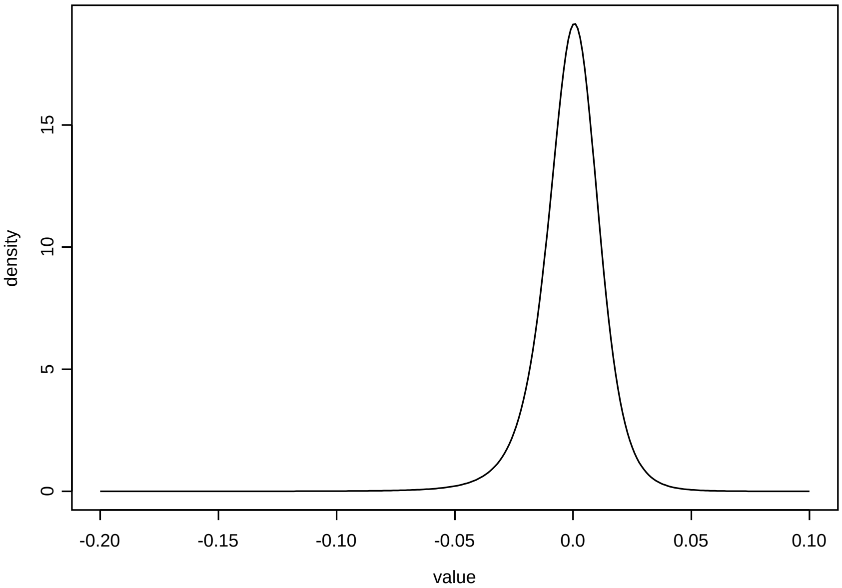
NIG and Skew Student's t: Two special cases of the Generalised Hyperbolic distribution



SAMBA/01/05

Kjersti Aas
Ingrid Hobæk Haff

January 2005

Norwegian Computing Center
APPLIED RESEARCH AND DEVELOPMENT

NR Note

Title: NIG and Skew Student's t:
Two special cases of the Generalised Hyperbolic distribution

Date: January 2005
Year: 2005
Note no.: SAMBA/01/05

Author: Kjersti Aas
Ingrid Hobæk Haff

Abstract: 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 fatter 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 towards the left tail.

The generalised hyperbolic (GH) distribution is a promising distribution for such returns. It was introduced by Barndorff-Nielsen (1977). An important aspect is that GH distributions embrace many special cases and limiting distributions. Some examples are the hyperbolic, the normal inverse Gaussian (NIG), the (skew) Student's t, and the normal distributions. All of these have been used to model financial returns. In this paper we focus on two of the special cases, the NIG and the skewed Student's t-distribution.

Keywords: Probability distribution, Generalised Hyperbolic, NIG, Skew Student's t, EM-algorithm, Simulation

Target group:

Availability: Closed

Project: Finansmarked

Project no.: 220195

Research field: Finance, insurance and power market

No. of pages: 14

Norwegian Computing Center
Gaustadalléen 23, P.O. Box 114 Blindern, NO-0314 Oslo, Norway
Telephone: +47 2285 2500, telefax: +47 2269 7660, `http://www.nr.no`

Contents

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 fatter 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 with a heavier left tail, indicating that big losses are more frequent than big gains.

The generalised hyperbolic (GH) distribution is a promising distribution for such returns. It was introduced by Barndorff-Nielsen (1977) in connection with a study of grains of sand. 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 heavier than those of the normal distribution. While several specific subclasses of the GH distribution have been applied in various situations, the distribution itself is very seldom used in practical applications. This is probably due to the fact that it is not particularly analytically tractable, and that it is very challenging to estimate its parameters, especially the parameter that determines the subclass. Even for very large sample sizes, it may be hard to make a distinction between different values of the subclass parameter because of the flatness of the GH likelihood function in this parameter. See for instance Prause (1999).

An important aspect is that GH distributions embrace many special cases and limiting distributions. Some examples are the hyperbolic, the normal inverse Gaussian (NIG), the (skewed) Student's t, and the normal distributions. All of these have been used to model financial returns. In this paper we focus on two of the special cases, the NIG and the skew Student's t-distribution.

