Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

A multidimensial mixture model for unsupervised tail estimation

Note no SAMBA/09/09

Authors Lars Holden
Ola Haug

Date 26th February 2009

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**A multidimensial mixture model for unsupervised tail estimation**
Authors	**Lars Holden** <lars.holden@nr.no> **Ola Haug** <ola.haug@nr.no>
Date	26th February 2009
Publication number	SAMBA/09/09

Abstract

This paper proposes a new method to combine several densities such that each density dominates a separate part of a joint distribution. The method is fully unsupervised, i.e. the parameters in the densities and the thresholds are simultaneously estimated. The approach uses cdf functions in the mixing. This makes it easy to estimate parameters and the resulting density is smooth. Our method may be used both when the tails are heavier and lighter than the rest of the distribution, and it is illustrated in 1D and 3D. The presented model is compared with other published models and a very simple model using a univariate transformation.

Keywords	Mixing functions, heavy and light tailed distributions, maximum likelihoods, mixture models
Target group	
Availability	Open
Project	SFI-finans
Project number	
Research field	Finance, insurance and commodities
Number of pages	32

Contents

1 Introduction

In many applications, the tail of a probability distribution is of particular interest, e.g. prediction of floods or estimation of financial reserves in insurance. Because extreme data are rare, it is difficult to fit tail models and to support parametric model choices convincingly. Most papers study this problem in one dimension assuming a heavy tail. The approach presented in this paper mix different distributions that describes different parts of a new joint distribution. The joint distribution may have heavier or lighter tail(s) compared to the tail(s) of the distribution that is used in the central part of the joint distribution. Moreover, it is possible to generalise to R^n.

Common practice in extreme value modelling is to fix a threshold u and to fit a distribution, e.g. a generalized Pareto Distribution (GPD), to the data exceeding u. There is a number of methods to estimate the parameters once u is fixed, see for instance (21), (7) and references therein. As is well known, the estimates depend significantly on the choice of the threshold, see for instance (10), Figure 6.2.8. In order to reduce model bias, the threshold u should be chosen large, but this often leaves very few data points for the estimation of the parameters. Hence, the resulting parameter estimates will have large variances. Moreover, the selection of an appropriate threshold is a difficult task in practice, see for instance (8), (22), (10), and (18). Often a supervised analysis is performed, selectively and off-line as part of a monitoring scheme. For practitioners, who usually need to perform their data analyses regularly, it would be convenient to have automatic and robust approaches that do not require an a priori tuning of a threshold. Such an unsupervised approach to tail estimation would be of particular relevance in automatic real-time monitoring of financial, industrial and environmental quantities, for instance for warning purposes.

Recently, (9) and (12) have proposed two new ways of addressing this question. The paper (9) suggests a robust model validation mechanism to guide the threshold selection. The procedure assigns weights between zero and one to each data point, where a high weight means that the point should be retained since a GPD model is fitting it well. The author suggests to start with a low threshold u and increase it, thus reducing the number of data points, until all data left have weights close to 1. This is a promising method, but thresholding is still needed at the level of the weights.

The paper (12) was the first to suggest a fully unsupervised approach to tail estimation. Their approach has three key ingredients. First, the model consists of two components, one representing the central part of the distribution and the other the tail. Second, all data are modelled in one mixture model, and finally, the parameters in the two distributions and the mixing parameters are simultaneously estimated.

Our method is based on the same ideas as those in (12), but we mix the cumulative distribution functions (cdfs) instead of densities. This makes our approach computationally more efficient, which in turn makes it easier to generalise to higher dimensions. Threshold estimation is more complicated in higher dimensions, since the threshold is a surface. However, in our approach, this is handled efficiently. We also show how to use several different distributions for different parts of the tail(s). Also the method presented

in (3) uses cdfs in the mixing. But our method gives a continuous density in contrast to the method presented in (3).

The proposed model is compared with a model based on a univariate transformation. The properties of the two models are quite similar. But when generalizing to several dimensions there are important differences. The model based on cdfs may combine different multivariate densities, but needs to calculate the cdfs and not only the densities in the evaluation. The model based on univariate transformation does not need to calculate the multivariate cdfs, but the properties are dominated by the properties of the chosen multivariate distribution. If we want to change the multivariate properties, the method may be combined with a copula approach.

In many applications it is needed to describe the entire distribution, not only the tail(s). In our methods, the user selects densities that he/she believes fits the different parts of the data. However, in some cases the density that is used for describing the tail also describes the central part of the distribution better than the density that is supposed to describe the central part of the density. Then the tail density may end up modelling most of the density leading to better overall match with data, but with poorer description of the tail. This is easy to identify from the estimated threshold and may for example be corrected by putting a prior on the threshold.

Assessments of the probabilities of multivariate extreme events are sought in a diversity of applications, see e.g. (23), (16), (20), (14), and (13). Our method differs from the mentioned papers in that we only mix different densities, with or without heavy tails. We define a mixing zone where the density changes smoothly from one density to another density. This is particularly relevant for densities with heavy tails where the tail is modelled by a different density than the rest of the distribution.

We first describe the models in one dimension in Sections 2 - 4, and then generalize to several dimensions in Sections 5 - 6. Then, in Section 7, the models are tested with synthetic one-dimensional and three-dimensional data and applied to real univariate and three-dimensional financial data. The ambition of the paper is not to suggest the best modelling of these data, only to illustrate the new methods on relevant data. Finally, Section 8 contains some concluding remarks.

2 The cdf-model in one dimension

In this section we start with two one-dimensional components in the mixture, and then we show how the model may be generalised to several components.

2.1 Two components

Let $x \in R$ and let $G(x;\theta_G)$ and $F(x;\theta_F)$ be two cdfs that we want to combine to a cdf denoted L. Define a threshold u and a mixing zone $(u-\varepsilon, u+\varepsilon)$ for $\varepsilon \geq 0$, and let the cdfs G and F determine the properties of L below and above the mixing zone, respectively. Further, let both cdfs influence L in the mixing zone. We will often mix truncated

distributions. Hence, we define the truncated densities where

$$g_t(x;\theta_G) = \begin{cases} g(x;\theta_G) & \text{if } x < u \\ 0 & \text{if } x \geq u \end{cases}$$

and

$$f_t(x;\theta_F) = \begin{cases} 0 & \text{if } x < u \\ f(x;\theta_F) & \text{if } x \geq u. \end{cases}$$

The corresponding truncated cdfs are defined as

$$G_t(x;\theta_G) = \int_{-\infty}^{x} g_t(t;\theta_G)dt$$

and

$$F_t(x;\theta_F) = \int_{-\infty}^{x} F_t(t;\theta_F)dt$$

where we do not require that $G_t(\infty;\theta_G) = 1$ and $F_t(\infty;\theta_F) = 1$. We then define the mixed cdf by

$$L(x;\theta_L) = \kappa(G_t(q(x;\theta_q);\theta_G) + F_t(p(x;\theta_p);\theta_F)) \tag{1}$$

where κ is defined such that $L(\infty;\theta_L) = 1$ and $q(x;\theta_q)$ and $p(x;\theta_q)$ are two monotone increasing mixing functions described below. The parameters of $L(x;\theta_L)$ include all the other parameters i.e. $\theta_L = (\theta_G, \theta_F, \theta_q, \theta_p)$. Equation (1) is a well-defined cdf when the truncated cdfs G_t and F_t and the mixing functions q and p satisfy the criteria specified below. The corresponding density $l(x;\theta_L)$ is given by

$$l(x;\theta_L) = \begin{cases} \kappa\, g(x;\theta_G) & \text{if } x < u - \varepsilon \\ \kappa\,(g(q(x;\theta_q);\theta_G)\, q'(x;\theta_q) + & \\ \qquad f(p(x;\theta_p);\theta_F)\, p'(x;\theta_p)) & \text{if } x \in (u-\varepsilon, u+\varepsilon) \\ \kappa\, f(x;\theta_F) & \text{if } x > u + \varepsilon \end{cases} \tag{2}$$

This requires that $q(x;\theta_q) = p(x;\theta_q) = x$ where it is applied for x outside the mixing zone. The mixing functions q and p determine how G and F influence L in the mixing zone. They are monotonously increasing functions defined on R and with range equal to R. If we set $\varepsilon = 0$ and the two mixing functions equal to the identity function i.e. $q(x;\theta_q) = p(x;\theta_q) = x$, then we get the standard approach where only data above the threshold is used in the tail estimation and the joint distribution is discontinuous. In our approach, we want all data to be used in the estimation of a continuous density $l(x;\theta_L)$. Then we set $\varepsilon > 0$ and the function q maps the interval $(-\infty, u+\varepsilon)$ onto $(-\infty, u)$ and p maps the interval $(u-\varepsilon, \infty)$ onto (u, ∞) as illustrated in Figure 1. In order to get the derivative of $l(x;\theta_L)$ to be continuous, we need to have $q'(u+\varepsilon;\theta_q) = q''(u+\varepsilon;\theta_q) = 0$ and $p'(u-\varepsilon;\theta_q) = p''(u-\varepsilon;\theta_q) = 0$. Further properties of the mixing function is discussed in Section 2.2. We have found that the two mixing functions $p(x;\theta_p)$ and $q(x;\theta_q)$ defined below work well. Define

$$q(x;\theta_q) = \begin{cases} x & x < u - \varepsilon \\ \frac{1}{2}(x + u - \varepsilon) + \frac{\varepsilon}{\pi}\cos(\frac{\pi(x-u)}{2\varepsilon}) & u - \varepsilon \leq x < u + \varepsilon \\ x - \varepsilon & u + \varepsilon \leq x, \end{cases} \tag{3}$$

$$p(x;\theta_p) = \begin{cases} x+\varepsilon & x < u-\varepsilon \\ \frac{1}{2}(x+u+\varepsilon) - \frac{\varepsilon}{\pi}\cos(\frac{\pi(x-u)}{2\varepsilon}) & u-\varepsilon \le x < u+\varepsilon \\ x & u+\varepsilon \le x. \end{cases} \quad , \tag{4}$$

where $\theta_q = \theta_p = (u, \varepsilon)$. Note that the properties of q and p imply that $g_t = g$ and $f_t = f$. Figure 2 shows an example where two Gaussian densities are mixed.

