Tail behaviour and tail dependence of generalized hyperbolic distributions

Ernst August v. Hammerstein

Abstract Generalized hyperbolic distributions have been well established in finance during the last two decades. However, their application often is computationally demanding because values of their distribution and quantile functions can only be determined by numerically integrating their densities. Moreover, they are, in general, not stable under convolution which makes the computation of quantiles in factor models driven by these distributions even more complicated. In the first part of the present paper, we take a closer look at the tail behaviour of univariate generalized hyperbolic distributions and their convolutions and provide asymptotic formulas for the quantile functions that allow for an approximative calculation of quantiles for very small resp. large probabilities. With help of these results, we then analyze the dependence structure of multivariate generalized hyperbolic distributions. In particular, we concentrate on the implied copula and determine its tail dependence coefficients. Our main result states that the generalized hyperbolic copula can only attain the two extremal values 0 or 1 for the latter, that is, it is either tail independent or completely dependent. We provide necessary conditions for each case to occur as well as a simpler criterion for tail independence. Possible limit distributions of the generalized hyperbolic family are also included in our investigations.

1 Introduction

Almost forty years ago, generalized hyperbolic distributions (henceforth GH) have been introduced in [4] in connection with the modeling of aeolian sand deposits and dune movements. Eighteen years later, they were introduced in finance by [12] where the hyperbolic subclass was used as a more realistic model for stock returns.

Ernst August v. Hammerstein
Department of Quantitative Finance, University of Freiburg,
Platz der Alten Synagoge, D-79098 Freiburg, Germany
e-mail: ernst.august.hammerstein@finance.uni-freiburg.de

The normal inverse Gaussian subclass followed shortly after in [5], and the general case was then considered in [10], [17], and [16]. Starting from early applications to stock price modeling and option pricing, GH distributions have been successfully used in various fields of finance during the last two decades, for example, in interest rate theory and the pricing of interest rate derivatives (see [18], [13], and [14]), currency markets ([15]), and portfolio credit risk models ([11]).

There are mainly two reasons for the widespread applicability of GH distributions: First, they are infinitely divisible and therefore allow to make use of the extensive theory of Lévy processes in continuous-time models based on them. The second is their convenient tail behaviour. On the one hand, the tails of the GH probability densities have considerably more mass than the ones of normal distributions. This means that, for example, extreme price movements which are observed more often nowadays are much more likely under the assumption of GH distributed asset returns, whereas such events are severely underestimated in models based on the normal distribution. On the other hand, the GH densities asymptotically still decay exponentially and therefore possess a moment generating function on some non-degenerate interval around the origin. This is an inevitable requirement for the construction of financial models of exponential Lévy type since stock or bond price models having infinite expectations are obviously not very realistic. Moreover, the existence of a moment generating function allows (under a mild additional assumption) for an easy way to explicitly determine a risk-neutral measure for derivative pricing via an Esscher transform.

Despite these advantages, GH distributions are, to some extent, computationally demanding in practical applications nevertheless because their distribution functions can neither be given in closed form, nor does there exist a well-known and quickly convergent series expansion for them. Therefore, the values of the distribution and quantile functions can only be determined by numerically integrating the corresponding densities. These procedures naturally become less stable and reliable if the arguments of the distribution function are extremely large resp. the probabilities inserted into the quantile function are very close to zero or one. The latter difficulty can occur in risk management, especially in credit risk, where one has to calculate values at risk or expected shortfalls for probabilities beyond 99% and to deal with small default probabilities. In the first part of this paper, we analyze the tail behaviour of univariate GH distributions in greater detail and derive asymptotic formulas for the distribution and quantile functions of GH distributions and their convolutions that enable a simple calculation of approximative values of the latter.

The other major topic we are concerned with in the second part of the paper is the dependence structure of multivariate GH distributions. In practice, correlation still seems to be the predominant dependence measure although it only provides a complete characterization of dependencies in case of a multivariate normal distribution. The dependence structure of the latter is indeed linear and fully described by the corresponding correlation matrix. However, the picture changes significantly if one departs from the normal world. In general, zero correlation does not imply independence, and maximal dependence (co- or countermonotonicity) can already occur for correlations with absolute value strictly smaller than one. We will show that the latter also holds for multivariate GH distributions.

Another dependence concept that has gained increasing attention, especially in credit portfolio modeling, is tail dependence. Roughly speaking, the tail dependence coefficients give the asymptotic probabilities of joint extremal events which may be, for example, multiple defaults in a credit portfolio within the same time interval or severe losses of different stocks at the same trading day. Tail dependence is solely determined by the implied copula which is inherent in every multivariate distribution and—in contrast to correlation—completely characterizes the dependence structure of the latter. The implied copula of a multivariate normal distribution is known to be tail independent, that is, extreme marginal outcomes occur (asymptotically) independent from each other. In credit and insurance risk modeling, this property often is not realistic, therefore dependence models in this area are usually based on copulas possessing tail dependence coefficients greater than zero like the t- or grouped t-copula (see [9]). To see whether the implied copula of a multivariate GH distribution provides a suitable model in this context, we determine the potential range of its tail dependence coefficients. It turns out that only the two extremal values 0 or 1 can be obtained, implying that the GH copula either is tail independent or completely dependent. For both cases, we derive explicit conditions on the GH parameters as well as a simpler criterion for tail independence.

The paper is structured as follows: In the next section, we recall the definition of univariate GH distributions as normal mean-variance mixtures, determine possible limit distributions and provide some useful facts on normal mean-variance mixtures in general as well as the mixing generalized inverse Gaussian distributions which will be required later on. Section 3 then is devoted to a thorough study of the tail behaviour of univariate GH distributions and their convolutions. Multivariate GH distributions and their weak limits are introduced in Section 4, where also the most important properties for the subsequent analysis of its dependence structure are discussed. The latter is done in Section 5 which finishes the paper.

2 Univariate GIG and GH distributions and some of their limits

Generalized hyperbolic distributions can be defined as normal mean-variance mixtures where the mixing distribution is a generalized inverse Gaussian (GIG) one. For the convenience of the reader, we first define normal mean-variance mixtures in general and provide some of their properties which might be of their own interest.

Definition 1. A real valued random variable X is said to have a *normal mean-variance mixture distribution* if

$$X \stackrel{d}{=} \mu + \beta Z + \sqrt{Z} W,$$

where $\mu, \beta \in \mathbb{R}$, $W \sim N(0,1)$ and $Z \sim G$ is a real-valued, non-negative random variable which is independent of W. Equivalently, a probability measure F on $(\mathbb{R}, \mathscr{B})$

is said to be a normal mean-variance mixture if

$$F(\mathrm{d}x) = \int_{\mathbb{R}_+} N(\mu + \beta y, y)(\mathrm{d}x)\, G(\mathrm{d}y),$$

where the mixing distribution G is a probability measure on $(\mathbb{R}_+, \mathscr{B}_+)$. We shall use the short hand notation $F = N(\mu + \beta y, y) \circ G$.

The most important facts about normal mean-variance mixtures are summarized in the following lemma. It especially shows that properties like stability under convolutions and weak convergence are inherited from the mixing distributions. A detailed proof can be found in [23, Lemmas 1.6 and 1.7].

Lemma 1. *Let $\mathbb{G}$ be a class of probability distributions on $(\mathbb{R}_+, \mathscr{B}_+)$ and suppose $G, G_1, G_2 \in \mathbb{G}$.*

a) If G possesses a moment generating function $M_G(u) = \int_{\mathbb{R}_+} e^{ux} G(\mathrm{d}x)$ on some open interval (a,b) with $a < 0 < b$, then $F = N(\mu + \beta y, y) \circ G$ also possesses a moment generating function and $M_F(u) = e^{\mu u} M_G(\frac{u^2}{2} + \beta u)$, $a < \frac{u^2}{2} + \beta u < b$.
b) If $G = G_1 * G_2 \in \mathbb{G}$, then $\big(N(\mu_1 + \beta y, y) \circ G_1\big) * \big(N(\mu_2 + \beta y, y) \circ G_2\big) = N(\mu_1 + \mu_2 + \beta y, y) \circ G$.
c) If $(\mu_n)_{n \geq 1}$ and $(\beta_n)_{n \geq 1}$ are convergent sequences of real numbers with finite limits $\mu, \beta < \infty$, and $(G_n)_{n \geq 1}$ is a weakly convergent sequence of mixing distributions with $G_n \xrightarrow{w} G$, then $N(\mu_n + \beta_n y, y) \circ G_n \xrightarrow{w} N(\mu + \beta y, y) \circ G$.

We now leave the general case and concentrate on a specific class $\mathbb{G}$ of mixing distributions, namely the generalized inverse Gaussian one mentioned above. This class was introduced more than 50 years ago (one of the first papers where its densities are mentioned is [22]) and rediscovered in [33], [34], and [4]. An extensive survey with statistical applications can be found in [25]. The density of a GIG distribution is as follows:

$$d_{GIG(\lambda,\delta,\gamma)}(x) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{1}{2K_\lambda(\delta\gamma)}\, x^{\lambda-1} e^{-\frac{1}{2}\left(\delta^2 x^{-1} + \gamma^2 x\right)}\, \mathbb{1}_{(0,\infty)}(x), \tag{1}$$

where $K_\lambda(x)$ denotes the modified Bessel function of third kind with index λ. Permitted parameters are

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

Parametrizations with $\delta = 0$ or $\gamma = 0$ have to be understood as limiting cases. Using the asymptotic behaviour

$$K_\lambda(x) \sim \frac{\Gamma(|\lambda|)}{2} \left(\frac{x}{2}\right)^{-|\lambda|}, \qquad x \downarrow 0,\ \lambda \neq 0, \tag{2}$$

of the Bessel functions where $\Gamma(x)$ denotes the Gamma function, the limit for $\lambda > 0$ is obtained as

$$\lim_{\delta\to 0} d_{GIG(\lambda,\delta,\gamma)}(x) = \left(\frac{\gamma^2}{2}\right)^{\lambda} \frac{x^{\lambda-1}}{\Gamma(\lambda)} e^{-\frac{\gamma^2}{2}x} \mathbb{1}_{(0,\infty)}(x) = d_{G(\lambda,\frac{\gamma^2}{2})}(x), \tag{3}$$

which is nothing but the density of a Gamma distribution $G(\lambda, \frac{\gamma^2}{2})$ with shape parameter λ and scale parameter $\frac{\gamma^2}{2}$. For $\lambda < 0$, we arrive at

$$\lim_{\gamma\to 0} d_{GIG(\lambda,\delta,\gamma)}(x) = \left(\frac{2}{\delta^2}\right)^{\lambda} \frac{x^{\lambda-1}}{\Gamma(-\lambda)} e^{-\frac{\delta^2}{2x}} \mathbb{1}_{(0,\infty)}(x) = d_{iG(\lambda,\frac{\delta^2}{2})}(x) \tag{4}$$

which equals the density of an inverse Gamma distribution $iG(\lambda, \frac{\delta^2}{2})$.

For $|\lambda| = \frac{1}{2}$, the Bessel function $K_\lambda(x)$ can be given in explicit form: We have $K_{-\frac{1}{2}}(x) = K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}\, e^{-x}$, thus

$$d_{GIG(-\frac{1}{2},\delta,\gamma)}(x) = \frac{\delta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2x}(\gamma x-\delta)^2} \mathbb{1}_{(0,\infty)}(x) \tag{5}$$

which is the density of an inverse Gaussian distribution $IG(\delta,\gamma)$, showing that the GIG distributions are, in fact, a natural extension of this subclass.

A distribution G on $(\mathbb{R}_+, \mathscr{B}_+)$ is completely characterized by its Laplace transform $\mathfrak{L}_G(u) = \int_{\mathbb{R}_+} e^{-ux} G(\mathrm{d}x)$ from which many properties of G can be easily derived. For the GIG class, we obtain the following representations (see [23, Proposition 1.9] for a proof).

Proposition 1. *The Laplace transforms of GIG distributions are given by*

$$\begin{aligned}
\mathfrak{L}_{GIG(\lambda,\delta,\gamma)}(u) &= \left(\frac{\gamma}{\sqrt{\gamma^2+2u}}\right)^{\lambda} \frac{K_\lambda\big(\delta\sqrt{\gamma^2+2u}\big)}{K_\lambda(\delta\gamma)}, && \delta,\gamma > 0,\\
\mathfrak{L}_{G(\lambda,\frac{\gamma^2}{2})}(u) &= \left(1+\frac{2u}{\gamma^2}\right)^{-\lambda}, && \lambda > 0,\\
\mathfrak{L}_{iG(\lambda,\frac{\delta^2}{2})}(u) &= \left(\frac{2}{\delta\sqrt{2u}}\right)^{\lambda} \frac{2K_\lambda\big(\delta\sqrt{2u}\big)}{\Gamma(-\lambda)}, && \lambda < 0.
\end{aligned}$$

With help of the preceding proposition and the fact that $\mathfrak{L}_{G_1}(u)\mathfrak{L}_{G_2}(u) = \mathfrak{L}_G(u)$ implies $G_1 * G_2 = G$, one can derive the subsequent convolution properties of GIG distributions:

$$\begin{aligned}
&\text{a)}\quad IG(\delta_1,\gamma) * IG(\delta_2,\gamma) = IG(\delta_1+\delta_2,\gamma),\\
&\text{b)}\quad IG(\delta_1,\gamma) * GIG\big(\tfrac{1}{2},\delta_2,\gamma\big) = GIG\big(\tfrac{1}{2},\delta_1+\delta_2,\gamma\big),\\
&\text{c)}\quad GIG(-\lambda,\delta,\gamma) * G\big(\lambda,\tfrac{\gamma^2}{2}\big) = GIG(\lambda,\delta,\gamma), \qquad \lambda > 0,\\
&\text{d)}\quad G\big(\lambda_1,\tfrac{\gamma^2}{2}\big) * G\big(\lambda_2,\tfrac{\gamma^2}{2}\big) = G\big(\lambda_1+\lambda_2,\tfrac{\gamma^2}{2}\big), \qquad \lambda_1,\lambda_2 > 0.
\end{aligned} \tag{6}$$

Further observe that all $GIG(\lambda,\delta,\gamma)$-distributions with $\gamma>0$ decay at an exponential rate for $x\to\infty$, so they possess moments of arbitrary order, and the moment generating functions are given by

$$M_{GIG(\lambda,\delta,\gamma)}(u)=\int_0^\infty e^{ux}\,d_{GIG(\lambda,\delta,\gamma)}(x)\,\mathrm{d}x=\mathfrak{L}_{GIG(\lambda,\delta,\gamma)}(-u),\quad u\in\big(-\infty,\tfrac{\gamma^2}{2}\big).\tag{7}$$

After these preliminaries, we can now study the class of generalized hyperbolic distributions which have been introduced in the seminal paper [4], motivated by empirical statistics of aeolian sand deposits. The GH distributions are defined as normal mean-variance mixtures with a GIG mixing distribution as follows:

$$GH(\lambda,\alpha,\beta,\delta,\mu):=N(\mu+\beta y,y)\circ GIG\big(\lambda,\delta,\sqrt{\alpha^2-\beta^2}\big).\tag{8}$$

The parameter restrictions for GIG distributions immediately imply that the GH parameters have to fulfill the constraints

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

As before, parametrizations with $\delta=0$ and $|\beta|=\alpha$ have to be understood as limiting cases which by Lemma 1 c) equal normal mean-variance mixtures with the corresponding GIG limit distributions. We defer a more precise introduction of the latter and first concentrate on GH distributions with parameters $\delta>0$ and $|\beta|<\alpha$. Their Lebesgue densities are given by

$$\begin{aligned}d_{GH(\lambda,\alpha,\beta,\delta,\mu)}(x)&=\int_0^\infty d_{N(\mu+\beta y,y)}(x)\,d_{GIG\left(\lambda,\delta,\sqrt{\alpha^2-\beta^2}\right)}(y)\,\mathrm{d}y\\&=a(\lambda,\alpha,\beta,\delta,\mu)\big(\delta^2+(x-\mu)^2\big)^{(\lambda-\frac12)/2}K_{\lambda-\frac12}\Big(\alpha\sqrt{\delta^2+(x-\mu)^2}\Big)\,e^{\beta(x-\mu)}\end{aligned}\tag{9}$$

with the norming constant

$$a(\lambda,\alpha,\beta,\delta,\mu)=\frac{\big(\alpha^2-\beta^2\big)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\,\alpha^{\lambda-\frac12}\,\delta^\lambda K_\lambda\big(\delta\sqrt{\alpha^2-\beta^2}\,\big)}.\tag{10}$$

A closer look at the densities reveals that the influence of the parameters is as follows: α determines the shape, β the skewness, μ is a location parameter, and δ serves for scaling. λ characterizes certain subclasses and considerably influences the size of mass contained in the tails. Setting $\lambda=-\frac12$ leads to the subclass of normal inverse Gaussian distributions (NIG). By (8), these are the normal mean-variance mixtures arising from inverse Gaussian mixing distributions which explains their name. With the symmetry relation $K_{-\lambda}(x)=K_\lambda(x)$ and the aforementioned representation of $K_{-\frac12}(x)$, its densities are obtained from (9) and (10) as

$$d_{NIG(\alpha,\beta,\delta,\mu)}(x) = \frac{\alpha\delta}{\pi}\,\frac{K_1\big(\alpha\sqrt{\delta^2+(x-\mu)^2}\big)}{\sqrt{\delta^2+(x-\mu)^2}}\,e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)}. \tag{11}$$