The NIG distribution was introduced by Barndorff-Nielsen (1997). It is able to model symmetric and asymmetric distributions with possibly long tails in both directions. Moreover, the NIG distribution possesses a number of attractive theoretical properties, among others its analytical tractability. For these reasons, it has been used repeatedly for applications in finance, both as the conditional distribution of a GARCH-model (Andersson, 2001; Forsberg and Bollerslev, 2002; Jensen and Lunde, 2001; Venter and de Jongh, 2002) and as the unconditional return distribution (Bølviken and Benth, 2000; Eberlein and Keller, 1995; Lillestøl, 2000; Prause, 1997; Rydberg, 1997). The tail behaviour of NIG is often classified as semi-heavy. That is, the tails are much heavier than in the Gaussian distribution, but it may not be adequate to deal with cases of extremely heavy tails, such as those of Pareto or non-Gaussian stable laws.

The skewed Student's t-distribution is a less studied subclass of the GH 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 is our opinion that that it deserves further attention. This distribution, hereby denoted the generalised hyperbolic skew Student's t-distribution (the GH skew Student's t-distribution for short), is it almost as analytically tractable as the NIG distribution. Moreover, it allows for very heavy tails and substantial skewness. Finally, 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 rest of this paper is organized as follows. Section 2 treats the GH distribution. First, we review some aspects of this distribution. Then, we describe an approach for estimating the parameters, and finally we give the algorithm for generating samples from the distribution. Sections 3 and 4 adress the special cases NIG and skew Student's t, respectively.

2 The generalised hyperbolic distribution

2.1 Density

The univariate generalised hyperbolic distribution can be parameterised in several ways. We follow Prause (1999) and let

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

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

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

The GH distribution may be represented as a normal variance-mean mixture with the generalised inverse Gaussian distribution as a mixing distribution (Barndorff-Nielsen and Blæsild, 1981). This means that a generalised hyperbolic variable X can be represented as

$$X = \mu + \beta\, Z + \sqrt{Z}\, Y, \tag{2}$$

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 (2) that $X|Z = z \sim N(\mu + \beta\, z, z)$. $GIG(\cdot)$ denotes the Generalised Inverse Gaussian (GIG) distribution (Barndorff-Nielsen, 1977) having density

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

The mean and variance of a GH distributed random variate X are (Barndorff-Nielsen and Stelzer, 2004)

$$\mathrm{E}(X) = \mu + \frac{\beta\,\delta}{\gamma}\,\frac{K_{\lambda+1}(\delta\,\gamma)}{K_\lambda(\delta\,\gamma)}$$

$$\mathrm{Var}(X) = \delta^2\left(\frac{K_{\lambda+1}(\delta\,\gamma)}{\delta\,\gamma\, K_\lambda(\delta\,\gamma)} + \frac{\beta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\,\gamma)}{K_\lambda(\delta\,\gamma)} - \left(\frac{K_{\lambda+1}(\delta\,\gamma)}{K_\lambda(\delta\,\gamma)}\right)^2\right)\right).$$

The GH distribution has semi-heavy tails, in particular

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

More specifically, the heaviest tail decays as

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

and the lightest as

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

The GH family of distributions is extremely flexible, and many of the special cases are known by alternative names. In what follows, we describe some of the special cases.

The NIG distribution Letting $\lambda = -\frac{1}{2}$, we obtain the normal inverse Gaussian distribution

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

where $\delta > 0$ and $0 < |\beta| < \alpha$. To obtain this density, we use the following properties of the modified Bessel function (Blæsild, 1981)

$$K_{\frac{1}{2}}(x) = 2^{-\frac{1}{2}}\,\sqrt{\pi}\,x^{-\frac{1}{2}}\exp(-x) \quad \text{and} \quad K_\nu(x) = K_{-\nu}(x).$$

Utilising 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 the NIG distribution has semi-heavy tails, i.e.

