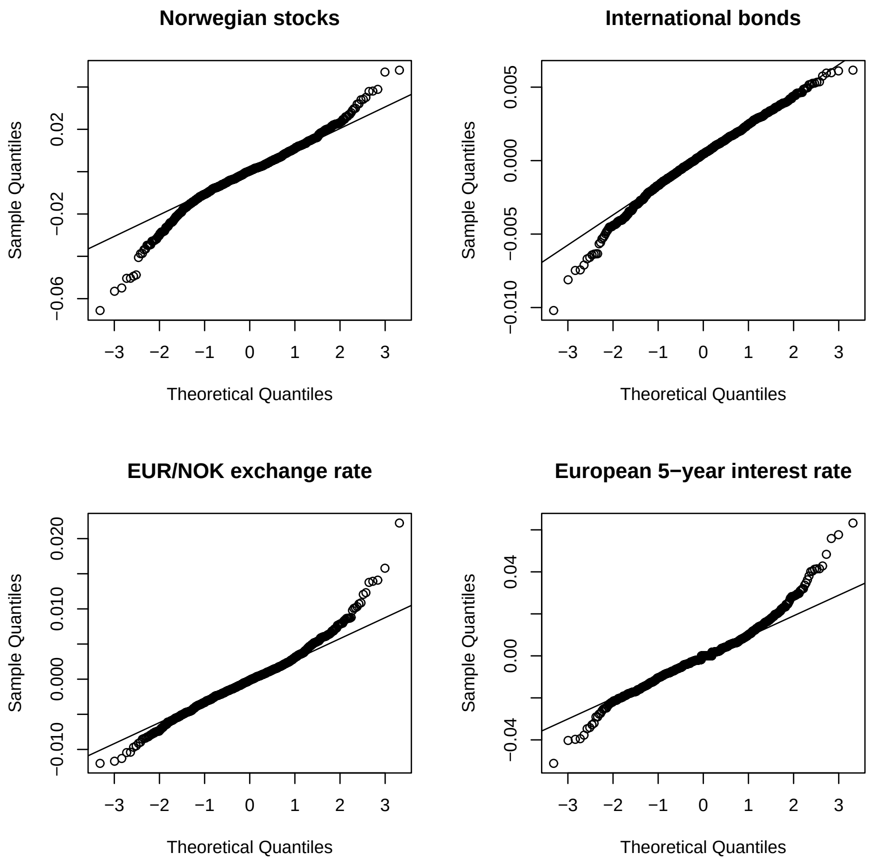

Figure 1. QQ-plots for log returns of selected financial market variables.

bonds, the NOK/EUR exchange rate (NOK is Norwegian Krones), and the EURIBOR 5-year interest rate. The historical time period used goes from 04.01.1999 to 08.07.2003, and corresponds to 1094 observations. Figure 1 shows normal QQ-plots for the corresponding logarithmic returns. For the Norwegian stock return distribution, both tails are heavier than the Gaussian, and the left tail is heavier than the right. The left tail of the international bond return distribution is heavy also, while the right tail is lighter than the Gaussian distribution. For the NOK/EUR exchange rate distribution the right tail is heavy, and the left tail that is closer to the Gaussian distribution. Finally, for the European 5-year interest data, both tails are heavy, but the right tails is heavier than the left.

Hence, all distributions are clearly skewed, having one heavy, and one semi-heavy, or more Gaussian-like tail. This motivates for the use of the GH skewed Student's t-distribution, which has one tail determined by polynomial, and the other by exponential behaviour.

3. A review of skew and heavy-tailed distributions

This section gives an overview of other skew distributions with heavy or semi-heavy tails, more specifically, the NIG distribution and various definitions of skew Student's distributions.

3.1. NIG

The normal inverse Gaussian (NIG) distribution is a generalised hyperbolic distribution with $\lambda = -\frac{1}{2}$. Its density is

$$f_x(x) = \frac{\delta\,\alpha\,\exp\left(\delta\sqrt{\alpha^2-\beta^2}\right)\,K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\,\exp\left(\beta\,(x-\mu)\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}},$$

where $\delta > 0$ and $0 < |\beta| \leq \alpha$. The parameters μ and δ determine the location and scale, respectively, while α and β control the shape of the density. In particular, $\beta = 0$ corresponds to a symmetric distribution.

It can be shown that in the tails, the NIG distribution behaves as

$$f_x(x) \sim \mathrm{const}|x|^{-3/2}\exp\left(-\alpha|x| + \beta\,x\right) \qquad \text{as } x \to \pm\infty. \tag{1}$$

More specifically, the heaviest tail decays as

$$f_x(x) \sim \mathrm{const}|x|^{-3/2}\exp\left(-\alpha|x| + |\beta|\,|x|\right) \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad -\infty, \\ \beta > 0 & \text{and} \quad x \to \quad +\infty, \end{cases}$$

and the lightest as

$$f_x(x) \sim \mathrm{const}|x|^{-3/2}\exp\left(-\alpha|x| - |\beta|\,|x|\right) \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad +\infty, \\ \beta > 0 & \text{and} \quad x \to \quad -\infty. \end{cases}$$

Thus, the two tails behave differently, but they are both semi-heavy. One would therefore expect NIG to model skewness rather well, at least in cases where the tails are not too heavy.

3.2. Other skew Student's t-distributions

There are several definitions presented in the literature that can be regarded as competing skew Student's t-distributions. In this section review three of the most popular alternatives. For simplicity, we only give the central, non-scaled versions.