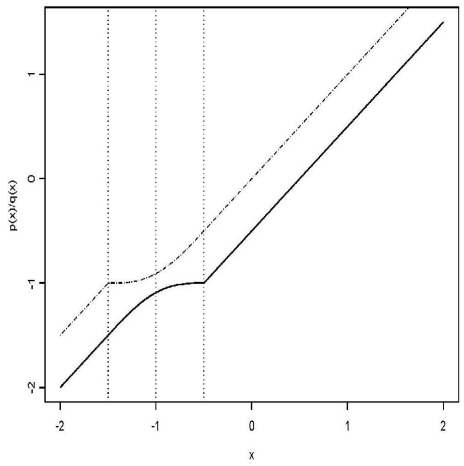

Figure 1. The mixing functions $p(x;\theta_p)$ (dotted line) and $q(x;\theta_q)$ (solid line) for $u = -1$ and $\varepsilon = 0.5$. The vertical bars correspond to $u-\varepsilon$, u and $u+\varepsilon$, respectively.

2.2 Several components

Equation (1) may easily be generalised to a mixture of several truncated cdfs $G_1, \ldots, G_k$ with parameters $\theta_{G_1}, \ldots, \theta_{G_k}$. Define a threshold u_i and a mixing zone $(u_{i-1}-\varepsilon_{i-1}, u_i+\varepsilon_i)$, where $\varepsilon_i \ge 0$ and the truncated density $g_i(x;\theta_G) > 0$ only if $u_{i-1} < x < u_i$ for each component i. We assume that $u_0 = -\infty$ and $u_k = \infty$. The resulting cdf is given by

$$L(x;\theta_L) = \kappa \sum_{i=1}^{k} G_i(q_i(x;\theta_{q_i});\theta_{G_i}), \tag{5}$$

where κ is defined such that $L(\infty;\theta_L) = 1$. Let $G_i(u_{i-1};\theta_{G_i}) = 0$ and

$$G_i(x;\theta_{G_i}) = \int_{u_{i-1}}^{x} g_i(t;\theta_G)dt$$

for $i = 1, \cdots, k$.

The density $l(x;\theta_L)$ corresponding to the cdf $L(x;\theta_L)$ in Equation (5) is given by

$$l(x;\theta_L) = \begin{cases} \kappa\, g_i(x;\theta_{G_i}) & \text{if } x < u_i - \varepsilon_i \\ \kappa\,(g_i(q_i(x;\theta_{q_i});\theta_{G_i})\, q_i'(x;\theta_{q_i}) + & \\ \qquad g_{i+1}(q_{i+1}(x;\theta_{q_{i+1}});\theta_{G_{i+1}})\, q_{i+1}'(x;\theta_{q_{i+1}})) & \text{if } x \ge u_i - \varepsilon_i \end{cases} \tag{6}$$

assuming we have $x \in (u_{i-1}+\varepsilon_{i-1}, u_i+\varepsilon_i)$ for a value of i. The first expression denotes the density between two consecutive mixing zones, and the other within a mixing zone.

Each mixing function q_i, with parameters $\theta_{q_i} = (u_{i-1}, \varepsilon_{i-1}, u_i, \varepsilon_i)$, maps the interval $(u_{i-1} - \varepsilon_{i-1}, u_i + \varepsilon_i)$ onto (u_{i-1}, u_i). The mixing functions must be continuous and monotonously increasing. Moreover, in order to ensure a smooth transition between the densities, we should have

$$q_i'(x; \theta_{q_i}) + q_{i+1}'(x; \theta_{q_{i+1}}) = 1, \tag{7}$$

in the mixing zone, and each q_i, $i = 1, \cdots, k$ should satisfy

$$q_i'(u_{i-1} - \varepsilon_{i-1}; \theta_{q_i}) = 0, \quad q_i'(u_{i-1} + \varepsilon_{i-1}; \theta_{q_i}) = 1, \tag{8}$$

$$q_i'(u_i - \varepsilon_i; \theta_{q_i}) = 1 \quad \text{and} \quad q_i'(u_i + \varepsilon_i; \theta_{q_i}) = 0. \tag{9}$$

We avoid breakpoints in the density corresponding to the cdf L in Equation (5) by also requiring

$$q_i''(u_{i-1} - \varepsilon_{i-1}; \theta_{q_i}) = 0, \quad q_i''(u_{i-1} + \varepsilon_{i-1}; \theta_{q_i}) = 0, \tag{10}$$

$$q_i''(u_i - \varepsilon_i; \theta_{q_i}) = 0 \quad \text{and} \quad q_i''(u_i + \varepsilon_i; \theta_{q_i}) = 0. \tag{11}$$

One of the major problems in extreme value theory is to estimate the threshold u. We reduce this problem by defining the threshold u_i from the equation

$$g_i(u_i; \theta_{G_i}) = g_{i+1}(u_i; \theta_{G_{i+1}}). \tag{12}$$

If there are several values of u_i that satisfies the equation, we may select the supremum or infinum of these values. Equation (12) ensures that there are not large changes in $l(x; \theta_L)$ in the mixing zones. Letting the threshold be a function of the other parameters instead of a separate parameter, reduces the number of parameters. In Section 7.1 we also show that this makes the estimation of the parameters in the model more stable. Equations (8) - (9) ensure that $l(x; \theta_L)$ has continuous derivative without requiring Equation (12). If $g_i, g_{i+1}, \cdots, g_k$ have heavier and heavier tails or lighter and lighter tails, Equation (12) is particularly natural. If the tails are heavier, then κ is slightly less than 1 and if the tails are lighter, then κ is slightly larger than 1.

There are several possible definitions for q_i that satisfy the requirements given in Equations (7)-(9). We use

$$q_i(x; \theta_{q_i}) = \begin{cases} x + \varepsilon_{i-1} & x < u_{i-1} - \varepsilon_{i-1} \\ \frac{1}{2}(x + u_{i-1} + \varepsilon_{i-1}) - \frac{\varepsilon_{i-1}}{\pi} \cos(\frac{\pi(x - u_{i-1})}{2\varepsilon_{i-1}}) & u_{i-1} - \varepsilon_{i-1} \le x < u_{i-1} + \varepsilon_{i-1} \\ x & u_{i-1} + \varepsilon_{i-1} \le x < u_i - \varepsilon_i \\ \frac{1}{2}(x + u_i - \varepsilon_i) + \frac{\varepsilon_i}{\pi} \cos(\frac{\pi(x - u_i)}{2\varepsilon_i}) & u_i - \varepsilon_i \le x < u_i + \varepsilon_i \\ x - \varepsilon_i & u_i + \varepsilon_i \le x. \end{cases} \tag{13}$$

Figure 3 shows the q_i-function.

3 The transformation model in one dimension

An alternative to the model described in the previous section is to transform data to a known density like what is done in a normal score transform. We will present a method

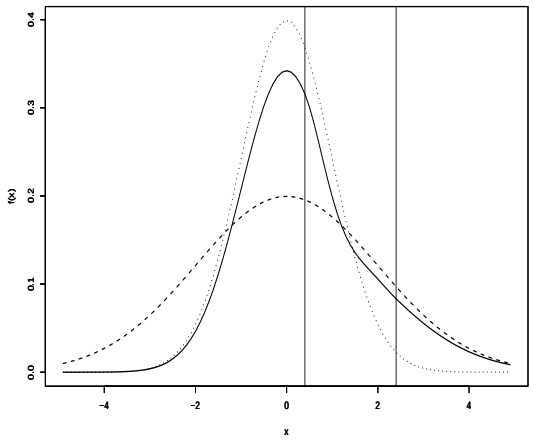

Figure 2. The two normal densities $g(x;\theta_G) \sim N(0,1)$ and $f(x;\theta_F) \sim N(0,4)$ are mixed with the mixing zone $(0.4, 2.4)$ The resulting mixture density $l(x;\theta_L)$ is given by the solid black line.

of this type where we focus on the tail behaviour and make it quite similar to the method presented in the previous section. Since we focus on the tails where there are few data points, we use a parametric transformation instead of an empiric transformation. The authors are well aware that the main argument for this model is that it is mathematically convenient and not that it is based on classical statistical principles. There are many similarities between this approach and the method described in the previous section, and the results are as good as for the other method.

Let $x \in R$ and let $G(x;\theta_G)$ be a cdf where we want to change the tail behaviour. Let $q(x;\theta_q)$ be a monotone increasing function and define the new cdf by the function

$$L(x;\theta_L) = G(q(x;\theta_q);\theta_G) \tag{14}$$

which is a valid cdf. The density is obviously

$$l(x;\theta_L) = g(q(x;\theta_q);\theta_G)q'(x;\theta_q) \tag{15}$$

where g and q' denote the derivative of G and q respectively. We get heavier tail if $|q(x;\theta_q)| < |x|$ and lighter tail if $|q(x;\theta_q)| > |x|$ for x in the tail of g. There is a large variety of alternatives for the function q. Using the same notation as in the previous section we define a mixing zone $(u-\varepsilon, u+\varepsilon)$ where we let $q(x) = x$ in the central part of the distribution and $q(x) = f(x)$ on the tail side (outside) of the mixing zone. We have found that $f(x) = u(x/u)^\beta$ gives good results. If we let G be the normal distribution, we see from Equation (15) that $l(x,\theta_L)$ get the asymptotic behaviour

$$|u|^{1-\beta}|x|^{\beta-1}\beta\exp(-|u|^{2-2\beta}|x|^{2\beta}).$$

We want $q'(x;\theta_q)$ continuously differentiable in order to get $l(x;\theta_L)$ continuous dif-

ferentiable. Therefore, we propose the following function

$$q(x;\theta_q) = \begin{cases} x & x < u-\varepsilon \\ \frac{c(x-u-\varepsilon)^{k_1}}{k_1(k_1-1)} + \frac{d(x-u-\varepsilon)^{k_2}}{k_2(k_2-1)} + x & u-\varepsilon \le x < u+\varepsilon \\ u(x/u)^{\beta} & u+\varepsilon \le x \end{cases} \tag{16}$$

where ε is a fixed constant determining the length of the transition zone and c, d, k_1 and k_2 are chosen in order to get $l(x;\theta_L)$ smooth. We have the following equations

$$d = \frac{f'(u+\varepsilon) - 1 - \frac{2\varepsilon}{k_1-1} f''(u+\varepsilon)}{(\frac{1}{k_2-1} - \frac{1}{k_1-1})(2\varepsilon)^{k_2-1}} \tag{17}$$

$$c = \frac{f''(u+\varepsilon) - d(2\varepsilon)^{k_2-2}}{(2\varepsilon)^{k_1-2}} \tag{18}$$

in order to get $q(x;\theta_q)$ twice continuously differentiable where the constants satisfy $k_1 > 3$ and $k_2 > 2$. The function q is smooth with $k_1 = 4$ and $k_2 = 3$. See Figure 4 for an illustration of a $q(x;\theta_q)$ function and the corresponding density. The parameters in the model, θ_G, β and u should be found from data. The length of the mixing zone should be set as a constant or connected to the variance of $G(x)$ since it is difficult to estimate this from data. Similarly to combine several components in (5), we may have several mixing zones in (16).