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

More specifically, the heaviest tail decays as

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

and the lightest as

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

The Cauchy distribution Letting $\lambda = -\frac{1}{2}$, $\beta = 0$, and $\alpha \to 0$, the Caucy distribution with parameters μ and δ is obtained. The tails of the Cauchy distribution decay as

$$f_x(x) \sim \text{const}\,|x|^{-2} \quad \text{as } x \to \pm\infty.$$

The Gaussian distribution Letting $\alpha \to \infty$, $\beta = 0$, and $\sigma^2 = \frac{\delta}{2}$, the NIG distribution reaches the Gaussian distribution with parameters μ and σ^2. The tail behaviour of the standard normal density is

$$f_x(x) \sim \text{const}\,\exp\left(-\frac{1}{2}x^2\right) \quad \text{as } x \to \pm\infty.$$

Note, that neither the Cauchy nor the Gaussian distribution belong to the class of GH distributions.

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

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

where $\delta > 0$. This limiting distribution is calculated using the following properties of the modified Bessel function (Abramowitz and Stegun, 1972)

$$K_\nu(x) = K_{-\nu}(x) \quad \text{and} \quad K_\nu(x) \sim \Gamma(\nu) \, 2^{\nu-1} \, x^{-\nu} \quad \text{for} \quad x \to 0, \nu > 0.$$

Utilising the following property of the modified Bessel function (Abramowitz and Stegun, 1972)

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

we have that the tails of this distribution decay like

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

Hence, the heaviest tail decays as

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

and the lightest as

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

The Student's t-distribution The Student's t-distribution with ν degrees of freedom is obtained as the limit of the skew Student's t-distribution above when $\beta \to 0$, $\mu = 0$ and $\delta = \sqrt{\nu}$:

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

To obtain the limit, we again use the properties of the modified Bessel function:

$$K_{\frac{\nu+1}{2}}\left(\beta \sqrt{\delta^2 + (x-\mu)^2}\right) \underset{\beta \to 0}{\longrightarrow} \Gamma\left(\frac{\nu+1}{2}\right) 2^{\frac{\nu-1}{2}} \left(\beta \sqrt{\delta^2 + (x-\mu)^2}\right)^{-\frac{\nu+1}{2}}.$$

The tails of the Student's t-distribution decline as a power function, i.e.

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

2.2 Estimation using the EM-algorithm

Estimating the parameters of the generalised hyperbolic distributions is a difficult task, even for specified subclasses, uniquely defined by λ. The problem is to maximize

$$\log L(\lambda, \alpha, \beta, \delta, \mu; X_1, ..., X_n) = \sum_{i=1}^{n} \log f_x(x; \lambda, \alpha, \beta, \delta, \mu), \tag{3}$$

where $f_x(x; \lambda, \alpha, \beta, \delta, \mu)$ is a more detailed notation for the density in Equation (1). It becomes easier if we exploit the normal variance-mean mixture of the GH distribution. Then, we 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 describe how the EM-algorithm can be used for estimating the parameters of the GH distribution.

The joint density of X and Z in Equation (2) is given by

$$f_{x,z}(x, z) = f_{x|z}(x|z)\, f_z(z).$$

Hence, one may construct the log-likelihood

$$\begin{aligned} \log \mathrm{L}(\lambda, \alpha, \beta, \delta, \mu; X_1, ..., X_n, Z_1, ..., Z_n) &= \sum_{i=1}^{n} \log f_{x|z}(x_i|z_i; \mu, \beta) + \sum_{i=1}^{n} \log f_z(z_i; \lambda, \delta, \gamma) \\ &= \log \mathrm{L}_1(\mu, \beta) + \log \mathrm{L}_2(\lambda, \delta, \gamma). \end{aligned} \tag{4}$$

Here

$$\log \mathrm{L}_1(\mu, \beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \log z_i - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu - \beta\, z_i)^2}{z_i} \tag{5}$$

and

$$\log \mathrm{L}_2(\lambda, \delta, \gamma) = n\,\lambda \log \frac{\gamma}{\delta} + (\lambda - 1) \sum_{i=1}^{n} \log z_i - n \log\left(2\, K_\lambda(\delta\,\gamma)\right) - \frac{\delta^2}{2} \sum_{i=1}^{n} \frac{1}{z_i} - \frac{\gamma^2}{2} \sum_{i=1}^{n} z_i, \tag{6}$$

with $\gamma = \sqrt{\alpha^2 + \beta^2}$. The EM-algorithm consists in iterating two steps; the expectation step (E-step) and the maximization step (M-step).

