Journal of Financial Econometrics, 2006, Vol. 4, No. 2, 275–309

The Generalized Hyperbolic Skew Student's *t*-Distribution

KJERSTI AAS
Norwegian Computing Center

INGRID HOBÆK HAFF
Norwegian Computing Center

ABSTRACT

In this article we argue for a special case of the generalized hyperbolic (GH) family that we denote as the GH skew Student's *t*-distribution. This distribution has the important property that one tail has polynomial and the other exponential behavior. Further, it is the only subclass of the GH family of distributions 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 behavior has not been addressed. This article presents empirical evidence of exponential/polynomial tail behavior in skew financial data, and demonstrates the superiority of the GH skew Student's *t*-distribution with respect to data fit compared with some of its competitors. Through VaR and expected shortfall calculations we show why the exponential/polynomial tail behavior is important in practice.

KEYWORDS: EM algorithm, generalized hyperbolic distribution, NIG, skew probability distribution, skew Student's *t*

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 nonnormality. 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 semiheavy 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

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. We also thank two anonymous referees and René Garcia (the editor) for comments and suggestions that helped to improve the article. Address correspondence to Kjersti Aas, Norwegian Computing Center, Gaustadallen 23, P.O.Box 114, Blindern, NO-0314 Oslo, Norway, or e-mail: kjersti.aas@nr.no.

doi:10.1093/jjfinec/nbj006
Advance Access publication January 27, 2006

(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 return distribution [Eberlein and Keller (1995), Prause (1997), Rydberg (1997), Bølviken and Benth (2000), Lillestøl (2000)]. The two tails of the NIG distribution behave differently, but they are both semiheavy. 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 modeling skew and heavy-tailed data is the skew extensions to the Student's *t*-distribution. Hansen (1994) was the first to propose such a distribution for modeling financial returns. Since then, several other articles have studied different skew *t*-type distributions for financial and other applications [see, e.g., Fernandez and Steel (1998), Branco and Dey (2001), Bauwens and Laurent (2002), Venter and de Jongh (2002), Azzalini and Capitanio (2003), Jones and Faddy (2003), Sahu, Dey and Branco (2003), Patton (2004)]. All these distributions have two tails behaving as polynomials. This means that they fit heavy-tailed data well, but they do not handle substantial skewness. By substantial skewness we mean cases with one heavy tail and one nonheavy tail. Hence our definition relates to the relative fatness of the two tails of the density rather than some threshold for the skewness coefficient.

The NIG distribution is a subclass of the generalized hyperbolic (GH) distribution [Barndorff-Nielsen (1977)]. The GH distribution possesses a number of attractive properties; for example, it is closed under conditioning, marginalization, and affine transformations. It can be both symmetric and skew, and its tails are generally semiheavy. 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 estimate the parameter determining the subclass. The latter is due to the flatness of the GH likelihood function in this parameter [Prause (1999)].

In this article we argue for a special case of the GH 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), Demarta and McNeil (2004), and Mencia and Sentana (2004). However, it is not well known, and specifically its special tail behavior 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 behavior. 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, Laird, and Rubin (1977)], making it very useful for financial applications.

The remainder of this article is organized as follows. Section 4.1 presents empirical evidence for the exponential/polynomial tail behavior of skew financial data. Section 1 reviews other skew distributions with heavy or semiheavy tails. Section 2 provides the definition of the GH skew Student's t-distribution. Section 3 gives the details of the EM algorithm for the estimation of its parameters. In Section 4 we fit the GH skew Student's t-distribution to the financial market variables presented in Section 4.1 and compare the results with the fit of the alternative distributions presented in Section 1. The practical importance of the exponential/polynomial tail behavior of the GH skew Student's t-distribution is highlighted through VaR and expected shortfall calculations in Section 5. Finally, Section 6 contains some concluding remarks.

1 A REVIEW OF SKEW AND HEAVY-TAILED DISTRIBUTIONS

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

1.1 NIG

The NIG distribution is a GH distribution with $\lambda = -\frac{1}{2}$ (see Section 2 for a definition of the GH distribution). Its density is

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

where $\delta > 0$ and $0 < |\beta| \leq \alpha$. In the above expression, K_j is the modified Bessel function of the third kind of order j [Abramowitz and Stegun (1972)]. 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.

Utilizing the following property of the modified Bessel function [Abramowitz and Stegun (1972)],

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}}\exp(-x) \quad \text{for } x \to \pm\infty,$$

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

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

More specifically, the heaviest tail decays as

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

and the lightest as

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

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

1.2 The Skew Exponential Power Distribution

A second alternative is the skew exponential power distribution proposed by Ayebo and Kozubowski (2003). Its density is given by

$$f_x(x) = \frac{\nu\beta}{\delta\Gamma\left(\frac{1}{\nu}\right)(1+\beta^2)}\exp\left[-\left(\frac{\beta(x-\mu)}{\delta}\right)^{\nu} I(x>\mu) - \left(\frac{\mu-x}{\beta\delta}\right)^{\nu} I(x<\mu)\right] \text{if } x \neq \mu,$$

and

$$f_x(x) = \frac{\nu\beta}{\delta\Gamma\left(\frac{1}{\nu}\right)(1+\beta^2)} \quad \text{if } x = \mu.$$

When $\beta = 1$, the distribution is symmetric about μ. The parameter v controls the tails; smaller values of v correspond to heavier tails. For instance, $v \approx 2$ gives a near-normal tail, and letting $v = 1$, leads to the skew Laplace distribution. It can be shown [see, e.g., Snoussi and Idier (2005)] that an asymmetric Laplace distribution is obtained by a GH distribution with $\lambda = 1$ and $\delta \to 0$ (see Section 2 for a definition of the GH distribution). The tails of the skew exponential power distribution are not power laws. Thus all moments exist.

1.3 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 we review three of the most popular alternatives.

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); a very similar version was introduced independently by Hansen (1994)]. The density is of the form

$$f_x(x) = \frac{2\beta}{(\beta^2+1)}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}\sqrt{\nu}\delta}\left[1 + \frac{\left(\frac{x-\mu}{\delta}\right)^2}{\nu}\left(\frac{1}{\beta^2}I(x \geq \mu) + \beta^2 I(x<\mu)\right)\right]^{-\frac{\nu+1}{2}},$$

where $I(\cdot)$ is the indicator function and $\beta > 0$. When $\beta = 1$, f reduces to the noncentral Student's t-distribution with ν degrees of freedom, expectation μ, and variance $\delta^2 \frac{\nu}{\nu-2}$. The tail behavior is that of the $t_\nu(\cdot)$ distribution, that is,

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

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

$$f_x(x) = \frac{\Gamma(\nu+\beta)2^{1-\nu-\beta}}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu}{2}+\beta\right)\sqrt{\nu+\beta}\delta}\left(1+\frac{\left(\frac{x-\mu}{\delta}\right)}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\delta}\right)^2}}\right)^{\frac{\nu}{2}+\frac{1}{2}}$$
$$\left(1-\frac{\left(\frac{x-\mu}{\delta}\right)}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\delta}\right)^2}}\right)^{\frac{\nu}{2}+\beta+\frac{1}{2}},$$

where $\nu > 0$ and $\beta > -\nu/2$. When $\beta = 0$, f reduces to the noncentral Student's t-distribution with ν degrees of freedom, expectation μ, and variance $\delta^2 \frac{\nu}{\nu-2}$. When $\beta > 0$ or $\beta < 0$, f is negatively or positively skewed, respectively. In the tails, the density behaves as

$$f(x) \sim \text{const}|x|^{-\nu-1} \quad \text{as } x \to -\infty$$

and

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

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 of the form

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

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 noncentral Student's t-distribution with ν degrees of freedom, expectation μ, and variance $\delta^2 \frac{\nu}{\nu-2}$. The tail behavior is that of the $t_\nu(\cdot)$ distribution, that is,

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

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

$$f_x(x) = \frac{1}{\delta} t_\nu\Big(\frac{x-\mu}{\delta}\Big) 2T_\nu\Big(\beta\Big(\frac{x-\mu}{\delta}\Big)\Big),$$

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

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

and the lightest as

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

Hence all four 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 may not handle substantial skewness.

2 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 parameterized in several ways. We follow Prause (1999) and let

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

K_j is defined in Section 1.1, and the parameters must fulfill the conditions

$$\delta \geq 0, |\beta| < \alpha \quad \text{if } \lambda > 0$$
$$\delta > 0, |\beta| < \alpha \quad \text{if } \lambda = 0 \tag{4}$$

$$\delta > 0, |\beta| \leq \alpha \quad \text{if } \lambda < 0. \tag{5}$$

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

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

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

The GH skew Student's *t*-distribution may be represented as a normal variance-mean mixture with the generalized inverse gaussian (GIG) distribution as a mixing distribution [Barndorff-Nielsen and Blæsild (1981)], where the GIG distribution has the density [Barndorff-Nielsen (1977)]

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

This means that a GH 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

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

and

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

The density $f_X(x)$ in Equation (9) can be recognized as that of a noncentral Student's *t*-distribution with ν degrees of freedom, expectation μ, and variance $\delta^2/(\nu-2)$.

