CORRELATION AND DEPENDENCY IN RISK MANAGEMENT: PROPERTIES AND PITFALLS

PAUL EMBRECHTS, ALEXANDER MCNEIL, AND DANIEL STRAUMANN

Abstract. Modern risk management calls for an understanding of stochastic dependence going beyond simple linear correlation. This paper deals with the static (non-time-dependent) case and emphasizes the copula representation of dependence for a random vector. Linear correlation is a natural dependence measure for multivariate normally and, more generally, elliptically distributed risks but other dependence concepts like comonotonicity and rank correlation should also be understood by the risk management practitioner. Using counterexamples the falsity of some commonly held views on correlation is demonstrated; in general, these fallacies arise from the naive assumption that dependence properties of the elliptical world also hold in the non-elliptical world. In particular, the problem of finding multivariate models which are consistent with prespecified marginal distributions and correlations is addressed. Pitfalls are highlighted and simulation algorithms avoiding these problems are constructed.

1. Introduction

1.1. **Correlation in finance and insurance.** In financial theory the notion of correlation is central. The Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT) (Campbell, Lo, and MacKinlay 1997) use correlation as a measure of dependence between different financial instruments and employ an elegant theory, which is essentially founded on an assumption of multivariate normally distributed returns, in order to arrive at an optimal portfolio selection. Although insurance has traditionally been built on the assumption of independence and the law of large numbers has governed the determination of premiums, the increasing complexity of insurance and reinsurance products has led recently to increased actuarial interest in the modelling of dependent risks (Wang 1997); an example is the emergence of more intricate multi-line products.

The current quest for a sound methodological basis for integrated risk management also raises the issue of correlation and dependency. Although contemporary financial risk management revolves around the use of correlation to describe dependence between risks, the inclusion of non-linear derivative products invalidates many of the distributional assumptions underlying the use of correlation. In insurance these assumptions are even more problematic because of the typical skewness and heavy-tailedness of insurance claims data.

Recently, within the actuarial world, dynamic financial analysis (DFA) and dynamic solvency testing (DST) have been heralded as a way forward for integrated risk management of the investment and underwriting risks to which an insurer (or

Date: November 1998.

Key words and phrases. Risk management; correlation; elliptic distributions; rank correlation; dependency; copula; comonotonicity; simulation; Value-at-Risk; coherent risk measures.

The work of the third author was supported by RiskLab; the second author would like to thank Swiss Re for financial support.

bank) is exposed. DFA, for instance, is essentially a Monte Carlo or simulation-based approach to the joint modelling of risks (see e.g. CAS (1997) or Lowe and Stanard (1997)). This necessitates model assumptions that combine information on marginal distributions together with ideas on interdependencies. The correct implementation of a DFA-based risk management system certainly requires a proper understanding of the concepts of dependency and correlation.

1.2. **Correlation as a source of confusion.** But correlation, as well as being one of the most ubiquitous concepts in modern finance and insurance, is also one of the most misunderstood concepts. Some of the confusion may arise from the literary use of the word to cover any notion of dependency. To a mathematician correlation is only one particular measure of stochastic dependency among many. It is the canonical measure in the world of multivariate normal distributions, and more generally for spherical and elliptical distributions. However, empirical research in finance and insurance shows that the distributions of the real world are seldom in this class.

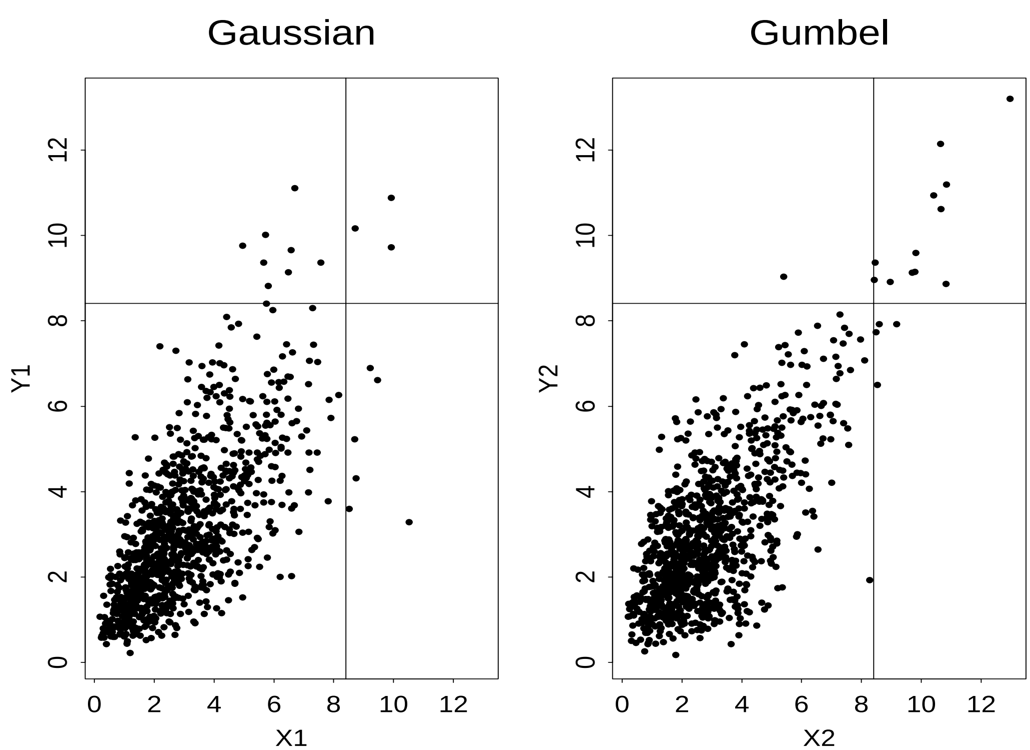

FIGURE 1. 1000 random variates from two distributions with identical Gamma(3,1) marginal distributions and identical correlation $\rho = 0.7$, but *different* dependence structures.

As motivation for the ideas of this paper we include Figure 1. This shows 1000 bivariate realisations from two *different* probability models for (X, Y). In both models X and Y have identical gamma marginal distributions and the linear correlation between them is 0.7. However, it is clear that the dependence between X and Y in the two models is qualitatively quite different and, if we consider the random variables to represent insurance losses, the second model is the more dangerous model from the point of view of an insurer, since extreme losses have a tendency to occur together. We will return to this example later in the paper, see Section 5; for the time-being we note that the dependence in the two models cannot be distinguished on the grounds of correlation alone.

The main aim of the paper is to collect and clarify the essential ideas of dependence, linear correlation and rank correlation that anyone wishing to model dependent phenomena should know. In particular, we highlight a number of important fallacies concerning correlation which arise when we work with models other than the multivariate normal. Some of the pitfalls which await the end-user are quite subtle and perhaps counter-intuitive.

We are particularly interested in the problem of constructing multivariate distributions which are consistent with given marginal distributions and correlations, since this is a question that anyone wanting to simulate dependent random vectors, perhaps with a view to DFA, is likely to encounter. We look at the existence and construction of solutions and the implementation of algorithms to generate random variates. Various other ideas recur throughout the paper. At several points we look at the effect of dependency structure on the Value-at-Risk or VaR under a particular probability model, i.e. we measure and compare risks by looking at quantiles. We also relate these considerations to the idea of a coherent measure of risk as introduced by Artzner, Delbaen, Eber, and Heath (1999).

We concentrate on the *static* problem of describing dependency between a pair or within a group of random variables. There are various other problems concerning the modelling and interpretation of serial correlation in stochastic processes and cross-correlation between processes; see Boyer, Gibson, and Loretan (1999) for problems related to this. We do not consider the statistical problem of estimating correlations and rank correlation, where a great deal could also be said about the available estimators, their properties and their robustness, or the lack of it.

1.3. **Organization of paper.** In Section 2 we begin by discussing joint distributions and the use of copulas as descriptions of dependency between random variables. Although copulas are a much more recent and less well known approach to describing dependency than correlation, we introduce them first for two reasons. First, they are the principal tool we will use to illustrate the pitfalls of correlation and second, they are the approach which in our opinion affords the best understanding of the general concept of dependency.

In Section 3 we examine linear correlation and define spherical and elliptical distributions, which constitute, in a sense, the natural environment of the linear correlation. We mention both some advantages and shortcomings of correlation. Section 4 is devoted to a brief discussion of some alternative dependency concepts and measures including comonotonicity and rank correlation. Three of the most common fallacies concerning linear correlation and dependence are presented in Section 5. In Section 6 we explain how vectors of dependent random variables may be simulated using correct methods.

2. Copulas

Probability-integral and quantile transforms play a fundamental role when working with copulas. In the following proposition we collect together some essential facts that we use repeatedly in this paper. The notation $X \sim F$ means that the random variable X has distribution function F.

Proposition 1. *Let X be a random variable with distribution function F. Let F^{-1} be the quantile function of F, i.e.*

$$F^{-1}(\alpha) = \inf\{x|F(x) \geq \alpha\},$$

$\alpha \in (0,1)$. *Then*

1. *For any standard-uniformly distributed $U \sim U(0,1)$ we have $F^{-1}(U) \sim F$. This gives a simple method for simulating random variates with distribution function F.*
2. *If F is continuous then the random variable $F(X)$ is standard-uniformly distributed, i.e. $F(X) \sim U(0,1)$.*

Proof. In most elementary texts on probability. □

2.1. **What is a copula?** The dependence between the real-valued random variables $X_1, \ldots, X_n$ is completely described by their joint distribution function

$$F(x_1, \ldots, x_n) = \mathbb{P}[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

The idea of separating F into a part which describes the dependence structure and parts which describe the marginal behaviour only, has led to the concept of a copula.

Suppose we transform the random vector $\mathbf{X} = (X_1, \ldots, X_n)^t$ component-wise to have standard-uniform marginal distributions, $\mathrm{U}(0,1)$[1]. For simplicity we assume to begin with that $X_1, \ldots, X_n$ have continuous marginal distributions $F_1, \ldots, F_n$, so that this can be achieved by using the probability-integral transformation $T : \mathbb{R}^n \to \mathbb{R}^n, (x_1, \ldots, x_n)^t \mapsto (F_1(x_1), \ldots, F_n(x_n))^t$. The joint distribution function C of $(F_1(X_1), \ldots, F_n(X_n))^t$ is then called the copula of the random vector $(X_1, \ldots, X_n)^t$ or the multivariate distribution F. It follows that

$$\begin{aligned} F(x_1, \ldots, x_n) &= \mathbb{P}[F_1(X_1) \leq F_1(x_1), \ldots, F_n(X_n) \leq F_n(x_n)] \\ &= C(F_1(x_1), \ldots, F_n(x_n)). \end{aligned} \tag{1}$$

Definition 1. A copula is the distribution function of a random vector in $\mathbb{R}^n$ with uniform-$(0,1)$ marginals. Alternatively a copula is any function $C : [0,1]^n \to [0,1]$ which has the three properties:

1. $C(x_1, \ldots, x_n)$ is increasing in each component x_i.
2. $C(1, \ldots, 1, x_i, 1, \ldots, 1) = x_i$ for all $i \in \{1, \ldots, n\}$, $x_i \in [0,1]$.
3. For all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in [0,1]^n$ with $a_i \leq b_i$ we have:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C(x_{1i_1}, \ldots, x_{ni_n}) \geq 0, \tag{2}$$

where $x_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, \ldots, n\}$.

These two alternative definitions can be shown to be equivalent. It is a particularly easy matter to verify that the first definition in terms of a multivariate distribution function with standard uniform marginals implies the three properties above: property 1 is clear; property 2 follows from the fact that the marginals are uniform-$(0,1)$; property 3 is true because the sum (2) can be interpreted as $\mathbb{P}[a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n]$, which is non-negative.

For any continuous multivariate distribution the representation (1) holds for a *unique* copula C. If $F_1, \ldots, F_n$ are not all continuous it can still be shown (see Schweizer and Sklar (1983), Chapter 6) that the joint distribution function can

[1] Alternatively one could transform to any other distribution, but $\mathrm{U}(0,1)$ is particularly easy.

always be expressed as in (1), although in this case C is *no longer* unique and we refer to it as a *possible* copula of F.

The representation (1), and some invariance properties which we will show shortly, suggest that we interpret a copula associated with $(X_1, \dots X_n)^t$ as being *the dependence structure.* This makes particular sense when all the F_i are continuous and the copula is unique; in the discrete case there will be more than one way of writing the dependence structure. Pitfalls related to non-continuity of marginal distributions are presented in Marshall (1996). A recent, very readable introduction to copulas is Nelsen (1999).

2.2. **Examples of copulas.** For independent random variables the copula trivially takes the form

$$C_{ind}(x_1, \dots, x_n) = x_1 \cdot \ldots \cdot x_n. \tag{3}$$

We now consider some particular copulas for non-independent pairs of random variables (X, Y) having continuous distributions. The Gaussian or normal copula is

$$C_\rho^{Ga}(x, y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ \frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)} \right\} dsdt, \tag{4}$$

where $-1 < \rho < 1$ and Φ is the univariate standard normal distribution function. Variables with standard normal marginal distributions and this dependence structure, i.e. variables with d.f. $C_\rho^{Ga}(\Phi(x), \Phi(y))$, are standard bivariate normal variables with correlation coefficient ρ. Another well-known copula is the Gumbel or logistic copula

$$C_\beta^{Gu}(x, y) = \exp\left[- \left\{ (-\log x)^{1/\beta} + (-\log y)^{1/\beta} \right\}^\beta \right], \tag{5}$$

where $0 < \beta \leq 1$ is a parameter which controls the amount of dependence between X and Y; $\beta = 1$ gives independence and the limit of C_β^{Gu} for $\beta \to 0+$ leads to perfect dependence, as will be discussed in Section 4. This copula, unlike the Gaussian, is a copula which is consistent with bivariate extreme value theory and could be used to model the limiting dependence structure of component-wise maxima of bivariate random samples (Joe (1997), Galambos (1987)).

The following is a simple method for generating a variety of copulas which will be used later in the paper. Let $f, g : [0,1] \to \mathbb{R}$ with $\int_0^1 f(x)dx = \int_0^1 g(y)dy = 0$ and $f(x)g(y) \geq -1$ for all $x, y \in [0,1]$. Then $h(x,y) = 1 + f(x)g(y)$ is a bivariate density function on $[0,1]^2$. Consequently,

$$C(x, y) = \int_0^x \int_0^y h(u, v)dudv = xy + \left(\int_0^x f(u)du \right) \left(\int_0^y g(v)dv \right) \tag{6}$$

is a copula. If we choose $f(x) = \alpha(1-2x)$, $g(y) = (1-2y)$, $|\alpha| \leq 1$, we obtain, for example, the Farlie-Gumbel-Morgenstern copula $C(x,y) = xy[1 + \alpha(1-x)(1-y))]$. Many copulas and methods to construct them can be found in the literature; see for example Hutchinson and Lai (1990) or Joe (1997).

2.3. **Invariance.** The following proposition shows one attractive feature of the copula representation of dependence, namely that the dependence structure as summarized by a copula is *invariant* under increasing and continuous transformations of the marginals.

Proposition 2. *If* $(X_1, \dots, X_n)^t$ *has copula* C *and* $T_1, \dots, T_n$ *are increasing continuous functions, then* $(T_1(X_1), \dots, T_n(X_n))^t$ *also has copula* C.

Proof. Let $(U_1, \dots, U_n)^t$ have distribution function C (in the case of continuous marginals F_{X_i} take $U_i = F_{X_i}(X_i)$). We may write

$$\begin{aligned}
&C(F_{T_1(X_1)}(x_1), \dots, F_{T_n(X_n)}(x_n))\\
&= \mathbb{P}[U_1 \le F_{T_1(X_1)}(x_1), \dots, U_n \le F_{T_n(X_n)}(x_n)]\\
&= \mathbb{P}[F^{-1}_{T_1(X_1)}(U_1) \le x_1, \dots, F^{-1}_{T_n(X_n)}(U_n) \le x_n]\\
&= \mathbb{P}[T_1 \circ F^{-1}_{X_1}(U_1) \le x_1, \dots, T_n \circ F^{-1}_{X_n}(U_n) \le x_n]\\
&= \mathbb{P}[T_1(X_1) \le x_1, \dots, T_n(X_n) \le x_n].
\end{aligned}$$

□

Remark 1. The continuity of the transformations T_i is necessary for general random variables $(X_1, \dots, X_n)^t$ since, in that case, $F^{-1}_{T_i(X_i)} = T_i \circ F^{-1}_{X_i}$. In the case where all marginal distributions of $\mathbf{X}$ are continuous it suffices that the transformations are increasing (see also Chapter 6 of Schweizer and Sklar (1983)).