M-step: In the M-step, one starts by maximising $\log \mathrm{L}_1(\mu, \beta)$ with respect to the parameters μ and β. At the kth iteration of the algorithm, the estimates for β and μ are

$$\beta^{(k+1)} = \frac{\sum_{i=1}^{n} \frac{x_i}{z_i} - \bar{x} \sum_{i=1}^{n} \frac{1}{z_i}}{n - \bar{z} \sum_{i=1}^{n} \frac{1}{z_i}} \quad \text{and} \quad \mu^{(k+1)} = \bar{x} - \beta^{(k+1)}\, \bar{z}. \tag{7}$$

Next, $\log \mathrm{L}_2(\lambda, \delta, \gamma)$ is maximized with respect to the parameters λ, δ, and γ. In the general case, this maximization must be performed numerically. However, for certain values of λ, the estimates of δ and γ have closed form expressions. This is for instance the case for the NIG distribution ($\lambda = \frac{1}{2}$), see Section 3.2.

In practice, we do not know the values of the variables $z_1, ..., z_n$ (these are the "missing values"). Hence, when computing the estimates in the M-step, we must replace z_i, z_i^{-1} and $\log z_i$ with $E(Z_i|X_i = x_i)$, $E(Z_i^{-1}|X_i = x_i)$ and $E(\log Z_i|X_i = x_i)$, respectively. Performing the E-step amounts to estimating these quantities.

E-step It can be shown (Karlis, 2002) that if $Z \sim GIG\,(\lambda, \delta, \gamma)$ and $[X|Z = z] \sim N(\mu + \beta\, z, z)$, then $Z|X \sim GIG(\lambda - \frac{1}{2}, \sqrt{\delta^2 + (x-\mu)^2}, \alpha)$. It can also be shown (Karlis, 2002) that the moments of the $GIG(\lambda, \delta, \gamma)$ distribution are given by

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

Further, we have (Mencia and Sentana, 2004)

$$\mathrm{E}(\log Z) = \left.\frac{\partial\,\mathrm{E}(Z^r)}{\partial\, r}\right|_{r=0},$$

where

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

Using the fact that

$$\frac{\partial}{\partial\, r} K_{\lambda+r}(\delta\,\gamma) = \frac{\partial}{\partial\,(\lambda+r)} K_{\lambda+r}(\delta\,\gamma)\, \frac{\partial}{\partial\, r}(\lambda + r)$$

and setting $r = 0$, gives

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

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 by Mencia and Sentana (2004), or they might be approximated numerically.

Starting with the E-step, alternating between the steps, the EM-algorithm produces improved parameter estimates in each step (in the sense that the value of the original likelihood in Equation (3) is continually increased). The two steps are iterated until some convergence criterion is satisfied. 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 starting values.

It should be noted that although the above-described procedure in principle can be used to fit the parameter λ, determining λ turns out to be a difficult task. Even for very large sample sizes it may be hard to make a distinction between $\lambda = 1$ and $\lambda = -1/2$ due to the flatness of the GH likelihood function in λ. See for instance Prause (1999).

2.3 Simulation

The most natural way of simulating generalised hyperbolic variables is by exploiting their normal variance-mean mixture structure. Since the mixing distribution is the generalised inverse Gaussian, the resulting algorithm reads as follows

- Sample Z from $GIG\,(\lambda, \delta, \gamma)$.
- Sample Y from N(0,1).
- Return $X = \mu + \beta\, Z + \sqrt{Z}\, Y$.

Sampling from the GIG-distribution is not straightforward. See Atkinson (1982) and Dagpunar (1989) for two different algoritms for this purpose. The first has been implemented in the `HyperbolicDist` package for R (R Development Core Team, 2004), and the latter in `UNU.RAN`, a library of C functions for non-uniform random number generation, developed at the Vienna University of Economics (see `http://statistik.wu-wien.ac.at/unuran/`).