Figure 1 shows the density proposed in Equation (8) for a few different choices of parameter sets given in rows 1–4 of Table 1.

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

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

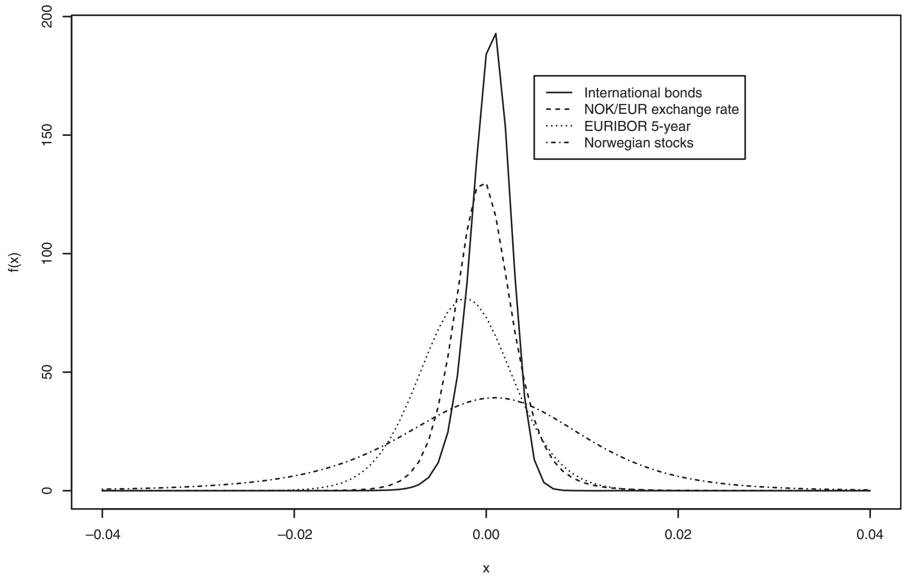

Figure 1 The density proposed in Equation (8) for the parameter combinations in Table 1.

Table 1 ML 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 (0.00062, 0.00393)	0.02102 (0.01720, 0.02800)	−14.06736 (−30.31, −4.00)	4.78729 (3.75, 7.20)
International bonds	0.00244 (0.00128, 0.00465)	0.00798 (0.00562, 0.01263)	−511.90690 (−1134.66, −214.81)	17.42587 (9.64, 41.09)
NOK/EUR exchange rate	−0.00082 (−0.00157, −0.00044)	0.00713 (0.00576, 0.01041)	60.11458 (30.34, 132.05)	6.02776 (4.60, 11.40)
EURIBOR 5-year	−0.00258 (−0.00527, −0.00127)	0.02028 (0.01563,0.02868)	17.61363 (5.92, 38.73)	4.84912 (3.68, 8.23)

95% confidence intervals are in parentheses.

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) (see Appendix A) 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

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

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

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

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

Hence the heaviest tail decays as

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

and the lightest as

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

Thus the GH skew Student's *t*-distribution has one heavy and one semiheavy tail. It is the only member of the GH family of distributions having this property. This can be seen as follows. Equation (6) shows that the only way of obtaining one heavy and one semiheavy 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 behavior of the GH skew Student's *t*-distribution also distinguishes it from the alternative skew Student's *t*-distributions reviewed in Section 1.3, which all have two tails decaying as polynomials. This makes it unique for modeling substantially skewed and heavy-tailed data.

3 PARAMETER ESTIMATION USING THE EM-ALGORITHM

The parameters of the GH skew Student's *t*-distribution can be estimated using maximum-likelihood (ML) estimation. The ML estimates might be directly obtained using a numerical optimization algorithm. However, the optimization is challenging, as some parameters are hard to separate and the likelihood function may have several local maxima. If one exploits the normal variance-mean mixture structure of the GH skew Student's t-distribution, one may apply the EM-algorithm [Dempster, Laird, and Rubin (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. Moreover, this algorithm is easily programmable, it surely converges to the maximum, and it provides interesting insight into the model. In what follows, we present 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 interating two steps: the expectation step (E-step) and the maximization step (M-step). In the E-step, one computes the expectation of the sufficient statistics 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.

3.1 E-Step

The E-step consists of 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, $Z|X \sim GIG\left(\lambda - \frac{1}{2}, \sqrt{\delta^2 + (x-\mu)^2}, \alpha\right)$. Hence, in the GH skew Student's t case, $Z|X \sim GIG\left(-\frac{(\upsilon+1)}{2}, \sqrt{\delta^2 + (x-\mu)^2}, |\beta|\right)$. 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,

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

and

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

Further, we have

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

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.

3.2 M-Step

In the M-step, one computes the parameter estimates resulting from maximizing the likelihood of $f_x(x, z) = f_{(x\,|\,z)}(x\,|\,z)\, f_z(z)$, replacing Z, Z^{-1}, and $\log Z$ by the pseudo values ρ_i, ξ_i, and χ_i from the E-step. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\bar{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$. In 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,

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

3.3 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 $\bar{m}_1$,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

$$
\begin{aligned}
\tilde{\mu} &= \bar{m}_1 - \frac{\tilde{\beta}\tilde{\delta}^2}{\tilde{\nu}-2} \\
\tilde{\beta} &= \operatorname{sign}(\bar{m}_3) \cdot \frac{(\tilde{\nu}-2)^{1/2}(\tilde{\nu}-4)^{1/2}\left[\bar{m}_2(\tilde{\nu}-2)-\tilde{\delta}^2\right]^{1/2}}{2^{1/2}\tilde{\delta}^2} \\
\tilde{\delta}^2 &= \frac{6(\tilde{\nu}-2)^2(\tilde{\nu}-4)\bar{m}_2}{3\tilde{\nu}^2-2\tilde{\nu}-32} \\
&\quad \left(1-\sqrt{1-\frac{(3\tilde{\nu}^2-2\tilde{\nu}-32)(12(5\tilde{\nu}-22)-(\tilde{\nu}-6)(\tilde{\nu}-8)\bar{m}_4)}{216(\tilde{\nu}-2)^2(\tilde{\nu}-4)}}\right).
\end{aligned}
$$

The moment estimate for v 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

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

4 NUMERICAL EXAMPLES

To study how suitable the GH skew Student's t-distribution is for modeling financial data, we have fit it to the four log return series described in Section 4.1. Moreover, we compare the fit of the GH skew Student's t-distribution to that of the NIG distribution, and the skew Student's t-distribution of Azzalini and Capitanio presented in Section 1.3. The latter is hereafter denoted Azzalini's skew Student's t-distribution. The reasons for selecting just these two distributions from the ones reviewed in Section 1 are as follows. The skew exponential power distribution has exponential tails like the NIG distribution, and hence the two distributions are believed to give similar results. Our preference for the NIG distribution is due to its analytical tractability. All four skew Student's t-distributions presented in Section 1.3 have two tails behaving like polynomials. Hence they are also likely to produce approximately the same fit to financial data.

Azzalini's version was a natural choice, as there exists a ready-made estimation routine in the software package *R*.

4.1 Data

The dataset studied in this article consists of four different kinds of market variables: the total index for Norwegian stocks (TOTX), the SSBWG hedged bond index for international bonds, the NOK/EUR exchange rate (NOK is Norwegian kroner), and the EURIBOR 5-year interest rate. The historical time period used goes from January 4, 1999, to July 8, 2003, and corresponds to 1094 daily observations. Our full sample actually runs until January 21, 2005, but we save the last 387 observations for an out-of-sample study (see Section 5). Figure 2 shows normal QQ plots of the corresponding logarithmic returns. The Norwegian