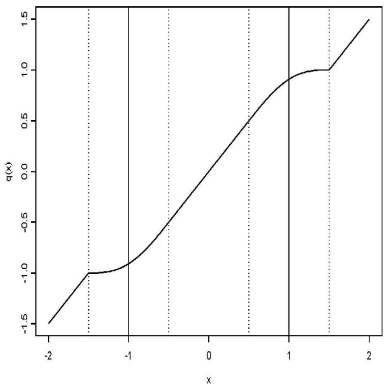

Figure 3. The transition function $q_i(x;\theta_{q_i})$ with $u_{i-1} = -1$, $u_i = 1$, and $\varepsilon_{i-1} = \varepsilon_i = 0.5$. The thresholds are given by vertical solid lines and the mixing zones are delimited by the vertical dotted lines. Note that the transition function maps the interval $(u_{i-1}-\varepsilon_{i-1}, u_i+\varepsilon_i)$ onto (u_{i-1}, u_i).

4 Comparison with other one dimensional models

The traditional method in extreme value modelling is to fix one or two thresholds, and use only values further out in the tails than these thresholds for parameter estimation. By using Equation (5) with three components, fixing the u_i's in advance, and letting $\varepsilon_i = 0$ for all components, the cdf-method proposed is identical with the traditional one.

Equation (1) bears resemblance with the mixed model

$$l_2(x;\theta_l) = \frac{1}{Z(\theta_l)}(p(x;\theta_p)g(x;\theta_G) + (1-p(x;\theta_p))f(x;\theta_F)), \tag{19}$$

proposed by (12). Here f and g are the densities of F and G respectively, and $Z(\theta_l)$ is an integrational constant. The integrational constant is generally found by numerical integration, which is likely to make the maximum likelihood estimation unstable and computationally expensive. By mixing the cdfs instead of the densities, we often get analytic expressions for the integrational constant, and the parameter estimation becomes more stable. Otherwise, it makes little difference whether the mixing is based on the densities or the cdfs. However, the increased efficiency of our model as compared to Equation (19) makes it more manageable to use in several dimensions.

In (3) it is proposed to use the cdf

$$L(x;\theta_L) = \begin{cases} G(x;\theta_G) & x < u \\ G(u;\theta_G) + (1 - G(u;\theta_G))F(x;\theta_F) & x \geq u \end{cases} \tag{20}$$

This is identical with the mixing model (1) if we assume there is no mixing zone, i.e. $\varepsilon = 0$ and $f(x)$ is replaced with $cf(x)$ for a constant c such that $\kappa = 1$. By introducing a mixing zone we obtain a continuous density. As shown in the example, this does not imply an increase in the number of parameters that need to be estimated. It only makes the result more plausible and offers more stable estimation of the parameters since the density is smooth.

In the recent preprint (5) [1] another model of the same type is proposed in the context of neural networks. It is proposed to mix a normal distribution with a GPD distribution with the restriction on the parameters such that the density and the derivative of the density are the same on both sides of the thresholds. This gives a smooth density without a mixing zone. Their model has one parameter less than the GPD-normal model presented in this paper since the requirement of a continuous derivative of the density eliminates the scaling parameter in the GPD density. This implies that variance of the normal distribution is connected to the scaling of the tail in the mixed model. It is not easy to generalize their model to other densities than GPD or several dimensions.

5 The cdf-model in several dimensions

This Section shows how to generalise the cdf-mixing model to higher dimensions. Let $\boldsymbol{x} \in R^n$. We assume the state space is divided into disjoint regions $A_i; i = 1, \ldots k$, with corresponding truncated cdfs and densities, G_i and g_i, respectively. We further assume that $g_i(\boldsymbol{x};\theta_{G_i}) = 0$, except for $\boldsymbol{x} \in A_i$. We define a mixing zone M_i at the border between A_i and A_{i+1}, and require that the corresponding transition function q_i is monotonously increasing and maps $A_i \cup M_{i-1} \cup M_i$ onto A_i. We define the multivariate cdf as

$$L(\boldsymbol{x};\theta_L) = \kappa \sum_{i=1}^{k} G_i(\boldsymbol{q}_i(\boldsymbol{x};\theta_{q_i});\theta_{G_i}), \tag{21}$$

1. Added in the referee process

where κ is defined such that $L(\boldsymbol{x};\theta_L)$ is a valid cdf. The density $l(\boldsymbol{x};\theta_L)$ corresponding to the cdf in Equation (21) is

$$l(\boldsymbol{x};\theta_L) = \begin{cases} \kappa\, g_i(\boldsymbol{x};\theta_{G_i}) & \text{if } \boldsymbol{x} \in A_i \setminus (M_{i-1} \cup M_i) \\ \kappa(\frac{\partial^n G_i(\boldsymbol{q}_i(\boldsymbol{x};\theta_{q_i});\theta_{G_i})}{\partial x_1 \cdots \partial x_n} + & \\ \qquad \frac{\partial^n G_{i+1}(\boldsymbol{q}_{i+1}(\boldsymbol{x};\theta_{q_{i+1}});\theta_{G_{i+1}})}{\partial x_1 \cdots \partial x_n}) & \text{if } \boldsymbol{x} \in M_i. \end{cases} \tag{22}$$

The regions A_i, mixing zones M_i and transition functions $\boldsymbol{q}_i(\boldsymbol{x};\theta_{q_i})$ may be chosen in several different ways. We present two different alternatives. In Section 5.1 we mix densities along one axis only and in Section 5.2 we mix distributions radially.

5.1 Mixing along one axis

First, we mix densities along one axis only. We let the density g_i vanish except between two parallel planes that we for simplicity assume are normal to the x_1-axis. A vector $\boldsymbol{x}$ is said to be inside a mixing zone if $u_i - \varepsilon_i < x_1 < u_i + \varepsilon_i$, and between two mixing zones if $u_{i-1} + \varepsilon_{i-1} < x_1 < u_i - \varepsilon_i$. Figure 5 illustrates this alternative for R^2 with three components. For the general case R^n, assume we mix in the direction of the vector $\boldsymbol{e}_1 = (1, 0, 0, \cdots, 0)$. Define the transition function q_i^M as

$$\boldsymbol{q}_i^M(\boldsymbol{x};\theta_{q_i}) = \boldsymbol{x} + (q_i(x_1;\theta_{q_i}) - x_1)\ \boldsymbol{e}_1, \tag{23}$$

where q_i is defined in Equation (13).

In this case, the density $l(\boldsymbol{x};\theta_L)$ is equal to

$$l(\boldsymbol{x};\theta_L) = \begin{cases} \kappa\, g_i(\boldsymbol{x};\theta_{G_i}) & \text{if } \boldsymbol{x} \in A_i \setminus (M_{i-1} \cup M_i) \\ \kappa\,(g_i(\boldsymbol{q}_i(\boldsymbol{x};\theta_{q_i});\theta_{G_i})\, \frac{\partial \boldsymbol{q}_i(\boldsymbol{x};\theta_{q_i})}{\partial x_1} + & \\ \qquad g_{i+1}(\boldsymbol{q}_{i+1}(\boldsymbol{x};\theta_{q_{i+1}});\theta_{G_{i+1}})\, \frac{\partial \boldsymbol{q}_{i+1}(\boldsymbol{x};\theta_{q_{i+1}})}{\partial x_1}) & \text{if } \boldsymbol{x} \in M_i. \end{cases} \tag{24}$$

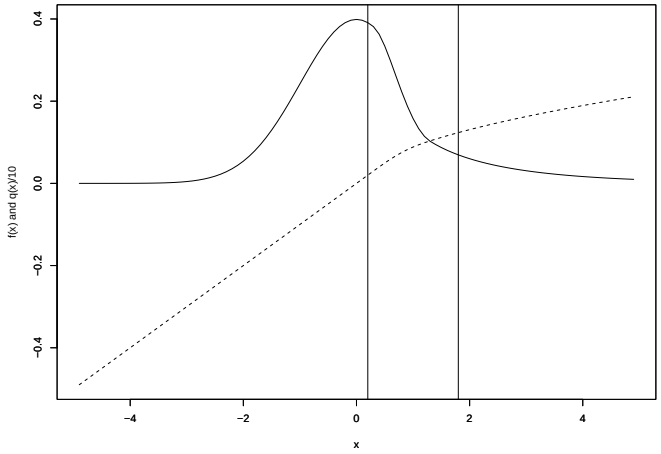

Figure 4. The density in the transformed normal model with $g(x;\theta_G) \sim N(0,1)$. The figure also shows $q(x,\theta_q)/10$ with mixing zone $(0.3, 1.3)$ and $q(x,\theta_q) = (x/u)^{0.5}$

5.2 Mixing radially

If we have only two components, and want to have one distribution in the centre and another distribution for more rare observations, the following approach may seem more natural. This example is developed further in Section 7.3. Assume that we have two distributions G and F, and that $G(\boldsymbol{x};\theta_G)$ has positive density for

$$r(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{\mu}) < u,$$

for a constant μ and $F(\boldsymbol{x};\theta_F)$ has positive density for $r(\boldsymbol{x}) > u$.

Figure 6 shows an example of the case with two dimensions, $\boldsymbol{\mu}$ equal to the zero vector and $\boldsymbol{\Sigma}$ diagonal. The overall cdf L is determined by G only for $r(\boldsymbol{x}) < u-\varepsilon$ (the area $A_G \setminus M$), by F for $r(\boldsymbol{x}) > u+\varepsilon$ (the area $A_F \setminus M$), and by both distributions in the mixing zone $u-\varepsilon \le r(\boldsymbol{x}) < u+\varepsilon$ (the area M).