3 The NIG distribution

3.1 Density

As shown in Section 2.1, 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 $0 < |\beta| < \alpha$. If we follow Karlis (2002) and Venter and de Jongh (2002), and let

$$p(x) = \delta\sqrt{\alpha^2-\beta^2} + \beta\,(x-\mu), \quad \text{and} \quad q(x) = \sqrt{\delta^2+(x-\mu)^2},$$

the density may be written as

$$f_x(x) = \frac{\delta\,\alpha}{\pi\,q(x)}\,\exp\left(p(x)\right)\,K_1\left(\alpha\,q(x)\right).$$

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. The steepness parameter, defined by

$$\xi = \left(1+\delta\sqrt{\alpha^2-\beta^2}\right)^{-1/2}, \tag{8}$$

determines the heaviness of the tails. The closer ξ is to 1, the heavier are the tails. The parameter

$$\chi = \beta\xi/\alpha,$$

determines the skewness of the distribution. For $\chi < 0$, the left tail is heavier than the right, for $\chi = 0$, the distribution is symmetric, and for $\chi > 0$, the right tail is heavier. Due to the restrictions on δ, β, and α, we have $0 \le |\chi| < \xi < 1$. The mean and variance of the NIG distribution are given by

$$\mathrm{E}(X) = \mu + \frac{\delta\,\beta}{\sqrt{\alpha^2-\beta^2}} \quad \text{and} \quad \mathrm{Var}(X) = \frac{\delta\,\alpha^2}{(\alpha^2-\beta^2)^{3/2}}.$$

The skewness and kurtosis are given by

$$s = 3\,\frac{\beta}{\alpha}\,\frac{1}{\delta^{1/2}\,(\alpha^2-\beta^2)^{1/4}}$$

$$k = 3\left(1+4\left(\frac{\beta}{\alpha}\right)^2\right)\frac{1}{\delta\,(\alpha^2-\beta^2)^{1/2}}.$$

It follows that the kurtosis-skewness pairs must satisfy $|s| \le \sqrt{\frac{3}{5}\,k}$.

3.2 Estimation using EM-algorithm

The basic methodology is as presented in Section 2.2. Here, we only describe the issues that are specific for the NIG distribution.

M-step $\log \mathrm{L}_1(\mu, \beta)$ is identical to that in Section 2.2. That is,

$$\beta^{(k+1)} = \frac{\sum_{i=1}^n \frac{x_i}{z_i} - \bar{x} \sum_{i=1}^n \frac{1}{z_i}}{n - \bar{z} \sum_{i=1}^n \frac{1}{z_i}} \quad \text{and} \quad \mu^{(k+1)} = \bar{x} - \beta^{(k+1)}\, \bar{z}. \tag{9}$$

For the NIG distribution $Z \sim GIG(-\frac{1}{2}, \delta, \gamma) = IG(\delta, \gamma)$, where $IG(\cdot)$ denotes the inverse Gaussian distribution with density (Karlis, 2002)

$$f_z(z) = \frac{\delta}{\sqrt{2\,\pi}} \exp(\delta\,\gamma) z^{-3/2} \exp\left\{ -\frac{1}{2} \left(\frac{\delta^2}{z} + \gamma^2\, z \right) \right\}.$$

This means that $\log \mathrm{L}_2(\lambda, \delta, \gamma) = \log \mathrm{L}_2(\delta, \gamma)$ is given by

$$\log \mathrm{L}_2(\delta, \gamma) = n \log \delta + n\delta\,\gamma - \frac{\delta^2}{2} \sum_{i=1}^n \frac{1}{z_i} - \frac{\gamma^2}{2} \sum_{i=1}^n z_i. \tag{10}$$

Differentiating $\log \mathrm{L}_2(\delta, \gamma)$ with respect to δ and γ gives

$$\frac{\partial \log \mathrm{L}_2(\delta, \gamma)}{\partial\, \delta} = \frac{n}{\delta} + n\,\gamma - \delta \sum_{i=1}^n \frac{1}{z_i} = 0 \qquad \frac{\partial \log \mathrm{L}_2(\delta, \gamma)}{\partial\, \gamma} = n\,\delta - \gamma \sum_{i=1}^n z_i = 0.$$

Solving these equations with respect to δ and γ gives

$$\delta^{(k+1)} = \sqrt{\frac{n}{\sum_{i=1}^n \frac{1}{z_i} - \frac{n}{\bar{z}}}} \quad \text{and} \quad \gamma = \delta^{(k+1)}/\bar{z}.$$