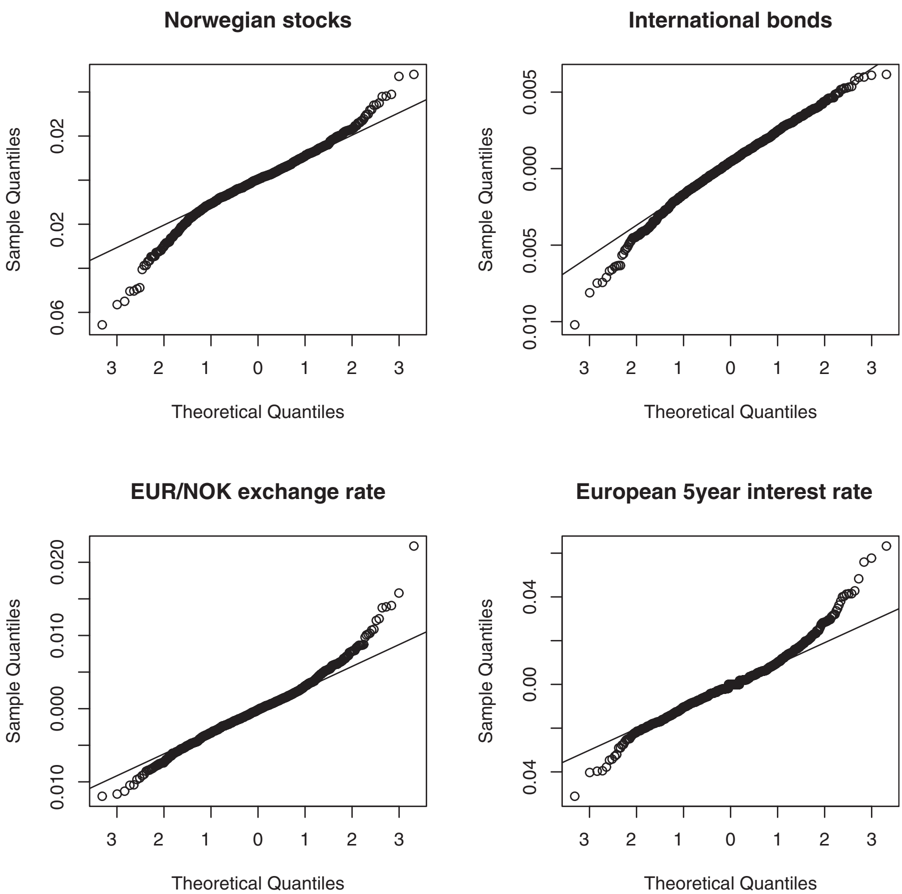

Figure 2 QQ plots for log returns of selected financial market variables.

stock return distribution has both tails heavier than the gaussian distribution, and the left tail is heavier than the right. The left tail of the international bond return distribution is also heavy, while the right tail is lighter than the gaussian distribution. The NOK/EUR exchange rate distribution has a heavy right tail, while the left tail is closer to the gaussian distribution. Finally, for the European 5-year interest data, both tails are heavy, but the right tail 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 the use of the GH skewed Student's t-distribution, which has one tail determined by polynomial and the other by exponential behavior.

4.2 Parameter Estimates

Tables 1–4 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 datasets. In order to find the 95% confidence intervals for the parameters given in parentheses, we bootstrapped the data. Since they are correlated over time, we performed a blocked bootstrap, according to Kuench (1989). The choice of block size is a trade-off between bias and variance. If the blocks are too small, they are rather correlated. On the other hand, if the blocks

Table 2 Moment estimates resulting when fitting the GH skew Student's t-distribution to each risk factor.

Risk factor	μ	δ	β	v
Norwegian stocks	0.00483	0.03230	−34.25245	9.00163
International bonds	0.00289	0.00853	−642.44490	20.23618
NOK/EUR exchange rate	−0.00153	0.00925	120.93460	9.05843
EURIBOR 5-year	−0.00524	0.03099	37.90472	9.03945

Table 3 Parameter estimates resulting when fitting the NIG distribution to each risk factor

Risk factor	μ	δ	β	α
Norwegian stocks	0.00190 (0.00060, 0.00375)	0.01350 (0.01102, 0.01755)	−13.87510 (−28.39, −3.54)	87.93290 (69.66, 126.23)
International bonds	0.00210 (0.00118, 0.00391)	0.00490 (0.00349, 0.00784)	−424.49800 (−923.62, −194.52)	1243.93400 (847.72, 2105.50)
NOK/EUR exchange rate	−0.00080 (−0.00142, −0.00041)	0.00450 (0.00362, 0.00663)	57.04440 (28.47, 125.50)	359.80620 (288.64, 574.76)
EURIBOR 5-year	−0.00250 (−0.00478, −0.00121)	0.01280 (0.01008, 0.01780)	17.31610 (5.85, 35.02)	91.92160 (72.70, 141.97)

Table 4 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 (0.00001, 0.00679)	0.01016 (0.00883, 0.01235)	−0.46150 (−0.91985, −0.00396)	4.65220 (3.60, 6.90)
International bonds	0.00110 (0.00030, 0.00190)	0.00206 (0.00181, 0.00243)	−0.45124 (−1.07257, −0.00051)	9.19201 (6.87, 29.14)
NOK/EUR exchange rate	−0.00043 (−0.00189, −0.00004)	0.00289 (0.00263, 0.00344)	0.10002 (0.00085, 0.77846)	5.17002 (4.16, 10.03)
EURIBOR 5-year	−0.00072 (−0.00778, −0.00052)	0.00916 (0.00813, 0.01223)	0.01551 (0.00513, 1.01641)	4.47262 (3.57, 7.87)

95% confidence intervals are in parentheses.

are too large compared with the length of the data series, the number of blocks is so small that a block may occur many times in the same bootstrap sample. We started by generating 1000 bootstrap samples with a set of block sizes between 1 and 200, and computed the standard deviation of the sample mean for each block size. Based on these and a plot of the autocorrelation function of the absolute values of the data, we chose a block size of 20. Moreover, we used 500 bootstrap samples to estimate the confidence intervals. As can be seen from the tables, the parameter β is significantly different from zero for all three distributions. Hence all distributions are clearly skewed.

For the GH skew Student's *t*-distribution, we used the EM-algorithm described in Section 3 (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 are shown in Table 2. They are somewhat different from the ML estimates, especially the estimate of ν. This is not surprising, since the moment estimates do not exist when $\nu \leq 8$. In order to examine the behavior of the EM-algorithm a bit closer, we ran it from several sets of initial values. The estimates obtained with the different sets were very similar. Moreover, 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 ML estimation scheme given in Azzalini and Capitanio (2003), which has been implemented in the sn-package for *R*. The CPU time per iteration was approximately 0.02 for GH skew Student's *t*, 0.01 for NIG, and 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 other two.

4.3 Goodness-of-Fit

To study the fit of the various distributions, we used the Kolmogorov-Smirnov's goodness-of-fit test [Conover (1971: 309–314)]. The Kolmogorov-Smirnov test showed very similar results for all distributions. The null hypothesis that the distribution fits the data was not rejected for any combination of distribution and risk factor at the 1% significance level, while at the 5% significance level, it was rejected for all three distributions for the EURIBOR five-year interest rate.

The Kolmogorov-Smirnov distance emphasizes deviations around the median of the fitted distribution. Since our main interest is the tails of the distribution, we also used graphical logarithmic left and right tail tests to examine the fit in the tails. The graphical tests were performed as follows. Let $\left(X_{(1)}, \ldots, X_{(N)}\right)$ denote the order statistic of the historical data and $\hat{F}(x)$ be the estimated cumulative distribution function of the fitted distribution, computed by numerical integration [for the NIG and GH skew Student's t-distribution, we used the method of Muhammad and Mori (2003), and for Azzalini's skew Student's distribution we used the function pst () in the sn-package for R]. A plot of $\log\left(\hat{F}\left(X_{(t)}\right)\right)$ against $X_{(t)}$ superimposed on a plot of $\log(t/(N+1))$ against $X_{(t)}$ shows the left tail fit of the fitted distribution, and a plot of $\log(1-\hat{F}(X_{(t)}))$ against $X_{(t)}$, superimposed on a plot of $\log((N+1-t)/(N+1))$ shows the right tail fit.

Figures 3–6 show the plots. The upper panel in each figure shows the left tail fit and the lower panel shows the right tail fit. The circles correspond to the empirical data and the three dotted lines to the NIG, Azzalini's skew Student's t-, and the GH skew Student's t-distribution. The solid line shows 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. On the other hand, Azzalini's skew Student's distribution slightly underestimates the left tail and overestimates the right. For the NOK/EUR exchange rate data, the NIG distribution underestimates the right tail. On the other hand, Azzalini's skew Student's t-distribution underestimates the right tail and overestimates the left. The GH skew student's t-distribution fits both tails better than the two other distributions. Finally, for the European five-year interest data, the NIG distribution underestimates the right tail, while 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 fits both tails quite well. Hence the GH skew Student's t-distribution provides the best overall fit for all four financial market variables.