For this mixing problem, the cdf in Equation (21) can be simplified to

$$L(\boldsymbol{x};\theta_L) = \kappa(G(\boldsymbol{q}^R(\boldsymbol{x};\theta_q);\theta_G) + F(\boldsymbol{p}^R(\boldsymbol{x};\theta_p);\theta_F)). \tag{25}$$

In this case, the following choices of $\boldsymbol{q}^R$ and $\boldsymbol{p}^R$ seem appropriate;

$$\boldsymbol{q}^R(\boldsymbol{x};\theta_q) = (\boldsymbol{x}-\boldsymbol{\mu})\frac{q(r(\boldsymbol{x});\theta_q)}{r(\boldsymbol{x})} + \boldsymbol{\mu} \tag{26}$$

and

$$\boldsymbol{p}^R(\boldsymbol{x};\theta_p) = (\boldsymbol{x}-\boldsymbol{\mu})\frac{p(r(\boldsymbol{x});\theta_p)}{r(\boldsymbol{x})} + \boldsymbol{\mu}, \tag{27}$$

where the functions p and q are given by Equations (3) and (4), respectively. The function $\boldsymbol{q}^R$ maps the set $A_G \cup M = \{\boldsymbol{x}; r(\boldsymbol{x}) < u+\varepsilon\}$ onto $A_G = \{\boldsymbol{x}; r(\boldsymbol{x}) < u\}$ and $\boldsymbol{p}^R$ maps the set $A_F \cup M = \{\boldsymbol{x}; r(\boldsymbol{x}) > u-\varepsilon\}$ onto $A_F = \{\boldsymbol{x}; r(\boldsymbol{x}) > u\}$.

In this case, the density $l(\boldsymbol{x};\theta_L)$ is equal to

$$l(\boldsymbol{x};\theta_L) = \begin{cases} \kappa\, g(\boldsymbol{x};\theta_G) & \text{if } \boldsymbol{x} \in A_G \setminus M \\ \kappa(\frac{\partial^n G(\boldsymbol{q}^R(\boldsymbol{x};\theta_q);\theta_G)}{\partial x_1 \cdots \partial x_n} + \frac{\partial^n F(\boldsymbol{p}^R(\boldsymbol{x};\theta_p);\theta_F)}{\partial x_1 \cdots \partial x_n}) & \text{if } \boldsymbol{x} \in M \\ \kappa\, f(\boldsymbol{x};\theta_F) & \text{if } \boldsymbol{x} \in A_F \setminus M \end{cases} \tag{28}$$

The density in the mixing zone is quite complicated to derive. However, if $g(\boldsymbol{x};\theta_G) \approx f(\boldsymbol{x};\theta_F)$ when $r(\boldsymbol{x};\theta_q) = u$, and the mixing zone is not too wide, we may use the approximation

$$l(\boldsymbol{x};\theta_L) \approx \kappa\,(g(\boldsymbol{q}^R(\boldsymbol{x};\theta_q);\theta_G)\, q'(r(\boldsymbol{x});\theta_q) + \, f(\boldsymbol{p}^R(\boldsymbol{x};\theta_p);\theta_F)p'(r(\boldsymbol{x});\theta_p)).$$

If we want G and F to be truncated multivariate Gaussian distributions, the truncation becomes much easier if both distributions have the same expectation vector and the covariance matrices only differ by a scaling factor.

6 The transformation model in several dimensions

It is easy to combine the univariate transformation (16) with multivariate distributions. The disadvantage is that the same multivariate distribution dominates in the entire domain. It may therefore be needed to combine this approach with a Copula, see (11). It is

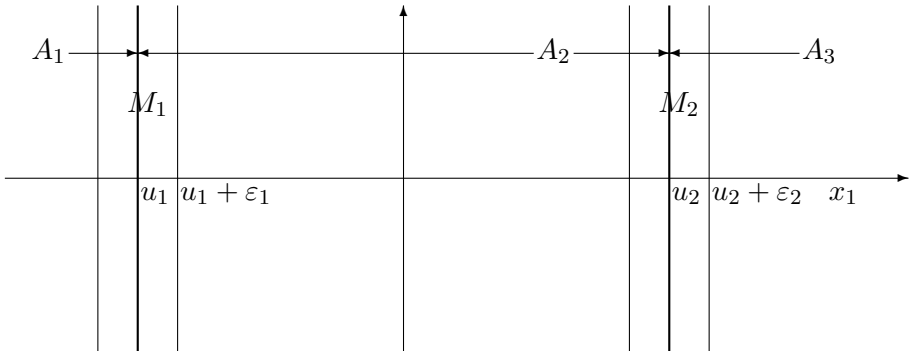

Figure 5. The cdf L is determined by only G_i in region $A_i \setminus (M_{i-1} \cup M_i)$, and by both G_i and G_{i+1} in the mixing zones M_i. The threshold u_i is in the center of the mixing zone M_i.

possible to have other transformations than univariate. But this leads to very complex formulas for the density.

The asymptotic properties in multivariate distributions are much more complex than in one dimension, see (13). It is possible, however, to obtain the different asymptotic properties by an appropriate choice of univariate transformation.

7 Numerical applications

We illustrate the proposed models on both synthetic data and real financial data. All parameter estimation is performed by maximizing the log-likelihood. The maximisation is done numerically using the routine *nlminb* in R. This seems to work very well in all tests performed.

7.1 Synthetic data, 1D

In this section, we test three different models. We generate synthetic data from one of the models and then estimate parameters and quantiles in all the proposed models. The first two models are mixtures of the form defined by Equation (5) with $k = 3$ and the normal distribution as the central distribution. The first model has the generalized Pareto Distribution (GPD) distribution in both tails and the second has the Weibull distribution in both tails. The GPD cdf is

$$G(x; \xi, \sigma) = 1 - (1 + \frac{\xi x}{\sigma})^{-\frac{1}{\xi}}$$

assuming $\xi > 0$, $\sigma > 0$ and $x > 0$, and the cdf in the Weibull distribution is

$$G(x; \beta, \lambda) = 1 - \exp(-(x\lambda)^{\beta})$$

for $\beta > 0$, $\lambda > 0$ and $x > 0$. This gives 10 parameters in the model described by Equation (5), 2 in each of the three distributions in addition to u_1, u_2, ε_1 and ε_2. The thresholds u_1 and u_2 are determined from Equation (12). This reduces the number of parameters

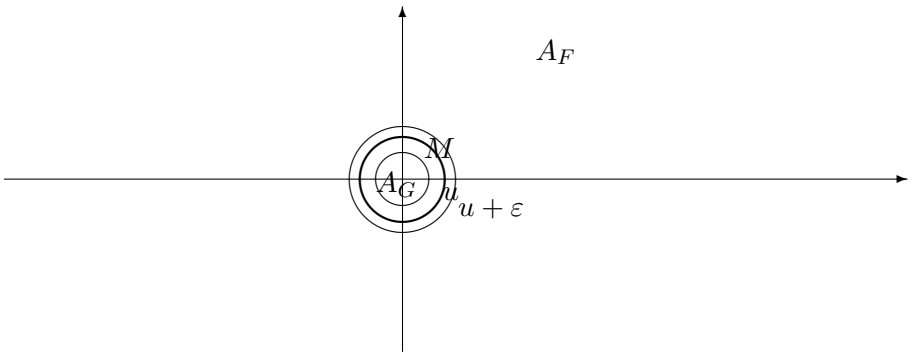

Figure 6. The bold circle corresponds to the threshold u. The areas A_G and A_F are inside and outside this circle, respectively. The cdf L is determined by G only in the area $A_G \setminus M$, by F only in the area $A_F \setminus M$, and by both G and F in the area M.

and also gives smoother distributions. The length of the transition intervals, $2\varepsilon_i$, are not critical in the estimation and not easy to estimate. Hence, we set $\varepsilon_i = \sigma_2$ for $i = 1$ and $i = 2$ where σ_2 is the standard deviation of the central normal distribution. In all the tests we set the expectation in the central density equal to 0 leaving 5 unknown parameters.

The third model we test is the transformation model described in Equation (14) with the polynomial function for q given in Equation (16). Also here we set $\varepsilon_i = \sigma_2$ for $i = 1$ and $i = 2$ in order to make the same choice as in the previous model. We denote σ_2 as the standard deviation of the normal distribution in order to use the same symbol with the corresponding parameter in the other models. Also here there are 5 unknown parameters.

In the tests we simulate $m = 1000$ samples from each of the models in turn and then estimate parameters in all the models. In addition, we also test with $10m = 10.000$ samples with the same model as is used in the simulation. This is repeated $k = 500$ where we estimate the parameters in each of the three models, the corresponding $0.001, 0.01, 0.99$ and 0.999 quantiles, the difference in L_1 norm, and the log-likelihood value. The tables give the average and the standard deviation of the estimated parameters/quantiles/values. The difference in L_1 norm is estimated by dividing the state space into 100 intervals. The difference in L_1 norm is half the sum of the absolute value of the difference in probability between the estimated and the original density in each interval. Simulation from a distribution where the density differs in the L_1 norm by 0.01 compared to the correct density, implies that 0.01 of the samples are from a wrong distribution.

The parameters are estimated by maximizing the log-likelihood. The model with GPD distributions in the tails is left with the following 5 parameters: $\xi_1, \sigma_1, \sigma_2, \xi_3, \sigma_3$ and the model with Weibull distribution in the tails has the parameters $\beta_1, \lambda_1, \sigma_2, \beta_3, \lambda_3$. In these two models the thresholds u_1 and u_2 are found from Equation (12) based on the other parameters. The model with transformations has the parameters $u_1, u_2, \beta_1, \sigma_2, \beta_3$.

Table 1. Average of estimated parameters and standard deviation of the estimates when simulated using a GPD-Normal-GPD distribution. $10m$ indicates the use of 10.000 data points in the sample.