Having determined $\beta^{(k+1)}$ and γ, the estimate for α is given by

$$\alpha^{(k+1)} = \sqrt{\gamma^2 + \left(\beta^{(k+1)}\right)^2}.$$

E-step For the NIG-distribution, only $E(Z_i|X_i = x_i)$ and $E(Z_i^{-1}|X_i = x_i)$ are needed. In the NIG case $Z|X \sim GIG(-1, q(x), \alpha)$, where $q(x)$ is as defined in Section 3.1. Hence,

$$E(Z_i|X_i = x_i) = \frac{q(x_i)}{\alpha} \frac{K_0\left(\alpha\, q(x_i)\right)}{K_1\left(\alpha\, q(x_i)\right)} \quad \text{and} \quad E(Z_i^{-1}|X_i = x_i) = \frac{\alpha}{q(x_i)} \frac{K_2\left(\alpha\, q(x_i)\right)}{K_1\left(\alpha\, q(x_i)\right)}.$$

Starting values The moment estimates may be used as starting values. Let $\bar{m}_1$, $\bar{m}_2$, $\bar{m}_3$ and $\bar{m}_4$ be the sample mean, standard deviation, skewness, and kurtosis, respectively, of the data, and define

$$\hat{\gamma} = \frac{3}{\bar{m}_2 \sqrt{3\,\bar{m}_4 - 5\,\bar{m}_3^2}}. \tag{11}$$

Then, the moment estimators are given by

$$\begin{aligned} \hat{\mu} &= \bar{m}_1 - \hat{\beta}\hat{\delta}/\hat{\gamma} \\ \hat{\beta} &= (\bar{m}_3\, \bar{m}_2\, \hat{\gamma}^2)/3 \\ \hat{\delta} &= (\bar{m}_2^2 \hat{\gamma}^3)/(\hat{\beta}^2 + \hat{\gamma}^2) \\ \hat{\alpha} &= (\hat{\gamma}^2 + \hat{\beta}^2)^{1/2} \end{aligned}$$

From Equation (11) we see that the moment estimators do not exist if $\bar{m}_3 > \sqrt{\frac{3}{5}\,\bar{m}_4}$.

3.3 Simulation

Rydberg (1997) suggested a fast and very efficient algorithm for generating independent univariate NIG distributed variates. In what follows, the algorithm is described.

- Sample Z from $IG\,(\delta, \gamma)$.
- Sample Y from N(0,1).
- Return $X = \mu + \beta\, Z + \sqrt{Z}\, Y$.

An IG variate Z can be sampled as follows. Set

$$\tau = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}}, \qquad \lambda = \delta^2 \quad \text{and sample} \quad v \sim \chi_1^2.$$

Compute

$$z_1 = \tau + \frac{\tau}{2\,\lambda}\left(\tau\, v - \sqrt{4\,\tau\, v\,\lambda + (\tau\, v)^2}\right), \qquad \text{and} \qquad z_2 = \frac{\tau^2}{z_1}.$$

Let

$$Z = \begin{cases} z_1 & \text{with probability } p, \\ z_2 & \text{with probability } 1 - p \end{cases}$$

where

$$p = \tau/(\tau + z_1).$$

4 The skew Student's t-distribution

4.1 Density

As shown in Section 2, the skew Student's t-distribution is obtained letting $\lambda = -\nu/2$ and $\alpha \to |\beta|$ in the GH distribution. Its density is given by

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

and

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

In the last case we use the fact that

$$K_{\frac{\nu+1}{2}}\left(\beta\,\sqrt{\delta^2 + (x-\mu)^2}\right) \underset{\beta\to 0}{\longrightarrow} \Gamma\left(\frac{\nu+1}{2}\right)\,2^{\frac{\nu-1}{2}}\,\left(\beta\,\sqrt{\delta^2 + (x-\mu)^2}\right)^{-\frac{\nu+1}{2}}.$$

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

$$\mathrm{E}(X) = \mu + \frac{\beta\,\delta^2}{\nu - 2} \tag{12}$$

and

$$\mathrm{Var}(X) = \frac{2\,\beta^2\,\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}. \tag{13}$$