5 APPLICATION TO RISK ESTIMATION

5.1 Unconditional Distribution

In this section we use the estimated distributions from Section 4 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 asset price. In the

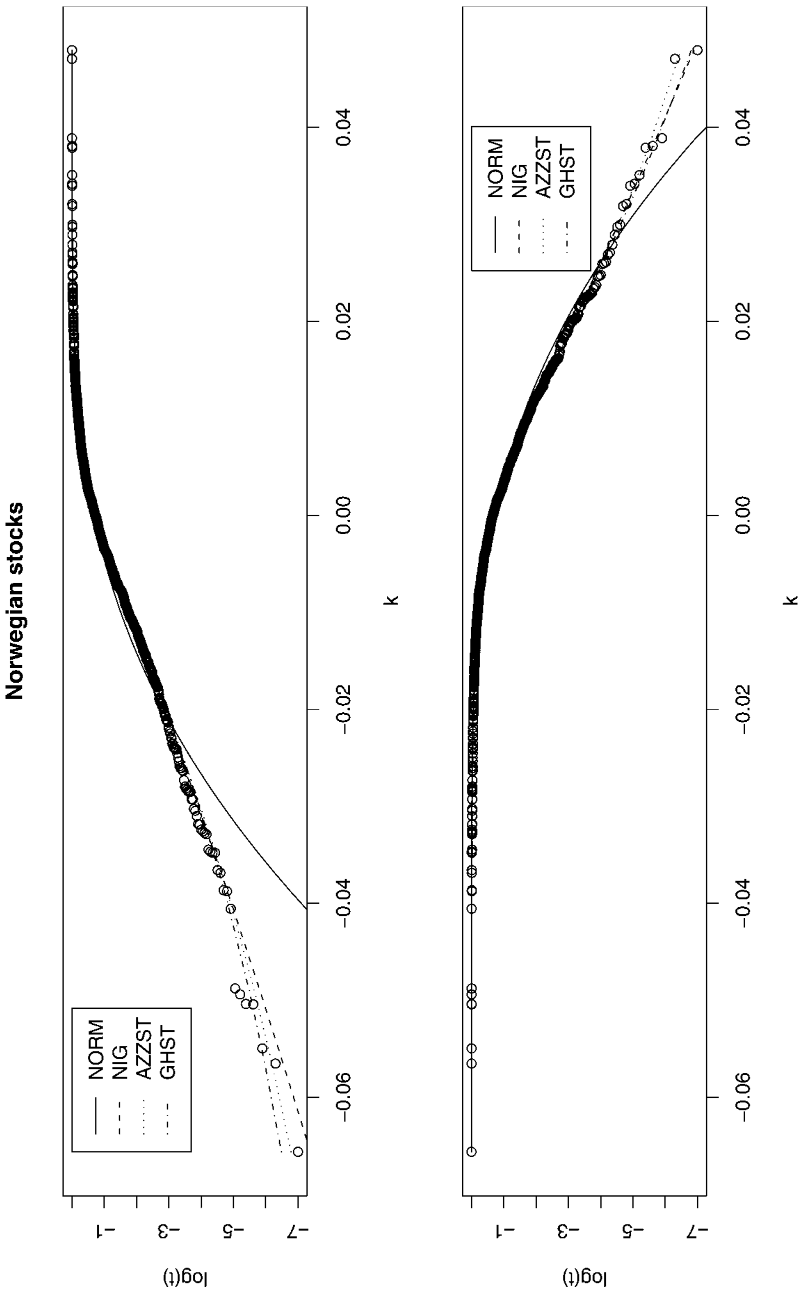

Figure 3 Left and right tail plots for Norwegian stocks.

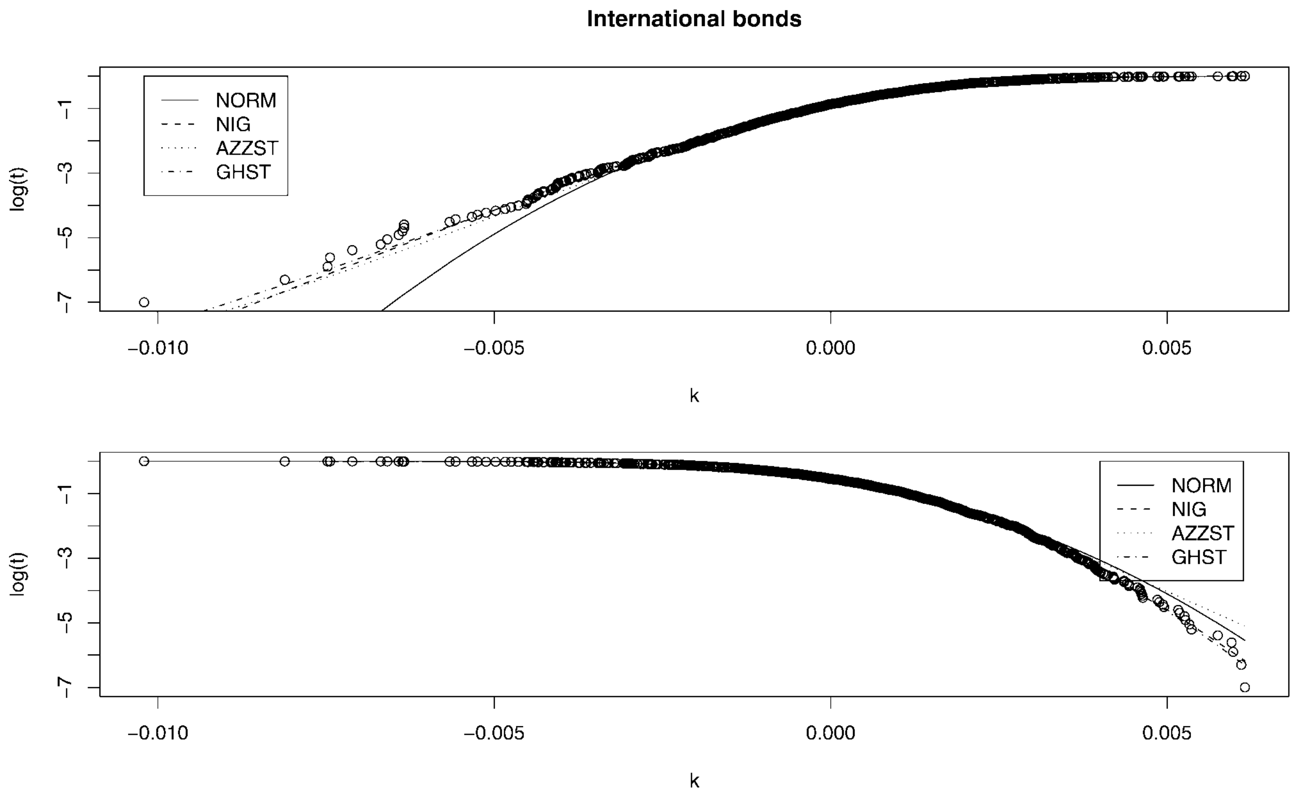

Figure 4 Left and right tail plots for international bonds.

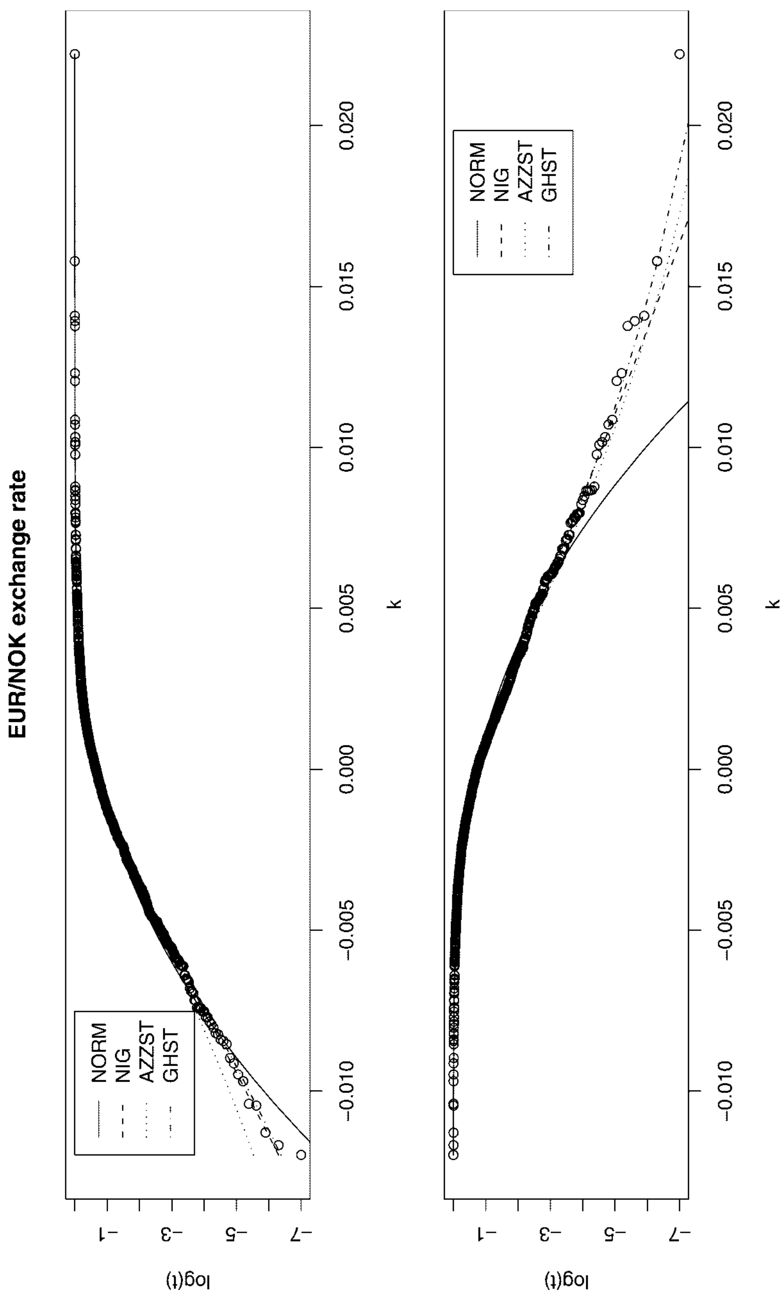

Figure 5 Left and right tail plots for NOK/EUR exchange rate.