A first alternative, is to skew the symmetric Student's t-distribution by continuously piecing together two differently scaled halves of the symmetric base distribution, see e.g. Fernandez and Steel (1998). The density is on the form

$$f(x) = \frac{2\beta}{1+\beta^2}\left[t_\nu(\beta\,x)\,I(x<0) + t_\nu\left(\frac{x}{\beta}\right)\,I(x \geq 0)\right],$$

where $I(\cdot)$ is the indicator function, $\beta > 0$, and $t_\nu(\cdot)$ is the density of the standard Student's t-distribution with ν degrees of freedom. When $\beta = 1$, f reduces to the standard Student's t-distribution with ν degrees of freedom. The tail behaviour is that of the $t_\nu(\cdot)$ distribution, i.e.

$$f_x(x) \sim \text{const}|x|^{-\nu-1} \qquad \text{as } x \to \pm\infty.$$

A second alternative is the skew Student's t-distribution based on order statistics, recently introduced by Jones and Faddy (2003). Its density is given by

$$f(x) = \frac{1}{B(\alpha,\beta)(\alpha+\beta)^{\frac{1}{2}} 2^{\alpha+\beta-1}} \left(1 + \frac{x}{\sqrt{\alpha+\beta+x^2}}\right)^{\alpha+\frac{1}{2}} \left(1 - \frac{x}{\sqrt{\alpha+\beta+x^2}}\right)^{\beta+\frac{1}{2}},$$

where $B(\cdot,\cdot)$ denotes the beta function, and $\alpha, \beta > 0$. When $\alpha = \beta$, f corresponds to the standard Student's t-distribution with 2α degrees of freedom. When $\alpha < \beta$ or $\alpha > \beta$, f is negatively or positively skewed respectively. In the tails, the density behaves as

$$f(x) \sim \text{const}|x|^{-2\alpha-1} \qquad \text{as } x \to -\infty$$

and

$$f(x) \sim \text{const}|x|^{-2\beta-1} \qquad \text{as } x \to +\infty.$$

A third alternative is the skew Student's t-distribution proposed by Azzalini and Capitanio (2003) (which coincides with the skew t-distribution of Branco and Dey (2001)), having a density on the form

$$f_x(x) = t_\nu\,(x)\; 2\,T_{\nu+1}\left(\beta\, x\,\sqrt{\frac{\nu+1}{x^2+\nu}}\right), \tag{2}$$

where $t_\nu(\cdot)$ is the density of the standard Student's t-distribution with ν degrees of freedom and $T_{\nu+1}(\cdot)$ is the distribution function of the standard Student's t-distribution with $\nu+1$ degrees of freedom. When $\beta = 0$, Equation (2) is reduced to the standard Student's t-distribution with ν degrees of freedom. The tail behaviour is that of the $t_\nu(\cdot)$ distribution, i.e.

$$f_x(x) \sim \text{const}|x|^{-\nu-1} \qquad \text{as } x \to \pm\infty.$$

Note that for the closely related density (Jones and Faddy, 2003), given by

$$f_x(x) = t_\nu\,(x)\; 2\,T_\nu\,(\beta\, x)\,,$$

the tail behaviour is slightly different. The heaviest tail decays as

$$f_x(x) \sim \text{const}|x|^{-\nu-1} \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad -\infty, \\ \beta > 0 & \text{and} \quad x \to \quad +\infty, \end{cases}$$

and the lightest as

$$f_x(x) \sim \text{const}|x|^{-2\nu-1} \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad +\infty, \\ \beta > 0 & \text{and} \quad x \to \quad -\infty. \end{cases}$$

Hence, all the three versions of the skew t-distribution presented in this section have two tails behaving as polynomials. This means that they should fit heavy-tailed data well, but they may not handle substantial skewness.

4. The GH skew Student's t-distribution

The GH skew Student's t-distribution is a limiting case of the GH distribution and we find it appropriate to start with a short description of the latter before we give the definition of the first. The univariate GH distribution can be parameterised in several ways. We follow Prause (1999), and let

$$f_x(x) = \frac{(\alpha^2-\beta^2)^{\lambda/2}\, K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\, \exp\left(\beta(x-\mu)\right)}{\sqrt{2\,\pi}\alpha^{\lambda-1/2}\delta^\lambda\, K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)\,\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{1/2-\lambda}}. \tag{3}$$

In the above expression, K_j is the modified Bessel function of the third kind of order j (Abramowitz and Stegun, 1972) and the parameters must fulfill the conditions

$$\begin{aligned} \delta \geq 0, |\beta| < \alpha \quad & \text{if } \lambda > 0 \\ \delta > 0, |\beta| < \alpha \quad & \text{if } \lambda = 0 \\ \delta > 0, |\beta| \leq \alpha \quad & \text{if } \lambda < 0. \end{aligned} \tag{4}$$

$$\tag{5}$$

It can be shown that in the tails, the GH distribution behaves as

$$f_x(x) \sim \text{const}|x|^{\lambda-1}\, \exp\left(-\alpha|x| + \beta\, x\right) \qquad \text{as } x \to \pm\infty, \tag{6}$$

for all values of λ. Hence, as long as $|\beta| \neq \alpha$, the GH distribution has two semi-heavy tails.

The GH skew Student's t-distribution may be represented as a normal variance-mean mixture with the Generalised Inverse Gaussian (GIG) distribution as a mixing distribution (Barndorff-Nielsen and Blæsild, 1981), where the GIG distribution has the density (Barndorff-Nielsen, 1977)

$$f(z;\lambda,\delta,\gamma) = \left(\frac{\gamma}{\delta}\right)^\lambda \frac{z^{\lambda-1}}{2\,K_\lambda(\gamma\,\delta)}\,\exp\left\{-\frac{1}{2}(\delta^2\,z^{-1}+\gamma^2\,z)\right\}.$$

This means that a generalised hyperbolic variable X can be represented as

$$X = \mu + \beta\, Z + \sqrt{Z}\, Y, \tag{7}$$

where $Y \sim N(0,1)$, $Z \sim GIG\,(\lambda,\delta,\gamma)$, with Y and Z independent and $\gamma = \sqrt{\alpha^2-\beta^2}$. It follows from Equation (7) that $X|Z = z \sim N(\mu + \beta\, z, z)$.

Letting $\lambda = -\nu/2$ and $\alpha \to |\beta|$ in Equation (3), we obtain the GH skew Student's t-distribution. Its density is given by

$$f_x(x) = \frac{2^{\frac{1-\nu}{2}}\,\delta^\nu\,|\beta|^{\frac{\nu+1}{2}}\,K_{\frac{\nu+1}{2}}\left(\sqrt{\beta^2\,(\delta^2+(x-\mu)^2)}\right)\,\exp\left(\beta\,(x-\mu)\right)}{\Gamma\left(\frac{\nu}{2}\right)\,\sqrt{\pi}\,\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{\frac{\nu+1}{2}}}, \qquad \beta \neq 0, \tag{8}$$

and

$$f_x(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\delta\,\Gamma(\frac{\nu}{2})}\left[1+\frac{(x-\mu)^2}{\delta^2}\right]^{-(\nu+1)/2}, \qquad \beta = 0. \tag{9}$$

$f_x(x)$ in Equation (9) can be recognised as the density of a non-central (scaled) Student's t-distribution with ν degrees of freedom.