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. They are given by

$$s = \frac{2(\nu-4)^{1/2}\beta\delta}{\left[2\beta^2\delta^2 + (\nu-2)(\nu-4)\right]^{3/2}}\left[3(\nu-2) + \frac{8\beta^2\delta^2}{\nu-6}\right] \tag{14}$$

and

$$k = \frac{6}{\left[2\,\beta^2\,\delta^2 + (\nu-2)(\nu-4)\right]^2}\left[(\nu-2)^2(\nu-4) + \frac{16\,\beta^2\delta^2(\nu-2)(\nu-4)}{\nu-6} + \frac{8\,\beta^4\delta^4(5\,\nu-22)}{(\nu-6)(\nu-8)}\right]. \tag{15}$$

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

4.2 Estimation using EM-algorithm

The basic methodology is as documented in Section 2.2. Here, we only describe the issues that are specific for the skew Student's t-distribution.

M-step $\log L_1(\mu, \beta)$ is identical to that in Section 2.2. That is,

$$\beta^{(k+1)} = \frac{\sum_{i=1}^n \frac{x_i}{z_i} - \bar{x}\sum_{i=1}^n \frac{1}{z_i}}{n - \bar{z}\sum_{i=1}^n \frac{1}{z_i}} \quad \text{and} \quad \mu^{(k+1)} = \bar{x} - \beta^{(k+1)}\,\bar{z}. \tag{16}$$

For the skew Student's t-distribution $Z \sim GIG(-\frac{\nu}{2}, \delta, 0)$. If we use the fact that

$$K_{\frac{\nu}{2}}\,(\delta\,\gamma) \underset{\gamma\to 0}{\longrightarrow} \Gamma\left(\frac{\nu}{2}\right)\,2^{\frac{\nu}{2}-1}\,(\delta\,\gamma)^{-\frac{\nu}{2}},$$

the density of Z can be written as

$$f_z(z) = \frac{\delta^\nu\,z^{-\frac{\nu}{2}-1}}{\Gamma\left(\frac{\nu}{2}\right)\,2^{\frac{\nu}{2}}}\,\exp\left\{-\frac{\delta^2}{2}\,z^{-1}\right\}.$$

$f_z(z)$ can be recognised as the density of the Inverse Gamma distribution with parameters $\nu/2$ and $\delta^2/2$. This means that $\log L_2(\lambda, \delta, \gamma) = \log L_2(\nu, \delta)$ is given by

$$\log L_2(\nu, \delta) = -n\,\log\Gamma\left(\frac{\nu}{2}\right) - \frac{n\,\log 2}{2}\,\nu + n\,\nu\,\log\delta - \frac{\nu+2}{2}\sum_{i=1}^n \log z_i - \frac{\delta^2}{2}\sum_{i=1}^n \frac{1}{z_i}. \tag{17}$$

Differentiating $\log \mathrm{L}_2(\nu, \delta)$ with respect to δ gives

$$\frac{\partial\,\log \mathrm{L}_2(\delta, \gamma)}{\partial\,\delta} = \frac{n\,\nu}{\delta} - \delta\sum_{i=1}^n \frac{1}{z_i} = 0.$$

Solving this equation with respect to δ gives

$$\delta^{(k+1)} = \sqrt{\frac{n\,\hat{\nu}}{\sum_{i=1}^{n} \frac{1}{z_i}}}. \tag{18}$$

Differentiating $\log \mathrm{L}_2(\nu, \delta)$ with respect to ν gives

$$\frac{\partial \log \mathrm{L}_2(\delta, \gamma)}{\partial \nu} = -\frac{n \log 2}{2} + n \log \delta - \frac{1}{2} \sum_{i=1}^{n} \frac{1}{z_i} - n \frac{\partial \log \Gamma\left(\frac{\nu}{2}\right)}{\partial \nu} = 0.$$

We have that

$$\frac{\partial \log \Gamma\left(\frac{\nu}{2}\right)}{\partial \nu} = \frac{1}{2} \Psi\left(\frac{\nu}{2}\right),$$

where $\Psi(\cdot)$ is the Digamma function. Utilising this, and inserting $\delta^{(k+1)}$ from Equation (18) gives the following equation for $\nu^{(k+1)}$