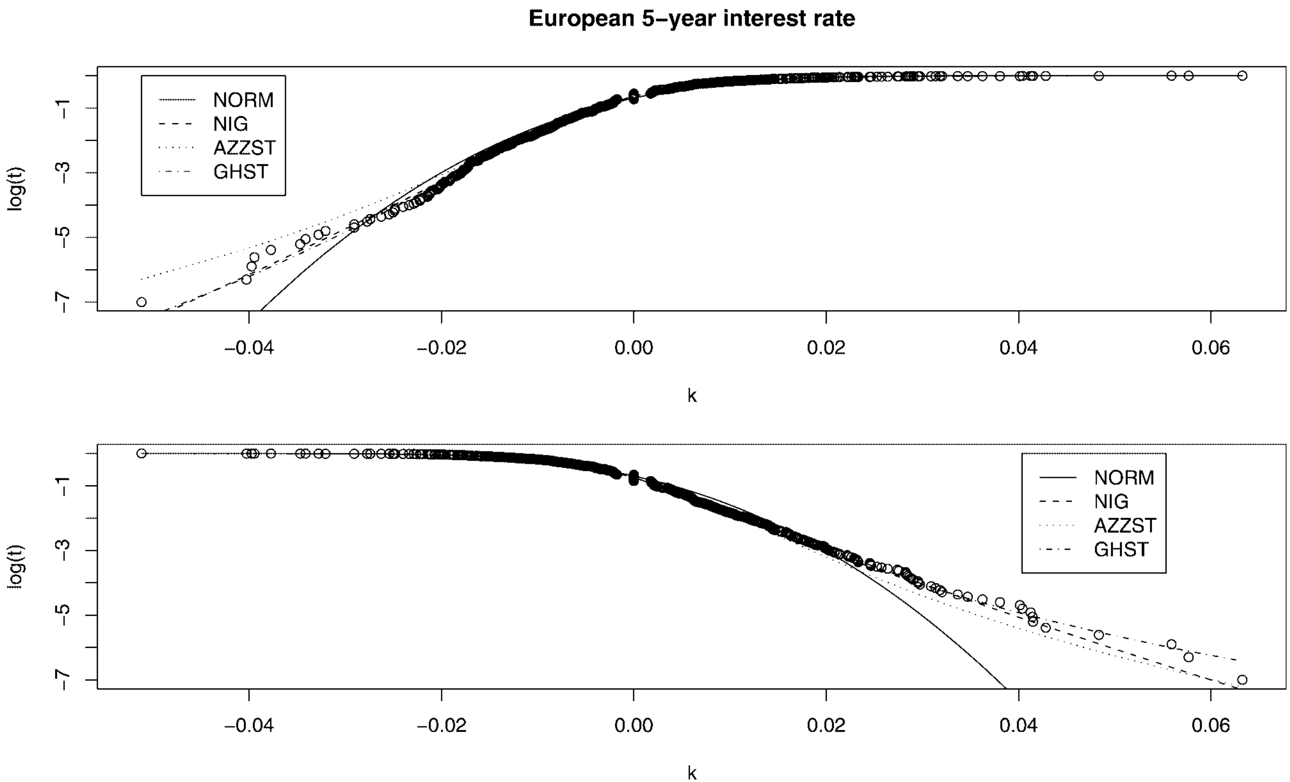

Figure 6 Left and right tail plots for European five-year interest rate.

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 July 9, 2003, to January 21, 2005, corresponding to 387 observations. For each day in the test period we predict the one-day VaR and ES at levels 0.005, 0.01, 0.05, 0.95, 0.99, and 0.995. The first three levels are used to measure the risk of long positions and the last three the risk of the short ones. We use the likelihood ratio statistic by Kupiec (1995) to verify whether the VaR predictions are correct. The method consists of calculating the number of times $x(\alpha)$ the observed returns fall below (long positions) or above (short positions) the VaR estimate at level α, that is, $R_t < \widehat{VaR}(\alpha)$ 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 equal to α. Under the null hypothesis, the likelihood ratio statistic given by

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

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

Table 5 shows the observed number of violations of VaR for each distribution and level. The corresponding p-values are shown in Table 6. 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.

To backtest the predicted ES value at confidence level α, $\widehat{ES}(\alpha)$, we use the measure proposed by Embrechts, Kaufmann, and Patie (2004). Define $\delta_t(\alpha) = R_t - \widehat{ES}(\alpha)$ and denote $\kappa(\alpha)$ as the set of time points for which a violation of $\widehat{VaR}(\alpha)$ occurs. Further, let $y(\alpha)$ be the number of times $\delta_t(\alpha)$ is less than (long positions) or greater than (short positions) its empirical α-quantile, and $\tau(\alpha)$ the set of time points for which this happens. The measure is given by

Table 5 NOK/EUR exchange rate, unconditional distribution: number of violations of VaR for each distribution and 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
Expected number	1.9	3.9	19.4	19.4	3.9	1.9

Table 6 NOK/EUR exchange rate, unconditional distribution: p-values from the Kupiec test for each distribution and 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**

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

where

$$D_1(\alpha) = \frac{1}{x(\alpha)} \sum_{t \in \kappa(\alpha)} \delta_t(\alpha)$$

and

$$D_2(\alpha) = \frac{1}{y(\alpha)} \sum_{t \in \tau(\alpha)} \delta_t(\alpha).$$

$D_1(\alpha)$ is the standard backtesting measure for expected shortfall estimates. Its weakness is that it depends strongly on the VaR estimates without adequately reflecting the correctness of these values. To correct for this, it is combined with a penalty $D_2(\alpha)$. A good estimation of expected shortfall will lead to a low value of $D(\alpha)$. In Table 7 we show the $D(\alpha)$ 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 all six cases. Hence, for the prediction of the expected shortfall of our test data, it is superior to the other distributions.

5.2 Conditional Distribution

As an example, we have also modeled the conditional distribution of the geometric returns for the NOK/EUR exchange rate. For simplicity we assume that

Table 7 NOK/EUR exchange rate, unconditional distribution: backtest measure of expected shortfall predictions for each distribution and 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

the variance of the geometric returns follow a univariate GARCH(1,1) [Bollerslev (1986)], without verifying whether this model fits the data better than more advanced models. That is, we have the model

$$\begin{aligned} r_t &= c + \sigma_t z_t \\ \mathrm{E}[z_t] &= 0 \text{ and } \mathrm{var}[z_t] = 1 \\ \sigma_t^2 &= a_0 + a\, \varepsilon_{t-1}^2 + b\, \sigma_{t-1}^2. \end{aligned} \tag{16}$$

First, we included an AR(1) term in the conditional mean equation, but the coefficient was shown to be not significantly different from zero, and hence the conditional mean equation has been restricted to a constant. The standardized returns, z_t, are modeled by the three alternative distributions: GH skew Student's t, NIG, and Azzalinis skew Student's t. To be used as the conditional distribution, these distributions must be standardized, that is, reparameterized to have mean zero and variance one. For the GH skew Student's t-distribution this is achieved by setting

$$\mu = -\frac{\delta\beta}{\nu - 2}, \tag{17}$$

$$\delta^2 = \frac{(\nu-2)(\nu-4)}{4\beta^2}\left(-1 + \sqrt{1 + \frac{8\beta^2}{\nu-4}}\right). \tag{18}$$

For the NIG distribution, we let

$$\mu = \frac{-\delta\beta}{\sqrt{\alpha^2 - \beta^2}}, \tag{19}$$

$$\delta = \frac{\left(\alpha^2 - \beta^2\right)^{3/2}}{\alpha^2}. \tag{20}$$

Finally, the standardized version of Azzalini's skew Student's t-distribution, is obtained by setting

$$\mu = -\sqrt{\frac{\nu}{\pi}}\frac{\delta\beta}{\sqrt{1+\beta^2}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \tag{21}$$

$$\delta = \sqrt{\frac{\nu - 2}{\nu\left(1 - \frac{\nu-2}{\pi}\frac{\alpha^2}{1+\alpha^2}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2\right)}}. \tag{22}$$

Estimation of the parameters of the conditional density along with the parameters of the GARCH model may be challenging. Therefore we use a sequential approach for estimating the parameters of our model. In the first step we estimate the parameters of the GARCH model for the conditional variance, using the pseudo-maximum-likelihood (PML) procedure advocated by Bollerslev and Wooldridge (1992), among others. In the second step, the parameters of the conditional distribution are determined using the methods described in the previous section. It can be shown that the PML method, under certain conditions, yields consistent estimators of the conditional mean and variance parameters [see, for instance, chapter 4 of Gouriéroux (1997)].

