R. Schmidt et al., 2006, Multivariate distribution models with generalized hyperbolic margins, *Comput. Statist. Data Anal.* **50**, 2065–2096.

Multivariate distribution models with generalized hyperbolic margins

Rafael Schmidt [1]

Department of Economic and Social Statistics, University of Cologne.
Department of Statistics, The London School of Economics & Political Science.

Tomas Hrycej, Eric Stützle

Department of Information Mining, DaimlerChrysler AG - Research and Technology.

Abstract

Multivariate generalized hyperbolic distributions represent an attractive family of distributions (with exponentially decreasing tails) for multivariate data modelling. However, in a limited data environment, robust and fast estimation procedures are rare. An alternative class of multivariate distributions (with exponentially decreasing tails) is proposed which comprises affine-linearly transformed random vectors with stochastically independent and generalized hyperbolic marginals. The latter distributions possess good estimation properties and have attractive dependence structures which are explored in detail. In particular, dependencies of extreme events (tail dependence) can be modelled within this class of multivariate distributions. In addition the necessary estimation and random-number generation procedures are provided. Various advantages and disadvantages of both types of distributions are discussed and illustrated via a simulation study.

Key words: Multivariate distributions, Generalized hyperbolic distributions, Affine-linear transforms, Copula, Tail dependence, Estimation, Random number generation

Email address: `rafael.schmidt@uni-koeln.de` (Rafael Schmidt).
URL: `www.uni-koeln.de/wiso-fak/wisostatsem/autoren/schmidt` (Rafael Schmidt).
[1] Department of Economic and Social Statistics, University of Cologne, Albertus-Magnus-Platz, 50923 Koeln, Germany. Tel.: +49 (0) 221 470 2283, Fax: +49 (0) 221 470 5074. This work was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

1 Introduction

Data analysis with generalized hyperbolic distributions has become quite popular in various areas of theoretical and applied statistics. Originally, Barndorff-Nielsen (1977) utilized this class of distributions to model grain size distributions of wind-blown sand (cf. also Barndorff-Nielsen and Blæsild (1981) and Olbricht (1991)). In econometrical finance the latter family of distributions has been used for multidimensional asset-return modelling. In this context, the generalized hyperbolic distribution replaces the Gaussian distribution, which is not able to describe the fat tails and the (distributional) skewness of most financial asset-return data. References are Eberlein and Keller (1995), Prause (1999), Bauer (2000), Bingham and Kiesel (2001), and Eberlein (2001).

Multivariate generalized hyperbolic distributions (in short: MGH distributions) were introduced and investigated by Barndorff-Nielsen (1978) and Blæsild and Jensen (1981). These distributions have attractive analytical and statistical properties whereas robust and fast parameter estimation turns out to be difficult in higher dimensions. Furthermore, MGH distribution functions possess no parameter constellation for which they are the product of their marginal distribution functions. However, many applications require the multivariate distribution function to model both: Marginal dependence and independence. Because of these and other shortcomings (see also Section 4) we introduce and explore a new class of multivariate distributions, the so called multivariate affine generalized hyperbolic distributions (in short: MAGH distributions). This class of distributions has an appealing stochastic representation and, in contrast to the MGH distributions, the estimation and simulation algorithms are easier. Moreover, our simulation study reveals that the goodness-of-fit of the MAGH distribution is comparable to that of the MGH distribution. The one-dimensional marginals of an MAGH distribution are even more flexible due to more flexibility in the parameter choice.

After a brief introduction of MGH and MAGH distributions in Sections 2 and 3, we start with a discussion of advantages and disadvantages of both types of distributions in data modelling and data analysis (see Section 4).

The analysis of the paper is divided into four different **stages**:

(1) Elaboration of statistical-mathematical properties of MGH and MAGH distributions (Section 5),

(2) Computational procedures for the parameter estimation of MGH and MAGH distributions (Section 6),

(3) Random number generation for MGH and MAGH distributions (Section 7),

(4) Simulation study and real data analysis (Sections 8 and 9).

The paper's main contributions on each stage are as follows: At the first stage we concentrate on the dependence structure of both distributions by utilizing the theory of copulae. In particular, we show that the dependence structure of an MAGH distribution is very appealing for data modelling. This is because the correlation matrix as an important dependence measure is more intuitive and easier to handle for an MAGH distribution than for an MGH distribution. Further, certain parameter constellations imply independent margins of the MAGH distribution whereas the margins of an MGH distribution do not have this property. Moreover, in contrast to MGH distributions, MAGH distributions can model dependencies of extreme events (so called tail dependence) which is an important property for financial risk analysis. At the next stage we show that the parameters of the MAGH distribution can be estimated in a simple two-stage method. This procedure reduces to the estimation of the covariance matrix and the parameters which are related to the univariate marginal distributions. Thus, generally speaking, the estimation simplifies to a one-dimensional estimation problem. In contrast, this type of estimation for MGH distributions can only be perform within the subclass of elliptically contoured distributions. The third stage establishes a fast and simple random-number generator which is based on the well-known rejection algorithm. In particular, we provide the explicit algorithm, while avoiding difficult minimization routines, which outperforms the known algorithms in the literature for MGH distributions under particular parameter constellations. Finally, at the fourth stage we present a detailed simulation study to illustrate the suitability of MAGH distributions for data modelling. The Appendix contains various results and proofs which are mainly related to the tail behavior of MGH and MAGH distributions.

2 Multivariate generalized hyperbolic model (MGH)

In the first place, a subclass of MGH distributions, namely the hyperbolic distributions, has been introduced via so-called variance-mean mixtures of inverse Gaussian distributions. This subclass suffers from not having hyperbolic distributed marginals, i.e., the subclass is not closed with respect to passing to the marginal distributions. Therefore and because of other theoretical aspects, Barndorff-Nielsen (1977) extended this class to the family of MGH distributions. Many different parametric representations of MGH density functions are provided in the literature, see e.g. Blæsild and Jensen (1981). The following density representation is appropriate in our context.

Definition 1 (MGH distribution) *An n-dimensional random vector X is said to have a* multivariate generalized hyperbolic *(MGH) distribution with location vector $\mu \in \mathbb{R}^n$ and scaling matrix $\Sigma \in \mathbb{R}^{n\times n}$ if it has the stochastic representation $X \stackrel{d}{=} A'Y + \mu$ for some lower triangular matrix $A' \in \mathbb{R}^{n\times n}$ such that $A'A = \Sigma$ is positive-definite and Y has a density function of the form $(y \in \mathbb{R}^n)$:*

$$f_Y(y) = c\frac{K_{\lambda-n/2}(\alpha\sqrt{1+y'y})}{(1+y'y)^{n/4-\lambda/2}}e^{\alpha\beta' y}, \quad \textit{with } c = \frac{\alpha^{n/2}\,(1-\beta'\beta)^{\lambda/2}}{(2\pi)^{n/2}K_\lambda(\alpha\sqrt{1-\beta'\beta})}. \tag{1}$$

K_ν denotes the modified Bessel-function of the third kind with index ν (cf. Magnus, Oberhettinger, and Soni (1966), pp. 65) and the parameter domain is $\|\beta\|_2 < 1$, $\alpha > 0$ and $\lambda \in \mathbb{R}$ ($\|\cdot\|_2$ denotes the Euclidian norm). The family of n-dimensional generalized hyperbolic distributions is denoted by $MGH_n(\mu, \Sigma, \omega)$, where $\omega := (\lambda, \alpha, \beta)$.

An important property of the above parameterization of the MGH density function is its invariance under affine-linear transformations. For $\lambda = (n+1)/2$ we obtain the *multivariate hyperbolic* density and for $\lambda = -1/2$ the *multivariate normal inverse Gaussian* density. Hence $\lambda = 1$ leads to hyperbolically distributed one-dimensional margins. It can be shown that MGH distributions with $\lambda = 1$ are closed with respect to passing to the marginal distributions and under affine-linear transformations. The latter subclass turns out to be important for practical applications (see also Section 8.1).

An MGH distribution belongs to the class of elliptically contoured distributions if and only if $\beta = (0, \ldots, 0)'$. In this case the density function of X can be represented as

$$f_X(x) = |\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu)), \quad x \in \mathbb{R}^n, \tag{2}$$

for some density generator function $g : I\!R_+ \to I\!R_+$. Consequently, the random vector Y in Definition 1 is spherically distributed. The density generator g in (2) is given by $g(u) = cK_{\lambda-n/2}(\alpha\sqrt{1+u})/(1+u)^{n/4-\lambda/2},\ u \in I\!R$, with some normalizing constant c. For a detailed treatment of elliptically contoured distributions, see Fang, Kotz, and Ng (1990) or Cambanis, Huang, and Simons (1981).

Remark. Usually the following representation of an MGH density is given in the literature:

$$\bar{f}_X(x) = \bar{c}\frac{K_{\bar{\lambda}-n/2}(\bar{\alpha}\sqrt{\bar{\delta}^2 + (x-\bar{\mu})'\bar{\Sigma}^{-1}(x-\bar{\mu})})}{\left(\bar{\alpha}^{-1}\sqrt{\bar{\delta}^2 + (x-\bar{\mu})'\bar{\Sigma}^{-1}(x-\bar{\mu})}\right)^{n/2-\bar{\lambda}}} e^{\bar{\beta}'(x-\bar{\mu})}, \quad x \in I\!R^n, \tag{3}$$

with some normalizing constant $\bar{c}$. The domain of variation [2] of the parameter vector $\bar{\omega} = (\bar{\lambda}, \bar{\alpha}, \bar{\delta}, \bar{\beta})$ is as follows: $\bar{\lambda}, \bar{\alpha} \in I\!R$, $\bar{\beta}, \bar{\mu} \in I\!R^n$, $\bar{\delta} \in I\!R_+$, $\bar{\beta}'\bar{\Sigma}\bar{\beta} < \bar{\alpha}^2$ and $\bar{\Sigma} \in I\!R^{n\times n}$ being a positive-definite matrix with determinant $|\bar{\Sigma}| = 1$. The one-to-one mapping between the parameter vector ω corresponding to (1) and $\bar{\omega}$ corresponding to (3) is given by: $\lambda = \bar{\lambda},\ \mu = \bar{\mu},\ \alpha = \bar{\alpha}\bar{\delta},\ \beta = 1/\bar{\alpha} \cdot \bar{A}\bar{\beta},\ \bar{A}'\bar{A} = \bar{\Sigma}$, and $\Sigma = \bar{\delta}^2\bar{\Sigma}$.

3 Multivariate affine generalized hyperbolic model (MAGH)

A disadvantage of multivariate generalized hyperbolic distributions (and of many other families of multivariate distributions) is that the margins X_i of $X = (X_1, \ldots, X_n)'$ are not mutually independent for some choice of the scaling matrix Σ. In other words, they do not allow the modelling of phenomena where random variables result as the sum of independent random variables. This shortcoming is serious since the independence may be an undisputable property of the problem for which the stochastic model is sought. Furthermore, in case of asymmetry (i.e., $\beta \neq 0$) the covariance matrix is in a complex relationship with the matrix Σ, which is shown in the next section.

Therefore we propose an alternative concept. Instead of a multivariate generalized hyperbolic distribution, a distribution is considered which is composed of n independent margins with univariate generalized hyperbolic distributions with zero location and unit scaling. Such a canonical random vector is then subject to an affine-linear transformation. As a consequence, the transformation matrix can be modelled proportionally to the square root of the covariance-matrix inverse even in the asymmetric case. This property holds, for example, for multivariate normal distributions.

Definition 2 (MAGH distribution) *An n-dimensional random vector X is said to be* multivariate affine generalized hyperbolic *(MAGH) distributed with location vector $\mu \in I\!R^n$ and scaling matrix $\Sigma \in I\!R^{n\times n}$ if it has the following stochastic representation $X \stackrel{d}{=} A'Y + \mu$ for some lower triangular matrix $A \in I\!R^{n\times n}$ such that $A'A = \Sigma$ is positive-definite and the random vector $Y = (Y_1, \ldots, Y_n)'$ consists of mutually independent random variables $Y_i \in MGH_1(0, 1, \omega_i),\ i = 1, \ldots, n$. In particular the one-dimensional margins of Y are generalized hyperbolic distributed. The family of n-dimensional affine generalized hyperbolic distributions is denoted by $MAGH_n(\mu, \Sigma, \omega)$, where $\omega := (\omega_1, \ldots, \omega_n)$ and $\omega_i := (\lambda_i, \alpha_i, \beta_i)',\ i = 1, \ldots, n$.*

[2] This representation omits the limiting distributions obtained at the boundary of the parameter space; see e.g. Blæsild and Jensen (1981)

Observe that an MAGH distribution has independent margins if the scaling matrix Σ equals the identity matrix I. However, no MAGH distribution belongs to the class of elliptically contoured distributions for dimension $n \geq 2$ which is illustrated by the density contour-plots in Figure 2.