As a simple illustration of the relevance of this result, suppose we have a probability model (multivariate distribution) for dependent insurance losses of various kinds. If we decide that our interest now lies in modelling the logarithm of these losses, the copula will not change. Similarly if we change from a model of percentage returns on several financial assets to a model of logarithmic returns, the copula will not change, only the marginal distributions.

3. Linear Correlation

3.1. **What is correlation?** We begin by considering pairs of real-valued, non-degenerate random variables X, Y with finite variances.

Definition 2. The linear correlation coefficient between X and Y is

$$\rho(X,Y) = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\sigma^2[X]\sigma^2[Y]}},$$

where $\mathrm{Cov}[X,Y]$ is the covariance between X and Y, $\mathrm{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ and $\sigma^2[X], \sigma^2[Y]$ denote the variances of X and Y.

The linear correlation is a measure of *linear* dependence. In the case of independent random variables, $\rho(X,Y) = 0$ since $\mathrm{Cov}[X,Y] = 0$. In the case of perfect linear dependence, i.e. $Y = aX + b$ a.s. or $\mathbb{P}[Y = aX + b] = 1$ for $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}$, we have $\rho(X,Y) = \pm 1$. This is shown by considering the representation

$$\rho(X,Y)^2 = \frac{\sigma^2[Y] - \min_{a,b} \mathbb{E}[(Y - (aX+b))^2]}{\sigma^2[Y]}. \tag{7}$$

In the case of imperfect linear dependence, $-1 < \rho(X,Y) < 1$, and this is the case when misinterpretations of correlation are possible, as will later be seen in Section 5. Equation (7) shows the connection between correlation and simple linear regression. The right hand side can be interpreted as the relative reduction in the variance of Y by linear regression on X, The regression coefficients a_R, b_R, which minimise the squared distance $\mathbb{E}[(Y-(aX+b))^2]$ are given by

$$\begin{aligned} a_R &= \frac{\mathrm{Cov}[X,Y]}{\sigma^2[X]}, \\ b_R &= \mathbb{E}[Y] - a_R \mathbb{E}[X]. \end{aligned}$$

Correlation fulfills the linearity property

$$\rho(\alpha X + \beta, \gamma Y + \delta) = \mathrm{sgn}(\alpha \cdot \gamma)\rho(X,Y),$$

when $\alpha, \gamma \in \mathbb{R} \setminus \{0\}$, $\beta, \delta \in \mathbb{R}$. Correlation is thus invariant under positive affine transformations, i.e. strictly increasing *linear* transformations.

The generalisation of correlation to more than two random variables is straightforward. Consider vectors of random variables $\mathbf{X} = (X_1, \ldots, X_n)^t$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^t$ in $\mathbb{R}^n$. We can summarise all pairwise covariances and correlations in $n \times n$ matrices $\mathrm{Cov}[\mathbf{X}, \mathbf{Y}]$ and $\rho(\mathbf{X}, \mathbf{Y})$. As long as the corresponding variances are finite we define

$$\begin{aligned} \mathrm{Cov}[\mathbf{X}, \mathbf{Y}]_{ij} &:= \mathrm{Cov}[X_i, Y_j], \\ \rho(\mathbf{X}, \mathbf{Y})_{ij} &:= \rho(X_i, Y_j) \quad 1 \leq i, j \leq n. \end{aligned}$$

It is well known that these matrices are symmetric and positive semi-definite. Often one considers only pairwise correlations between components of a single random vector; in this case we set $\mathbf{Y} = \mathbf{X}$ and consider $\rho(\mathbf{X}) := \rho(\mathbf{X}, \mathbf{X})$ or $\mathrm{Cov}[\mathbf{X}] := \mathrm{Cov}[\mathbf{X}, \mathbf{X}]$.

The popularity of linear correlation can be explained in several ways. Correlation is often straightforward to calculate. For many bivariate distributions it is a simple matter to calculate second moments (variances and covariances) and hence to derive the correlation coefficient. Alternative measures of dependence, which we will encounter in Section 4 may be more difficult to calculate.

Moreover, correlation and covariance are easy to manipulate under linear operations. Under affine linear transformations $A : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Ax + a$ and $B : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Bx + b$ for $A, B \in \mathbb{R}^{m \times n}$, $a, b \in \mathbb{R}^m$ we have

$$\mathrm{Cov}[A\mathbf{X} + a, B\mathbf{Y} + b] = A\mathrm{Cov}[\mathbf{X}, \mathbf{Y}]B^t.$$

A special case is the following elegant relationship between variance and covariance for a random vector. For every linear combination of the components $\alpha^t \mathbf{X}$ with $\alpha \in \mathbb{R}^n$,

$$\sigma^2[\alpha^t \mathbf{X}] = \alpha^t \mathrm{Cov}[\mathbf{X}]\alpha.$$

Thus, the variance of any linear combination is fully determined by the pairwise covariances between the components. This fact is commonly exploited in portfolio theory.

A third reason for the popularity of correlation is its naturalness as a measure of dependence in multivariate normal distributions and, more generally, in multivariate spherical and elliptical distributions, as will shortly be discussed. First, we mention a few disadvantages of correlation.

3.2. **Shortcomings of correlation.** We consider again the case of two real-valued random variables X and Y.

- The variances of X and Y must be finite or the linear correlation is not defined. This is not ideal for a dependency measure and causes problems when we work with heavy-tailed distributions. For example, the covariance and the correlation between the two components of a bivariate t_ν-distributed random vector are not defined for $\nu \leq 2$. Non-life actuaries who model losses in different business lines with infinite variance distributions must be aware of this.

- Independence of two random variables implies they are uncorrelated (linear correlation equal to zero) but zero correlation does not in general imply independence. A simple example where the covariance disappears despite strong dependence between random variables is obtained by taking $X \sim \mathcal{N}(0,1)$, $Y = X^2$, since the third moment of the standard normal distribution is zero. Only in the case of the *multivariate* normal is it permissable to interpret uncorrelatedness as implying independence. This implication is no longer valid when only the marginal distributions are normal and the *joint* distribution is non-normal, which will also be demonstrated in Example 1. The class of spherical distributions model uncorrelated random variables but are not, except in the case of the multivariate normal, the distributions of independent random variables.
- Linear correlation has the serious deficiency that it is *not* invariant under non-linear strictly increasing transformations $T : \mathbb{R} \to \mathbb{R}$. For two real-valued random variables we have in general

$$\rho(T(X), T(Y)) \neq \rho(X, Y).$$

If we take the bivariate standard normal distribution with correlation ρ and the transformation $T(x) = \Phi(x)$ (the standard normal distribution function) we have

$$\rho(T(X), T(Y)) = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right), \tag{8}$$

see Joag-dev (1984). In general one can also show (see Kendall and Stuart (1979), page 600) for bivariate normally-distributed vectors and arbitrary transformations $T, \widetilde{T} : \mathbb{R} \to \mathbb{R}$ that

$$|\rho(T(X), \widetilde{T}(Y))| \leq |\rho(X, Y)|,$$

which is also true in (8).

3.3. **Spherical and elliptical distributions.** The spherical distributions extend the standard multivariate normal distribution $\mathcal{N}_n(0, I)$, i.e. the distribution of independent standard normal variables. They provide a family of symmetric distributions for uncorrelated random vectors with mean zero.

Definition 3.
A random vector $\mathbf{X} = (X_1, \ldots, X_n)^t$ has a spherical distribution if for every orthogonal map $U \in \mathbb{R}^{n \times n}$ (i.e. maps satisfying $UU^t = U^tU = I_{n\times n}$)

$$U\mathbf{X} =_{\mathrm{d}} \mathbf{X}. \text{ [2]}$$

The characteristic function $\psi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^t\mathbf{X})]$ of such distributions takes a particularly simple form. There exists a function $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that $\psi(\mathbf{t}) = \phi(\mathbf{t}^t\mathbf{t}) = \phi(t_1^2 + \ldots + t_n^2)$. This function is the *characteristic generator* of the spherical distribution and we write

$$\mathbf{X} \sim S_n(\phi).$$

If $\mathbf{X}$ has a density $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ then this is equivalent to $f(\mathbf{x}) = g(\mathbf{x}^t\mathbf{x}) = g(x_1^2 + \ldots + x_n^2)$ for some function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, so that the spherical distributions are best interpreted as those distributions whose density is constant on spheres.

[2] We standardly use $=_{\mathrm{d}}$ to denote equality in distribution.

Some other examples of densities in the spherical class are those of the multivariate t-distribution with ν degrees of freedom $f(\mathbf{x}) = c(1+\mathbf{x}^t\mathbf{x}/\nu)^{-(n+\nu)/2}$ and the logistic distribution $f(\mathbf{x}) = c\exp(-\mathbf{x}^t\mathbf{x})/[1+\exp(-\mathbf{x}^t\mathbf{x})]^2$, where c is a generic normalizing constant. Note that these are the distributions of uncorrelated random variables but, contrary to the normal case, *not* the distributions of independent random variables. In the class of spherical distributions the multivariate normal is the only distribution of independent random variables, see Fang, Kotz, and Ng (1987), page 106.

The spherical distributions admit an alternative stochastic representation. $\mathbf{X} \sim S_n(\phi)$ if and only if

$$\mathbf{X} =_{\mathrm{d}} R \cdot \mathbf{U}, \tag{9}$$

where the random vector $\mathbf{U}$ is uniformly distributed on the unit hypersphere $S_{n-1} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^t\mathbf{x} = 1\}$ in $\mathbb{R}^n$ and $R \geq 0$ is a positive random variable, independent of $\mathbf{U}$ (Fang, Kotz, and Ng (1987), page 30). Spherical distributions can thus be interpreted as mixtures of uniform distributions on spheres of differing radius in $\mathbb{R}^n$. For example, in the case of the standard multivariate normal distribution the *generating variate* satisfies $R \sim \sqrt{\chi_n^2}$, and in the case of the multivariate t-distribution with ν degrees of freedom $R^2/n \sim F(n,\nu)$ holds, where $F(n,\nu)$ denotes an F-distribution with n and ν degrees of freedom.

Elliptical distributions extend the multivariate normal $\mathcal{N}_n(\mu,\Sigma)$, i.e. the distribution with mean μ and covariance matrix Σ. Mathematically they are the affine maps of spherical distributions in $\mathbb{R}^n$.

Definition 4. Let $T : \mathbb{R}^n \to \mathbb{R}^n, \mathbf{x} \mapsto A\mathbf{x} + \mu$, $A \in \mathbb{R}^{n\times n}$, $\mu \in \mathbb{R}^n$ be an affine map. $\mathbf{X}$ has an elliptical distribution if $\mathbf{X} = T(\mathbf{Y})$ and $\mathbf{Y} \sim S_n(\phi)$.

Since the characteristic function can be written as

$$\begin{aligned} \psi(\mathbf{t}) &= \mathbb{E}[\exp(i\mathbf{t}^t\mathbf{X})] = \mathbb{E}[\exp(i\mathbf{t}^t(A\mathbf{Y}+\mu))] \\ &= \exp(i\mathbf{t}^t\mu)\exp(i(A^t\mathbf{t})^t\mathbf{Y}) = \exp(i\mathbf{t}^t\mu)\phi(\mathbf{t}^t\Sigma\mathbf{t}), \end{aligned}$$

where $\Sigma := AA^t$, we denote the elliptical distributions

$$\mathbf{X} \sim \mathrm{E}_n(\mu,\Sigma,\phi).$$

For example, $\mathcal{N}_n(\mu,\Sigma) = \mathrm{E}_n(\mu,\Sigma,\phi)$ with $\phi(t) = \exp(-t^2/2)$. If $\mathbf{Y}$ has a density $f(\mathbf{y}) = g(\mathbf{y}^t\mathbf{y})$ and if A is regular ($\det(A) \neq 0$ so that Σ is strictly positive-definite), then $\mathbf{X} = A\mathbf{Y} + \mu$ has density

$$h(\mathbf{x}) = \frac{1}{\sqrt{\det(\Sigma)}} g((\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)),$$

and the contours of equal density are now ellipsoids.

Knowledge of the distribution of $\mathbf{X}$ does not completely determine the elliptical representation $\mathrm{E}_n(\mu,\Sigma,\phi)$; it uniquely determines μ but Σ and ϕ are only determined up to a positive constant[3]. In particular Σ can be chosen so that it is directly interpretable as the covariance matrix of $\mathbf{X}$, although this is not always standard. Let $\mathbf{X} \sim \mathrm{E}_n(\mu,\Sigma,\phi)$, so that $\mathbf{X} =_{\mathrm{d}} \mu + A\mathbf{Y}$ where $\Sigma = AA^t$ and $\mathbf{Y}$ is a random vector satisfying $\mathbf{Y} \sim S_n(\phi)$. Equivalently $\mathbf{Y} =_{\mathrm{d}} R\cdot\mathbf{U}$, where $\mathbf{U}$ is uniformly distributed on S^{n-1} and R is a positive random variable independent of $\mathbf{U}$. If $\mathbb{E}[R^2] < \infty$ it follows

[3]If X is elliptical and non-degenerate there exists μ, A and $\mathbf{Y} \sim S_n(\phi)$ so that $\mathbf{X} =_{\mathrm{d}} A\mathbf{Y} + \mu$, but for any $\lambda \in \mathbb{R}\setminus\{0\}$ we also have $\mathbf{X} =_{\mathrm{d}} (A/\lambda)\lambda\mathbf{Y} + \mu$ where $\lambda\mathbf{Y} \sim S_n(\widetilde{\phi})$ and $\widetilde{\phi}(u) := \phi(\lambda^2 u)$. In general, if $X \sim \mathrm{E}_n(\mu,\Sigma,\phi) = \mathrm{E}_n(\tilde{\mu},\widetilde{\Sigma},\widetilde{\phi})$ then $\mu = \tilde{\mu}$ and there exists $c > 0$ so that $\widetilde{\Sigma} = c\Sigma$ and $\widetilde{\phi}(u) = \phi(u/c)$ (see Fang, Kotz, and Ng (1987), page 43).

that $\mathbb{E}[\mathbf{X}] = \mu$ and $\mathrm{Cov}[\mathbf{X}] = AA^t\mathbb{E}[R^2]/n = \Sigma\mathbb{E}[R^2]/n$ since $\mathrm{Cov}[\mathbf{U}] = I_{n\times n}/n$. By starting with the characteristic generator $\widetilde{\phi}(u) := \phi(u/c)$ with $c = n/\mathbb{E}[R^2]$ we ensure that $\mathrm{Cov}[\mathbf{X}] = \Sigma$. An elliptical distribution is thus fully described by its mean, its covariance matrix and its characteristic generator.

We now consider some of the reasons why correlation and covariance are natural measures of dependence in the world of elliptical distributions. First, many of the properties of the multivariate normal distribution are shared by the elliptical distributions. Linear combinations, marginal distributions and conditional distributions of elliptical random variables can largely be determined by linear algebra using knowledge of covariance matrix, mean and generator. This is summarized in the following properties.

- Any linear combination of an elliptically distributed random vector is also elliptical with the same characteristic generator ϕ. If $\mathbf{X} \sim \mathrm{E}_n(\mu, \Sigma, \phi)$ and $B \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, then
$$B\mathbf{X} + b \sim \mathrm{E}_m(B\mu + b, B\Sigma B^t, \phi).$$
It is immediately clear that the components $X_1, \dots, X_n$ are all symmetrically distributed random variables of the same type[4].
- The marginal distributions of elliptical distributions are also elliptical with the same generator. Let $\mathbf{X} = \begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix} \sim \mathrm{E}_n(\Sigma, \mu, \phi)$ with $\mathbf{X}_1 \in \mathbb{R}^p$, $\mathbf{X}_2 \in \mathbb{R}^q$, $p + q = n$. Let $\mathbb{E}[\mathbf{X}] = \mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$, $\mu_1 \in \mathbb{R}^p$, $\mu_2 \in \mathbb{R}^q$ and $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$, accordingly. Then
$$\mathbf{X}_1 \sim \mathrm{E}_p(\mu_1, \Sigma_{11}, \phi),\ \ \mathbf{X}_2 \sim \mathrm{E}_q(\mu_2, \Sigma_{22}, \phi).$$
- We assume that Σ is strictly positive-definite. The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is also elliptical, although in general with a different generator $\widetilde{\phi}$:
$$\mathbf{X}_1|\mathbf{X}_2 \sim \mathrm{E}_p(\mu_{1.2}, \Sigma_{11.2}, \widetilde{\phi}), \tag{10}$$
where $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)$, $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The distribution of the generating variable $\widetilde{R}$ in (9) corresponding to $\widetilde{\phi}$ is the conditional distribution
$$\sqrt{(\mathbf{X} - \mu)^t\Sigma^{-1}(\mathbf{X} - \mu) - (\mathbf{X}_2 - \mu_2)^t\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)}\ \Big|\mathbf{X}_2.$$
Since in the case of multivariate normality uncorrelatedness is equivalent to independence we have $\widetilde{R} =_{\mathrm{d}} \sqrt{\chi_p^2}$ and $\widetilde{\phi} = \phi$, so that the conditional distribution is of the same type as the unconditional; for general elliptical distributions this is not true. From (10) we see that
$$\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2] = \mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2),$$
so that the best prediction of $\mathbf{X}_1$ given $\mathbf{X}_2$ is linear in $\mathbf{X}_2$ and is simply the linear regression of $\mathbf{X}_1$ on $\mathbf{X}_2$. In the case of multivariate normality we have additionally
$$\mathrm{Cov}[\mathbf{X}_1|\mathbf{X}_2] = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

[4]Two random variables X und Y are of the same type if we can find $a > 0$ and $b \in \mathbb{R}$ so that $Y =_{\mathrm{d}} aX + b$.

which is independent of $\mathbf{X}_2$. The independence of the conditional covariance from $\mathbf{X}_2$ is also a characterisation of the multivariate normal distribution in the class of elliptical distributions (Kelker 1970).