	θ_1	θ_2	σ_2	θ_4	θ_5	u_1	u_2
Simulation GPD-N-GPD	0.300	0.400	1.000	0.200	0.400	-2.166	2.415
Est. GPD-N-GPD $10m$	0.299	0.401	1.000	0.200	0.400	-2.165	2.417
St.dev. GPD-N-GPD $10m$	0.019	0.017	0.010	0.021	0.020	0.067	0.081
Est. GPD-N-GPD	0.295	0.404	0.996	0.195	0.403	-2.131	2.390
St. dev. GPD-N-GPD	0.068	0.058	0.035	0.063	0.061	0.260	0.311
Est. Weibull-N-Weibull	0.511	0.211	1.000	0.606	0.253	-2.391	2.513
St.dev. Weibull-N-Weibull	0.058	0.061	0.031	0.061	0.059	0.170	0.315
Est. transf. N	-	0.410	1.060	-	0.489	-1.755	1.914
St.dev. transf. N	-	0.069	0.031	-	0.104	0.190	0.251

Table 2. Quantiles, L_1-error and log-likelihood using GPD-N-GPD.

	$q_{0.001}$	$q_{0.01}$	$q_{0.99}$	$q_{0.999}$	L_1	loglikeh.
Simulation GPD-N-GPD	-9.15	-3.92	3.00	5.91	-	-1583.5
Est. GPD-N-GPD $10m$	-9.14	-3.92	3.00	5.91	0.005	-15833.0
St.dev. GPD-N-GPD $10m$	0.54	0.12	0.08	0.31	0.002	99.5
Est. GPD-N-GPD	-9.20	-3.90	2.99	5.88	0.016	-1581.2
St. dev. GPD-N-GPD	1.96	0.41	0.24	0.96	0.007	30.5
Est. Weibull-N-Weibull	-9.34	-4.12	3.10	6.15	0.015	-1583.6
St.dev. Weibull-N-Weibull	1.70	0.49	0.28	0.85	0.008	33.2
Est. transf. N	-8.41	-4.08	3.22	5.95	0.025	-1583.8
St.dev. transf. N	1.66	0.50	0.32	1.01	0.006	33.4

Table 3. Average of estimated parameters and standard deviation of the estimates when simulated using a Weibull-Normal-Weibull distribution.

	θ_1	θ_2	σ_2	θ_4	θ_5	u_1	u_2
Simulation Weibull-N-Weibull	0.500	0.200	1.000	0.600	0.250	-2.394	2.487
Est. Weibull-N-Weibull $10m$	0.502	0.202	0.999	0.600	0.251	-2.388	2.483
St.dev. Weibull-N-Weibull $10m$	0.019	0.021	0.010	0.021	0.021	0.052	0.063
Est. Weibull-N-Weibull	0.511	0.211	1.000	0.606	0.253	-2.390	2.514
St.dev. Weibull-N-Weibull	0.058	0.061	0.031	0.061	0.059	0.170	0.315
Est. GPD-N-GPD	0.383	0.331	1.000	0.333	0.298	-2.314	2.590
St.dev. GPD-N-GPD	0.091	0.071	0.035	0.125	0.111	0.265	0.440
Est. transf. N	-	0.410	1.060	-	0.489	1.755	1.914
St.dev. transf. N	-	0.069	0.031	-	0.104	0.190	0.251

Table 4. Quantiles, L_1-error and log-likelihood using Weibull-Normal-Weibull.

	$q_{0.001}$	$q_{0.01}$	$q_{0.99}$	$q_{0.999}$	L_1	loglikeh.
Simulation Weibull-N-Weibull	-9.45	-4.18	3.12	6.21	-	-1585.7
Est. Weibull-N-Weibull $10m$	-9.45	-4.18	3.13	6.22	0.005	- 15874.9
St.dev. Weibull-N-Weibull $10m$	0.51	0.14	0.08	0.28	0.002	99.4
Est. Weibull-N-Weibull	-9.34	-4.12	3.10	6.15	0.015	-1583.6
St.dev. Weibull-N-Weibull	1.70	0.49	0.278	0.848	0.008	33.2
Est. GPD-N-GPD	-11.41	-4.08	3.11	7.82	0.017	-1584.7
St.dev. GPD-N-GPD	3.02	0.51	0.34	2.18	0.008	33.4
Est. transf. N	-8.41	-4.08	3.22	5.95	0.025	-1583.8
St.dev. transf. N	1.66	0.50	0.32	1.01	0.006	33.4

Table 5. Average of estimated parameters and standard deviation of the estimates when simulated using a transformed normal distribution. The transformed distribution uses u_i as one of the five estimated parameters (instead of θ_1 and θ_4) while u_i depend on the other parameters in the other models.

	θ_1	θ_2	σ_2	θ_4	θ_5	u_1	u_2
Simulation transf. N	-	0.450	1.000	-	0.600	-1.500	1.500
Est. transf. N $10m$	-	0.450	1.000	-	0.600	-1.498	1.500
St.dev. transf. N $10m$	-	0.015	0.008	-	0.021	0.034	0.052
Est. transf. N	-	0.458	0.996	-	0.605	1.484	1.489
St.dev. transf. N	-	0.056	0.028	-	0.075	0.118	0.198
Est. GPD-N-GPD	0.343	0.360	0.928	0.241	0.387	-1.985	2.001
St.dev. GPD-N-GPD	0.075	0.056	0.038	0.116	0.106	0.221	0.467
Est. Weibull-N-Weibull	0.554	0.256	0.926	0.685	0.339	-2.074	2.028
St.dev. Weibull-N-Weibull	0.055	0.059	0.033	0.078	0.080	0.159	0.298

In Tables 1, 3, and 5, the parameters are denoted $\theta_1, \theta_2,\ \sigma_2,\ \theta_3, \theta_4,\ u_1$, and u_2 where $\theta_1, \theta_2, \theta_3, \theta_4$, have a different interpretation in the different models. The results of the simulations are shown in Tables 1 - 6. We see that there are quite good estimates for all the parameters. The standard deviation is comparable with estimation of σ in the normal distribution with the same sample size. Only 1-4% of the $m = 1000$ samples are from the tails and in the mixing zone these are mixed with data points from the central distribution. The standard deviations for the different parameters have about the same size including σ_2, the standard deviation in the central distribution. The estimates for the thresholds u_i have a larger standard deviation than the other parameters, indicating that the threshold should be determined implicitly by the other parameters. We have also tried to estimate these simultaneously with the other parameters. Then all parameters have larger uncertainty. Also the quantiles and the density measured in the L_1 error are quite close to the quantiles and density that were used in the simulation. The uncertainty in the $q_{0.001}$ and $q_{0.999}$ quantiles in the GPD distribution is larger than for the other distributions since this has heavier tails than the two other distributions. As expected, we always get better estimates when we fit the same model as is used in the simulation and when we increase the number of samples to $10m = 10.000$.

7.2 Financial data, 1D

We want to illustrate the use of the models on real data and have selected three stock market indices; the European, the American and the Japanese. It is not the ambition to make the best possible model for these data. That would require a more complex model including for example handling of stochastic volatility which is outside the topic of this paper. We first study each of the stock markets independently using the methods from Sections 2 and 3, and then we model the portfolio using the methods described in Section 5 and 6. The results from the multivariate tests are given in Section 7.4.

We assume that the three stock markets can be represented by the corresponding Morgan Stanley (MSCI) price indices in local currency neglecting the currency risk in the portfolio. We use index data from the period 01.01.1987 to 28.05.2002 for model estimation. This period corresponds to $m = 4065$ observations. The return series are shown in Figure 7. Let $x_{i,t}$ denote the original indices, $i = 1, 2, 3$. We use the data $r_{i,t} = log(x_{i+1,t}/x_{i,t}) - \mu_i$ where μ_i is determined such that $\sum_i r_{i,t} = 0$.

Figure 8 shows normal QQ-plots for the standardised logarithmic residuals $r_{i,t}/\sigma_i$ for each of the three markets where σ_i is the standard deviation of $r_{i,t}$. As can be seen from the figure, all distributions are doubly heavy-tailed. Moreover, they are clearly skewed, having one tail heavier than the other. This motivates for the use of a mixture distribution with three components, one for the left tail, one for the centre of the distribution, and one for the right tail, respectively. Hence, we use the distribution given in Equation (5) with three components. Exactly as in Section 7.1 we test with the GPD and the Weibull density in both tails and with the normal distribution in the centre. In addition, we test with the transformed normal distribution given in Equation (14).

In all cases we have the same 5 parameters as described in Section 7.1 that are estim-

Table 6. Quantiles, L_1-error and log-likelihood using a transformed normal distribution.

	$q_{0.001}$	$q_{0.01}$	$q_{0.99}$	$q_{0.999}$	L_1	loglikeh.
Simulation transf. N	-7.51	-4.00	3.13	5.02	-	-1548.2
Est. transf. N $10m$	-7.51	-4.00	3.13	5.02	0.005	-15483.9
St.dev. transf. N $10m$	0.33	0.11	0.07	0.17	0.002	95.8
Est. transf. N	-7.52	-3.99	3.11	5.02	0.018	-1545.7
St.dev. transf. N	1.17	0.38	0.20	0.54	0.008	31.1
Est. GPD-N-GPD	-10.2	-3.94	3.12	6.79	0.028	- 1550.2
St.dev. GPD-N-GPD	2.31	0.42	0.29	1.76	0.007	31.5
Est. Weibull-N-Weibull	-8.44	-3.98	3.10	5.71	0.025	1548.0
St.dev. Weibull-N-Weibull	1.27	0.40	0.24	0.63	0.006	31.3