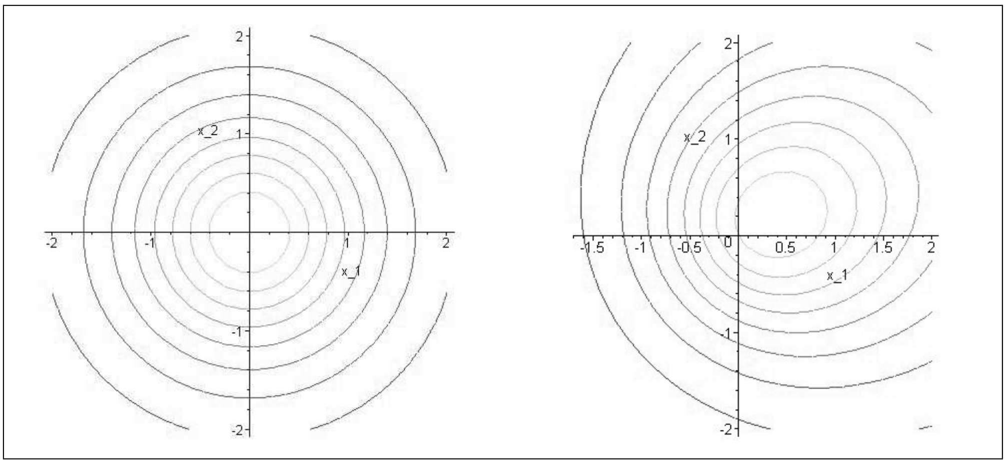


Fig. 1. Contour-plots of the bivariate density function of an $MGH_2(0, I, \omega)$ distribution with parameters $\lambda = 1$, $\alpha = 1$ and $\beta = (0, 0)'$ (left figure), $\beta = (0.5, 0.25)'$ (right figure)

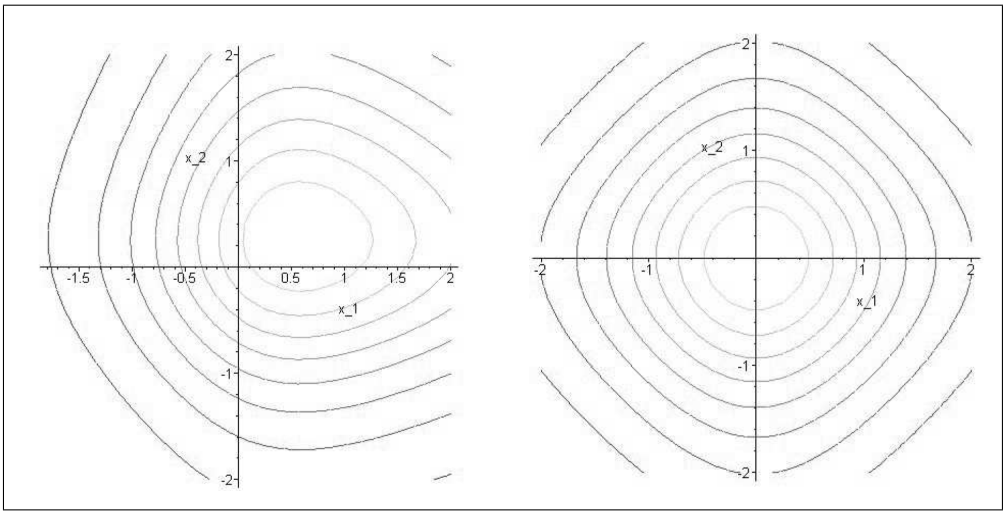


Fig. 2. Contour-plots of the bivariate density function of an $MAGH_2(0, I, \omega)$ distribution with parameters $\lambda = (1, 1)'$, $\alpha = (1, 1)'$ and $\beta = (0, 0)'$ (left figure), $\beta = (0.5, 0.25)'$ (right figure).

General affine transformations. The consideration of the lower triangular matrix A' in the stochastic representations of Definitions 1 and 2 is essential since any other decomposition of the scaling matrix Σ would lead to a different class of distributions. This phenomenon and a possible extension are discussed below.

Only the elliptically contoured subclass of the MGH distributions is invariant with respect to different decompositions $A'A = \Sigma$. In particular, all decompositions of the scaling matrix Σ lead to the same distribution since they enter the characteristic function via the form $\Sigma = A'A$. Equation (2) also justifies the latter property. However, in the asymmetric or general affine case this equivalence does not hold anymore. In this case, for example, the matrix A can be sought via a singular value decomposition $A = UWV'$, where W is a diagonal matrix having the square roots of eigenvalues of $\Sigma = A'A$ on its diagonal and where the matrix V consists of the corresponding eigenvectors of Σ. The matrices W and V are directly determined from Σ whereas the matrix U might be some arbitrary matrix with orthonormal columns (rotation and flip). However, the most common case, of course, is $U = I$. Here the matrix A is directly computed from Σ utilizing its eigenvalues and eigenvectors. Consequently, every margin of Y is distributed according to a linear combination of the margins of X determined by the principal components (PC) (i.e., the eigenvectors) of the covariance matrix Σ: $Y = A'^{-1}X = W^{-1}V'X$.

4 MGH versus MAGH: Advantages and disadvantages

In this section we list and compare some advantages and disadvantages of MGH distributions and MAGH distributions. We start with the *distributional flexibility* to fit real data. An outstanding property of MAGH distributions is that, after an affine-linear transformation, all one-dimensional margins can be fitted separately via different generalized hyperbolic distributions. In contrast to this, the one-dimensional margins of MGH distributions are not that flexible since the parameters α and λ relate to the entire multivariate distribution and determine a strong structural behavior (see Definition 1). However, this structure causes a large subclass of MGH distributions to belong to the family of elliptically contoured distributions which inherit many useful statistical and analytical properties from multivariate normal distributions. For example, the family of elliptically contoured distribution is closed under linear regression and passing to the marginal distributions (see Cambanis, Huang, and Simons (1981)).

Regarding the *dependence structure*, the MAGH distributions may have independent margins for some parameter constellation (see Theorem 4). In particular, they support models which are based on a linear combination of independent factors. In contrast, the MGH distributions are not capable of modelling independent margins. They even yield "extremal" dependencies for bivariate distributions having correlation zero. Moreover, the correlation matrix of MAGH distributions is proportional to the scaling matrix Σ within a large subclass of asymmetric MAGH distributions (see Theorem 6), whereas Σ is hardly to interpret for skewed MGH distributions. Further, the copula of MAGH distributions, being the dependence structure of an affine-linearly transformed random vector with independent components, is quite illustrative and possesses many appealing modelling properties. On the other hand, the copula structure of MGH distributions may suffer from inflexibility. Regarding the dependence of extreme events, the MAGH distributions can model tail dependence whereas MGH distributions are always tail independent. Therefore, MAGH distributions are suitable especially within the field of risk management.

Sections 6 and 8 reveal that in contrast to MGH distributions, *parameter estimation* for MAGH distributions is considerably simpler and more robust. Even in an asymmetric environment, the parameters of MAGH distributions can be identified in a two-stage procedure which has a considerable computational advantage in higher dimensions. The same procedure can be applied for elliptically contoured MGH distributions ($\beta = 0$). The random vector *generation* algorithms for MGH and MAGH distributions turn out to be equally efficient and fast, irrespectively of the dimension.

The simulations in Sections 8 and 9 show that both distributions fit simulated and real data well. Thus, summarizing the above advantages and disadvantages, the MAGH distributions have much to recommend them regarding their parameter estimation, dependence structure, and random vector generation. However, it depends also on the kind of application and the user's taste which model to prefer.

5 Some properties of MGH and MAGH distributions

In this section we investigate and compare several statistical-mathematical properties of the MGH distribution and the MAGH distribution. In this context, the dependence structures will be of our particular interest. It has been already mentioned that the one-dimensional marginals of both types of distributions are quite flexible. The copula technique combines these marginals with the respective dependence structure leading to a multidimensional MGH or MAGH distribution. The dependence structure of affine-linearly transformed distributions, such as MAGH distributions, in terms of copulae has not attracted much attention in the literature yet. However, the usefulness of copulae has been shown for many applications especially in finance (see Cherubini, Luciano, and Vecchiato (2004) for an overview). The main contribution of this section is a detailed analysis of the dependence structure of MGH and MAGH distributions. We consider the respective copulae and various dependence measures, such as the covariance and correlation coefficients, Kendall's tau, and tail dependence. It turns out that the dependence structure of MAGH distributions is quite different to the respective MGH counterpart, although for example, the contour plot in Figure 2 does not reflect this fact. In particular, we show that the behavior of common extreme events is different and that for any parameter constellation the margins of the MGH distributions cannot be independent.

It is precisely the copula which encodes all information on the dependence structure unencumbered by the information on marginal distributions and which couples the marginal distributions to give the joint distribution. In particular if $X = (X_1, \ldots, X_n)'$ has joint distribution F with continuous marginals $F_1, \ldots, F_n$, then the distribution function of the vector $(F_1(X_1), \ldots, F_n(X_n))'$ is a copula C, and $F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$.

The MGH and MAGH copulae. According to Definitions 1 and 2, the MGH and MAGH distributions are represented by affine-linear transformations of random vectors following a standardized MGH distribution and MAGH distribution, respectively. Leaving the affine-linear transformation aside, we are interested in the dependence structure (copula) of the underlying random vector. In particular we set the scaling matrix $\Sigma = I$ and $\mu = 0$. Note that the copula of an MGH or MAGH distribution does not depend on the location vector, i.e., μ is not a copula parameter.

Theorem 3 *Let $Y \in MGH_n(0, I, \omega)$. Then the copula density function of Y is given by*

$$c(u_1, \ldots, u_n) = c \frac{K_{\lambda-n/2}(\alpha\sqrt{1+y'y})}{(1+y'y)^{n/4-\lambda/2}} \prod_{i=1}^{n} \frac{(1+y_i^2)^{1/4-\lambda/2}}{K_{\lambda-1/2}(\alpha\sqrt{1+y_i^2})} \frac{\exp(\alpha\beta' y)}{\exp(\prod_{i=1}^{n} \alpha\beta_i y_i)} \Bigg|_{y_i = F_i^{-1}(u_i)},$$

for $u_i \in [0,1],\ i = 1, \ldots, n$, and some normalizing constant c. Here F_i refers to the distribution function of the one-dimensional margin $Y_i,\ i = 1, \ldots, n$. Let $Y \in MAGH_n(0, I, \omega)$. Then the corresponding copula equals the independence copula, i.e, the copula density function is given by

$$c(u_1, \ldots, u_n) \equiv 1, \quad u_i \in [0,1],\ i = 1, \ldots, n.$$

Proof. The first part follows from the copula definition. For the second part note that Y has independent margins if and only if Y possesses the independence copula $C(u_1, \ldots, u_n) = u_1 \cdot \ldots \cdot u_n$ according to Theorem 2.10.14 in Nelsen (1999). □

Figure 3 clearly reveals that, in contrast to the MGH distribution, the copula of an MAGH distribution shows strong dependence in the limiting corners of the respective quadrants. This property is investigated in more detail later when we discuss the concept of tail dependence.

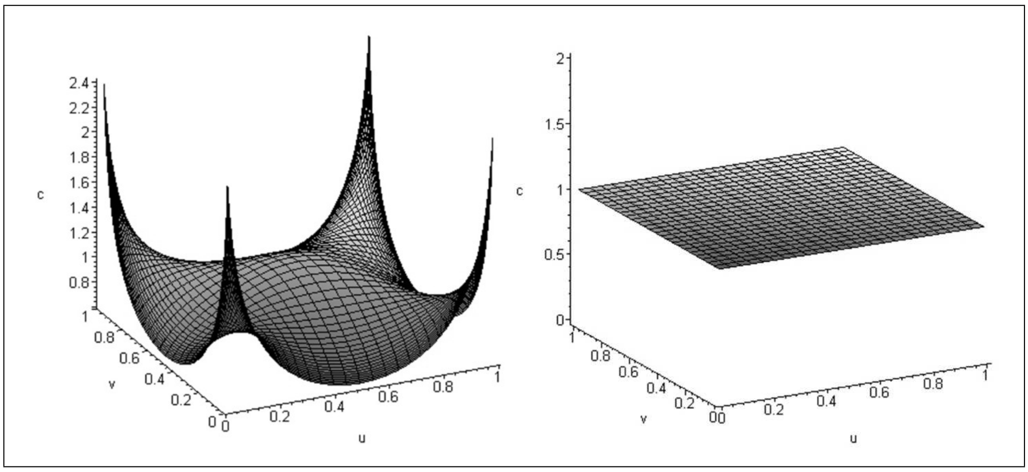


Fig. 3. Copula-density function $c(u,v)$ of an $MGH_2(0,I,\omega)$ distribution (left figure) with parameters $\lambda=1,\ \alpha=1$ and $\beta=(0,0)'$ and that of an $MAGH_2(0,I,\omega)$ distribution (right figure) with arbitrary parameter constellation.