From Lemma 1 a), Proposition 1, and equation (7) we conclude that all GH distributions with parameters $\delta > 0$ and $|\beta| < \alpha$ possess a moment generating function of the following form:

$$\begin{aligned} M_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) &= e^{\mu u} M_{GIG(\lambda,\delta,\sqrt{\alpha^2-\beta^2})}\big(\tfrac{u^2}{2}+\beta u\big) \\ &= e^{\mu u}\,\mathfrak{L}_{GIG(\lambda,\delta,\sqrt{\alpha^2-\beta^2})}\big(-\tfrac{u^2}{2}-\beta u\big) \\ &= e^{\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\frac{\lambda}{2}} \frac{K_\lambda\big(\delta\sqrt{\alpha^2-(\beta+u)^2}\big)}{K_\lambda\big(\delta\sqrt{\alpha^2-\beta^2}\big)}, \end{aligned} \tag{12}$$

which is defined for all $u \in (-\alpha-\beta, \alpha-\beta)$. The characteristic functions of GH distributions are easily obtained via the relation

$$\phi_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) = \int_{\mathbb{R}} e^{iux} d_{GH(\lambda,\alpha,\beta,\delta,\mu)}(x)\,\mathrm{d}x = M_{GH(\lambda,\alpha,\beta,\delta,\mu)}(iu). \tag{13}$$

The limit distributions emerging in the case of $\lambda > 0$ and $\delta \to 0$ are also known as Variance Gamma distributions (VG). By Lemma 1 c) and equation (3), they are normal mean-variance mixtures of the following form:

$$VG(\lambda,\alpha,\beta,\mu) = N(\mu+\beta y, y)\circ G\big(\lambda, \tfrac{\alpha^2-\beta^2}{2}\big). \tag{14}$$

Using the asymptotic relationship (2), the corresponding densities can be obtained as pointwise limits (for $x-\mu \neq 0$) of the GH densities:

$$\begin{aligned} d_{VG(\lambda,\alpha,\beta,\mu)}(x) &= \lim_{\delta\to 0} d_{GH(\lambda,\alpha,\beta,\delta,\mu)}(x) \\ &= \frac{(\alpha^2-\beta^2)^\lambda}{\sqrt{\pi}\,(2\alpha)^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\,|x-\mu|^{\lambda-\frac{1}{2}}\,K_{\lambda-\frac{1}{2}}(\alpha|x-\mu|)\,e^{\beta(x-\mu)}. \end{aligned} \tag{15}$$

This class was introduced in [27] (symmetric case $\beta = \theta = 0$) and [26] (general case), but with a different parametrization $VG(\sigma,\nu,\theta,\tilde{\mu})$. The latter is obtained by

$$\sigma^2 = \frac{2\lambda}{\alpha^2-\beta^2}, \qquad \nu = \frac{1}{\lambda}, \qquad \theta = \beta\sigma^2 = \frac{2\beta\lambda}{\alpha^2-\beta^2}, \qquad \tilde{\mu} = \mu.$$

Lemma 1 a), Proposition 1, and equation (7) imply that all VG distributions possess a moment generating function which is given by

$$M_{VG(\lambda,\alpha,\beta,\mu)}(u) = e^{\mu u}\mathfrak{L}_{GIG(\lambda,0,\sqrt{\alpha^2-\beta^2})}\big(-\tfrac{u^2}{2}-\beta u\big) = e^{\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda} \tag{16}$$

for all $u \in (-\alpha-\beta, \alpha-\beta)$.

For $\lambda < 0$, there are two possible limit cases. If $\alpha, \beta \to 0$, Lemma 1 c) and equation (4) imply that the limit distributions are normal mean-variance mixtures

$$t(\lambda, \delta, \mu) = N(\mu, y) \circ iG\big(\lambda, \tfrac{\delta^2}{2}\big) \tag{17}$$

which equal scaled and shifted t-distributions with $f = -2\lambda$ degrees of freedom (the usual Student's t-distribution is obtained with $\delta^2 \equiv -2\lambda$). The associated densities can again be obtained as pointwise limits of the GH densities:

$$d_{t(\lambda,\delta,\mu)}(x) = \lim_{\alpha,\beta \to 0} d_{GH(\lambda,\alpha,\beta,\delta,\mu)}(x) = \frac{\Gamma\big(-\lambda + \frac{1}{2}\big)}{\sqrt{\pi\delta^2}\,\Gamma(-\lambda)} \left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda - \frac{1}{2}}. \tag{18}$$

The other class of limit distributions for $\lambda < 0$ is obtained by letting $|\beta| \to \alpha > 0$. Again by Lemma 1 c) and equation (4), these are normal mean-variance mixtures given by

$$GH(\lambda, \alpha, \pm\alpha, \delta, \mu) = N(\mu \pm \alpha y, y) \circ iG\big(\lambda, \tfrac{\delta^2}{2}\big) \tag{19}$$

and possessing the density

$$\begin{aligned} d_{GH(\lambda,\alpha,\pm\alpha,\delta,\mu)}(x) = {}& \frac{2^{\lambda+\frac{1}{2}}}{\sqrt{\pi}\,\alpha^{\lambda-\frac{1}{2}}\,\delta^{2\lambda}\,\Gamma(-\lambda)} \big(\delta^2 + (x-\mu)^2\big)^{(\lambda-\frac{1}{2})/2} \\ & \times K_{\lambda-\frac{1}{2}}\big(\alpha\sqrt{\delta^2 + (x-\mu)^2}\big)\, e^{\pm\alpha(x-\mu)}. \end{aligned} \tag{20}$$

This type of distribution was called generalized hyperbolic skew Student t-distribution and applied to financial data in [1].

Let us close this section by remarking that also the normal and GIG distributions themselves can emerge as potential limits of univariate GH distributions. But the tail behaviour and tail dependence of the former are already well-known, and for GIG distributions there does not seem to exist a natural multivariate version of which the tail dependence could be studied, therefore we tacitly ignore these two limiting cases here and in the following.

3 Tail behaviour of GH distributions and their convolutions

From the existence of a moment generating function one can already conclude that the tails of the GH densities with $0 \le |\beta| < \alpha$ decay at an exponential rate. More precisely, for $|x| \to \infty$ we have $\delta^2 + (x-\mu)^2 \sim x^2$, and the asymptotic behaviour of the Bessel functions

$$K_\lambda(x) \sim \sqrt{\frac{\pi}{2x}}\, e^{-x}, \quad x \to \infty, \tag{21}$$

further implies $K_{\lambda-\frac{1}{2}}\big(\alpha\sqrt{\delta^2+(x-\mu)^2}\big) \sim \sqrt{\frac{\pi}{2\alpha}}\,|x|^{-1/2}\,e^{-\alpha|x|}$, so we obtain from equation (9)

$$d_{GH(\lambda,\alpha,\beta,\delta,\mu)}(x) \sim c\,|x|^{\lambda-1}\,e^{-\alpha|x|+\beta x}, \quad x\to\pm\infty, \tag{22}$$

where $c=\sqrt{\frac{\pi}{2\alpha}}a(\lambda,\alpha,\beta,\delta,\mu)$, and $a(\lambda,\alpha,\beta,\delta,\mu)$ is the norming constant from (10). Completely analogously, we infer from equations (15) and (21) that

$$d_{VG(\lambda,\alpha,\beta.\mu)}(x) \sim \tilde{c}\,|x|^{\lambda-1}\,e^{-\alpha|x|+\beta x}, \quad x\to\pm\infty, \tag{23}$$

where $\tilde{c}=\frac{(\alpha^2-\beta^2)^\lambda}{(2\alpha)^\lambda\Gamma(\lambda)}$. Thus the GH and VG densities have semi-heavy tails in the sense of the following

Definition 2. A probability density f with support $\mathbb{R}$ has *semi-heavy tails* if there exist some constants $a_1,a_2\in\mathbb{R}$ and $b_1,b_2,c_1,c_2>0$ such that

$$f(x)\sim c_1\,|x|^{a_1}\,e^{-b_1|x|}, \quad x\to-\infty, \quad \text{and} \quad f(x)\sim c_2\,x^{a_2}\,e^{-b_2x}, \quad x\to+\infty.$$

From the above definition, it can be easily deduced that every probability distribution F having a Lebesgue density f with semi-heavy tails also possesses a moment generating function which is defined at least on the open interval $(-b_1,b_2)$. In case of GH and VG distributions we have $a_1=a_2=\lambda-1$, $b_1=\alpha+\beta$, $b_2=\alpha-\beta$ and $c_1=c_2=c$ resp. $c_1=c_2=\tilde{c}$.

A remarkable and probably surprising property of densities with semi-heavy tails is that the tail behaviour of the corresponding distribution functions is the same up to a multiplicative constant, which is shown in the next proposition.

Proposition 2. *Let f be a probability density with semi-heavy tails characterized by a_1,a_2,b_1,b_2,c_1,c_2, $F(x):=\int_{-\infty}^x f(y)\,\mathrm{d}y$ be the associated distribution function and $\bar{F}(x):=1-F(x)$. Then $f(x)\sim b_1\,F(x)$ as $x\to-\infty$ and $f(x)\sim b_2\,\bar{F}(x)$ as $x\to+\infty$.*

Proof. Let us consider the right tail $\bar{F}(x)$ first. From the assumptions we get, using partial integration,

$$\bar{F}(x)=\int_x^\infty f(y)\,\mathrm{d}y \sim c_2\int_x^\infty y^{a_2}\,e^{-b_2y}\,\mathrm{d}y=\frac{c_2}{b_2}\,x^{a_2}\,e^{-b_2x}+\frac{c_2a_2}{b_2}\int_x^\infty y^{a_2-1}\,e^{-b_2y}\,\mathrm{d}y.$$

The claim now follows if we can show that $\big(\int_x^\infty y^{a_2-1}\,e^{-b_2y}\mathrm{d}y\big)\big(x^{a_2}e^{-b_2x}\big)^{-1}\to 0$ as $x\to\infty$. But the latter quotient equals

$$\frac{1}{x}\int_x^\infty\left(\frac{y}{x}\right)^{a_2-1}e^{-b_2(y-x)}\mathrm{d}y=\frac{1}{x}\int_0^\infty\left(\frac{y+x}{x}\right)^{a_2-1}e^{-b_2y}\,\mathrm{d}y$$

and thus converges to zero as $x\to\infty$ because the existence of an integrable majorant ensures that the integral on the right hand side remains bounded. Possible majorants are $g(y)=(y+1)^{a_2-1}e^{-b_2y}$ if $a_2>1$ and $g(y)=e^{-b_2y}$ if $a_2\le 1$. Using the change of variables $z=-y$ we see that for $x\to-\infty$

$$F(x) \sim c_1 \int_{-\infty}^{x} |y|^{a_1}\, e^{-b_1|y|}\,\mathrm{d}y = c_1 \int_{|x|}^{\infty} z^{a_1}\, e^{-b_1 z}\,\mathrm{d}z,$$

hence the assertion for the left tail immediately follows from what we have proven above. □

The tail behaviour of the t-distributions, however, can be derived much easier. The asymptotics of the corresponding densities are easily seen from (18) to equal $d_{t(\lambda,\delta,\mu)}(x) \sim \bar{c}\,|x|^{2\lambda-1}$, $x \to \pm\infty$. Hence, $F_{t(\lambda,\delta,\mu)}(x) \sim \frac{\bar{c}}{2\lambda-1}|x|^{2\lambda}$. The knowledge of the tail behaviour allows to derive the asymptotic behaviour of the associated quantile functions as well. This is of particular importance for GH- and VG-distributions whose distribution functions cannot be given in closed form, and a reliable and rapidly convergent series expansion for these is not known either. To determine quantiles of the former, one therefore has to resort to numerical integration of their densities. This may—depending on the quality of the integration routine used—lead to more or less inaccurate and unstable results for p-quantiles if p is very close to 0 or 1. The quantile asymptotics are summarized in the following lemma which is a slightly modified version of [3, Lemma 3.1]. Due to its importance for the derivation of the tail dependence coefficients in Section 5, we also provide a short proof here.

Lemma 2. *Suppose $F : \mathbb{R} \to [0,1]$ is a continuous and strictly increasing distribution function.*

a) If $F(x) \sim c_1|x|^{-a_1}$ as $x \to -\infty$ and $1-F(x) \sim c_2 x^{-a_2}$ as $x \to \infty$ for some $a_1, a_2, c_1, c_2 > 0$, then $F^{-1}(u) \sim -\big(\frac{c_1}{u}\big)^{\frac{1}{a_1}}$ and $F^{-1}(1-u) \sim \big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}$ for $u \downarrow 0$.

b) If instead $F(x) \sim c_1|x|^{a_1}e^{-b_1|x|}$ as $x \to -\infty$ and $1-F(x) \sim c_2 x^{a_2} e^{-b_2 x}$ as $x \to \infty$ for some $a_1, a_2 \in \mathbb{R}$ and $b_1, b_2, c_1, c_2 > 0$, then $F^{-1}(u) \sim \frac{\log(u)}{b_1}$ and $F^{-1}(1-u) \sim -\frac{\log(u)}{b_2}$ for $u \downarrow 0$.

Proof. a) If $1-F(x) \sim c_2 x^{-a_2}$ as $x \to \infty$, then for any $r > 0$

$$\lim_{u\downarrow 0} \frac{1-F\Big(r\big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}\Big)}{u} = r^{-a_2}.$$

For $r < 1$, the right hand side of the above equation is greater than one, so we conclude that in this case $1-F\Big(r\big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}\Big) > u$ for sufficiently small u and hence $F^{-1}(1-u) > r\big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}$ (note that the assumptions on F imply $F^{-1}\big(F(y)\big) = y$ for all $y \in \mathbb{R}$). If $r > 1$, then we similarly obtain $1-F\Big(r\big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}\Big) < u$ and thus $F^{-1}(1-u) < r\big(\frac{c_2}{u}\big)^{\frac{1}{a_2}}$ for sufficiently small u. This proves the assertion for $F^{-1}(1-u)$, and the asymptotic behaviour of $F^{-1}(u)$ for $u \downarrow 0$ can be shown analogously.

b) If $1-F(x) \sim c_2 x^{a_2} e^{-b_2 x}$ as $x \to \infty$, then we have

$$\lim_{u\downarrow 0} \frac{1-F\big(-\frac{r\log(u)}{b_2}\big)}{u} = \lim_{u\downarrow 0} c_2 \left(-\frac{r\log(u)}{b_2}\right)^{a_2} u^{r-1} = \begin{cases} \infty, & r < 1, \\ 0, & r > 1. \end{cases}$$

With the same reasoning as before we conclude $F^{-1}(1-u) \sim -\frac{\log(u)}{b_2}$ for $u \downarrow 0$, and the corresponding result for $F^{-1}(u)$ is easily obtained along the same lines. $\square$

Observe that the tails of the densities $d_{GH(\lambda,\alpha,\pm\alpha,\delta,\mu)}(x)$ of the generalized hyperbolic skew Student t-distribution behave completely different for large arguments. If $\beta = \alpha$, then by (21) the asymptotic behaviour is as follows:

$$\begin{aligned} d_{GH(\lambda,\alpha,\alpha,\delta,\mu)}(x) &\sim \tilde{c}_1 \, |x|^{\lambda-1} \, e^{-2\alpha|x|}, \quad & x \to -\infty, \\ d_{GH(\lambda,\alpha,\alpha,\delta,\mu)}(x) &\sim \tilde{c}_2 \, |x|^{\lambda-1}, & x \to +\infty, \end{aligned} \tag{24}$$

and the other way round if $\beta = -\alpha$. Hence, they have one semi-heavy and one heavy (power) tail, so the asymptotic behaviour of their distribution functions and quantiles is obtained by combining the corresponding results above. Further, it is easily seen from equation (1) that the GIG densities possess a semi-heavy right tail with parameters $a_2 = \lambda - 1$, $b_2 = \frac{\gamma^2}{2}$, and $c_2 = \frac{\gamma^\lambda}{2\delta^\lambda K_\lambda(\delta\gamma)}$, so the above lemma and proposition can also be applied here.