Since the type of all marginal distributions is the same, we see that an elliptical distribution is *uniquely* determined by its mean, its covariance matrix and knowledge of this type. Alternatively the dependence structure (copula) of a continuous elliptical distribution is *uniquely* determined by the correlation matrix and knowledge of this type. For example, the copula of the bivariate t-distribution with ν degrees of freedom and correlation ρ is

$$C^t_{\nu,\rho}(x,y) = \int_{-\infty}^{t_\nu^{-1}(x)} \int_{-\infty}^{t_\nu^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{1 + \frac{(s^2 - 2\rho st + t^2)}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} dsdt, \tag{11}$$

where $t_\nu^{-1}(x)$ denotes the inverse of the distribution function of the standard univariate t-distribution with ν degrees of freedom. This copula is seen to depend only on ρ and ν.

An important question is, which univariate types are possible for the marginal distribution of an elliptical distribution in $\mathbb{R}^n$ for any $n \in \mathbb{N}$? Without loss of generality, it is sufficient to consider the spherical case (Fang, Kotz, and Ng (1987), pages 48–51). F is the marginal distribution of a spherical distribution in $\mathbb{R}^n$ for any $n \in \mathbb{N}$ if and only if F is a mixture of centred normal distributions. In other words, if F has a density f, the latter is of the form,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sigma} \exp\Big(-\frac{x^2}{2\sigma^2}\Big) G(d\sigma),$$

where G is a distribution function on $[0,\infty)$ with $G(0) = 0$. The corresponding spherical distribution has the alternative stochastic representation

$$X =_{\mathrm{d}} S \cdot Z,$$

where $S \sim G$, $Z \sim \mathcal{N}_n(0, I_{n\times n})$ and S and Z are independent. For example, the multivariate t-distribution with ν degrees of freedom can be constructed by taking $S \sim \sqrt{\nu}/\sqrt{\chi^2_\nu}$.

3.4. **Covariance and elliptical distributions in risk management.** A further important feature of the elliptical distributions is that these distributions are amenable to the *standard* approaches of risk management. They support both the use of *Value-at-Risk* as a measure of risk and the mean-variance (Markowitz) approach (see e.g. Campbell, Lo, and MacKinlay (1997)) to risk management and portfolio optimization.

Suppose that $\mathbf{X} = (X_1, \dots, X_n)^t$ represents n risks with an elliptical distribution and that we consider *linear* portfolios of such risks

$$\{Z = \sum_{i=1}^n \lambda_i X_i \mid \lambda_i \in \mathbb{R}\}$$

with distribution F_Z. The Value-at-Risk (VaR) of portfolio Z at probability level α is given by

$$\mathrm{VaR}_\alpha(Z) = F_Z^{-1}(\alpha) = \inf\{z \in \mathbb{R} : F_Z(z) \geq \alpha\};$$

i.e. it is simply an alternative notation for the quantile function of F_Z evaluated at α and we will often use $\mathrm{VaR}_\alpha(Z)$ and $F_Z^{-1}(\alpha)$ interchangeably.

In the elliptical world the use of VaR as a measure of the risk of a portfolio Z makes sense because VaR is a coherent risk measure in this world. A coherent risk measure in the sense of Artzner, Delbaen, Eber, and Heath (1999) is a real-valued function ϱ on the space of real-valued random variables[5] which fulfills the following (sensible) properties:

A1. (Positivity). For any positive random variable $X \geq 0$: $\varrho(X) \geq 0$.

A2. (Subabdditivity). For any two random variables X and Y we have

$$\varrho(X+Y) \leq \varrho(X) + \varrho(Y).$$

A3. (Positive homogeneity). For $\lambda \geq 0$ we have that $\varrho(\lambda X) = \lambda \varrho(X)$.

A4. (Translation invariance).

$$\text{For any } a \in \mathbb{R} \text{ we have that } \varrho(X+a) = \varrho(X) + a.$$

In the elliptical world the use of any positive homogeneous, translation-invariant measure of risk to rank risks or to determine optimal risk-minimizing portfolio weights under the condition that a certain *return* is attained, is equivalent to the Markowitz approach where the variance is used as risk measure. Alternative risk measures such as VaR_α or expected shortfall, $\mathbb{E}[Z|Z > \mathrm{VaR}_\alpha(Z)]$, give different numerical values, but have no effect on the management of risk. We make these assertions more precise in Theorem 1.

Throughout this paper for notational and pedagogical reasons we use VaR in its most simplistic form, i.e. disregarding questions of appropriate horizon, estimation of the underlying profit-and-loss distribution, etc. However, the key messages stemming from this oversimplified view carry over to more concrete VaR calculations in practice.

Theorem 1. *Suppose* $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$ *with* $\sigma^2[X_i] < \infty$ *for all* i. *Let*

$$\mathcal{P} = \{Z = \sum_{i=1}^{n} \lambda_i X_i \mid \lambda_i \in \mathbb{R}\}$$

be the set of all linear portfolios. Then the following are true.

1. *(Subadditivity of VaR.) For any two portfolios* $Z_1, Z_2 \in \mathcal{P}$ *and* $0.5 \leq \alpha < 1$,
$$VaR_\alpha(Z_1 + Z_2) \leq VaR_\alpha(Z_1) + VaR_\alpha(Z_2).$$
2. *(Equivalence of variance and positive homogeneous risk measurement.) Let* ϱ *be a real-valued risk measure on the space of real-valued random variables which depends only on the distribution of a random variable* X. *Suppose this measure satisfies A3. Then for* $Z_1, Z_2 \in \mathcal{P}$
$$\varrho(Z_1 - \mathbb{E}[Z_1]) \leq \varrho(Z_2 - \mathbb{E}[Z_2]) \iff \sigma^2[Z_1] \leq \sigma^2[Z_2].$$
3. *(Markowitz risk-minimizing portfolio.) Let* ϱ *be as in 2 and assume that A4 is also satisfied. Let*
$$\mathcal{E} = \{Z = \sum_{i=1}^{n} \lambda_i X_i \mid \lambda_i \in \mathbb{R}, \sum_{i=1}^{n} \lambda_i = 1, \mathbb{E}[Z] = r\}$$
be the subset of portfolios giving expected return r. *Then*
$$argmin_{Z \in \mathcal{E}}\ \varrho(Z) = argmin_{Z \in \mathcal{E}}\ \sigma^2[Z].$$

Proof. The main observation is that $(Z_1, Z_2)^t$ has an elliptical distribution so Z_1, Z_2 and $Z_1 + Z_2$ all have distributions of the same type.

[5]Positive values of these random variables should be interpreted as *losses*; this is in contrast to Artzner, Delbaen, Eber, and Heath (1999), who interpret *negative* values as losses.

1. Let q_α be the α-quantile of the standardised distribution of this type. Then
$$\begin{aligned}\mathrm{VaR}_\alpha(Z_1) &= \mathbb{E}[Z_1] + \sigma[Z_1]q_\alpha, \\ \mathrm{VaR}_\alpha(Z_2) &= \mathbb{E}[Z_2] + \sigma[Z_2]q_\alpha, \\ \mathrm{VaR}_\alpha(Z_1+Z_2) &= \mathbb{E}[Z_1+Z_2] + \sigma[Z_1+Z_2]q_\alpha.\end{aligned}$$
Since $\sigma[Z_1+Z_2] \leq \sigma[Z_1]+\sigma[Z_2]$ and $q_\alpha \geq 0$ the result follows.
2. Since Z_1 and Z_2 are random variables of the same type, there exists an $a > 0$ such that $Z_1 - \mathbb{E}[Z_1] =_{\mathrm{d}} a(Z_2 - \mathbb{E}[Z_2])$. It follows that
$$\varrho(Z_1 - \mathbb{E}[Z_1]) \leq \varrho(Z_2 - \mathbb{E}[Z_2]) \iff a \leq 1 \iff \sigma^2[Z_1] \leq \sigma^2[Z_2].$$
3. Follows from 2 and the fact that we optimize over portfolios with identical expectation.

□

While this theorem shows that in the elliptical world the Markowitz variance-minimizing portfolio minimizes popular risk measures like VaR and expected shortfall (both of which are coherent in this world), it can also be shown that the Markowitz portfolio minimizes some other risk measures which do not satisfy A3 and A4. The partial moment measures of downside risk provide an example. The kth (upper) partial moment of a random variable X with respect to a threshold τ is defined to be
$$\mathrm{LPM}_{k,\tau}(X) = \mathbb{E}\left[\{(X-\tau)_+\}^k\right], \ k \geq 0, \tau \in \mathbb{R}.$$

Suppose we have portfolios $Z_1, Z_2 \in \mathcal{E}$ and assume additionally that $r \leq \tau$, so that the threshold is set above the expected return r. Using a similar approach to the preceding theorem it can be shown that
$$\sigma^2[Z_1] \leq \sigma^2[Z_2] \iff (Z_1 - \tau) =_{\mathrm{d}} a(Z_2 - \tau) - (1-a)(\tau - r),$$
with $0 < a \leq 1$. It follows that
$$\mathrm{LPM}_{k,\tau}(Z_1) \leq \mathrm{LPM}_{k,\tau}(Z_2) \iff \sigma^2[Z_1] \leq \sigma^2[Z_2],$$
from which the equivalence to Markowitz is clear. See Harlow (1991) for an empirical case study of the change in the optimal asset allocation when $\mathrm{LPM}_{1,\tau}$ (target shortfall) and $\mathrm{LPM}_{2,\tau}$ (target semi-variance) are used.

4. Alternative dependence concepts

We begin by clarifying what we mean by the notion of perfect dependence. We go on to discuss other measures of dependence, in particular rank correlation. We concentrate on pairs of random variables.

4.1. **Comonotonicity.** For every copula the well-known Fréchet bounds apply (Fréchet (1957))
$$\underbrace{\max\{x_1+\cdots+x_n+1-n, 0\}}_{C_\ell(x_1,\dots,x_n)} \leq C(x_1,\dots,x_n) \leq \underbrace{\min\{x_1,\dots,x_n\}}_{C_u(x_1,\dots,x_n)}; \qquad (12)$$
these follow from the fact that every copula is the distribution function of a random vector $(U_1,\dots,U_n)^t$ with $U_i \sim \mathrm{U}(0,1)$. In the case $n=2$ the bounds C_ℓ and C_u are themselves copulas since, if $U \sim \mathrm{U}(0,1)$, then
$$\begin{aligned}C_\ell(x_1,x_2) &= \mathbb{P}[U \leq x_1, 1-U \leq x_2] \\ C_u(x_1,x_2) &= \mathbb{P}[U \leq x_1, U \leq x_2],\end{aligned}$$

so that C_ℓ and C_u are the bivariate distribution functions of the vectors $(U, 1-U)^t$ and $(U, U)^t$ respectively.

The distribution of $(U, 1-U)^t$ has all its mass on the diagonal between $(0,1)$ and $(1,0)$, whereas that of $(U, U)^t$ has its mass on the diagonal between $(0,0)$ and $(1,1)$. In these cases we say that C_ℓ and C_u describe perfect positive and perfect negative dependence respectively. This is formalized in the following theorem.

Theorem 2. *Let $(X, Y)^t$ have one of the copulas C_ℓ or C_u.*[6] *(In the former case this means $F(x_1, x_2) = \max\{F_1(x_1) + F_2(x_2) - 1, 0\}$; in the latter $F(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\}$.) Then there exist two monotonic functions $u, v : \mathbb{R} \to \mathbb{R}$ and a real-valued random variable Z so that*

$$(X, Y)^t =_d (u(Z), v(Z))^t,$$

with u increasing and v decreasing in the former case and with both increasing in the latter. The converse of this result is also true.

Proof. The proof for the second case is given essentially in Wang and Dhaene (1998). A geometrical interpretation of Fréchet copulas is given in Mikusinski, Sherwood, and Taylor (1992). We consider only the first case $C = C_\ell$, the proofs being similar. Let U be a $\mathrm{U}(0,1)$-distributed random variable. We have

$$(X, Y)^t =_{\mathrm{d}} (F_1^{-1}(U), F_2^{-1}(1-U))^t = (F_1^{-1}(U), F_2^{-1} \circ g\,(U))^t,$$

where $F_i^{-1}(q) = \inf_{x \in \mathbb{R}}\{F_i(x) \geq q\}$, $q \in (0,1)$ is the quantile function of F_i, $i = 1, 2$, and $g(x) = 1 - x$. It follows that $u := F_1^{-1}$ is increasing and $v := F_2^{-1} \circ g$ is decreasing. For the converse assume

$$(X, Y)^t =_{\mathrm{d}} (u(Z), v(Z))^t,$$

with u and v increasing and decreasing respectively. We define $A := \{Z \in u^{-1}((-\infty, x])\}$, $B := \{Z \in v^{-1}((-\infty, y])\}$. If $A \cap B \neq \emptyset$ then the monotonicity of u and v imply that

$$\mathbb{P}[A \cup B] = \mathbb{P}[\Omega] = 1 = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

and hence $\mathbb{P}[A \cap B] = \mathbb{P}[u(Z) \leq x, v(Z) \leq y] = F_1(x) + F_2(y) - 1$. If $A \cap B = \emptyset$, then $F_1(x) + F_2(y) - 1 \leq 0$. In all cases we have

$$\mathbb{P}[u(Z) \leq x, v(Z) \leq y] = \max\{F_1(x) + F_2(y) - 1, 0\}.$$

□

We introduce the following terminology.

Definition 5. [Yaari (1987)] If $(X, Y)^t$ has the copula C_u (see again footnote 6) then X and Y are said to be *comonotonic*; if it has copula C_ℓ they are said to be *countermonotonic*.

In the case of continuous distributions F_1 and F_2 a stronger version of the result can be stated:

$$C = C_\ell \iff Y = T(X) \text{ a.s.}, \; T = F_2^{-1} \circ (1 - F_1) \text{ decreasing}, \tag{13}$$

$$C = C_u \iff Y = T(X) \text{ a.s.}, \; T = F_2^{-1} \circ F_1 \text{ increasing}. \tag{14}$$

[6] If there are discontinuities in F_1 or F_2 so that the copula is not unique, then we interpret C_ℓ and C_u as being *possible copulas*.

4.2. **Desired properties of dependency measures.** A measure of dependence, like linear correlation, summarises the dependency structure of two random variables in a single number. We consider the properties that we would like to have from this measure. Let $\delta(\cdot,\cdot)$ be a dependency measure which assigns a real number to any pair of real-valued random variables X and Y. Ideally, we desire the following properties:

P1. $\delta(X, Y) = \delta(Y, X)$ (symmetry).
P2. $-1 \le \delta(X, Y) \le 1$ (normalisation).
P3. $\delta(X, Y) = 1 \iff X, Y$ comonotonic;
$\delta(X, Y) = -1 \iff X, Y$ countermonotonic.
P4. For $T : \mathbb{R} \to \mathbb{R}$ strictly monotonic on the range of X:

$$\delta(T(X), Y) = \begin{cases} \delta(X, Y) & \text{T increasing,} \\ -\delta(X, Y) & \text{T decreasing.} \end{cases}$$

Linear correlation fulfills properties P1 and P2 only. In the next Section we see that rank correlation also fulfills P3 and P4 if X and Y are continuous. These properties obviously represent a selection and the list could be altered or extended in various ways (see Hutchinson and Lai (1990), Chapter 11). For example, we might like to have the property

P5. $\delta(X, Y) = 0 \iff X, Y$ are independent.