Theorem 4 (Limiting cases) *Let $\Sigma^{(m)} := (\sigma_{ij}^{(m)})_{i,j=1,2}$ be a sequence of symmetric positive-definite matrices and $\rho^{(m)} := \sigma_{12}^{(m)}/\sqrt{\sigma_{11}^{(m)}\sigma_{22}^{(m)}}$. Suppose that $X^{(m)} \in MGH_2(\mu,\Sigma^{(m)},\omega)$ or $X^{(m)} \in MAGH_2(\mu,\Sigma^{(m)},\omega)$ for every $m \in I\!N$, and let $C^{(m)}$ denote the corresponding copula. Denote with $W(u_1,u_2)=\max(u_1+u_2-1,0)$ and $M(u_1,u_2)=\min(u_1,u_2)$ the well-known lower and upper Frechet copula bounds and let $\Pi(u_1,u_2)=u_1\cdot u_2$ be the product or independence copula. Then*

i) $C^{(m)} \to M$ *pointwise if* $\rho^{(m)} \to 1,\ \sigma_{ij}^{(m)} \to \sigma_{ij} \neq 0,\ i,j=1,2,$ *as* $m \to \infty$,

ii) $C^{(m)} \to W$ *pointwise if* $\rho^{(m)} \to -1,\ \sigma_{ij}^{(m)} \to \sigma_{ij} \neq 0,\ i,j=1,2,$ *as* $m \to \infty$,

iii) if $X^{(m)} \in MGH_2(\mu,\Sigma^{(m)},\omega)$, *then* $C^{(m)} \neq \Pi$ *for each parameter constellation, and*

iv) if $X^{(m)} \in MAGH_2(\mu,\Sigma^{(m)},\omega)$, *then* $C^{(m)} = \Pi$ *if and only if* $\Sigma^{(m)} = I$.

Proof. i)+ii) Consider the Cholesky decomposition of $\Sigma^{(m)}$ and use the fact that the copula C of X is invariant under strictly increasing transformations of the margins since X possesses continuous marginal distribution functions. iii) Suppose $X^{(m)} \in MGH_2(\mu,\Sigma^{(m)},\omega)$. Then $X^{(m)}$ possesses the product copula if and only if it has independent margins (see Theorem 2.4.2 in Nelsen (1999)). According to Definition 1, $X^{(m)}$ does not have independent margins if $\Sigma^{(m)}$ is not a diagonal matrix. Thus it suffices to consider $\Sigma^{(m)}=I$. Further we can put $\beta=0$ since β has no influence on the factorization of the density function of an MGH distribution (see formula (1)). However, in that case $X^{(m)}$ belongs to the family of elliptically contoured distributions. Therefore Theorem 4.11 in Fang, Kotz, and Ng (1990) implies that $X^{(m)}$ possesses independent margins if and only if $X^{(m)}$ has a bivariate normal distribution. Since normal distributions and MGH distributions are disjoint classes of distributions, the assertion follows. Part iv) follows with the definition of MAGH distributions. □

Remark. The results of Theorem 4 can be extended to n-dimensional MGH and MAGH distributions. However, for $n \geq 3$ the lower Fréchet bound is not a copula function anymore; see Theorem 2.10.13 in Nelsen (1999) for an interpretation of the lower Fréchet bound in that case.

The covariance and correlation matrix. Among the large number of dependence measures for multivariate random vectors the covariance and the correlation matrix are still the most favorite ones in most practical applications. However, according to Embrechts, McNeil, and Straumann (2002), these dependence measures should be considered with care for non-elliptically contoured distributions such as MAGH distributions.

Theorem 5 (Mean and covariance for MGH distributions)
Let $X \in MGH_n(\mu, \Sigma, \omega)$ and define $R_{\lambda,i}(x) := \frac{K_{\lambda+i}(x)}{x^i K_\lambda(x)}$. Then the mean vector and the covariance matrix of X are given by

$$E[X] = \mu + \alpha R_{\lambda,1}\Big(\sqrt{\alpha^2(1-\beta'\beta)}\Big)\, A'\beta \quad \text{and}$$

$$Cov[X] = R_{\lambda,1}\Big(\sqrt{\alpha^2(1-\beta'\beta)}\Big)\, \Sigma + \Big[R_{\lambda,2}\Big(\sqrt{\alpha^2(1-\beta'\beta)}\Big) - R^2_{\lambda,1}\Big(\sqrt{\alpha^2(1-\beta'\beta)}\Big)\Big]\, \frac{A'\beta\beta' A}{1-\beta'\beta}.$$

For the symmetric case $\beta = (0, \ldots, 0)'$ and $\lambda = 1$, the mean vector and the covariance matrix of X simplify to $E[X] = 0$ and $Cov[X] = K_2(\alpha)/(\alpha K_1(\alpha)) \cdot \Sigma$.

Proof. Let $X \in MGH_n(\bar{\mu}, \bar{\Sigma}, \bar{\omega})$ with parameter representation as in (3). Then X is distributed according to a variance-mean mixture of a multivariate normal distribution, i.e., $X|(Z = z) \sim N(\bar{\mu} + z\bar{\Sigma}\bar{\beta}, z\bar{\Sigma})$, where the mixing random variable Z is distributed according to a generalized inverse Gaussian distribution $GIG(\bar{\lambda}, \bar{\delta}, \sqrt{\bar{\delta}^2(\bar{\alpha}^2 - \bar{\beta}'\bar{\Sigma}\bar{\beta})})$ (see e.g. Barndorff-Nielsen, Kent, and Sørensen (1982)). The respective mean vector and covariance can be calculated via this representation (see Eberlein and Prause (2002) for more details) utilizing the parameter mapping $\bar{\delta}^2(\bar{\alpha}^2 - \bar{\beta}'\bar{\Sigma}\bar{\beta}) = \alpha^2(1-\beta'\beta)$, $\bar{\Sigma}\bar{\beta}\bar{\delta}^2 = \Sigma\alpha A^{-1}\beta$ and $\bar{\delta}^4\bar{\Sigma}\bar{\beta}\bar{\beta}'\bar{\Sigma} = A'\beta\beta' A$. □

Theorem 6 (Mean and covariance for MAGH distributions)
Let $X \in MAGH_n(\mu, \Sigma, \omega)$. Then the mean vector and the covariance matrix are given by

$$E[X] = A'e_Y + \mu \quad \text{and} \quad Cov[X] = A'CA,$$

where $e_Y = (E[Y_1], \ldots, E[Y_n])'$ with $E[Y_i] = R_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)})\alpha_i\beta_i$ and $C = diag(c_{11}, \ldots, c_{nn})$ with

$$c_{ii} = R_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)}) + \Big[R_{\lambda_i,2}(\sqrt{\alpha_i^2(1-\beta_i^2)}) - R^2_{\lambda_i,1}(\sqrt{\alpha_i^2(1-\beta_i^2)})\Big]\frac{\beta_i^2}{1-\beta_i^2}.$$

The covariance matrix $Cov[X]$ is proportional to Σ if $\alpha = \alpha_i$, $\beta = \beta_i$ and $\lambda = \lambda_i$ for all $i = 1, \ldots, n$.

Proof. The assertion follows immediately from Theorem 5 and Definition 2. □

Kendall's tau. The correlation coefficient is a measure of linear dependence between two random variables and therefore it is not invariant under monotone increasing transformations. However, not only does "scale-invariance" present an undisputable requirement for a proper dependence measure in general (cf. Joe (1997), Chapter 5), but also in practice "scale-invariant" dependence measures play an increasing role in dependence modelling. Kendall's tau is the most famous one and therefore we determine it for MGH and MAGH distributions.

Definition 7 (Kendall's tau) *Let $X = (X_1, X_2)'$ and $\bar{X} = (\bar{X}_1, \bar{X}_2)'$ be independent bivariate random vectors with common continuous distribution function F and copula C. Kendall's tau is defined by*

$$\tau = \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) > 0) - \mathbb{P}((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) < 0) = 4 \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1.$$

Theorem 8 *Let $\Sigma = (\sigma_{ij})_{i,j=1,2}$ be a positive-definite matrix and $\rho := \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$.*
i) If $X \in MGH_2(\mu, \Sigma, \omega)$ with $\beta = 0$, then $\tau = \frac{2}{\pi}\arcsin(\rho)$.
ii) If $X \in MAGH_2(\mu, \Sigma, \omega)$ with stochastic representation $X \stackrel{d}{=} A'Y + \mu$, $A'A = \Sigma$, then for $\rho \neq 0$

$$\tau = \frac{4}{|c|} \int_{\mathbb{R}^2} f_{Y_1}(x_1) f_{Y_2}\left(\frac{x_2 - x_1}{c}\right) \cdot \int_{-\infty}^{x_1} F_{Y_2}\left(\frac{x_2 - z}{c}\right) f_{Y_1}(z)\, dz d(x_1, x_2) - 1, \tag{4}$$

where $c := sgn(\rho)\sqrt{1/\rho^2 - 1}$. Further $\tau = 0$ for $\rho = 0$.

The *proof* is given in the Appendix.

Remark. An explicit expression of Kendall's tau for MAGH distributions (given by (4)) cannot be expected. However, since the density functions of Y_1 and Y_2 are available, formula (4) yields a tractable numerical solution. Further, the values of Kendall's tau and the correlation coefficient for the class of bivariate MGH and MAGH distributions cover the entire interval $[-1, 1]$. For Kendall's tau, this can be seen from part i) and ii) in Theorem 4 (note that Kendall's tau is a measure of concordance). For the correlation coefficient of an MGH distribution, Theorem 5 implies that the parameter $\rho \in [-1, 1]$ corresponds to the correlation coefficient if $\beta = 0$. The conditions where ρ equals the correlation coefficient for an MAGH distribution are stated in Theorem 6.

Tail dependence. A strong emphasis in this paper is put on the dependence structure of MGH and MAGH distributions. In this context we establish the following result about the dependence structure of extreme events (tail dependence or extremal dependence) related to the latter types of distributions. The importance of tail dependence especially in financial risk management is addressed in Hauksson et al. (2001) and Embrechts, Lindskog, and McNeil (2003). The following definition (according to Joe (1997), p. 33) represents one of many possible definitions of tail dependence.

Definition 9 (Tail dependence) *Let $X = (X_1, X_2)'$ be a 2-dimensional random vector. We say that X is upper tail-dependent if*

$$\lambda_U := \lim_{u \to 1^-} \mathbb{P}(X_1 > F_1^{-1}(u) \mid X_2 > F_2^{-1}(u)) > 0, \tag{5}$$

where the limit is assumed to exist and F_1^{-1}, F_2^{-1} denote the generalized inverse distribution functions of X_1, X_2, respectively. Consequently, we say that X is upper tail-independent if λ_U equals 0. Similarly, X is said to be lower tail-dependent (lower tail-independent) if $\lambda_L > 0$ ($\lambda_L = 0$), where $\lambda_L := \lim_{u \to 0^+} \mathbb{P}(X_1 \leq F_1^{-1}(u) \mid X_2 \leq F_2^{-1}(u))$ provided they exist.

Figure 3 reveals that bivariate standardized MGH distributions show more evidence of dependence in the upper-right and lower-left quadrant of its distribution function than MAGH distributions. However, the following theorem shows that MGH distributions are always tail independent whereas non-standardized MAGH distributions can even model tail dependence. For the sake of simplicity we restrict ourselves to the symmetric case $\beta = 0$.

Theorem 10 *Let $\Sigma = (\sigma_{ij})_{i,j=1,2}$ be a positive-definite matrix and $\rho := \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$. Suppose $\beta = 0$. Then*
i) the $MGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-independent,
ii) the $MAGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-independent if
$\alpha_2 < \alpha_1\sqrt{1/\rho^2 - 1}$ or $\rho \leq 0$, and

iii) the $MAGH_2(\mu, \Sigma, \omega)$ distributions are upper and lower tail-dependent if $\alpha_2 > \alpha_1\sqrt{1/\rho^2 - 1}$ and $\rho > 0$.

The *proof* is given in the Appendix.

Remark. Additionally to Theorem 10 it can be shown that an $MGH_2(\mu, \Sigma, \omega)$ distribution (with $\beta = 0$) is tail independent if $\alpha_2 = \alpha_1\sqrt{1/\rho^2 - 1}$ and $\lambda_2 < \lambda_1$.

Other properties. It has been already mentioned that the family of MGH distributions is closed with respect to the marginal distributions. Similarly, if $X = (X_1, \ldots, X_n)'$ is an MAGH-distributed random vector, then the partitions $X_{(1)} = (X_1, \ldots, X_k)'$ and $X_{(2)} = (X_{k+1}, \ldots, X_n)'$ are MAGH distributed. Further, the conditional distribution of $X_{(2)}$ given $X_{(1)}$ and regular affine-linear transformations of X are MAGH, too. For general partitions of X and singular affine-linear transformations, the latter properties hold also if the definition of an MAGH distribution (see Definition 2) allows for the stochastic representation $X = A'Y + \mu$, where Y is not necessarily n-dimensional. However, in this case the stochastic representation is not unique anymore (which is comparable to elliptically contoured distributions). We may also allow for general decompositions of the scaling matrix $\Sigma = A'A$ in Definition 2 in order to remain in the class of MAGH distributions after a permutation of the margins of X (cf. the discussion in Section 3).