But not only the tail behaviour of single GH distributions, also that of convolutions of the latter is of interest in finance. Think, for example, of factor models for credit portfolios where for each portfolio constituent a state variable $X_i = \sqrt{\rho}M + \sqrt{1-\rho}Z_i$, $0 \le \rho \le 1$, with a systematic factor M and an independent idiosyncratic factor Z_i is defined. The portfolio loss distribution derived from this approach then entails the quantile function $F_{X_i}^{-1}(p_d)$ of the distribution of X_i which has to be evaluated for typically very small default probabilities p_d. If the factor distributions F_M and F_{Z_i} are not stable under convolution, the distribution of X_i is usually unknown, therefore the quantiles $F_{X_i}^{-1}(p_d)$ can only be determined by either time-consuming simulations or advanced numerical methods. Precisely this is the case if one assumes the factors to be GH distributed. From Lemma 1 b) and equation (6), one can deduce the following convolution properties of the GH family:

$$\begin{aligned} &\text{a)} \quad NIG(\alpha,\beta,\delta_1,\mu_1) * NIG(\alpha,\beta,\delta_2,\mu_2) = NIG(\alpha,\beta,\delta_1+\delta_2,\mu_1+\mu_2), \\ &\text{b)} \quad NIG(\alpha,\beta,\delta_1,\mu_1) * GH\big(\tfrac{1}{2},\alpha,\beta,\delta_2,\mu_2\big) = GH\big(\tfrac{1}{2},\alpha,\beta,\delta_1+\delta_2,\mu_1+\mu_2\big), \\ &\text{c)} \quad GH(-\lambda,\alpha,\beta,\delta,\mu) * VG\big(\lambda,\alpha,\beta,\mu_2\big) = GH(\lambda,\alpha,\beta,\delta,\mu_1+\mu_2),\ \lambda > 0, \\ &\text{d)} \quad VG(\lambda_1,\alpha,\beta,\mu_1) * VG(\lambda_2,\alpha,\beta,\mu_2) = VG(\lambda_1+\lambda_2,\alpha,\beta,\mu_1+\mu_2),\ \lambda_1,\lambda_2 > 0. \end{aligned} \tag{25}$$

Inspecting the Laplace transforms of GIG distributions given in Proposition 1 more closely, one can deduce that the list of GIG convolution formulas (6) is complete, that is, no other convolution of two GIG distributions will yield a distribution that itself is contained in the GIG class. Consequently, there do not exist more than the four convolution formulas above for the GH family either. In particular, a convolution of two GH distributions with different parameters α and/or β cannot be GH distributed itself. This fact makes the application of generalized hyperbolic factor models computationally demanding, therefore some (approximate) formulas for $F_{X_i}^{-1}(p_d)$, at least for small probabilities p_d, which are faster and easier to evaluate would be desirable here. For a more thorough introduction to GH factor models, we refer to [11] and [23, Chapter 3]; there the quantiles of the convolution were calculated with help of Fourier inversion.

The behaviour of GH convolution tails is described in Proposition 3 below. The latter applies, in fact, to an even slightly more general framework where both factors belong to $\mathscr{L}_{a,b}$, the class of distributions with exponential tails with rates a and b, which we define as follows:

Definition 3. A distribution function F is said to have *exponential tails* with rates $a>0$ and $b>0$ $(F \in \mathscr{L}_{a,b})$ if for all $y \in \mathbb{R}$

$$\lim_{x\to-\infty} \frac{F(x-y)}{F(x)} = e^{-ay} \quad \text{and} \quad \lim_{x\to\infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = e^{by}.$$

Note that most definitions of exponential tails only use one index which characterizes the behaviour of the right tail $\bar{F}(x)$. This is due to the fact that these arose from extreme value theory or more generally actuarial sciences where one typically works with probability distributions on $\mathbb{R}_+$. The above is a natural generalization to distributions having support $\mathbb{R}$ we are concerned with.

The class $\mathscr{L}_{a,b}$ is closely related to the class $\mathscr{R}_p$ of regularly varying functions to be introduced in

Definition 4. A measurable function g is *regularly varying* with exponent $p \in \mathbb{R}$ $(g \in \mathscr{R}_p)$ if $\lim_{t\to\infty} \frac{g(st)}{g(t)} = s^p$ for all $s>0$.

We have $F \in \mathscr{L}_{a,b}$ iff $F(-\ln(x)) \in \mathscr{R}_{-a}$ and $\bar{F}(\ln(x)) \in \mathscr{R}_{-b}$. To see this, put $s=e^y$ and $t=e^{-x}$, then

$$e^{-ay} = \lim_{x\to-\infty} \frac{F(x-y)}{F(x)} \iff s^{-a} = \lim_{t\to\infty} \frac{F(-\ln(t)-\ln(s))}{F(-\ln(t))} = \lim_{t\to\infty} \frac{F(-\ln(st))}{F(-\ln(t))},$$

and the assertion for the right tails follows analogously with $s=e^{-y}$ and $t=e^x$.

Using Definition 2 and Proposition 2, it is immediately seen that for a probability distribution F possessing a density f with semi-heavy tails we have

$$\lim_{x\to-\infty} \frac{F(x-y)}{F(x)} = \lim_{x\to-\infty} \frac{f(x-y)}{f(x)} = \lim_{x\to-\infty} \left(\frac{|x-y|}{|x|}\right)^{a_1} e^{-b_1(|x-y|-|x|)} = e^{-b_1 y}$$

and an analogous limit is obtained for the right tails, hence $F \in \mathscr{L}_{b_1,b_2}$. In particular, we see that $GH(\lambda,\alpha,\beta,\delta,\mu)$- and $VG(\lambda,\alpha,\beta,\mu)$-distributions belong to the class $\mathscr{L}_{\alpha+\beta,\alpha-\beta}$. The asymptotic behaviour of the densities of the t-distributions, however, is easily seen from (18) to equal $d_{t(\lambda,\delta,\mu)}(x) \sim \bar{c}\,|x|^{2\lambda-1}$, $x\to\pm\infty$. Hence, $F_{t(\lambda,\delta,\mu)}(x) \sim \frac{\bar{c}}{2\lambda-1}|x|^{2\lambda}$, and thus $F_{t(\lambda,\delta,\mu)}(-x), \bar{F}_{t(\lambda,\delta,\mu)}(x) \in \mathscr{R}_{2\lambda}$. We defer the latter for a moment and first consider convolutions of factors with exponential tails. An easy solution occurs if the factors of the convolution have semi-heavy tails which decay at different rates: the convolution tails are determined by the factor with the heavier left (respectively right) tail.

Proposition 3. *Suppose that $F_1 \in \mathscr{L}_{b_1,b_2}$, $F_2 \in \mathscr{L}_{\tilde{b}_1,\tilde{b}_2}$ with moment generating functions $M_{F_1}(u)$ and $M_{F_2}(u)$. If $b_1 < \tilde{b}_1$ and $b_2 < \tilde{b}_2$, then $F_1 * F_2 \in \mathscr{L}_{b_1,b_2}$ and*

$$\lim_{x\to-\infty} \frac{(F_1 * F_2)(x)}{F_1(x)} = M_{F_2}(-b_1), \qquad \lim_{x\to\infty} \frac{\overline{(F_1 * F_2)}(x)}{\bar{F}_1(x)} = M_{F_2}(b_2).$$

A detailed proof of this result can be found in [23, Proposition 1.16]. The assumption above that both tails of F_1 are heavier than those of F_2 was just made for notational convenience. As it is easily seen, in general we have $F_1 * F_2 = \mathscr{L}_{b_1\wedge\tilde{b}_1, b_2\wedge\tilde{b}_2}$, that is, one factor may determine the left tail of the convolution and the other one the right tail. In [19, Theorem 3 b)] it has been shown that if the right tails of F_1 and F_2 are both exponential with the same rate a, then the right tail of $F_1 * F_2$ is also exponential with rate a, so we may conclude that $F_1 * F_2 = \mathscr{L}_{b_1\wedge\tilde{b}_1, b_2\wedge\tilde{b}_2}$ remains valid if $b_1 = \tilde{b}_1$ and/or $b_2 = \tilde{b}_2$. Summing up, we have the following

Corollary 1. *Let F_1, F_2 be the distribution functions of $GH(\lambda_1, \alpha_1, \beta_1, \delta_1, \mu_1)$ resp. $GH(\lambda_2, \alpha_2, \beta_2, \delta_2, \mu_2)$, and $F = F_1 * F_2$. If $\alpha_1 + \beta_1 \neq \alpha_2 + \beta_2$ and $\alpha_1 - \beta_1 \neq \alpha_2 - \beta_2$, then*

$$F(x) \sim M_{F^l_{max}}(-b_1) F^l_{max}(x),\ x \to -\infty, \ \ \textit{and} \ \ \bar{F}(x) \sim M_{F^r_{max}}(b_2) \bar{F}^r_{max}(x),\ x \to \infty,$$

where $b_1 = \min(\alpha_1 + \beta_1, \alpha_2 + \beta_2)$, $b_2 = \min(\alpha_1 - \beta_1, \alpha_2 - \beta_2)$, and $F^l_{max}(x), F^r_{max}(x)$ are the distribution functions of the GH distribution whose parameters α_i, β_i attain the value b_1 resp. b_2. The assertions remain valid if one or both factors are VG distributed instead.

If $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ or $\alpha_1 - \beta_1 = \alpha_2 - \beta_2$, the left resp. right tail behaviour cannot be precisely specified, and one only has the weaker result $F \in \mathscr{L}_{b_1,b_2}$.

Since the convolution tails are asymptotically equivalent to the tail of one factor distribution, multiplied by a constant, approximate quantile values of the convolution for probabilities close to zero or one can be computed with help of Lemma 2 b). It can be shown that the latter also applies under the weaker assumption $F \in \mathscr{L}_{b_1,b_2}$. Therefore, we obtain that the asymptotic behaviour of the quantile function of a convolution of GH distributions with $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ is given by $F^{-1}(u) \sim \frac{\log(u)}{\alpha_1+\beta_1}$, $u \downarrow 0$, and similarly, $F^{-1}(u) \sim -\frac{\log(u)}{\alpha_1-\beta_1}$ for $u \uparrow 1$ if $\alpha_1 - \beta_1 = \alpha_2 - \beta_2$.

A corresponding result for the regularly varying tails of the t-distributions can be obtained by applying [6, Theorem 1.1 and the Theorem on p. 54] which yields

Corollary 2. *Let F_1, F_2 be the distribution functions of $t(\lambda_1, \delta_1, \mu_1)$ and $t(\lambda_2, \delta_2, \mu_2)$ with corresponding densities f_1, f_2, then*

$$\lim_{|x|\to\infty} \frac{(f_1 * f_2)(x)}{f_1(x) + f_2(x)} = 1 \ \ \textit{and} \ \lim_{x\to-\infty} \frac{(F_1 * F_2)(x)}{F_1(x) + F_2(x)} = \lim_{x\to\infty} \frac{\overline{(F_1 * F_2)}(x)}{\bar{F}_1(x) + \bar{F}_2(x)} = 1.$$

If $\lambda_1 < \lambda_2$, then with the above notations we have $f_1(x) = o\big(f_2(x)\big)$ as $|x| \to \infty$ and $F_1(x) = o\big(F_2(x)\big), x \to -\infty$, as well as $\bar{F}_1(x) = o\big(\bar{F}_2(x)\big), x \to \infty$, consequently

$$\lim_{|x|\to\infty} \frac{(f_1 * f_2)(x)}{f_2(x)} = 1 \;\text{ and }\; \lim_{x\to-\infty} \frac{(F_1 * F_2)(x)}{F_2(x)} = \lim_{x\to\infty} \frac{\overline{(F_1 * F_2)}(x)}{\bar{F}_2(x)} = 1$$

(see also [6, Theorem 2.1]). Hence, also in this case the tail behaviour of the convolution and the asymptotic behaviour of the convolution density is determined by the factor with the heavier tails. Approximate quantile values can then, similarly as before, be calculated using Lemma 2 a).

4 Multivariate normal mean-variance mixtures and GH distributions

Let us first fix some notations which will be used throughout the rest of the paper: The vectors $u = (u_1,\dots,u_d)^\top$ and $x = (x_1,\dots,x_d)^\top$ are elements of $\mathbb{R}^d$, the superscript $^\top$ stands for the transpose of a vector or matrix. $\langle u,x\rangle = u^\top x = \sum_{i=1}^d u_i x_i$ denotes the scalar product of the vectors u,x and $\|u\| = (u_1^2+\cdots+u_d^2)^{1/2}$ the Euclidean norm of u. If A is a real-valued $d\times d$-square matrix, then $\det(A)$ denotes the determinant of A. The $d\times d$-identity matrix is labeled I_d. In contrast to u and x, the letters y, s and t are reserved for univariate real variables, that is, we assume $y,s,t \in \mathbb{R}$ or $\mathbb{R}_+$. To properly distinguish between the real number zero and the zero vector, we write $0 \in \mathbb{R}$ and $\mathbf{0} := (0,\dots,0)^\top \in \mathbb{R}^d$. Note that here and in the following $d \geq 2$ indicates the dimension, whereas n is usually used as running index for all kinds of sequences. In particular, the notation $N_d(\mu,\Delta)$ will be used for the d-dimensional normal distribution with mean vector μ and covariance matrix Δ.

With these preliminaries, we can formulate the multivariate version of Definition 1 as follows:

Definition 5. An $\mathbb{R}^d$-valued random variable X is said to have a *multivariate normal mean-variance mixture distribution* if

$$X \stackrel{d}{=} \mu + Z\beta + \sqrt{Z}AW,$$

where $\mu,\beta \in \mathbb{R}^d$, A is a real-valued $d\times d$-matrix such that $\Delta := AA^\top$ is positive definite, W is a standard normal distributed random vector ($W \sim N_d(\mathbf{0},I_d)$) and $Z \sim G$ is a real-valued, non-negative random variable independent of W.
Equivalently, a probability measure F on $(\mathbb{R}^d,\mathscr{B}^d)$ is said to be a multivariate normal mean-variance mixture if

$$F(\mathrm{d}x) = \int_{\mathbb{R}_+} N_d(\mu + y\beta, y\Delta)(\mathrm{d}x)\, G(\mathrm{d}y),$$

where the mixing distribution G is a probability measure on $(\mathbb{R}_+,\mathscr{B}_+)$. We shall use the short hand notation $F = N_d(\mu + y\beta, y\Delta)\circ G$.

Remark 1. Note that one can further assume w.l.o.g. $|\det(A)| = \det(\Delta) = 1$, since a (positive) multiplicative constant can always be included within the variable Z. More precisely, let $\bar{A} = |\det(A)|^{-1/d}A$, $\bar{\beta} = |\det(A)|^{-2/d}\beta$ and $\bar{Z} = |\det(A)|^{2/d}Z$, then $|\det(\bar{A})| = 1$ and $\mu + Z\beta + \sqrt{Z}AW = \mu + \bar{Z}\bar{\beta} + \sqrt{\bar{Z}}\bar{A}W$. Equivalently, if $\bar{\Delta} = \bar{A}\bar{A}^\top$ and $\bar{G} = \mathscr{L}(\bar{Z})$, then also $\det(\bar{\Delta}) = 1$ and $N_d(\mu + y\beta, y\Delta) \circ G = N_d(\mu + y\bar{\beta}, y\bar{\Delta}) \circ \bar{G}$. Further observe that the use of a single univariate mixing variable Z causes dependencies between all entries of X, as we shall see in Section 5.

The straightforward generalization of Lemma 1 in Section 2 is

Lemma 3. *Let $\mathbb{G}$ be a class of probability distributions on $(\mathbb{R}_+, \mathscr{B}_+)$ and suppose $G, G_1, G_2 \in \mathbb{G}$.*

a) If G possesses a moment generating function $M_G(y)$ on some open interval (a,b) with $a < 0 < b$, then $F = N_d(\mu + y\beta, y\Delta) \circ G$ also possesses a moment generating function $M_F(u) = e^{\langle u,\mu\rangle} M_G\big(\frac{\langle u,\Delta u\rangle}{2} + \langle u,\beta\rangle\big)$ that is defined for all $u \in \mathbb{R}^d$ with $a < \frac{\langle u,\Delta u\rangle}{2} + \langle u,\beta\rangle < b$.
b) If $G = G_1 * G_2 \in \mathbb{G}$, then $\big(N_d(\mu_1 + y\beta, y\Delta) \circ G_1\big) * \big(N_d(\mu_2 + y\beta, y\Delta) \circ G_2\big) = N_d(\mu_1 + \mu_2 + y\beta, y\Delta) \circ G$.
c) If $(\mu_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ are convergent sequences of real vectors with finite limits $\mu, \beta \in \mathbb{R}^d$ (that is, $\|\mu\|, \|\beta\| < \infty$), and $(G_n)_{n\geq 1}$ is a sequence of mixing distributions with $G_n \xrightarrow{w} G$, then $N_d(\mu_n + y\beta_n, y\Delta) \circ G_n \xrightarrow{w} N_d(\mu + y\beta, y\Delta) \circ G$.

For further reference, we also briefly highlight the relationship between multivariate normal mean-variance mixtures and elliptical distributions. From a financial point of view, the latter are of some interest because they have the nice property that within this class the Value-at-Risk (VaR) is a coherent risk measure in the sense of [2] (this has been shown in [20, Theorem 1], see also [29, Theorem 6.8]).

Definition 6. An $\mathbb{R}^d$-valued random vector X has an *elliptical distribution* if there exists a function $\psi : \mathbb{R}_+ \to \mathbb{R}$, a symmetric, positive semidefinite $d \times d$-matrix Σ and some $\mu \in \mathbb{R}^d$ such that the characteristic function $\phi_X(u) = E[e^{i\langle u,X\rangle}]$ of X admits the representation

$$\phi_X(u) = e^{i\langle u,\mu\rangle}\psi\big(\langle u, \Sigma u\rangle\big) \qquad \forall u \in \mathbb{R}^d.$$

The elliptical distribution $\mathscr{L}(X)$ then is denoted by $E_d\big(\mu, \Sigma, \psi(t)\big)$.