Unfortunately, this contradicts property P4 as the following shows.

Proposition 3. *There is no dependency measure satisfying P4 and P5.*

Proof. Let $(X, Y)^t$ be uniformly distributed on the unit circle S^1 in $\mathbb{R}^2$, so that $(X, Y)^t = (\cos\phi, \sin\phi)^t$ with $\phi \sim \mathrm{U}(0, 2\pi)$. Since $(-X, Y)^t =_{\mathrm{d}} (X, Y)^t$, we have

$$\delta(-X, Y) = \delta(X, Y) = -\delta(X, Y),$$

which implies $\delta(X, Y) = 0$ although X and Y are dependent. With the same argumentation it can be shown that the measure is zero for any *spherical* distribution in $\mathbb{R}^2$. $\square$

If we require P5, then we can consider dependency measures which only assign positive values to pairs of random variables. For example, we can consider the amended properties,

P2b. $0 \le \delta(X, Y) \le 1$.
P3b. $\delta(X, Y) = 1 \iff X, Y$ comonotonic or countermonotonic.
P4b. For $T : \mathbb{R} \to \mathbb{R}$ strictly monotonic $\delta(T(X), Y) = \delta(X, Y)$.

If we restrict ourselves to the case of continuous random variables there are dependency measures which fulfill all of P1, P2b, P3b, P4b and P5, although they are in general measures of theoretical rather than practical interest. We introduce them briefly in the next Section. A further measure which satisfies all of P1, P2b, P3b, P4b and P5 (with the exception of the implication $\delta(X, Y) = 1 \implies X, Y$ comonotonic or countermonotonic) is *monotone correlation*,

$$\delta(X, Y) = \sup_{f,g} \rho(f(X), g(Y)),$$

where ρ represents linear correlation and the supremum is taken over all monotonic f and g such that $0 < \sigma^2(f(X)), \sigma^2(g(Y)) < \infty$ (Kimeldorf and Sampson 1978). The disadvantage of all of these measures is that they are constrained to give non-negative values and as such cannot differentiate between positive and negative dependence and that it is often not clear how to estimate them. An overview of dependency measures and their statistical estimation is given by Tjøstheim (1996).

4.3. **Rank correlation.**

Definition 6. Let X and Y be random variables with distribution functions F_1 and F_2 and joint distribution function F. Spearman's rank correlation is given by

$$\rho_S(X,Y) = \rho(F_1(X), F_2(Y)), \tag{15}$$

where ρ is the usual linear correlation. Let (X_1, Y_1) and (X_2, Y_2) be two independent pairs of random variables from F, then Kendall's rank correlation is given by

$$\rho_\tau(X,Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{16}$$

For the remainder of this Section we assume that F_1 and F_2 are continuous distributions, although some of the properties of rank correlation that we derive could partially be formulated for discrete distributions. Spearman's rank correlation is then seen to be the correlation of the copula C associated with $(X,Y)^t$. Both ρ_S and ρ_τ can be considered to be measures of the degree of monotonic dependence between X and Y, whereas linear correlation measures the degree of linear dependence only. The generalisation of ρ_S and ρ_τ to $n > 2$ dimensions can be done analogously to that of linear correlation: we write pairwise correlations in a $n \times n$-matrix.

We collect together the important facts about ρ_S and ρ_τ in the following theorem.

Theorem 3. *Let X and Y be random variables with continuous distributions F_1 and F_2, joint distribution F and copula C. The following are true:*

1. $\rho_S(X,Y) = \rho_S(Y,X)$, $\rho_\tau(X,Y) = \rho_\tau(Y,X)$.
2. *If X and Y are independent then* $\rho_S(X,Y) = \rho_\tau(X,Y) = 0$.
3. $-1 \le \rho_S(X,Y), \rho_\tau(X,Y) \le +1$.
4. $\rho_S(X,Y) = 12 \int_0^1 \int_0^1 \{C(x,y) - xy\} dx dy$.
5. $\rho_\tau(X,Y) = 4 \int_0^1 \int_0^1 C(u,v) dC(u,v) - 1$.
6. *For $T : \mathbb{R} \to \mathbb{R}$ strictly monotonic on the range of X, both ρ_S and ρ_τ satisfy P4.*
7. $\rho_S(X,Y) = \rho_\tau(X,Y) = 1 \iff C = C_u \iff Y = T(X)$ *a.s. with T increasing.*
8. $\rho_S(X,Y) = \rho_\tau(X,Y) = -1 \iff C = C_\ell \iff Y = T(X)$ *a.s. with T decreasing.*

Proof. 1., 2. and 3. are easily verified.

4. Use of the identity, due to Höffding (1940)

$$\mathrm{Cov}[X,Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F(x,y) - F_1(x)F_2(y)\}\, dx dy \tag{17}$$

which is found, for example, in Dhaene and Goovaerts (1996). Recall that $(F_1(X), F_2(Y))^t$ have joint distribution C.

5. Calculate

$$\begin{aligned}
\rho_\tau(X,Y) &= 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \\
&= 2 \cdot 2 \iiiint_{\mathbb{R}^4} \mathbf{1}_{\{x_1 > x_2\}} \mathbf{1}_{\{y_1 > y_2\}}\, dF(x_2, y_2)\, dF(x_1, y_1) - 1 \\
&= 4 \iint_{\mathbb{R}^2} F(x_1, y_1)\ dF(x_1, y_1) - 1 \\
&= 4 \iint C(u,v)\, dC(u,v) - 1.
\end{aligned}$$

6. Follows since ρ_τ and ρ_S can both be expressed in terms of the copula which is invariant under strictly increasing transformations of the marginals.
7. From 4. it follows immediately that $\rho_S(X,Y) = +1$ iff $C(x,y)$ is maximized iff $C = C_u$ iff $Y = T(X)$ a.s. Suppose $Y = T(X)$ a.s. with T increasing, then the continuity of F_2 ensures $\mathbb{P}[Y_1 = Y_2] = \mathbb{P}[T(X_1) = T(X_2)] = 0$, which implies $\rho_\tau(X,Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] = 1$. Conversely $\rho_\tau(X,Y) = 1$ means $\mathbb{P} \otimes \mathbb{P}[(\omega_1, \omega_2) \in \Omega \times \Omega | (X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) > 0\}] = 1$. Let us define sets $A = \{\omega \in \Omega | X(w) \leq x\}$ and $B = \{\omega \in \Omega | Y(w) \leq y\}$. Assume $\mathbb{P}[A] \leq \mathbb{P}[B]$. We have to show $\mathbb{P}[A \cap B] = \mathbb{P}[A]$. If $\mathbb{P}[A \setminus B] > 0$ then also $\mathbb{P}[B \setminus A] > 0$ and $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) < 0$ on the set $(A \setminus B) \times (B \setminus A)$, which has measure $\mathbb{P}[A \setminus B] \cdot \mathbb{P}[B \setminus A] > 0$, and this is a contradiction. Hence $\mathbb{P}[A \setminus B] = 0$, from which one concludes $\mathbb{P}[A \cap B] = \mathbb{P}[A]$.
8. We use a similar argument to 7.

□

In this result we have verified that rank correlation does have the properties P1, P2, P3 and P4. As far as P5 is concerned, the spherical distributions again provide examples where pairwise rank correlations are zero, despite the presence of dependence.

Theorem 3 (part 4) shows that ρ_S is a scaled version of the signed volume enclosed by the surfaces $S_1 : \; z = C(x,y)$ and $S_2 : \; z = xy$. The idea of measuring dependence by defining suitable distance measures between the surfaces S_1 and S_2 is further developed in Schweizer and Wolff (1981), where the three measures

$$\begin{aligned}
\delta_1(X,Y) &= 12 \int_0^1 \int_0^1 |C(u,w) - uv| du dv \\
\delta_2(X,Y) &= \left(90 \int_0^1 \int_0^1 |C(u,w) - uv|^2 du dv\right)^{1/2} \\
\delta_3(X,Y) &= 4 \sup_{u,v \in [0,1]} |C(u,v) - uv|
\end{aligned}$$

are proposed. These are the measures that satisfy our amended set of properties including P5 but are constrained to give non-negative measurements and as such cannot differentiate between positive and negative dependence. A further disadvantage of these measures is statistical. Whereas statistical estimation of ρ_S and ρ_τ from data is straightforward (see Gibbons (1988) for the estimators and Tjøstheim (1996) for asymptotic estimation theory) it is much less clear how we estimate measures like $\delta_1, \delta_2, \delta_3$.

The main advantages of rank correlation over ordinary correlation are the invariance under monotonic transformations and the sensible handling of perfect dependence. The main disadvantage is that rank correlations do not lend themselves to the same elegant variance-covariance manipulations that were discussed for linear correlation; they are not moment-based correlations. As far as calculation is concerned, there are cases where rank correlations are easier to calculate and cases where linear correlations are easier to calculate. If we are working, for example, with multivariate normal or t-distributions then calculation of linear correlation is easier, since first and second moments are easily determined. If we are working with a multivariate distribution which possesses a simple closed-form copula, like the Gumbel or Farlie-Gumbel-Morgenstern, then moments may be difficult to determine and calculation of rank correlation using Theorem 3 (parts 4 and 5) may be easier.

4.4. **Tail Dependence.** If we are particularly concerned with extreme values an asymptotic measure of tail dependence can be defined for pairs of random variables X and Y. If the marginal distributions of these random variables are continuous then this dependency measure is also a function of their copula, and is thus invariant under strictly increasing transformations.

Definition 7. Let X and Y be random variables with distribution functions F_1 and F_2. The *coefficient of (upper) tail dependence* of X and Y is

$$\lim_{\alpha \to 1-} \mathbb{P}[Y > F_2^{-1}(\alpha) \mid X > F_1^{-1}(\alpha)] = \lambda,$$

provided a limit $\lambda \in [0,1]$ exists. If $\lambda \in (0,1]$ X and Y are said to be *asymptotically dependent* (in the upper tail); if $\lambda = 0$ they are *asymptotically independent*.

As for rank correlation, this definition makes most sense in the case that F_1 and F_2 are continuous distributions. In this case it can be verified, under the assumption that the limit exists, that

$$\begin{aligned}
&\lim_{\alpha \to 1-} \mathbb{P}[Y > F_2^{-1}(\alpha) \mid X > F_1^{-1}(\alpha)] \\
= &\lim_{\alpha \to 1-} \mathbb{P}[Y > \mathrm{VaR}_\alpha(Y) \mid X > \mathrm{VaR}_\alpha(X)] = \lim_{\alpha \to 1-} \frac{\overline{C}(\alpha,\alpha)}{1-\alpha},
\end{aligned}$$

where $\overline{C}(u,u) = 1 - 2u + C(u,u)$ denotes the survivor function of the unique copula C associated with $(X,Y)^t$. Tail dependence is best understood as an asymptotic property of the copula.

Calculation of λ for particular copulas is straightforward if the copula has a simple closed form. For example, for the Gumbel copula introduced in (5) it is easily verified that $\lambda = 2 - 2^\beta$, so that random variables with this copula are asymptotically dependent provided $\beta < 1$.

For copulas without a simple closed form, such as the Gaussian copula or the copula of the bivariate t-distribution, an alternative formula for λ is more useful. Consider a pair of uniform random variables $(U_1, U_2)^t$ with distribution $C(x,y)$, which we assume is differentiable in both x and y. Applying l'Hospital's rule we obtain

$$\lambda = -\lim_{x \to 1-} \frac{d\overline{C}(x,x)}{dx} = \lim_{x \to 1-} \Pr[U_2 > x \mid U_1 = x] + \lim_{x \to 1-} \Pr[U_1 > x \mid U_2 = x].$$

Furthermore, if C is an *exchangeable* copula, i.e. $(U_1,U_2)^t =_{\mathrm{d}} (U_2,U_1)^t$, then

$$\lambda = 2 \lim_{x \to 1-} \Pr[U_2 > x \mid U_1 = x].$$

It is often possible to evaluate this limit by applying the same quantile transform F_1^{-1} to both marginals to obtain a bivariate distribution for which the conditional probability is known. If F_1 is a distribution function with infinite right endpoint then

$$\begin{aligned}
\lambda = 2 \lim_{x \to 1-} \Pr[U_2 > x \mid U_1 = x] &= 2 \lim_{x \to \infty} \Pr[F_1^{-1}(U_2) > x \mid F_1^{-1}(U_1) = x] \\
&= 2 \lim_{x \to \infty} \Pr[Y > x \mid X = x],
\end{aligned}$$

where $(X,Y)^t \sim C(F_1(x), F_1(y))$.

For example, for the Gaussian copula C_ρ^{Ga} we would take $F_1 = \Phi$ so that $(X,Y)^t$ has a standard bivariate normal distribution with correlation ρ. Using the fact that

$Y \mid X = x \sim N(\rho x, 1-\rho^2)$, it can be calculated that

$$\lambda = 2 \lim_{x\to\infty} \overline{\Phi}(x\sqrt{1-\rho}/\sqrt{1+\rho}).$$

Thus the Gaussian copula gives asymptotic independence, provided that $\rho < 1$. Regardless of how high a correlation we choose, if we go far enough into the tail, extreme events appear to occur independently in each margin. See Sibuya (1961) or Resnick (1987), Chapter 5, for alternative demonstrations of this fact.

The bivariate t-distribution provides an interesting contrast to the bivariate normal distribution. If $(X, Y)^t$ has a standard bivariate t-distribution with ν degrees of freedom and correlation ρ then, conditional on $X = x$,

$$\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{Y-\rho x}{\sqrt{1-\rho^2}} \sim t_{\nu+1}.$$

This can be used to show that

$$\lambda = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right),$$

where $\bar{t}_{\nu+1}$ denotes the tail of a univariate t-distribution. Provided $\rho > -1$ the copula of the bivariate t-distribution is asymptotically dependent. In Table 1 we tabulate the coefficient of tail dependence for various values of ν and ρ. Perhaps surprisingly, even for negative and zero correlations, the t-copula gives asymptotic dependence in the upper tail. The strength of this dependence increases as ν decreases and the marginal distributions become heavier-tailed.

$\nu \setminus \rho$	-0.5	0	0.5	0.9	1
2	0.06	0.18	0.39	0.72	1
4	0.01	0.08	0.25	0.63	1
10	0.0	0.01	0.08	0.46	1
∞	0	0	0	0	1

TABLE 1. Values of λ for the copula of the bivariate t-distribution for various values of ν, the degrees of freedom, and ρ, the correlation. Last row represents the Gaussian copula.

In Figure 2 we plot exact values of the conditional probability $\mathbb{P}[Y > \mathrm{VaR}_\alpha(Y) \mid X = \mathrm{VaR}_\alpha(X)]$ for pairs of random variables $(X, Y)^t$ with the Gaussian and t-copulas, where the correlation parameter of both copulas is $\rho = 0.9$ and the degrees of freedom of the t-copula is $\nu = 4$. For large values of α the conditional probabilities for the t-copula dominate those for the Gaussian copula. Moreover the former tend towards a non-zero asymptotic limit, whereas the limit in the Gaussian case is zero.

4.5. **Concordance.** In some situations we may be less concerned with measuring the strength of stochastic dependence between two random variables X and Y and we may wish simply to say whether they are *concordant* or *discordant*, that is, whether the dependence between X and Y is *positive* or *negative*. While it might seem natural to define X and Y to be positively dependent when $\rho(X, Y) > 0$ (or when $\rho_S(X, Y) > 0$ or $\rho_\tau(X, Y) > 0$), stronger conditions are generally used and we discuss two of these concepts in this Section.