The mean and variance of a GH skew Student's t-distributed random variate X are

$$\mathrm{E}(X) = \mu + \frac{\beta\,\delta^2}{\nu - 2} \tag{10}$$

and

$$\mathrm{Var}(X) = \frac{2\,\beta^2\,\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}. \tag{11}$$

The variance is only finite when $\nu > 4$, as opposed to the symmetric Student's t-distribution which only requires $\nu > 2$. The derivation of the skewness and kurtosis is relatively straightforward (but cumbersome!) due to the normal mixture structure of the distribution. These are given by

$$s = \frac{2(\nu-4)^{1/2}\beta\delta}{\left[2\beta^2\delta^2 + (\nu-2)(\nu-4)\right]^{3/2}}\left[3(\nu-2) + \frac{8\beta^2\delta^2}{\nu-6}\right] \tag{12}$$

and

$$k = \frac{6}{\left[2\,\beta^2\,\delta^2 + (\nu-2)(\nu-4)\right]^2}\left[(\nu-2)^2(\nu-4) + \frac{16\,\beta^2\delta^2(\nu-2)(\nu-4)}{\nu-6} + \frac{8\,\beta^4\delta^4(5\,\nu-22)}{(\nu-6)(\nu-8)}\right]. \tag{13}$$

The skewness and kurtosis do not exist when $\nu \leq 6$, and $\nu \leq 8$, respectively.

It follows from Equation (6) that in the tails, the skew Student's t-density is given by

$$f_x(x) \sim \mathrm{const}|x|^{-\nu/2-1}\exp\left(-|\beta||x| + \beta\,x\right) \qquad \text{as } x \to \pm\infty.$$

Hence, the heaviest tail decays as

$$f_x(x) \sim \mathrm{const}|x|^{-\nu/2-1} \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad -\infty, \\ \beta > 0 & \text{and} \quad x \to \quad +\infty, \end{cases}$$

and the lightest as

$$f_x(x) \sim \mathrm{const}|x|^{-\nu/2-1}\exp\left(-2\,|\beta|\,|x|\right) \qquad \text{when} \begin{cases} \beta < 0 & \text{and} \quad x \to \quad +\infty, \\ \beta > 0 & \text{and} \quad x \to \quad -\infty. \end{cases}$$

Thus, the GH skew Student's t-distribution has one heavy, and one semi-heavy tail. It is the only member of the GH family of distributions having this property. This can be seen as follows. From Equation (6) we have that the only way of obtaining one heavy and one semi-heavy tail, is to let $\alpha \to |\beta|$. According to the parameter conditions given in Equation (4), this requires $\lambda < 0$. Finally, if $\lambda < 0$, and $\alpha = |\beta|$ we obtain the GH skew Student's t-distribution independent of the magnitude of $\lambda < 0$.

The tail behaviour of the GH skew Student's t-distribution also distinguishes it from the alternative skew Student's t-distributions reviewed in Section 3.2, which all have two tails with polynomial behaviour. This makes it unique for modelling substantially skew and heavy-tailed data.

5. Parameter estimation using the EM-algorithm

The parameters of the GH skew Student's t-distribution can be estimated using maximum likelihood estimation. The maximisation problem becomes easier if one exploits its normal variance-mean mixture structure. Then, one may apply the EM-algorithm (Dempster et al., 1977), which is a powerful algorithm for ML estimation on data containing missing values. It is particularly suitable for mixture distributions, since the mixing operation in a sense produces missing data; the mixing variables. In what follows, we will provide an the EM-algorithm for estimating the parameters of the GH skew Student's t-distribution.

We assume that the true data are made of an observable part X and an unobservable part Z. The EM-algorithm consists in iterating two steps; the expectation step (E-step) and the maximization step (M-step). In the E-step, one computes the expectation of the unobservable part, given the current values of the parameters, and in the M-step the likelihood of $f_x(x,z) = f_{x|z}(x|z)\, f_z(z)$ is computed using the expectations from the E-step.

E-step The E-step consists in computing the conditional expectation of the sufficient statistics of the GIG distribution, which are Z, Z^{-1} and $\log Z$. It can be shown (Barndorff-Nielsen, 1997) that for the GH distribution, $Z|X \sim GIG(\lambda - \frac{1}{2}, \sqrt{\delta^2 + (x-\mu)^2}, \alpha)$. Hence, in the GH skew Student's t-case, $Z|X \sim GIG(-\frac{(\nu+1)}{2}, \sqrt{\delta^2 + (x-\mu)^2}, |\beta|)$. The moments of the $GIG(\lambda, \delta, \gamma)$ distribution are given by (Karlis, 2002)

$$\mathrm{E}(Z^r) = \left(\frac{\delta}{\gamma}\right)^r \frac{K_{\lambda+r}(\delta\,\gamma)}{K_\lambda(\delta\,\gamma)}.$$

Define $q(x_i) = \sqrt{\delta^2 + (x_i - \mu)^2}$. Then, for the GH skew Student's t-distribution

$$\xi_i = \mathrm{E}(Z_i|X_i = x_i) = \frac{q(x_i)}{|\beta|} \frac{K_{\frac{1-\nu}{2}}\left(|\beta|\, q(x_i)\right)}{K_{\frac{\nu+1}{2}}\left(|\beta|\, q(x_i)\right)}$$

and

$$\rho_i = \mathrm{E}(Z_i^{-1}|X_i = x_i) = \frac{|\beta|}{q(x_i)} \frac{K_{\frac{\nu+3}{2}}\left(|\beta|\, q(x_i)\right)}{K_{\frac{\nu+1}{2}}\left(|\beta|\, q(x_i)\right)}.$$