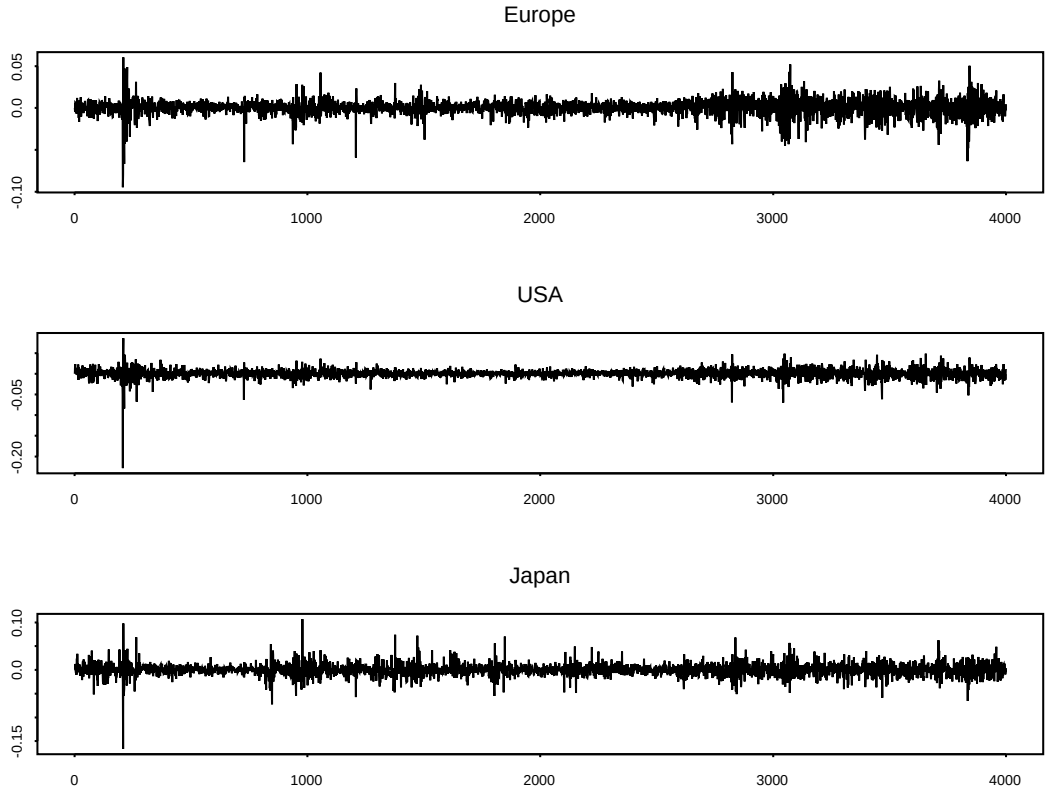

Figure 7. European, American and Japanese geometric return series for the period 01.01.1987 – 28.05.2002.

Table 7. Parameter estimates for the GPD-N-GPD mixture model using residuals. We have $\varepsilon_i = \sigma_2$.

Parameter	Europe	USA	Japan
ξ_1	0.266	0.156	0.432
σ_1	0.00395	0.00564	0.00831
σ_2	0.00409	0.00320	0.00377
ξ_3	0.0735	0.198	0.0724
σ_3	0.00498	0.00680	0.00801
log-likelihood	13609	13227	12339

ated by maximizing the likelihood

$$\prod_{t=1}^{m} l(r_{i,t};\theta_L), \tag{29}$$

where $l(r;\theta_L)$ is given by Equations (6) and (15) respectively. The results are shown for each of the models in Tables 7 - 9. For the transformed normal model, three of the threshold values ended up equal to the limit ± 0.005. This indicates that we get best fit with using the transformation for all negative/positive values. The parameter σ_2 is then only used to set the density for $x = 0$ and influences the joint density in the mixing zones $(-0.01, 0)$ and $(0, 0.01)$. We get best fit using GPD-N-GPD, then Weibull-N-Weibull and then transformed normal density. The estimated quantiles are given in Table 17 together with the corresponding results from the multivariate distributions.

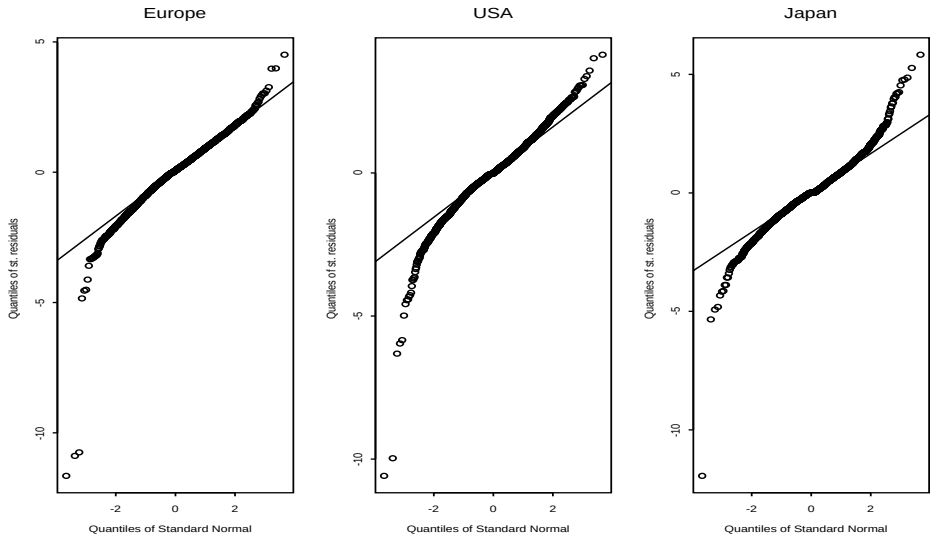

Figure 8. QQ-plots of the standardized residuals fitted against the normal distribution.

Table 8. Parameter estimates for the Weibull-N-Weibull mixture model using residuals. We have $\varepsilon_i = \sigma_2$.

Parameter	Europe	USA	Japan
β_1	0.686	0.801	0.987
λ_1	0.00357	0.00546	0.00873
σ_2	0.00454	0.00344	0.00368
β_3	0.885	0.977	0.948
λ_3	0.00464	0.00664	0.00845
log-likelihood	13612	13223	12334

7.2.1 Parameter estimates for NIG distribution

We have chosen to compare the results using the mixture model with the ones obtained using the Normal Inverse Gaussian (NIG) distribution. This distribution has been used for financial applications, both as the conditional distribution of a GARCH-model, see (2), and as the unconditional return distribution, see (4). The paper (24) compares different probability distributions for the innovations in one-dimensional processes. The authors consider a NIG distribution, a skewed Student's t-distribution and a non-parametric kernel approximation. They report that the NIG distribution provides the best fit overall for the models considered.

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

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

where $\delta > 0$ and $0 < |\beta| \leq \alpha$. In the above expression, K_1 is the modified Bessel function of the third kind of order 1, see (1). The parameters μ and δ determine the location and scale, respectively, while α and β control the shape of the density. In particular, $\beta = 0$ corresponds to a symmetric distribution.

The parameters of the NIG distribution are estimated using the EM-algorithm described in (15), with the moment estimates as starting values. The parameter estimates are shown in Table 10.

7.2.2 Comparing the models in the tails

We have used graphical logarithmic left and right hand tail tests to examine the fit in the tails. The graphical tests were performed as follows. Let $(X_{(1)}, ..., X_{(N)})$ denote the order statistic of the historical data, and $\hat{F}(x)$ the estimated cumulative distribution function of the fitted distribution. For the NIG distribution this is calculated using the method described in (19). A plot of $\log(\hat{F}(X_{(t)}))$ against $X_{(t)}$ superimposed on a plot of $\log\left(t/(N+1)\right)$ against $X_{(t)}$ shows the left tail fit for the fitted distribution, and a plot of $\log(1-\hat{F}(X_{(t)}))$ against $X_{(t)}$, superimposed on a plot of $\log\left((N+1-t)/(N+1)\right)$, the right tail fit.

Figure 9 shows the plots. All the models give quite similar results but the mixture

Table 9. Parameter estimates for the transformed normal model using residuals. We have $\varepsilon_i = 0.005$.

Parameter	Europe	USA	Japan
u_1	-0.007	-0.006	- 0.005
u_2	0.0085	0.005	0.005
β_1	0.550	0.565	0.665
σ_2	0.00674	0.00659	0.00778
β_3	0.583	0.665	0.655
log-likelihood	13591	13188	12300

Table 10. Parameter estimates for NIG distributions for logarithmic residuals.

Parameter	Europe	USA	Japan
μ	0.00134	0.000746	-0.000308
δ	0.00678	0.00733	0.00945
α	78.3	67.2	57.0
β	-12.6	-3.85	1.17

distribution with GPD tails followed by the NIG distribution seems slightly better than the others.

7.3 Synthetic data, 3D

Also in 3D we simulate data from a multivariate distribution and then estimate the parameters from the simulated data. We test mixing five multivariate normal distributions, and the multivariate normal distribution of univariate transformed variables. In both tests we assume that the expectation is zero.

The mixture of five multivariate normal distributions is a further development of the model described in Section 5.2. See Figure 10 for an illustration in 2D. All the multivariate normal distributions have the same parameters except a scaling of the correlation matrix. The multivariate normal distributions are truncated by two ellipsoids that are outside each other and a plane that is normal to the largest axis of the ellipsoid. At the ellipsoid all the multivariate normal distributions have constant density and the densities are symmetric at both sides of the plane. In the centre there is a multivariate normal distribution truncated by the inner ellipsoid. Between the two ellipsoids there are two multivariate normal distributions, one on each side of the plane. Outside the outer ellipsoid there are two multivariate normal distributions, also one on each side of the plane. Since the ellipsoids are determined by the correlation matrix of the multivariate normal distributions, it is possible to make exact truncation by the use of the Chi-square distribution. In the mixing zone around each ellipsoid it is possible to compensate for the volume effect of the transformation making it possible to calculate the mixed density exactly.