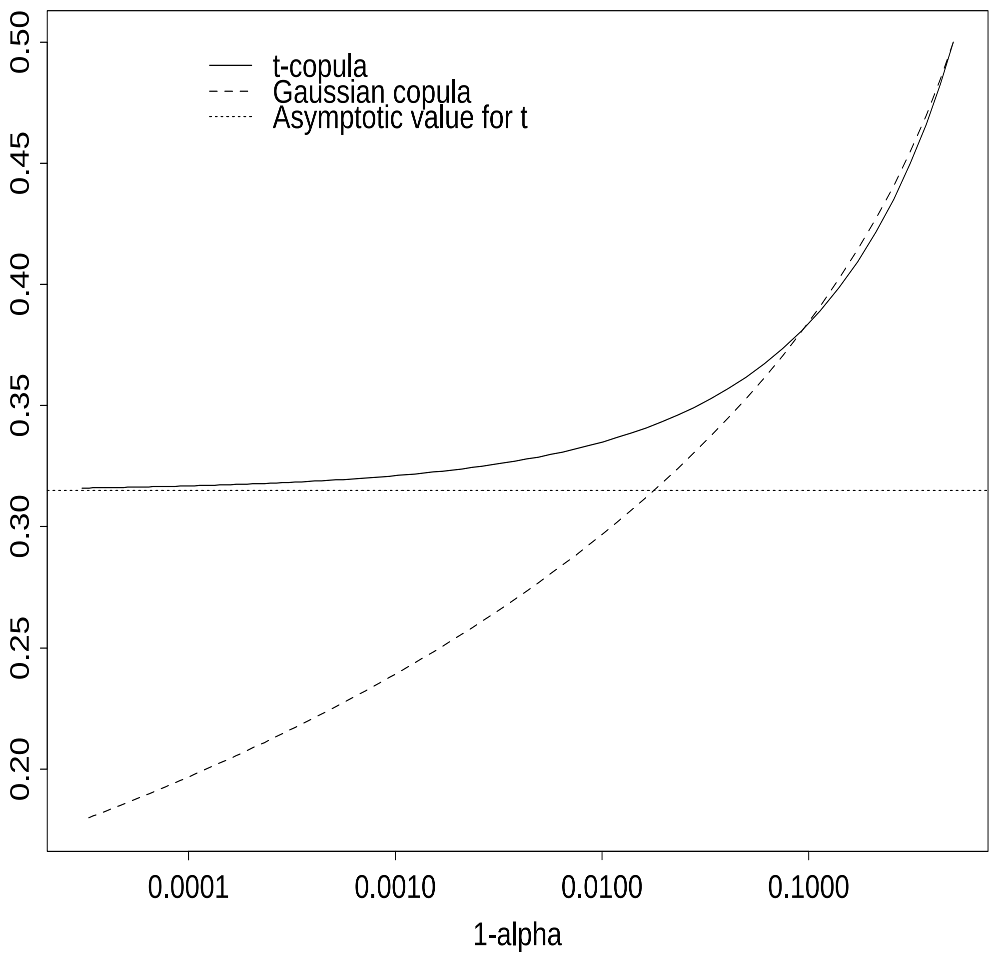

Figure 2. Exact values of the conditional probability $\mathbb{P}[Y > \mathrm{VaR}_\alpha(Y) \mid X = \mathrm{VaR}_\alpha(X)]$ for pairs of random variables $(X, Y)^t$ with the Gaussian and t-copulas, where the correlation parameter in both copulas is $\rho = 0.9$ and the degrees of freedom of the t-copula is $\nu = 4$.

Definition 8. Two random variables X and Y are *positive quadrant dependent* (PQD), if

$$\mathbb{P}[X > x, Y > y] \geq \mathbb{P}[X > x]\mathbb{P}[Y > y] \quad \text{for all } x, y \in \mathbb{R}. \tag{18}$$

Since $\mathbb{P}[X > x, Y > y] = 1 - \mathbb{P}[X \leq x] + \mathbb{P}[Y \leq y] - \mathbb{P}[X \leq x, Y \leq y]$ it is obvious that (18) is equivalent to

$$\mathbb{P}[X \leq x, Y \leq y] \geq \mathbb{P}[X \leq x]\mathbb{P}[Y \leq y] \quad \text{for all } x, y \in \mathbb{R}.$$

Definition 9. Two random variables X and Y are *positively associated* (PA), if

$$\mathbb{E}[g_1(X, Y)g_2(X, Y)] \geq \mathbb{E}[g_1(X, Y)]\mathbb{E}[g_2(X, Y)] \tag{19}$$

for all real-valued, measurable functions g_1 und g_2, which are increasing in both components and for which the expectations above are defined.

For further concepts of positive dependence see Chapter 2 of Joe (1997), where the relationships between the various concepts are also systematically explored. We note that PQD and PA are invariant under increasing transformations and we verify that the following chain of implications holds:

$$\text{Comonotonicity} \Rightarrow \text{PA} \Rightarrow \text{PQD} \Rightarrow \rho(X,Y) \geq 0, \rho_S(X,Y) \geq 0, \rho_\tau(X,Y) \geq 0. \quad (20)$$

If X and Y are comonotonic, then from Theorem 2 we can conclude that $(X,Y) =_{\mathrm{d}} (F_1^{-1}(U), F_2^{-1}(U))$, where $U \sim \mathrm{U}(0,1)$. Thus the expectations in (19) can be written as

$$\mathbb{E}[g_1(X,Y)g_2(X,Y)] = \mathbb{E}[\widetilde{g}_1(U)\widetilde{g}_2(U)]$$

and

$$\mathbb{E}[g_1(X,Y)] = \mathbb{E}[\widetilde{g}_1(U)] \ , \ \mathbb{E}[g_2(X,Y)] = \mathbb{E}[\widetilde{g}_2(U)],$$

where $\widetilde{g}_1$ and $\widetilde{g}_2$ are increasing. Lemma 2.1 in Joe (1997) shows that

$$\mathbb{E}[\widetilde{g}_1(U)\widetilde{g}_2(U)] \geq \mathbb{E}[\widetilde{g}_1(U)]\mathbb{E}[\widetilde{g}_2(U)],$$

so that X and Y are PA. The second implication follows immediately by taking

$$\begin{aligned} g_1(u,v) &= \mathbf{1}_{\{u>x\}} \\ g_2(u,v) &= \mathbf{1}_{\{v>y\}}. \end{aligned}$$

The third implication PQD $\Rightarrow \rho(X,Y) \geq 0, \rho_S(X,Y) \geq 0$ follows from the identity (17) and the fact that PA and PQD are invariant under increasing transformations. PQD $\Rightarrow \rho_\tau(X,Y) \geq 0$ follows from Theorem 2.8 in Joe (1997).

In the sense of these implications (20), comonotonicity is the *strongest* type of concordance or positive dependence.

5. Fallacies

Where not otherwise stated, we consider bivariate distributions of the random vector $(X,Y)^t$.

Fallacy 1. Marginal distributions and correlation determine the joint distribution.

This is true if we restrict our attention to the multivariate normal distribution or the elliptical distributions. For example, if we know that $(X,Y)^t$ have a bivariate normal distribution, then the expectations and variances of X and Y and the correlation $\rho(X,Y)$ uniquely determine the joint distribution. However, if we only know the marginal distributions of X and Y and the correlation then there are many possible bivariate distributions for $(X,Y)^t$. The distribution of $(X,Y)^t$ is not uniquely determined by F_1, F_2 and $\rho(X,Y)$. We illustrate this with examples, interesting in their own right.

Example 1. Let X and Y have standard normal distributions and let assume $\rho(X,Y) = \rho$. If $(X,Y)^t$ is *bivariate* normally distributed, then the distribution function F of $(X,Y)^t$ is given by

$$F(x,y) = C_\rho^{Ga}(\Phi(x), \Phi(y)).$$

We have represented this copula earlier as a double integral in (4). Any other copula $C \neq C_\rho^{Ga}$ gives a bivariate distribution with standard normal marginals which is *not* bivariate normal with correlation ρ. We construct a copula C of the type (6) by taking

$$\begin{aligned} f(x) &= \mathbf{1}_{\{(\gamma,1-\gamma)\}}(x) + \frac{2\gamma-1}{2\gamma}\mathbf{1}_{\{(\gamma,1-\gamma)^c\}}(x) \\ g(y) &= -\mathbf{1}_{\{(\gamma,1-\gamma)\}}(y) - \frac{2\gamma-1}{2\gamma}\mathbf{1}_{\{(\gamma,1-\gamma)^c\}}(y), \end{aligned}$$

where $\frac{1}{4} \leq \gamma \leq \frac{1}{2}$. Since $h(x,y)$ disappears on the square $[\gamma, 1-\gamma]^2$ it is clear that C for $\gamma < \frac{1}{2}$ and $F(x,y) = C(\Phi(x), \Phi(y))$ is *never* bivariate normal; from symmetry considerations ($C(u,v) = C(1-u,v)$, $0 \leq u, v \leq 1$) the correlation irrespective of γ is zero. There are uncountably many bivariate distributions with standard normal marginals and correlation zero. In Figure 3 the density of F is shown for $\gamma = 0.3$; this is clearly very different from the joint density of the standard bivariate normal distribution with zero correlation.

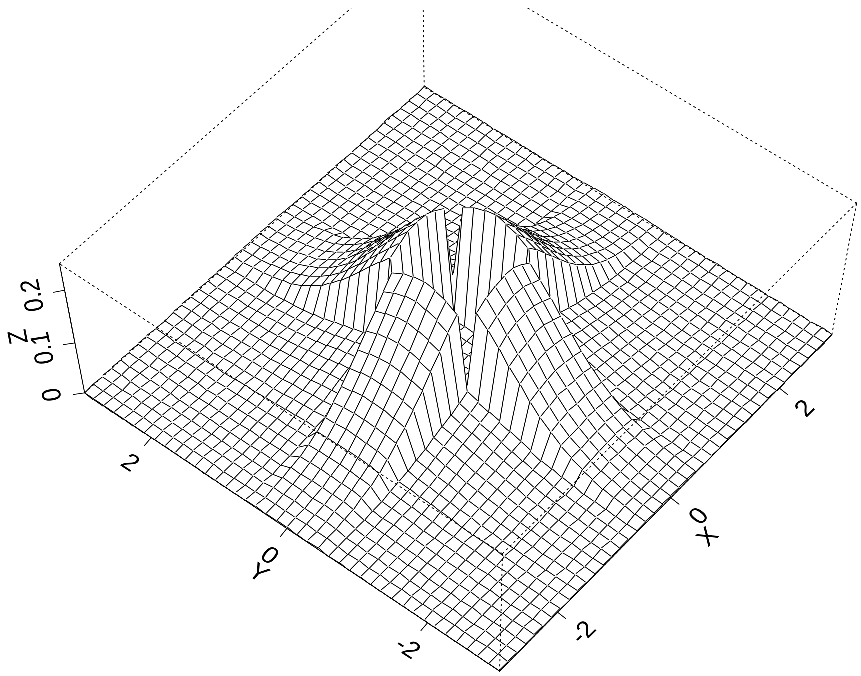

FIGURE 3. Density of a non-bivariate normal distribution which has standard normal marginals.

Example 2. A more realistic example for risk management is the motivating example of the Introduction. We consider two bivariate distributions with Gamma(3,1) marginals (denoted $\mathrm{G}_{3,1}$) and the same correlation $\rho = 0.7$, but with different dependence structures, namely

$$
\begin{aligned}
F_{Ga}(x,y) &= C_{\tilde{\rho}}^{Ga}(\mathrm{G}(x), \mathrm{G}(y)), \\
F_{Gu}(x,y) &= C_{\alpha}^{Gu}(\mathrm{G}(x), \mathrm{G}(y)),
\end{aligned}
$$

where $C_{\tilde{\rho}}^{Ga}$ is the Gaussian dependence structure and C_{β}^{Gu} is the Gumbel copula introduced in (5). To obtain the desired linear correlation the parameter values were set to be $\tilde{\rho} = 0.71$ and $\beta = 0.54$ [7].

In Section 4.4 we showed that the two copulas have quite different tail dependence; the Gaussian copula is asymptotically independent if $\tilde{\rho} < 1$ and the Gumbel copula is asymptotically dependent if $\beta < 1$. At finite levels the greater tail dependence of the Gumbel copula is apparent in Figure 1. We fix $u = \mathrm{VaR}_{0.99}(X) = \mathrm{VaR}_{0.99}(Y) =$

[7] These numerical values were determined by stochastic simulation.

$G_{3,1}^{-1}(0.99)$ and consider the conditional exceedance probability $\mathbb{P}[Y > u \mid X > u]$ under the two models. An easy empirical estimation based on Figure 1 yields

$$\begin{aligned}\widehat{\mathbb{P}}_{F_{Ga}}[Y > u \mid X > u] &= 3/9,\\ \widehat{\mathbb{P}}_{F_{Gu}}[Y > u \mid X > u] &= 12/16.\end{aligned}$$

In the Gumbel model exceedances of the threshold u in one margin tend to be accompanied by exceedances in the other, whereas in the Gaussian dependence model joint exceedances in both margins are rare. There is less "diversification" of large risks in the Gumbel dependence model.

Analytically it is difficult to provide results for the Value-at-Risk of the sum $X + Y$ under the two bivariate distributions,[8] but simulation studies confirm that $X + Y$ produces more large outcomes under the Gumbel dependence model than the Gaussian model. The difference between the two dependence structures might be particularly important if we were interested in losses which were triggered only by joint extreme values of X and Y.

Example 3. The Value-at-Risk of linear portfolios is certainly not uniquely determined by the marginal distributions and correlation of the constituent risks. Suppose $(X, Y)^t$ has a bivariate normal distribution with standard normal marginals and correlation ρ and denote the bivariate distribution function by F_ρ. Any mixture $F = \lambda F_{\rho_1} + (1 - \lambda)F_{\rho_2}$, $0 \le \lambda \le 1$ of bivariate normal distributions F_{ρ_1} and F_{ρ_2} also has standard normal marginals and correlation $\lambda\rho_1 + (1 - \lambda)\rho_2$. Suppose we fix $-1 < \rho < 1$ and choose $0 < \lambda < 1$ and $\rho_1 < \rho < \rho_2$ such that $\rho = \lambda\rho_1 + (1 - \lambda)\rho_2$. The sum $X + Y$ is longer tailed under F than under F_ρ. Since

$$\mathbb{P}_F[X + Y > z] = \lambda\overline{\Phi}\left(\frac{z}{2(1 + \rho_1)}\right) + (1 - \lambda)\overline{\Phi}\left(\frac{z}{2(1 + \rho_2)}\right),$$

and

$$\mathbb{P}_{F_\rho}[X + Y > z] = \overline{\Phi}\left(\frac{z}{2(1 + \rho)}\right),$$

we can use Mill's ratio

$$\overline{\Phi}(x) = 1 - \Phi(x) = \phi(x)\left(\frac{1}{x} + \mathrm{O}\left(\frac{1}{x^2}\right)\right)$$

to show that

$$\lim_{z\to\infty} \frac{\mathbb{P}_F[X + Y > z]}{\mathbb{P}_{F_\rho}[X + Y > z]} = \infty.$$

Clearly as one goes further into the respective tails of the two distributions the Value-at-Risk for the mixture distribution F is larger than that of the original distribution F_ρ. By using the same technique as Embrechts, Mikosch, and Klüppelberg (1997) (Example 3.3.29) we can show that, as $\alpha \to 1-$,

$$\begin{aligned}\mathrm{VaR}_{\alpha,F}(X + Y) &\sim 2(1 + \rho_2)\left(-2\log(1 - \alpha)\right)^{1/2}\\ \mathrm{VaR}_{\alpha,F_\rho}(X + Y) &\sim 2(1 + \rho)\left(-2\log(1 - \alpha)\right)^{1/2},\end{aligned}$$

so that

$$\lim_{\alpha\to 1-} \frac{\mathrm{VaR}_{\alpha,F}(X + Y)}{\mathrm{VaR}_{\alpha,F_\rho}(X + Y)} = \frac{1 + \rho_2}{1 + \rho} > 1,$$

[8]See Müller and Bäuerle (1998) for related work on stop-loss risk measures applied to bivariate portfolios under various dependence models.

irrespective of the choice of λ.

Fallacy 2. Given marginal distributions F_1 and F_2 for X and Y, all linear correlations between -1 and 1 can be attained through suitable specification of the joint distribution.

This statement is not true and it is simple to construct counterexamples.

Example 4. Let X and Y be random variables with support $[0, \infty)$, so that $F_1(x) = F_2(y) = 0$ for all $x, y < 0$. Let the right endpoints of F_1 and F_2 be infinite, $\sup_x\{x|F_1(x) < 1\} = \sup_y\{y|F_2(y) < 1\} = \infty$. Assume that $\rho(X, Y) = -1$, which would imply $Y = aX + b$ a.s., with $a < 0$ and $b \in \mathbb{R}$. It follows that for all $y < 0$,

$$\begin{aligned} F_2(y) &= \mathbb{P}[Y \le y] = \mathbb{P}[X \ge (y-b)/a] \ge \mathbb{P}[X > (y-b)/a] \\ &= 1 - F_1((y-b)/a) > 0, \end{aligned}$$

which contradicts the assumption $F_2(y) = 0$.

The following theorem shows which correlations are possible for given marginal distributions.