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

E-step For the skew Student's t-distribution, all of $E(Z_i|X_i = x_i)$, $E(Z_i^{-1}|X_i = x_i)$ and $E(\log Z_i|X_i = x_i)$ are needed. In this case case, $Z|X \sim GIG(-\frac{(\nu+1)}{2}, \sqrt{\delta^2 + (x-\mu)^2}, |\beta|)$. Define $q(x_i) = \sqrt{\delta^2 + (x_i - \mu)^2}$. Then,

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

Moreover, we have

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

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. The EM-algorithm can be programmed in any statistical package supporting Bessel functions with fractional order, e.g. R (R Development Core Team, 2004).

Starting values The moment estimates may be used as starting values. Let $\bar{m}_1$, $\bar{m}_2$, $\bar{m}_3$ and $\bar{m}_4$ be the sample mean, standard deviation, skewness, and kurtosis, respectively, of the data. Then, according to Equations (12)-(15), the moment estimates for μ, β and δ are given by

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

There are two possible values for δ^2. We always choose the smallest of these (the solution corresponding to the minus sign). The moment estimate for ν has to be determined numerically. It is the solution of the equation

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

where κ is given by

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

The search for ν is conducted within a limited area. The search for ν is conducted within a limited area, determined by the following conditions.

For the kurtosis to exist, we must have

$$\nu > 8. \tag{19}$$

For $\hat{\delta}^2$ to exist, we must have

$$1 - \frac{(3\nu^2-2\nu-32)\,(12(5\nu-22)-(\nu-6)(\nu-8)\bar{m}_4)}{216(\nu-2)^2(\nu-4)} > 0,$$

which gives the following equation in ν:

$$3\bar{m}_4^3 + (36 - 20\bar{m}_4)\nu^2 - (528 + 20\bar{m}_4)\nu + (1488 + 192\bar{m}_4) > 0. \tag{20}$$

For $\hat{\delta}^2$ to be positive, we must have

$$1 - \frac{(3\nu^2-2\nu-32)\,(12(5\nu-22)-(\nu-6)(\nu-8)\bar{m}_4)}{216(\nu-2)^2(\nu-4)} < 1,$$

which gives the following equation in ν:

$$\frac{30}{\bar{m}_4} + 1 - \frac{1}{\bar{m}_4}\sqrt{\bar{m}_4^2 + 156\bar{m}_4 + 900} < \nu < \frac{30}{\bar{m}_4} + 1 + \frac{1}{\bar{m}_4}\sqrt{\bar{m}_4^2 + 156\bar{m}_4 + 900}. \tag{21}$$

For $\hat{\beta}$ to exist, we must have

$$\bar{m}_2(\nu-2) - \hat{\delta}^2 > 0,$$

which gives the following condition for ν:

$$\nu > \max\left(8, \frac{6}{\bar{m}_4} + 4\right). \tag{22}$$

All four conditions given by Equations (19)-(22) must be fulfilled for the moment estimators to exist.

4.3 Simulation

- Sample Z from Inverse Gamma$(\nu/2, \delta^2/2)$.
- Sample Y from N(0,1).
- Return $X = \mu + \beta\, Z + \sqrt{Z}\, Y$.

A sample Z from the Inverse Gamma distribution is obtained by sampling $R \sim$ Gamma$(\nu/2, \delta^2/2)$ and setting $Z = 1/R$.

References

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

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

A. C. Atkinson. The simulation of generalized inverse gaussian and hyperbolic random variables. *SIAM Journal of Scientific & Statistical Computing*, 3:502–515, 1982.

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

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

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

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

O. E. Barndorff-Nielsen and R. Stelzer. Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse gaussian lévy processes. *Submitted for publication*, 2004.

P. Blæsild. The two-dimensional hyperbolic distribution and related distributions, with applications to Johannsen's bean data. *Biometrika*, 68:251–263, 1981.

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

J. S. Dagpunar. An easily implemented generalized inverse gaussian generator. *Communications in Statistics - Simulations*, 18:703–710, 1989.

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

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

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

L. Forsberg and T. Bollerslev. 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, 2002.

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

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

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

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

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

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

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

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. URL `http://www.R-project.org`. ISBN 3-900051-00-3.

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

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