In this mixing model we are able to handle heavier tails than the normal distribution and different properties for extreme values in the positive and negative direction determ-

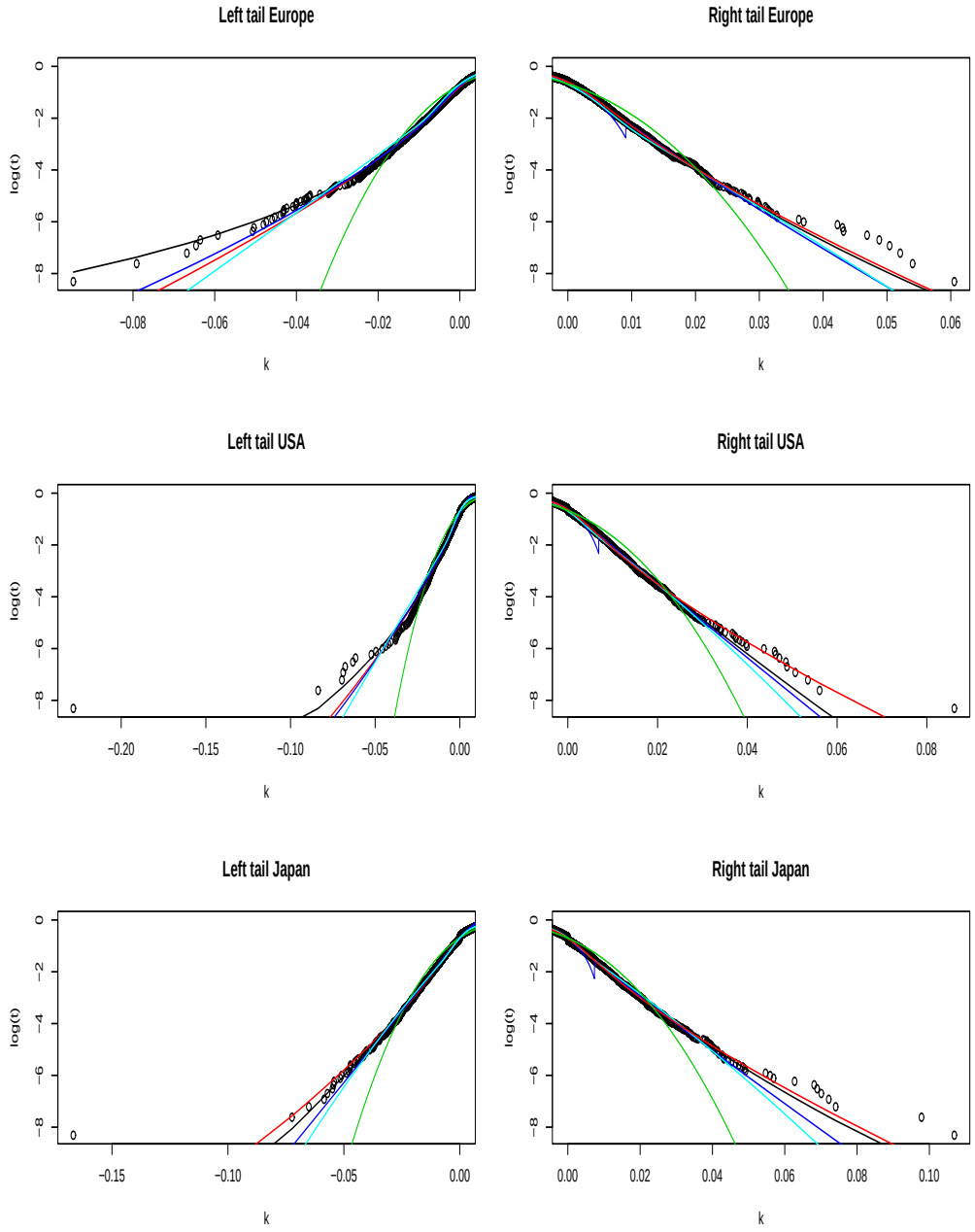

Figure 9. Plot of the tail behaviour in the five models. The circles correspond to the empirical data, the light-blue line to the mixture distribution with Weibull tails, the black line to the mixture distribution with GPD tail, the red to the NIG distribution and the blue line to the transformed normal. For reference, a normal fit is also included shown in green.

ined by the plane normal to the main axis of the correlation ellipsoid. The two truncation values are determined from an equation similar to (12). The size of the ellipsoids are described by the parameters u_1 and u_2, which is the intersection between the positive x_1 axis and the two ellipsoids. The inner truncation values u_1 is at the ellipsoid where the value of the estimated normal distribution in the centre is equal to average of the values of the estimated densities outside this ellipsoid. The outer truncation value u_2 are determined where the average of the values of the estimated densities inside and outside this ellipsoid are equal. The integrational constant κ is set such that the mixed density is a proper density with integral equal to 1. In model we have the following parameters:

- the standard deviations in the normal distribution, σ_i, 3 variables.
- the correlation between the three variables, ρ_i, 3 variables.
- the parameters $F_{i,-}, F_{i,+}$ for $i = 1, 2$ scaling the correlation matrix between the two ellipsoids and outside the outer ellipsoid, on both sides of the truncation plane giving 4 parameters.

In total there are 10 parameters that are estimated by maximum likelihood. In addition, we have the length of the mixing zones that is not as critical as the above parameters. The mixing zone is $(1 - \varepsilon, 1 + \varepsilon)$ relative to the distance to the ellipsoid from the origin where $\varepsilon = 0.06$. The 3D simulated data gives the result in Table 11 based on 100 samples with 1000 data points in each sample. The estimated values are close to the values used in the simulation and the standard deviations are reasonable.

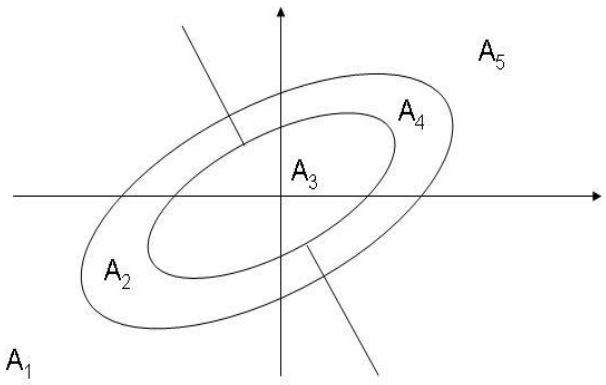

Figure 10. The areas of the five multinormal densities; inside inner ellipsoid, between the ellipsoid on each side of plane and outside outer ellipsoid on both side of plane.

We have combined the multivariate normal model with the same univariate transformations as we used in 1D in Section 7.1. The method is described in Sections 3 and 6. We have the following parameters

- the standard deviations in the normal distribution σ_i, 3 variables.
- the correlation between the three variables, ρ_i, 3 variables.

Table 11. The parameter in the mixture of five multivariate normal distributions. The correlation is listed in the order (1,2), (2,3) and (1,3). We first show the parameters that are used in the simulation and then the average of the estimation based on 100 realizations with 1000 samples. The standard deviation of the estimated parameters are shown in parenthesis.

i	σ_i	ρ_i	$F_{i,-}$	$F_{i,+}$	likelihood
Simulation					
1	1.00	0.71	1.40	1.40	-6071
2	1.00	0.82	3.00	3.00	-
3	1.00	0.58	-	-	-
Estimation					
1	0.99 (.04)	0.70 (.02)	1.40 (.07)	1.40 (.07)	- 6063
2	0.99 (.04)	0.81 (.02)	2.99 (.05)	2.99 (.05)	-
3	0.99 (.04)	0.57 (.03)	-	-	-

- the thresholds u_1 and u_2 where the univariate transformation is not the identity for $x_i < u_1$ and $x_i > u_2$ for $i = 1, 2, 3$, giving 2 variables

- the transformation parameters $\beta_{i,j}$ for both tails in each of the three variables, 6 parameters.

In total this gives 14 parameters that are estimated by maximum likelihood. Also here the length of the mixing zones is not as critical as the above parameters. The length of the mixing zones is set to $\varepsilon = 0.5$. The 3D simulated data gives the result in Table 12 based on 100 samples with 1000 data points in each sample. The estimated values are close to the values used in the simulation and the standard deviation is reasonable.

7.4 Finance data, 3D

Also in the multivariate example, we use the data set described in Section 7.2 with $n = 3$ variables and the two models described in the previous section in addition to the multivariate normal density. We first show some multivariate analyses of the data.

The estimated parameters in the multivariate normal density with the finance data are shown in Table 13. If we exclude the 0.5% of the data where the value of the multivariate normal density with the estimated parameters is smallest, the correlation in the remaining data set is, except for a slight reduction between Europe and Japan, almost identical with the correlation in the entire data set. This is shown in the lower line of Table 13. Hence, the extreme values have the same correlation as the rest of the data set. Excluding the 0.5% most extreme data points corresponds to $\mathbf{x}\boldsymbol{\Sigma}\mathbf{x} < 0.001$ where $\boldsymbol{\Sigma}$ is the estimated correlation matrix from the full data set. Using the Chi-square distribution this corresponds to 8.4e-6 of the probability mass in the estimated distribution. The difference between the empirical truncation and the truncation in the multivariate normal model is much larger by excluding 0.5% instead of for example 1% of the data. This shows that about 0.5% of the data has extreme values according to the estimated multivariate normal distribution. Note that the standard deviations are smaller without the extreme data

Table 12. The parameter in the multivariate normal of transformed variables. The correlation is listed in the order (1,2), (2,3) and (1,3). We first show the parameters that are used in the simulation and then the average of the estimation based on 100 realizations with 1000 samples. The standard deviation of the estimated parameters are shown in parenthesis.

i	σ_i	ρ_i	u_i	$\beta_{i,1}$	$\beta_{i,3}$	likelihood
Simulation						
1	1.00	0.71	1.00	0.50	0.50	-3669
2	1.00	0.82	1.00	0.70	0.70	
3	1.00	0.58	-	0.50	0.70	
Estimation						
1	1.00 (.02)	0.72 (.02)	0.98 (.09)	0.50 (.05)	0.50 (.05)	-3663
2	1.00 (.02)	0.82 (.01)	0.99 (.10)	0.71 (.05)	0.71 (.06)	
3	1.00 (.03)	0.58 (.02)	-	0.51 (.04)	0.71 (.07)	

Table 13. The parameters in the multivariate normal. The correlations are listed in the order (1,2), (2,3) and (1,3). The upper line is with all the data. In the lower line 0.5% of the most extreme data is excluded.

Data	σ_1	σ_2	σ_3	ρ_1	ρ_2	ρ_3
Full data set	0.00962	0.01090	0.01298	0.425	0.120	0.326
Excluded 0.5% of data	0.00905	0.00974	0.01233	0.425	0.128	0.300

points. This corresponds to increasing the density at the origin by a factor 1.53. We have also checked the simultaneous extreme values, see Table 14. The table shows much higher frequencies in the two corners where all three variables have the same sign.

We test the two multivariate models described in the previous section; mixing of five multivariate normal distributions, and multivariate normal distribution of univariate transformed variables, in addition to the multivariate normal distribution. In all the tests we assume that the expectation is zero. In the mixture of the five multivariate normal distributions we are able to get heavier tails than in the multivariate normal distribution. We utilize that the correlations for the entire data set and the data set excluding the extreme values are the same. The estimated parameters are shown in Table 15. In the multivariate normal distribution of univariate transformed variables we get the parameters shown in Table 16.

Table 14. The number of observations in the corners where all three dimensions have values larger than L times the standard deviation in the multivariate normal distribution. The + and - sign indicate the corners.