Further, we have

$$\chi_i = \mathrm{E}(\log Z_i|X_i = x_i) = \log\left(\frac{q(x_i)}{|\beta|}\right) + \frac{1}{K_{\frac{\nu+1}{2}}\left(|\beta|\, q(x_i)\right)} \frac{\partial K_{\frac{\nu+1}{2}}\left(|\beta|\, q(x_i)\right)}{\partial(\frac{\nu+1}{2})},$$

which follows from (Mencia and Sentana, 2004)

$$\mathrm{E}(\log Z) = \left.\frac{\partial\,\mathrm{E}(Z^r)}{\partial\, r}\right|_{r=0} = \log\left(\frac{\delta}{\gamma}\right) + \frac{1}{K_\lambda(\delta\,\gamma)} \frac{\partial}{\partial\,\lambda} K_\lambda(\delta\,\gamma).$$

The derivatives of the modified Bessel function $K_\lambda(\cdot)$ of the third kind with respect to the order λ may be computed using the analytical formulas provided in (Mencia and Sentana, 2004). However, these are very complex, such that a numerical approximation may be preferable.

M-step In the M-step, one computes the parameter estimates resulting from maximizing the likelihood of $f_x(x,z) = f_{x|z}(x|z)\, f_z(z)$, using the pseudo values ρ_i, ξ_i, and χ_i from the M-step. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\bar{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$. At the kth iteration of the algorithm, the estimates for β and μ are updated as

$$\beta^{(k+1)} = \frac{\sum_{i=1}^n x_i\, \rho_i - \bar{x}\sum_{i=1}^n \rho_i}{n - \bar{\xi}\sum_{i=1}^n \rho_i} \tag{14}$$

$$\mu^{(k+1)} = \bar{x} - \beta^{(k+1)}\, \bar{\xi}. \tag{15}$$

The parameter ν is given as the solution of the following equation

$$\log\frac{n}{2} - \log\left(\sum_{i=1}^n \rho_i\right) - \frac{1}{n}\sum_{i=1}^n \chi_i = \Psi\left(\frac{\nu^{(k+1)}}{2}\right) - \log\nu^{(k+1)},$$

where $\Psi(\cdot)$ is the Digamma function. Finally,

$$\delta^{(k+1)} = \sqrt{\frac{n\,\nu^{(k+1)}}{\sum_{i=1}^n \rho_i}}.$$

Initial values Convergence of the algorithm to the ML estimates is guaranteed since it is a standard EM-algorithm. However, it may be caught in a local maximum, and it is important to choose appropriate initial values. We use the moment estimates. Let $\bar{m}_1$, $\bar{m}_2$, $\bar{m}_3$ and $\bar{m}_4$ be the sample mean, standard deviation, skewness, and kurtosis of the data, respectively. Then, according to Equations (10)-(13), the moment estimates for μ, β and δ are given by

$$\tilde{\mu} = \bar{m}_1 - \frac{\tilde{\beta}\,\tilde{\delta}^2}{\tilde{\nu}-2}$$

$$\tilde{\beta} = \operatorname{sign}(\bar{m}_3)\cdot\frac{(\tilde{\nu}-2)^{1/2}(\tilde{\nu}-4)^{1/2}\left[\bar{m}_2(\tilde{\nu}-2)-\tilde{\delta}^2\right]^{1/2}}{2^{1/2}\tilde{\delta}^2}$$

$$\tilde{\delta}^2 = \frac{6(\tilde{\nu}-2)^2(\tilde{\nu}-4)\bar{m}_2}{3\tilde{\nu}^2-2\tilde{\nu}-32}\left(1-\sqrt{1-\frac{(3\tilde{\nu}^2-2\tilde{\nu}-32)\,(12(5\tilde{\nu}-22)-(\tilde{\nu}-6)(\tilde{\nu}-8)\bar{m}_4)}{216(\tilde{\nu}-2)^2(\tilde{\nu}-4)}}\right).$$

The moment estimate for ν is the solution of the equation

$$[4-6(\tilde{\nu}+2)(\tilde{\nu}-2)*\kappa]\,\sqrt{2}\sqrt{\tilde{\nu}-4}\sqrt{1-6(\tilde{\nu}-2)(\tilde{\nu}-4)*\kappa} - \bar{m}_3\,(\tilde{\nu}-6) = 0,$$

where κ is given by

$$\kappa = \frac{1}{3\tilde{\nu}^2-2\tilde{\nu}-32}\cdot\left(1-\sqrt{1-\frac{(3\tilde{\nu}^2-2\tilde{\nu}-32)\,(12(5\tilde{\nu}-22)-(\tilde{\nu}-6)(\tilde{\nu}-8)\bar{m}_4)}{216(\tilde{\nu}-2)^2(\tilde{\nu}-4)}}\right).$$

6. Numerical examples

We have fitted the GH skew Student's t-distribution to the four log return series from Section 2. Moreover, we have compared the fit of the GH skew Student's t-distribution

Table 1. Parameter estimates resulting when fitting the GH skew Student's t-distribution to each risk factor.

Risk factor	μ	δ	β	ν
Norwegian stocks	0.00193	0.02102	-14.06736	4.78729
International bonds	0.00244	0.00798	-511.90690	17.42587
NOK/EUR exchange rate	-0.00082	0.00713	60.11458	6.02776
EURIBOR 5-year	-0.00258	0.02028	17.61363	4.84912

Table 2. Parameter estimates resulting when fitting the NIG distribution to each risk factor.

Risk factor	μ	δ	β	α
Norwegian stocks	0.00190	0.01350	-13.87510	87.93290
International bonds	0.00210	0.00490	-424.49800	1243.93400
NOK/EUR exchange rate	-0.00080	0.00450	57.04440	359.80620
EURIBOR 5-year	-0.00250	0.01280	17.31610	91.92160

to the fit of the NIG distribution, and the skew Student's t-distribution of Azzalini and Capitanio presented in Section 3.2. The latter is hereafter denoted Azzalini's skew Student's t-distribution. We focus on the tails, which usually are the most important parts of the distribution in financial applications.