The parameters of the GARCH model and the three conditional distributions are given in Tables 8 and 9, respectively. The GARCH standard errors are computed using the default covariance estimate, while jackknife 95% confidence intervals are given in parentheses in the latter. The confidence intervals were generated by bootstrapping 500 different versions of the original dataset, each time computing new parameter estimates. Since the standardized returns, z_t, contrary to the original returns, r_t, do not exhibit serial dependence, it is not necessary to use a block bootstrap in this case.

Table 8 NOK/EUR exchange rate: estimated parameters for the GARCH model with standard errors in parentheses.

Risk factor	c	a_0	a	b
NOK/EUR	−1.65e−004 (1.75e−007)	5.87e−007 (1.75e−007)	0.085 (0.014)	0.871 (0.023)

Table 9 NOK/EUR exchange rate: parameter estimates resulting when fitting the distributions to the standardized returns.

Distribution	μ	δ	β	v
GH skew	−0.2129	2.2898	0.2055	7.3194
Student's t	(−0.0603, −0.4371)	(1.8649, 3.1222)	(0.0391, 0.4547)	(5.3764, 12.3917)
	μ	δ	β	α
NIG	−0.2020	1.4697	0.2058	1.5115
	(−0.0403, 0.3993)	(1.2027, 1.9808)	(0.0403, 0.4207)	(1.2267, 2.0295)
	μ	δ	β	v
Azzalini's skew t	−0.3646	0.9018	0.4838	7.1129
	(−0.6527, −0.0621)	(0.8204, 1.0809)	(0.0631, 0.9449)	(5.2259, 13.5977)

95% confidence intervals are in parentheses.

We use the same training and test set as in the previous section, and the test procedure is as follows. For each day t in the test set:

1. Compute the one-step-ahead forecast of σ_t, given information up to time t.
2. Given the parameters of the distribution of z_t, determine the parameters of the distribution of $r_t = c + \sigma_t z_t$.
3. For confidence levels $\alpha \in \{0.005, 0.01, 0.05, 0.95, 0.99, 0.995\}$, predict the one-day $\text{VaR}_t(\alpha)$ and $\text{ES}_t(\alpha)$.

To compute the distribution of $r_t = c + \sigma_t z_t$, we use the following properties

a. If $X \sim GH(\mu, \delta, \beta, v)$, then $Y = b + aX \sim GH(b + a\mu, a\delta, \beta/a, v)$

b. If $X \sim NIG(\mu, \delta, \beta, \alpha)$, then $Y = b + aX \sim NIG(b + a\mu, a\delta, \beta/a, \alpha/a)$

c. If $X \sim AZZT(\mu, \delta, \beta, v)$, then $Y = b + aX \sim AZZT(b + a\mu, a\delta, \beta, v)$

Figure 7 shows the tail plot for the standardized returns of the NOK/EUR exchange rate. As can be seen in the figure, all distributions, except the gaussian, give approximately the same fit to the training data. This is also verified by Table 10, which shows that the number of violations of VaR for each distribution and level are the same. The corresponding p-values are shown in Table 11. If we use a 5% level for the Kupiec test, the null hypothesis is never rejected for any of the distributions. To backtest the ES value we use the test proposed by McNeil and Frey (2000). We define the residuals as

$$S_t(\alpha) = \frac{R_t - \text{ES}_t(\alpha)}{\sigma_t}.$$

Under our model, these residuals are i.i.d. and, conditional on $R_t < \text{ES}_t(\alpha)$ for $\alpha \in \{0.005, 0.01, 0.05\}$ or $R_t > ES_t(\alpha)$ for $\alpha \in \{0.95, 0.99, 0.995\}$, they have expected zero value. To test the hypothesis of mean zero, we use a bootstrap test that makes no assumption of the underlying distribution of the residuals [see Efron and Tibshirani (1993: 224)]. In Table 12 we show p-values from the test for each distribution and level. If we use a 5% level, the null hypothesis is rejected twice for Azzalini's skew Student's t-distribution and never for the NIG and GH skew Student's t-distribution.

The above test has the evident weakness that it strongly depends on VaR estimates without adequately reflecting the correctness of these values. Hence we also use the measure proposed by Embrechts, Kaufmann, and Patie (2004), described in Section 5.1. In Table 13, we show the $D(\alpha)$ values for each distribution and level. As can be seen in the table, the GH skew Student's t-distribution

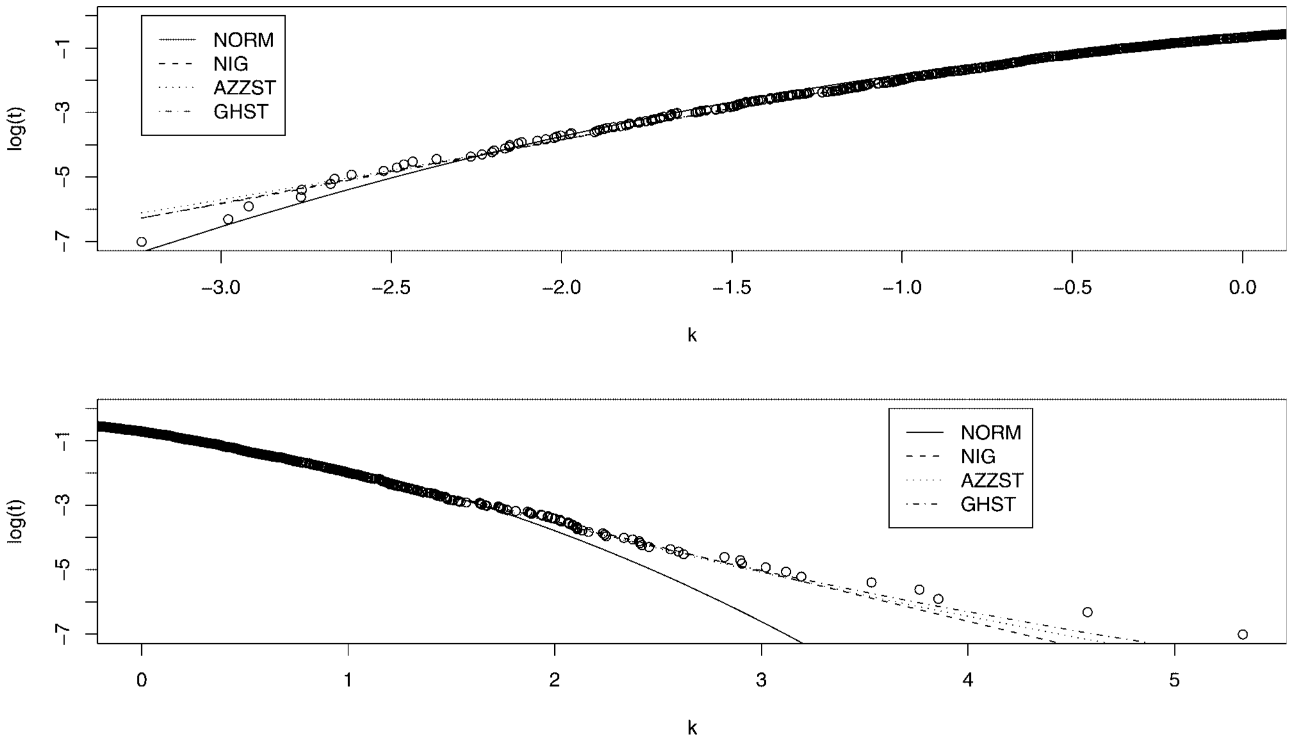

Figure 7 Left and right tail plots for the standardized returns of the NOK/EUR exchange rate.