6 Parameter estimation

Following a brief setup of the estimation procedure, we show that the parameters of the MAGH distribution can be estimated with a simple two-stage method. This method refers to the estimation of the covariance matrix and the parameters which correspond to the univariate marginal distributions. Thus, generally speaking, the estimation reduces to a one-dimensional estimation problem. In contrast, for the MGH distributions one can only perform this type of estimation within the subclass of elliptically contoured distributions.

Minimizing the cross entropy. A common descriptive statistics to measure the similarity between two (multivariate) distribution or density functions f^* and f, respectively, is given by the (directed) *Kullback divergence* (see Kullback (1959) and Ullah (1996)) which is defined as

$$H_K(f, f^*) := \int_{\mathbb{R}^n} f^*(x) \log \frac{f^*(x)}{f(x)} dx = \int_{\mathbb{R}^n} f^*(x) \log f^*(x) dx - \int_{\mathbb{R}^n} f^*(x) \log f(x) dx.$$

The Kullback divergence H_K is zero if and only if the densities f and f^* coincide. It is also additive across the marginal densities if the marginal distributions are stochastically independent. This is one desired property of a pseudo-distance measure for multidimensional random vectors. In our context, we use the Kullback divergence to measure the similarity between the "true" density f^* and its approximation f. For example, minimizing the Kullback divergence by varying f is a common way to find a good approximation (or good fit) of the "true" density f^*. This kind of descriptive statistics is frequently used if common goodness-of-fit tests such as the χ^2-square test turn out to be too complicated. The latter is often the case for high-dimensional distributions. In Sections 8 and 9 we investigate the problem of goodness-of-fit for simulated and real data by means of Kullback divergence and χ^2-square tests.

Note that the first term of H_K is constant and can be dropped. The resulting expression is called *cross entropy.* Because f^* is unknown, we approximate it by its empirical counterpart. In this case, minimizing the cross entropy (or Kullback divergence) coincides with the concept of maximum likelihood.

For computational reasons it may be advantageous to work with the standardized vector $y = B(x-\mu)$ where $B := A'^{-1}$. The density transformation theorem yields that the density function f_X of $X \stackrel{d}{=} A'Y + \mu \in MGH(\mu, \Sigma, \omega)$ is given by $f_X(x) = f_{A'Y+\mu}(x) = |B| f_Y(y)$, with $|B| > 0$ being the determinant of B and parameters $\eta = (\mu, B, \omega)$ and $\omega = (\lambda, \alpha, \beta)$.

While the location vector μ can take arbitrary values in $\mathbb{R}^n$, B and ω are subject to various constraints. The matrix B must be triangular with positive diagonal such that $A'A$ is positive-definite. The parameter α is supposed to be positive and the vector β must fulfill $||\beta||_2 \leq 1$.

Many approaches are possible regarding the latter optimization problem, inter alia we mention two methods:

- constrained nonlinear optimization methods, and
- unconstrained nonlinear optimization methods after suitable parameter transformations.

We prefer the unconstrained approach due to reasons of robustness. The following parameter transformations are appropriate. The matrix B can be sought of the form $B = UD$ where D is some diagonal matrix having strictly positive elements and U is a triangular matrix having only ones on its diagonal. In order to enforce the strict positivity of the diagonal elements $d_{ii}, i = 1, \ldots, n$, of D, the following transformations are applied:

$$b_{ii} = d_{ii} = e^{\nu_i}, \quad i = 1, \ldots, n,$$
$$b_{ij} = d_{ii} u_{ij} = e^{\nu_i} u_{ij}, \quad i = 1, \ldots, n-1, \quad j = 2, \ldots, n, \quad j > i,$$

with unknown parameters ν_i and u_{ij}. The parameter α is estimated via the same exponential map. For the vector β we utilize the smooth transformation

$$\beta = \gamma \cdot \frac{1}{1 + \exp(-\|\gamma\|_2)} \cdot \frac{1}{\|\gamma\|_2}. \tag{6}$$

The optimization algorithm we use belongs to the probabilistic ones (note that the objective function is non-convex) and consists of two characteristic phases:

1. In a *global phase* the objective function is evaluated at a random number of points being part of the search space.
2. In a *local phase* the samples are transformed to become candidates for local optimization routines.

Results concerning the convergence of probabilistic algorithms to the global minimum are, for example, given in Rinnooy Kan and Timmer (1987). The latter reference motivates us to use a Multi-Level Single-Linkage procedure. Here the global optimization method generates random samples from the search space by identifying the samples belonging to the same objective function attractor and eliminating multiple ones. For the remaining samples in the local phase a conjugate gradient method (see Fletcher (1987)) is started. Further, a Bayesian stopping rule is applied in order to assess the probability that all attractors have been explored. For an account on the Bayesian stopping rule we refer the reader to Boender and Rinnooy Kan (1987).

Estimation of MAGH distributions. A big advantage of the MAGH distribution is that its parameters can be easily identified in a two-stage procedure comprising the following steps:

(1) Compute the sample covariance matrix S of the random vector X. Transform X to the vector $Y = BX$ with independent margins. The matrix B is received via Cholesky decomposition $S^{-1} = B'B$.

(2) Identify the parameters λ_i, α_i and β_i which belong to the univariate marginal distributions of Y. The location vector μ can be received via $B^{-1}e$ with e_i being the location parameter of Y_i (see Theorem 5). The scaling matrix Σ equals $B^{-1'}DB^{-1}$ where the diagonal matrix D is determined by the scaling parameters of Y_i on its diagonal.

The latter procedure considerably simplifies the complexity of the numerical optimization. Note that the parameters of the one-dimensional MAGH distributions can be estimated via unconstrained optimization as explained above. We remark that the two-stage estimation may affect the (asymptotic) efficiency of the estimation. However, a theoretical analysis of the loss of efficiency goes beyond the scope of this article. Though, our empirical results show that this two-stage estimation is quite robust with respect to the finite-sample volatility of the corresponding estimators.

It is important for applications that the univariate densities are not necessarily identically parameterized. This means that the margins may have different parameters $\lambda_i, \alpha_i, \beta_i,\ i = 1, \ldots, n$. In other words, there is a considerable freedom of choosing the parameters λ, α, and β. Therefore, in addition to a parameterization similar to that for MGH distributions (i.e., the same λ and α for all one-dimensional margins and different β_i) two further extreme alternatives are possible:

- Minimum parameterization: Equal parameters λ, α and β for all one-dimensional margins.
- Maximum parameterization: Individual parameters λ_i, α_i and β_i for all one-dimensional margins.

The appropriate parameterization depends on the kind of application and the available data volume. The optimization procedure presented in this section is a special case of an identification-algorithm for conditional distributions explored in Stützle and Hrycej (2001, 2002a, 2002b).

7 Sampling from MGH and MAGH distributions

Complementing the question of estimation as described in Section 6, we provide now an efficient and self-contained generator of random vectors for the families of MGH and MAGH distributions. The generator, which is based on a rejection algorithm, comprises several features which to our knowledge have not been published yet. In particular, we provide an explicit algorithm, while avoiding difficult minimization routines, which outperforms the known algorithms in the literature for MGH distributions with general parameter constellations. In principle, the generation of random vectors reduces to the generation of one-dimensional random numbers. For

$MGH_n(\mu, \Sigma, \omega)$ distributions this is possible via the following variance-mean mixture representation. Let the random variable Z be distributed according to a generalized inverse Gaussian distribution with parameters λ, χ, and ψ. In particular the latter family is referred to as the $GIG(\lambda, \chi, \psi)$ distributions. Then, $X \in MGH_n(\mu, \Sigma, \omega)$ is conditionally normally distributed with mixing random variable Z, i.e., $X|(Z = z) \sim N_n(\mu + z\tilde{\beta}, z\Delta)$, where $\Delta \in \mathbb{R}^{n\times n}$ is a symmetric positive-definite matrix with determinant $|\Delta| = 1$ and $\mu, \tilde{\beta} \in \mathbb{R}^n$. The parameters Σ and $\omega = (\lambda, \alpha, \beta)$ are given by

$$\alpha = \sqrt{(\psi + \tilde{\beta}'\Delta\tilde{\beta})\chi},\ \beta = 1/\sqrt{\psi + \tilde{\beta}'\Delta\tilde{\beta}}L\tilde{\beta},\ \Sigma = \chi \cdot \Delta,$$

with Cholesky decomposition $L'L = \Delta$. The inverse map is given by

$$\chi = |\Sigma|^{1/n},\ \psi = \alpha^2/|\Sigma|^{1/n} \cdot (1 - \beta'\beta),\ \Delta = \Sigma/|\Sigma|^{1/n}, \text{ and } \tilde{\beta} = \alpha \cdot (A)^{-1}\beta,$$

with Cholesky decomposition $A'A = \Sigma$.

The sampling algorithm is now of the following form: A pseudo random number is sampled from a random variable Z having a generalized inverse Gaussian distribution with parameters λ, χ, and ψ. Then an n-dimensional random vector X being conditionally normally distributed with mean vector $\mu + Z\Delta\tilde{\beta}$ ("drift") and covariance matrix $Z\Delta$ (determinant $|\Delta| = 1$) is generated.

The density function of the generalized inverse Gaussian distribution $GIG(\lambda, \chi, \psi)$, is given by

$$f_Z(x) = c \cdot x^{\lambda-1} \exp\Big(-\frac{\chi}{2x} - \frac{\psi x}{2}\Big), \quad x > 0, \tag{7}$$

with normalizing constant $c = (\psi/\chi)^{\lambda/2}/(2K_\lambda(\sqrt{\psi\chi}))$. The range of the parameters is given by

i) $\chi > 0,\ \psi \geq 0$ if $\lambda < 0$, or

ii) $\chi > 0,\ \psi > 0$ if $\lambda = 0$, or

iii) $\chi \geq 0,\ \psi > 0$ if $\lambda > 0$.

The following algorithm is formulated for generalized hyperbolic distributions $MGH_n(\mu, \Sigma, \omega)$ with parameter $\lambda = 1$. Section 8.1 justifies the restriction to this class of distributions. In this context the $GIG(\lambda, \chi, \psi)$ distribution is referred to as inverse Gaussian distribution. However, the algorithm can be extended to general λ. Our empirical study shows that the algorithm outperforms the efficiency of the sampling algorithm proposed by Atkinson (1982) which suites to a larger class of distributions (see also Prause (1999), Section 4.6). Moreover, the algorithm avoids tedious minimization routines and time-consuming evaluations of the Bessel function K_λ. The generation utilizes a rejection method (see Ross (1997), pp. 565) with a three part rejection-envelop. We define the envelop $d : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$d(x) := \begin{cases} d_1(x) = ca_1 \exp(b_1 x), & \text{if } 0 < x < x_1, \\ d_2(x) = ca_2, & \text{if } x_1 \leq x < x_2, \\ d_3(x) = ca_3 \exp(-b_3 x), & \text{if } x_2 \leq x < \infty, \end{cases} \tag{8}$$

with $a_i > 0,\ i = 1, \ldots, 3,\ b_i > 0,\ i = 1, 3$, and $x_1 \leq x_2 \leq x_3$ to be defined later. Let $z_i,\ i = 1, 3$ denote the inflection points and $z_2 = \sqrt{\chi/\psi}$ denote the mode of the unimodal density f_Z. Further we require $d_1(z_1) = f_Z(z_1),\ d_2(z_2) = f_Z(z_2),\ d_3(z_3) = f_Z(z_3)$. The points $x_1 > 0$ and $x_2 > 0$ correspond to the intersection points of d_1, d_2 and d_2, d_3, respectively, i.e. $d_1(x_1) = d_2(x_1),\ d_2(x_2) = d_3(x_2)$.

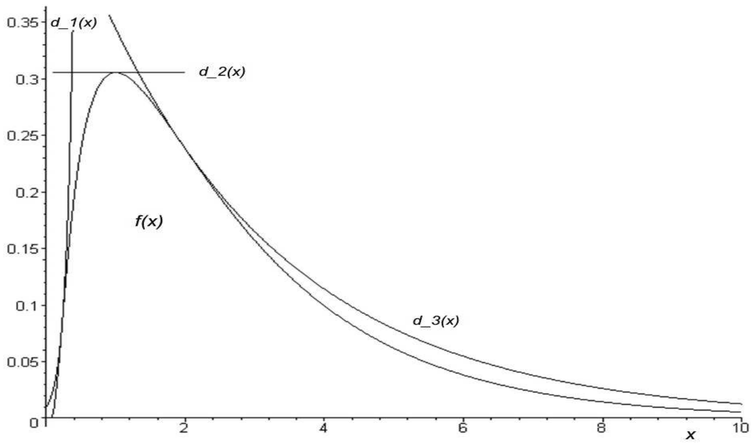


Fig. 4. Three part envelop d for the inverse Gaussian density function f_Z with parameters $\chi = 1$, $\psi = 1$.