6.1. Parameter estimates

Tables 1-3 show the parameter estimates resulting from fitting the GH skew Student's t-distribution, the NIG distribution and Azzalini's skew Student's t-distribution to the four data sets. For the GH skew Student's t-distribution we used the EM-algorithm described in Section 5 (it can be programmed in any statistical package supporting Bessel functions with fractional order, e.g. R) and stopped iterating when the maximum absolute relative difference in any parameter was smaller than 0.0001 in two successive iterations. The moment estimates were found to be very good initial values for the EM-algorithm.

For the NIG distribution, we used the EM-algorithm described in Karlis (2002), with the moment estimates as initial values. The convergence criterion was the same as for the GH skew Student's t-distribution. For Azzalini's skew Student's t-distribution, we used the numerical maximum likelihood estimation scheme given in Azzalini and Capitanio (2003), which has been implemented in the `sn`-package for R. The CPU time pr. iteration was approximately 0.01 for NIG, approximately 0.02 for GH skew Student's t and approximately 0.04 for Azzalini's skew Student's t, whereas the number of iterations needed until convergence was slightly larger for the GH skew Student's t than for the two others.

Table 3. Parameter estimates resulting when fitting Azzalini's skew Student's t-distribution to each risk factor.

Risk factor	μ	δ	β	ν
Norwegian stocks	0.00388	0.01016	-0.46150	4.65220
International bonds	0.00110	0.00206	-0.45124	9.19201
NOK/EUR exchange rate	-0.00043	0.00289	0.10002	5.17002
EURIBOR 5-year	-0.00072	0.00916	0.01551	4.47262

6.2. Goodness of fit

Since our main interest is the tails of the distributions, we use graphical logarithmic left and right tail tests for examining the fit in the tails. The graphical tests were performed as follows. Let $\hat{F}(x)$ denote the estimated cumulative distribution function of the fitted distribution, computed by numerical integration, and $(X_{(1)}, ..., X_{(N)})$ the order statistic of the historical data. A plot of $\log(\hat{F}(X_{(t)}))$ against $X_{(t)}$ superimposed onto a plot of $\log(1/(N+1))$ against $X_{(t)}$ shows the left tail fit for the fitted distribution, and a plot of $\log(1-\hat{F}(X_{(t)}))$ against $X_{(t)}$, superimposed onto a plot of $\log((N+1-t)/(N+1))$, the right tail fit.

Figures 2-5 show the plots. The upper panel in each figure shows the left tail fit, and the lower panel the right tail fit. The circles corresponds to the empirical data, the light-blue line corresponds to the GH skew Student's t-distribution, the red to the NIG distribution and the dark-blue to Azzalini's skew Student's t-distribution. The green line corresponds to the Gaussian distribution, which is included as a reference. All distributions, except the Gaussian, fit the Norwegian stock return distribution quite well. For the international bond return distribution NIG provides almost as good fit as the GH skew Student's t. Azzalini's skew Student's distribution on the other hand, slightly underestimates the left tail and overestimates the right. For the NOK/EUR exchange rate data, the NIG distribution underestimates the right tail, while Azzalini's skew Student's t-distribution on the other hand, underestimates the right tail and overestimates the left. The GH skew Student's t-distribution fit both tails better than the two other distributions. Finally, for the European 5-year interest data, the NIG distribution underestimates the right tail and Azzalini's skew Student's distribution underestimates the right tail and overestimates the left. In this case also, the GH skew Student's t-distribution fit both tails quite well. Hence, the GH skew Student's t-distribution provides the best overall fit for all four financial market variables.

7. Application to risk estimation

In this section we use the estimated distributions from Section 6 to determine the risk for long and short trading positions of the NOK/EUR exchange rate. For the first kind of positions, the risk is connected to potential drops in the asset price. In the second case, the trader looses money when the price increases. Correspondingly, one focuses on the left side of the return distribution for long positions and on the right side for short ones.

To measure risk, we use VaR and expected shortfall (ES) (Artzner et al., 1997) at different confidence levels. The reason for including ES is that VaR only measures a quantile of the distribution, and hence ignores important information regarding the tails of the distribution beyond this quantile. ES, defined as the conditional expectation of the return, given that it is beyond the VaR level, describes the tail risk better.

We define a test period from 09.07.2003 to 21.01.2005, corresponding to 387 observations. For each day in the test period, we predict the 1-day VaR and ES at levels 0.005, 0.01, 0.05, 0.95, 0.99, 0.995. The first three levels are used to measure the risk of long positions and the last three of short. We use the likelihood ratio statistic by Kupiec (1995) to verify whether the VaR predictions are correct. The method consists in calculating the number of times x^{α} the observed returns fall below (long positions) or above (short positions) the VaR estimate at level α, i.e. $R_t < \widehat{VaR}^{\alpha}$ or $R_t > \widehat{VaR}^{\alpha}$, and comparing it to the expected number of violations. The null hypothesis is that the expected proportion of violations is