It can be shown that if an elliptical distribution has a density f, then it must necessarily be of the form

$$f(x) = \frac{1}{\sqrt{\det(\Sigma)}} h\big(\langle x-\mu, \Sigma^{-1}(x-\mu)\rangle\big)$$

for some measurable function $h : \mathbb{R} \to \mathbb{R}_+$. The level sets of such a density obviously are the ellipsoids $\{x \in \mathbb{R}^d \,|\, \langle x-\mu, \Sigma^{-1}(x-\mu)\rangle = \bar{c}\}$, $\bar{c} > 0$, which explains where the name of this class of distributions stems from. Combining [29, Theorem 3.22 and Definition 3.26], we get the following characterization of elliptically distributed random vectors.

Proposition 4. *$X \sim E_d\big(\mu,\Sigma,\psi(t)\big)$ if and only if*

$$X \stackrel{d}{=} \mu + RAS$$

where R is an $\mathbb{R}_+$-valued random variable, S is a random vector which is independent of R and uniformly distributed on the unit sphere $\mathscr{S} := \{\xi \in \mathbb{R}^d \,|\, \|\xi\| = 1\}$, and A is a $d\times d$-matrix fulfilling $AA^\top = \Sigma$.

The connection between elliptical distributions and multivariate normal mean-variance mixtures is given in

Corollary 3. *A normal mean-variance mixture $F = N_d(\mu + y\beta, y\Delta)\circ G$ is an elliptical distribution if and only if $\beta = \mathbf{0}$, that is, if and only if it is a normal variance mixture.*

Proof. The characteristic function of F can be shown to have the form $\phi_F(u) = e^{i\langle u,\mu\rangle}\mathfrak{L}_G\big(\frac{\langle u,\Delta u\rangle}{2} - i\langle u,\beta\rangle\big)$ which evidently has the representation $e^{i\langle u,\mu\rangle}\psi\big(\langle u,\Sigma u\rangle\big)$ required by Definition 6 with $\Sigma = \Delta$ and $\psi(t) = \mathfrak{L}_G\big(\frac{t}{2}\big)$ if and only if $\beta = \mathbf{0}$. □

Now we leave the general theory and turn our attention to the multivariate GH distributions. These have already been introduced as a natural generalization of the univariate case at the end of the seminal paper [4] and were investigated further in [7] and [8]. They are defined as normal mean-variance mixtures with GIG mixing distributions in the following way:

$$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta) := N_d(\mu + y\Delta\beta, y\Delta)\circ GIG\big(\lambda,\delta,\sqrt{\alpha^2 - \langle\beta,\Delta\beta\rangle}\big), \quad (26)$$

where it is usually assumed without loss of generality (see p. 14) that $\det(\Delta) = 1$ which we shall also do in the following if not stated otherwise. Due to the parameter restrictions of GIG distributions, the other GH parameters have to fulfil the constraints

$$\lambda \in \mathbb{R},\ \alpha,\delta \in \mathbb{R}_+,\ \beta,\mu \in \mathbb{R}^d \quad \text{and} \quad \begin{array}{ll} \delta \ge 0,\ 0 \le \sqrt{\langle\beta,\Delta\beta\rangle} < \alpha, & \text{if } \lambda > 0, \\ \delta > 0,\ 0 \le \sqrt{\langle\beta,\Delta\beta\rangle} < \alpha, & \text{if } \lambda = 0, \\ \delta > 0,\ 0 \le \sqrt{\langle\beta,\Delta\beta\rangle} \le \alpha, & \text{if } \lambda < 0. \end{array}$$

The meaning and influence of the parameters is essentially the same as in the univariate case (see p. 6). Again, parametrizations with $\delta = 0$, $\alpha = 0$ or $\sqrt{\langle\beta,\Delta\beta\rangle} = \alpha$ have to be understood as limiting cases.

Note that the above definition of multivariate GH distributions as normal mean-variance mixtures of the form $N_d(\mu + y\Delta\beta, y\Delta)\circ G$ is of course equivalent to the representation $N_d(\mu + y\tilde{\beta}, y\Delta)\circ G$ used in the previous section because the $d\times d$-matrix Δ is always regular by assumption. The modification of the mean term just simplifies some formulas as we shall see below. For notational consistency with Section 2, the term $GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$ will be reserved for multivariate GH distributions with $\beta,\mu \in \mathbb{R}^d$, whereas $GH(\lambda,\alpha,\beta,\delta,\mu)$ denotes a univariate GH distribution with $\beta,\mu \in \mathbb{R}$ as before.

If $\delta > 0$ and $\sqrt{\langle\beta,\Delta\beta\rangle} < \alpha$, the density of $GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$ is given by

$$\begin{aligned} d_{GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)}(x) &= \int_0^\infty d_{N_d(\mu+y\Delta\beta,y\Delta)}(x)\, d_{GIG\left(\lambda,\delta,\sqrt{\alpha^2-\langle\beta,\Delta\beta\rangle}\right)}(y)\,\mathrm{d}y \\ &= \frac{(\alpha^2-\langle\beta,\Delta\beta\rangle)^{\frac{\lambda}{2}}}{(2\pi)^{\frac{d}{2}}\alpha^{\lambda-\frac{d}{2}}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2-\langle\beta,\Delta\beta\rangle})}\left(\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2\right)^{(\lambda-\frac{d}{2})/2} \\ &\qquad \times K_{\lambda-\frac{d}{2}}\left(\alpha\sqrt{\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2}\right)e^{\langle\beta,x-\mu\rangle} \end{aligned} \tag{27}$$

Remark 2. If the $d\times d$-matrix Δ is replaced by a matrix $\bar{\Delta}$ of the same dimensions with $\det(\bar{\Delta}) \neq 1$, then the normal density $d_{N_d(\mu+y\bar{\Delta}\beta,y\bar{\Delta})}(x)$ has an additional factor $\det(\bar{\Delta})^{-1/2}$ which will be incorporated in the norming constant of $d_{GH_d(\lambda,\alpha,\beta,\delta,\mu,\bar{\Delta})}(x)$. Suppose $\bar{\Delta} = c^{1/d}\Delta$ for some $c>0$, then $\det(\bar{\Delta}) = c$, and if we also replace $\lambda,\alpha,\beta,\delta,\mu$ by the barred parameters

$$\bar{\lambda} := \lambda, \quad \bar{\alpha} := c^{\frac{1}{2d}}\alpha, \quad \bar{\beta} := \beta, \quad \bar{\delta} := c^{-\frac{1}{2d}}\delta, \quad \bar{\mu} := \mu,$$

then it is easily seen from (27) that the densities of $GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$ and $GH_d(\bar{\lambda},\bar{\alpha},\bar{\beta},\bar{\delta},\bar{\mu},\bar{\Delta})$ and thus both distributions coincide. Note that these considerations also remain true for all subsequently defined limit distributions. This again shows that the assumption $\det(\Delta) = 1$ is not an essential restriction. The barred parameters will be used later at some points in Section 5 to indicate that $\det(\bar{\Delta}) = 1$ is not assumed there.

If multivariate GH distributions would have been defined as a mixture of the form $N_d(\mu+y\beta,y\Delta)\circ GIG\left(\lambda,\delta,\sqrt{\alpha^2-\langle\beta,\Delta\beta\rangle}\right)$ (see the remark on the previous page), then the last factor of the density (27) would be $e^{\langle\Delta^{-1}\beta,x-\mu\rangle}$ instead of $e^{\langle\beta,x-\mu\rangle}$, and $\bar{\beta}$ would have to be defined by $\bar{\beta} = c^{1/d}\beta$.

With the special choice $\lambda = -\frac{1}{2}$, one obtains the multivariate normal inverse Gaussian distribution $NIG_d(\alpha,\beta,\delta,\mu,\Delta)$ possessing the density

$$\begin{aligned} d_{NIG_d(\alpha,\beta,\delta,\mu,\Delta)}(x) &= \sqrt{\frac{2}{\pi}}\,\frac{\delta\alpha^{\frac{d+1}{2}}e^{\delta\sqrt{\alpha^2-\langle\beta,\Delta\beta\rangle}}}{(2\pi)^{\frac{d}{2}}}\left(\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2\right)^{-\frac{d+1}{4}} \\ &\qquad \times K_{\frac{d+1}{2}}\left(\alpha\sqrt{\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2}\right)e^{\langle\beta,x-\mu\rangle}. \end{aligned} \tag{28}$$

Let us briefly mention possible weak limits of multivariate GH distributions here. If $\lambda > 0$ and $\delta \to 0$, then by equations (26), (3), and Lemma 3 c) we have convergence to a multivariate Variance-Gamma distribution

$$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)\xrightarrow{w}N_d(\mu+y\Delta\beta,y\Delta)\circ G\left(\lambda,\tfrac{\alpha^2-\langle\beta,\Delta\beta\rangle}{2}\right)=VG_d(\lambda,\alpha,\beta,\mu,\Delta)$$

which has the density

$$d_{VG_d(\lambda,\alpha,\beta,\mu,\Delta)}(x) = \frac{(\alpha^2 - \langle\beta,\Delta\beta\rangle)^\lambda}{(2\pi)^{\frac{d}{2}}\alpha^{\lambda-\frac{d}{2}}2^{\lambda-1}\Gamma(\lambda)}\left(\langle x-\mu,\Delta^{-1}(x-\mu)\rangle\right)^{(\lambda-\frac{d}{2})/2} \times K_{\lambda-\frac{d}{2}}\left(\alpha\sqrt{\langle x-\mu,\Delta^{-1}(x-\mu)\rangle}\right)e^{\langle\beta,x-\mu\rangle}. \tag{29}$$

For $\lambda < 0$ and $\alpha \to 0$ as well as $\beta \to \mathbf{0}$, we arrive at the multivariate scaled and shifted t-distribution with $f = -2\lambda$ degrees of freedom:

$$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta) \xrightarrow{w} N_d(\mu, y\Delta)\circ iG\left(\lambda,\tfrac{\delta^2}{2}\right) = t_d(\lambda,\delta,\mu,\Delta).$$

It has the density

$$d_{t_d(\lambda,\delta,\mu,\Delta)}(x) = \frac{\Gamma\left(-\lambda+\frac{d}{2}\right)}{(\delta^2\pi)^{\frac{d}{2}}\Gamma(-\lambda)}\left(1+\frac{\langle x-\mu,\Delta^{-1}(x-\mu)\rangle}{\delta^2}\right)^{\lambda-\frac{d}{2}}. \tag{30}$$

If $\lambda < 0$, but $\langle\beta,\Delta\beta\rangle \to \alpha^2$, then we have weak convergence to the normal mean-variance mixture

$$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta) \xrightarrow{w} N_d(\mu + y\Delta\beta, y\Delta)\circ iG\left(\lambda,\tfrac{\delta^2}{2}\right)$$

possessing the density

$$d_{GH_d(\lambda,\sqrt{\langle\beta,\Delta\beta\rangle},\beta,\delta,\mu,\Delta)}(x) = \frac{2^{\lambda+1-\frac{d}{2}}\,\delta^{-2\lambda}}{\pi^{\frac{d}{2}}\Gamma(-\lambda)\alpha^{\lambda-\frac{d}{2}}}\left(\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2\right)^{(\lambda-\frac{d}{2})/2} \times K_{\lambda-\frac{d}{2}}\left(\alpha\sqrt{\langle x-\mu,\Delta^{-1}(x-\mu)\rangle+\delta^2}\right)e^{\langle\beta,x-\mu\rangle}, \tag{31}$$

where $\alpha = \sqrt{\langle\beta,\Delta\beta\rangle}$.

The most important properties of multivariate GH distributions are summarized in the following theorem which goes back to [7, Theorem 1], see also [8, p. 49f]. It shows that this distribution class is closed under forming marginals, conditioning and affine transformations.

Theorem 1. *Suppose $X \sim GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$. Let $(X_1,X_2)^\top$ be a partition of X where X_1 has the dimension r and X_2 the dimension $k = d-r$, and let $(\beta_1,\beta_2)^\top$ and $(\mu_1,\mu_2)^\top$ be similar partitions of β and μ. Furthermore, let*

$$\Delta = \begin{pmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{pmatrix}$$

be a partition of Δ such that Δ_{11} is an $r\times r$-matrix. Then the following holds:

a) $X_1 \sim GH_r(\lambda^*,\alpha^*,\beta^*,\delta^*,\mu^*,\Delta^*)$ with starred parameters given by $\lambda^* = \lambda$, $\alpha^* = \det(\Delta_{11})^{-\frac{1}{2r}}\sqrt{\alpha^2 - \langle\beta_2,(\Delta_{22}-\Delta_{21}\Delta_{11}^{-1}\Delta_{12})\beta_2\rangle}$, $\beta^* = \beta_1 + \Delta_{11}^{-1}\Delta_{12}\beta_2$, $\delta^* = \det(\Delta_{11})^{\frac{1}{2r}}\delta$, $\mu^* = \mu_1$, and $\Delta^* = \det(\Delta_{11})^{-\frac{1}{r}}\Delta_{11}$.

b) The conditional distribution of X_2 given $X_1 = x_1$ is $GH_k(\tilde{\lambda}, \tilde{\alpha}, \tilde{\beta}, \tilde{\delta}, \tilde{\mu}, \tilde{\Delta})$ with tilded parameters $\tilde{\lambda} = \lambda - \frac{r}{2}$, $\tilde{\alpha} = \det(\Delta_{11})^{\frac{1}{2k}} \alpha$, $\tilde{\beta} = \beta_2$, $\tilde{\delta} = \det(\Delta_{11})^{-\frac{1}{2k}} \times \sqrt{\delta^2 + \langle x_1 - \mu_1, \Delta_{11}^{-1}(x_1 - \mu_1)\rangle}$, $\tilde{\mu} = \mu_2 + \Delta_{21}\Delta_{11}^{-1}(x_1 - \mu_1)$, and $\tilde{\Delta} = \det(\Delta_{11})^{\frac{1}{k}} \times (\Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12})$.

c) Suppose $Y = BX + b$ where B is a regular $d \times d$-matrix and $b \in \mathbb{R}^d$, then $Y \sim GH_d(\hat{\lambda}, \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu}, \hat{\Delta})$ where $\hat{\lambda} = \lambda$, $\hat{\alpha} = |\det(B)|^{-\frac{1}{d}} \alpha$, $\hat{\beta} = (B^{-1})^\top \beta$, $\hat{\delta} = |\det(B)|^{\frac{1}{d}} \delta$, $\hat{\mu} = B\mu + b$, and $\hat{\Delta} = |\det(B)|^{-\frac{2}{d}} B\Delta B^\top$.

Remark 3. An important fact we want to stress here is that the above theorem remains also valid for all multivariate GH limit distributions considered before. Thus, one can in particular conclude from part b) that the limiting subclass of VG distributions itself is, in contrast to the t limit distributions, not closed under conditioning. This holds because the parameter $\tilde{\delta}$ of the conditional distribution in general is greater than zero, and the parameter $\tilde{\lambda} = \lambda - \frac{r}{2}$ may become negative if the subdimension r is sufficiently large.

Moreover, all margins of $t_d(\lambda, \delta, \mu, \Delta)$ are again t distributed $t_r(\lambda, \delta^*, \mu^*, \Delta^*)$ because if the joint distribution has the parameters $\alpha = 0$ and $\beta = \mathbf{0}$, part a) of the theorem implies that $\alpha^* = 0$ and $\beta^* = \mathbf{0}$ for every marginal distribution. Similarly, all margins of $VG_d(\lambda, \alpha, \beta, \mu, \Delta)$ are again VG distributions because if $\delta = 0$, then also $\delta^* = 0$. In addition it can be shown that all margins of $GH_d\big(\lambda, \sqrt{\langle \beta, \Delta\beta\rangle}, \beta, \delta, \mu, \Delta\big)$-distributions are of the same limiting type as their joint distribution, too.