Primarily, the rejection method requires the generation of random numbers with density $s \cdot d(x)$ where the scaling factor s has to be computed in order to obtain a density function $s \cdot d(x), x > 0$. This scaling factor is derived below.

Pseudo algorithm for generating an inverse Gaussian random number:

(1) Compute the zeros z_1, z_2 for $\psi^2 z^4 - 2\chi\psi z^2 - 4\chi z + \chi^2 = 0$.

(2) Set $b_1 = (\chi/z_1^2 - \psi)/2$ and $a_1 = \exp(-\chi/z_1)$.

(3) Set $a_2 = \exp(-\sqrt{\chi\psi})$.

(4) If $(\psi - \chi/z_2^2)/2 > 0$ then
Set $b_3 = (\psi - \chi/z_2^2)/2$ and $a_3 = \exp(-\chi/z_2)$.
Else Set $b_3 = \psi/2$ and $a_3 = 1$.

(5) Set $x_1 = \ln(a_2/a_1)/b_1$ and $x_2 = -\ln(a_2/a_3)/b_3$.

(6) Set
$$s = \Big(\frac{a_1}{b_1}\exp(b_1 x_1) - \frac{a_1}{b_1} + (x_2 - x_1)a_2 + \frac{a_3}{b_3}\exp(-b_3 x_2)\Big).$$

(7) Set
$$k_1 = \frac{1}{s}\Big(\frac{a_1}{b_1}\exp(b_1 x_1) - \frac{a_1}{b_1}\Big) \quad \text{and} \quad k_2 = k_1 + \frac{1}{s}(x_2 - x_1)a_2.$$

(8) Generate independent and uniformly distributed random numbers U and V on the interval $[0, 1]$.

(9) If $U \le k_1$ goto step 10.
ElseIf $k_1 < U \le k_2$ goto step 11.
Else goto step 12.

(10) Set
$$x = \frac{1}{b_1}\ln\big(\frac{b_1}{a_1} sU + 1\big).$$

If

$$V \le \frac{f_Z(x)}{d_1(x)} = \frac{1}{a_1}\exp\Big(-(\frac{\chi x^{-1}+\psi x}{2}+b_1 x)\Big),$$

Then Return x

Else goto step 8.

(11) Set

$$x = \frac{sU}{a_2} - \frac{a_1}{b_1 a_2}\Big(\exp(b_1 x_1) - 1\Big) + x_1.$$

If

$$V \le \frac{f_Z(x)}{d_2(x)} = \frac{1}{a_2}\exp\Big(-(\frac{\chi x^{-1}+\psi x}{2})\Big),$$

Then Return x.

Else goto step 8.

(12) Set

$$x = -\frac{1}{b_3}\ln\Big[-\frac{b_3}{a_3}\Big\{sU - \frac{a_1}{b_1}(e^{b_1 x_1}-1) - (x_2-x_1)a_2 - \frac{a_3}{b_3}e^{-b_3 x_2}\Big\}\Big].$$

If

$$V \le \frac{f_Z(x)}{d_3(x)} = \frac{1}{a_3}\exp\Big(-(\frac{\chi x^{-1}+\psi x}{2}) + b_3 x\Big),$$

Then Return x.

Else goto step 8.

Remark. In order to generate a sequence of inverse Gaussian random numbers repeat step 8.

So far we have generated random numbers from an univariate inverse Gaussian distribution. We turn now to the generation of multivariate generalized hyperbolic random vectors. For this we exploit the above introduced mixture representation.

Pseudo algorithm for generating an MGH vector:

(1) Set $\Delta = L'L$ via Cholesky decomposition.

(2) Generate an inverse Gaussian random number Z with parameters χ and ψ.

(3) Generate a standard normal random vector N.

(4) Return $X = \mu + Z\Delta\tilde{\beta} + \sqrt{Z}L'N$.

Pseudo algorithm for generating an MAGH vector:

(1) Set $\Sigma = A'A$ via Cholesky decomposition.

(2) Generate a random vector Y with independent $MGH_1(0,1,\omega_i),\ i = 1,\dots,n,$ distributed components (see above).

(3) Return $X = \mu + A'Y$.

Table 1 presents the empirical efficiency of the MGH random vector generator for various parameter constellations. In our framework, efficiency is defined by the following ratio

$$\text{Efficiency} = \frac{\sharp \text{ of generated samples}}{\sharp \text{ of algorithm-passes including rejections}}.$$

χ/ψ	0.1	0.5	1	2	5	10
0.1	0.94	0.913	0.904	0.886	0.877	0.872
0.5	0.916	0.884	0.877	0.865	0.859	0.852
1	0.901	0.877	0.867	0.863	0.857	0.851
2	0.889	0.866	0.857	0.856	0.856	0.845
5	0.876	0.858	0.854	0.852	0.847	0.849
10	0.866	0.86	0.851	0.853	0.842	0.847

Table 1
Empirical efficiency of the MGH random number generator for $\lambda = 1$ and $10,000$ generated samples.

8 Simulation and empirical study

A series of computational experiments with simulated data is performed in this section. The experiments disclose that

- MGH density functions with parameter $\lambda \in I\!R$ seem to have very close counterparts in the MGH-subclass with parameter $\lambda = 1$ (in short: MH distribution),
- The MAGH distribution can closely approximate the MGH distribution with similar parameter values.

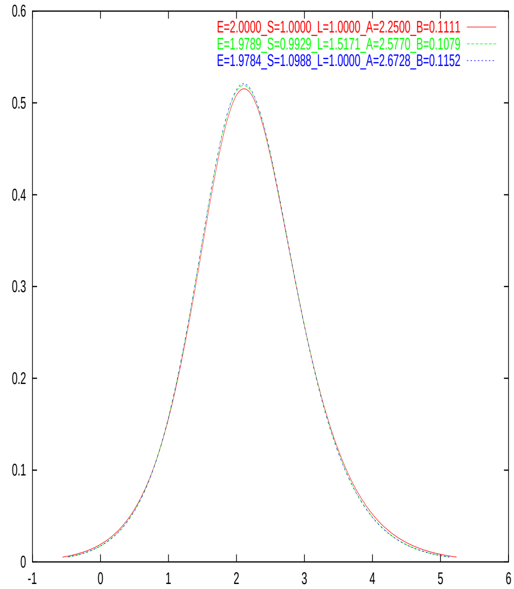
8.1 General MGH distributions versus MGH distributions with parameter $\lambda = 1$ *(MH)*

Figure 5 shows the identification results of two univariate MGH distributions and two univariate MH distributions (i.e., $\lambda = 1$). All four plots simultaneously show the respective fit via MGH distribution (bright dotted line) and via MH distribution (dark dotted line). In both plots on the left side of the figure, samples were drawn from an MH distribution. The fit illustrates the phenomenon that although the identification procedure with MGH densities frequently produces $\lambda \neq 1$, the approximation of the original MH density function remains good. The plots on the right side illustrate the opposite case, namely, a good approximation of the MGH density function ($\lambda \neq 1$) via an MH density function ($\lambda = 1$).

Consider now a bivariate $MGH_2(\mu, \Sigma, \omega)$ distribution with $\Sigma = (\sigma_{ij})_{i,j=1,2}$. The mutual tradeoff between λ, α, and the scaling parameters $S_1 := \sqrt{\sigma_{11}}$ and $S_2 := \sqrt{\sigma_{22}}$ of the corresponding distribution function is shown in Table 2. While all samples were drawn from an MGH distribution with parameter $\lambda = 1$ (MH distrubution), MGH identification usually leads to an overestimation of λ which is traded off by lower values of α, S_1, and S_2. In contrast to that, the parameter identifications via MH distributions are close to the reference values. However, the differences between the cross entropies are hardly discernible, showing that both parameter combinations correspond to densities which are close to each other.

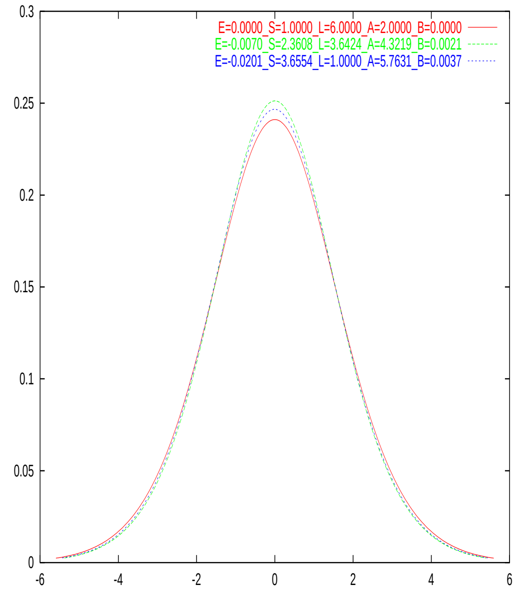
The following conclusions can be drawn:

- The fit of both distributions, MGH and MH distribution, measured by cross entropy and visual closeness of the density plots, is satisfying. This implies the existence of multiple parameter



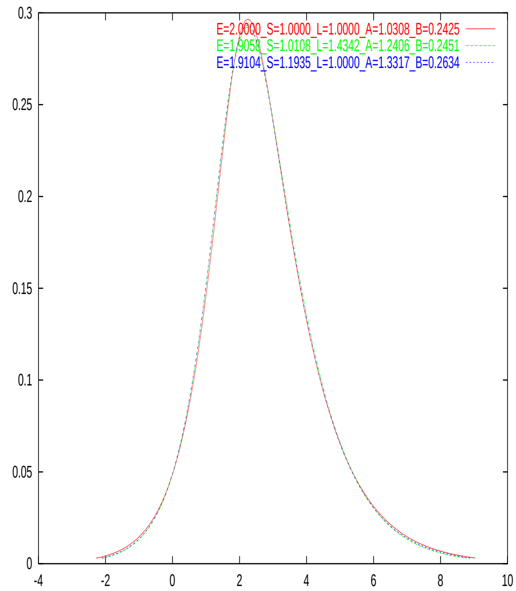
$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(2.00, 1.00, 1.00, 2.25, 0.11)'$ (solid line)

$(m(\hat{\mu}), m(\hat{\sqrt{\Sigma}}), m(\hat{\lambda}), m(\hat{\alpha}), m(\hat{\beta}))' =$
$(1.98, 0.99, 1.51, 2.58, 0.11)'$ (bright dotted line)
$(1.98, 1.10, 1.00, 2.67, 0.12)'$ (dark dotted line)



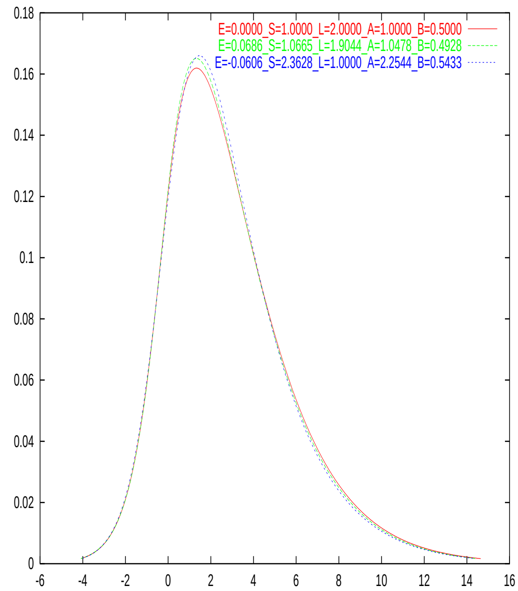
$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(0.00, 1.00, 6.00, 2.00, 0.00)'$ (solid line)

$(m(\hat{\mu}), m(\hat{\sqrt{\Sigma}}), m(\hat{\lambda}), m(\hat{\alpha}), m(\hat{\beta}))' =$
$(0.01, 2.36, 3.64, 4.32, 0.00)'$ (bright dotted line)
$(0.02, 3.66, 1.00, 5.76, 0.00)'$ (dark dotted line)



$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(2.00, 1.00, 1.00, 1.03, 0.24)'$ (solid line)

$(m(\hat{\mu}), m(\hat{\sqrt{\Sigma}}), m(\hat{\lambda}), m(\hat{\alpha}), m(\hat{\beta}))' =$
$(1.91, 1.01, 1.43, 1.24, 0.24)'$ (bright dotted line)
$(1.91, 1.19, 1.00, 1.33, 0.26)'$ (dark dotted line)



$(\mu, \sqrt{\Sigma}, \lambda, \alpha, \beta)' =$
$(0.00, 1.00, 2.00, 1.00, 0.50)'$ (solid line)

$(m(\hat{\mu}), m(\hat{\sqrt{\Sigma}}), m(\hat{\lambda}), m(\hat{\alpha}), m(\hat{\beta}))' =$
$(0.07, 1.07, 1.90, 1.05, 0.49)'$ (bright dotted line)
$(0.06, 2.36, 1.00, 2.25, 0.54)'$ (dark dotted line)

Fig. 5. Univariate MGH and MH distributions. Reference densities (solid lines) and identified MGH and MH densities (bright and dark dotted lines) averaged from 100 random samples for various parameter constellations. The values $m(\cdot)$ denote the sample mean of the respective parameters.

constellations for MGH distributions which lead to quite similar density functions. Similar results have been observed for the corresponding tail functions.