L	+ + +	+ + -	+ - +	+ - -	- + +	- + -	- - +	- - -	sum
1	37	12	6	4	0	12	12	61	144
1.5	14	4	0	1	0	4	2	21	46
2	6	0	0	1	0	2	0	7	16

Table 15. The parameter in the mixture of five multivariate normal distributions from the 3D finance data. The correlation is listed in the order (1,2), (2,3) and (1,3). The log-likelihood is 39341

i	σ_i	ρ_i	$F_{i,-}$	$F_{i,+}$
1	0.00471	0.358	1.471	1.449
2	0.00539	0.111	1.975	1.836
3	0.00650	.286	-	-

Table 16. The parameter in the multivariate normal of transformed variables from the 3D finance data. The correlation is listed in the order (1,2), (2,3) and (1,3).

i	σ_i	ρ_i	u_i	$\beta_{i,1}$	$\beta_{i,3}$
1	0.00662	0.395	-0.0066	0.573	0.645
2	0.00711	0.123	0.0070	0.588	0.636
3	0.00853	0.298	-	0.664	0.637

The three methods are compared in Table 17 where we estimate the quantiles $q_{\alpha,0.01}$, $q_{\alpha,0.99}$ for each region $\alpha =$ E(urope), U(SA)and J(apan) and multivariate quantiles $q_{A,0.01}$, $q_{A,0.99}$ where $x_i > q_{A,0.01}$ for all the variables $i = 1, 2, 3$. The Table also shows the number of parameters and the log-likelihood from the estimation. The multivariate normal distribution of univariate transformed variables gave the highest likelihood but has more parameters than the mixture of five multivariate normal distributions.

Table 18 gives a summary of a comparison between the empiric quantiles and the quantiles estimated by the different models. The multivariate normal of the transformed variables gives the best estimates for the 0.01/0.99 quantiles and the two new models are equally good on the 0.001/0.999 quantiles. It is possible to improve the quantile estimates by other transformations or a mixture of more normal distributions in the other model.

Table **??** shows the densities of the three different multivariate models at the origin and the most extreme data points for the estimated multinormal distribution. Notice that the multivariate normal density has very small values at some data points. The two other models gives higher densities for these extreme data points. In the mixing of 5 multivariate normal densities and the multivariate normal of the transformed variables, we do not see such extreme densities. When there are such extreme data points the maximum likelihood estimation makes a trade off between densities in the extreme data points and the densities in the many data points in the middle of the distribution. The trade off implies that the model is neither satisfactory in the middle, nor for the extreme data points. It is possible to influence this trade off by putting a prior on the threshold.

8 Summary and conclusions

In this paper we present a new method to mix densities from different models. The method is inspired by (12). But we mix cdfs instead of densities since this is much more

Table 17. Comparing the quantiles and likelihood between the different models and the data. The first four lines are the empirical data. The models are multivariate normal (MN), mixing of multivariate normal densities (MMN), multivariate normal of transformed variables (MNT) and the univariate models using GPD, Weibull and univariate transformation that assumes independence between the three variables. NOP is number of parameters, $q_{\alpha,p}$ the quantiles and loglik. the log-likelihood of the estimated variables.

Model	NOP	p	$q_{E,p}$	$q_{U,p}$	$q_{J,p}$	$q_{A,p}$	loglik.
Data		0.01	-0.0293	-0.0273	-0.0345	-0.0114	
		0.99	0.0235	0.0272	0.0359	0.0110	
		0.001	-0.0645	-0.0688	-0.0585	-0.0171	
		0.999	0.0504	0.0506	0.0721	0.0285	
MN	6	0.01	-0.0223	-0.0253	-0.0302	-0.0093	38235
		0.99	0.0223	0.0253	0.0302	0.0093	
		0.001	-0.0282	-0.0309	-0.0389	-0.0166	
		0.999	0.0282	0.0309	0.0389	0.0166	
MMN	10	0.01	-0.0280	-0.0319	-0.0389	-0.00836	39342
		0.99	0.0261	0.0300	0.0358	0.00858	
		0.001	-0.0391	-0.0448	-0.0543	-0.0215	
		0.999	0.0361	0.0417	0.0499	0.0202	
MNT	14	0.01	-0.0296	-0.0320	-0.0350	-0.0101	39649
		0.99	0.0240	0.0273	0.0361	0.0108	
		0.001	-0.0484	-0.0517	-0.0536	-0.0123	
		0.999	0.0371	0.0425	0.0563	0.0132	
U.GPD-N-GPD	3x5	0.01	-0.0302	-0.0310	-0.0357	- 0.00580	39178
		0.99	0.0244	0.0280	0.0365	0.00626	
		0.001	-0.0683	-0.0600	-0.0595	-0.0113	
		0.999	0.0413	0.0454	0.0632	-0.0115	
U. W-N-W	3x5	0.01	-0.0299	-0.0311	-0.0349	-0.000569	39168
		0.99	0.0241	0.0276	0.0358	0.00511	
		0.001	-0.0558	-0.0547	-0.0557	-0.0115	
		0.999	0.0391	0.0439	0.0582	0.0117	
U. NT	3x5	0.01	-0.0306	-0.0325	-0.0350	-0.00875	39079
		0.99	0.0244	0.0273	0.0360	0.00895	
		0.001	-0.0512	-0.0535	-0.0535	-0.0105	
		0.999	0.0397	0.0418	0.0553	0.0111	

Table 18. The table gives the arithmetic average absolute distance between the empirical quantiles and the quantiles estimated by the different models. The models are multivariate normal (MN), mixing of multivariate normal densities (MMN), and multivariate normal of transformed variables (MNT).

Quantiles	MN	MMN	MNT
0.01 and 0.99	0.0030	0.0026	0.0012
0.001 and 0.999	0.0255	0.0166	0.0163

computationally stable and efficient making it possible to extend to several dimensions in contrast to the method described in (12). The paper (3) also combines cdfs but by introducing a mixing zone we obtain continuous densities. We also show how univariate transformations may be used in order to represent tail behaviour.

The different models are tested by simulation from one model and then estimate parameters and quantiles from all the models. The suitability of each model depend on the data in each case. We compare the different models on a 3D financial data set, evaluating likelihood and tail behaviours, both univariate and multivariate. In the univariate test the different models seem to behave quite similarly. Three different multidimensional models are compared on the data set. The multivariate normal model of transformed variables gave the highest likelihood, but the mixture of 5 multivariate normal distributions also gave satisfactory results with a smaller number of parameters.

Before we select a model we should analyse the data and then find a model with suitable tail behaviour. In the multivariate case we need to find the univariate tail behaviour and the correlation in the tails in the data.

Acknowledgments

This work is partly sponsored by the Norwegian fund Finansmarkedsfondet and partly by the Norwegian Research Council. We thank Kjersti Aas and Håvard Rue for valuable comments.

References

[1] Abramowitz, M. and Stegun, I. A., (1972) Handbook of Mathematical Function, Dover, New York.

[2] Andersson, J., (2001) On the Normal Inverse Gaussian Stochastic Volatility model, Journal of Business and Economic Statistics, **19** , 44 - - 54.

[3] Behrens C.N.A., Lopes, H. & Gamerman, D. (2002), Finite mixture distributions, sequential likelihood and the em algorithm, Technical report, Federal University of Rio de Janerio, Department of Statistical Methods.

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

[5] Carreau, J. and Bengio, Y. (2006), 'A Hybrid Pareto Model for Conditional Asymmetric Fat-Tailed' Distribution', *Preprint Departement d'Informatique et Recherche Operationnelle, University of Montreal, Canada* **1**, 55–79.

[6] Conover, W. J. (1971), *Practical nonparametric statistics*, John Wiley & Sons, New York.

[7] Crovella, M. and Taqqu, M. (1999), 'Estimating the heavy tail index from scaling properties', *Methodology and Computing in Applied Probability* **1**, 55–79.

[8] Davison, A. C. and Smith, R. L. (1990), 'Models for exceedances over high thresholds (with discussion)', *Journal of the Royal Statistical Society, Series B* **5**(3), 393–442.

[9] Dupuis, D. J. (1999), 'Exceedances over high thresholds: A guide to threshold selection', *Extremes* **1**(3), 251–261.

[10] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling extremal events*, Springer-Verlag, Berlin Heidelberg.

[11] Embrechts, P., Lindskog, F., and McNeil, AJ. (2003), *Modelling dependence with copula and applications to risk management*, In S.T. Rachev, editor, Handbook of Heavy Tailed distributions in Finance, Elsevier, North-Holland.

[12] Frigessi, A., Haug, O. and Rue, H. (2002), 'A dynamic mixture model for unsupervised tail estimation without threshold selection', *Extremes* **5**, 219–235.

[13] Heffernan, J. and Tawn, J. A. (2004), 'A conditional approach for multivariate extreme values', *J. R. Statist. Soc. B* **66**, 497–546.

[14] Hsing, T., Klüppelberg, C. and Kuhn, G. (2004), 'Dependence Estimation and Visualization in Multivariate Extremes with Applications to Financial Data', *Extremes* **7**(2), 99–121.

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

[16] Longin, F. and Solnik, B. (2001), 'Extreme correlation of international equity markets', *The Journal of Finance* **56**(2), 649–676.

[17] Marshall, A. W. and Olkin, I. (1967), 'A multivariate exponential distribution', *Journal of the American Statistical Association* **62**, 30–44.

[18] McNeil, A. J. and Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**, 271–300.

[19] Muhammad, M. and Mori, M. (2003), 'Double exponential formulas for numerical indefinite integration', *Journal of Computational and Applied Mathematics* **161**(2), 431–448.

[20] Poon, S., Rockinger, M. and Tawn, J. A. (2004), 'Extreme-value dependence in financial markets: Diagnostics, models and financial implications', *Rev. Finan. Stud.* **17**, 581–610.

[21] Resnick, S. I. (1997), 'Heavy tail modeling and teletraffic data', *The Annals of Statistics* **25**(5), 1805–1869.

[22] Rootzen, H. and Tajvidi, N. (1997), 'Extreme value statistics and wind storm losses: A case study', *Scandinavian Actuarial Journal* **1**, 70–94.

[23] Starica, C. (2000), 'Multivariate extremes for models with constant conditional correlation', *Journal of Empirical Finance* **6**, 513–553.

[24] Venter, J. H. and de Jongh, P. J. (2004), 'Selecting an innovation distribution for garch models to improve efficiency of risk and volatility estimation', *Journal of Risk* **6**(3), 27–53.