Theorem 4. [*Höffding (1940) and Fréchet (1957)*] *Let $(X, Y)^t$ be a random vector with marginals F_1 and F_2 and unspecified dependence structure; assume $0 < \sigma^2[X], \sigma^2[Y] < \infty$. Then*

1. *The set of all possible correlations is a closed interval $[\rho_{\min}, \rho_{\max}]$ and for the extremal correlations $\rho_{\min} < 0 < \rho_{\max}$ holds.*
2. *The extremal correlation $\rho = \rho_{\min}$ is attained if and only if X and Y are countermonotonic; $\rho = \rho_{\max}$ is attained if and only if X and Y are comonotonic.*
3. *$\rho_{\min} = -1$ iff X and $-Y$ are of the same type; $\rho_{\max} = 1$ iff X and Y are of the same type.*

Proof. We make use of the identity (17) and observe that the Fréchet inequalities (12) imply

$$\max\{F_1(x) + F_2(y) - 1, 0\} \le F(x, y) \le \min\{F_1(x), F_2(y)\}.$$

The integrand in (17) is minimized pointwise, if X and Y are countermonotonic and maximized if X and Y are comonotonic. It is clear that $\rho_{\max} \ge 0$. However, if $\rho_{\max} = 0$ this would imply that $\min\{F_1(x), F_2(y)\} = F_1(x)F_2(y)$ for all x, y. This can only occur if F_1 or F_2 is degenerate, i.e. of the form $F_1(x) = \mathbf{1}_{\{x \ge x_0\}}$ or $F_2(y) = \mathbf{1}_{\{y \ge y_0\}}$, and this would imply $\sigma^2[X] = 0$ or $\sigma^2[Y] = 0$ so that the correlation between X and Y is undefined. Similarly we argue that $\rho_{\min} < 0$. If $F_\ell(x_1, x_2) = \max\{F_1(x) + F_2(y) - 1, 0\}$ and $F_u(x_1, x_2) = \min\{F_1(x), F_2(y)\}$ then the mixture $\lambda F_\ell + (1-\lambda)F_u$, $0 \le \lambda \le 1$ has correlation $\lambda\rho_{\min} + (1-\lambda)\rho_{\max}$. Using such mixtures we can construct joint distributions with marginals F_1 and F_2 and with arbitrary correlations $\rho \in [\rho_{\min}, \rho_{\max}]$. This will be used in Section 6 □

Example 5. Let $X \sim \text{Lognormal}(0, 1)$ and $Y \sim \text{Lognormal}(0, \sigma^2)$, $\sigma > 0$. We wish to calculate $\rho_{\min}$ and $\rho_{\max}$ for these marginals. Note that X and Y are not of the same type although $\log X$ and $\log Y$ are. It is clear that $\rho_{\min} = \rho(e^Z, e^{-\sigma Z})$ and $\rho_{\max} = \rho(e^Z, e^{\sigma Z})$, where $Z \sim \mathcal{N}(0, 1)$. This observation allows us to calculate $\rho_{\min}$ and $\rho_{\max}$ analytically:

$$\rho_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}},$$

$$\rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}}.$$

These maximal and mininal correlations are shown graphically in Figure 4. We observe that $\lim_{\sigma \to \infty} \rho_{\min} = \lim_{\sigma \to \infty} \rho_{\max} = 0$.

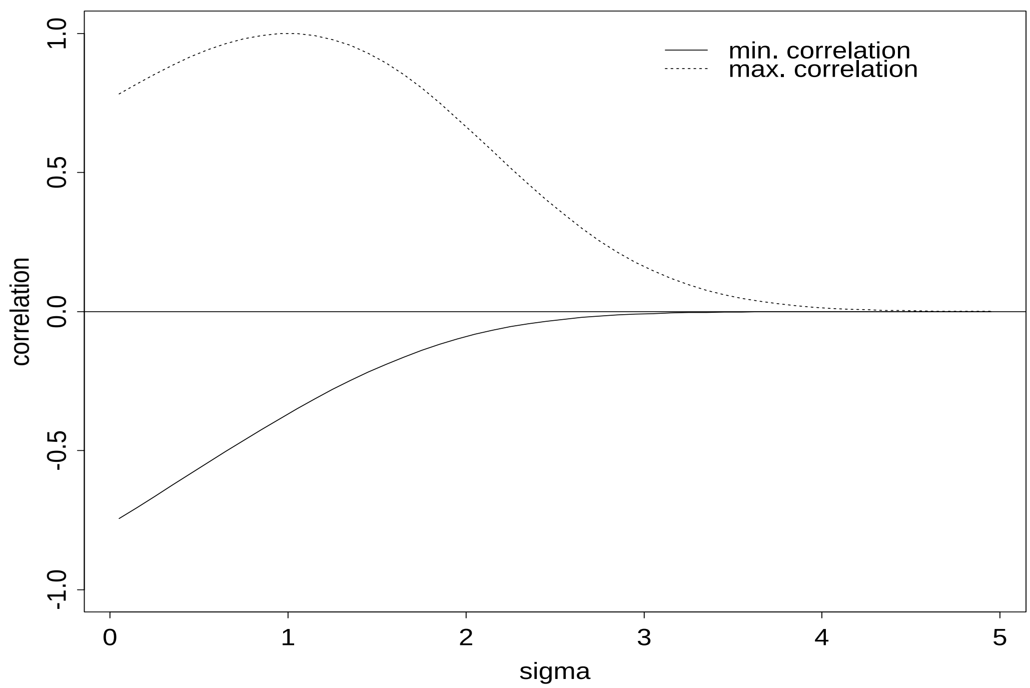

FIGURE 4. $\rho_{\min}$ and $\rho_{\max}$ graphed against σ.

This example shows it is possible to have a random vector $(X, Y)^t$ where the correlation is almost zero, even though X and Y are comonotonic or countermonotonic and thus have the strongest kind of dependence possible. This seems to contradict our intuition about probability and shows that small correlations cannot be interpreted as implying weak dependence between random variables.

Fallacy 3. The worst case VaR (quantile) for a linear portfolio $X + Y$ occurs when $\rho(X, Y)$ is maximal, i.e. X and Y are comonotonic

As we had discussed in Section 3.3 it is common to consider variance as a measure of risk in insurance and financial mathematics and, whilst it is true that the variance of a linear portfolio, $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2\rho(X,Y)\sigma(X)\sigma(Y)$, is maximal when the correlation is maximal, it is in general not correct to conclude that the Value-at-Risk is also maximal. For elliptical distributions it is true, but generally it is false.

Suppose two random variables X and Y have distribution functions F_1 and F_2 but that their dependence structure (copula) is unspecified. In the following theorem we give an upper bound for $\mathrm{VaR}_\alpha(X + Y)$.

Theorem 5. [*Makarov (1981) and Frank, Nelsen, and Schweizer (1987)*]

1. *For all* $z \in \mathbb{R}$,

$$\mathbb{P}[X + Y \le z] \ge \sup_{x+y=z} C_\ell(F_1(x), F_2(y)) =: \psi(z).$$

This bound is sharp in the following sense: Set $t = \psi(z-) = \lim_{u \to z-} \psi(u)$. *Then there exists a copula, which we denote by* $C^{(t)}$, *such that under the distribution with distribution function* $F(x, y) = C^{(t)}(F_1(x), F_2(y))$ *we have that* $\mathbb{P}[X + Y < z] = \psi(z-)$.[9]

2. *Let* $\psi^{-1}(\alpha) := \inf\{z \mid \psi(z) \geq \alpha\}$, $\alpha \in (0, 1)$, *be the generalized inverse of* ψ. *Then*
$$\psi^{-1}(\alpha) = \inf_{C_\ell(u,v)=\alpha} \{F_1^{-1}(u) + F_2^{-1}(v)\}.$$
3. *The following upper bound for Value-at-Risk holds:*
$$VaR_\alpha(X + Y) \leq \psi^{-1}(\alpha).$$
This bound is best-possible.

Proof. 1. For any $x, y \in \mathbb{R}$ with $x + y = z$ application of the lower Fréchet bound (12) yields
$$\mathbb{P}[X + Y \leq z] \geq \mathbb{P}[X \leq x, Y \leq y] \geq C_\ell(F_1(x), F_2(y)).$$
Taking the supremum over $x + y = z$ on the right hand side shows the first part of the claim.

The proof of the second part will be a sketch. We merely want to show how $C^{(t)}$ is chosen. For full mathematical details we refer to Frank, Nelsen, and Schweizer (1987). We restrict ourselves to continuous distribution functions F_1 and F_2. Since copulas are distributions with uniform marginals we transform the problem onto the unit square by defining $A = \{(F_1(x), F_2(y)) | x + y \geq z\}$ the boundary of which is $s = \{(F_1(x), F_2(y)) | x + y = z\}$. We need to find a copula $C^{(t)}$ such that $\iint_A dC^{(t)} = 1 - t$. Since F_1 and F_2 are continuous, we have that $\psi(z-) = \psi(z)$ and therefore $t \geq u + v - 1$ for all $(u, v) \in s$. Thus the line $u + v - 1 = t$ can be considered as a tangent to s and it becomes clear how one can choose $C^{(t)}$. $C^{(t)}$ belongs to the distribution which is uniform on the line segments $\overline{(0,0)(t,t)}$ and $\overline{(t,1)(1,t)}$. Therefore
$$C^{(t)}(u, v) = \begin{cases} \max\{u + v - 1, t\} & (u, v) \in [t, 1] \times [t, 1], \\ \min\{u, v\} & \text{otherwise.} \end{cases} \tag{21}$$
Since the set $\overline{(t,1)(1,t)} \subset A$ has probability mass $1 - t$ we have under $C^{(t)}$ that $\mathbb{P}[X + Y \geq z] = \iint_A dC^{(t)} \geq 1 - t$ and therefore $\mathbb{P}[X + Y < z] \leq t$. But since t is a lower bound for $\mathbb{P}[X + Y < z]$ it is necessary that $\mathbb{P}[X + Y < z] = t$.

2. This follows from the duality theorems in Frank and Schweizer (1979).
3. Let $\epsilon > 0$. Then we have
$$\mathbb{P}[X + Y \leq \psi^{-1}(\alpha) + \epsilon] \geq \psi(\psi^{-1}(\alpha) + \epsilon) \geq \alpha.$$
Taking the limit $\epsilon \to 0+$ this yields $\mathbb{P}[X + Y \leq \psi^{-1}(\alpha)] \geq \alpha$ and therefore $\mathrm{VaR}_\alpha(X + Y) \leq \psi^{-1}(\alpha)$. This upper bound cannot be improved. Again, take $\epsilon > 0$. Then if $(X, Y)^t$ has copula $C^{(\psi^{-1}(\alpha) - \epsilon/2)}$ one has
$$\begin{aligned} \mathbb{P}[X + Y \leq \psi^{-1}(\alpha) - \epsilon] &\leq \mathbb{P}[X + Y < \psi^{-1}(\alpha) - \epsilon/2] \\ &= \psi((\psi^{-1}(\alpha) - \epsilon/2)-) \leq \psi(\psi^{-1}(\alpha) - \epsilon/2) < \alpha \end{aligned}$$
and therefore $\mathrm{VaR}_\alpha(X + Y) > \psi^{-1}(\alpha) - \epsilon$.

□

[9] In general there is no copula such that $\mathbb{P}[X + Y \leq z] = \psi(z)$, not even if F_1 and F_2 are both continuous; see Nelsen (1999).

Remark 2. The results in Frank, Nelsen, and Schweizer (1987) are more general than Theorem 5 in this paper. Frank, Nelsen, and Schweizer (1987) give lower *and* upper bounds for $\mathbb{P}[L(X,Y) \le z]$ where $L(\cdot,\cdot)$ is continuous and increasing in each coordinate. Therefore a best-possible *lower* bound for $\mathrm{VaR}_\alpha(X+Y)$ also exists. Numerical evaluation methods of ψ^{-1} are described in Williamson and Downs (1990). These two authors also treat the case where we restrict attention to particular subsets of copulas. By considering the sets of copulas $\mathcal{D} = \{C | C(u,v) \ge u\,v,\ 0 \le u,v \le 1\}$, which has minimal copula $C_{ind}(u,v) = u\,v$, we can derive bounds of $\mathbb{P}[X+Y \le z]$ under positive dependence (PQD as defined in Definition 8). Multivariate generalizations of Theorem 5 can be found in Li, Scarsini, and Shaked (1996).

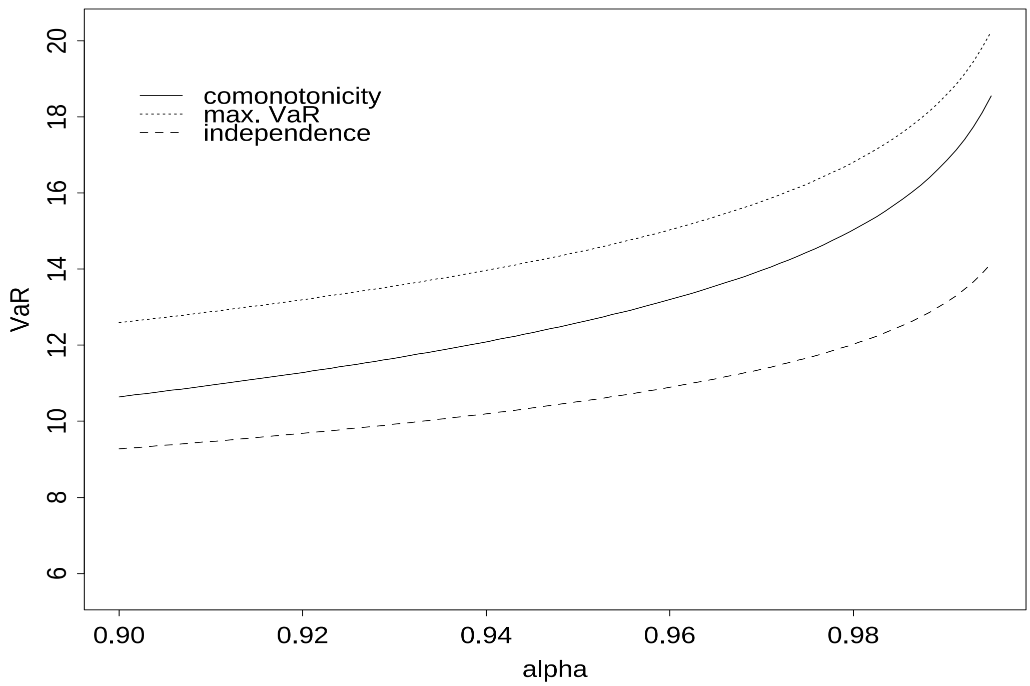

FIGURE 5. $\psi^{-1}(\alpha)$(max. VaR) graphed against α.

In Figure 5 the upper bound $\psi^{-1}(\alpha)$ is shown for $X \sim \mathrm{Gamma}(3,1)$ and $Y \sim \mathrm{Gamma}(3,1)$, for various values of α. Notice that $\psi^{-1}(\alpha)$ can easily be analytically computed analytically for this case since for α sufficiently large

$$\psi^{-1}(\alpha) = \inf_{u+v-1=\alpha} \{F_1^{-1}(u) + F_2^{-1}(v)\} = F_1^{-1}((\alpha+1)/2) + F_1^{-1}((\alpha+1)/2).$$

This is because $F_1 = F_2$ and the density of $\mathrm{Gamma}(3,1)$ is unimodal, see also Example 6. For comparative purposes $\mathrm{VaR}_\alpha(X+Y)$ is also shown for the case where X, Y are independent and the case where they are comonotonic. The latter is computed by addition of the univariate quantiles since under comonotonicity $\mathrm{VaR}_\alpha(X+Y) = \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$. [10] The example shows that for a fixed

[10] This is also true when X or Y do not have continuous distributions. Using Proposition 4.5 in Denneberg (1994) we deduce that for comonotonic random variables $X+Y = (u+v)(Z)$ where u and v are *continuous* increasing functions and $Z = X+Y$. Remark 1 then shows that $\mathrm{VaR}_\alpha(X+Y) = (u+v)(\mathrm{VaR}_\alpha(Z)) = u(\mathrm{VaR}_\alpha(Z)) + v(\mathrm{VaR}_\alpha(Z)) = \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$.

$\alpha \in (0,1)$ the maximal value of $\mathrm{VaR}_\alpha(X+Y)$ is considerably larger than the value obtained in the case of comonotonicity. This is not surprising since we know that VaR is not a subadditive risk measure (Artzner, Delbaen, Eber, and Heath 1999) and there are situations where $\mathrm{VaR}_\alpha(X+Y) > \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$. In a sense, the difference $\psi^{-1}(\alpha) - (\mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y))$ quantifies the amount by which VaR *fails* to be subadditive for particular marginals and a particular α. For a coherent risk measure ϱ, we must have that $\varrho(X+Y)$ attains its maximal value in the case of comonotonicity and that this value is $\varrho(X) + \varrho(Y)$ (Delbaen 1999). The fact that there are situations which are worse than comonotonicity as far as VaR is concerned, is another way of showing that VaR is not a coherent measure of risk.