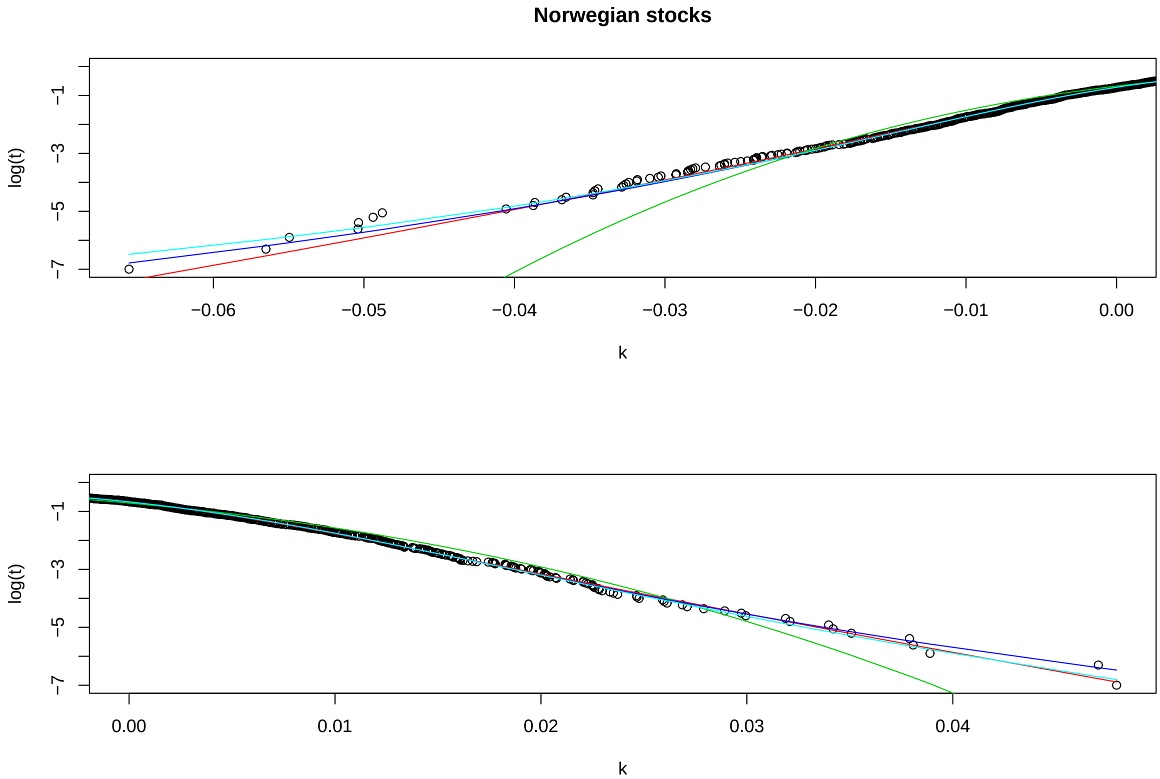

Figure 2. Left and right tail plots for Norwegian stocks.

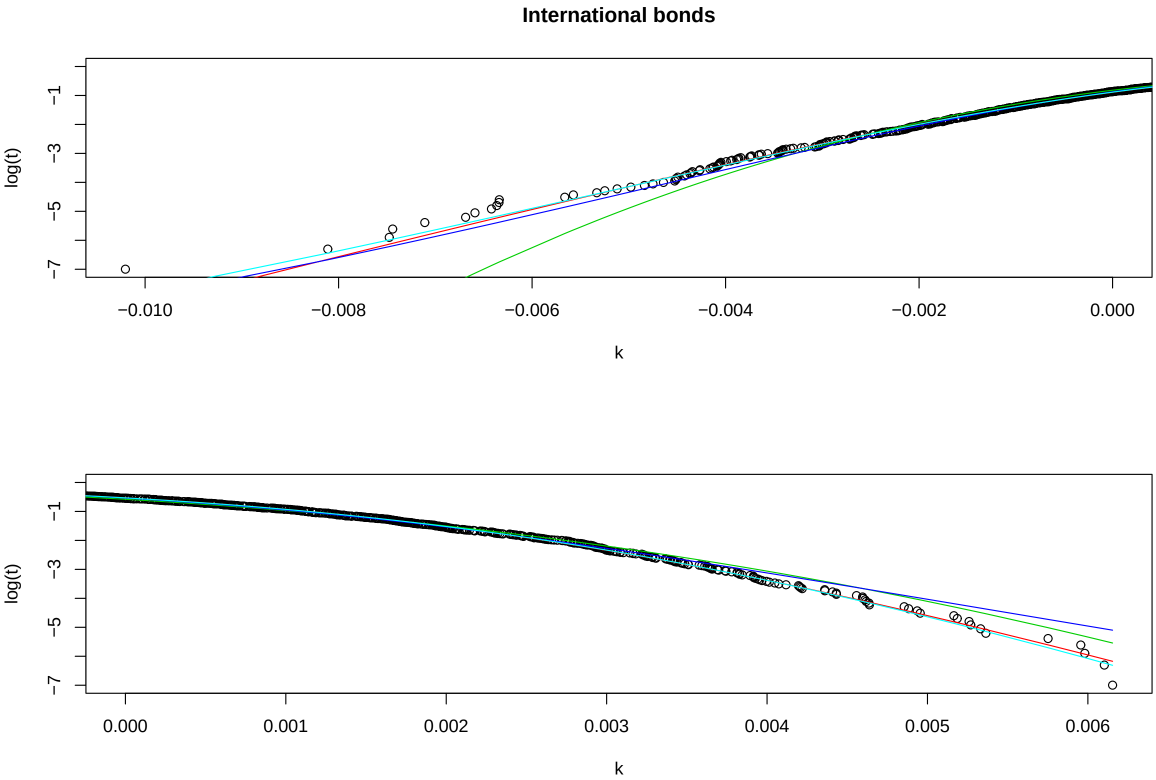

Figure 3. Left and right tail plots for international bonds.