Let us finally take a closer look at the moments of multivariate GH distributions. By Definition 5, every random variable X possessing a multivariate normal mean-variance mixture distribution admits the stochastic representation $X \overset{d}{=} \mu + Z\beta + \sqrt{Z}AW$ with independent random variables Z and $W \sim N_d(\mathbf{0}, I_d)$. The standardization of W and its independence from Z imply that

$$\begin{aligned} E(X) &= \mu + E(Z)\beta \\ \operatorname{Cov}(X) &= E\big[(X - E(X))(X - E(X))^\top\big] = E(Z)\Delta + \operatorname{Var}(Z)\beta\beta^\top \end{aligned} \tag{32}$$

with $\Delta = AA^\top$, provided that $E(|Z|), \operatorname{Var}(Z) < \infty$. If $X \sim GH_d(\lambda, \alpha, \beta, \delta, \mu, \Delta)$, then by (26) $X \overset{d}{=} \mu + Z\Delta\beta + \sqrt{Z}AW$ and $Z \sim GIG\big(\lambda, \delta, \sqrt{\alpha^2 - \langle \beta, \Delta\beta\rangle}\big)$. Using Proposition 1 and equation (7), one obtains explicit expressions for $E(Z)$ and $\operatorname{Var}(Z)$ which can be inserted into the general equations above to finally obtain

$$\begin{aligned} E[GH_d(\lambda, \alpha, \beta, \delta, \mu, \Delta)] &= \mu + \frac{\delta^2}{\zeta} \frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} \beta, \\ \operatorname{Cov}[GH_d(\lambda, \alpha, \beta, \delta, \mu, \Delta)] &= \frac{\delta^2}{\zeta} \frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} \Delta + \frac{\delta^4}{\zeta^2} \left(\frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \frac{K_{\lambda+1}^2(\zeta)}{K_\lambda^2(\zeta)} \right) \beta\beta^\top \end{aligned} \tag{33}$$

with $\zeta = \delta\sqrt{\alpha^2 - \langle\beta,\Delta\beta\rangle}$. In case of the Variance-Gamma limits we have

$$\begin{aligned} E[VG_d(\lambda,\alpha,\beta,\mu,\Delta)] &= \mu + \frac{2\lambda}{\alpha^2 - \langle\beta,\Delta\beta\rangle}\,\beta, \\ \mathrm{Cov}[VG_d(\lambda,\alpha,\beta,\mu,\Delta)] &= \frac{2\lambda}{\alpha^2 - \langle\beta,\Delta\beta\rangle}\,\Delta + \frac{4\lambda}{(\alpha^2 - \langle\beta,\Delta\beta\rangle)^2}\,\beta\beta^\top. \end{aligned} \tag{34}$$

Observe that by Lemma 3 both multivariate GH and VG distributions possess moment generating functions and hence finite moments of arbitrary order because the mixing GIG and Gamma distributions do have this property. This is no longer true for the limit distributions with $\lambda < 0$ because the corresponding inverse Gamma mixing distributions only have finite moments up to order $r < -\lambda$. By Theorem 1 a), the marginal distributions of $t_d(\lambda,\delta,\mu,\Delta)$ are given by $t(\lambda,\sqrt{\Delta_{ii}}\delta,\mu_i)$, $1 \le i \le d$ (recall that $\alpha = 0$ and $\beta = \mathbf{0}$ in this case), and from their tail behaviour (see p. 9) one can easily conclude that mean vector and covariance matrix of the t limit distributions are well defined and finite only if $\lambda < -\frac{1}{2}$ resp. $\lambda < -1$. If these constraints are fulfilled, then

$$E[t_d(\lambda,\delta,\mu,\Delta)] = \mu \qquad \text{and} \qquad \mathrm{Cov}[t_d(\lambda,\delta,\mu,\Delta)] = \frac{\delta^2}{-2\lambda-2}\,\Delta. \tag{35}$$

In the other limiting case where $\langle\beta,\Delta\beta\rangle = \alpha^2 > 0$, equations (32) state that necessary and sufficient conditions for the existence of a mean vector and covariance matrix of the limit distributions are that the inverse Gamma mixing distributions have finite means and variances which holds true if and only if $\lambda < -1$ and $\lambda < -2$, respectively. If λ is appropriately small, then

$$\begin{aligned} E\big[GH_d\big(\lambda,\sqrt{\langle\beta,\Delta\beta\rangle},\beta,\delta,\mu,\Delta\big)\big] &= \mu + \frac{\delta^2}{-2\lambda-2}\,\beta, \\ \mathrm{Cov}\big[GH_d\big(\lambda,\sqrt{\langle\beta,\Delta\beta\rangle},\beta,\delta,\mu,\Delta\big)\big] &= \frac{\delta^2}{-2\lambda-2}\,\Delta + \frac{\delta^4}{4(\lambda+1)^2(-\lambda-2)}\,\beta\beta^\top. \end{aligned} \tag{36}$$

5 On the dependence structure of multivariate GH distributions

Correlation is probably the most established dependence measure due to its simplicity and its predominant role within the normal world where it characterizes dependencies almost completely. This follows from the fact that the components W_i, $1 \le i \le d$, of a standard normal distributed random vector $W \sim N_d(\mathbf{0},I_d)$ are independent from each other (the joint density is just the product of the marginal ones in this case) and the stochastic representation

$$X \sim N_d(\mu,\Delta) \quad \Longleftrightarrow \quad X \stackrel{d}{=} \mu + AW \ \text{ where } W \sim N_d(\mathbf{0},I_d) \text{ and } AA^\top = \Delta.$$

Since X in distribution is nothing but a linear transform of a random vector W with independent (normal distributed) entries, the components of X can, roughly speaking, exhibit at most linear dependencies, and exactly these are specified and quantified by the pairwise correlations. However, things completely change if we depart from normality and consider normal variance mixtures instead. Suppose

$$X \sim N_d(\mu, y\Delta) \circ G, \ \text{ that is, } \ X \stackrel{d}{=} \mu + \sqrt{Z}AW$$

where $\mathscr{L}(Z) = G$, $W \sim N_d(\mathbf{0}, I_d)$ and $AA^\top = \Delta$ according to Definition 5. As we already remarked on p. 15, the mixing variable Z causes dependencies between the components of X, but these are typically not captured by correlation as the following lemma shows. It is a slightly more general version of [29, Lemma 3.5] which we adopt here since—in our opinion—the result is as simple as illustrative.

Lemma 4. *Suppose that $X \stackrel{d}{=} \mu + \sqrt{Z}AW$ has a normal variance mixture distribution where $E(Z) < \infty$ and $\Delta = AA^\top$ is a $d \times d$-diagonal matrix such that* $\mathrm{Cov}(X_i, X_j) = 0$, $1 \le i, j \le d$, $i \ne j$, *by* (32). *Then the X_i, $1 \le i \le d$, are independent if and only if Z is almost surely constant, that is, if and only if X is multivariate normal distributed.*

Proof. Because Δ is diagonal (and positive definite by Definition 5), we can assume without loss of generality that also the matrix A is diagonal and $A_{ii} = \sqrt{\Delta_{ii}}$, $1 \le i \le d$. The independence of Z and W and Jensen's inequality then imply

$$\begin{aligned} E\left(\prod_{i=1}^d |X_i - \mu_i|\right) &= E\left((\sqrt{Z})^d \prod_{i=1}^d |\sqrt{\Delta_{ii}} W_i|\right) = E\big((\sqrt{Z})^d\big) \prod_{i=1}^d E\big(|\sqrt{\Delta_{ii}} W_i|\big) \\ &\ge E\big(\sqrt{Z}\big)^d \prod_{i=1}^d E\big(|\sqrt{\Delta_{ii}} W_i|\big) = \prod_{i=1}^d E\big(|X_i - \mu_i|\big). \end{aligned}$$

Since the function $f(x) = x^d$ is strictly convex on $\mathbb{R}_+$ for $d \ge 2$, equality throughout holds if and only if Z is constant almost surely. □

Remark 4. The above result can even be extended: If $X \stackrel{d}{=} \mu + Z\beta + \sqrt{Z}AW$ has a normal mean-variance mixture distribution with $0 < \mathrm{Var}(Z) < \infty$, $\Delta = AA^\top$ is a $d \times d$-diagonal matrix and $\mathrm{Cov}(X_i, X_j) = 0$ for some $1 \le i \ne j \le d$, then X_i and X_j are not independent either. This can be seen as follows: Since Δ is diagonal and $\mathrm{Var}(Z) > 0$, by (32) $\mathrm{Cov}(X_i, X_j) = 0$ implies that $(\beta\beta^\top)_{ij} = 0$. This means, either $\beta_i = 0$ or $\beta_j = 0$ (or both, but then we would be within the setting of Lemma 4 again). Suppose $\beta_i \ne 0$ and $\beta_j = 0$, then we calculate using similar arguments as above

$$\begin{aligned} E\big((X_i - \mu_i)\,|X_j - \mu_j|\big) &= E\big((\beta_i Z + \sqrt{Z}\sqrt{\Delta_{ii}} W_i)\,|\sqrt{Z}\sqrt{\Delta_{jj}} W_j|\big) \\ &= E\big((\beta_i Z^{\frac{3}{2}} + Z\sqrt{\Delta_{ii}} W_i)\big) E\big(|\sqrt{\Delta_{jj}} W_j|\big) = \beta_i E\big(Z^{\frac{3}{2}}\big)\, E\big(|\sqrt{\Delta_{jj}} W_j|\big) \\ &> \beta_i E(Z)^{\frac{3}{2}}\, E\big(|\sqrt{\Delta_{jj}} W_j|\big) = E(\beta_i Z)\, E(Z)^{\frac{1}{2}}\, E\big(|\sqrt{\Delta_{jj}} W_j|\big) \\ &> E(\beta_i Z)\, E\big(Z^{\frac{1}{2}}\big) E\big(|\sqrt{\Delta_{jj}} W_j|\big) = E(X_i - \mu_i)\, E\big(|X_j - \mu_j|\big), \end{aligned}$$

and the inequalities are strict because $f(x) = x^{\frac{3}{2}}$ and $g(x) = \sqrt{x}$ are strictly convex resp. concave and $\mathscr{L}(Z)$ is non-degenerate by assumption.

Thus, in general zero correlation within multivariate normal mean-variance mixture models must not be interpreted as independence. In particular, the components X_i of a generalized hyperbolic distributed random vector $X \sim GH_d(\lambda, \alpha, \beta, \delta, \mu, \Delta)$ can never be independent because Theorem 1 b) states that the conditional distribution $\mathscr{L}\big(X_i \,|\, (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)^\top = \bar{x}\big) = GH(\tilde{\lambda}, \tilde{\alpha}, \tilde{\beta}, \tilde{\delta}, \tilde{\mu})$ always depends on the vector $\bar{x}$ (at least the parameter $\tilde{\delta}$ does so) for every $1 \le i \le d$. Moreover, it should be observed that for generalized hyperbolic distributed random variables the maximal attainable absolute correlation is usually strictly smaller than one: the Cauchy–Schwarz inequality states that $|\mathrm{Corr}(X_1, X_2)| = 1$ can occur if and only if $X_2 = aX_1 + b$ almost surely for some $a, b \in \mathbb{R}$ and $a \neq 0$, but if $X_1 \sim GH(\lambda_1, \alpha_1, \beta_1, \delta_1, \mu_1)$ and $X_2 \sim GH(\lambda_2, \alpha_2, \beta_2, \delta_2, \mu_2)$, the required linear relationship imposes some conditions on the GH parameters. Recall that $aX_1 + b \sim GH(\lambda, \frac{\alpha}{|a|}, \frac{\beta}{a}, \delta|a|, a\mu + b)$ by Theorem 1 c). Thus, using the scale- and location-invariant parameters $\zeta_i = \delta_i(\alpha_i^2 - \beta_i^2)^{\frac{1}{2}}$ and $\rho_i = \frac{\beta_i}{\alpha_i}$, $i = 1, 2$, we conclude that $X_2 = aX_1 + b$ can hold only if $\zeta_1 = \zeta_2$, $|\rho_1| = |\rho_2|$ and $\lambda_1 = \lambda_2$.

Having seen that correlation is in general not the tool to precisely describe and measure dependencies in multivariate models, one may ask if there exists a more powerful notion for this purpose. The answer is provided by

Definition 7. A *d-dimensional copula C* is a distribution function on $[0,1]^d$ with standard uniform marginal distributions, that is, $C : [0,1]^d \to [0,1]$ has the following properties:

a) $C(u) = C(u_1, \ldots, u_d)$ is increasing in each argument u_i,
b) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $1 \le i \le d$ and $u_i \in [0,1]$,
c) For all $(a_1, \ldots, a_d)^\top, (b_1, \ldots, b_d)^\top \in [0,1]^d$ with $a_i \le b_i$, $1 \le i \le d$, we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1 + \cdots + i_d} C(u_{1i_1}, \ldots, u_{di_d}) \ge 0,$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $1 \le j \le d$.

Properties a) and b) immediately follow from the definition of $C(u)$ as a distribution function with identically uniformly distributed marginals on $[0,1]$, c) essentially is a reformulation of the fact that if $U = (U_1, \ldots, U_d)^\top$ is a random vector possessing the distribution function $C(u)$, then necessarily $P(a_1 \le U_1 \le b_1, \ldots, a_d \le U_d \le b_d) \ge 0$. It can also be shown that these properties are sufficient, that is, every function $C : [0,1]^d \to [0,1]$ fulfilling a), b) and c) is a copula. Clearly, the k-dimensional margins of a copula C are also copulas for every $2 \le k < d$.

The central role of copulas in the study of multivariate distributions is highlighted by the following fundamental result which goes back to [35]. It not only shows that copulas are inherent in every multivariate distribution, but also that the latter can be constructed by plugging the desired marginal distributions into a suitably chosen copula. A short and elegant proof of Sklar's Theorem which is based on the distributional transform can be found in [31].

Theorem 2 (Sklar's Theorem). *Let F be a d-dimensional distribution function with margins $F_1,\dots,F_d$. Then there exists a copula $C:[0,1]^d \to [0,1]$ such that for all $x=(x_1,\dots,x_d)^\top \in [-\infty,\infty]^d$*

$$F(x_1,\dots,x_d) = C\big(F_1(x_1),\dots,F_d(x_d)\big). \tag{37}$$

If $F_1,\dots,F_d$ are all continuous, then C is unique, otherwise C is uniquely determined on $F_1(\mathbb{R})\times\cdots\times F_d(\mathbb{R})$ where $F_i(\mathbb{R})$ denotes the range of F_i.
Conversely, if $C:[0,1]^d \to [0,1]$ is a copula and $F_1,\dots,F_d$ are univariate distribution functions, then the function $F(x)$ defined by (37) *is a multivariate distribution function with margins $F_1,\dots,F_d$.*

If all marginal distribution functions F_i of F are continuous and their generalized inverses F_i^{-1} are defined by $F_i^{-1}(u_i) := \inf\{y \,|\, F_i(y) \geq u_i\}$ (with the usual convention $\inf\emptyset = \infty$), then $F_i\big(F_i^{-1}(u_i)\big) = u_i$. Thus it immediately follows from (37) by inserting $x_i = F_i^{-1}(u_i)$, $u_i \in [0,1]$, $1 \leq i \leq d$, that in this case the unique copula $C_F(u)$ contained in F is given by

$$C_F(u_1,\dots,u_d) = F\big(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d)\big). \tag{38}$$

The computation of this so-called implied copula $C_F(u)$ is in general numerically demanding if the distribution function $F(x)$ is not known explicitly. Suppose for example that only the density $f(x)$ of F can be expressed in closed form, then already the determination of a single value $F(x_0)$ requires to evaluate a d-dimensional integral which especially for greater dimensions d can hardly be done sufficiently precise in reasonable time. But for multivariate normal mean-variance mixtures it is sometimes possible to significantly reduce the numerical complexity: Suppose that $F = N_d(\mu + y\beta, y\Delta)\circ G$ with known margins F_i possessing Lebesgue densities f_i as above, and let O be an orthogonal $d\times d$-matrix such that $O\Delta O^\top$ is diagonal, then

$$\begin{aligned} C_F(u_1,\dots,u_d) &= F\big(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d)\big) \\ &= \int_{-\infty}^{F_d^{-1}(u_d)} \dots \int_{-\infty}^{F_1^{-1}(u_1)} \int_0^\infty d_{N_d(\mu+y\beta,y\Delta)}(x_1,\dots,x_d)\, G(\mathrm{d}y)\,\mathrm{d}x_1 \dots \mathrm{d}x_d \\ &= \int_0^\infty \prod_{i=1}^d \Phi_{((O(\mu+y\beta))_i,(O\Delta O^\top)_{ii})}\Big(\big(O(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d))^\top\big)_i\Big)\, G(\mathrm{d}y), \end{aligned} \tag{39}$$

where $\Phi_{(\mu,\sigma^2)}$ denotes the (univariate) distribution function of $N(\mu,\sigma^2)$. The last expression can be evaluated much easier on a computer since it only requires the calculation of one-dimensional integrals (possibly more than one because the values $F_i^{-1}(u_i)$ of the marginal quantile functions may only be obtained by integrating the corresponding densities $f_i(x_i)$ numerically).