Suppose we define a measure of diversification by

$$D = (\mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)) - \mathrm{VaR}_\alpha(X+Y),$$

the idea being that comonotonic risks are undiversifiable ($D = 0$) but that risks with weaker dependence *should* be diversifiable ($D > 0$). Unfortunately, Theorem 5 make it clear that we can always find distributions with linear correlation strictly less than the (comonotonic) maximal correlation (see Theorem 4) that give negative diversification ($D < 0$). This weakens standard diversification arguments, which say that "low correlation means high diversification". As an example Table 2 gives the numerical values of the correlations of the distributions yielding maximal $\mathrm{VaR}_\alpha(X+Y)$ for $X, Y \sim \mathrm{Gamma}(3,1)$.

α	0.25	0.5	0.75	0.8	0.85	0.9	0.95	0.99
ρ	-0.09	0.38	0.734	0.795	0.852	0.901	0.956	0.992

TABLE 2. Correlations of the distributions giving maximal $\mathrm{VaR}_\alpha(X+Y)$.

It might be supposed that VaR is in some sense asymptotically subadditive, so that negative diversification disappears as we let α tend to one, and comonotonicity becomes the worst case. The following two examples show that this is also wrong.

Example 6. The quotient $\mathrm{VaR}_\alpha(X+Y)/(\mathrm{VaR}_\alpha(X)+\mathrm{VaR}_\alpha(Y))$ can be made arbitrarily large. In general we do *not* have $\lim_{\alpha\to 1^-} \psi^{-1}(\alpha)/(\mathrm{VaR}_\alpha(X)+\mathrm{VaR}_\alpha(Y)) = 1$. To see this consider Pareto marginals $F_1(x) = F_2(x) = 1 - x^{-\beta}$, $x \geq 1$, where $\beta > 0$. We have to determine $\inf_{u+v-1=\alpha}\{F_1^{-1}(u) + F_2^{-1}(v)\}$. Since $F_1 = F_2$, the function

$$g : (\alpha, 1) \to \mathbb{R}_{\geq 0}, \ u \mapsto F_1^{-1}(u) + F_2^{-1}(\alpha + 1 - u)$$

is symmetrical with respect to $(\alpha+1)/2$. Since the Pareto density is decreasing, the function g is decreasing on $(\alpha, (\alpha+1)/2]$ and increasing on $[(\alpha+1)/2, 1)$; hence $g((\alpha+1)/2) = 2F_1^{-1}((\alpha+1)/2)$ is the minimum of g and $\psi^{-1}(\alpha) = 2F_1^{-1}((\alpha+1)/2)$. Therefore

$$\begin{aligned}\frac{\mathrm{VaR}_\alpha(X+Y)}{\mathrm{VaR}_\alpha(X)+\mathrm{VaR}_\alpha(Y)} &\leq \frac{\psi^{-1}(\alpha)}{\mathrm{VaR}_\alpha(X)+\mathrm{VaR}_\alpha(Y)} \\ &= \frac{F_1^{-1}((\alpha+1)/2)}{F_1^{-1}(\alpha)} = \frac{(1-\frac{\alpha+1}{2})^{-1/\beta}}{(1-\alpha)^{-1/\beta}} = 2^{1/\beta}.\end{aligned}$$

The upper bound $2^{1/\beta}$, which is irrespective of α, can be reached.

Example 7. Let X and Y be independent random variables with identical distribution $F_1(x) = 1 - x^{-1/2}$, $x \geq 1$. This distribution is extremely heavy-tailed with no finite mean. Consider the risks $X + Y$ and $2X$, the latter being the sum of comonotonic risks. We can calculate

$$\mathbb{P}[X + Y \leq z] = 1 - \frac{2\sqrt{z-1}}{z} < \mathbb{P}[2X \leq z],$$

for $z > 2$. It follows that

$$\mathrm{VaR}_\alpha(X + Y) > \mathrm{VaR}_\alpha(2X) = \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$$

for $\alpha \in (0, 1)$, so that, from the point of view of VaR, independence is worse than perfect dependence no matter how large we choose α. VaR is not sub-additive for this rather extreme choice of distribution and diversification arguments do not hold; one is better off taking one risk and doubling it than taking two independent risks. Diversifiability of two risks is not only dependent on their dependence structure but also on the choice of marginal distribution. In fact, for distributions with $F_1(x) = F_2(x) = 1 - x^{-\kappa}$, $\kappa > 0$, we do have asymptotic subadditivity in the case $\kappa > 1$. That means $\mathrm{VaR}_\alpha(X+Y) < \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$ if α large enough. To see this use lemma 1.3.1 of Embrechts, Mikosch, and Klüppelberg (1997) and the fact that $1 - F_1$ is regularly varying of index $-\kappa$ (for an introduction to regular variation theory see the appendix of the same reference).

6. Simulation of Random Vectors

There are various situations in practice where we might wish to simulate dependent random vectors $(X_1, \ldots, X_n)^t$. In finance we might wish to simulate the future development of the values of assets in a portfolio, where we know these assets to be dependent in some way. In insurance we might be interested in multiline products, where payouts are triggered by the occurrence of losses in one or more dependent business lines, and wish to simulate typical losses. The latter is particularly important within DFA. It is very tempting to approach the problem in the following way:

1. Estimate marginal distributions $F_1, \ldots, F_n$,
2. Estimate matrix of pairwise correlations $\rho_{ij} = \rho(X_i, X_j), i \neq j$,
3. Combine this information in some simulation procedure.

Unfortunately, we now know that step 3 represents an attempt to solve an ill-posed problem. There are two main dangers. Given the marginal distributions the correlation matrix is subject to certain restrictions. For example, each ρ_{ij} must lie in an interval $[\rho_{\min}(F_i, F_j), \rho_{\max}(F_i, F_j)]$ bounded by the minimal and maximal attainable correlations for marginals F_i and F_j. It is possible that the estimated correlations are not consistent with the estimated marginals so that no corresponding multivariate distribution for the random vector exists. In the case where a multivariate distribution exists it is often not unique.

The approach described above is highly questionable. Instead of considering marginals and correlations separately it would be more satisfactory to attempt a direct estimation of the multivariate distribution. It might also be sensible to consider whether the question of interest permits the estimation problem to be reduced to a one-dimensional one. For example, if we are really interested in the behaviour of the sum $X_1 + \cdots + X_n$ we might consider directly estimating the univariate distribution of the sum.

6.1. **Given marginals and linear correlations.** Suppose, however, we are required to construct a multivariate distribution F in $\mathbb{R}^n$ which is consistent with given marginals distributions $F_1, \dots, F_n$ and a linear correlation matrix ρ. We assume that ρ is a *proper* linear correlation matrix, by which we mean in the remainder of the paper that it is a symmetric, positive semi-definite matrix with $-1 \le \rho_{ij} \le 1,\ i,j = 1,\dots,n$ and $\rho_{ii} = 1,\ i = 1,\dots,n$. Such a matrix will always be the linear correlation matrix of *some* random vector in $\mathbb{R}^n$ but we must check it is compatible with the given marginals. Our problem is to find a multivariate distribution F so that if $(X_1,\dots,X_n)^t$ has distribution F the following conditions are satisfied:

$$X_i \sim F_i,\ i = 1,\dots,n, \tag{22}$$

$$\rho(X_i, X_j) = \rho_{ij},\ i,j = 1,\dots,n. \tag{23}$$

In the bivariate case, provided the prespecified correlation is attainable, the construction is simple and relies on the following.

Theorem 6. *Let F_1 and F_2 be two univariate distributions and $\rho_{\min}$ and $\rho_{\max}$ the corresponding minimal and maximal linear correlations. Let $\rho \in [\rho_{\min}, \rho_{\max}]$. Then the bivariate mixture distribution given by*

$$F(x_1,x_2) = \lambda F_\ell(x_1,x_2) + (1-\lambda)F_u(x_1,x_2), \tag{24}$$

where $\lambda = (\rho_{\max} - \rho)/(\rho_{\max} - \rho_{\min})$, $F_\ell(x_1,x_2) = \max\{F_1(x_1) + F_2(x_2) - 1, 0\}$ and $F_u(x_1,x_2) = \min\{F_1(x_1), F_2(x_2)\}$, has marginals F_1 and F_2 and linear correlation ρ.

Proof. Follows easily from Theorem 4. □

Remark 3. A similar result to the above holds for rank correlations when $\rho_{\min}$ and $\rho_{\max}$ are replaced by -1 and 1 respectively.

Remark 4. Also note that the mixture distribution is not the unique distribution satisfying our conditions. If $\rho \ge 0$ the distribution

$$F(x_1,x_2) = \lambda F_1(x_1)F_2(x_2) + (1-\lambda)F_u(x_1,x_2), \tag{25}$$

with $\lambda = (\rho_{\max} - \rho)/\rho_{\max}$ also has marginals F_1 and F_2 and correlation ρ. Many other mixture distributions (e.g. mixtures of distributions with Gumbel copulas) are possible.

Simulation of one random variate from the mixture distribution in Theorem 6 is achieved with the following algorithm:

1. Simulate U_1, U_2 independently from standard uniform distribution,
2. If $U_1 \le \lambda$ take $(X_1,X_2)^t = (F_1^{-1}(U_2), F_2^{-1}(1-U_2))^t$,
3. If $U_1 > \lambda$ take $(X_1,X_2)^t = (F_1^{-1}(U_2), F_2^{-1}(U_2))^t$.

Constructing a multivariate distribution in the case $n \ge 3$ is more difficult. For the existence of a solution it is certainly necessary that $\rho_{\min}(F_i,F_j) \le \rho_{ij} \le \rho_{\max}(F_i,F_j),\ i \ne j$, so that the pairwise constraints are satisfied. In the bivariate case this is sufficient for the existence of a solution to the problem described by (22) and (23), but in the general case it is not sufficient as the following example shows.

Example 8. Let F_1, F_2 and F_3 be Lognormal$(0,1)$ distributions. Suppose that ρ is such that ρ_{ij} is equal to the minimum attainable correlation for a pair of Lognormal$(0,1)$ random variables (≈ -0.368) if $i \ne j$ and $\rho_{ij} = 1$ if $i = j$. This is both a proper correlation matrix and a correlation matrix satisfying the pairwise constraints for lognormal random variables. However, since ρ_{12}, ρ_{13} and ρ_{23} are all minimum attainable correlations, Theorem 4 implies that X_1, X_2 and X_3 are pairwise countermonotonic random variables. Such a situation is unfortunately impossible as is is clear from the following proposition.

Proposition 4. *Let X, Y and Z be random variables with joint distribution F and continuous marginals F_1, F_2 and F_3.*

1. *If (X, Y) and (Y, Z) are comonotonic then (X, Z) is also comonotonic and $F(x, y, z) = \min\{F_1(x), F_2(y), F_3(z)\}$.*
2. *If (X, Y) is comonotonic and (Y, Z) is countermonotonic then (X, Z) is countermonotonic and $F(x, y, z) = \max\{0, \min\{F_1(x), F_2(y)\} + F_3(z) - 1\}$.*
3. *If (X, Y) and (Y, Z) are countermonotonic then (X, Z) is comonotonic and $F(x, y, z) = \max\{0, \min\{F_1(x), F_3(z)\} + F_2(y) - 1\}$*

Proof. We show only the first part of the proposition, the proofs of the other parts being similar. Using (14) we know that $Y = S(X)$ a.s. and $Z = T(Y)$ a.s. where $S, T : \mathbb{R} \to \mathbb{R}$ are increasing functions. It is clear that $Z = T \circ S(X)$ a.s. with $T \circ S$ increasing, so that X and Z are comonotonic. Now let $x, y, z \in \mathbb{R}$ and because also (X, Z) is comonotonic we may without loss of generality assume that $F_1(x) \leq F_2(y) \leq F_3(z)$. Assume for simplicity, but without loss of generality, that $Y = S(X)$ and $Z = T(Y)$ (i.e. ignore almost surely). It follows that $\{X \leq x\} \subseteq \{Y \leq y\}$ and $\{Y \leq y\} \subseteq \{Z \leq z\}$ so that

$$F(x, y, z) = \mathbb{P}[X \leq x] = F_1(x).$$

□

Example 9. Continuity of the marginals is an essential assumption in this proposition. It does not necessarily hold for discrete distributions as the next counterexample shows. Consider the multivariate two-point distributions given by

$$\begin{aligned} \mathbb{P}[(X, Y, Z)^t = (0, 0, 1)^t] &= 0.5, \\ \mathbb{P}[(X, Y, Z)^t = (1, 0, 0)^t] &= 0.5. \end{aligned}$$

(X, Y) and (Y, Z) are comonotonic but (X, Z) is countermonotonic.

The proposition permits us now to state a result concerning existence and uniqueness of solutions to our problem given by in the special case where random variables are either pairwise comonotonic or countermonotonic.

Theorem 7. [*Tiit (1996)*] *Let $F_1, \ldots, F_n$, $n \geq 3$, be continuous distributions and let ρ be a (proper) correlation matrix satisfying the following conditions for all $i \neq j$, $i \neq k$ and $j \neq k$:*

- $\rho_{ij} \in \{\rho_{\min}(F_i, F_j), \rho_{\max}(F_i, F_j)\}$,
- *If $\rho_{ij} = \rho_{\max}(F_i, F_j)$ and $\rho_{ik} = \rho_{\max}(F_i, F_k)$ then $\rho_{jk} = \rho_{\max}(F_j, F_k)$,*
- *If $\rho_{ij} = \rho_{\max}(F_i, F_j)$ and $\rho_{ik} = \rho_{\min}(F_i, F_k)$ then $\rho_{jk} = \rho_{\min}(F_j, F_k)$,*
- *If $\rho_{ij} = \rho_{\min}(F_i, F_j)$ and $\rho_{ik} = \rho_{\min}(F_i, F_k)$ then $\rho_{jk} = \rho_{\max}(F_j, F_k)$.*

Then there exists a unique distribution with marginals $F_1, \ldots, F_n$ and correlation matrix ρ. This distribution is known as an extremal distribution. In $\mathbb{R}^n$ there are 2^{n-1} possible extremal distributions.

Proof. Without loss of generality suppose

$$\rho_{1j} = \begin{cases} \rho_{\max}(F_1, F_j) & \text{for } 2 \leq j \leq m \leq n, \\ \rho_{\min}(F_1, F_j) & \text{for } m < j \leq n, \end{cases}$$

for some $2 \leq m \leq n$. The conditions of the theorem ensure that the pairwise relationship of any two margins is determined by their pairwise relationship to the first margin. The margins for which ρ_{1j} takes a maximal value form an equivalence class, as do the margins for which ρ_{1j} takes a minimal value. The joint distribution must be such that $(X_1, \ldots, X_m)$ are pairwise comonotonic, $(X_{m+1}, \ldots, X_n)$ are pairwise comonotonic, but two random variables taken from different groups are countermonotonic. Let $U \sim U(0,1)$. Then the random vector

$$(F_1^{-1}(U), F_2^{-1}(U), \ldots, F_m^{-1}(U), F_{m+1}^{-1}(1-U), \ldots, F_n^{-1}(1-U))^t,$$

has the required joint distribution. We use a similar argument to the Proposition 4 and assume, without loss of generality, that

$$\min_{1 \leq i \leq m} \{F_i(x_i)\} = F_1(x_1), \ \min_{m < i \leq n} \{F_i(x_i)\} = F_{m+1}(x_1).$$

It is clear that the distribution function is

$$\begin{aligned} F(x_1, \ldots, x_n) &= \mathbb{P}[X_1 \leq x_1, X_{m+1} \leq x_{m+1}] \\ &= \max\{0, \min_{1 \leq i \leq m} \{F_i(x_i)\} + \min_{m \leq i \leq n} \{F_i(x_i)\} - 1\}, \end{aligned}$$

which in addition shows uniqueness of distributions with pairwise extremal correlations. □

Let $G_j, j = 1, \ldots, 2^{n-1}$ be the extremal distributions with marginals $F_1, \ldots, F_n$ and correlation matrix ρ_j. Convex combinations

$$G = \sum_{j=1}^{2^{n-1}} \lambda_j G_j, \ \lambda_j \geq 0, \ \sum_{j=1}^{2^{n-1}} \lambda_j = 1,$$

also have the same marginals and correlation matrix given by $\rho = \sum_{j=1}^{2^{n-1}} \lambda_j \rho_j$. If we can decompose an arbitrary correlation matrix ρ in this way, then we can use a convex combination of extremal distributions to construct a distribution which solves our problem. In Tiit (1996) this idea is extended to quasi-extremal distributions. Quasi-extremal random vectors contain sub-vectors which are extremal as well as sub-vectors which are independent.