Table 10 NOK/EUR exchange rate, conditional distribution: number of violations of VaR for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	2	6	20	16	7	2
NIG	2	6	20	16	7	2
Azzalini's skew Student's *t*	2	6	20	16	7	2
Expected number	1.9	3.9	19.4	19.4	3.9	1.9

Table 11 NOK/EUR exchange rate, conditional distribution: p-values from the Kupiec test for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	0.96	0.31	0.88	0.42	0.15	0.96
NIG	0.96	0.31	0.88	0.42	0.15	0.96
Azzalini's skew Student's *t*	0.96	0.31	0.88	0.42	0.15	0.96

Table 12 NOK/EUR exchange rate, conditional distribution: p-values from the McNeil and Frey test for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	0.78	0.90	0.05	0.89	0.54	0.74
NIG	0.23	0.12	0.07	0.89	0.68	0.75
Azzalini's skew Student's *t*	0.00	0.47	0.50	0.49	0.54	0.00

Table 13 NOK/EUR exchange rate, conditional distribution: backtest measure of expected shortfall predictions for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	0.0005	0.0007	0.0004	0.0006	0.0007	0.0022
NIG	0.0034	0.0020	0.0004	0.0007	0.0010	0.0031
Azzalini's skew Student's *t*	0.0011	0.0009	0.0004	0.0007	0.0008	0.0043

gives lower values than the two other distributions in five of the six cases. In the last case, all values are equal.

5.3 Conditional Distribution of a Portfolio

A commerical bank and a regulator are interested in the aggregate VaR across several trading activities. One may either aggregate profit and loss data, and proceed with a univariate model for the aggregate, or start with disaggregate data. The latter approach has the advantage of better capturing the structure of risk within the portfolio, but Berkowitz and O'Brien (2002) have shown that in some cases aggregation and modeling issues may make the univariate approach more desireable. In this section we repeat the experiment from Section 5.2, but for a portfolio consisting of equal proportions of Norwegian stocks, international bonds, and the NOK/EUR exchange rate. We follow Berkowitz and O'Brien (2002) and model the aggregate return of this portfolio by the univariate GARCH model in Equation (16). The parameters of the GARCH model and the three conditional distributions are given in Tables 14 and 15, respectively. Figure 8 shows the tail plot for the standardized returns of the portfolio. As can be seen in the figure, all distributions, except the gaussian, give approximately the same fit to the training data.

We assume that the portfolio weights are the same as in the training period for each day in the test period, and follow the same test procedure as described in Section 5.2. Table 16 shows the number of violations of VaR for each distribution and level and Table 17 the corresponding *p*-values. If we use a 5% level for the Kupiec test, the null hypothesis is rejected twice for the NIG distribution and once for the two other distributions. The standard deviation of the portfolio return is much lower in the test period than in

Table 14 Portfolio: estimated parameters for the GARCH model with standard errors in parentheses.

Risk factor	c	a_0	a	b
Portfolio return	8.16e−005 (1.19e−004)	9.31e−007 (2.39e−007)	0.131 (0.018)	0.819 (0.024)

Table 15 Portfolio: parameter estimates resulting when fitting the distributions to the standardized returns.

Distribution	μ	δ	β	ν
GH skew	0.1966	2.3537	−0.2337	7.5993
Student's t	(0.0140, 0.5662)	(1.8629, 3.778)	(−0.6568, −0.0435)	(5.4219, 14.0094)
	μ	δ	β	α
NIG	0.1813	1.5034	−0.1840	1.5363
	(0.0110, 0.4909)	(1.1990, 2.0828)	(−0.5279, −0.0111)	(1.2213, 2.2287)
	μ	δ	β	ν
Azzalini's	0.3867	0.9343	−0.5856	7.4814
skew t	(0.0555, 0.6773)	(0.8272, 1.1264)	(−1.1294, −0.1177)	(5.3714, 13.9432)

95% confidence intervals are in parentheses.

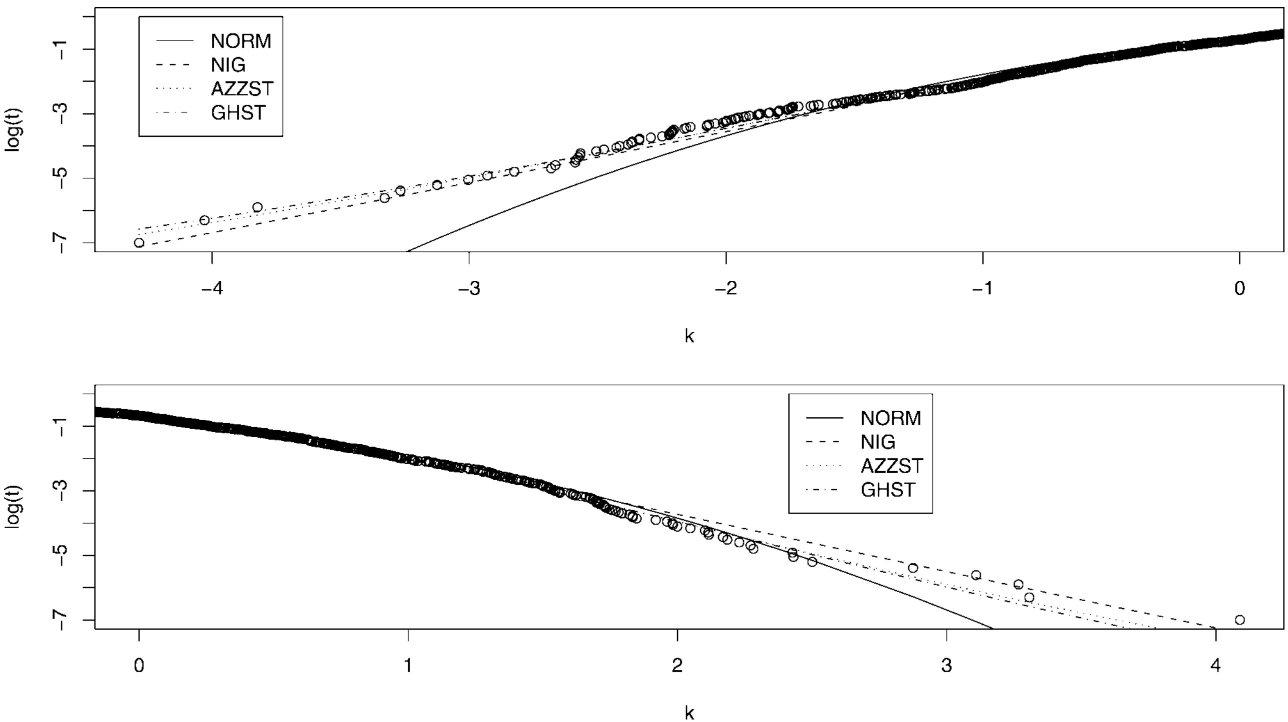

Figure 8 Left and right tail plots for the standardized returns of the portfolio.

Table 16 Portfolio: number of violations of VaR for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's t	0	1	11	13	2	1
NIG	0	1	11	10	1	0
Azzalini's skew Student's t	0	1	11	13	2	1
Expected number	1.9	3.9	19.4	19.4	3.9	1.9

Table 17 Portfolio: p-values from the Kupiec test for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's t	0.05	0.08	0.03	0.12	0.29	0.46
NIG	0.05	0.08	0.03	0.02	0.08	0.05
Azzalini's skew Student's t	0.05	0.08	0.03	0.12	0.29	0.46

the estimation period (0.0028 versus 0.0042), which is the reason for the relatively small p-values. Figure 9 shows the absolute value of the portfolio return for the entire time period January 4, 1999, to January 21, 2005, with the estimation and test period separated by the vertical line. The horizontal line corresponds to the volatility of the estimation period. As can be seen in the figure, the volatility of the test period is very low compared to that of the estimation period. This difference in volatility is mainly caused by the corresponding difference in volatility for the Norwegian stocks (0.0086 versus 0.0127). For international bonds, the difference is smaller (0.0022 versus 0.0019), while for the NOK/EUR exchange rate the volatility of the test period actually is higher than that of the training period (0.0036 versus 0.0039).

Table 18 shows the p-values from the the McNeil and Frey test. A "—" in the table means that the p-value cannot be computed because the number of violations is too low. If we use a 5% level, the null hypothesis is never rejected. Finally, the $D(\alpha)$ values from the Embrechts test are given in Table 19. Since the number of violations of VaR is very low (zero to two) for all levels except 5% and 95%, only the results for these two levels should be trusted.