- Generalized hyperbolic densities seem to have very close counterparts in the class of MH distributions (even for large λ). Therefore, the class of MH distributions will be sufficiently rich for our considerations.

In view of the above results, only MH distributions and the corresponding MAH distributions (multivariate affine generalized hyperbolic distributions with $\lambda = 1$) will be considered in the next section.

8.2 MH distributions versus MAH distributions

Parameter estimates are compared for the following three classes of bivariate distributions:

(1) MH distributions.

(2) MAH distributions with minimal parameter configuration (same value of α and β for each margin).

(3) MAH distributions with maximal parameter configuration (different values of α_i and β_i for each margin).

All types of distributions in Table 3 have been identified from data sampled from an MH distribution. The two-stage algorithm introduced in Section 6 has been used for the identification of the MAHmax model (determining first the sample correlation matrix, then transforming the variables, and finally identifying the univariate distributions). The identification results are provided in Table 3.

The following conclusions can be drawn:

- For all three models, most parameters show an acceptable fit regarding the sample bias and the sample standard deviation. The relative variability of the estimates increases with decreasing α (fatter tailed distributions). Such fatter tailed distributions seem to be more ill-posed with respect to the estimation of individual parameters.

- The differences between the parameter estimates obtained either for the MH, the MAHmin, or the MAHmax distribution are negligible (although the data are drawn from an MH distribution).

- The fit in terms of the cross entropy does not differ significantly between the various models. As expected, the MAHmax estimates are closer to the MH reference distribution than are the MAHmin estimates in terms of the cross entropy (Note that in one case they are even better than the MH estimates). The fitting capability of the MAHmax model comes at the expense of a larger variability and a sometimes larger bias (”overlearning effect”).

Table 2
Bivariate MGH and MH distribution. For each parameter λ, α, $S_1 = \sqrt{\sigma_{11}}$, and $S_2 = \sqrt{\sigma_{22}}$ the table lists the reference value, the sample mean $m(\cdot)$, and the sample standard deviation $\sigma(\cdot)$ of the parameter estimations derived from 100 samples of sample-size 1000 each. In the last column, the sample mean and the sample standard deviation of the corresponding cross entropy H are provided.

	λ	$m(\hat{\lambda})$	$\sigma(\hat{\lambda})$	α	$m(\hat{\alpha})$	$\sigma(\hat{\alpha})$	S_1	$m(\hat{S}_1)$	$\sigma(\hat{S}_1)$	S_2	$m(\hat{S}_2)$	$\sigma(\hat{S}_2)$	$m(\hat{H})$	$\sigma(\hat{H})$
MGH	1.0000	1.1061	0.0571	0.3200	0.2724	0.0867	0.3200	0.2574	0.0735	0.3200	0.2595	0.0746	3.4525	0.0438
MH	1.0000	1.0000	0.0000	0.3200	0.3286	0.1228	0.3200	0.3229	0.1078	0.3200	0.3243	0.1119	3.4177	0.3450
MGH	1.0000	1.3745	0.1651	1.1456	0.8758	0.2085	1.0000	0.7096	0.1341	1.0000	0.7141	0.1343	3.8683	0.0455
MH	1.0000	1.0000	0.0000	1.1456	1.1609	0.2922	1.0000	0.9954	0.1912	1.0000	1.0037	0.1931	3.8683	0.0454
MGH	1.0000	2.1779	0.5777	2.2400	1.8217	0.5151	1.0000	0.6897	0.1603	1.0000	0.6931	0.1621	2.5389	0.0411
MH	1.0000	1.0000	0.0000	2.2400	2.4543	0.6451	1.0000	1.0475	0.1859	1.0000	1.0522	0.1862	2.5390	0.0411

Table 3
Bivariate MH and MAH distributions. Reference value, sample bias and sample standard deviation (in brackets) are provided for the respective parameter estimates derived from 100 samples of sample-size 1000 each. Denote Dep.Par.:= $\sigma_{12}/(S_1 S_2)$. The last column lists the sample mean and the sample standard deviation of the cross entropy H.

Distribution	μ_1	μ_2	S_1	S_2	α_1	α_2	β_1	β_2	Dep.Par.	Cross Ent. H
	0.000 value:	0.000 value:	1.000 value:	1.000 value:	1.000 value:	1.000 value:	0.000 value:	0.000 value:	0.000 value:	
MAHmin	0.004 (0.085)	0.002 (0.087)	0.019 (0.208)	0.025 (0.213)	0.040 (0.265)	0.040 (0.265)	0.002 (0.032)	0.002 (0.032)	0.005 (0.037)	3.769 (0.039)
MAHmax	0.005 (0.111)	0.014 (0.112)	0.034 (0.323)	0.060 (0.300)	0.064 (0.398)	0.102 (0.385)	0.006 (0.045)	0.003 (0.044)	0.005 (0.048)	3.768 (0.039)
MH	0.003 (0.095)	0.007 (0.094)	0.023 (0.163)	0.029 (0.168)	0.042 (0.210)	0.042 (0.210)	0.005 (0.040)	0.001 (0.039)	0.003 (0.036)	3.753 (0.038)
	0.000 value:	0.000 value:	0.320 value:	0.320 value:	0.320 value:	0.320 value:	0.000 value:	0.000 value:	0.000 value:	
MAHmin	0.022 (0.830)	0.051 (0.644)	0.001 (0.126)	0.003 (0.131)	0.008 (0.137)	0.008 (0.137)	0.004 (0.115)	0.004 (0.115)	0.019 (0.109)	3.457 (0.112)
MAHmax	0.060 (0.464)	0.015 (0.200)	0.000 (0.205)	0.031 (0.194)	0.011 (0.214)	0.017 (0.203)	0.020 (0.111)	0.008 (0.122)	0.004 (0.214)	3.382 (0.142)
MH	0.071 (0.697)	0.100 (1.015)	0.003 (0.108)	0.004 (0.112)	0.009 (0.123)	0.009 (0.123)	0.004 (0.035)	0.007 (0.067)	0.012 (0.105)	3.418 (0.345)
	0.000 value:	0.000 value:	1.155 value:	1.155 value:	2.236 value:	2.236 value:	0.000 value:	0.000 value:	0.500 value:	
MAHmin	0.001 (0.120)	0.001 (0.075)	0.139 (0.238)	0.143 (0.191)	0.077 (0.626)	0.077 (0.626)	0.003 (0.042)	0.003 (0.042)	0.055 (0.027)	2.687 (0.040)
MAHmax	0.002 (0.110)	0.002 (0.127)	0.085 (0.294)	0.056 (0.320)	0.318 (1.083)	0.270 (0.969)	0.004 (0.064)	0.001 (0.053)	0.004 (0.145)	2.686 (0.040)
MH	0.003 (0.118)	0.002 (0.083)	0.159 (0.224)	0.128 (0.177)	0.128 (0.603)	0.128 (0.603)	0.003 (0.058)	0.000 (0.048)	0.054 (0.026)	2.680 (0.040)
	2.000 value:	4.000 value:	1.155 value:	1.155 value:	1.258 value:	1.258 value:	0.229 value:	0.512 value:	0.500 value:	
MAHmin	0.147 (0.201)	0.759 (0.129)	0.098 (0.253)	0.006 (0.234)	0.224 (0.279)	0.224 (0.279)	0.145 (0.034)	0.138 (0.034)	0.050 (0.029)	4.210 (0.050)
MAHmax	1.470 (0.173)	0.498 (0.171)	0.504 (0.511)	0.521 (0.192)	0.134 (0.483)	0.400 (0.303)	0.154 (0.039)	0.050 (0.014)	0.144 (0.123)	4.175 (0.048)
MH	0.201 (0.117)	0.260 (0.082)	0.031 (0.214)	0.186 (0.174)	0.182 (0.240)	0.182 (0.240)	0.128 (0.038)	0.042 (0.013)	0.007 (0.027)	4.128 (0.045)

9 Application to financial data

The MGH and MAGH distributions are now fitted to various asset-return data: Dax/Cac and Nikkei/Cac returns (both comprising 3989 samples) and Dax/Dow (1254 samples). In particular, the following distributions are used:

(1) MGH/MH,

(2) MAGH/MAH with minimum parameterization (denoted by (min)), that is, with all margins equally parameterized,

(3) MAGH/MAH with maximum parameterization (denoted by (max)), that is, with each margin individually parameterized.

For some of these distributions we have also estimated the symmetric pendant (denoted by (sym)), i.e., $\beta = 0$. Further, we illustrate estimations following the affine-linear transformation method provided at the end of Section 3 (denoted by (PC)). The results are presented in Table 4. The dependence parameter Dep.Par. refers to the sub-diagonal elements of the normed matrix Σ, i.e., Dep.Par.:= $\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} = \sigma_{ij}/(S_i S_j)$.

We have already mentioned that the Kullback divergence or cross entropy is a widespread measure of the goodness-of-fit of multidimensional distributions (in particular, lower cross entropy signals a better fit), see Ullah (1996) for an overview. For small divergences between the distributions (i.e., good fits), cross entropy is approximately equal to another popular divergence measure, the χ^2 divergence:

$$\int f(x) \log \frac{f(x)}{g(x)} dx = \int f(x) \log \left(\frac{f(x) - g(x)}{g(x)} + 1 \right) dx \approx \int \frac{(f(x) - g(x))^2}{g(x)} dx.$$

Thus, one cannot expect substantially different results if alternative divergence measures are applied.

Taking additional parameters or degrees of freedom, such as

- $\lambda \neq 1$ (MAGH or MGH) instead of $\lambda = 1$ (MAH or MH),
- $\beta \neq 0$ (MAGH or MGH) instead of $\beta = 0$ (symmetric MAGH or MGH),
- multiple λ, α, and β parameters (maximum parametrizations) instead of single ones (minimum parametrizations)

can be justified by means of the individual cross entropies. Consider the cross entropies H_i and H_j of variants i and j with variant i having k additional parameters in contrast to variant j. The fact that the cross entropy times the sample size n is equal to the negative log-likelihood leads to the following likelihood ratio L between the variant j and the variant i :

$$L_n(i, j) = n(H_j - H_i).$$

The statistics $2L_n(i, j)$ possesses a χ^2-distribution with k degrees of freedom (see Stuart and Ord (1994)) and can be used as a test for additional parameters.

Table 4
Estimations from two-dimensional asset-return data (Dax-Cac, Dax-Dow, Nikkei-Cac).