If in addition to the marginal distributions F_i also F itself possesses a Lebesgue density $f(x)$, a further simplification can be achieved by using the (implied) copula

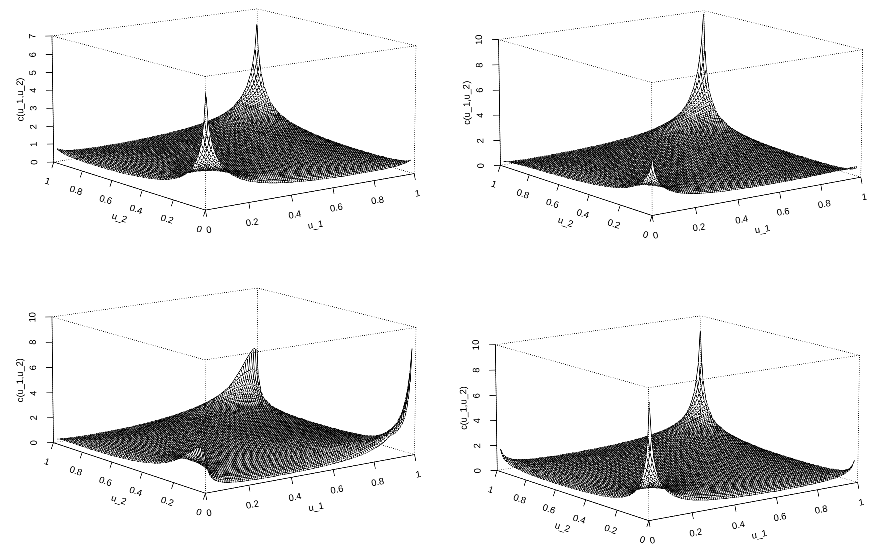

Fig. 1 Densities of implied copulas of bivariate GH distributions and their limits. The underlying distributions are as follows: *top left*: symmetric $NIG_2(10, \mathbf{0}, 0.2, \mathbf{0}, \bar{\Delta})$, *top right*: skewed $NIG_2(10, \binom{4}{1}, 0.2, \mathbf{0}, \bar{\Delta})$, *bottom left*: skewed $NIG_2(4, \binom{3}{-2}, 0.2, \mathbf{0}, \bar{\Delta})$, *bottom right*: $t_2(-2, 2, \mathbf{0}, \bar{\Delta})$. For all distributions $\bar{\Delta} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho = 0.3$.

density $c_F(u)$ which is defined by

$$c_F(u_1, \ldots, u_d) := \frac{\partial C_F(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} = \frac{f\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big)}{f_1\big(F_1^{-1}(u_1)\big) \cdots f_d\big(F_d^{-1}(u_d)\big)}, \tag{40}$$

where the last equation immediately follows from (38). Combining (40) and Theorem 1 a) allows to calculate the copula densities $c_{GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)}(u)$ of all multivariate GH distributions including the aforementioned limits. Some results for the bivariate case are visualized in Figure 1 above. Note that the choice of $\rho = 0.3$ implies $\det(\bar{\Delta}) = 1 - \rho^2 < 1$, so the parameters of the t- and NIG distributions are the barred ones $(\bar{\lambda}, \bar{\alpha}, \bar{\beta}, \bar{\delta}, \bar{\mu})$ defined in the remark on page 17. If $\bar{\beta} = \beta = \mathbf{0}$, then by equations (33)–(35) $\bar{\Delta}$ equals the correlation matrix of the related distribution.

Apart from being inherent in every multivariate distribution, the importance of copulas relies on the fact that they encode the dependencies between the margins F_i of F. Many popular dependence measures like, for example, Kendall's tau, Spearman's rho, or the Gini coefficient can be expressed and calculated solely in terms of the associated copulas (see [29, Proposition 5.29], and [30, Corollary 5.1.13]). Thus the assertion of Sklar's Theorem might alternatively be stated in the following way: Every multivariate distribution can be split up into two parts, the marginal distributions and the dependence structure. The next proposition shows that copulas and hence all dependence measures that can be derived from them are invariant under strictly increasing transformations of the margins. A proof can be found in, e.g., [29, Proposition 5.6].

Proposition 5. *Suppose that* $(X_1,\dots,X_d)^\top$ *is a random vector with joint distribution function* F*, continuous margins* F_i*,* $1\le i\le d$*, and implied copula* C_F *given by* (38)*. Let* $T_1,\dots,T_d$ *be strictly increasing functions and* G *be the joint distribution function of* $\big(T_1(X_1),\dots,T_d(X_d)\big)^\top$*. Then the implied copulas of* F *and* G *coincide, that is,* $C_F = C_G$*.*

From the above proposition it especially follows that the correlation of two random variables does not depend on the inherent copula of their joint distribution alone because correlation is invariant under (strictly) increasing linear transformations only, but not under arbitrary increasing mappings. Correlation is also linked to the marginal distributions since it requires them to possess finite second moments to be well defined, whereas by Sklar's Theorem a copula of the joint distribution always exists without imposing any conditions on the margins.

We now turn to the dependence measure we shall be concerned with for the rest of the present section, the coefficients of tail dependence, which are formally defined by

Definition 8. Let F be the joint distribution function of the bivariate random vector $(X_1,X_2)^\top$ and F_1,F_2 be the marginal distribution functions of X_1 and X_2, then the *coefficient of upper tail dependence* of F resp. X_1 and X_2 is

$$\lambda_u := \lambda_u(F) = \lambda_u(X_1,X_2) = \lim_{q\uparrow 1} P\big(X_2 > F_2^{-1}(q)\,|\,X_1 > F_1^{-1}(q)\big),$$

provided a limit $\lambda_u \in [0,1]$ exists. If $0<\lambda_u\le 1$, then F resp. X_1 and X_2 are said to be *upper tail dependent*; if $\lambda_u = 0$, they are called *upper tail independent* or asymptotically independent in the upper tail. Similarly, the *coefficient of lower tail dependence* is

$$\lambda_l := \lambda_l(F) = \lambda_l(X_1,X_2) = \lim_{q\downarrow 0} P\big(X_2 \le F_2^{-1}(q)\,|\,X_1 \le F_1^{-1}(q)\big),$$

again provided a limit $\lambda_l \in [0,1]$ exists. If $\lambda_u = \lambda_l = 0$, then F resp. X_1 and X_2 are *tail independent*.

If the distribution functions F_1 and F_2 are not continuous and strictly increasing, F_1^{-1} and F_2^{-1} in the previous definition again have to be understood as generalized inverses as defined on page 24.

The larger (or less) q, the more rare is the event $\{X_i > F_i^{-1}(q)\}$ (respectively $\{X_i \le F_i^{-1}(q)\}$). Thus the coefficients of tail dependence are nothing but the limits of the conditional probabilities that the second random variable takes extremal values given the first one also does so. In other words, they may be regarded as the probabilities of joint extremal outcomes of X_1 and X_2. This concept also is of some importance in finance: Suppose, for example, that X_1 and X_2 represent two risky assets. If their joint distribution is lower tail dependent, the possibility that both of them suffer severe losses at the same time cannot be neglected. In portfolio credit risk models, X_1 and X_2 may be the state variables of two different firms or credit instruments, and the coefficient of lower tail dependence can then be interpreted as the probability of a joint default. Tail dependence is a copula property, which is illustrated by the subsequent

Proposition 6. *Let $(X_1,X_2)^\top$ be a bivariate random vector with joint distribution function F, continuous margins F_1,F_2, and implied copula C_F as defined in* (38). *Then the following holds:*

a) The coefficients of lower and upper tail dependence can be calculated by

$$\lambda_l = \lim_{q\downarrow 0}\frac{C_F(q,q)}{q} \qquad \text{and} \qquad \lambda_u = \lim_{q\uparrow 1}\frac{1-2q+C_F(q,q)}{1-q}.$$

b) If in addition F_1,F_2 are strictly increasing, λ_l and λ_u can be obtained by

$$\lambda_l = \lim_{q\downarrow 0} P\big(X_2 \le F_2^{-1}(q)\,|\,X_1 = F_1^{-1}(q)\big) + \lim_{q\downarrow 0} P\big(X_1 \le F_1^{-1}(q)\,|\,X_2 = F_2^{-1}(q)\big),$$
$$\lambda_u = \lim_{q\uparrow 1} P\big(X_2 > F_2^{-1}(q)\,|\,X_1 = F_1^{-1}(q)\big) + \lim_{q\uparrow 1} P\big(X_1 > F_1^{-1}(q)\,|\,X_2 = F_2^{-1}(q)\big).$$

The assertion of part a) of the proposition can be found in many textbooks on copulas and dependence, and part b) essentially follows from the ideas of [29, pp. 197 and 210]. A detailed proof can be found in [23, Proposition 2.22].

With the help of these preliminaries, we are now able to give a complete answer to the question which members of the multivariate GH family show tail dependence and which do not. To our knowledge, only symmetric GH distributions have been considered in this regard in the literature so far. By equation (26) and Corollary 3, every multivariate GH distribution with parameter $\beta = \mathbf{0}$ belongs to the class of elliptical distributions, thus the tail independence of $GH_d(\lambda,\alpha,\mathbf{0},\delta,\mu,\Delta)$ (apart from the t limit case with $\alpha = 0$) can be deduced from the more general result below of [24, Theorem 4.3]. It uses the representation $X \overset{d}{=} \mu + RAS$ of an elliptically distributed random vector X which was introduced in Proposition 4.

Theorem 3. *Let $X \overset{d}{=} \mu + RAS \sim E_d\big(\mu,\Sigma,\psi(t)\big)$ be an elliptically distributed random vector with $\Sigma_{ii} > 0$, $1 \le i \le d$, and $|\rho_{ij}| := |\Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}| < 1$ for all $i \ne j$. Then the following statements are equivalent:*

a) The distribution function F_R of R is regularly varying with exponent $p < 0$, that is, $F_R \in \mathscr{R}_p$ (see Definition 4).
b) $(X_i,X_j)^\top$ is tail dependent for all $i \ne j$.

Moreover, if $F_R \in \mathscr{R}_p$ with $p < 0$, then for all $i \ne j$

$$\lambda_u(X_i,X_j) = \lambda_l(X_i,X_j) = \frac{\int_{(\pi/2-\arcsin(\rho_{ij}))/2}^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t}{\int_0^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t}\,.$$

If $X \sim N_d(\mu, y\Delta) \circ G$ has a normal variance mixture distribution which is elliptical by Corollary 3, then X admits the two stochastic representations $\mu + \sqrt{Z}AW \stackrel{d}{=} X \stackrel{d}{=} \mu + RAS$ where the vector μ and the $d \times d$-matrix A on the left and right hand side coincide. This equation suggests that the tail behaviour of the distribution F_R of R is mainly influenced by the distribution G of Z and vice versa. Indeed, one can show that F_R is regularly varying with exponent $2p < 0$ ($F_R \in \mathscr{R}_{2p}$) if and only if $G \in \mathscr{R}_p$ (see [29, pp. 92 and 295f]).

Suppose now $X \sim GH_d(\lambda, \alpha, \mathbf{0}, \delta, \mu, \Delta)$ (excluding the t limiting case for a moment), then by equation (1) the density of the corresponding mixing distribution $GIG(\lambda, \delta, \alpha)$ has a semi-heavy right tail in the sense of Definition 2 with constants $a_2 = \lambda - 1$, $b_2 = \frac{\alpha^2}{2}$ and $c_2 = \frac{(\alpha/\delta)^\lambda}{2K_\lambda(\delta\alpha)}$. (In case of the VG limit, the density of the mixing Gamma distribution $G\big(\lambda, \frac{\alpha^2}{2}\big)$ also has a semi-heavy right tail with the same constants a_2 and b_2, but $c_2 = \frac{(\alpha^2/2)^\lambda}{\Gamma(\lambda)}$.) By Proposition 2 and Definition 3, the distribution functions of $GIG(\lambda, \delta, \alpha)$ and $G\big(\lambda, \frac{\alpha^2}{2}\big)$ both have an exponential right tail with rate b_2. In view of Definition 4 and the subsequent remark, distribution functions with exponential right tails can be regarded as regularly varying with exponent $-\infty$. Consequently, for the distribution function F_R of R in the representation $X \stackrel{d}{=} \mu + RAS$ we have $F_R \in \mathscr{R}_{-\infty}$ as well. Applying Theorem 3 yields

$$\lambda_u(X_i, X_j) = \lambda_l(X_i, X_j) = \lim_{p \to -\infty} \frac{\int_{(\pi/2 - \arcsin(\rho_{ij}))/2}^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t}{\int_0^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t} = 0,$$

showing the tail independence of all symmetric $GH_d(\lambda, \alpha, \mathbf{0}, \delta, \mu, \Delta)$-distributions with parameter $\alpha > 0$.

Remark 5. The convergence of the ratio of the two integrals can be justified as follows: Since $h : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}_-$ with $h(x) = \log(\cos(x))$ has an absolute maximum at $x_0 = 0$ and $h''(x) = -\cos^{-2}(x)$, an application of Laplace's method shows that for all $0 < b \le \frac{\pi}{2}$

$$\int_0^b \cos^{|p|}(t)\,\mathrm{d}t = \int_0^b e^{|p|h(t)}\,\mathrm{d}t \sim \sqrt{\frac{\pi}{-2|p|h''(0)}}\, e^{|p|h(0)} = \sqrt{\frac{\pi}{2|p|}}, \quad |p| \to \infty,$$

consequently

$$\lim_{p \to -\infty} \frac{\int_{(\pi/2 - \arcsin(\rho_{ij}))/2}^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t}{\int_0^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t} = 1 - \lim_{p \to -\infty} \frac{\int_0^{(\pi/2 - \arcsin(\rho_{ij}))/2} \cos^{|p|}(t)\,\mathrm{d}t}{\int_0^{\pi/2} \cos^{|p|}(t)\,\mathrm{d}t} = 0.$$

In the t limiting case, however, we have $X \stackrel{d}{=} \mu + \sqrt{Z}AW \sim t_d(\lambda, \delta, \mu, \Delta)$ with $Z \sim iG\big(\lambda, \frac{\delta^2}{2}\big)$, and from equation (4) it is easily seen that the density $d_{iG(\lambda, \delta^2/2)}$ is regularly varying with exponent $\lambda - 1$. Hence $G = F_Z \in \mathscr{R}_\lambda$ and thus, as pointed out above, $F_R \in \mathscr{R}_{2\lambda}$, so we conclude from Theorem 3 that $\lambda_u(X_i, X_j) = \lambda_l(X_i, X_j) > 0$

for all t distributions $t_d(\lambda,\delta,\mu,\Delta)$. The coefficients are quantified more accurately in Theorem 4 below.

This main result of the present section shows that the dependence behaviour can change dramatically if we move from symmetric to skewed GH distributions with parameter $\beta \neq \mathbf{0}$: in addition to tail independence also complete dependence can occur, that is, both of the coefficients λ_l and λ_u may be equal to one.

Theorem 4. *Let $X \sim GH_2(\lambda,\alpha,\beta,\delta,\mu,\Delta)$ and define $\rho := \frac{\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{22}}}$ as well as $\bar\beta_i := \sqrt{\Delta_{ii}}\beta_i$ for $i=1,2$. Then the following holds:*

a) If $0 \leq \sqrt{\langle\beta,\Delta\beta\rangle} < \alpha$, then the GH distribution (including possible VG limits) is tail independent if $-1<\rho\leq 0$. If $0<\rho<1$, then

$$\lambda_l(X_1,X_2) = \lambda_u(X_1,X_2) = \begin{cases} 0, & c_*, c_*^{-1} > \rho, \\ 1, & \min(c_*, c_*^{-1}) < \rho, \end{cases}$$

where $c_* := \frac{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_2^2}+\bar\beta_1+\rho\bar\beta_2}{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_1^2}+\bar\beta_2+\rho\bar\beta_1}$.

b) If $\lambda<0$ and $\alpha=0$, then $X \sim t_2(\lambda,\delta,\mu,\Delta)$ and

$$\lambda_u(X_1,X_2) = \lambda_l(X_1,X_2) = 2F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}\left(-\sqrt{\frac{(-2\lambda+1)(1-\rho)}{1+\rho}}\right),$$

where $F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}$ denotes the distribution function of the univariate Student's t-distribution $t\big(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0\big)$ with $f=-2\lambda+1$ degrees of freedom.

c) Let $\lambda<0$ and $0<\sqrt{\langle\beta,\Delta\beta\rangle}=\alpha$. If $(\bar\beta_1+\rho\bar\beta_2)(\bar\beta_2+\rho\bar\beta_1)<0$, then

$$\lambda_u(X_1,X_2) = \lambda_l(X_1,X_2) = \begin{cases} 0, & \rho<0, \\ 1, & \rho>0. \end{cases}$$

If $(\bar\beta_1+\rho\bar\beta_2)(\bar\beta_2+\rho\bar\beta_1)>0$, then

$$\lambda_u(X_1,X_2) = \lambda_l(X_1,X_2) = \begin{cases} 0, & c_*, c_*^{-1} > \rho, \\ 1, & \min(c_*, c_*^{-1}) < \rho, \end{cases} \quad \text{where } c_* := \frac{\bar\beta_1+\rho\bar\beta_2}{\bar\beta_2+\rho\bar\beta_1}.$$

Proof. Propositions 6 and 5 state that tail dependence is a copula property and therefore invariant under strictly increasing transformations of X_1 and X_2. But if $X \sim GH_2(\lambda,\alpha,\beta,\delta,\mu,\Delta)$, the linear transformation $Y = \begin{pmatrix} 1/\sqrt{\Delta_{11}} & 0 \\ 0 & 1/\sqrt{\Delta_{22}} \end{pmatrix}(X-\mu)$ obviously is strictly increasing in each component, and Theorem 1 c) implies that $Y \sim GH_2(\bar\lambda,\bar\alpha,\bar\beta,\bar\delta,\mathbf{0},\bar\Delta)$ with $\bar\lambda=\lambda$, $\bar\alpha=\alpha$, $\bar\beta = \begin{pmatrix} \sqrt{\Delta_{11}} & 0 \\ 0 & \sqrt{\Delta_{22}} \end{pmatrix}\beta$, $\bar\delta=\delta$, $\bar\Delta = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho := \Delta_{12}/\sqrt{\Delta_{11}\Delta_{22}}$. Note that we here use the barred parameters defined in Remark 2 because in general $\det(\bar\Delta) = 1-\rho^2 < 1$. As already pointed out in Remarks 2 and 3 on pages 17 and 18, these considerations remain also valid for all GH limit distributions. Hence we can and will always assume $X \sim GH_2(\lambda, \alpha, \bar{\beta}, \delta, \mathbf{0}, \bar{\Delta})$ in the following. The fact that Δ is supposed to be positive definite with $\det(\Delta) = 1$ by definition implies the inequality $0 < \frac{1}{\Delta_{11}\Delta_{22}} = \frac{\Delta_{11}\Delta_{22} - \Delta_{12}^2}{\Delta_{11}\Delta_{22}} = 1 - \rho^2$, thus $|\rho| < 1$.