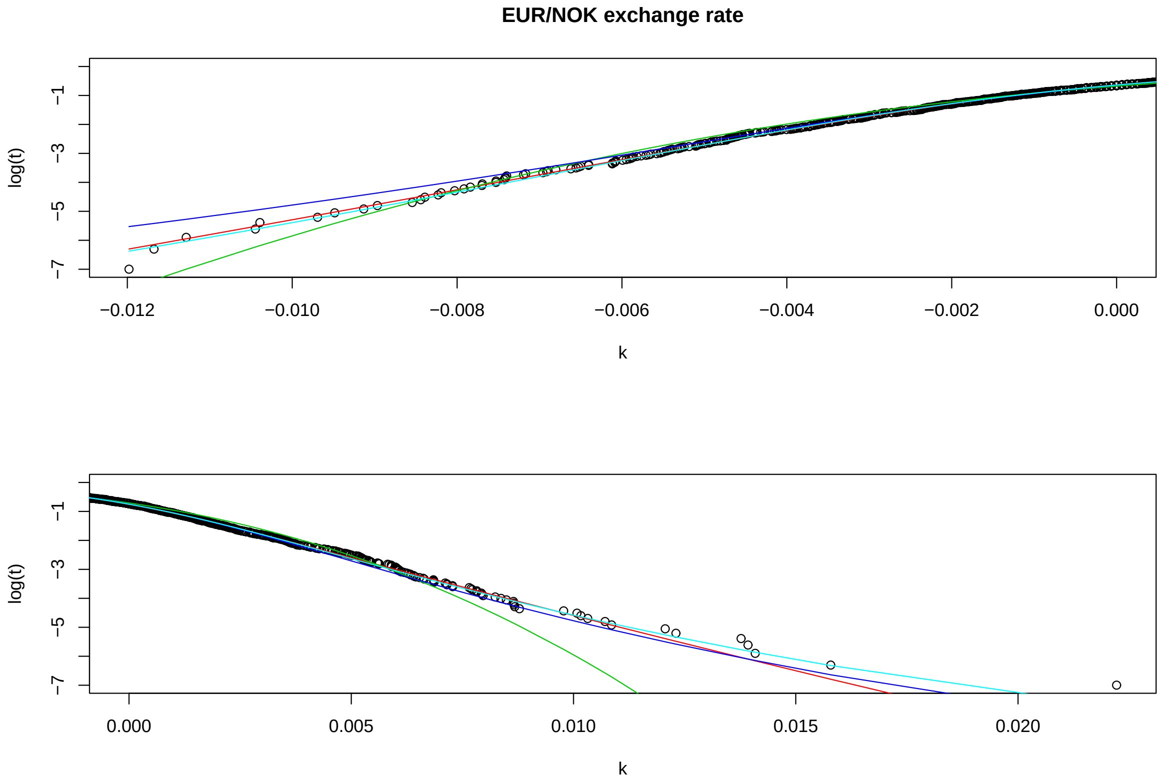

Figure 4. Left and right tail plots for NOK/EUR exchange rate.

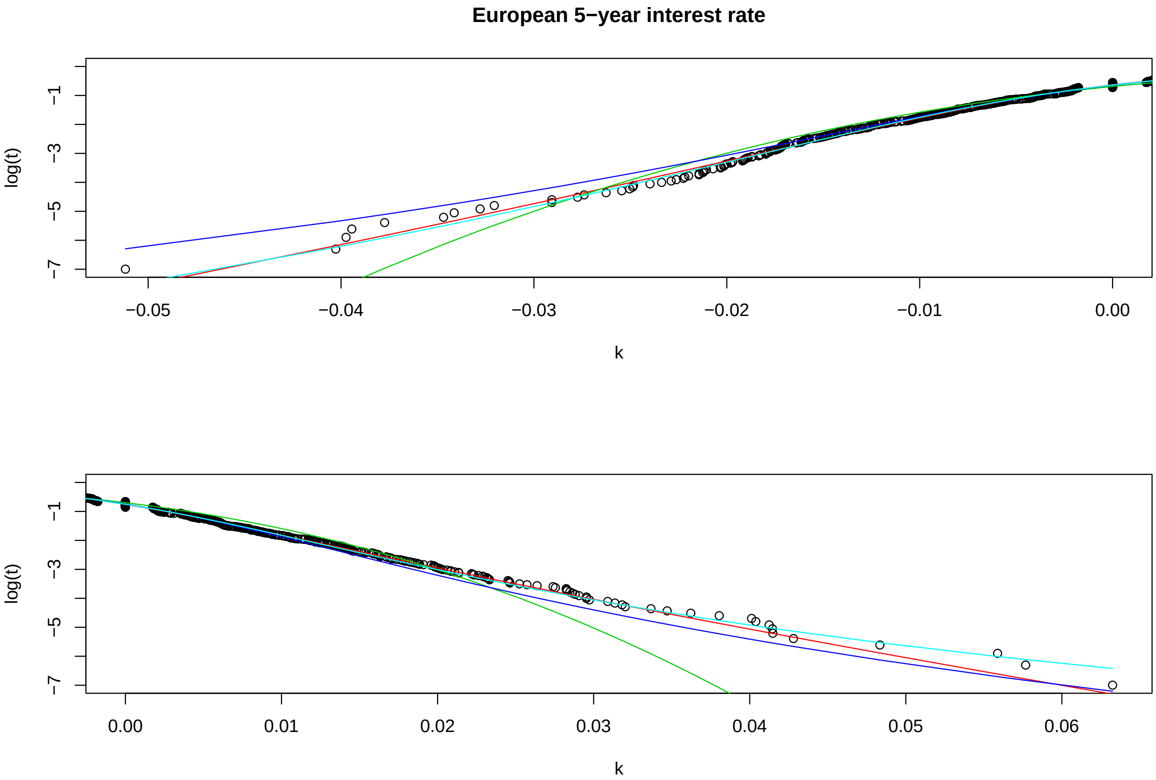

Figure 5. Left and right tail plots for European 5-year interest rate.

Table 4. Number of violations of VaR for each distribution and each level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's t	2	5	22	19	6	3
NIG	2	5	21	18	6	6
Azzalini's skew Student's t	1	3	19	20	9	6

Table 5. P-values for the Kupiec test for each distribution and each level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's t	0.96	0.58	0.54	0.93	0.31	0.48
NIG	0.96	0.58	0.70	0.75	0.31	**0.02**
Azzalini's skew Student's t	0.46	0.64	0.93	0.88	**0.03**	**0.02**

equal to α. Under the null hypothesis, the likelihood ratio statistic given by

$$2\ln\left(\left(\frac{x^\alpha}{N}\right)^{x^\alpha}\left(1-\frac{x^\alpha}{N}\right)^{N-x^\alpha}\right)-2\ln\left(\alpha^{x^\alpha}(1-\alpha)^{N-x^\alpha}\right),$$

where N is the length of the sample, is asymptotically distributed as $\chi^2(1)$.

Table 4 shows the observed number of violations of VaR for each distribution and each level. The corresponding p-values are shown in Table 5. If we use a 5% level for the Kupiec test, the null hypothesis is rejected twice for Azzalini's skew Student's t-distribution, once for the NIG distribution and never for the GH skew Student's t-distribution.