Data	Model	Location μ	Scaling S	Dep.Par.	Corr.	λ	α	β	CrossEnt
DaxCac	MAGHmin	0.0012 0.0008	0.0062 0.0060	0.689	0.709	0.889	0.675	-0.034	-6.183
DaxCac	MAHmin	0.0011 0.0008	0.0062 0.0059	0.688	0.709	1.000	0.699	-0.029	-6.183
DaxCac	MAGHmin sym.	0.0004 0.0003	0.0056 0.0054	0.689	0.709	1.002	0.631	0.000	-6.182
DaxCac	MAGHmaxPC	0.0012 0.0005	0.0103 0.0046	0.941	0.709	0.812 0.732	0.192 1.363	0.037 0.000	-6.208
DaxCac	MAHmaxPC	0.0015 0.0007	0.0095 0.0042	0.898	0.709	1.000 1.000	0.261 1.315	0.039 0.000	-6.208
DaxCac	MAGHmaxPC sym.	0.0002 0.0001	0.0107 0.0048	0.999	0.709	0.635 0.645	0.050 1.427	0.000 0.000	-6.275
DaxCac	MGH	0.0010 0.0005	0.0045 0.0044	0.672	0.709	1.134	0.535	-0.038 -0.017	-6.213
DaxCac	MH	0.0011 0.0006	0.0059 0.0058	0.673	0.709	1.000	0.689	-0.044 -0.019	-6.213
DaxCac	MGH sym.	0.0005 0.0003	0.0043 0.0042	0.672	0.709	1.143	0.504	0.000 0.000	-6.212
DaxDow	MAGHmin	0.0025 0.0013	0.0137 0.0097	0.505	0.498	0.891	1.285	-0.067	-5.743
DaxDow	MAHmin	0.0021 0.0011	0.0124 0.0088	0.503	0.498	1.000	1.169	-0.056	-5.743
DaxDow	MAGHmin sym.	0.0002 0.0001	0.0137 0.0097	0.500	0.498	0.759	1.208	0.000	-5.741
DaxDow	MAGHmaxPC	0.0009 0.0006	0.0145 0.0092	0.287	0.498	0.694 0.654	1.014 1.623	-0.001 0.000	-5.746
DaxDow	MAHmaxPC	0.0012 0.0002	0.0122 0.0082	0.373	0.498	1.000 1.000	0.868 1.561	0.032 0.000	-5.746
DaxDow	MAGHmaxPC sym.	0.0000 0.0001	0.0147 0.0093	0.277	0.498	0.650 0.657	1.013 1.634	0.000 0.000	-5.746
DaxDow	MGH	0.0020 0.0003	0.0135 0.0097	0.489	0.498	1.087	1.359	-0.077 -0.010	-5.757
DaxDow	MH	0.0021 0.0004	0.0137 0.0098	0.489	0.498	1.000	1.346	-0.079 -0.015	-5.757
DaxDow	MGH sym.	0.0001 0.0001	0.0130 0.0093	0.488	0.498	1.097	1.289	0.000 0.000	-5.755
NikkeiCac	MAGHmin	0.0000 0.0004	0.0028 0.0027	0.196	0.236	0.870	0.272	-0.011	-5.884
NikkeiCac	MAHmin	-0.0000 0.0004	0.0024 0.0023	0.199	0.236	1.000	0.245	-0.008	-5.884
NikkeiCac	MAGHmin sym.	-0.0001 0.0003	0.0028 0.0027	0.197	0.236	0.880	0.278	0.000	-5.883
NikkeiCac	MAGHmaxPC	0.0008 0.0007	0.0084 0.0069	0.170	0.236	0.646 0.902	0.700 0.957	0.009 0.000	-5.839
NikkeiCac	MAHmaxPC	0.0008 0.0008	0.0067 0.0058	0.352	0.236	1.000 1.000	0.564 0.859	0.008 0.000	-5.838
NikkeiCac	MAGHmaxPC sym.	0.0000 0.0003	0.0090 0.0076	0.278	0.236	0.652 0.639	0.714 1.020	0.000 0.000	-5.838
NikkeiCac	MGH	0.0004 0.0005	0.0032 0.0031	0.217	0.236	1.117	0.355	-0.026 -0.014	-5.885
NikkeiCac	MH	0.0004 0.0005	0.0043 0.0041	0.216	0.236	1.000	0.459	-0.028 -0.017	-5.885
NikkeiCac	MGH sym.	-0.0000 0.0003	0.0032 0.0031	0.217	0.236	1.118	0.349	0.000 0.000	-5.885

In our case, a single additional parameter such as λ or β is justified on a significance level of 0.05 with $H_j - H_i > 3.84146/3989 \approx 0.001$ for Dax/Cac and Nikkei/Cac data and with $H_j - H_i > 3.84146/1254 \approx 0.003$ for Dax/Dow data. However, this is not the case for minimally parameterized variants, in particular generalized asymmetric variants are not justified on this significance level.

Three additional parameters (λ, α, and β in the maximally parameterized variants) are justified on a significance level 0.05 with $H_j - H_i > 7.81473/3989 \approx 0.002$ for Dax/Cac and Nikkei/Cac data and with $H_j - H_i > 7.81473/1254 \approx 0.006$ for Dax/Dow data. On the significance level 0.01, the critical differences are 0.003 for Dax/Cac and Nikkei/Cac data and 0.009 for Dax/Dow data. These significance thresholds are exceeded for Dax/Cac data - the flexibility of the maximally parameterized variants is obviously valuable for the modelling in this case.

To evaluate the goodness-of-fit of the respective distribution models to the financial data, the χ^2 test has been performed. While there are diverse recommendations (see Stuart and Ord (1994)) concerning the choice of the class intervals in the univariate case, the choice remains difficult for multivariate distributions. Unfortunately, the results of the test depend essentially on this choice. We have used some simple choices: 6^2, 8^2, and 10^2 intervals of width 0.005 or 0.01. For 6^2 and 8^2 intervals, the hypotheses that Dax/Dow data do not arise from the individual MGH and MAGH distributions cannot be rejected on the one per cent significance level (for MGH distributions and 8^2 intervals the hypothesis cannot be rejected on the five per cent significance level). For the other data sets, the hypothesis could be rejected on the one percent significance level. However, according to our analysis the hypothesis must be rejected for any other common parametric multidimensional distribution which is different from the hyperbolic family.

Due to the total or approximate symmetry of some distribution models, the dependence parameters can be roughly interpreted as "correlation coefficients" for MGH and MAGHmin distributions. They are even close to the corresponding sample correlation coefficient (column "Corr."). Note that such an interpretation is not possible for the MAGHmax model due to the different parameterization of the one-dimensional margins.

Conclusion

Summarizing the results we have investigated an interesting new class of multidimensional distributions with exponentially decreasing tails to analyze high-dimensional data, namely the multivariate affine generalized hyperbolic distributions. These distributions are attractive regarding the estimation of unknown parameters and random vector generation. We illustrated that this class of distributions possesses an appealing dependence structure and we derived several dependence properties. Finally, an extensive simulation study showed the flexibility and robustness of the introduced model. Thus, the use of multivariate affine generalized hyperbolic distributions is recommended for multidimensional data modelling in statistical and financial applications.

Appendix

Proof (Theorem 8). i) Suppose $X \in MGH_2(\mu, \Sigma, \omega)$ with parameter $\beta = 0$. Then X belongs to the family of elliptically contoured distributions and the assertion follows by Theorem 2 in Lindskog, McNeil, and Schmock (2003).
ii) Suppose $X \in MAGH_2(\mu, \Sigma, \omega)$ with stochastic representation $X \stackrel{d}{=} A'Y + \mu,\ A'A = \Sigma$, copula C and parameter ρ. For $\rho = 0$, the assertion follows by Theorem 5.1.9 in Nelsen (1999) due to the independence of X_1 and X_2. For the remaining case we can assume $\rho > 0$ as for $\rho < 0$ the assertion is shown similarly. According to Theorem 5.1.3 in Nelsen (1999), Kendall's tau is a copula property what justifies the assumption $\mu = 0$. Further, Kendall's tau is invariant under strictly increasing transformations of the margins (see Theorem 5.1.9 in Nelsen (1999)) and therefore we may set

$$A' = \begin{pmatrix} 1 & 0 \\ 1 & c \end{pmatrix} \quad \text{with } c := \sqrt{1/\rho^2 - 1}.$$

Consequently, the distribution function of $X = (X_1, X_2)'$ has the form

$$F_X(x_1, x_2) = \mathbb{P}(Y_1 \le x_1, Y_1 + cY_2 \le x_2) = \int_{-\infty}^{x_1} F_{Y_2}\left(\frac{x_2 - z}{c}\right) f_{Y_1}(z)\, dz,$$

and the corresponding density function is $f_X(x_1, x_2) = f_{Y_2}\left(\frac{x_2 - x_1}{c}\right) f_{Y_1}(x_1)/|c|$. Thus, using the fact that

$$\tau = 4 \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1 = 4 \int_{\mathbb{R}^2} F_X(x_1, x_2) f_X(x_1, x_2)\, d(x_1, x_2) - 1,$$

formula (4) is shown. □

In order to prove Theorem 10 we first investigate the tail behavior of the univariate symmetric MGH and MAGH distributions. The tail of the distribution function F, as always, is denoted by $\overline{F} := 1 - F$.

Definition 11 (Semi-heavy tails) *A continuous (symmetric) function* $g : \mathbb{R} \to (0, \infty)$ *is called semi-heavy tailed (or exponentially tailed) if it satisfies*

$$g(x) \sim c|x|^\nu \exp(-\eta|x|) \quad \text{as } x \to \pm\infty, \tag{9}$$

with $\nu \in \mathbb{R}$, $\eta > 0$ *and some positive constant* c. *The class of (symmetric) semi-heavy tailed functions is denoted by* $L_{\nu,\eta}$.

Lemma 12 *Let* f *be a density function such that* $f \in L_{\nu,\eta}$, $\nu \in \mathbb{R}$, $\eta > 0$. *Then the corresponding distribution function* F *possesses the same asymptotic behavior as its density, i.e.,* $F(x) \sim \bar{c}|x|^\nu \exp(-\eta|x|)$ *as* $x \to -\infty$ *and* $\bar{F}(x) \sim \bar{c}x^\nu \exp(-\eta x)$ *as* $x \to \infty$ *for some positive constant* $\bar{c}$; *write* $F \in L_{\nu,\eta}$.

Proof. Consider e.g. the tail function $\bar{F}$. Applying partial integration we obtain

$$\bar{F}(x) = \int_x^\infty f(u)\, du \sim c \int_x^\infty u^\nu \exp(-\eta u)\, du = c\eta x^\nu \exp(-\eta x) + \eta c\nu \int_x^\infty u^{\nu-1} \exp(-\eta u)\, du.$$

Thus, the proof is complete if we show that $\int_x^\infty u^{\nu-1} \exp(-\eta u)\, du \big/ x^\nu \exp(-\eta x) = o(1)$ as $x \to \infty$. Rewriting the latter quotient yields

$$0 \le \frac{1}{x} \int_x^\infty \left(\frac{u}{x}\right)^{\nu-1} \exp(-\eta(u - x))\, du = \frac{1}{x} \int_0^\infty \left(\frac{u + x}{x}\right)^{\nu-1} \exp(-\eta u)\, du.$$

The assertion is now immediate because

$$\left(\frac{u}{x} + 1\right)^{\nu-1} \le (u + 1)^{\nu-1} \text{ for } \nu \ge 1 \text{ and } \left(\frac{u}{x} + 1\right)^{\nu-1} \le 1 \text{ for } \nu < 1,$$

and the corresponding integrals exist. □

The next lemma is quite useful; it states that the tail of the convolution of two semi-heavy tailed distributions is determined by the heavier tail.

Lemma 13 *Let F_1 and F_2 be distribution functions with $F_1 \in L_{\nu_1,\eta_1}$ and $F_2 \in L_{\nu_2,\eta_2}$ where $0 < \eta_2 < \eta_1$, $\nu_1, \nu_2 \in I\!R$. Then $F_1 * F_2 \in L_{\nu_2,\eta_2}$ and, moreover,*

$$\lim_{t\to\infty} \overline{F_1 * F_2}(t)/\overline{F}_2(t) = m_1 := \int_{-\infty}^{\infty} e^{\eta_2 u} dF_1(u). \tag{10}$$

Proof. For some fixed $s > 1$, we have

$$\begin{aligned}\overline{F_1 * F_2}(t) &= \int_{-\infty}^{\infty} \overline{F}_2(t-u)dF_1(u) = \int_{-\infty}^{t/s} \overline{F}_2(t-u)dF_1(u) - \int_{-\infty}^{t-t/s} \overline{F}_2(u)dF_1(t-u) \\ &= \int_{-\infty}^{t/s} \overline{F}_2(t-u)dF_1(u) + \int_{-\infty}^{t-t/s} \overline{F}_1(t-u)dF_2(u) + \overline{F}_2(t-t/s)\overline{F}_1(t/s),\end{aligned}$$

where the last equality follows by partial integration. Thus, dominated convergence yields

$$\lim_{t\to\infty} \overline{F_1 * F_2}(t)/\overline{F}_2(t) = \lim_{t\to\infty} \int_{-\infty}^{t/s} \overline{F}_2(t-u)/\overline{F}_2(t)dF_1(u) = \int_{-\infty}^{\infty} e^{\eta_2 u} dF_1(u) =: m_1 < \infty,$$

because

$$0 \le \int_{-\infty}^{t-t/s} \overline{F}_1(t-u)/\overline{F}_2(t)dF_2(u) \le \frac{\overline{F}_1(t/s)}{\overline{F}_2(t)} \cdot \overline{F}_2(t-t/s) \to 0 \quad \text{as } t \to \infty.$$

A consequence of the symmetric tails of F_1 and F_2 is that $\lim_{t\to-\infty} F_1 * F_2(t)/F_2(t) = m_1$. Hence, $F_1 * F_2 \in L_{\nu_2,\eta_2}$ is proven. □

According to Barndorff-Nielsen and Blæsild (1981), the univariate MGH distributions have semi-heavy tails, in particular

$$MGH_1(0, 1, \omega) \sim c|x|^{\lambda-1} \exp((\mp\alpha + \alpha\beta)x) \text{ as } x \to \pm\infty, \tag{11}$$

with some positive constant c. Hence, in the symmetric case $\beta = 0$ we obtain $MGH_1(0, 1, \omega) \in L_{\nu,\eta}$ with $\nu = \lambda - 1$ and $\eta = \alpha$. Now we are ready to prove Theorem 10.