a) If $X \sim GH_2(\lambda, \alpha, \bar{\beta}, \delta, \mathbf{0}, \bar{\Delta})$ and $0 \leq \sqrt{\langle \beta, \Delta\beta \rangle} < \alpha$, then by Theorem 1 a) the marginal distributions are $X_1 \sim GH\big(\lambda, (\alpha^2 - (1-\rho^2)\bar{\beta}_2^2)^{1/2}, \bar{\beta}_1 + \rho\bar{\beta}_2, \delta, 0\big)$ and $X_2 \sim GH\big(\lambda, (\alpha^2 - (1-\rho^2)\bar{\beta}_1^2)^{1/2}, \bar{\beta}_2 + \rho\bar{\beta}_1, \delta, 0\big)$. To simplify notations we set $\hat{\alpha}_1 := (\alpha^2 - (1-\rho^2)\bar{\beta}_2^2)^{1/2}$, $\hat{\beta}_1 := \bar{\beta}_1 + \rho\bar{\beta}_2$, and $\hat{\alpha}_2 := (\alpha^2 - (1-\rho^2)\bar{\beta}_1^2)^{1/2}$, $\hat{\beta}_2 := \bar{\beta}_2 + \rho\bar{\beta}_1$, then we obtain $\hat{\alpha}_1^2 - \hat{\beta}_1^2 = \hat{\alpha}_2^2 - \hat{\beta}_2^2 = \alpha^2 - \langle \beta, \Delta\beta \rangle > 0$. Thus the densities of $\mathscr{L}(X_1)$ and $\mathscr{L}(X_2)$ both have semi-heavy tails (see Definition 2 and the remark thereafter), and Proposition 2 (or equivalently Corollary 1) implies that the corresponding distribution functions F_1 and F_2 fulfill the assumptions of Lemma 2 b) with $b_1 = \hat{\alpha}_i + \hat{\beta}_i$ and $b_2 = \hat{\alpha}_i - \hat{\beta}_i$, $i = 1,2$. From this we conclude that $F_1^{-1}(q) \sim c_l F_2^{-1}(q)$ for $q \downarrow 0$ as well as $F_1^{-1}(q) \sim c_u F_2^{-1}(q)$ for $q \uparrow 1$ where $c_l := \frac{\hat{\alpha}_2 + \hat{\beta}_2}{\hat{\alpha}_1 + \hat{\beta}_1} > 0$ and $c_u := \frac{\hat{\alpha}_2 - \hat{\beta}_2}{\hat{\alpha}_1 - \hat{\beta}_1} > 0$. Note that $c_l c_u = \frac{\hat{\alpha}_2^2 - \hat{\beta}_2^2}{\hat{\alpha}_1^2 - \hat{\beta}_1^2} = 1$ and thus $c_u = c_l^{-1}$. All this also holds in the VG limit case with $\delta = 0$ because Theorem 1 a) still applies there and the univariate VG marginal densities have semi-heavy tails, too (see p. 9).

Theorem 1 b) states that the conditional distribution of X_i given $X_j = x_j$ (where here and in the following $i, j \in \{1,2\}$ as well as $i \neq j$) is given by $P(X_i \,|\, X_j = x_j) = GH\big(\lambda - \frac{1}{2}, \alpha(1-\rho^2)^{-1/2}, \bar{\beta}_i, \sqrt{\delta^2 + x_j^2}\sqrt{1-\rho^2}, \rho x_j\big)$, and part c) of the same theorem then yields

$$\begin{aligned} &P\left(\frac{X_i - \rho x_j}{\sqrt{\delta^2 + x_j^2}\sqrt{1-\rho^2}} \,\middle|\, X_j = x_j \right) \\ &\qquad = GH\Big(\lambda - \tfrac{1}{2}, \alpha\sqrt{\delta^2 + x_j^2}, \bar{\beta}_i\sqrt{\delta^2 + x_j^2}\sqrt{1-\rho^2}, 1, 0\Big) \\ &\qquad =: GH_{i|j}^*\big(\lambda - \tfrac{1}{2}, \alpha, \bar{\beta}_i, \delta, \rho, x_j\big). \end{aligned}$$

Again, this also remains true in the VG limit case (see Remark 3 on p. 19). Let $F_{i|j}^q$ denote the distribution function of $GH_{i|j}^*\big(\lambda - \frac{1}{2}, \alpha, \bar{\beta}_i, \delta, \rho, F_j^{-1}(q)\big)$ and set

$$h_{i|j}(q) := (1-\rho^2)^{-\frac{1}{2}} \frac{F_i^{-1}(q) - \rho F_j^{-1}(q)}{\sqrt{\delta^2 + \big(F_j^{-1}(q)\big)^2}} \qquad \text{for } q \in (0,1),$$

then we have

$$\begin{aligned} \lim_{q \downarrow 0} P\big(X_i \leq F_i^{-1}(q) \,\big|\, X_j = F_j^{-1}(q)\big) &= \lim_{q \downarrow 0} F_{i|j}^q\big(h_{i|j}(q)\big), \\ \lim_{q \uparrow 1} P\big(X_i > F_i^{-1}(q) \,\big|\, X_j = F_j^{-1}(q)\big) &= \lim_{q \uparrow 1} 1 - F_{i|j}^q\big(h_{i|j}(q)\big). \end{aligned}$$

Moreover, if $\alpha > |\beta| \geq 0$, then $GH(\lambda, r\alpha, r\beta, \delta, \mu) \xrightarrow{w} \varepsilon_\mu$ for $r \to \infty$ because

$$\lim_{r\to\infty} \phi_{GH(\lambda,r\alpha,r\beta,\delta,\mu)}(u) =$$

$$= \lim_{r\to\infty} e^{iu\mu} \left(\frac{(r\alpha)^2 - (r\beta)^2}{(r\alpha)^2 - (r\beta + iu)^2} \right)^{\frac{\lambda}{2}} \frac{K_\lambda\left(\delta\sqrt{(r\alpha)^2 - (r\beta + iu)^2}\right)}{K_\lambda\left(\delta\sqrt{(r\alpha)^2 - (r\beta)^2}\right)}$$

$$= \lim_{r\to\infty} e^{iu\mu} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + \frac{iu}{r})^2} \right)^{\frac{\lambda}{2}} \frac{K_\lambda\left(r\delta\sqrt{\alpha^2 - (\beta + \frac{iu}{r})^2}\right)}{K_\lambda\left(r\delta\sqrt{\alpha^2 - \beta^2}\right)} = e^{iu\mu}$$

which implies that $GH^*_{i|j}\left(\lambda - \frac{1}{2}, \alpha, \bar{\beta}_i, \delta, \rho, F_j^{-1}(q)\right)$ converges weakly to the degenerate distribution ε_0 if $q \downarrow 0$ or $q \uparrow 1$. From the asymptotic relations of the quantile functions $F_1^{-1}(q)$ and $F_2^{-1}(q)$ we further obtain

$$\lim_{q\downarrow 0} h_{i|j}(q) = (1-\rho^2)^{-\frac{1}{2}}\left(\rho - c_l^{j-i}\right), \quad \text{and} \quad \lim_{q\uparrow 1} h_{i|j}(q) = (1-\rho^2)^{-\frac{1}{2}}\left(c_l^{i-j} - \rho\right)$$

(remember $c_u = c_l^{-1}$), consequently

$$\lim_{q\downarrow 0} P\big(X_i \leq F_i^{-1}(q)\,\big|\,X_j = F_j^{-1}(q)\big) = F_{\varepsilon_0}\left(\frac{\rho - c_l^{j-i}}{\sqrt{1-\rho^2}}\right) = \begin{cases} 0, & c_l^{j-i} > \rho, \\ 1, & c_l^{j-i} < \rho, \end{cases}$$

as well as

$$\lim_{q\uparrow 1} P\big(X_i > F_i^{-1}(q)\,\big|\,X_j = F_j^{-1}(q)\big) = 1 - F_{\varepsilon_0}\left(\frac{c_l^{i-j} - \rho}{\sqrt{1-\rho^2}}\right) = \begin{cases} 0, & c_l^{i-j} > \rho, \\ 1, & c_l^{i-j} < \rho, \end{cases}$$

and Proposition 6 b) finally implies that $\lambda_l(X_1, X_2) = \lambda_u(X_1, X_2) = 0$ if and only if $c_l, c_l^{-1} > \rho$. Since $c_l > 0$, the conditions are trivially met if $\rho \leq 0$. If $0 < \rho < 1$, then at most one of the quantities c_l and c_l^{-1} can be smaller than ρ (note that the convergence to a well-defined limit cannot be assured if $c_l^{j-i} = \rho > 0$, therefore we exclude these possibilities in our considerations). This completes the proof of a).

c) Because Theorem 1 a) still applies if $X \sim GH_2(\lambda, \alpha, \bar{\beta}, \delta, \mathbf{0}, \bar{\Delta})$, $\lambda < 0$, and $0 < \sqrt{\langle \beta, \Delta\beta \rangle} = \alpha$, we have, using the notations from above, $X_i \sim GH(\lambda, \hat{\alpha}_i, \hat{\beta}_i, \delta, 0)$, $i = 1, 2$. However, in this case $\hat{\alpha}_i^2 - \hat{\beta}_i^2 = \alpha^2 - \sqrt{\langle \beta, \Delta\beta \rangle} = 0$, hence both marginal distributions are univariate GH limit distributions with $\lambda < 0$ and $\hat{\alpha}_i = |\hat{\beta}_i|$. If $\hat{\beta}_i > 0$, we conclude from equations (20), (21), and Proposition 2 that the tail behaviour of the distribution function is given by $F_i(y) \sim c_{i1}|y|^{\lambda-1}e^{-2\hat{\alpha}_i|y|}$ for $y \to -\infty$ and $1 - F_i(y) \sim c_{i2}|y|^\lambda$ as $y \to \infty$ where

$$c_{i1} = \frac{2^{\lambda-1}}{\hat{\alpha}_i^{\lambda+1}\delta^{2\lambda}\Gamma(|\lambda|)} \qquad \text{and} \qquad c_{i2} = \frac{2^\lambda}{|\lambda|\,\hat{\alpha}_i^\lambda\,\delta^{2\lambda}\Gamma(|\lambda|)}.$$

Lemma 2 now states that $F_i^{-1}(q) \sim \frac{\log(q)}{2\hat{\alpha}_i}$ for $q \downarrow 0$ and $F_i^{-1}(q) \sim \left(\frac{c_{i2}}{1-q}\right)^{\frac{1}{|\lambda|}}$ for $q \uparrow 1$. If $\hat{\beta}_i < 0$, then we analogously obtain $F_i^{-1}(q) \sim -\left(\frac{c_{i2}}{q}\right)^{\frac{1}{|\lambda|}}$ as $q \downarrow 0$ and $F_i^{-1}(q) \sim -\frac{\log(1-q)}{2\hat{\alpha}_i}$ as $q \uparrow 1$. Because the case $\hat{\beta}_i = \bar{\beta}_i + \rho\bar{\beta}_j = 0$ is ruled out by assumption, the equality $0 = \hat{\alpha}_i^2 - \hat{\beta}_i^2 = \alpha^2 - (1-\rho^2)\bar{\beta}_j^2 - (\bar{\beta}_i + \rho\bar{\beta}_j)^2$ implies that $\alpha > \sqrt{1-\rho^2}|\bar{\beta}_i|$. Thus, we can proceed along the same lines as in the proof of part a) and get

$$\lim_{q \downarrow 0} P\big(X_i \le F_i^{-1}(q) \,\big|\, X_j = F_j^{-1}(q)\big) = F_{\varepsilon_0}\Big(\lim_{q \downarrow 0} h_{i|j}(q)\Big),$$
$$\lim_{q \uparrow 1} P\big(X_i > F_i^{-1}(q) \,\big|\, X_j = F_j^{-1}(q)\big) = 1 - F_{\varepsilon_0}\Big(\lim_{q \uparrow 1} h_{i|j}(q)\Big)$$

if we again exclude the cases where $h_{i|j}(q) \to 0$ for the same reasons as above.

Suppose $\hat{\beta}_1, \hat{\beta}_2 > 0$, then $F_1^{-1}(q) \sim c_l F_2^{-1}(q)$ with $c_l = \frac{\hat{\alpha}_2}{\hat{\alpha}_1} = \frac{\hat{\beta}_2}{\hat{\beta}_1} > 0$ as $q \downarrow 0$ and $F_1^{-1}(q) \sim c_u F_2^{-1}(q)$ with $c_u = \left(\frac{c_{12}}{c_{22}}\right)^{1/|\lambda|} = \left(\frac{\hat{\alpha}_2^\lambda}{\hat{\alpha}_1^\lambda}\right)^{1/|\lambda|} = \frac{\hat{\beta}_1}{\hat{\beta}_2} = c_l^{-1}$ for $q \uparrow 1$. Consequently, we again have

$$\lim_{q \downarrow 0} h_{i|j}(q) = (1-\rho^2)^{-\frac{1}{2}}\left(\rho - c_l^{j-i}\right), \qquad \lim_{q \uparrow 1} h_{i|j}(q) = (1-\rho^2)^{-\frac{1}{2}}\left(c_l^{i-j} - \rho\right)$$

and conclude, analogously as before, that $\lambda_l(X_1, X_2) = \lambda_u(X_1, X_2) = 0$ if and only if $c_l, c_l^{-1} > \rho$. If $\hat{\beta}_1, \hat{\beta}_2 < 0$, the tail behaviour of the quantile functions is just exchanged ($c_l \rightsquigarrow c_l^{-1}$ and $c_u = c_l^{-1} \rightsquigarrow c_l$), hence the assertion remains also valid in this case.

Finally, let $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$, then $F_1^{-1}(q) \sim \frac{\log(q)}{2\hat{\alpha}_1}$ and $F_2^{-1}(q) \sim -\left(\frac{c_{22}}{q}\right)^{\frac{1}{|\lambda|}}$ as $q \downarrow 0$, thus $\lim_{q \downarrow 0} \frac{F_1^{-1}(q)}{F_2^{-1}(q)} = 0$ and

$$\lim_{q \downarrow 0} h_{i|j}(q) = \begin{cases} (1-\rho^2)^{-\frac{1}{2}}\rho, & i-j=-1, \\ -\infty, & i-j=1, \end{cases}$$

hence $\lambda_l(X_1, X_2) = 0$ if and only if $\rho < 0$. Further $F_1^{-1}(q) \sim \left(\frac{c_{12}}{1-q}\right)^{\frac{1}{|\lambda|}}$ and $F_2^{-1}(q) \sim -\frac{\log(1-q)}{2\hat{\alpha}_1}$ for $q \uparrow 1$, consequently $\lim_{q \uparrow 1} \frac{F_2^{-1}(q)}{F_1^{-1}(q)} = 0$ and

$$\lim_{q \uparrow 1} h_{i|j}(q) = \begin{cases} -(1-\rho^2)^{-\frac{1}{2}}\rho, & i-j=1, \\ \infty, & i-j=-1, \end{cases}$$

which implies that also $\lambda_u(X_1, X_2) = 0$ if and only if $\rho < 0$. Trivially, all conclusions remain true if $\hat{\beta}_1 < 0$ and $\hat{\beta}_2 > 0$.

b) The proof of this part goes back to [20], see also [29, p. 211]. If $\lambda < 0$ and $\alpha = 0$, we can assume $X \sim GH_2(\lambda, 0, \mathbf{0}, \delta, \mathbf{0}, \bar{\Delta}) = t_2(\lambda, \delta, \mathbf{0}, \bar{\Delta})$, and the marginal distributions are given by $\mathscr{L}(X_1) = \mathscr{L}(X_2) = GH(\lambda,0,0,\delta,0) = t(\lambda,\delta,0)$ according to Theorem 1 a), hence we have $F_1^{-1}(q) = F_2^{-1}(q)$ for all $q \in (0,1)$ in this case. By Theorem 1 b), the conditional distributions also coincide, that is, $P(X_2 \,|\, X_1 = x) = P(X_1 \,|\, X_2 = x) = t\big(\lambda, \sqrt{\delta^2+x^2}\sqrt{1-\rho^2}, \rho x\big)$, and part c) of the same theorem implies

$$P\left(\frac{\sqrt{-2\lambda+1}}{\sqrt{1-\rho^2}}\,\frac{X_2-\rho x}{\sqrt{\delta^2+x^2}}\,\middle|\, X_1 = x\right) = P\left(\frac{\sqrt{-2\lambda+1}}{\sqrt{1-\rho^2}}\,\frac{X_1-\rho x}{\sqrt{\delta^2+x^2}}\,\middle|\, X_2 = x\right)$$
$$= t\big(\lambda - \tfrac{1}{2}, \sqrt{-2\lambda+1}, 0\big).$$