For backtesting the predicted ES-value for confidence level α, $\widehat{ES}^\alpha$, we use the measure proposed by Embrechts et al. (2004). It is given by

$$D^\alpha = (|D_1^\alpha| + |D_2^\alpha|)/2,$$

where

$$D_1^\alpha = \frac{1}{x^\alpha}\sum_{t\in\kappa^\alpha}(R_t - \widehat{ES}^\alpha).$$

Here κ^α is the set of time points for which a violation of $\widehat{VaR}^\alpha$ occur. Define $\delta_t^\alpha = R_t - \widehat{ES}^\alpha$. Further, let y^α be the number of times δ_t^α is less than (long positions) or greater than (short positions) its empirical α-quantile, and τ^α the set of time points for which this happens. Then,

$$D_2^\alpha = \frac{1}{y^\alpha}\sum_{t\in\tau^\alpha}(R_t - \widehat{ES}^\alpha).$$

D_1^α is the standard backtesting measure for expected shortfall estimates. Its weakness is that it depends strongly on the VaR estimates without adequately reflecting the goodness/badness of these values. To correct for this, it is combined with a penalty D_2^α. A good estimation of expected shortfall will lead to a low value of D^α. In Table 6, we show the D^α-values for each distribution and level. As can be seen from the table, the GH skew Student's t-distribution gives lower values than the two other distributions in 17 out of 18 cases. Hence, it is superior to the other distributions in predicting expected shortfall for our test data.

Table 6. Backtest-measure of expected shortfall predictions for each distribution and each level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's t	0.0115	0.0005	0.0002	0.0005	0.0005	0.0006
NIG	0.0136	0.0007	0.0002	0.0007	0.0014	0.0014
Azzalini's skew Student's t	0.0310	0.0020	0.0004	0.0012	0.0014	0.0023

8. Conclusions

In this paper we have argued for a special case of the generalised hyperbolic distribution that we denote the GH skew Student's t-distribution. This distribution has the important property that one tail is determined by polynomial, and the other by exponential behaviour. This makes it different from other skew Student's t-distributions proposed in the literature, that have two heavy tails. It is also the only member of the GH family of distributions having this property. Moreover, is it almost as analytically tractable as the NIG distribution, and due to the normal mean-variance mixture structure, we may apply the powerful EM-algorithm for parameter estimation. Hence, the GH skew Student's t-distribution is very useful for financial applications.

We have fitted the GH skew Student's t-distribution to four types of financial market variables. For heavy-tailed data it provides better overall fit than the more well-known NIG distribution. If the data in addition are very skewed, it is also superior to the skew Student's t-distribution provided by Azzalini and Capitanio (2003). We have also predicted out-of-sample 1-day VaR and expected shortfall at levels 0.005, 0.01, 0.05, 0.95, 0.99, 0.995 for the NOK/EUR exchange rate. Backtesting shows that the GH skew Student's t-distribution outperforms the NIG and skew Student's t-distribution provided by Azzalini and Capitanio (2003) when expected shortfall is used as a risk measure, and is also slightly better at predicting VaR.

Acknowledgements

This work is sponsored by the Norwegian fund Finansmarkedsfondet. The authors acknowledge the support and guidance of colleagues at the Norwegian Computing Center, in particular Professor Håvard Rue.

References

Abramowitz, M. and I. A. Stegun (1972). *Handbook of Mathematical Function.* New York: Dover.

Andersson, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business and Economic Statistics 19*, 44–54.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heat (1997). Thinking coherently. *Risk 10*(11), 68–71.

Azzalini, A. and A. Capitanio (2003). Distributions generated by pertubation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society B 65*, 579–602.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A 353*, 409–419.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*, 1–13.

Barndorff-Nielsen, O. E. and P. Blæsild (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. *Statistical Distributions in Scientific Work 4*, 19–44.

Barndorff-Nielsen, O. E. and N. Shepard (2001). Normal modified stable processes. *Theory of probability and Mathematical Statistics 65*, 1–19.

Bauwens, L. and S. Laurent (2002). A new class of multivariate skew densities, with application to GARCH models. CORE Discussion Paper 20. Forthcoming in Journal of Business & Economic Statistics.

Bølviken, E. and F. E. Benth (2000). Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution. In *Proceedings of the AFIR 2000 Colloquium, Tromsø, Norway*, pp. 87–98.

Branco, M. D. and D. K. Dey (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis 79*, 99–113.

Demarta, S. and A. J. McNeil (2004). The t copula and related copulas. Technical report, ETH Zurich. Forthcoming in International Statical Review.

Dempster, A. P., N. M. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data using the EM algorithm. *Journal Roy. Statist. Soc. B 39*, 1–38.

Eberlein, E. and U. Keller (1995). Hyberbolic distributions in finance. *Bernoulli, 1*(3), 281–299.

Embrechts, P., R. Kaufmann, and P. Patie (2004). Strategic long-term financial risks: Single risk factors. To appear in A special issue of Computational Optimization and Applications.

Fernandez, C. and M. Steel (1998). On bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association 93*, 359–371.

Forsberg, L. and T. Bollerslev (2002). Bridging the gap between the distribution of realized (ecu) volatility and ARCH modeling (of the euro): The GARCH-NIG model. *Journal of Applied Econometrics 17*(5), 535–548.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review 35*, 705–730.

Jensen, M. B. and A. Lunde (2001). The NIG-S & ARCH model: A fat-tailed stochastic, and autoregressive conditional heteroscedastic volatility model. *Econometrics Journal 4*, 319–342.

Jones, M. C. and M. J. Faddy (2003). A skew extension of the t distribution, with applications. *J. Roy. Statist. Soc., Ser. B 65*(2), 159–174.

Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters 57*, 43–52.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of derivatives 2*, 173–184.

Lillestøl, J. (2000). Risk analysis and the NIG distribution. *Journal of Risk 2*(4), 41–56.

Mencia, F. J. and E. Sentana (2004). Estimation and testing of dynamic models with generalised hyperbolic innovations. CMFI Working Paper 0411, Madrid, Spain.

Patton, A. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics 2*(1), 130–168.

Prause, K. (1997). Modelling financial data using generalized hyperbolic distributions. FDM preprint 48, University of Freiburg.

Prause, K. (1999). The generalized hyperbolic models: Estimation, financial derivatives and risk measurement. PhD Thesis, Mathematics Faculty, University of Freiburg.

Rydberg, T. H. (1997). The normal inverse Gaussian Levy process: Simulation and approximation. *Commun. Statist.-Stochastic Models 34*, 887–910.

Sahu, S. K., D. K. Dey, and M. D. Branco (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *The Canadian Journal of Statistics 31*, 129–150.

Venter, J. H. and P. J. de Jongh (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk 4*(2), 1–24.