Proof (Theorem 10). We only show upper tail-dependence and upper tail-independence, respectively, as the lower pendant is obtained similarly. Recall that tail dependence is a copula property and therefore we may put $\mu = 0$.

i) Let $X \in MGH_2(0, \Sigma, \omega)$ with $\beta = 0$. In that case X belongs to the family of elliptically contoured distributions. According to Theorem 6.8 in Schmidt (2002) the assertion follows because of the exponentially-tailed density generator.

ii) Let $X \in MAGH_2(0, \Sigma, \omega)$ with stochastic representation $X \stackrel{d}{=} A'Y$ and Cholesky matrix $A' = (a_{ij})_{i,j=1,2}$. Note that $a_{11}, a_{22} > 0$. Then

$$\mathbb{P}(X_2 > F_{X_2}^{-1}(v) \mid X_1 > F_{X_1}^{-1}(v)) = \frac{\mathbb{P}(a_{11}Y_1 > F_{X_1}^{-1}(v), a_{12}Y_1 + a_{22}Y_2 > F_{X_2}^{-1}(v))}{\mathbb{P}(a_{11}Y_1 > F_{X_1}^{-1}(v))}$$

$$= \frac{1}{1-v} \int_{F_{Y_1}^{-1}(v)}^{\infty} \mathbb{P}(Y_2 > (F_{X_2}^{-1}(v) - a_{12}y)/a_{22}) f_{Y_1}(y)\, dy =: I,$$

because Y_1 and Y_2 are independent random variables. If $\rho \le 0$ then $a_{12} \le 0$ and upper tail-independence immediately follows by dominated convergence.

Consider now $\rho > 0$ and therefore $a_{12} > 0$. Let $\alpha_2 < \alpha_1\sqrt{1/\rho^2 - 1}$. Then $\alpha_2/a_{22} < \alpha_1/a_{12}$. For all $\varepsilon \in (0,1]$ we conclude with $u := 1 - v$ that

$$\begin{aligned} I &\le \varepsilon + \frac{1}{u} \int_{F_{Y_1}^{-1}(1-u)}^{F_{Y_1}^{-1}(1-u\varepsilon)} \mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}y)/a_{22}) f_{Y_1}(y)\, dy \\ &\le \varepsilon + (1-\varepsilon)\mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}). \end{aligned} \tag{12}$$

Due to (11) we know that $F_{a_{12}Y_1} \in L_{\nu_1,\eta_1}$ with $\nu_1 = \lambda_1 - 1$ and $\eta_1 = \alpha_1/a_{12}$. Thus, Lemma 13 gives $F_{X_2} \in L_{\nu_2,\eta_2}$ with $\nu_2 = \lambda_2 - 1$ and $\eta_2 = \alpha_2/a_{22}$ as $0 < \eta_2 < \eta_1$. Then the probability in (12) converges to zero as $u \to 0^+$ if $F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon) = F_{a_{12}Y_1+a_{22}Y_2}^{-1}(1-u) - F_{a_{12}Y_1}^{-1}(1-u\varepsilon) \to \infty$ as $u \to 0^+$. Put $x_u := F_{a_{12}Y_1}^{-1}(1-u\varepsilon)$ and $y_u := F_{a_{12}Y_1+a_{22}Y_2}^{-1}(1-u)$. Then

$$u = \frac{1}{\varepsilon}\overline{F}_{a_{12}Y_1}(x_u) = \overline{F}_{X_2}(y_u) \sim \frac{c_1}{\varepsilon} x_u^{\nu_1} \exp(-\eta_1 x_u) \sim c_2 y_u^{\nu_2} \exp(-\eta_2 y_u), \tag{13}$$

as $u \to 0^+$ and therefore $x_u, y_u \to \infty$. The asymptotic behavior (13) implies $y_u - x_u \to \infty$ as $u \to 0^+$ because $0 < \eta_2 < \eta_1$. Hence, upper tail-independence is shown.

iii) Now suppose $\rho > 0$ and $\alpha_2 > \alpha_1\sqrt{1/\rho^2 - 1}$ which yields $a_{12} > 0$ and $\alpha_2/a_{22} > \alpha_1/a_{12}$. According to Lemma 13 we have $F_{a_{12}Y_1} \in L_{\nu_1,\eta_1}$ and $F_{X_2} \in L_{\nu_1,\eta_1}$ with $\nu_1 = \lambda_1 - 1$ and $\eta_1 = \alpha_1/a_{12}$. Notice that with $u := 1 - v$

$$I \ge \mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}). \tag{14}$$

Further $u = \overline{F}_{a_{12}Y_1}(x_u) = \overline{F}_{X_2}(y_u) \sim c_1 x_u^{\nu_1} \exp(-\eta_1 x_u) \sim c_2 y_u^{\nu_1} \exp(-\eta_1 y_u)$. Hence

$$\frac{c_2}{c_1}\left(\frac{y_u}{x_u}\right)^{\nu_1} \exp(-\eta_1(y_u - x_u)) \to 1 \text{ as } u \to 0^+.$$

Suppose that $\limsup_{u\to 0^+}(y_u - x_u) = \infty$. If $\nu_1 \ge 0$, then the fact

$$\liminf_{u\to 0^+} \frac{c_2}{c_1}\left(\frac{y_u}{x_u}\right)^{\nu_1} \exp(-\eta_1(y_u - x_u)) \le \liminf_{u\to 0^+} \frac{c_2}{c_1}(2(y_u - x_u))^{\nu_1} \exp\Big(-\eta_1(y_u - x_u)\Big) = 0,$$

as $u \to 0^+$ would lead to a contradiction. On the other hand, if $\nu_1 < 0$ then

$$\liminf_{u\to 0^+} \frac{c_2}{c_1}\left(\frac{y_u}{x_u}\right)^{\nu_1} \exp(-\eta_1(y_u - x_u)) \le \liminf_{u\to 0^+} \frac{c_2}{c_1} \exp\Big(-\eta_1(y_u - x_u)\Big) = 0 \text{ as } u \to 0^+$$

would also lead to a contradiction. Therefore we conclude that $\liminf_{u\to 0^+} \mathbb{P}(Y_2 > (F_{X_2}^{-1}(1-u) - a_{12}F_{Y_1}^{-1}(1-u\varepsilon))/a_{22}) > 0$ as Y_2 is supported on $\mathbb{R}$. Finally, the limit $\lim_{v\to 1-} \mathbb{P}(X_2 > F_{X_2}^{-1}(v) \mid X_1 > F_{X_1}^{-1}(v))$ exists due to the convexity of $\overline{F}_{X_1}$ and $\overline{F}_{X_2}$ for large arguments. □

Acknowledgement. The authors would like to thank Professor Friedrich Schmid and Professor Ulrich Stadtmüller for very helpful discussions. Further, the authors are grateful to Professor Gholamreza Nakhaeizadeh for supporting this work. The authors thank two anonymous referees for very constructive comments which improved an earlier version of this paper.

References

Atkinson, A.C., 1982, The simulation of generalized inverse Gaussian and hyperbolic random variables, *Siam J. Sci. Stat. Comput* 3, 502–515.

Barndorff-Nielsen, O. E., 1977, Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society London* A, 401–419.

Barndorff-Nielsen, O. E., 1978, Hyperbolic distributions and distributions on hyperbolae, *Scand. J. Statist.* pp. 151–157.

Barndorff-Nielsen, O. E., and P. Blæsild, 1981, Hyperbolic distributions and ramifications: Contributions to theory and application. In: C. Taillie, G. Patil, and B. Baldessari (Eds.), *Statistical Distributions in Scientific Work*, Dordrecht Reidel vol. 4, pp. 19–44.

Barndorff-Nielsen, O. E., J. Kent, and M. Sørensen, 1982, Normal Variance-Mean Mixtures and z Distributions, *International Statistical Review* 50, 145–159.

Bauer, C., 2000, Value at risk using hyperbolic distributions, *J. Economics and Business* 52, 455–467.

Bingham, N. H., and R. Kiesel, 2001, Modelling asset returns with hyperbolic distributions. In: J. Knight and S. Satchell (Eds.), *Asset return distributions*, Butterworth-Heinemann, 1–20.

Blæsild, P., and J. L. Jensen, 1981, Multivariate distributions of hyperbolic type. In: C. Taillie, G. Patil, and B. Baldessari (Eds.), *Statistical distributions in scientific work*, Dordrecht Reidel, 4, 45–66.

Boender, C.G.E., and A.H.G. Rinnooy Kan, 1987, Bayesian stopping rules for multistart global optimization methods, *Mathematical Programming* 37, 59–80.

Cambanis, S., S. Huang, and G. Simons, 1981, On the Theory of Elliptically Contoured Distributions, *Journal of Multivariate Analysis* 11, 368–85.

Cherubini, U., E. Luciano, and W. Vecchiato, 2004, *Copula Methods in Finance.* (John Wiley & Sons).

Eberlein, E., 2001, Application of generalized hyperbolic Lévy motions to finance. In: O.E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes: Theory and Applications,* 319–337.

Eberlein, E., and U. Keller, 1995, Hyperbolic distributions in finance, *Bernoulli* 1, 281–299.

Eberlein, E., and K. Prause, 2002, The generalized hyperbolic model: financial derivatives and risk measures In: H. Geman, D. Madan, S. Pliska, and T. Vorst (Eds.), *Mathematical Finance-Bachelier Congress 2000,* 245–267.

Embrechts, P., F. Lindskog, and A. McNeil, 2003, Modelling Dependence with Copulas and Applications to Risk Management. In: S. Rachev (Ed.) *Handbook of Heavy Tailed Distibutions in Finance*, Elsevier, Chapter 8, 329–384.

Embrechts, P., A. McNeil, and D. Straumann, 2002, Correlation and Dependency in Risk Management: Properties and Pitfalls. In: M.A.H. Dempster (Ed.), *Risk Management: Value at Risk and Beyond,* Cambridge University Press, Cambridge, 176–223.

Fang, K.T., S. Kotz, and K.W. Ng, 1990, *Symmetric Multivariate and Related Distributions.* (Chapman and Hall London).

Fletcher, Roger, 1987, *Practical Methods of Optimization.* (Wiley New York).

Hauksson, H.A., M. Dacorogna, T. Domenig, U. Mueller, and G. Samorodnitsky, 2001, Multivariate extremes, aggregation and risk estimation, *Quantitative Finance* 1, 79–95.

Joe, H., 1997, *Multivariate Models and Dependence Concepts.* (Chapman and Hall London).

Kullback, S, 1959, *Information theory and statistics.* (Wiley New York).

Lindskog, F., A. McNeil, and U. Schmock, 2003, Kendall's tau for elliptical distributions. In: G. Bohl, G. Nakhaeizadeh, S.T. Rachev, T. Ridder and K.H. Vollmer (Eds.), *Credit Risk: Measurement, Evaluation and Management*, Physica-Verlag Heidelberg, 649–676.

Magnus, W., F. Oberhettinger, and R.P. Soni, 1966, *Formulas and Theorems for the Special Functions of Mathematical Physics* vol. 52 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen.* (Springer Verlag, Berlin).

Nelsen, R.B., 1999, *An Introduction to Copulas.* (Springer Verlag).

Olbricht, W., 1991, On mergers of distributions and distributions with exponential tails, *Computational Statistics & Data Analysis* 12, 315–326.

Parzen, E., 1962, On estimation of probability density function and mode, *Ann. Math. Stat.* 35, 1065–1076.

Prause, K., 1999, The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures, PhD Thesis, Albert-Ludwigs-Universität Freiburg.

Rinnooy Kan, A.H.G., and G.T. Timmer, 1987, Stochastic global optimization methods, Part I: Clustering methods, Part II: Multi-level methods, *Mathematical Programming* 39, 26–78.

Ross, S.M., 1997, *Introduction to Probability models.* (Academic Press, San Diego) 6 edn.

Schmidt, R., 2002, Tail dependence for elliptically contoured distributions, *Mathematical Methods of Operations Research* 55, 301–327.

Stuart, Alan, and Keith Ord, 1994, *Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory.* (Arnold London).

Stuart, Alan, and Keith Ord, 1994, *Kendall's Advanced Theory of Statistics, Volume II: Classical Inference and the Linear Model.* (Arnold London).

Stützle, Eric A., and Tomas Hrycej, 2001, Forecasting of Conditional Distributions - An Application to the Spare Parts Demand Forecast. In: *Proc. 2001 IASTED Int. Conf. on Artificial Intelligence and Soft Computing* Cancun, Mexico.

Stützle, Eric A., and Tomas Hrycej, 2002a, Estimating Multivariate Conditional Distributions via Neural Networks and Global Otimization. In: *Proc. 2001 IEEE Int. Joint Conference on Neural Networks* Honolulu, Hawaii.

Stützle, Eric A., and Tomas Hrycej, 2002b, Modelling Future Demand by Estimating the Multivariate Conditional Distribution via the Maximum Likelihood Principle and Neural Networks. In: *Proc. 2002 IASTED Int. Conf. on Modelling, Identification and Control* Innsbruck, Austria.

Ullah, A., 1996, Entropy, divergence and distance measuser with econometric applications, *Journal of Statistica Planning and Inference* 49, 137–162.