(Note that, in principle, the additional scaling factor $\sqrt{-2\lambda+1}$ is not necessary, but leads to the relation $\delta^2 = -2\lambda+1 = -2(\lambda-\frac{1}{2})$ of the parameters of the conditional distribution which therewith becomes a classical Student's t-distributiom with $f = -2\lambda+1$ degrees of freedom.) If we set

$$h(q) := \frac{\sqrt{-2\lambda+1}}{\sqrt{1-\rho^2}}\,\frac{F_2^{-1}(q)-\rho F_1^{-1}(q)}{\sqrt{\delta^2+(F_1^{-1}(q))^2}} \quad \text{for } q \in (0,1),$$

we get, using that $F_1^{-1}(q) = F_2^{-1}(q)$,

$$\lim_{q\downarrow 0} h(q) = -\frac{\sqrt{-2\lambda+1}\,(1-\rho)}{\sqrt{1-\rho^2}} = -\sqrt{\frac{(-2\lambda+1)(1-\rho)}{1+\rho}} = -\lim_{q\uparrow 1} h(q),$$

consequently

$$\lim_{q\downarrow 0} P\big(X_2 \le F_2^{-1}(q) \,\big|\, X_1 = F_1^{-1}(q)\big) = \lim_{q\downarrow 0} P\big(X_1 \le F_1^{-1}(q) \,\big|\, X_2 = F_2^{-1}(q)\big)$$
$$= \lim_{q\downarrow 0} F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}\big(h(q)\big) = F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}\left(-\sqrt{\frac{(-2\lambda+1)(1-\rho)}{1+\rho}}\right)$$

and

$$\lim_{q\uparrow 1} P\big(X_2 > F_2^{-1}(q) \,\big|\, X_1 = F_1^{-1}(q)\big) = \lim_{q\uparrow 1} P\big(X_1 > F_1^{-1}(q) \,\big|\, X_2 = F_2^{-1}(q)\big)$$
$$= \lim_{q\uparrow 1} 1 - F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}\big(h(q)\big)$$
$$= 1 - F_{t(\lambda-\frac{1}{2},\sqrt{-2\lambda+1},0)}\left(\sqrt{\frac{(-2\lambda+1)(1-\rho)}{1+\rho}}\right).$$

The symmetry relation $F_{t(\lambda-1/2,\sqrt{-2\lambda+1},0)}(-x) = 1 - F_{t(\lambda-1/2,\sqrt{-2\lambda+1},0)}(x)$ and Proposition 6 b) now yield the desired result. □

The conditions $c_* > \rho$ and $c_*^{-1} > \rho$ in Theorem 4 a) are trivially fulfilled if $\bar{\beta}_1 = \bar{\beta}_2$, because then $c_* = c_*^{-1} = 1$. This, in particular, includes the case $\beta = \mathbf{0}$

which provides an alternative proof for the tail independence of symmetric GH distributions (apart from the t limit case). In general, however, it might seem to be a little bit cumbersome to check these conditions. The following corollary provides a simpler criterion for tail independence of GH distributions.

Corollary 4. *Suppose that $X \sim GH_2(\lambda,\alpha,\beta,\delta,\mu,\Delta)$ and $\rho := \frac{\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{22}}} > 0$. Then we have $\lambda_l(X_1,X_2) = \lambda_u(X_1,X_2) = 0$ if either $\sqrt{\langle\beta,\Delta\beta\rangle} < \alpha$ and $\beta_1\beta_2 \geq 0$, or $0 < \sqrt{\langle\beta,\Delta\beta\rangle} = \alpha$ and $\beta_1\beta_2 > 0$.*

Proof. According to Theorem 4 a) and c), we just have to show that the conditions $\beta_1\beta_2 \geq 0$ resp. > 0 imply $c_*, c_*^{-1} > \rho$. Assume $\sqrt{\langle\beta,\Delta\beta\rangle} < \alpha$ first. If both $\beta_1,\beta_2 \geq 0$, then so are $\bar\beta_1 = \sqrt{\Delta_{11}}\,\beta_1$ and $\bar\beta_2 = \sqrt{\Delta_{22}}\,\beta_2$. Since $\rho > 0$, we see from the inequality $0 < \alpha^2 - \langle\beta,\Delta\beta\rangle = \alpha^2 - \bar\beta_1^2 - 2\rho\bar\beta_1\bar\beta_2 - \bar\beta_2^2$ that $\bar\beta_i < \alpha$, $i = 1,2$. Therewith we obtain

$$c_* = \frac{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_2^2}+\bar\beta_1+\rho\bar\beta_2}{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_1^2}+\bar\beta_2+\rho\bar\beta_1} > \frac{\sqrt{\alpha^2-(1-\rho^2)\alpha^2}+\rho\bar\beta_1+\rho\bar\beta_2}{\alpha+\bar\beta_1+\bar\beta_2} = \rho,$$

and an analogous estimate shows that also $c_*^{-1} > \rho$. If $\beta_1 \leq 0$ and $\beta_2 \leq 0$, we use the fact that c_*^{-1} may alternatively be represented by $c_*^{-1} = \frac{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_2^2}-\bar\beta_1-\rho\bar\beta_2}{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_1^2}-\bar\beta_2-\rho\bar\beta_1}$ and similarly conclude that $c_*, c_*^{-1} > \rho$.

Now, let $0 < \sqrt{\langle\beta,\Delta\beta\rangle} = \alpha$ and note that the condition $\beta_1\beta_2 > 0$ implies $(\bar\beta_1+\rho\bar\beta_2)(\bar\beta_2+\rho\bar\beta_1) > 0$. If both $\beta_1,\beta_2 > 0$, then $c_* = \frac{\bar\beta_1+\rho\bar\beta_2}{\bar\beta_2+\rho\bar\beta_1} > \frac{\rho\bar\beta_1+\rho\bar\beta_2}{\bar\beta_2+\bar\beta_1} = \rho$, and $c_*^{-1} > \rho$ follows analogously. If $\beta_1,\beta_2 < 0$, the same result is obtained by using the representation $c_* = \frac{-\bar\beta_1-\rho\bar\beta_2}{-\bar\beta_2-\rho\bar\beta_1}$. □

An immediate consequence of the preceding corollary is that complete dependence ($\lambda_l(X_1,X_2) = \lambda_u(X_1,X_2) = 1$) within bivariate GH distributions can only occur if the parameters β_1 and β_2 have opposite signs, and one might conjecture that the conditions $c_*, c_*^{-1} > \rho$ are also always fulfilled in these cases such that a two-dimensional GH distribution would be tail independent for almost any choice of parameters. However, this is not true, and it is fairly easy to construct counterexamples: Take $\alpha = 4$, $\bar\beta_1 = 3$, $\bar\beta_2 = -2$, and $\rho = 0.3$, then $\alpha^2 - \langle\beta,\Delta\beta\rangle = \alpha^2 - \bar\beta_1^2 - 2\rho\bar\beta_1\bar\beta_2 - \bar\beta_2^2 = 6.6$ and

$$c_*^{-1} = \frac{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_1^2}+\bar\beta_2+\rho\bar\beta_1}{\sqrt{\alpha^2-(1-\rho^2)\bar\beta_2^2}+\bar\beta_1+\rho\bar\beta_2} \approx 0.286 < \rho.$$

The corresponding copula density is shown in Figure 1. In view of Theorem 4, the densities displayed there represent all possible tail dependencies of GH distributions: $NIG_2(10,\mathbf{0},0.2,\mathbf{0},\bar\Delta)$ and $NIG_2(10,\binom{4}{1},0.2,\mathbf{0},\bar\Delta)$ are tail independent, $NIG_2(4,\binom{3}{-2},0.2,\mathbf{0},\bar\Delta)$ is completely dependent, and $t_2(-2,2,\mathbf{0},\bar\Delta)$ lies in between.

The fact that for GH distributions the coefficients of tail dependence can only take the most extreme values 0 and 1 may surely be surprising at first glance, but this phenomenon can also be observed in other distribution classes (making it possibly less astonishing). For example, [3] found a similar behaviour for the upper tail dependence coefficient $\lambda_u(X_1,X_2)$ of a skewed grouped t distribution. An alternative derivation and discussion of their results can also be found in [21].

Thus, the dependence structure of multivariate GH distributions is fairly strict in some sense since it neither allows independent components nor non-trivial values of the tail dependence coefficients. A possible way to relax these restrictions is to consider affine mappings of random vectors with independent GH distributed components: If $Y \overset{d}{=} AX+\mu$, where $\mu \in \mathbb{R}^d$, A is a lower triangular $d\times d$-matrix, and $X=(X_1,\dots,X_d)^\top$ with independent $X_i \sim GH(\lambda_i,\alpha_i,\beta_i,1,0)$, $1\le i\le d$, then Y is said to have a multivariate affine GH distribution. Dependent on the choice of A, $\mathscr{L}(Y)$ can either possess independent margins or show upper and lower tail dependence. [32] provide a thorough discussion of this model.

Finally, we want to remark that the dependence structure of factor models for credit portfolios which have already been mentioned on p. 11 significantly differs from that of multivariate distributions discussed above. Recall that the state variables X_i in general factor model are given by

$$X_i := \sqrt{\rho}\,M + \sqrt{1-\rho}\,Z_i, \qquad 0\le\rho<1, \quad i=1,\dots,N, \tag{41}$$

where $M,Z_1,\dots,Z_N$ are assumed to be independent and, in addition, the Z_i are identically distributed (hence so are the X_i). The corresponding distribution functions are denoted by F_M, F_Z, F_X and are usually supposed to be continuous and strictly increasing on $\mathbb{R}$. If M and the Z_i are standard normal distributed ($M,Z_i\sim N(0,1)$), then also the joint distribution of the X_i is a multivariate normal distribution with the associated implied copula. However, if we assume the factors M and Z_i to follow a GH distribution ($M\sim GH(\lambda_M,\alpha_M,\beta_M,\delta_M,\mu_M)$, $Z_i\sim GH(\lambda_Z,\alpha_Z,\beta_Z,\delta_Z,\mu_Z)$ for all $1\le i\le N$), then the distribution of the random vector $X=(X_1,\dots,X_N)^\top$ is not a multivariate generalized hyperbolic one. This can easily deduced from the fact that the X_i in general are not GH distributed due the lack of stability under convolutions of the GH class, whereas a multivariate GH distribution must always have univariate GH margins according to Theorem 1 a). Consequently, the implied copula C_{G_X} of the distribution G_X of the vector X also differs from the implied copula of a multivariate GH distribution. The factor copula C_{G_X} can be calculated by

$$\begin{aligned} C_{G_X}(u_1,\dots,u_N) &= G_X\big(F_X^{-1}(u_1),\dots,F_X^{-1}(u_N)\big) \\ &= E\big[P\big(X_1\le F_X^{-1}(u_1),\dots,X_N\le F_X^{-1}(u_N)\,|\,M\big)\big] \\ &= \int_{\mathbb{R}} \prod_{i=1}^N F_Z\left(\frac{F_X^{-1}(u_i)-\sqrt{\rho}\,y}{\sqrt{1-\rho}}\right) F_M(\mathrm{d}y) \end{aligned} \tag{42}$$

and admits tail dependence ($\lambda_u(X_i,X_j),\lambda_l(X_i,X_j)>0$, $1\le i\ne j\le N$) if and only if the M is heavy tailed, that is, $F_M\in\mathscr{R}_p$ for some $-\infty<p<0$ (see Definition 4). This has been shown in [28]. Hence, we can conclude that factor models with GH distributions can show tail dependence if and only if $F_M = t(\lambda, \delta, \mu)$ or $F_M = GH(\lambda, \alpha, \pm\alpha, \delta, \mu)$.

References

1. Aas, K., Haff, I.H.: The generalized hyperbolic skew Student's t-distribution. J. Fin. Econ. **4**, 275–309 (2006)
2. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Math. Fin. **9**(3), 203–228 (1999)
3. Banachewicz, K., van der Vaart, A.: Tail dependence of skewed grouped t-distributions. Stat. & Prob. Letters **78**, 2388–2399 (2008)
4. Barndorff-Nielsen, O.E.: Exponentially decreasing distributions for the logarithm of particle size. Proc. Roy. Soc. London A **353**, 401–419 (1977)
5. Barndorff-Nielsen, O.E.: Processes of normal inverse Gaussian type. Fin. Stoch. **2**, 41–68 (1998)
6. Bingham, N.H., Goldie, C.M., Omey, E.: Regularly varying probability densities. Publ. Inst. Math. Beograd **80(94)**, 47–57 (2006)
7. Blæsild, P.: The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. Biometrika **68**, 251–263 (1981)
8. Blæsild, P., Jensen, J.L.: Multivariate distributions of hyperbolic type. In: Taillie, C., Patil, G.P., Baldessari, B. (eds.) Statistical Distributions in Scientific Work 4, pp. 45–66, Reidel, Dordrecht (1981)
9. Daul, S., De Giorgi, E., Lindskog, F., McNeil, A.: Using the grouped t-copula. Risk (November 2003), 73–76
10. Eberlein, E.: Application of generalized hyperbolic Lévy motions to finance. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S. (eds.) Lévy Processes: Theory and Applications, pp. 319–336, Birkhäuser, Boston (2001)
11. Eberlein, E., Frey, R., v. Hammerstein, E.A.: Advanced credit portfolio modeling and CDO pricing. In: Jäger, W., Krebs, H.-J. (eds.) Mathematics - Key Technology for the Future: Joint Projects between Universities and Industry 2004–2007, pp. 253–280. Springer, Berlin (2008)
12. Eberlein, E., Keller, U.: Hyperbolic distributions in finance. Bernoulli **1**, 281–299 (1995)
13. Eberlein, E., Kluge, W.: Exact pricing formulae for caps and swaptions in a Lévy term structure model. J. Comp. Fin. **9**, 99–125 (2006)
14. Eberlein, E., Kluge, W.: Calibration of Lévy term structure models. In: Fu, M.C., Jarrow, R.A., Yen, J.-Y., Elliott, R.J. (eds.) Advances in Mathematical Finance, pp. 147–172, Birkhäuser, Boston (2007)
15. Eberlein, E., Koval, N.: A cross-currency Lévy market model. Quant. Fin. **6**, 465–480 (2006)
16. Eberlein, E., Özkan, F.: Time consistency of Lévy models. Quant. Fin. **3**, 40–50 (2003)
17. Eberlein, E., Prause, K.: The generalized hyperbolic model: financial derivatives and risk measures. In: Geman, H., Madan, D.B., Pliska, S., Vorst, T. (eds.) Mathematical Finance – Bachelier Congress 2000, pp. 245–267, Springer, Berlin (2002)
18. Eberlein, E., Raible, S.: Term structure models driven by general Lévy processes. Math. Fin. **9**, 31–54 (1999)
19. Embrechts, P., Goldie, C.M.: On closure and factorization properties of subexponential distributions. J. Austr. Math. Soc. A **29**, 243–256 (1980)
20. Embrechts, P., McNeil, A.J., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M.A.H. (ed.) Risk Management: Value at Risk and Beyond, pp. 176–223, Cambridge University Press, Cambridge (2002)
21. Fung, T., Seneta, E.: Tail dependence for two skew t distributions. Stat. & Prob. Letters **80**, 784–791 (2010)

22. Good, I.J.: On the population frequencies of species and the estimation of population parameters. Biometrika **40**, 237–264 (1953)
23. v. Hammerstein, E.A.: Generalized hyperbolic distributions: Theory and applications to CDO pricing, PhD Thesis, University of Freiburg (2011). Available at http://www.freidok.uni-freiburg.de/volltexte/7974/
24. Hult, H., Lindskog, F.: Multivariate extremes, aggregation and dependence in elliptical distributions. Adv. Appl. Prob. **34**, 587–608 (2002)
25. Jørgensen, B.: Statistical Properties of the Generalized Inverse Gaussian distribution. Lecture Notes in Statistics **9**, Springer, New York (1982)
26. Madan, D.B., Carr, P.P., Chang, E.C.: The variance gamma process and option pricing. Europ. Fin. Rev. **2**, 79–105 (1998)
27. Madan, D.B., Seneta, E.: The variance gamma (V.G.) model for share market returns. J. Business **63**, 511–524 (1990)
28. Malevergne, Y., Sornette, D.: How to account for extreme co-movements between individual stocks and the market. J. Risk **6**, 71–116 (2004)
29. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, Princeton (2005)
30. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics **139**, Springer, New York (1999)
31. Rüschendorf, L.: On the distributional transform, Sklar's theorem, and the empirical copula process. J. Stat. Plan. Inf. **139**(11), 3921–3927 (2009)
32. Schmidt, R., Hrycej, T., Stützle, E.: Multivariate distribution models with generalized hyperbolic margins. Comp. Stat. & Data An. **50**, 2065–2096 (2006)
33. Sichel, H.: Statistical valuation of diamondiferous deposits. J. South. Afric. Inst. Min. Metal. **73**, 235–243 (1973)
34. Sichel, H.: On a distribution representing sentence-length in written prose. J. Roy. Stat. Soc. A. **137**, 25–34 (1974)
35. Sklar, A.: Fonctions de répartition à *n* dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)