A disadvantage of the extremal (and quasi-extremal) distributions is the fact that they have no density, since they place all their mass on edges in $\mathbb{R}^n$. However, one can certainly think of practical examples where such distributions might still be highly relevant.

Example 10. Consider two portfolios of credit risks. In the first portfolio we have risks from country A, in the second risks from country B. Portfolio A has a profit-and-loss distribution F_1 and portfolio B a profit-and-loss distribution F_2. With probability p the results move in the same direction (comonotonicity); with probability $(1-p)$ they move in opposite directions (countermonotonicity). This situation can be modelled with the distribution

$$F(x_1, x_2) = p \cdot \min\{F_1(x_1), F_2(x_2)\} + (1-p) \cdot \max\{F_1(x_1) + F_2(x_2) - 1, 0\},$$

and of course generalized to more than two portfolios.

6.2. **Given marginals and Spearman's rank correlations.** This problem has been considered in Iman and Conover (1982) and their algorithm forms the basis of the @RISK computer program (Palisade 1997).

It is clear that a Spearman's rank correlation matrix is also a linear correlation matrix (Spearman's rank being defined as the linear correlation of ranks). It is not known to us whether a linear correlation matrix is necessarily a Spearman's rank correlation matrix. That is, given an arbitrary symmetric, positive semi-definite matrix with unit elements on the diagonal and off-diagonal elements in the interval $[-1, 1]$, can we necessarily find a random vector with continuous marginals for which this is the rank correlation matrix, or alternatively a multivariate distribution for which this is the linear correlation matrix of the copula? If we estimate a rank correlation matrix from data, is it guaranteed that the estimate is itself a rank correlation matrix? A necessary condition is certainly that the estimate is a linear correlation matrix, but we do not know if this is sufficient.

If the given matrix is a true rank correlation matrix, then the problem of the existence of a multivariate distribution with prescribed marginals is solved. The choice of marginals is in fact irrelevant and imposes no extra consistency conditions on the matrix.

Iman and Conover (1982) do not attempt to find a multivariate distribution which has exactly the given rank correlation matrix ρ. They simulate a standard multivariate normal variate $(X_1, \ldots, X_n)^t$ with linear correlation matrix ρ and then transform the marginals to obtain $(Y_1, \ldots, Y_n)^t = (F_1^{-1}(\Phi(X_i)), \ldots, F_n^{-1}(\Phi(X_n)))^t$.The rank correlation matrix of $\mathbf{Y}$ is identical to that of $\mathbf{X}$. Now because of (8)

$$\rho_S(Y_i, Y_j) = \rho_S(X_i, X_j) = \frac{6}{\pi} \arcsin \frac{\rho(X_i, X_j)}{2} \approx \rho(X_i, X_j),$$

and, in view of the bounds for the absolute error,

$$\left| \frac{6}{\pi} \arcsin \frac{\rho}{2} - \rho \right| \leq 0.0181, \;\; \rho \in [-1, 1],$$

and for the relative error,

$$\frac{\left| \frac{6}{\pi} \arcsin \frac{\rho}{2} - \rho \right|}{|\rho|} \leq \frac{\pi - 3}{\pi},$$

the rank correlation matrix of $\mathbf{Y}$ is very close to that which we desire. In the case when the given matrix belongs to an extremal distribution (i.e. comprises only elements 1 and -1) then the error disappears entirely and we have constructed the unique solution of our problem.

This suggests how we can find a sufficient condition for ρ to be a Spearman's rank correlation matrix and how, when this condition holds, we can construct a distribution that has the required marginals and exactly this rank correlation matrix. We define the matrix $\tilde{\rho}$ by

$$\tilde{\rho}_{ij} = 2 \sin \frac{\pi \rho_{ij}}{6}, \tag{26}$$

and check whether this is a proper linear correlation matrix. If so, then the vector $(Y_1, \ldots, Y_n)^t = (F_1^{-1}(\Phi(X_i)), \ldots, F_n^{-1}(\Phi(X_n)))^t$ has rank correlation matrix ρ, where $(X_1, \ldots, X_n)^t$ is a standard multivariate normal variate with linear correlation matrix $\tilde{\rho}$.

In summary, a necessary condition for ρ to be a rank correlation matrix is that it is a linear correlation matrix and a sufficient condition is that $\tilde{\rho}$ given by (26) is a linear correlation matrix. We are not aware at present of a necessary and sufficient condition.

A further problem with the approach described above is that we only ever construct distributions which have the dependency structure of the multivariate normal distribution. This dependency structure is limited as we observed in Example 2; it does not permit asymptotic dependency between random variables.

6.3. **Given marginals and copula.** In the case where marginal distributions $F_1, \dots, F_n$ and a copula $C(u_1, \dots, u_n)$ are specified a unique multivariate distribution with distribution function $F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$ satisfying these specifications can be found. The problem of simulating from this distribution is no longer the theoretical one of whether a solution exists, but rather the technical one of how to perform the simulation. We assume the copula is given in the form of a parametric function which the modeller has chosen; we do not consider the problem of how copulas might be estimated from data, which is certainly more difficult than estimating linear or rank correlations.

Once we have simulated a random vector $(U_1, \dots, U_n)^t$ from C, then the random vector $(F_1^{-1}(U_1), \dots, F_n^{-1}(U_n))^t$ has distribution F. We assume that efficient univariate simulation presents no problem and refer to Ripley (1987),Gentle (1998) or Devroye (1986) for more on this subject. The major technical difficulty lies now in simulating realisations from the copula.

Where possible a transformation method can be applied; that is, we make use of multivariate distributions with the required copula for which a multivariate simulation method is already known. For example, to simulate from the bivariate Gaussian copula it is trivial to simulate $(Z_1, Z_2)^t$ from the standard bivariate normal distribution with correlation ρ and then to transform the marginals with the univariate distribution function so that $(\Phi(Z_1), \Phi(Z_2))^t$ is distributed according to the desired copula. For the bivariate Gumbel copula a similar approach can be taken.

Example 11. Consider the Weibull distribution having survivor function $\overline{F}_1(x) = 1 - F_1(x) = \exp\left(-x^\beta\right)$ for $\beta > 0,\ x \geq 0$. If we apply the Gumbel copula to this survivor function (not to the distribution function) we get a bivariate distribution with Weibull marginals and survivor function

$$\overline{F}(z_1, z_2) = \mathbb{P}[Z_1 > z_1, Z_2 > z_2] = C(\overline{F}_1(z_1), \overline{F}_1(z_2)) = \exp\left[-(z_1 + z_2)^\beta\right].$$

Lee (1979) describes a method for simulating from this distribution. We take $(Z_1, Z_2)^t = (US^{1/\beta}, (1-U)S^{1/\beta})^t$ where U is standard uniform and S is a mixture of Gamma distributions with density $h(s) = (1 - \beta + \beta s)\exp(-s)$ for $s \geq 0$. Then $(\overline{F}_1(Z_1), \overline{F}_1(Z_2))^t$ will have the desired copula distribution.

Where the transformation method cannot easily be applied, another possible method involves recursive simulation using univariate conditional distributions. We consider the general case $n > 2$ and introduce the notation

$$C_i(u_1, \dots, u_i) = C(u_1, \dots, u_i, 1, \dots, 1),\ i = 2, \dots, n-1$$

to represent i–dimensional marginal distributions of $C(u_1, \dots, u_n)$. We write $C_1(u_1) = u_1$ and $C_n(u_1, \dots, u_n) = C(u_1, \dots, u_n)$. Let us suppose now that $(U_1, \dots, U_n)^t \sim C$; the conditional distribution of U_i given the values of the first $i-1$ components of $(U_1, \dots, U_n)^t$ can be written in terms of derivatives and densities of the i–dimensional marginals

$$\begin{aligned} C_i(u_i \mid u_1, \ldots, u_{i-1}) &= \mathbb{P}[U_i \le u_i \mid U_1 = u_1, \ldots, U_{i-1} = u_{i-1}] \\ &= \frac{\partial^{i-1} C_i(u_1, \ldots, u_i)}{\partial u_1 \cdots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C_{i-1}(u_1, \ldots, u_{i-1})}{\partial u_1 \cdots \partial u_{i-1}}, \end{aligned}$$

provided both numerator and denominator exist. This suggests that in the case where we can calculate these conditional distributions we use the algorithm:

- Simulate a value u_1 from $U(0,1)$,
- Simulate a value u_2 from $C_2(u_2 \mid u_1)$,
- Continue in this way,
- Simulate a value u_n from $C_n(u_n \mid u_1, \ldots, u_{n-1})$.

To simulate a value from $C_i(u_i \mid u_1, \ldots, u_{i-1})$ we would in general simulate u from $U(0,1)$ and then calculate $C_i^{-1}(u \mid u_1, \ldots, u_{i-1})$, if necessary by numerical root finding.

7. Conclusions

In this paper we have shown some of the problems that can arise when the concept of linear correlation is used with non-elliptical multivariate distributions. In the world of elliptical distributions correlation is a natural and elegant summary of dependence, which lends itself to algebraic manipulation and the standard approaches of risk management dating back to Markowitz. In the non-elliptical world our intuition about correlation breaks down and leads to a number of fallacies. The first aim of this paper has been to suggest that practitioners of risk management must be aware of these pitfalls and must appreciate that a deeper understanding of dependency is needed to model the risks of the real world.

The second main aim of this paper has been to address the problem of simulating dependent data with given marginal distributions. This question arises naturally when one contemplates a Monte Carlo approach to determining the risk capital required to cover dependent risks. We have shown that the ideal situation is when the multivariate dependency structure (in the form of a copula) is fully specified by the modeller. Failing this, it is preferable to be given a matrix of rank correlations than a matrix of linear correlations, since rank correlations are defined at a copula level, and we need not worry about their consistency with the chosen marginals. Both correlations are, however, scalar-valued dependence measures and if there is a multivariate distribution which solves the simulation problem, it will not be the unique solution. The example of the Introduction showed that two distributions with the same correlation can have qualitatively very different dependency structures and, ideally, we should consider the whole dependence structure which seems appropriate for the risks we wish to model.

Achnowledgements. We would like to thank Eduardo Vilela and Rüdiger Frey for fruitful discussions and Roger Kaufmann for careful reading.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): "Coherent measures of risk," To appear in Mathematical Finance.

Boyer, B. H., M. S. Gibson, and M. Loretan (1999): "Pitfalls in tests for changes in correlations," International Finance Papers Number 597, Board of Governors of the Federal Reserve System.

CAMPBELL, J. Y., A. W. LO, and A. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

CAS (1997): *CAS Forum Summer 1997: DFA Call Papers*, Casualty Actuarial Society.

DELBAEN, F. (1999): "Coherent risk measures on general probability spaces," Preprint ETH Zürich.

DENNEBERG, D. (1994): *Non-additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht.

DEVROYE, L. (1986): *Non-uniform Random Variate Generation*. Springer, New York.

DHAENE, J., and M. J. GOOVAERTS (1996): "Dependency of risks and stop-loss order," *ASTIN Bulletin*, 26(2), 201–212.

EMBRECHTS, P., T. MIKOSCH, and C. KLÜPPELBERG (1997): *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

FANG, K.-T., S. KOTZ, and K.-W. NG (1987): *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.

FRANK, M., R. B. NELSEN, and B. SCHWEIZER (1987): "Best–possible bounds for the distribution of a sum — a problem of Kolmogorov," *Probability Theory and Related Fields*, 74, 199–211.

FRANK, M., and B. SCHWEIZER (1979): "On the duality of generalized infimal and supremal convolutions," *Rendiconti di Matematica*, 12(1), 1–23.

FRÉCHET, M. (1957): "Les tableaux de corrélation dont les marges sont données," *Annales de l'Université de Lyon, Sciences Mathématiques et Astronomie, Série A*, 4, 13–31.

GALAMBOS, J. (1987): *The Asymptotic Theory of Extreme Order Statistics*. Kreiger Publishing Co., Melbourne, FL.

GENTLE, J. E. (1998): *Random number generation and Monte Carlo methods*. Springer, New York.

GIBBONS, J. D. (1988): *Nonparametric Statistical Inference*. Dekker, New York.

HARLOW, W. (1991): "Asset allocation in a downside-risk framework," *Financial Analysts Journal*, 47(5), 28–40.

HÖFFDING, W. (1940): "Massstabinvariante Korrelationstheorie," *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin*, 5, 181–233.

HUTCHINSON, T. P., and C. D. LAI (1990): *Continuous Bivariate Distributions, Emphasizing Applications*. Rumsby Scientific Publishing, Adelaide.

IMAN, R. L., and W. CONOVER (1982): "A distribution–free approach to inducing rank correlation among input variables," *Communications in Statistics — Simulation and Computation*, 11, 311–334.

JOAG-DEV, K. (1984): "Measures of Dependence," in *Handbook of Statistics*, ed. by P. R. Krishnaiah, vol. 4, pp. 79–88. North–Holland/Elsevier, New York.

JOE, H. (1997): *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

KELKER, D. (1970): "Distribution theory of spherical distributions and a location–scale parameter generalization," *Sankhiă A*, 32, 419–430.

KENDALL, M., and A. STUART (1979): *Handbook of Statistics*. Griffin & Company, London.

KIMELDORF, G., and A. R. SAMPSON (1978): "Monotone dependence," *Annals of Statistics*, 6, 895–903.

LEE, L. (1979): "Multivariate distributions having Weibull properties," *Journal of Multivariate Analysis*, 9, 267–277.

LI, H., M. SCARSINI, and M. SHAKED (1996): "Bounds for the distribution of a multivariate Sum," in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendorff, B. Schweizer, and M. D. Taylor, pp. 198–212, Hayward, CA. Institute of Mathematical Statistics.

LOWE, S. P., and J. N. STANARD (1997): "An integrated dynamic financial analysis and decision system for a property catastrophe insurer," *ASTIN Bulletin*, 27(2), 339–371.

MAKAROV, G. (1981): "Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed," *Theory of Probability and its Applications*, 26, 803–806.

MARSHALL, A. W. (1996): "Copulas, marginals and joint distributions," in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendorff, B. Schweizer, and M. D. Taylor, pp. 213–222, Hayward, CA. Institute of Mathematical Statistics.

MIKUSINSKI, P., H. SHERWOOD, and M. TAYLOR (1992): "The Fréchet bounds revisited," *Real Analysis Exchange*, 17, 759–764.

Müller, A., and N. Bäuerle (1998): "Modelling and comparing dependencies in multivariate risk portfolios," *ASTIN Bulletin*, 28(1), 59–76.

Nelsen, R. B. (1999): *An Introduction to Copulas.* Springer, New York.

Palisade (1997): *Manual for @RISK.* Palisade Corporation, Newfield, NY.

Resnick, S. I. (1987): *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

Ripley, B. D. (1987): *Stochastic Simulation.* Wiley, New York.

Schweizer, B., and A. Sklar (1983): *Probabilistic Metric Spaces.* North–Holland/Elsevier, New York.

Schweizer, B., and E. Wolff (1981): "On nonparametric measures of dependence for random variables," *Annals of Statistics*, 9, 879–885.

Sibuya, M. (1961): "Bivariate extreme statistics," *Annals of Mathematical Statistics*, 11, 195–210.

Tiit, E. (1996): "Mixtures of multivariate quasi–extremal distributions having given marginals," in *Distributions with Fixed Marginals and Related Topics*, ed. by L. Rüschendorff, B. Schweizer, and M. D. Taylor, pp. 337–357, Hayward, CA. Institute of Mathematical Statistics.

Tjøstheim, D. (1996): "Measures of dependence and tests of independence," *Statistics*, 28, 249–284.

Wang, S. (1997): "Aggregation of correlated Risk Portfolios: Models and Algorithms," Preprint Casualty Actuarial Society (CAS).

Wang, S., and J. Dhaene (1998): "Comonoticity, correlation order and premium principles," *Insurance Mathematics and Economics*, 22, 235–242.

Williamson, R. C., and T. Downs (1990): "Probabilistic Arithmetic: Numerical methods for calculating convolutions and dependency bounds," *Journal of Approximate Reasoning*, 4, 89–158.

Yaari, M. (1987): "The dual theory of choice under risk," *Econometrica*, 55, 95–115.

(Paul Embrechts) Departement Mathematik, ETHZ, CH-8092 Zürich, Switzerland
E-mail address: embrechts@math.ethz.ch

(Alexander McNeil) Departement Mathematik, ETHZ, CH-8092 Zürich, Switzerland
E-mail address: mcneil@math.ethz.ch

(Daniel Straumann) Departement Mathematik, ETHZ, CH-8092 Zürich, Switzerland
E-mail address: strauman@math.ethz.ch