6 CONCLUSION

In this article we have argued for a special case of the GH 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 behavior. This makes it different from other skew Student's t-distributions

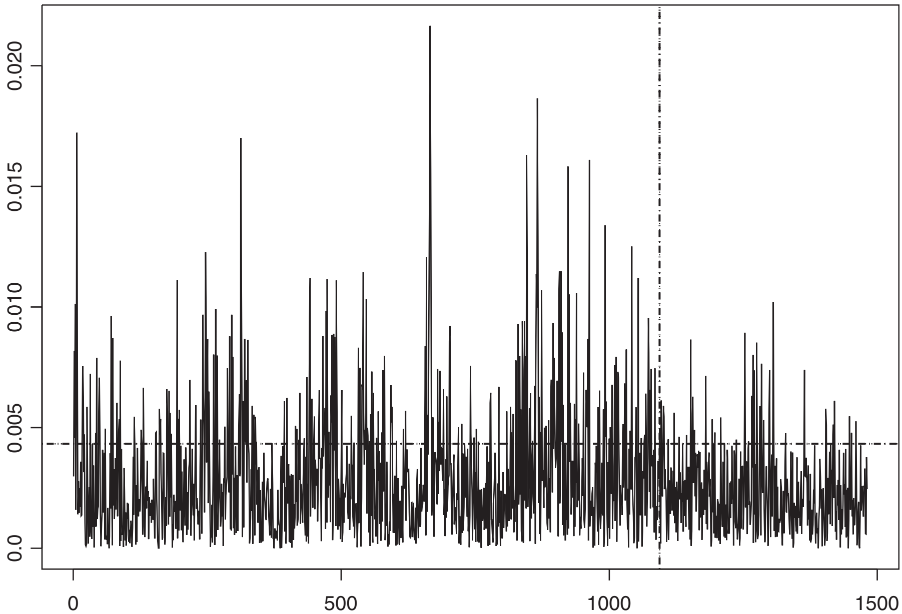

Figure 9 The absolute value of the portfolio return for the entire time period—January 4, 1999, to January, 21, 2005—with the estimation and test period separated by the vertical line. The horizontal line corresponds to the volatility of the estimation period.

Table 18 Portfolio: *p*-values from the McNeil and Frey test for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	—	—	0.97	0.13	0.25	—
NIG	—	—	0.92	0.13	—	—
Azzalini's skew Student's *t*	—	—	0.98	0.10	0.24	—

Table 19 Portfolio: backtest measure of expected shortfall predictions for each distribution and level.

Distribution	0.5%	1%	5%	95%	99%	99.5%
GH skew Student's *t*	—	0.0065	0.0011	0.0006	0.0010	0.0014
NIG	—	0.0021	0.0009	0.0008	0.0005	—
Azzalini's skew Student's *t*	—	0.0031	0.0012	0.0006	0.0012	0.0017

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, it is almost as analytically tractable as the NIG distribution, and due to the normal mean-variance mixture structure, we can 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. In addition, if the data are very skewed, it is also superior to Azzalini's skew Student's t-distribution. We have also predicted out-of-sample one-day VaR and expected shortfall at levels 0.005, 0.01, 0.05, 0.95, 0.99, and 0.995 for the unconditional distribution of the log returns of the NOK/EUR exchange rate. Backtesting shows that the GH skew Student's t-distribution outperforms the NIG and Azzalini's skew Student's t-distribution when expected shortfall is used as a risk measure, and is also slightly better for predicting VaR. Moreover, we have fitted a GARCH(1,1) to the log returns of the NOK/EUR exchange rate and predicted out-of-sample one-day VaR and expected shortfall using the GH skew Student's t, the NIG, and Azzalini's skew Student's t-distribution as conditional distributions. Even in this case, the GH skew Student's t-distribution outperforms the two other distributions. Finally, we have fitted a GARCH(1,1) to the aggregate log returns of a portfolio consisting of equal proportions of Norwegian stocks, international bonds, and the NOK/EUR exchange rate. The GH skew Student's t-distribution also seems to fit such returns quite well.

APPENDIX A: SKEWNESS AND KURTOSIS OF GH SKEW STUDENT'S *t*-DISTRIBUTION

The skewness of a variable X is given by

$$E\left[(X-E[X])^3\right]\Big/V[X]^{3/2} = \left(E[X^3]-3E[X]E[X^2]+2(E[X])^3\right)\Big/V[X]^{3/2}, \quad (23)$$

where $E[X]$ and $V[X]$ are the expectation and variance of the variable, respectively. Equation (23) is obtained by inserting the expressions for the mean and variance of a GH skew Student's t-distributed random variate given in Equations (10) and (11), and $E[X^3]$ given by

$$E\left[X^3\right] = \mu^3 + \frac{\delta^2\left(3\mu^2\beta+3\nu\right)}{\nu-2} + \frac{\delta^4\left(3\beta+3\mu\beta^2\right)}{(\nu-2)(\nu-4)} + \frac{\delta^6\beta^3}{(\nu-2)(\nu-4)(\nu-6)}, \quad (24)$$

into Equation (23).

The kurtosis of a variable X is given by

$$E\left[(X-E[X])^4\right]\Big/V[X]^2, \quad (25)$$

where

$$E\left[(X - E[X])^4\right] = E\left[X^4\right] - 3\left(E\left[X^2\right]\right)^2 - 4E[X]E[X^3] + 12(E[X])^2 E\left[X^2\right] - 6(E[X])^4.$$

Inserting the expressions for the mean and variance from Equations (10) and (11), as well as the third moment from Equation (24) and $E[X^4]$ given by

$$E\left[X^4\right] = \mu^4 + \frac{\delta^2\left(6\mu^2 + 4\mu^3\beta\right)}{\nu - 2} + \frac{\delta^4\left(3 + 12\mu\beta + 6\mu^2\beta^2\right)}{(\nu - 2)(\nu - 4)}$$
$$+ \frac{\delta^6\left(6\beta^2 + 4\mu\beta^3\right)}{(\nu - 2)(\nu - 4)(\nu - 6)} + \frac{\delta^8\beta^4}{(\nu - 2)(\nu - 4)(\nu - 6)(\nu - 8)}$$

into Equation (25) results in Equation (13) for the kurtosis.

Received March 10, 2005; revised August 11, 2005; accepted December 21, 2005

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.

Ayebo, A., and T. J. Kozubowski. (2003). "An Asymmetric Generalization of Gaussian and Laplace Laws." *Journal of Probability and Statistical Science 1*, 187–210.

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." *Proceedings of the Royal Society of London, Series 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*.

Berkowitz, J., and J. O'Brien. (2002). "How Accurate are the Value-at-Risk Models at Commercial Banks." *Journal of Finance 57*, 1093–1111.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics 31*, 307–327.

Bollerslev, T., and J. Wooldridge. (1992). "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances." *Economic Reviews 11*, 143–172.

Bølviken, E., and F. E. Benth. (2000). "Quantification of Risk in Norwegian Stocks via the Normal Inverse Gaussian Distribution." *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.

Conover, W. J. (1971). *Practical Nonparametric Statistics*. New York: John Wiley & Sons.

Demarta, S., and A. J. McNeil. (2004). The *t* Copula and Related Copulas. Technical report, ETH Zurich. Forthcoming in *International Statistical Review*.

Dempster, A. P., N. M. Laird, and D. Rubin. (1977). "Maximum Likelihood from Incomplete Data Using the EM Algorithm." *Journal of the Royal Statistical Society B 39*, 1–38.

Eberlein, E., and U. Keller. (1995). "Hyberbolic Distributions in Finance." *Bernoulli 1*, 281–299.

Efron, B., and I. Tibshirani. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Embrechts, P., R. Kaufmann, and P. Patie. (2004). "Strategic Long-Term Financial Risks: Single Risk Factors. Forthcoming in *Computational Optimization and Applications*.

Fernandex, 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*, 535–548.

Gouriéroux, C. (1997). *ARCH-Models and Financial Applications*. Berlin: Springer.

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." *Journal of the Royal Statistical Society, Series B 65*, 159–174.

Karlis, D. (2002). "An EM Type Algorithm for Maximum Likelihood Estimation of the Normal-Inverse Gaussian Distribution." *Statistics and Probability Letters 57*, 43–52.

Kuench, H. R. (1989). "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics 17*, 1217–1241.

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.

McNeil, A. J., and R. Frey. (2000). "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach." *Journal of Empirical Finance 7*, 271–300.

Mencia, F. J., and E. Sentana. (2004). "Estimation and Testing of Dynamic Models with Generalised Hyperbolic Innovations." CMFI Working Paper 0411, Madrid, Spain.

Muhammad, M., and M. Mori. (2003). "Double Exponential Formulas for Numerical Indefinite Integration." *Journal of Computational and Applied Mathematics 161*, 431–448.

Patton, A. (2004). "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation." *Journal of Financial Econometrics 2*, 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 dissertation, University of Freiburg.

Rydberg, T. H. (1997). "The Normal Inverse Gaussian Levy Process: Simulation and Approximation. *Communications in Statistics-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. *Canadian Journal of Statistics 31*, 129–150.

Snoussi, H., and J. Idier. (2005). "Blind Separation of Generalized Hyperbolic Processes: Unifying Approach to Stationary Non Gaussianity and Gaussian Non Stationarity. *Proceedings of the 30th International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Philadelphia, 2005.

Venter, J. H., and P. J. de Jongh. (2002). "Risk Estimation Using the Normal Inverse Gaussian Distribution." *Journal of Risk 4(2)*, 1–24.